POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

Indeks 363189
ISSN 0867-6747

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

TOM 41 — ZESZYT 1

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1995

Kwartalnik jest wydawany z funduszu Polskiej Akademii Nauk i dofinansowany z instytucji:

— Komitet Badań Naukowych
— Politechnika Warszawska, Wydział Elektroniki i Technik Informacyjnych
— Politechnika Gdańska, Wydział Elektryczny
— Politechnika Szczecińska, Instytut Elektroniki i Informatyki
— Ministerstwo Łączności
— 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 41 — ZESZYT 1

WYDAWNICTWO NAUKOWE PWN

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. 628 89 81 21007-737

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

WYDAWNICTWO NAUKOWE PWN
Warszawa, ul. Miodowa 10

Ark. wyd. 11,25 Ark. druk. 9.0 + wkł. kred. Podpisano do druku w kwietniu 1995 r.

Papier offsetowy kl. III 70 g. B-1 Druk ukończono w maju 1995 r.

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

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 1. ss. 3–4

Z głębokim ubolewaniem zawiadamiamy, że 17 grudnia 1994 r. w wieku 91 lat zmarł

Profesor Witold Nowicki

Profesor Witold Nowicki był założycielem i wieloletnim redaktorem naczelnym (1955–1985) Rozpraw Elektrotechnicznych — Kwartalnika Polskiej Akademii Nauk, którego kontynuatorem jest obecnie Kwartalnik Elektroniki i Telekomunikacji.

W okresie przedwojennym Witold Nowicki był pracownikiem dydaktycznym w Politechnice Warszawskiej i Lwowskiej oraz nauczycielem w Państwowej Szkole Teletechnicznej w Warszawie. Był także pracownikiem naukowym w Laboratorium Teletechnicznym przy Ministerstwie Poczt i Telegrafów a następnie w Państwowym Instytucie Telekomunikacji.

W czasie wojny kontynuował swoją pracę nauczycielską pracując jednocześnie w Spółdzielni „Grupa Techniczna". Prowadził działalność konspiracyjną. Został aresztowany przez Gestapo i po pobycie na Pawiaku był osadzony w obozie koncentracyjnym w Sztuthofie.

Po odzyskaniu niepodległości objął stanowisko wicedyrektora Państwowego Instytutu Telekomunikacyjnego, które zajmował do roku 1949. Jednocześnie w roku 1945 został organizatorem i kierownikiem nowoutworzonej Katedry Przenoszenia Przewodowego Politechniki Warszawskiej, przemianowanej później na Katedrę Teletransmisji Przewodowej. Od 1947 r. — jako profesor nadzwyczajny — kierował tą Katedrą do roku 1970, kiedy została ona włączona do uczelnianego Instytutu Teleelektroniki nazwanego później Instytutem Telekomunikacji. Był dyrektorem tego instytutu do roku 1973, to jest do czasu otrzymania statusu profesora emerytowanego.

Profesor zwyczajny doktor inżynier Witold Nowicki był wybitnym organizatorem badań naukowych, znakomitym dydaktykiem, nauczycielem i wychowawcą wielu pokoleń specjalistów telekomunikacji, w tym 16 profesorów i docentów. Był autorem ponad 160 publikacji i skryptów, licznych książek i wielu artykułów zamieszczanych w czasopismach naukowych, technicznych oraz w prasie codziennej. Publikacje te dotyczyły czterech głównych obszarów zainteresowań Profesora: podstaw teoretycznych i technicznych teletransmisji, zagadnień społecznych i gospodarczych telekomunikacji, problemów rozwoju małoseryjnego przemysłu precyzyjnego, w szczególności

telekomunikacyjnego, oraz zagadnień językoznawczych — w tym terminologicznych, ze szczególnym uwzględnieniem terminologii telekomunikacji.

Profesor Witold Nowicki był także bardzo aktywny społecznie, działał w Stowarzyszeniu Elektryków Polskich, które wyróżniło Go tytułem Członka Honorowego. Był członkiem rzeczywistym Towarzystwa Naukowego Warszawskiego oraz brał udział w wielu komitetach naukowych i technicznych, radach naukowych i zespołach redakcyjnych.

Działalność Profesora była wysoko oceniana. Został wyróżniony wieloma nagrodami i otrzymał liczne odznaczenia państwowe, resortowe i społeczne, w tym Krzyż Komandorski Orderu Odrodzenia Polski i Medal im. Mikołaja Kopernika przyznany przez Polską Akademię Nauk.

Nad grobem Profesora Witolda Nowickiego na Warszawskich Powązkach zgromadziło się wielu Jego przyjaciół, uczniów i współpracowników. Żegnali twórcę polskiej teletransmisji, współtwórcę polskiej telekomunikacji, naukowca, technika, działacza gospodarczego i humanistę.

Żegnamy także twórcę i wieloletniego redaktora naszego kwartalnika.

Cześć Jego pamięci!

Rada Redakcyjna

Komitet Redakcyjny

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 1. ss. 5–29

Identyfikacja praktyczna modeli transmitancyjnych na podstawie początkowej fazy odpowiedzi skokowej

STANISŁAW SKOCZOWSKI

Instytut Automatyki Przemysłowej, Politechnika Szczecińska

Otrzymano 1995.01.02

Autoryzowano do druku 1995.02.14

Przedstawiono uproszczoną metodę identyfikacji modeli transmitacyjnych opartą na dyskretnych pomiarach początkowej fazy odpowiedzi skokowej procesu. Przy założeniu liniowości i znajomości opóźniania, metoda pozwala na oszacowanie rzędu procesu i w konsekwencji także pozostałych parametrów transmitacji na podstawie dyskretnych wartości pary próbek. W ogólności, gdy nieznane są rzędy licznika (m) i mianownika (n) transmitacji, metoda pozwala na identyfikację rzędu „zredukowanego" r=n−m. Rozważono szczegółowo szacowanie parametrów modeli wieloineracyjnych, szacując także „rozrzut" stałych czasowych. Wskazano możliwość identyfikacji parametrów modelu z jednokrotnym różniczkowaniem rzędu (n) oraz modelu oscylacyjnego. Przytoczono wyniki badań eksperynemtalnych oraz symulacji.

Słowa kluczowe: modele transmitancyjne, modele wieloineracyjune, obiekty elektrotermiczne.

1. WPROWADZENIE

Celem pracy jest wykazanie możliwości praktycznej identyfikacji modeli transmitancyjnych za pomocą efektywnej metody [6] pozwalającej już w pierwszej fazie odpowiedzi skokowej obiektu rzeczywistego, w czasie od zera do osiągnięcia punktu przegięcia, oszacować na podstawie dyskretnych próbek odpowiedzi rząd i pozostałe parametry modelu, przy założeniu liniowości obiektu. W rozważaniach teoretycznych pomija się wpływ zakłóceń nałożonych w rzeczywistości na mierzoną odpowiedź skokową. Proponuje się jednak [8] zwielokrotnienie pomiarów i obliczeń oraz uśrednienie wyników tych obliczeń, co jak wykazują badania symulacyjne daje dobre rezultaty w przypadku obecności zakłóceń. Przytacza się wyniki symulacji oraz badań eksperymentalnych rzeczywistych obiektów elektrotermicznych.

Jakkolwiek rozwój współczesnych metod identyfikacji idzie w kierunku analizy statystycznej procesów stochastycznych i tu poczyniono olbrzymie postępy, to jednak nie oznacza, że metody deterministyczne jak np. oparte na badaniu odpowiedzi skokowej powinny zostać zapomniane, zarzucone i nie mogą być udoskonalane. W tym przekonaniu utwierdza fakt stosowania w rozwiązaniach konstrukcyjnych regulatorów przemysłowych samonastrajających właśnie metod deterministycznych identyfikacji uproszczonej [12]. Takie uproszczone metody mogą mieć zastosowania praktycznie dla celów inżynierskich [3] i mogą być wykorzystywane w konstrukcjach mikroprocesorowych regulatorów przemysłowych.

2. MODELE TRANSMITANCYJNE

Przy założeniu liniowości lub po linearyzacji, można wyróżnić szereg typowych praktycznych modeli procesów stabilnych, reprezentowanych w ogólności przez transmitancję

$$G(s)=\frac{b_m s^m+\ldots+b_1 s+b_0}{s^n+a_{n-1}s^{n-1}+\ldots+a_1 s+a_0}\, e^{-s\tau}. \tag{1}$$

Przy tym, definiuje się pojęcie [13] średniej stałej czasowej $\bar{T}$, która spełnia relację

$$\frac{1}{\bar{T}}=-\bar{p}=\frac{-(p_1+\ldots+p_n)}{n}=\frac{-(\delta_1+\ldots+\delta_n)}{n}, \tag{2}$$

gdzie $\delta_j=R_e[p_j]$, zaś p_j są biegunami transmitancji (1). Przy tym, dla obiektu stabilnego, ważne są zależności

$$\bar{T}=\frac{n}{a_{n-1}}=\frac{n}{\sum_{i=1}^{n}(-p_i)}=\frac{n}{\sum_{i=1}^{n}\frac{1}{T_i}} \tag{3}$$

$$\frac{1}{T_{\min}}>\frac{1}{\bar{T}}>\frac{1}{T_{\max}} \tag{4}$$

$$\left.\begin{aligned} T_{\max}&=\max\left[-\frac{1}{\delta_1},\ldots,-\frac{1}{\delta_n}\right]\\ T_{\min}&=\min\left[-\frac{1}{\delta_1},\ldots,-\frac{1}{\delta_n}\right]\end{aligned}\right\} \tag{5}$$

Definiuje się też współczynnik rozrzutu λ, w postaci

$$\lambda=\sqrt[n-1]{\frac{T_{\min}}{T_{\max}}},\qquad 0<\lambda\leqslant 1 \tag{6}$$

co oznacza, że poszczególne stałe czasowe T_i są „rozpięte" pomiędzy T_{min} i T_{max} wg zależności

$$T_i = \lambda^{i-1} T_{max}, \quad i = 1, 2, ..., n. \tag{7}$$

Zauważyć należy, że w rzeczywistości, poszczególne stałe czasowe T_i w transmitancji (1) mogą być zupełnie inaczej niż wg (7) rozłożone, co istotnie komplikuje ten problem w przypadku rzędów $n > 2$. Podkreślić też trzeba, że operując pojęciem średniej stałej czasowej $\bar{T}$ oraz λ, narzuca się z góry położenie biegunów transmitancji (1). W szczególności, dla procesów wieloinercyjnych

$$p_i = -\frac{1}{T_1 \lambda^{i-1}}, \quad i = 1, 2, ..., n, \quad \text{gdzie } T_1 = T_{max}.$$

Zakładając w dalszych rozważaniach, że opóźnienie τ w transmitancji (1) jest znane, można zaniedbując je wyróżnić 5 typowych modeli transmitancyjnych, „wywodzących się" z (1).

Istnieje szeroka klasa obiektów wieloinercyjnych pozbawionych z natury rzeczy różniczkowania, które mogą być w ogólności charakteryzowane transmitancją

$$M(s) = \frac{k}{\prod_{i=1}^{n} (1 + sT_i)}. \tag{8}$$

Najprostszą popularną aproksymacją (8), rozpowszechnioną dla celów inżynierskich, jest trójparametrowy model Küpfmüllera

$$K(s) = \frac{k e^{-s\tau_z}}{1 + sT_z}. \tag{9}$$

Model (9) ignoruje zjawisko wieloinercyjności, wysoki rząd procesu zastępuje rzędem $n = 1$, wprowadza sztucznie zastępcze opóźnienia τ_z i zastępczą stałą czasową T_z.

Sprawia to, że model ten może obowiązywać tylko w bardzo ograniczonym zakresie — np. nie może to być jednocześnie model dla niskich i wysokich częstotliwości. Te mankamenty nie występują w innym, porównywalnie prostym modelu Strejca

$$S(s) = \frac{k}{(1 + s\bar{T})^n}, \tag{10}$$

który jest tym lepszym uproszczeniem transmitancji (8), im mniejszy jest rozrzut stałych czasowych, to jest im λ jest bliższe jedności. Dla $T_i = \bar{T}$, $i = 1...n$, modele te stają się identyczne.

W przypadku procesów, w których występuje różniczkowanie, najprostszym i z praktycznych względów najważniejszym będzie model transmitancyjny z jednokrotnym różniczkowaniem,

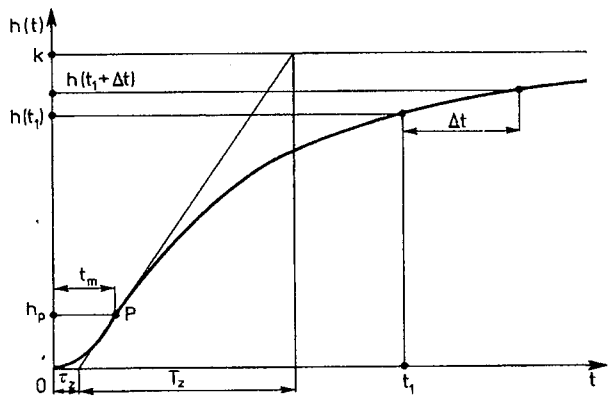
$$D(s)=\frac{b_1 s+b_0}{(1+s\bar{T})^n}, \tag{11}$$

a procesy oscylacyjne, reprezentować może model

$$O(s)=\frac{k\omega_0^2}{s^2+2\beta\omega_0 s+\omega_0^2}. \tag{12}$$

Zauważymy, że w przypadku istnienia znanego opóźnienia τ, modele (8), (9), (10), (11) i (12) mogą być prosto skorygowane przez uwzględnienie w transmitancji dodatkowego członu $e^{-s\tau}$.

Z uwagi na popularność modelu (9) warto zwrócić uwagę na relacje parametrów modeli (9) i (10) [10]. Na Rys. 1 przedstawiono typową odpowiedź skokową procesu wieloinercyjnego.



Rys. 1. Typowa odpowiedź skokowa procesu wieloinercyjnego. Styczna w punkcie przegięcia P wyznacza τ_z i T_z modelu (9)

Przyjmując *a priori* dla procesu z Rys. 1 model (9), z góry przesądza się o tym, że zastępcza stała czasowa T_z jest duża, opóźnienie zastępcze τ_z jest duże i że zastępuje ono zupełnie inaczej w rzeczywistości przebiegający początek odpowiedzi skokowej $h(t)$. Gdyby pokazanej na Rys. 1 odpowiedzi $h(t)$ przypisać model (10), to na podstawie [4] będzie w przybliżeniu

$$n_0 \approx 1+2\pi\left(\frac{t_m}{T_z}\right)^2 \tag{13}$$

zaokrąglając n_0 do liczby całkowitej n, będzie

$$\left.\begin{array}{ll} \text{dla} & n=1 \quad \bar{T}=T_z \\ \text{dla} & n>1 \quad \bar{T}\approx\dfrac{T_z}{\sqrt{2\pi(n-1)}} \end{array}\right\}. \tag{14}$$

Z Rys. 1 wynikają proste zależności

$$\tau_z \approx t_m - \frac{h_p}{\left(\frac{\Delta h}{\Delta t}\right)_{\max}} \tag{15}$$

$$\left(\frac{\Delta h}{\Delta t}\right)_{\max} \approx \frac{k}{T_z}. \tag{16}$$

Dla modelu (10), wartość h_p jest określona w postaci

$$h_p = k\left\{1 - e^{-(n-1)}\left[1 + (n-1) + \ldots + \frac{(n-1)^{n-1}}{(n-1)!}\right]\right\}. \tag{17}$$

Po uwzględnieniu w (15), zależności (13), (16) i (17), będzie w przybliżeniu

$$\tau_z \approx T_z\left\{\sqrt{\frac{n-1}{2\pi}} - 1 + e^{-(n-1)}\left[1 + (n-1) + \ldots + \frac{(n-1)^{n-1}}{(n-1)!}\right]\right\}. \tag{18}$$

Wzory (13), (14) i (18) wyrażają przybliżone relacje między modelem (9) i modelem (10). Dane tych relacji, dla rzędów $n=1$, 2, 3 i 4 pokazano w Tablicy 1, gdzie pod kreską podano też wartości dokładne, uzyskane na drodze obliczeń numerycznych [1].

Tablica I

Relacje między parametrami modeli (9) i (10), w postaci stosunku $\bar{T}/T_z$ i τ_z/T_z dla różnych wartości rzędu

	$n=1$	$n=2$	$n=3$	$n=4$
$\frac{\bar{T}}{T_z}$	$\frac{1}{1}$	$\frac{0{,}398}{0{,}368}$	$\frac{0{,}282}{0{,}271}$	$\frac{0{,}230}{0{,}225}$
$\frac{\tau_z}{T_z}$	$\frac{0}{0}$	$\frac{0{,}135}{0{,}104}$	$\frac{0{,}240}{0{,}218}$	$\frac{0{,}340}{0{,}319}$

Należy podkreślić, że w przypadku dużego rozrzutu stałych czasowych w (8), silnie maleje $\frac{\tau_z}{T_z}$ oraz określony na podstawie (13) rząd n_0. Jednocześnie obniża się punkt przegięcia $P-c$ i maleje wartość h_p i czas t_m. Duży rozrzut stałych czasowych T_i, $i=1,\ldots,n$, powoduje silne zmniejszenie $\bar{T}$ modelu (10), co ilustruje przykładowo Tablica II dla rzędu $n=3$.

Tablica II

Zależność średniej stałej czasowej $\bar{T}$ od rozrzutu λ dla modelu (10) przy $n=3$

$T_1/T_2/T_3$	10/10/10	15/10/5	20/9/1	30/3/0,3
$\bar{T}$	10	8,18	2,58	0,81
λ	1	0,577	0,224	0,1

3. SZACOWANIE PARAMETRÓW MODELI

Odpowiedź skokowa modelu liniowego o transmitancji (1) jest odwrotną transformatą Laplace'a jak niżej

$$h(t)=L^{-1}\left\{\frac{G(s)}{s}\right\}. \tag{19}$$

Wykonując [13] dzielenie licznika transmitancji (1) przez mianownik, uzyskuje się w rezultacie szereg, którego wyrazy po podzieleniu przez s poddaje się transformacji odwrotnej. W efekcie uzyskuje się szereg. Zaniedbując opóźnienie τ, będzie

$$h(t)=b_m\frac{t^r}{r!}+(b_{m-1}-b_m a_{n-1})\frac{t^{r+1}}{(r+1)!}+[b_{m-2}-b_{m-1}a_{n-1}+b_m(a_{n-1}^2-a_{n-2})]\frac{t^{r+2}}{(r+2)!}+\ldots \tag{20}$$

gdzie $r=n-m$ nazwiemy rzędem zredukowanym modelu (1). Szereg ten, dla niewielkich wartości argumentu t jest silnie zbieżny.

Biorąc stosunek dwu dyskretnych wartości odpowiedzi skokowej i przyjmując dla uproszczenia $t_1=\Delta$, $t_2=2\Delta$, gdzie Δ jest okresem próbkowania mierzonej odpowiedzi, uzyskuje się przybliżoną zależność [7]

$$\frac{h(2\Delta)}{h(\Delta)}\approx 2^r\Phi, \tag{21}$$

gdzie funkcja Φ jest określona

$$\Phi=\frac{1-\left[1-\frac{b_{m-1}}{b_m}\left(\frac{\bar{T}}{n}\right)\right]\frac{2n}{r+1}\left(\frac{\Delta}{\bar{T}}\right)+\left[1-\frac{b_{m-1}}{b_m}\left(\frac{\bar{T}}{n}\right)+\left(\frac{b_{m-2}}{b_m}-a_{n-2}\right)\left(\frac{\bar{T}}{n}\right)^2\right]\frac{4n^2}{(r+1)(r+2)}\left(\frac{\Delta}{\bar{T}}\right)^2+\ldots}{1-\left[1-\frac{b_{m-1}}{b_m}\left(\frac{\bar{T}}{n}\right)\right]\frac{n}{r+1}\left(\frac{\Delta}{\bar{T}}\right)+\left[1-\frac{b_{m-1}}{b_m}\left(\frac{\bar{T}}{n}\right)+\left(\frac{b_{m-2}}{b_m}-a_{n-2}\right)\left(\frac{\bar{T}}{n}\right)^2\right]\frac{n^2}{(r+1)(r+2)}\left(\frac{\Delta}{\bar{T}}\right)^2+\ldots} \tag{22}$$

a przybliżenie wynika z istoty metody, związanej z dyskretyzacją odpowiedzi $h(t)$.

Logarytmując (21) można napisać

$$r-r_0\approx\frac{\ln\Phi}{\ln 2}=\delta r, \tag{23}$$

gdzie

$$r_0=\frac{\ln\frac{h(2\Delta)}{h(\Delta)}}{\ln 2}. \tag{24}$$

Ponieważ funkcja Φ dla niewielkich relatywnie okresów próbkowania $(\Delta/\bar{T})$ jest w przybliżeniu równa 1, to wzór (24) pozwala na dość dokładne oszacowanie rzędu zredukowanego r modelu (1), dzięki zaokrągleniu r_0 do liczby całkowitej. W rzeczywi-

stości błąd δ_r, określany zależnością (23) może posłużyć do uzyskania dodatkowych informacji np. o wielkości rozrzutu w przypadku procesów wieloinercyjnych, dla których $m=0$, $r=n$.

Obliczając w granicy n_0 wg. (24), gdy $\Delta t \to 0$, można napisać w ogólności (dla dowolnego t_1 i t_2)

$$n_0 = \lim_{\Delta t \to 0} \frac{\ln \dfrac{h(t+\Delta t)}{h(t)}}{\ln \dfrac{t+\Delta t}{t}} = \lim_{\Delta t \to 0} \frac{\ln\left(1+\dfrac{\Delta h(t)}{h(t)}\right)}{\ln\left(1+\dfrac{\Delta t}{t}\right)}$$

zatem, w przybliżeniu $n_0 \approx \dfrac{t}{h(t)} \lim_{\Delta t \to 0} \dfrac{\Delta h(t)}{\Delta t}$, skąd

$$n_0 = h'(t)\frac{t}{h(t)}, \quad \text{gdzie} \quad h'(t) = \frac{dh(t)}{dt}. \tag{25}$$

Zależność (25) wskazuje na związek n_0 z pochodną odpowiedzi skokowej. Można zatem sądzić o pochodnej znając n_0 lub odwrotnie. Musi być jednak spełniony warunek aby czas t był relatywnie mały, o czym będzie mowa w dalszej części pracy.

Bazując na rozwinięciu (20), można także uzyskać zależności wiążące niektóre parametry modelu (1) ze stosunkiem próbek $h(2\Delta)/h(\Delta)$. Po prostych przekształceniach, uwzględniając tylko 2 pierwsze wyrazy rozwinięcia (20), można oszacować średnią inercję

$$\bar{T} \approx \frac{n}{\dfrac{r+1}{\Delta}\left[2^r \dfrac{h(\Delta)}{h(2\Delta)} - 1\right] + \dfrac{b_{m-1}}{b_m}}, \tag{26}$$

a współczynnik b_m licznika może być oszacowany

$$b_m \approx \frac{h^2(\Delta)}{h(2\Delta)}\left(\frac{2}{\Delta}\right)^r r!. \tag{27}$$

Łącząc wzór (26) ze wzorem (21), można napisać

$$b = \frac{b_{m-1}}{b_m} \approx \frac{n}{\bar{T}} - \frac{r+1}{\Delta}(1-\Phi). \tag{28}$$

Analizując budowę funkcji Φ, określonej zależnością (22), można zauważyć, że zarówno licznik jak i mianownik tej funkcji może być z dużą dokładnością przedstawiony przez zastąpienie funkcją wykładniczą typu e^{-x}. Można bowiem napisać

$$\Phi \approx \frac{e^{-\left[1-\frac{b_{m-1}}{b_m}\left(\frac{T}{n}\right)\right]\frac{2n}{r+1}\left(\frac{\Delta}{T}\right)}}{e^{-\left[1-\frac{b_{m-1}}{b_m}\left(\frac{T}{n}\right)\right]\frac{2n}{r+1}\left(\frac{\Delta}{T}\right)}} = e^{-\left[1-\frac{b_{m-1}}{b_m}\left(\frac{T}{n}\right)\right]\frac{n}{r+1}\left(\frac{\Delta}{T}\right)}. \tag{29}$$

Co uwzględniając w zależności (21), po zlogarytmowaniu pozwala wyrazić błąd δ_r jako funkcję o selektywnie rozdzielonym wpływie zarówno średniej stałej czasowej $\bar{T}$ jak i współczynnika $b=b_{m-1}/b_m$, dla modelu o transmitancji (1).

$$\delta_r = r - r_0 \approx \frac{1}{\ln 2}\left[\frac{n}{r+1}\frac{\Delta}{\bar{T}} - b\frac{\Delta}{r+1}\right]. \tag{30}$$

Inaczej przekształcając (30), uzyskuje się odpowiednik (28)

$$b \approx \frac{n}{\bar{T}} - \frac{\delta_r}{\Delta}(r+1)\ln 2. \tag{31}$$

Jakkolwiek wartość δ_r można wyznaczyć na drodze pomiarów i identyfikacji rzędu r, to jednak bez znajomości stopnia różniczkowania m i jednej z nieznanych wielkości $\bar{T}$ albo b, nie jest możliwe oszacowanie parametrów modelu (1) dlatego, że dla wyznaczenia $\bar{T}$ ze wzoru (26) niezbędna jest znajomość b i odwrotnie — dla określenia b ze wzoru (28) lub (31), niezbędna jest znajomość $\bar{T}$. Można zauważyć, iż w przypadku różniczkowania ($m>0$), może być $\delta_r \leqslant 0$ zaś w szczególności, w przypadku braku różniczkowania ($m=0$), dla procesów wieloinercyjnych, $\delta_r > 0$. Stąd jedynie wtedy, gdy wiadomo, że proces jest pozbawiony różniczkowania istnieje pewność co do sposobu zaokrąglenia r_0 do górnej liczby całkowitej. Mając informację o tym, że w procesie występuje działanie różniczkujące, należy dokonać zaokrąglenia do najbliższej liczby całkowitej.

3.1. SZACOWANIE PARAMETRÓW MODELI WIELOINERCYJNYCH [9]

Klasę tych modeli reprezentują transmitancje (8) i (10). W tym przypadku $n=r$, $m=0$, $b=0$, zaś oznaczony na podstawie pomiaru próbek rząd n_0 wg (24) spełnia relację

$$n = \mathrm{Ent}\{n_0\}. \tag{32}$$

Błąd δ_n, na podstawie (30) jest okreslony zależnością

$$\delta_n = n - n_0 \approx \frac{1}{\ln 2}\frac{n}{n+1}\frac{\Delta}{\bar{T}} \tag{33}$$

zaś $\bar{T}$, na podstawie (26) może być oszacowane w przybliżeniu

$$\bar{T} \approx \frac{n}{\dfrac{n+1}{\Delta}\left[2^n \dfrac{h(\Delta)}{h(2\Delta)} - 1\right]}. \tag{34}$$

Na podstawie (27) z uwzględnieniem (10), znając $\bar{T}$ można oszacować wzmocnienie modelu (10).

$$k \approx \frac{h^2(\Delta)}{h(2\Delta)}\left(\frac{2}{\Delta}\right)^n n!\bar{T}^n. \tag{35}$$

Zauważmy, że średnia stała czasowa, może być też oszacowana na podstawie (33), gdy znany jest błąd δ_n.

Dla oszacowania parametrów modelu (8), niezbędne jest oszacowanie współczynnika rozrzutu stałych czasowych λ oraz dodatkowo, niezbędna jest znajomość jednej ze stałych czasowych T_i np. $T_{\min}$ lub $T_{\max}$. Często w praktyce wymóg ten może być spełniony, bowiem przykładowo stała czasowa termopary umieszczonej w komorze pieca jako obiektu elektrotermicznego może być znana lub prosto określona. $T_{\min}$ może być także z grubsza oszacowana na podstawie zależności wynikającej z (3) i (6). Można bowiem napisać

$$\left.\begin{aligned}\frac{1}{\bar{T}}&=\frac{1}{T_{\max}}\frac{1+\frac{1}{\lambda}+\ldots+\frac{1}{\lambda^{n-1}}}{n}\\ \frac{1}{\bar{T}}&=\frac{1}{T_{\min}}\frac{1+\lambda+\ldots+\lambda^{n-1}}{n}\end{aligned}\right\} \tag{36}$$

Na podstawie (36), dla $\lambda\to 0$, $\bar{T}\to nT_{\min}$, stąd jest

$$\bar{T}\geqslant T_{\min}>\frac{\bar{T}}{n}. \tag{37}$$

Jest to jednak oszacowanie tym mniej dokładne im mniejszy jest rozrzut (im bliższy jedności jest współczynnik λ) — patrz Tablica II. W takich przypadkach lepszym wyjściem wydaje się szacowanie stałej czasowej $T_{\max}$.

Dla oszacowania $T_{\max}$ niezbędna jest znajomość $h(\infty)$ to jest k. Można przyjąć [13], że gdy odpowiedź skokowa $h(t)$ zbliża się do k, to o jej przebiegu decyduje już tylko $T_{\max}$. Wzmocnienie k modelu (10) lub (8) można oszacować albo na podstawie (35), albo mierząc przyrosty $\left(\frac{\Delta h}{\Delta}\right)$ można stwierdzić wystąpienie $\left(\frac{\Delta h}{\Delta}\right)_{\max}=\alpha$ i można określić odpowiadający temu czas t_m. Można wtedy napisać [9, 10] w przybliżeniu

$$k\approx\alpha t_m\sqrt{\frac{2\pi}{n_0-1}}. \tag{38}$$

Znając k oraz wartość $h(t_1)$ i $h(t_1+\Delta t)$ jak to pokazano na Rys. 1, można oszacować $T_{\max}$

$$T_{\max}\approx\frac{\Delta t}{\ln\frac{k-h(t_1)}{k-h(t_1+\Delta t)}}, \tag{39}$$

Dysponując rzędem n, średnią stałą czasową $\bar{T}$ oraz oszacowaniem albo $T_{\min}$, albo $T_{\max}$, można na podstawie (36) napisać równania odpowiednio:

$$\left.\begin{aligned} &\lambda^{n-1}+\ldots+\lambda+1-n\frac{T_{\min}}{\bar{T}}=0 \\ &\text{lub} \\ &\lambda^{n-1}\left(1-n\frac{T_{\max}}{\bar{T}}\right)+\ldots+\lambda+1=0 \end{aligned}\right\}, \tag{40}$$

które pozwalają oszacować współczynnik rozrzutu λ a zatem także oszacować poszczególne stałe czasowe modelu (8). Należy zauważyć, że łącząc wzory (33) i (36) można wykazać bezpośredni związek błędu δ_n z wielkością współczynnika rozrzutu λ. Przykładowo, dla $T_{\min}$ jest

$$\delta_n \approx \frac{1}{\ln 2}\left(\frac{\Delta}{T_{\min}}\right)\frac{1+\lambda+\ldots+\lambda^{n-1}}{n+1}. \tag{41}$$

Szczególnym przypadkiem modelu M(s), wymagającym specjalnego omówienia jest najprostszy, często przydatny model dwuinercyjny $M_2(s)$

$$M_2(s)=\frac{k}{(1+sT_1)(1+sT_2)}. \tag{42}$$

Jest on tym bardziej interesujący, że na podstawie dodatkowej informacji pomiarowej o wartości t_m w którym to czasie osiągnięta jest maksymalna pochodna odpowiedzi skokowej, można oszacować wartość współczynnika rozrzutu λ. Po dwukrotnym zróżniczkowaniu $h(t)$, uzyskuje się równanie

$$\frac{e^{-\frac{t_m}{T_2}}}{T_2}-\frac{e^{-\frac{t_m}{T_1}}}{T_1}=0, \tag{43}$$

na podstawie którego, stosując (3) i (6) jest

$$\lambda=e^{-\frac{2t_m}{\bar{T}}\left(\frac{1-\lambda}{1+\lambda}\right)} \tag{44}$$

lub

$$\frac{t_m}{\bar{T}}=\frac{1}{2}\frac{1+\lambda}{1-\lambda}\ln\frac{1}{\lambda}, \tag{45}$$

a biorąc pod uwagę (33), można także napisać dla modelu (42)

$$\frac{t_m\delta_m}{\Delta}=\frac{1}{3\ln 2}\frac{1+\lambda}{1-\lambda}\ln\frac{1}{\lambda}. \tag{46}$$

Należy zauważyć, że wielkości t_m, $\bar{T}$, δ_n mogą być uzyskane z pomiarów dyskretnych początkowej fazy odpowiedzi skokowej h(t). Jednak forma zależności (44), (45) i (46) jest uwikłana, stąd analityczne wyznaczenie współczynnika λ nie jest możliwe. Mając jednak na uwadze, że $0<\lambda\leqslant 1$, można zastosować przybliżenie dla funkcji logarytmicznej, co pozwala napisać w przybliżeniu

$$\left.\begin{aligned} &\frac{t_m}{\bar{T}} \approx 1+\frac{1}{3}\left(\frac{1-\lambda}{1+\lambda}\right)^2 \\ &\text{lub} \\ &\frac{t_m \delta_n}{\Delta}\frac{3}{2}\ln 2 \approx 1+\frac{1}{3}\left(\frac{1-\lambda}{1+\lambda}\right)^2 \end{aligned}\right\} \tag{47}$$

oznaczając $\gamma=\frac{t_m}{\bar{T}}=\frac{t_m\delta_n}{\Delta}\frac{3}{2}\ln 2$, można na podstawie (47) oszacować λ, dla modelu dwuinercyjnego

$$\lambda \approx \frac{3\gamma-2}{4-3\gamma}-\sqrt{\left(\frac{3\gamma-2}{4-3\gamma}\right)^2-1}. \tag{48}$$

Tak więc w przypadku $n=2$, stosując model (42), rozrzut może być oszacowany także w czasie początkowej fazy odpowiedzi skokowej — już w momencie osiągnięcia punktu przegięcia, dla $t=t_m$. Gdy rząd $n>2$, nie wydaje się możliwym efektywne rozwikłanie skomplikowanych zależności, prowadzące do rozwiązania analitycznego w postaci wzoru na λ, będącego odpowiednikiem np. (48). Przykładowo, dla $n=3$, uzyskuje się

$$\frac{e^{-\frac{3t_m}{T}\frac{\lambda}{1+\lambda+\lambda^2}}-e^{-\frac{3t_m}{T}\frac{1}{1+\lambda+\lambda^2}}}{e^{-\frac{3t_m}{T}\frac{\lambda^2}{1+\lambda+\lambda^2}}-e^{-\frac{3t_m}{T}\frac{\lambda}{1+\lambda+\lambda^2}}}=\lambda \tag{49}$$

a dla $n=4$ będzie

$$\frac{e^{-\frac{4t_m}{T}\frac{\lambda^2}{1+\lambda+\lambda^2+\lambda^3}}-e^{-\frac{4t_m}{T}\frac{\lambda}{1+\lambda+\lambda^2+\lambda^3}}}{e^{-\frac{4t_m}{T}\frac{\lambda^3}{1+\lambda+\lambda^2+\lambda^3}}-e^{-\frac{4t_m}{T}\frac{1}{1+\lambda+\lambda^2+\lambda^3}}}=\frac{\lambda(1-\lambda)}{1-\lambda^3}. \tag{50}$$

Zależności (49) i (50) mogą być wykorzystane do oszacowania λ na drodze obliczeń numerycznych lub do sprawdzenia prawidłowości oszacowania λ na innej drodze. Innym, mniej znanym modelem procesów wieloinercyjnych jest model w postaci transmitancji [11]

$$F(s)=\frac{k}{\prod_{i=1}^{n}\left(1+s\frac{T}{i}\right)}. \tag{51}$$

Ta sztuczna postać transmitancji narzuca z góry wartość współczynnika rozrzutu, uzależniając λ od rzędu n jedynie

$$\lambda=\frac{1}{\sqrt[n-1]{n}}, \text{ przy czym, zgodnie z (3)} \tag{52}$$

$$T=\frac{\bar{T}(n+1)}{2}. \tag{53}$$

Dla $F(s)$, odpowiedź skokowa jest określona prostą zależnością

$$h(t)=k\left(1-e^{-\frac{t}{\bar{T}}}\right)^n. \tag{54}$$

Przyrównując $h''(t)=0$, uzyskuje się wzór dla modelu $F(s)$

$$\frac{t_m}{\bar{T}}=\left(\frac{n+1}{2}\right)\ln n. \tag{55}$$

Zależność (55) może w trakcie identyfikacji dowolnego procesu wieloinercyjnego posłużyć do sprawdzenia czy model (51) „pasuje" do badanej odpowiedzi gdyż w momencie $t=t_m$ mogą być już znane parametry $\bar{T}$ i n. Dla porównania, w Tablicy III zestawiono przykładowo wartości $t_m/\bar{T}$, dla różnych modeli.

Tablica III

Przykładowe wartości stosunku $t_m/\bar{T}$ dla wybranych modeli wieloinercyjnych

Model	$t_m/\bar{T}$			Uwagi
	$n=2$	$n=3$	$n=4$	
$S(s)$	1	2	3	$\lambda=1$
$F(s)$	1,0397	2,197	3,466	λ narzucone, zależne od n
$M_2(s)$	1,207 gdy $\lambda=0{,}2$	—	—	Stosunek $t_m/\bar{T}$ silnie zależy od λ
	1,407 gdy $\lambda=0{,}9$			

3.2. SZACOWANIE PARAMETRÓW MODELI OBIEKTÓW WIELOINERCYJNYCH Z JEDNOKROTNYM RÓŻNICZKOWANIEM

Dla modelu transmitancyjnego określonego wzorem (11) odpowiedź skokową $h_d(t)$ można określić jako sumę

$$h_d(t)=b_1\frac{dh(t)}{dt}+b_0h(t), \tag{56}$$

$$\text{gdzie} \qquad h(t)=L^{-1}\left\{\frac{1}{s}\,\frac{1}{(1+s\bar{T})^n}\right\} \tag{57}$$

co uwzględniając, dla $t=\Delta$ będzie

$$h_d(\Delta)=b_1\,\frac{e^{-\frac{\Delta}{\bar{T}}}}{(n-1)!\bar{T}}\left(\frac{\Delta}{\bar{T}}\right)^{n-1}+b_0\left\{1-e^{-\frac{\Delta}{\bar{T}}}\left[1+\frac{\Delta}{\bar{T}}+\ldots+\left(\frac{\Delta}{\bar{T}}\right)^{n-1}\frac{1}{(n-1)!}\right]\right\}. \tag{58}$$

Dla wyznaczenia analitycznego czasu t_m, odpowiadającego maksymalnej wartości pochodnej $\left(\frac{dh_d(\Delta)}{d\Delta}\right)_{\max}=\alpha$, która to wielkość może być w procesie identyfikacji określona, można na podstawie (58), dwukrotnie różniczkując, napisać

$$\frac{d^2h_2(\Delta)}{d\Delta^2}=\left(\frac{\Delta}{\overline{T}}\right)^{n-3}\frac{e^{-\frac{\Delta}{\overline{T}}}}{\overline{T}^2(n-3)!}\left[\frac{b_1}{\overline{T}}+\frac{1}{(n-2)}\left(\frac{\Delta}{\overline{T}}\right)\left(b_0-\frac{2b_1}{\overline{T}}\right)-\frac{1}{(n-1)(n-2)}\left(\frac{\Delta}{\overline{T}}\right)^2\left(b_0-\frac{b_1}{\overline{T}}\right)\right] \tag{59}$$

skąd t_m jest

$$t_m=\frac{\overline{T}(n-1)}{2}\left(\frac{2-b\overline{T}}{1-b\overline{T}}\right)\pm\frac{\overline{T}^2(n-1)(n-2)}{2(1-b\overline{T})}\sqrt{\left(\frac{2-b\overline{T}}{\overline{T}(n-2)}\right)^2+4\frac{b\overline{T}-1}{\overline{T}^2(n-1)(n-2)}} \tag{60}$$

Zauważmy, że wartość t_m może być uzyskana na drodze pomiaru w trakcie identyfikacji. Także na drodze pomiaru, dla określonego Δ, może być określony błąd $\delta_r=r-r_0$. Stąd, korzystając z zależności (31) i (60), można napisać po przekształceniu równanie

$$\left.\begin{aligned}&\overline{T}^2\left(1-\frac{\delta t_m}{n-2}\right)+\overline{T}\left(t_m+\frac{\delta t_m^2}{(n-1)(n-2)}\right)-\frac{t_m^2}{n-2}=0\\&\text{gdzie } \delta=\frac{\delta_r}{\Delta}(r+1)\ln 2\end{aligned}\right\}. \tag{61}$$

Na podstawie (61) można oszacować średnią stałą czasową modelu (11) a mając $\overline{T}$, można na podstawie przekształconej zależności (59), obliczyć w przybliżeniu b

$$b=\frac{1}{\overline{T}}\frac{\frac{2}{n-2}\left(\frac{t_m}{\overline{T}}\right)-\frac{1}{(n-1)(n-2)}\left(\frac{t_m}{\overline{T}}\right)^2-1}{\frac{1}{n-2}\left(\frac{t_m}{\overline{T}}\right)-\frac{1}{(n-1)(n-2)}\left(\frac{t_m}{\overline{T}}\right)^2}. \tag{62}$$

Znając b i obliczając na podstawie (27) b_1, szacuje się $b_0=bb_1$, zgodnie z oznaczeniem przyjętym w (28).

Warto przeanalizować zależność błędu δ_r w przypadku identyfikacji modelu (11). Dla dowolnego rzędu n, przy jednokrotnym różniczkowaniu ($m=1$), odpowiedź modelu (11) jest określona przez (58). Zauważmy, że występujacy w nawiasie kwadratowym drugiego członu szereg jest uciętym na $(n-1)$ wyrazie, szeregiem rozwinięcia funkcji wykładniczej $e^{-\frac{\Delta}{\overline{T}}}$.

Można zatem w przybliżeniu zastąpić wyrażenie w nawiasie kwadratowym, wyrażeniem $e^{-\frac{\Delta}{\overline{T}}}-R_N$, gdzie $R_N=\frac{1}{n!}\left(\frac{\Delta}{\overline{T}}\right)^n$ jest resztą szeregu. Stosując zatem odpowiednie podstawienie

$$\left[1+\frac{\Delta}{\overline{T}}+\ldots+\left(\frac{\Delta}{\overline{T}}\right)^{n-1}\frac{1}{(n-1)!}\right]\approx e^{-\frac{\Delta}{\overline{T}}}-\frac{1}{n!}\left(\frac{\Delta}{\overline{T}}\right)^n \tag{63}$$

w wyrażeniu określającym $h_d(\Delta)$ i obliczając stosunek $h_d(2\Delta)/h_d(\Delta)$, będzie po prostych przekształceniach

$$\frac{h_d(2\Delta)}{h_d(\Delta)} \approx e^{-\frac{\Delta}{T}} 2^{n-1} \frac{1+\frac{b}{n}2\Delta}{1+\frac{b}{n}\Delta}. \tag{64}$$

Logarytmując (64), można określić błąd δ_r

$$\delta_r = r - r_0 \approx \frac{1}{\ln 2}\left[\frac{\Delta}{T} - \ln \frac{1+\frac{2b\Delta}{n}}{1+\frac{b\Delta}{n}}\right]. \tag{65}$$

Zależność (65), podobnie jak (30) ukazuje selektywny wpływ na wartość δ_r zarówno średniej stałej czasowej $\overline{T}$ a co z tym związane rozrzutu inercji, jak i współczynnika b, określającego „siłę różniczkowania". Jak widać, znak δ_r może być różny w zależności od tego, który z tych wpływów ma przewagę. Zależność (65) jest zgodna w przybliżeniu (jeśli zastąpić dla małych x funkcję $\ln(1+x)$ przez x) ze wzorem (30). Wynika z tego, że mając informację iż $m=1$, dla dowolnego n, identyfikowane r_0 należy zaokrąglić do najbliższej liczby całkowitej r. Reasumując, procedura identyfikacji przybliżonej modelu $D(s)$ jest następująca: przy znanym okresie próbkowania Δ, na podstawie pomiarów $h_d(\Delta)$ i $h_d(2\Delta)$ szacuje się r_0 i r oraz b_1. Następnie po uchwyceniu t_m, uznając np. za t_m chwilę, w której przestaje narastać $\Delta h_d/\Delta$, można oszacować z (61) średnią stałą czasową $\overline{T}$, oraz ze wzoru (62) b a następnie określić b_0 co kończy procedurę szacowania parametrów. Zauważmy, że mając oszacowaną wartość b, można w dalszych krokach, na podstawie (65) oszacować rozrzut stałych czasowych λ podobnie jak w p. 3.1.

3.3. SZACOWANIE PARAMETRÓW MODELU OSCYLACYJNEGO

Dysponując *a priori* informacją o tym, że proces identyfikowany jest oscylacyjny typu (12) lub zakładając, że dany proces może być aproksymowany modelem (12), można po sprawdzeniu tego założenia badając wg (24) rząd — oszacować parametry transmitancji (12) na podstawie pomiarów dyskretnych wartości odpowiedzi skokowej $h(t)$ oraz pomiaru czasu t_m, dla którego przyrosty $\Delta h/\Delta t$ są maksymalne.

Odpowiedź skokowa modelu (12) jest określona zależnością

$$h_0(t) = k\left[1 - e^{-\beta\omega_0 t}\left(\cos\omega_0\sqrt{1-\beta^2}\,t + \frac{\beta}{\sqrt{1-\beta^2}}\sin\omega_0\sqrt{1-\beta^2}\,t\right)\right]. \tag{66}$$

Stosując przybliżenia $e^x \approx 1+x$, $\cos \approx 1-\frac{x^2}{2}$, $\sin \approx x$ dla $t=\Delta$, gdzie Δ jest relatywnie małe, będzie

$$h_0(\Delta) \approx k(1+\beta^2)\frac{\omega_0^2\Delta^2}{2}, \tag{67}$$

gdzie nieznane są k oraz ω_0 i β. Średnia stała czasowa $\overline{T}$ na podstawie (3) i (12) jest

$$\overline{T}=\frac{1}{\beta\omega_0}, \tag{68}$$

można ją oszacować na podstawie zależności (33) lub (34) wyznaczając uprzednio rząd n_0 z (24).

Dodatkowe, niezbędne równanie można uzyskać obliczając ekstremum pochodnej po czasie odpowiedzi $h_0(t)$. Po zróżniczkowaniu (66) jest

$$\frac{dh_0(t)}{dt}=\frac{k\omega_0 e^{-\beta\omega_0 t}}{\sqrt{1-\beta^2}}\sin\omega_0\sqrt{1-\beta^2}t. \tag{69}$$

Funkcja ta osiąga maksimum dla $t=t_m$

$$t_m=\frac{\arcsin\sqrt{1-\beta^2}}{\omega_0\sqrt{1-\beta^2}}. \tag{70}$$

Czasowi t_m odpowiada wartość odpowiedzi $h_0(t_m)=h_m$ w punkcie przegięcia

$$h_m=k\left[1-2\beta e^{-\frac{\beta\arcsin\sqrt{1-\beta^2}}{\sqrt{1-\beta^2}}}\right]. \tag{71}$$

Uwzględniając w (70) zależność (68), można napisać dla modelu oscylacyjnego (12)

$$\frac{t_m}{\overline{T}}=\beta\frac{\arcsin\sqrt{1-\beta^2}}{\sqrt{1-\beta^2}}. \tag{72}$$

Na podstawie (67) z uwzględnieniem (68), można by obliczyć w przybliżeniu k jako funkcję pomierzonej $h(\Delta)$, obliczonego $\overline{T}$ i nieznanego β i wstawić do (71) z uwzględnieniem (72), uzyskuje się skomplikowane równanie 3-go stopnia ze względu na nieznany parametr β co nie jest satysfakcjonujące

$$2\beta^3 e^{-\frac{t_m}{\overline{T}}}-\beta^2\left(1-\frac{h_m\Delta^2}{2h(\Delta)\overline{T}^2}\right)+\frac{\Delta^2 h_m}{2h(\Delta)\overline{T}^2}=0,$$

gdzie h_m jest uzyskane w wyniku pomiaru dla $t=t_m$.

W tej sytuacji wydaje się lepszym rozwiązaniem zastąpienie prawej strony wzoru (72) przez przybliżenie szeregu będącego rozwinięciem funkcji arcsinx. Po przekształceniach można napisać prostą zależność przybliżoną

$$\beta\approx 0{,}78\frac{t_m}{\overline{T}}. \tag{73}$$

Ponieważ czas t_m jest oszacowany w drodze pomiarów a $\overline{T}$ jest oszacowane na podstawie (34), można w przybliżeniu oszacować β ze wzoru (73) oraz ω_0 na podstawie (68) a następnie z (67) można oszacować wzmocnienie k:

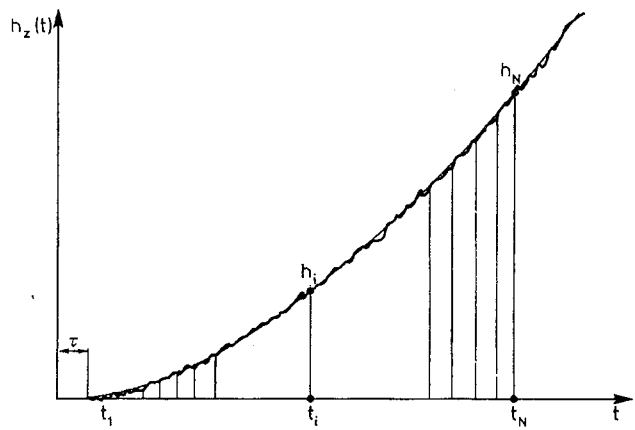
$$k \approx \frac{2h(\Delta)}{\Delta^2} \overline{T^2} \frac{\beta^2}{1+\beta^2}.$$

Zauważmy, że wzmocnienie k można również oszacować na podstawie (35) z uwzględnieniem (68)

$$k \approx \frac{h^2(\Delta)}{h(2\Delta)} \frac{8}{\Delta^2} \beta^2 \overline{T^2}. \tag{74}$$

4. UWAGI NA TEMAT WPŁYWU ZAKŁÓCEŃ ORAZ DOKŁADNOŚCI STOSOWANYCH PRZYBLIŻEŃ

Istotnym problemem jaki się pojawia przy stosowaniu w praktyce prezentowanej metody identyfikacji jest minimalizacja wpływu zakłóceń losowych, nałożonych na odpowiedź próbkowaną. W tym przypadku, można zastosować [8] zamiast jednej pary próbek odpowiedzi skokowej, całą serię uzyskaną w trakcie próbkowania jak to przedstawiono na Rys. 2.



Rys. 2. Ilustracja serii pomiarów zakłóconej odpowiedzi $h_z(t)$, dla minimalizacji wpływu zakłóceń

Dysponując serią N próbek, można uzyskać

$\frac{N(N-1)}{2}$ wyników obliczeń parametrów. Oznaczając $h_i = h(t_i)$ można napisać

$$n_{0_i} = \frac{\ln \frac{h(t_i)}{h(t_j)}}{\ln \frac{t_i}{t_j}}, \tag{75}$$

gdzie $i=k,\ k+1...k+N,\ j<i$. Należy zadbać o to, aby t_{k+N} nie było za duże. W szczególnym przypadku $k=1$, tak jak to przyjęto na Rys. 2.

Przykładowo, dla $N=10$, można uzyskać 45 wyników obliczeń n_0, które mogą zostać uśrednione po odrzuceniu wartości skrajnych.

Jednak w takim przypadku, zależności stosowane do obliczeń parametrów winny być odpowiednio zmodyfikowane jak to przykładowo uczyniono ze wzorem (24) podając (75). Jest to ważne w przypadku algorytmizacji obliczeń w procesie identyfikacji. Oczywiście, dla sterowania może być zastosowany zupełnie inny okres próbkowania.

Ten problem nie jest przedmiotem niniejszej pracy. W dalszej części pokazane zostaną jedynie przykładowe wyniki symulacji wpływu zakłóceń.

W celu minimalizacji wpływu zakłóceń, można również próbkować zakłóconą odpowiedź po przepuszczeniu jej przez filtr o znanej transmitancji (np. inercji I rzędu o stałej T_f) a w otrzymanym na drodze identyfikacji modelu, należy uwzględnić fakt dołączenia filtru. Przy tym powinna być spełniona nierówność $T_{\max}<T_f$. Problem doboru wielkości T_f stanowi odrębne zagadnienie. Może się okazać, że na skutek relatywnie dużej szybkości identyfikowanego procesu w stosunku do pasma częstotliwości zakłóceń, ten sposób staje się bezużyteczny. Innym ważnym problemem jest zagadnienie dokładności stosowanych przybliżeń z uwagi na to, że autorowi zależało na uzyskaniu w miarę możliwości prostych zależności analitycznych przydatnych do obliczeń inżynierskich i możliwych do interpretacji fizycznej. Istotnym dla tych rozważań jest problem dokładności przybliżenia (21). Zauważymy, że dla modelu $S(s)$, wartości stosunku $h(2\Delta)/h(\Delta)$ mogą być obliczone dokładnie, co pozwala na porównanie czynionych błędów, które to błędy mogą wynikać albo z uwzględnienia skończonej liczby wyrazów (20) dla modelu $S(s)$, lub też ze sposobu aproksymacji typu (29) lub (63). Zatem, dla $S(s)$ jest dokładnie

$$\frac{h(2\Delta)}{h(\Delta)}=\frac{1-e^{-\frac{2\Delta}{T}}\left[1+\frac{2\Delta}{T}+\frac{1}{2!}\left(\frac{2\Delta}{T}\right)^2+\ldots+\frac{1}{(n-1)!}\left(\frac{2\Delta}{T}\right)^{n-1}\right]}{1-e^{-\frac{\Delta}{T}}\left[1+\frac{\Delta}{T}+\frac{1}{2!}\left(\frac{\Delta}{T}\right)^2+\ldots+\frac{1}{(n-1)!}\left(\frac{\Delta}{T}\right)^{n-1}\right]}. \tag{76}$$

Natomiast, rozwijając w szereg wg (21) odpowiedź skokową i uwzględniając 3 wyrazy będzie w przybliżeniu dla $S(s)$

$$\frac{h(2\Delta)}{h(\Delta)}\approx 2^n\frac{1-\frac{2n}{n+1}\frac{\Delta}{T}+\frac{2n}{n+2}\left(\frac{\Delta}{T}\right)^2}{1-\frac{n}{n+1}\frac{\Delta}{T}+\frac{n}{2(n+2)}\left(\frac{\Delta}{T}\right)^2}, \tag{77}$$

a uwzględniając tylko 2 wyrazy, zależność ta ulegnie dalszemu uproszczeniu

$$\frac{h(2\Delta)}{h(\Delta)}\approx 2^n\frac{1-\frac{2n}{n+1}\frac{\Delta}{T}}{1-\frac{n}{n+1}\frac{\Delta}{T}}. \tag{78}$$

Natomiast, stosując aproksymację typu (29), jest w przybliżeniu

$$\frac{h(2\Delta)}{h(\Delta)} \approx 2^n e^{-\frac{n}{n+1}\frac{\Delta}{\bar{T}}} \tag{79}$$

zaś przyjmując podstawienie (63), będzie

$$\frac{h(2\Delta)}{h(\Delta)} \approx 2^n e^{-\frac{\Delta}{\bar{T}}}. \tag{80}$$

Przyjmując, że procentowa niedokładność przybliżenia, będzie mierzona za pomocą prostej zależności

$$\delta \frac{h(2\Delta)}{h(\Delta)} = \frac{\left(\frac{h(2\Delta)}{h(\Delta)}\right)_{\text{dokładne}} - \left(\frac{h(2\Delta)}{h(\Delta)}\right)_{\text{przybliżone}}}{\left(\frac{h(2\Delta)}{h(\Delta)}\right)_{\text{dokładne}}}, \tag{81}$$

można ocenić i porównać powyższe sposoby stosowanych przybliżeń. W poniższej tablicy zestawiono przykładowo wyniki obliczeń dla $n=3$, $\bar{T}=1$, $k=1$, dla kilku wybranych wartości $\Delta/\bar{T}$.

Tablica IV

Przykładowe wyniki błędów procentowych, wynikających ze stosowanych przekształceń, dla różnych $\Delta/\bar{T}$

	$\Delta/\bar{T}$	Zależność dokładna (76)	Przybliżenie 3-wyrazowe (77)	Przybliżenie 2-wyrazowe (78)	Przybliżenie wg (29) (79)	Przybliżenie wg. (63) (80)
$\frac{h(2\Delta)}{h(\Delta)}$	0,1	7,426173439	7,431016313	7,35135456	7,421947891	7,238699344
$\delta\frac{h(2\Delta)}{h(\Delta)}$	0,1	–	−0,0652%	1,0077%	0,0569%	2,524%
	0,25	–	−0,705%	7,549%	0,357%	6,356%
	0,5	–	−12,62%	42,6%	1,486%	13,06%

Zdecydowanie najlepsze jest przybliżenie (79). Można też zauważyć, iż maksymalny czas t_{k+N} (rys. 2) powinna spełniać relację

$$\frac{t_{k+N}}{\bar{T}} < 1 \tag{82}$$

co wynika także z analizy rozwinięcia (20).

5. PRZYKŁADY ZASTOSOWAŃ

Jako przykłady zastosowań prezentowanej metody zostaną przytoczone między innymi wyniki szacowania parametrów 2 obiektów przemysłowych elektrotermicz-

nych a także wyniki symulacji pracy regulatora samonastrajającego PID, opartego na prezentowanej metodzie szacowania parametrów.

5.1. PIEC ELEKTRYCZNY TYPU PEG at, 90 kW, PRODUKCJI Z-DU ELTERMA

Do badań eksperymentalnych wykorzystano dane pomiarowe tzw. charakterystyki rozgrzewu piecy produkowanych przez ELTERMĘ w Świebodzinie. Pomiary te były wykonywane zgodnie z PN-73/E-06209, i polegały w tym przypadku na rejestracji odpowiedzi $h(t)$ pieca od mometu $t=0$, w którym została włączona pełna moc grzejna, aż do momentu osiągnięcia temperatury znamionowej pieca, uwzględniając tzw. czas rozgrzewu pieca. Niezbędne dane pomiarowe przedstawiono poniżej

$h(\Delta)\approx 2{,}1^{\circ}\mathrm{C}$ $\quad \alpha\approx 7{,}69^{\circ}\mathrm{C/min}$ $\quad h(t_1)=870^{\circ}\mathrm{C}$
$h(2\Delta)\approx 7{,}0^{\circ}\mathrm{C}$ $\quad t_m\approx 20$ min $\quad h(t_1+\Delta t)=950^{\circ}\mathrm{C}$
$\Delta=5$ min $\quad \tau_z\approx 7{,}5$ min $\quad \Delta t=30$ min

W tym przypadku można było dokonać graficznej aproksymacji odpowiedzi $h(t)$ i w przybliżeniu oszacować na tej podstawie $k\approx 1150^{\circ}\mathrm{C}$ i $T_z\approx 140$ min. Wyniki obliczeń wg wzorów, których numery podano, umieszczono w Tablicy V.

Tablica V

Wyniki oszacowań parametrów modeli pieca PEGat.

Nr wzoru	13	24	32	34	35	38	39	40	7	48	36	36
Parametr	n_0	n_0	n	$\bar{T}$	k	k	T_{max}	λ	T_{min}	λ	T_{max}	T_{min}
Wartość parametru	1,13	1,737	2	16,7	56	449	89	0,103	9,2	0,122	76,8	9,37
Jednostka	–	–	–	min	°C	°C	min	–	min	–	min	min

Wyniki te pozwalają na przyjęcie alternatywnych modeli pieca PEG at 90 KW:

$$K(s)\approx\frac{1150e^{-7{,}5s}}{1+s140}, \qquad S(s)\approx\frac{1150}{(1+s16{,}7)^2}$$

$$M'(s)\approx\frac{1150}{(1+s9{,}2)(1+s89)}, \qquad M''(s)\approx\frac{1150}{(1+s9{,}37)(1+s76{,}8)}.$$

Tu, jak i w następnych przykładach, model $M'(s)$ uzyskano stosując zależności (39) i (40) dla oszacowania rozrzutu, zaś model $M''(s)$, stosując wzory (48) i (36). Wzmocnienie k we wszystkich modelach świadomie przyjęto takie, jakie wynikało z możliwej do przeprowadzenia aproksymacji graficznej — co może być oczywiście dyskusyjne. Zauważmy, że obliczając $\bar{T}$ ze wzoru (33), uzyskuje się $\bar{T}=18{,}3$ min, co powoduje wzrost wzmocnienia k wg. (35) do $67{,}1^{\circ}\mathrm{C}$. Zmienia się też wtedy nieco λ.

5.2. SUSZARKA ELEKTRYCZNA PRODUKCJI ELTRMA, TYPU SDM-1E, Z TERMOPARĄ BEZ OSŁONY

$h(\Delta) \approx 7{,}7^{\circ}\mathrm{C}$ $\quad \alpha \approx 5{,}0^{\circ}\mathrm{C/min}$ $\quad h(t_1) = 210^{\circ}\mathrm{C}$
$h(2\Delta) \approx 25{,}8^{\circ}\mathrm{C}$ $\quad t_m \approx 16{,}75$ min $\quad h(t_1 + \Delta t) = 241^{\circ}\mathrm{C}$
$\Delta = 4$ min $\quad \tau_z \approx 2{,}9$ min $\quad \Delta t = 10$ min

W tym przypadku również oszacowano graficznie aproksymując $k \approx 372$ C i $T_z \approx 70$ min. Wyniki obliczeń przedstawia Tablica VI.

Tablica VI

Zestawienie wyników oszacowań parametrów suszarki z termoparą bez osłony

Nr wzoru	13	24	32	34	35	38	39	40	7	48	36	36
Parametr	n_0	n_0	n	$\bar{T}$	k	k	T_{max}	λ	T_{min}	λ	T_{max}	T_{min}
Wartość parametru	1,36	1,745	2	13,76	218	350	57,2	0,136	7,78	0,107	71,5	7,6
Jednostka	–	–	–	min	°C	°C	min	–	min	–	min	min

Wyniki oszacowań, pozwalają na przyjęcie alternatywnych modeli suszarki:

$$K(s) \approx \frac{372e^{-2{,}9s}}{1+s70}, \qquad S(s) \approx \frac{372}{(1+s13{,}76)^2}$$

$$M'(s) \approx \frac{372}{(1+s7{,}78)(1+s57{,}2)}, \qquad M''(s) \approx \frac{372}{(1+s7{,}6)(1+s71{,}5)}$$

* Obliczając $\bar{T}$ na podstawie zależności (33) uzyskuje się wynik $\bar{T} \approx 15{,}08$ min.

5.3. SUSZARKA ELEKTRYCZNA PRODUKCJI ELTERMA, TYPU SDM-1E, Z TERMOPARĄ W FABRYCZNEJ OSŁONIE

Jest to ta sama suszarka jak w p. 5.2, ale z termoparą w osłonie metalowej.

$h(\Delta) \approx 7^{\circ}\mathrm{C}$ $\quad \alpha \approx 5{,}0^{\circ}\mathrm{C/min}$ $\quad h(t_1) = 202^{\circ}\mathrm{C}$
$h(2\Delta) \approx 31^{\circ}\mathrm{C}$ $\quad t_m \approx 17{,}5$ min $\quad h(t_1 + \Delta t) = 229^{\circ}\mathrm{C}$
$\Delta = 5$ min $\quad \tau_z \approx 4$ min $\quad \Delta t = 10$ min

Tablica VII

Wyniki oszacowań parametrów suszarki z termoparą w osłonie

Nr wzoru	13	24	32	34	35	38	39	40	7	33	37	36	36
Parametr	n_0	n_0	n	$\bar{T}$	k	k	T_{max}	λ	T_{min}	$\bar{T}$	T_{min}	λ	T_{max}
Wartość parametru	1,46	2,147	3	4,65	61	205	57,8	0,203	2,37	6,34	2,64	0,206	62
Jednostka	–	–	–	min	°C	°C	min	–	min	min	min	–	min

Również w tym przypadku oszacowano aproksymując graficznie $k \approx 372^{0}$C i $T_z \approx 68$ min. Wyniki obliczeń przedstawia Tablica VII.

Na podstawie uzyskanych wyników oszacowań, można przyjąć alternatywne modele przybliżone jak niżej podano. Stałą czasową T_2, obliczone na podstawie (7). Model $M''(s)$ uzyskano wychodząc z inercji $\bar{T}$ na podstawie (33) i szacując $T_{\min}$ na podstawie nierówności (37).

$$K(s) \approx \frac{372e^{-4s}}{1+s68}, \qquad S(s) \approx \frac{372}{(1+s4{,}65)^3}$$

$$M'(s) \approx \frac{372}{(1+s2{,}37)(1+s11{,}8)(1+s57{,}8)}, \qquad M''(s) \approx \frac{372}{(1+s2{,}64)(1+s12{,}8)(1+s62)}.$$

5.4. WYNIKI OSZACOWAŃ PARAMETRÓW MODELI PROCESÓW LINIOWYCH Z JEDNOKROTNYM RÓŻNICZKOWANIEM

W tym przypadku, dla sprawdzenia zależności przytoczonych w rozdziale 3.2, wykorzystano dane uzyskane na drodze symulacji liniowych obiektów o jednakowej stałej czasowej z inercją 3 i 4 rzędu oraz z jednokrotnym różniczkowaniem, w postaci transmitancji

$$D_1(s) = \frac{b_1 s+1}{(1+s)^3} \quad \text{i} \quad D_2(s) = \frac{b_1 s+1}{(1+s)^4},$$

przy zmianie współczynnika $b_1 = 0{,}1$, 0,25, 0,5. Wartość czasu t_m odczytano w przybliżeniu z wykresu $h(t)$ a wyniki obliczeń ujęto w Tablicy VIII dla $\Delta = 0{,}2$. Należy podkreślić, że wyniki mogłyby być lepsze, gdyby wzięto dane dla $\Delta = 0{,}1$. Zwraca uwagę fakt dużych błędów w oszacowaniu parametrów dla $n=4$ i małego b_1.

Tablica VIII

Wyniki oszacowań parametrów we? z jednokrotnym różniczkowaniem

rząd	$n=3$			$n=4$		
b_1	0,1	0,25	0,5	0,1	0,25	0,5
$h(\Delta)$	0,0028	0,0052	0,0093	0,0002	0,0003	0,0006
$h(2\Delta)$	0,0133	0,0213	0,0347	0,0015	0,0026	0,0044
δ_r (23)	−0,2479	−0,0138	0,1	0,0931	−0,1155	0,125
t_m	1,9	1,72	1,4	2,8	2,65	2,4
$\delta(61)$	−2,578	−0,143	1,04	1,291	−1,6	1,74
$\bar{T}$ (61)	1,04	1,05	0,9977	1,085	0,943	0,948
b (62)	8,47	3,152	1,962	2,62	5,76	2,346
b_1 (27)	0,118	0,254	0,4985	0,16	0,208	0,491
b_0 (28)	0,999	0,80	0,978	0,42	1,19	1,15

5.5. WYNIKI OSZACOWANIA PARAMETRÓW OBIEKTU OSCYLACYJNEGO

Podobnie jak w p. 5.4 posłużono się próbkami uzyskanymi na drodze symulacji liniowego obiektu oscylacyjnego (12) dla kilku wybranych wartości β i ω_0, dla wzmocnienia $k=1$ i $\Delta=0{,}2$ oraz $\Delta=0{,}5$. Wyniki obliczeń zestawiono w Tablicy IX.

Tablica IX

Wyniki oszacowań na podstawie danych z symulacji odpowiedzi skokowej obiektu oscylacyjnego (12)

rząd	$\beta=0{,}1$, $\omega_0=0{,}3$, $\bar{T}=33{,}3$		$\beta=0{,}2$, $\omega_0=0{,}2$, $\bar{T}25$		$\beta=0{,}4$, $\omega_0=0{,}2$, $\bar{T}=12{,}5$	
Δ	0,2	0,5	0,2	0,5	0,2	0,5
$h(\Delta)$	0,00179228	0,01111751	0,00079564	0,0049299	0,00079943	0,00486526
$h(2\Delta)$	0,0071349	0,0437848	0,00316445	0,01941268	0,00313115	0,01891
n_0	1,99245	1,977588	1,991758	1,97736	1,98416	1,95853
t_m	~4,9	~4,9	~7,0	~7,0	~6,3	~6,3
$\bar{T}$ (34)	27,36	21,60	23,43	21,41	12,08	11,44
$\bar{T}$ (33)	27,26	21,46	23,34	21,24	12,14	11,59
β (73)	0,14	0,17	0,23	0,25	0,407	0,429
ω_0 (68)	0,26	0,27	0,185	0,187	0,03	0,204
k (67)	1,29	1,165	1,10	1,063	0,82	0,79
k (74)	1,32	1,218	1,16	1,148	0,967	0,965

Trzeba podkreślić, że źródłem błędów oszacowania, może być zbyt grube przybliżenie dla wartości β — wzór (73).

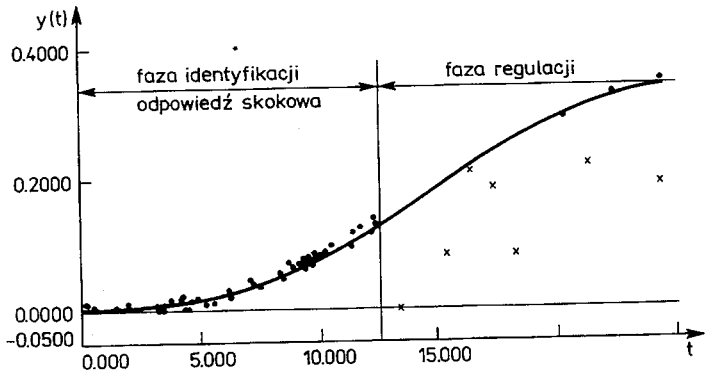
5.6. PRZYKŁAD ZASTOSOWANIA PREZENTOWANEJ METODY IDENTYFIKACJI UPROSZCZONEJ DO SAMONASTROJANIA REGULATORA PID [8]

Opracowano program symulacyjny układu regulacji stałowartościowej obiektu rzędu 1, 2 i 3, z możliwością nastawienia poszczególnych stałych czasowych T_1, T_2 i T_3, wyposażonego w model inercyjny elementu wykonawczego przed obiektem i w model filtru inercyjnego za obiektem, z nastawialnymi stałymi czasowymi, umożliwiającymi losowe zakłócenie próbek w trakcie identyfikacji i sterowania po samonastrojeniu. Program symulacyjny umożliwia również wprowadzenie nieliniowości przed lub za obiektem, wprowadzenie histerezy oraz błędów przetworników A/C i C/A. Możliwe jest też wprowadzenie dowolnego kryterium doboru nastaw regulatora PID dla samonastrojenia.

Samonastrojenie regulatora PID następuje automatycznie w pierwszej fazie rozruchu układu regulacji, dla danej wartości zadanej wielkości regulowanej.

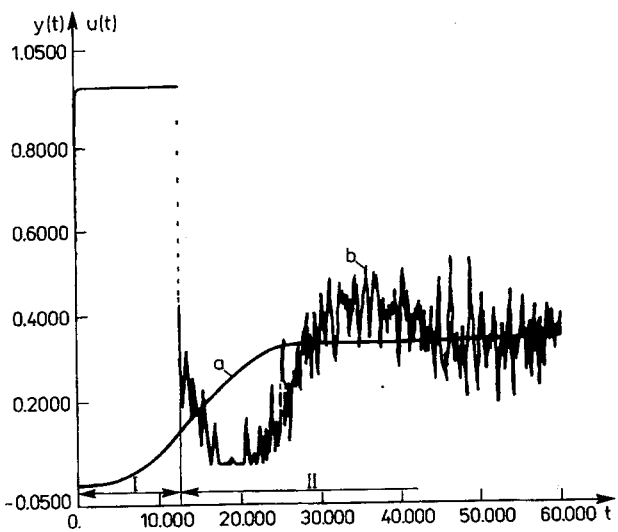
W programie symulacyjnym wykorzystuje się serię pomiarów zgodnie z ideą [8] opisaną w p. 4 a wyniki obliczeń są odpowiednio uśrednione a w nastepnym etapie

identyfikacji optymalizowane tak, aby odpowiedź modelu zidentyfikowanego i przebieg „rzeczywisty" wg. próbek były sobie najbliższe wg określonego kryterium [9]. Na Rys. 3 pokazano zakłócony przebieg odpowiedzi skokowej obiektu rzędu $n=3$ o stałych czasowych $T_1=12$, $T_2=10$ i $T_3=8$ i wzmocnieniu $k=1$. Wyniki identyfikacji parametrów bez optymalizacji $n=3$, $n_0=2{,}16$, $\overline{T}=14{,}404$, $k=215{,}265$ i po optymalizacji $n=3$, $\overline{T}=9{,}83$, $k=0{,}97$. W przypadku silnych zakłóceń, widoczne są efekty optymalizacji.



Rys. 3. Symulacja identyfikacji na podstawie odpowiedzi zakłóconej $n=3$; $n_0=2{,}16$; $\overline{T}=14{,}404$; $k=215{,}265$C

Na Rys. 4 pokazano przykładowo przebieg pracy układu symulacyjnego z samonastrojeniem regulatora PID, sterującego zakłócony obiekt 3 rzędu. Faza I zaznaczona na Rys. 4 stanowi fazę identyfikacji — w tym czasie sterujący jest stały, maksymalny. Krzywa a jest przebiegiem wielkości regulowanej a krzywa b ilustruje wielkość sterującą. Faza II jest fazą regulacji za pomocą regultora PID nastrojonego na podstawie identyfikacji przeprowadzonej w fazie I.



Rys. 4. Symulacja regulacji z samonastrojeniem

6. UWAGI KOŃCOWE

W pracy pokazano, że możliwe jest efektywne w praktyce oszacowanie rzędu zredukowanego $r=n-m$, modelu liniowego o transmitancji (1) oraz rzędu n procesów wieloinercyjnych lub oscylacyjnych, na podstawie odpowiedzi skokowej już na samym jej początku. Wskazano na ścisły związek tak identyfikowanego rzędu n_0 z pochodną odpowiedzi skokowej. Ponieważ prezentowana metoda oparta jest na wykorzystaniu dyskretnych wartości odpowiedzi, nadaje się ona szczególnie do zastosowań w układach dyskretnych. Wykazano, że na podstawie wartości oszacowania rzędu r_0 lub n_0, można sądzić o wielkości rozrzutu stałych czasowych i w efekcie, można te stałe czasowe oszacować.

Dzięki znajomości rzędu, dysponując dodatkowymi informacjami uzyskanymi na drodze pomiarów, w szczególności, znając wartość czasu t_{m^*}, dla którego odpowiedź identyfikowanego procesu osiąga maksymalną pochodną, można oszacować pozostałe parametry liniowych modeli. Przy wzroście rozrzutu stałych czasowych procesu maleje rząd n_0 i w szczególnie, gdy $T_{\min}\to 0$, następuje obniżenie rzędu o jeden. Ponieważ fizykalnie uzasadnione jest operowanie rzędem całkowitym, stąd zgodnie z zależnością (32), błędy metody w szacowaniu rzędu są bardzo małe, jeżeli tylko spełnione jest założenie o znajomości czystego opóźnienia identyfikowanego procesu [7]. Należy podkreślić, że istota prezentowanej metody opiera się na dyskretnym próbkowaniu co tworzy błąd δ_n, który powiększa się ze wzrostem rozrzutu gdyż wtedy maleje średnia stała czasowa $\overline{T}$. Stosunek $\Delta/\overline{T}$ winien spełniać nierówność $\Delta/\overline{T}\leqslant 0{,}5$ aby błędy odtwarzania $h(t)$ były małe.

Nalezy zauważyć, że przy wzroście rozrzutu stałych czasowych, silnie maleje średnia stała czasowa $\overline{T}$, co w konsekwencji powoduje gwałtowne (w potędzie n) obniżenie wzmocnienia. Można zatem, analogicznie jak o średniej stałej czasowej $\overline{T}$, mówić o średnim wzmocnieniu $\overline{k}$. W szczególności dla procesów wieloinercyjnych można tak interpretować słusznie podnoszone w [3] niezgodności $k=h(\infty)$ i T_z uzyskanych na podstawie zarejestrowanej odpowiedzi skokowej $h(t)$ z tymi parametrami, wyznaczonymi dla modelu (9) w drodze identyfikacji innymi metodami.

Dokładność prezentowanej metody w odniesieniu do innych parametrów niż r lub n, nie jest wysoka, ale może okazać się wystarczająca dla budowy przybliżonych modeli realnych procesów dla celów sterowania nimi. Zaletą metody może okazać się szybkość uzyskania oszacowań. Procedura identyfikacyjna może wykorzystać zamiast 1 pary próbek, serię wielu par, dających serię wynikow, co po uśrednieniu i optymalizacji może być skutecznym sposobem na minimalizację wpływu zakłóceń losowych nałożonych na przebieg próbkowanej odpowiedzi $h(t)$. Przytoczone przykłady zastosowań proponowanej metody do identyfikacji procesów elektrotermicznych dały wyniki pozytywne. Należy też podkreślić, że symulacje wykazały iż może ona być zastosowana dla celów adaptacji lub autotuningu regulatora.

BIBLIOGRAFIA

1. W. Findeisen: *Technika regulacji automatycznej*, PWN, Warszawa 1969
2. W. Łobodziński: *Wybrane zagadnienia estymacji parametrów uproszczonego modelu matematycznego pieców oporowych jako obiektów regulacji temperatury. Część 1 i 2.* Prace Przemysłowego Instytutu Elektroniki Nr 103, 1987, str. 53–81; Nr 106 1988, str. 19–64
3. M. Pachter, J.J. D'Azzo, L.E. Buzogany: *Second-order system models of high-order plants.* Int. J. Systems Sci. 1994, vol. 25 No 10 pp 1653–1662
4. S. Skoczowski: *Dwustawna regulacja temperatury.* Warszawa 1977
5. S. Skoczowski: *Einige Bemerkungen zur Approximierung von Regelstrecken mit Ausgleich.* Regelungstechnik. Vol. 31. 1983, pp. 321–234
6. S. Skoczowski: *Einfache Modelle der Temperaturregelstrecken und ihr Brauchberkeitsmass.* Proc. 10th UIE Congress, 1988, Stockholm, Sweden.
7. S. Skoczowski: *Metoda uproszczonej identyfikacji parametrów modelu Strejca na podstawie pierwszych próbek odpowiedzi skokowej.* Kwartalnik Elektroniki i Telekomunikacji, 1992, 38, z. 2/3, str. 193–303
8. S. Skoczowski: *Prosta metoda szybkiej identyfikacji rzędu modelu wybranych obiektów elektrotermicznych.* I Sympozjum Symulacje, Pomiary i Diagnostyka w Elektrotermii. Białystok 1992, Wydawnictwo Politechniki Białostockiej, str. 143–152
9. S. Skoczowski, A. Osadowski: *A simple identification method for the order of the Strejc model and its application to autotuning.* Preprints of the SICICA'94, IFAC, Budapest, Hungary, June 8–10, 1994, pp. 354–360
10. S. Skoczowski: *Estimation of the order and the spread of time constants for multi-time lag processes using step responses.* First International Symposium on Mathematical Models in Automation and Robotics. September 1–3, 1994. Międzyzdroje, Poland pp. 351–357
11. J. Sponer: *Kontinuierliche Steuerungen.* VT Berlin 1981
12. L. Trybus: *Regulatory wielofunkcyjne.* WNT, Warszawa 1993
13. N.-S. Xu: *Approximate evaluation of the sampled model parameters for a linear continuous-time system with relatively small sampling period*, Int. J. System Sci. vol. 20, 1989, pp. 811–838

S. SKOCZOWSKI

A SIMPLE METHOD FOR IDENTIFICATION OF TRANSFER FUNCTION BASED ON THE INITIAL PHASE OF STEP RESPONSES

Summary

An approximate method for parameter identification of linear model transfer function is presented. The order and consequently the remaining parameters of the model are determined from the process step response initial stage approximated by a parabola. Some test results of identification of selected electroheated control plants are presented. The method applicability to the autotuning of PID controlers is demonstrated by means of simulation techniques.

Key words: model transfer functions, process step response, electroheated control plants.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 1. ss. 31—44

Self-recovering property of convolutionally encoded continuous phase modulation

PIOTR TYCZKA

Instytut Elektroniki i Telekomunikacji, Politechnika Poznańska

WITOLD HOŁUBOWICZ

Francusko-Polska Wyższa Szkoła Nowych Technik Informatyczno-Komunikacyjnych

Received 1994.11.30

Authorized 1995.03.20

In this paper, we study convolutionally encoded Continuous Phase Modulation (CPM) schemes obtained using the matched encoding concept. We define the, so called, *self-recovering property* of a matched encoded scheme which is a useful feature from the practical point of view. It is shown that only some of the coded CPM schemes studied by us are self-recovering. The influence of the transmitter structure on the presence or absence of the self-recovering feature is discussed.

Key words: CPM schemes, convolutional codes, matched encoding technique, feedback-free modulator, transmitters.

1. INTRODUCTION

Trellis coding techniques, firstly applied to memoryless modulations such as M-ary PSK and M-ary QAM [1], have been more recently used also to modulations with memory. In particular, trellis-coded Continuous Phase Modulation (CPM) is often studied due to its constant envelope and small bandwidth requirements [2]. Combining coding with CPM allows to achieve an additional coding gain which is of practical interest, especially in satellite and mobile communication systems.

Among different encoding methods which appeared in the literature, we study one of them — the *matched convolutional encoding* technique [3]. This idea was developed

for MSK modulation [3] and then applied to Tamed Frequency Modulation (TFM) in [4, 5]. Matched encoding approach is based on the observation that for a given CPM modulator, combining with it different convolutional encoders with the same number of states may yield the optimum Viterbi receiver with different number of states. It means that the number of states in the optimum Viterbi receiver is determined not only by the number of modulator states and the number of encoder states, but also depends on the structure of the modulator and the encoder. This technique leads to defining encoders matched to a given CPM scheme in a given class of codes, i.e. codes having the same rate and constraint length. For a given modulation with memory, *matched code* is understood as a convolutional code choosen in such a way that, when combined with this modulation, the resulting receiver for the coded modulation has the smallest possible number of states. As it was shown in [5], in some cases it is even possible to find combinations which result in optimum receivers with smaller number of states than that of the uncoded modulation scheme.

The objective of this paper is to analyze the structure of the transmitter, treated as a finite-state machine, from the point of view of the sensitivity of its operation to an improper phase (or, in general, state) synchronization in the receiver. Similar problems have been earlier studied in the literature. One well-known example of a similar nature is the rotational invariance of signals, such as, for example, the trellis modulated signals (TCM) [6–8]. In that case, the codes and the modulated signals have been selected while keeping in mind the phase ambiguity typically present at the output of the carrier recovery circuit in the receiver. In the context of convolutionally encoded CPM signals, such as those analyzed in our paper, it has been earlier observed [4, 5] that matched coding allows the receiver to consider during the reception process only one part (usually 25% or 50%) of the full set of states[1] and not all of the as it is the case in a more traditional receiver. In order to exploit this feature in the receiver, the transmitter must be correctly initialized, so it must start its operation from one of them, so called, "basic states" (defined in the sequel of this paper). In this paper, we examine the situation when this initial synchronization has not been achieved. It is shown that two types of behaviour of the transmitter are possible. Either it automatically returns to one of the basic states or it permanently remains outside the set of basic states. In the former case, the scheme is called *self-recovering*. We analyze this phenomenon for convolutionally encoded CPM schemes with or without feedback and we show that the self-recovering feature is inherently dependent on the structure of the transmitter and is independent of the set of transmitted signals.

The outline of this paper is the following. Section 2 describes the coded CPM system analyzed and the concept of matched codes. Next Section defines the self-recovering property and discusses it in the context of a feedback-free CPM modulator. Section 4 is devoted to the problem of concatenation of an equivalent convolutional encoder with a modulator having feedback. A more general discussion of conditions required for the self-recovering property in different transmitter structures is conducted in Section 5. The last Section contains conclusions.

[1] The *full set of states* is understood here as the Cartesian product of the set of modulator states and the set of encoder states.

2. MATCHED ENCODING OF CPM

Let us consider a coded communication system with CPM. Throughout the paper we will assume TFM [9] as an example of a CPM scheme. Fig. 1 shows the block diagram of this system. **G** is a rate k/l binary convolutional encoder generating the code C_G. At time t, a k-tuple of bits $\mathbf{a}_t=(a_t^{(1)}, ..., a_t^{(k)})$ produces in l-tuple of bits $\mathbf{b}_t=(b_t^{(1)},...,b_t^{(1)})$, called a branch in the trellis of C_G. The code word $\mathbf{b}_t$ is serially transmitted to a TFM modulator which consists of a 3-tap transversal filter with the transfer function $T(D)=(1+D)^2/4$, a pulse shaping filter and a VCO of the continuous-phase type [5]. The TFM modulator can be represented as an 8-state sequential machine shown in Fig. 2 [5]. States are labeled by the excess-phase $\theta(t)$ values at the symbol transitions viewed modulo-2π and $\theta(t)$ is measured against the frequency corresponding to data sequence ...,−1,−1,−1,... (i.e. not against the center frequency). The phase takes on one of the six values: $\{0, \pm\pi/4, \pm 3\pi/4, \pi\}$. In two cases, when the same value of phase corresponds to two different states, the states have been labeled 0 and 0′, π and π'.

While trying to find good combinations of a convolutional code and CPM modulation, we must avoid catastrophic combinations which would result in a catastrophic combined state diagram. This issue is related to the presence of feedback in the CPM modulator [3]. Taking into account that the TFM modulator of Fig. 2 contains feedback and knowing that a non-catastrophic encoder combined with a feedback-free modulator always results in a non-catastrophic combined scheme, we can introduce precoding to the TFM modulator (Fig. 3(a)) [5]. In this way we obtain an equivalent feedback-free TFM modulator **M′**. The binary precoder **P** is described

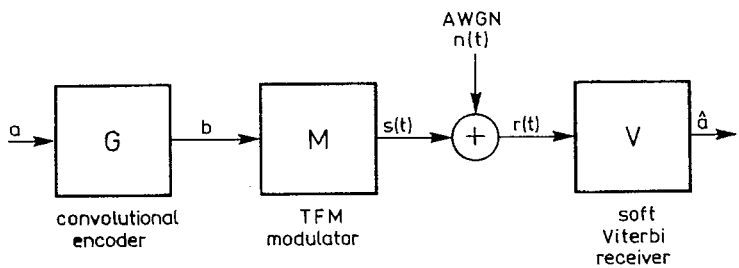


Fig. 1. Convolutionally encoded TFM system model

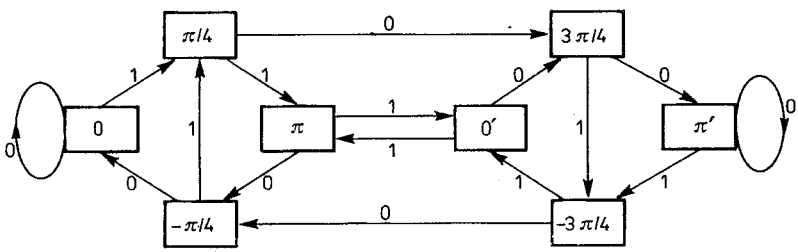


Fig. 2. The finite-state model of the (nonprecoded) TFM modulator

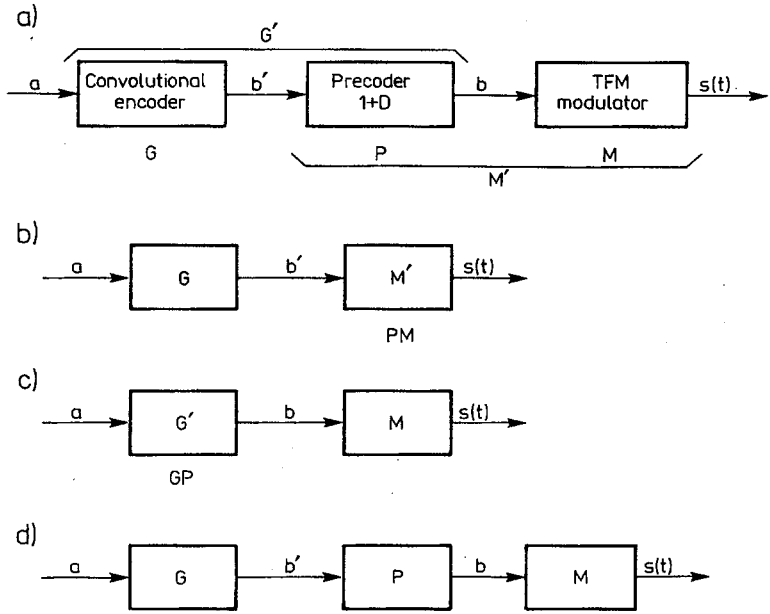


Fig. 3. a) Combined coding and precoded TFM system model
b) Convolutionally encoded feedback-free TFM modulator
c) Convolutionally encoded nonprecoded TFM modulator
d) Convolutionally encoded precoded TFM modulator

by the transfer function $T(D)=1+D$. The state diagram of the precoded TFM modulator **M'** is shown in Fig. 4.

In [3] a matched encoder for a given modulator **M** was defined as any rate-k/l constraint length-v convolutional encoder which corresponds to a combined trellis with the minimum possible number of states for that modulator, that code rate and that constraint length. For TFM, matched codes correspond to a combined trellis of $2S_G$ states [5]. An example of a matched code for the precoded TFM modulator **M'** of Fig. 4 is shown below.

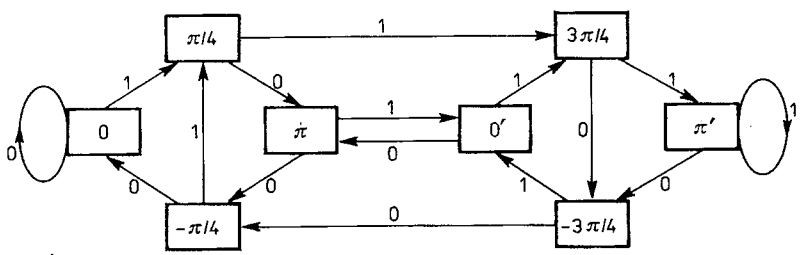


Fig. 4. The finite-state model of the precoded TFM modulator

Example 1:

Consider the rate-1/2 constraint length $v=1$ convolutional encoder given by:

$$\mathbf{G}=[1+D1]. \tag{1}$$

Let us define concatenated machine (encoder-modulator) state as an ordered pair $\sigma_V = (\sigma_G \sigma_{M'})$ where σ_G is the encoder state and $\sigma_{M'}$ is the modulator state. Let us assume now that our concatenated machine starts from a certain state, for instance, the all-zero state (0,0), and let us determine which combined states may this machine reach when the input sequence to the encoder ranges through all possible binary sequences (Fig. 5(a)). We see that even though the inputs range through all arbitrarily log binary sequences, only 4 states may be reached from the state (0,0). If we assume that the receiver knows the starting state of the transmitter, then the receiver needs only to distinguish among 4 states. The 4-state trellis diagram used by the receiver is shown in Fig. 5 (b). The four states shown in Fig. 5 are called the basic states[2] of the transmitter.

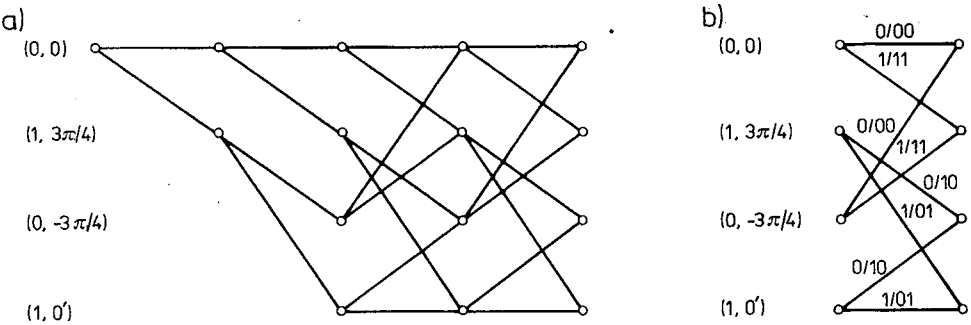


Fig. 5.a) Construction of the combined trellis for the scheme of Example 1.
b) Combined trellis for the coded signal of Example 1

3. SELF-RECOVERING PROPERTY

Let us consider again a concatenation of the code of Example 1 and the feedback-free TFM modulator described by the state diagram of Fig. 4 (Fig. 3(b)). We have seen that the reduction of the number of states reached via the matched encoding is obtained when a fixed (e.g. all-zero) initial state is assumed and under this assumption the receiver may operate only on a reduced set of basic states. In general however, there are $S_G \cdot S_{M'} = 2 \cdot 8 = 16$ possible states[3] of the concatenated machine since the set of states is the Cartesian product of all the encoder's states and all the modulator's states. The list of these 16 states is given in Table 1.

[2] A basic state in this paper is any state which is reachable from the all-zero state of the concatenated machine, assuming correct operation of this machine.

[3] A state is understood here as a minimum information concerning the past behaviour of the machine, which together with current and future inputs allow us to fully predict the future behaviour of the machine. Here, we do not make any assumptions about the initial state of the operation of the machine.

Table I

Full list of states for the combined trellis of Example 1

$\sigma_V=(\sigma_G\,\sigma_{M'})$	$\sigma_V=(\sigma_G\,\sigma_{M'})$	$\sigma_V=(\sigma_G\,\sigma_{M'})$	$\sigma_V=(\sigma_G\,\sigma_{M'})$
A=(0,0)	E=(1, π')	I=(0, 0')	M=(1, π)
B=(1, $3\pi/4$)	F=(1, $\pi/4$)	J=(0, $\pi/4$)	N=(1, $-3\pi/4$)
C=(0, $-3\pi/4$)	G=(0, $-\pi/4$)	K=(0, π')	P=(1, 0)
D=(1, 0')	H=(0, π)	L=(0, $3\pi/4$)	R=(1, $-\pi/4$)

Let us consider now a situation when the transmitter, due to instantaneous incorrect operation (or lack of initial synchronization), enters a state not indicated in Fig. 5(b), i.e. one of the remaining 12 states listed in Table 1. In the general case such a situation would be equivalent to a loss of synchronization in the receiver and would require a standard resynchronization procedure which, for example, would include bringing the transmitter to the all-zero state. We constructed a full 16-state transition diagram for this scheme, which is shown in Fig. 6. The 4 basic states from Fig. 5(b) are drawn with solid lines whereas the remaining 12 states from Table 1 are drawn using dashed lines. We find that all these 12 states are unstable, i.e. once in such a state, the transmitter automatically returns to one of the stable (i.e. recognized by the receiver) states after only at most two steps, causing at most two errors at the receiving side. We called this feature of a coded scheme the *self-recovering* property.

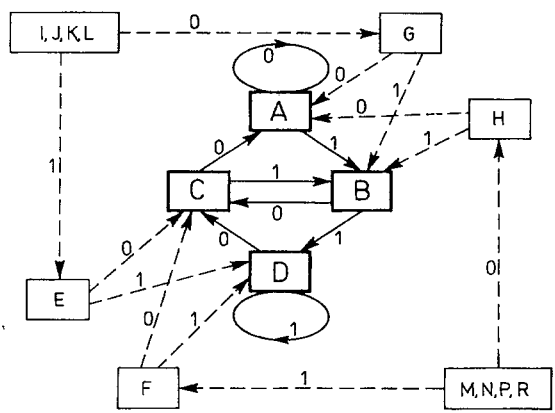


Fig. 6. Full 16-state transition diagram for the scheme of Example 1

We will show now that the self-recovering property is common to all coded schemes where both the encoder and the modulator are feedback-free. Our proof is based on Fig. 7 which shows a concatenation of an encoder **G** and a feedback-free modulator **M'**. Since **M'** is feedback-free it can be represented (like encoder **G**) as a shift register followed by a combinational circuit generating the output signal $s(t)$. Suppose now that the shift register of **G** is of length v and the shift register of **M'** is of length $v_{M'}$ and consider any state $(\sigma_G, \sigma_{M'})$. It is clear that the encoder **G** may reach any arbitrary state σ_G whereas the modulator **M'** states are determined by the previous

outputs **b** of the encoder. The reduction of the set of states, which is achieved when matched encoding is applied, occurs when some combinations (σ_G $\sigma_{M'}$) are missing in the list of basic states. Note that the non-basic states correspond to the improper choice of the state of **M'** for a given arbitrary state σ_G. Therefore, noting that $\sigma_{M'}$ is only depending on the sequence **b** and that any state σ_G is allowed in the description of basic states, once in a non-basic state, the concatenated machine **GM'** will return to one of the basic states after at most $\nu_{M'}$ bits **b** generated by the encoder. Hence, all non-basic states are unstable states[4]. To illustrate this general behaviour let us examine again the scheme of Example 1.

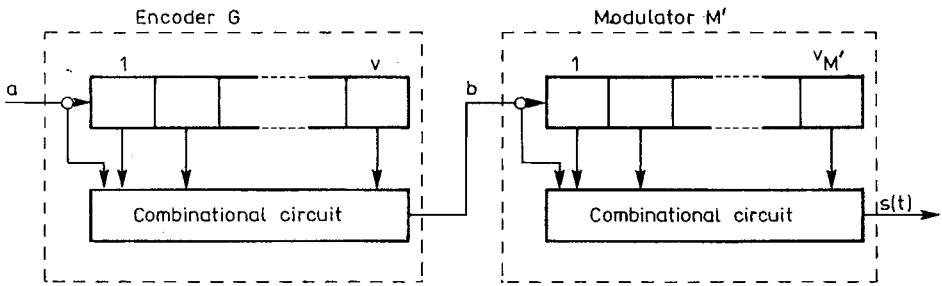


Fig. 7. Model of the concatenation of a convolutional encoder **G** and a feedback-free modulator **M'**

For the modulator **M'** of Fig. 4 any of its states $\sigma_{M'}$ is reached only by the paths which correspond to input sequences **b** having the same last three bits $b_{i-1}b_{i-2}b_{i-3}$ (e.g. $b_{i-1}b_{i-2}b_{i-3}=101$ for state 0'). Therefore, for the feedback-free TFM modulator the length of the shift register of Fig. 7 is $\nu_{M'}=3$ and its contents defines explicitly the modulator state. Hence, 4 receiver states of Example 1 correspond to 4 pairs ($\sigma_G\sigma_{M'}$) where $\sigma_{M'}$ represents here the contents of shift register of modulator **M'**. These pairs correspond to basic states and are listed in Table 2. The remaining 12 states are non-basic, and since

Table II

States recognized by the receiver for the combined trellis of Example 1

State	σ_G	$\sigma_{M'}=b_{i-1}b_{i-2}b_{i-3}$
A	0	000
B	1	110
C	0	011
D	1	101

M' is feedback-free, they are also non-stable. If the transmitter of Fig. 7 entered such a non-stable state then any input sequence **a** will cause that after 1 or 2 time intervals the contents of shift registers of **G** and **M'** will be one out of the pairs listed in Table 2 (Fig. 6). This corresponds to returning of the transmitter to its correct operation.

[4] A stable state in this paper is understood as one of a set of states, where (i.e. in this state set) the concatenated machine may stay for an arbitrarily long time period, while being able to reach any of these states.

In general, if the transmitter for a coded scheme is self-recovering it is not necessary to apply a resynchronization procedure in the case of transmitter operation error, in order to recover correct operation. It seems therefore that the self-recovering property might be an attractive feature for practical implementations and might increase the robustness of the system.

4. PERFORMANCE OF CONVOLUTIONALLY CODED FEEDBACK CPM

Search for the best codes for use with TFM made in [5] was conducted for concatenations **GM'** (Fig. 3(b)). While trying to avoid the catastrophic **GM'** combinations, only non-catastrophic encoders **G** were considered. It was assumed there that an encoder **G** matched to the precoded (feedback-free) TFM can be converted into a corresponding encoder **G'** which is matched to the non-precoded TFM. The conversion can be done by merging the encoder **G** with the precoder $1+D$ (Fig. 3(c)). For example, for rate-1/2 convolutional encoder **G** the generator polynomials $g'_{ij}(D)$ of the corresponding encoder **G'** are related to the generator polynomials $g_{ij}(D)$ of **G** in the following way:

$$\begin{aligned} g'_{11}(D) &= g_{12}(D)\cdot D \oplus g_{11}(D) \\ g'_{12}(D) &= g_{11}(D) \oplus g_{12}(D), \end{aligned} \tag{2}$$

where $\oplus$ denotes modulo-2 addition. Similar relations can be derived for other code rates. It should be noted here that the conversion of the rate-1/2 matched codes **G** into **G'** does not change the constraint length of the code. It is a consequence of the following structural conditions true for any rate-1/2 encoder matched to the machine of Fig. 4:

$$\begin{aligned} &\deg(g_{12}(D)) \leqslant v-1, \quad \text{if } v \geqslant 1, \\ &\deg(g_{12}(D)) = -\infty, \quad \text{if } v=0. \end{aligned} \tag{3}$$

This issue is described in detail in [5].

Table 3 shows all non-catastrophic rate-1/2 constraint length $v=1$ codes **G** matched to precoded TFM (**M'**) and their corresponding codes **G'**, matched to feedback

Table 3

Rate-1/2 constraint length $v=1$ (non-catastrophic) matched codes **G** for use with the precoded TFM and the corresponding codes **G'** for use with non-precoded TFM

G	G'
[1+D 1] [D 1] [D 0]	[1 D] [0 1+D](#) [D D]

(#) — catastrophic code

TFM (**M**). The number of basic states in the combined trellis (hence, the number of states in the Viterbi receiver) is here $S_V=4$. Clearly, codes **G'** have the same constraint length v as codes **G**. Mismatched codes **G** (for which $S_V=8$) do not meet conditions (3) and hence, after conversion the resulting codes **G'** have $v'=v+1$. However, these codes **G'** become matched codes to feedback TFM in the class of codes with constraint length $v'=v+1$. The mismatched codes **G** of class ($R=1/2$, $v=1$) and the corresponding codes **G'** are shown in Table 4.

Table 4

Rate-1/2 constraint length $v=1$ (non-catastrophic) mismatched codes **G** for use with the precoded TFM and the corresponding codes **G'** for use with non-precoded TFM

G	G'
[1+D D]	[1+D+D² 1]
[D 1+D]	[D² 1]
[D D]	[D+D² 0]
[1 1+D]	[1+D+D² D]
[1 D]	[1+D² 1+D]
[0 D]	[D² D]

Since matched codes **G'** have the same v as codes **G**, they have the same number of states. In addition, the conversion does not change the basic states of the concatenated machine, i.e. they are described by pairs $(\sigma_G\,\sigma_M)$ which are the same as pairs $(\sigma_G\,\sigma_{M'})$ in the scheme with precoded TFM. The behaviour of concatenation **G'M** in the situation of incorrect operatoin will be illustrated by the following Example.

Example 2:

Consider the rate-1/2 constraint length $v=1$ convolutional encoder given by:

$$\mathbf{G'}=[1D]. \tag{4}$$

This encoder **G'** is equal to the encoder **G** of Example 1 followed by the precoder **P**. The overall 16-state diagram of the concatenation of **G'** with non-precoded TFM modulator of Fig. 2 is drawn in Fig. 8 and states are listed in Table 5. It is seen that the diagram consists of two disjoint subdiagrams and hence, the self-recovering property is not preserved here. The upper subdiagram contains 4 basic receiver states (the only states reachable in normal operation) and exhibits self-recovering property. In this case, the four non-basic states J, K, M, R are stable. If the transmitter entered any

Table 5

Full list of states for the combined trellis of Example 2

$\sigma_V=(\sigma_{G'}\sigma_M)$	$\sigma_V=(\sigma_{G'}\sigma_M)$	$\sigma_V=(\sigma_{G'}\sigma_M)$	$\sigma_V=(\sigma_{G'}\sigma_M)$
A=(0,0)	E=(1, π')	I=(0, 0')	M=(1, π)
B=(1, $3\pi/4$)	F=(1, $\pi/4$)	J=(0, $\pi/4$)	N=(1, $-3\pi/4$)
C=(0, $-3\pi/4$)	G=(0, $-\pi/4$)	K=(0, π')	P=(1, 0)
D=(1, 0')	H=(0, π)	L=(0, $3\pi/4$)	R=(1, $-\pi/4$)

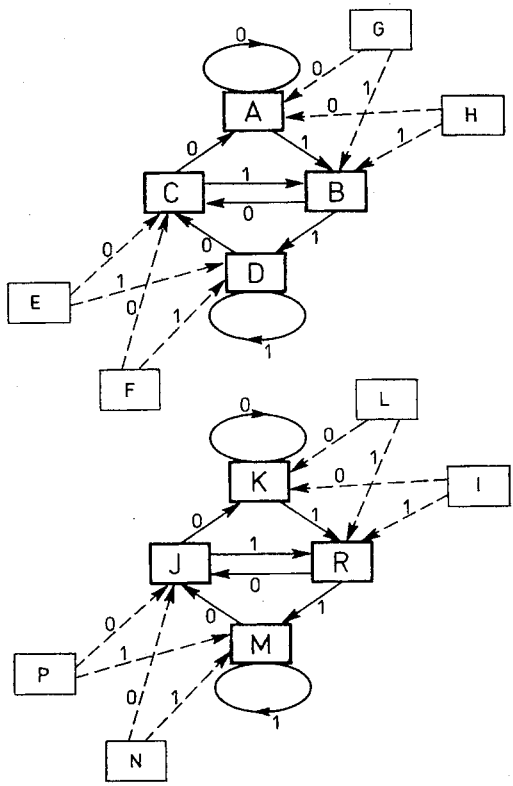


Fig. 8. Full 16-state transition diagram for the scheme of Example 2

state of the lower subdiagram then at the receiving side an infinitely long sequence of errors would occur. In particular, the lower subdiagram is a shifted by π version of the upper one. This would result in an output phase of a CPM transmitted signal shifted by π and therefore means a sign ambiguity.

Note that Examples 1 and 2 correspond to the same signal space code (signal constellation) and differ only in the transmitter representation (Fig. 3(b) vs. Fig 3(c)). Therefore, it is clear from the above examples that the self-recovering property is not determined by the signal space code but does depend on the transmitter structure. This dependency is studied in the next Section.

5. SELF-RECOVERING PROPERTY FROM THE VIEWPOINT OF THE TRANSMITTER STRUCTURE

Let us consider the precoded TFM modulator combined with the encoder **G** (Fig. 3(d)). The transmitter consists of the matched encoder **G** the precoder **P** (with the transfer function $T(D)=(1+D)^2/4$) and the feedback TFM modulator **M**. The total number of possible states of this scheme, defined as ordered triplets $(\sigma_G\ \sigma_P\ \sigma_M)$, is

$S_G \cdot S_P \cdot S_M$ which gives for the case of Examples 1 and 2 $2 \cdot 2 \cdot 8 = 32$. However, only 4 of them (the basic states) are used during correct operation. Construction of a full 32-state diagram of the transmitter is shown in Fig. 9. States which correspond to the same pairs of encoder and modulator states (σ_G σ_M) but differ only in the precoder state σ_P are labeled with a prime (e.g. the states $\mathbf{F}=(1,1,\pi/4)$ and $\mathbf{F}'=(1,0,\pi/4)$). Full list of states for the diagram of Fig. 9 is given in Table 6.

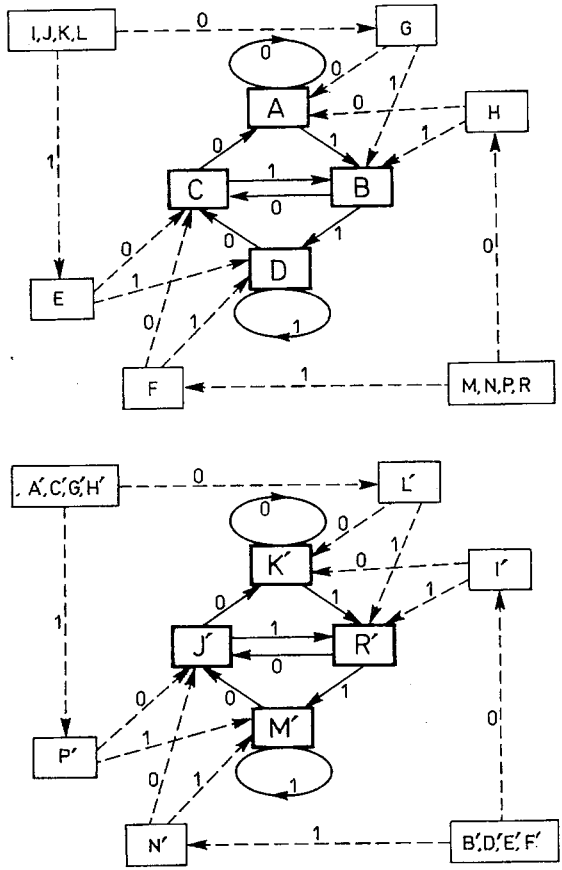


Fig. 9. Full 32-state transition diagram for the scheme of Fig. 3(d)

Table 6

Full list of states for the modulator of Fig. 3(d)

$\sigma_V=(\sigma_G\,\sigma_P\,\sigma_M)$	$\sigma_V=(\sigma_G\,\sigma_P\,\sigma_M)$	$\sigma_V=(\sigma_G\,\sigma_P\,\sigma_M)$	$\sigma_V=(\sigma_G\,\sigma_P\,\sigma_M)$
$A=(0,0,0)$	$E=(1,\ 1,\ \pi')$	$I=(0,\ 1,\ 0')$	$M=(1,\ 0,\ \pi)$
$A'=(0,\ 1,\ 0)$	$E'=(1,\ 0,\ \pi')$	$I'=(0,\ 0,\ 0')$	$M'=(1,\ 1,\ \pi)$
$B=(1,\ 1,\ 3\pi/4)$	$F=(1,\ 1,\ \pi/4)$	$J=(0,\ 1,\ \pi/4)$	$N=(1,\ 0,\ -3\pi/4)$
$B'=(1,\ 0,\ 3\pi/4)$	$F'=(1,\ 0,\ \pi/4)$	$J'=(0,\ 0,\ \pi/4)$	$N'=(1,\ 1,\ -3\pi/4)$
$C=(0,\ 0,\ -3\pi/4)$	$G=(0,\ 0,\ -\pi/4)$	$K=(0,\ 1,\ \pi')$	$P=(1,\ 0,\ 0)$
$C'=(0,\ 1,\ -3\pi/4)$	$G'=(0,\ 1,\ -\pi/4)$	$K'=(0,\ 0,\ \pi')$	$P'=(1,\ 1,\ 0)$
$D=(1,\ 1,\ 0')$	$H=(0,\ 0,\ \pi)$	$L=(0,\ 1,\ 3\pi/4)$	$R=(1,\ 0,\ -\pi/4)$
$D'=(1,\ 0,\ 0')$	$H'=(0,\ 1,\ \pi)$	$L'=(0,\ 0,\ 3\pi/4)$	$R'=(1,\ 1,\ -\pi/4)$

In Fig. 9 two disjoint 16-state diagrams appear which have identical structures. The upper diagram contains the 4 basic and stable states A, B, C, D and is self-recovering. It is easily seen that the diagram of Fig. 6 is a subdiagram of the overall diagram of Fig. 9 (the upper part of it). Hence, the upper part of the diagram of Fig. 9 represents the concatenation of the feedback-free modulator **M'** with the encoder **G**. The explanation for the existence of the lower part of this diagram can be given in the following way. A feedback-free TFM modulator of Fig. 4 was obtained by concatenation of the precoder $1+D$ and a standard (non-precoded) TFM modulator under assumption of zero state both in precoder and modulator at the starting point. This assumption leads to the situation where for a given modulator state we always have one and only one precoder state. Considering incorrect operation of the transmitter of Fig. 3(d) we may take into account also situations when the precoder state is *"forbidden"* for a given modulator state. Such combinations lead to the creation of the lower, disjoint diagram. Therefore, if the transmitter enters one of the states of the lower subdiagram, it will not recover to any of the basic states. The two subdiagrams are shifted by π in the phase of their corresponding states.

Let us compare now the diagram of Example 2 (Fig. 8) with the diagram of Fig. 9. It is seen that the diagram of the concatenation **G'M** shows similarity to the overall diagram of the **GPM** scheme and is a reduced version of that diagram. The reduction (of 16 states in comparison to Fig. 9) is due to the absence here of the precoder **P** and its states. The remaining states are described by the same pairs of encoder and modulator states. The self-recovering property is no longer present in the **G'M** scheme. This is a result of a feedback in modulator **M**, which cannot be represented as a shift register followed by a combinational circuit and consequently, it is not self-recovering. However, during correct operation the **G'M** scheme is equivalent to the **GM'** scheme.

From the above discussion it follows that the scheme **GPM** and its state diagram of Fig. 9 can be treated as a general case of such concatenations as **G'M** and **GM'**. It should be also pointed out that although the three schemes presented behave exactly in the same manner during the correct operation (operation on the 4-state basic trellis of Fig. 5(b)), only **GM'** scheme is self-recovering and therefore, able to return to the proper trellis.

As a conclusion, it is the precoder-modulator interface **P-M** in Fig. 3 (and not the **G-P** interface) which is directly responsible for the fact that the resulting scheme may be non-self-recovering. All the self-recovering schemes studied by us so far turned out to be based on the correct operation of the **M'=PM** pair and allowed only for incorrect behaviour on the **G-P** interface. Therefore, to avoid schemes which are not self-recovering, the best way seems to be to use directly a feedback-free equivalent CPM modulator **M'**, as shown in Fig. 4.

CONCLUSIONS

In this paper the problem of, so called, self-recovering properties of convolutionally coded (via matched encoding technique) CPM schemes was investigated. It turned

out that these schemes may in some cases be self-recovering with respect to the receiver-known states after entering by the transmitter an incorrect state.

Although the above considerations have been conducted on the base of only a few examples of rate-1/2 matched encoded TFM, they seen to be easily generalized to other CPM systems. The conclusion that can be drawn from our analysis is that the self-recovering of the scheme does not depend on the signal space code but rather on the inner structure of the transmitter. The self-recovering property appears in schemes where the modulator is feedback-free. If the feedback-free modulator is realized using the precoder, the scheme will only be self-recovering under the assumption of correct operation both of the precoder and the modulator. Coding of a feedback modulator does not preserve the self-recovering property of the scheme.

Since the self-recovering property strongly depends on the transmitter structure chosen, it seems that design of a feedback-free modulator is the best solution to this problem. However, the remaining schemes may also be suitable if the situation of incorrect operation of the transmitter would be recognized by the receiver. This would allow, for example, to switch the receiver to the alternative trellis. The price which would have to be paid in this case is additional controlling and complexity of the receiver.

This paper was presented in part at 2nd International Symposium on Communication Theory and Applications, Ambleside, UK, 11 – 16 July 1993.

REFERENCES

1. G. Ungerboeck: *Channel coding with multilevel/phase signals.* IEEE Trans. Inform. Theory, Jan. 1982, vol. IT-28, pp. 55 – 67
2. J.B. Anderson, T. Aulin, C.-E. Sundberg: *Digital Phase Modulation.* New York: Plenum Press, 1986
3. F. Morales-Moreno, S. Pasupathy: *Structure, optimization and realization of FFSK trellis codes.* IEEE Trans. Inform. Theory, July 1988, vol. IT-34, pp. 730-751
4. F. Morales-Moreno, W. Hołubowicz, S. Pasupathy: *Performance of trellis coded tamed frequency modulation.* Proc. 38th IEEE Vehicular Techn. Conf., Philadelphia, USA, June 1988, pp. 12 – 17
5. F. Morales-Moreno, W. Hołubowicz, S. Pasupathy: *Optimization of trellis coded TFM via matched codes.* IEEE Trans. Commun., Feb./March/April 1994, vol. 42, pp. 1586 – 1594
6. L.-F. Wei: *Rotationally invariant convolutional channel coding with expanded signal space. Part I: 180°.* IEEE J. Select. Areas Commun., Sept. 1984, vol. SAC-2, pp. 659 – 671
7. L.-F. Wei: *Rotationally invariant convolutional channel coding with expanded signal space. Part II: Nonlinear codes.* IEEE J. Select. Areas Commun., Sept. 1984, vol. SAC-2, pp. 672 – 686
8. Z.C. Zhu, A.P. Clark: *Rotationally invariant coded PSK signals.* IEE Proc., Part F, Feb. 1987, vol. 134, pp. 43 – 52
9. F. de Jager and C.B. Dekker: *Tamed frequency modulation, a novel method to achieve spectrum economy in digital transmission.* IEEE Trans. Commun., May 1978, vol. COM-26, pp. 534 – 542

P. TYCZKA, W. HOŁUBOWICZ

SAMOSYNCHRONIZACYJNE WŁAŚCIWOŚCI KODOWANYCH SPLOTOWO SYSTEMÓW Z CIĄGŁĄ MODULACJĄ FAZY

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

Artykuł poświęcony jest analizie własności układów połączenia kodera splotowego z modulatorem sygnałów z ciągłą modulacją fazy, otrzymywanych w wyniku zastosowania koncepcji kodowania splotowego dopasowanego. W artykule rozważono zagadnienie tzw. *samosynchronizacji* układów nadawczych kodowanych splotowo sygnałów CPM, która to cecha może być przydatna w praktycznej realizacji systemu. Przedyskutowano wpływ struktury nadajnika na występowanie jego zdolności samosynchronizujących.

Słowa kluczowe: kodery, modulatory sygnałów.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 1. ss. 45–68

Nowa metoda identyfikacji parametrów układu dynamicznego za pomocą sztucznej sieci neuronowej

ANDRZEJ MATERKA

Instytut Elektroniki, Politechnika Łódzka

Otrzymano 1994.11.08

Autoryzowano do druku 1995.01.15

W artykule zaproponowano użycie sztucznej sieci neuronowej typu „feedforward" do identyfikacji parametrów układów dynamicznych. Proponowana metoda nie wymaga prowadzenia obliczeń iteracyjnych i jest szczególnie przydatna do identyfikacji w czasie rzeczywistym. Przeprowadzono analizę błędów identyfikacji wywołanych zakłóceniami losowymi. Wykazano, że wariancja błędu jest nie większa od wariancji uzyskiwanej w przypadku rozwiązywania zagadnienia identyfikacji za pomocą standardowej metody najmniejszych kwadratów. Ponadto nowa metoda oferuje możliwość zmniejszenia wariancji błędu poprzez odpowiednią modyfikację algorytmu uczenia sieci. Przedstawiono wyniki symulacji komputerowej i pomiarów będące potwierdzeniem słuszności wniosków z analizy teoretycznej.

Słowa kluczowe: modelowanie średniokwadratowe, układy dynamiczne, sztuczne sieci neuronowe.

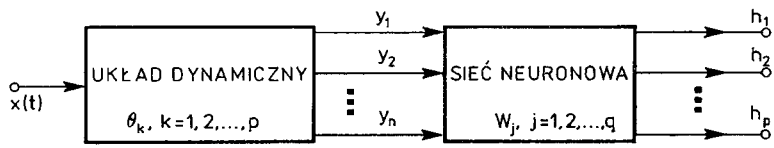
1. WSTĘP

Modelowanie, czyli poszukiwanie układu stosownych równań matematycznych do reprezentacji danego układu fizycznego jest podstawowym zagadnieniem teoretycznym w każdej prawie dziedzinie techniki oraz nauk stosowanych. Niezależnie od dziedziny zastosowania, podstawowe znaczenie dla praktyki modelowania ma identyfikacja parametrów modelu dla danego układu. W procesie identyfikacji wyznacza się wartości parametrów na podstawie pomiaru odpowiedzi układu na określone wymuszenie. Uzyskane wartości służą do diagnostyki układu.

W większości przypadków procedura identyfikacji polega na takim doborze parametrów, aby błąd między obliczoną odpowiedzią modelu i zmierzoną odpowiedzią układu na dane wymuszenie osiągnął wartość minimalną w sensie

przyjętego kryterium [1]. Sprowadza się to do numerycznego rozwiązania problemu minimalizacji funkcji wielu zmiennych. Najczęściej funkcja taka ma postać sumy kwadratów błędu a modelowanie nosi nazwę średniokwadratowego [2]. Odpowiedź układu jest na ogół nieliniową funkcją jego parametrów, nawet dla układów liniowych. W wyniku tego, identyfikacja jest procesem iteracyjnym, którego zbieżność zależy od wyboru punktu startowego i poziomu zakłóceń. Długi czas obliczeń powoduje, że identyfikacja parametrów rzadko może być realizowana w czasie rzeczywistym.

W niniejszej pracy wykazano, że identyfikacja parametrów modeli może być realizowana w czasie rzeczywistym przy zastosowaniu sztucznej sieci neuronowej. Istotę proponowanego rozwiązania zilustrowano na rys. 1. Identyfikacja polega na dołączeniu wektora próbek odpowiedzi $y(t)$ układu do wejść sieci neuronowej. Sygnał na wyjściach sieci jest estymatą wektora szukanych parametrów układu. W przypadku użycia sieci typu „*feedforward*" czas identyfikacji jest równy czasowi propagacji sygnału z wejścia sieci do jej wyjścia.



Rys. 1. Identyfikacja parametrów układu dynamicznego za pomocą sztucznej sieci neuronowej: $x(t)$ — wymuszenie, $\boldsymbol{y}=[y_1,y_2,...,y_n]^T$ — próbki odpowiedzi układu (obserwacje), $\theta=[\theta_1,\theta_2,...,\theta_p]^T$ — nieznane parametry, $\boldsymbol{h}=[h_1,h_2,...,h_p]^T$ — estymaty parametrów układu, $\boldsymbol{w}=[w_1, w_2,...,w_q]^T$ — wektor wagowy

Użyteczność proponowanej metody identyfikacji wykazano za pomocą symulacji komputerowej na przykładach układów rzędu drugiego [3] i układów nieliniowych z opóźnieniem [4]. (Wiadomo, że identyfikacja czasu opóźnienia odpowiedzi układu należy do szczególnie trudnych i czasochłonnych zadań modelowania metodą tradycyjną [1, 5]). Udowodniono też, że metoda może być zastosowana do pomiaru położenia obiektów w obrazie cyfrowym ze zwiększoną dokładnością (subpixel accuracy) [6], do diagnostyki uszkodzonych elementów w analogowych układach elektronicznych [7, 8] i do rozplotowej restauracji sygnałów [9]. Celem obecnych prac, zanim dojdzie do opracowania metod syntezy i projektowania sieci neuronowej, jest jej optymalizacja tak w sensie architektury jak i algorytmów uczenia.

2. IDENTYFIKACJA PARAMETRÓW UKŁADU DYNAMICZNEGO METODĄ NAJMNIEJSZYCH KWADRATÓW

Weźmy pod uwagę układ dynamiczny opisany modelem parametrycznym. Model taki może mieć postać układu równań różniczkowych, algebraicznych i/lub różnicowych z odpowiednimi warunkami początkowymi i brzegowymi. Współczynniki

równań tworzą p-wymiarowy wektor parametrów θ, który należy do $\Theta \subset \mathcal{R}p$. Niektóre elementy wektora θ mogą opisywać właściwości sygnału wymuszającego $x(t)$. Zmienna niezależna t reprezentuje czas, odległość, długość fali, częstotliwość, itp., stosownie do natury układu dynamicznego i badanych zjawisk fizycznych. Dla uproszczenia zostaną rozważone układy z jednym wejściem i jednym wyjściem.

Przypuśćmy, że dysponujemy n obserwacjami (t_i, y_i), $i=1,2,...,n$ odpowiedzi układu opisanej znaną z założenia zależnością funkcyjną $f(.)$:

$$y_i = f(t_i, \theta^*) + \varepsilon_i, \qquad i = 1, 2, ..., n \tag{1}$$

gdzie ε_i reprezentuje zakłócenia a θ^* jest prawdziwą wartością wektora parametrów θ. Problem identyfikacji parametrów polega na estymacji θ^* dla danego zbioru obserwacji (1), przy założeniu, że parametry układu nie zmieniają swych wartości w przedziale $[t_1, t_n]$. Najczęściej poszukuje się rozwiązania problemu identyfikacji za pomocą estymatora średniokwadratowego $\hat{\theta}$. Estymator taki minimalizuje sumę kwadratów błędu między odpowiedzią systemu y i odpowiedzią modelu f

$$S(\theta) = \sum_{i=1}^{n} [y_i - f(t_i, \theta)]^2 \tag{2}$$

dla $\theta \in \Theta$. W wielu pracach, między innymi w [10], wykazano przy następujących założeniach:

(a) zmienne losowe ε_i są niezależne oraz mają identyczne rozkłady z zerową wartością średnią i wariancją $\sigma^2 > 0$,

(b) dla każdego i, $f_i(\theta) = f(t_i, \theta)$ jest ciągłą funkcją θ dla $\theta \in \Theta$,

(c) Θ jest domkniętym podzbiorem $\mathcal{R}^p$,

że estymator $\hat{\theta}$ istnieje. Jeśli zatem założenia (a) — (c) są spełnione, to istnieje mierzalna funkcja $\hat{\theta} = g(y_1, y_2, ..., y_n)$, która minimalizuje (2), to jest taka, że

$$\hat{\theta} = \arg [\min_{\theta \in \Theta} S(\theta)] . \tag{3}$$

Ponadto rozkład prawdopodobieństwa estymatora średniokwadratowego $\hat{\theta}$ jest asymptocznie normalny, co oznacza, że przy $n \to \infty$ rozkład ten dąży do $N_p[\theta^*, \sigma^2 (F^T F)^{-1}]$, gdzie

$$\boldsymbol{F} = \left\| \begin{array}{cccc} \frac{\partial f_1}{\partial \theta_1} & \frac{\partial f_1}{\partial \theta_2} & \cdots & \frac{\partial f_1}{\partial \theta_p} \\ \frac{\partial f_2}{\partial \theta_1} & \frac{\partial f_2}{\partial \theta_2} & \cdots & \frac{\partial f_2}{\partial \theta_p} \\ \cdot & \cdot & \cdot & \cdot \\ \frac{\partial f_n}{\partial \theta_1} & \frac{\partial f_n}{\partial \theta_2} & \cdots & \frac{\partial f_n}{\partial \theta_p} \end{array} \right\|_{\hat{\theta}} \tag{4}$$

jest macierzą wrażliwości. Macierz ta, o wymiarach $n \times p$, jest używana do zdefiniowania warunków identyfikowalności układu. Okazuje się (patrz np.[11]), że dla

jednoznacznej identyfikacji wszystkich p parametrów wektora θ jest konieczne aby

$$| \boldsymbol{F}^T\boldsymbol{F} |_{\theta^*} \neq 0 . \tag{5}$$

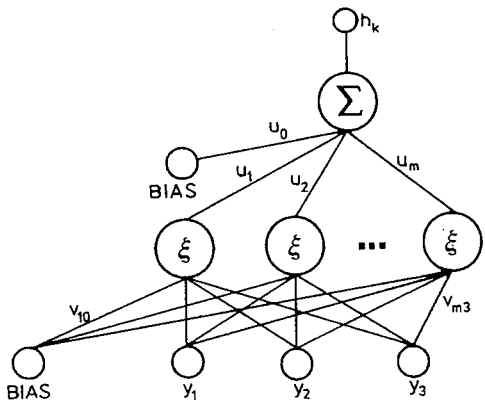
Nierówność (5) jest równoważna warunkowi, że dla układu identyfikowalnego kolumny macierzy wrażliwość $\boldsymbol{F}$ muszą być liniowo niezależne w otoczeniu $\theta^* \in \Theta$ [12]. Macierz F^TF w (5), o wymiarach $p \times p$, jest nazywana macierzą informacji [13]. Im większa jest wartość wyznacznika $| \boldsymbol{F}^T\boldsymbol{F} |$, tym większa jest informacja o nieznanych parametrach θ zawarta w danych eksperymentalnych (obserwacjach) i mniejszy wpływ zakłóceń losowych na wynik estymacji średniokwadratowej. Macierz informacji zależy od wyboru punktów pomiarowych t_i, $i=1,2,...,n$. Jest ona wykorzystywana do optymalnego planowania eksperymentu.

Wiadomo [2], że w przypadku nieliniowej funkcji $f(.)$, $S(\theta)$ może mieć wiele minimów lokalnych oprócz minimum globalnego dla $\hat{\theta}$. Jest to przyczyną trudności ze zbieżnością tradycyjnych algorytmów numerycznej minimalizacji funkcji (2) [1,2]. Ponadto minimalizacja ta, wymagająca na ogół wielu nakładów obliczeniowych, musi być realizowana na bieżąco, po każdej akwizycji wektora obserwacji układu. Z tego powodu dla wielu systemów identyfikacja metodą najmniejszych kwadratów nie może być realizowana w czasie rzeczywistym. Są to podstawowe wady stosowania tej metody do zagadnienia identyfikacji. Postuluje się [3], że zastosowanie sieci neuronowej do identyfikacji parametrów pozwala uniknąć wad tradycyjnego podejścia. Według proponowanej metody sieć neuronowa jest stosowana do aproksymacji funkcji wielu zmiennych $\theta = \boldsymbol{g}(y) = \boldsymbol{g}(y_1, y_2, ..., y_n)$, $\theta \in \Theta$. Chociaż funkcja ta nie jest zwykle znana, *przykłady* jej argumentów y_i, $i=1,2,...n$ oraz odpowiadające tym przykładom wartości θ mogą być obliczone z użyciem równań modelu $y_i = f(t_i, \theta)$ określonych wzorem (1). Ponieważ sieci neuronowe posiadają zdolność *uczenia się na przykładach* [14, 15], funkcja $\boldsymbol{g}(y)$ może zostać zrekonstruowana za pomocą odpowiedniego algorytmu uczenia. Jeśli funkcja ta jest ciągła i przedziałami jednoznaczna, to może być aproksymowana za pomocą sieci neuronowej z dowolną dokładnością, pod warunkiem odpowiedniego wyboru architektury sieci [16]. Koncepcja ta zostanie rozwinięta w dalszej części artykułu.

3. ZASTOSOWANIE SZTUCZNEJ SIECI NEURONOWEJ DO ZAGADNIENIA IDENTYFIKACJI

Rozważamy sieć neuronową zastosowaną do aproksymacji odwzorowania przestrzeni obserwacji układu dynamicznego na przestrzeni jego parametrów. Sposób połączenia sieci z układem pokazano na rys. 1. Sieć neuronowa jest w tym zastosowaniu układem statycznym bez sprzężeń zwrotnych (ang. *feedforward*). Najbardziej popularną siecią tego typu jest perceptron [14, 15]. Sieć taka może być zbudowana jako układ elektroniczny, analogowy lub cyfrowy, zawierający zestaw elementów przetwarzających (neuronów) ułożonych w warstwy. Warstwę najniższą

tworzą zaciski wejściowe sieci zaś warstwa najwyższa odpowiada jej zaciskom wyjściowym. Każdy z neuronów w warstwie wewnętrznej (inaczej ukrytej, ang. hidden layer) realizuje monotoniczną funkcję nieliniową, najczęściej sigmoidalną, np. $\xi(x)=1/(1+exp(-x))$. Sygnał wejściowy x neuronu w danej warstwie jest sumą ważoną sygnałów w węzłach warstwy bezpośrednio niższej. Istnieją zatem połączenia między warstwami, a każde z takich połączeń jest reprezentowane współczynnikiem wagowym $w_j, j=1,2,...,q$. Konkretna postać funkcji $h(y)$ realizowanej za pomocą sieci neuronowej zależy od wartości wektora wag $w=[w_1, w_2,...,w_q]$, dobieranego w procesie uczenia. Tak zwany algorytm wstecznej propagacji błędu (ang. error back propagation (BP) algorithm [14, 15]) jest prawdopodobnie najpopularniejszą metodą uczenia.



Rys. 2. Przykład sieci perceptronowej o trzech wejściach i jednym wyjściu

Przykład trójwarstwowej sieci perceptronowej o trzech wejściach i jednym wyjściu, z jedną warstwą ukrytą i jednym neuronem w warstwie wyjściowej, przedstawiono na rys. 2. Do identyfikacji układu p-parametrowego potrzeba p sieci tego rodzaju, przy czym każda z nich estymuje jeden z parametrów θ_k, $k=1,2,....,p$. Sygnał na k-tym wyjściu jest określony następującą zależnością

$$h_k(w,y)=u_0+\sum_{j=1}^{m} u_j\xi\left(v_{j0}+\sum_{i=1}^{n=3} v_{ji}y_i\right), \tag{6}$$

gdzie m jest liczbą neuronów w warstwie wewnętrznej. Całkowita liczba współczynników wagowych $w=[w_1, w_2,...,w_q]^T=[u_0, u_1,...,u_m, v_{10}, v_{11},...,v_{mn}]^T$, dostrajanych w procesie uczenia sieci z rys. 2, wynosi $q=(n+2)m+1$. W symulacji komputerowej omówionej bardziej szczegółowo w dalszej części artykułu, uczenie perceptronu polegało na minimalizacji (w przestrzeni współczyników wagowych $w\in\mathcal{R}^q$) następującej funkcji

$$P_k(w)=N^{-1}\sum_{l=1}^{N}(\theta_k^{(l)}-h_k(w,f(\theta^{(l)})))^2, \tag{7}$$

gdzie $f(\theta)=[f(t_1,\theta),...,f(t_n,\theta)]^T$ a N jest liczbą przykładów $\{f(\theta^{(i)}), \theta^{(i)}\}$ aproksymowanej zależności $g(.)$. Wielkość opisana wzorem (7) jest średnią kwadratów błędu przybliżenia funkcji $g(.)$ skojarzonej z układem dynamicznym za pomocą funkcji $h(.)$ realizowanej przez sieć neuronową.

Należy zaznaczyć, że preceptron trójwarstwowy nie jest jedynie siecią typu „*feedforward*", która może być użyta w zadaniach aproksymacji w czasie rzeczywistym. W wielu przypadkach nie jest to sieć najefektywniejsza, np. w sensie jak najmniejszej liczby współczynników wagowych q czy jak najkrótszego czasu niezbędnego do przeprowadzenia procesu uczenia. Problem wyboru optymalnej architektury sieci nuronowej do rozwiązania problemu identyfikacji pozostaje ciągle zagadnieniem otwartym. Przykład rozpowszechnionej sieci perceptronowej ma jednak dobre podstawy teoretyczne [16] i bez utraty ogólności rozważań może być użyty do zilustrowania tez niniejszego doniesienia.

Załóżmy zerową moc zakłóceń obserwacji $\sigma=0$, i przypuśćmy, że $h(y)=\theta^*$ dla danego wektora y. Rozważmy teraz mały przyrost wektora parametrów $\Delta\theta$, taki że $(\theta^*+\Delta\theta)\in\Theta$. Przyrost ten wywołuje odpowiednie przyrosty odchyleń wektorów y i h. Korzystając z rozwinięcia rozważanych funkcji w szereg Taylora, można napisać następującą zależność dla zacisku wyjściowego h_k, $k=1,2,...,p$:

$$\Delta h_k=h_k-\theta_k^*= \cong \sum_{i=1}^{n}\left.\frac{\partial h_k}{\partial y_i}\right|_{f^*}\Delta y_i \tag{8}$$

gdzie

$$\Delta y_i=y_i-f(t_j,\theta^*) \cong \sum_{k=1}^{p}\left.\frac{\partial f(t_i,\theta)}{\partial\theta_k}\right|_{\theta^*}\Delta\theta_k \tag{9}$$

oraz $f^*=f(\theta^*)$. Podstawienie zależności (9) do wzoru (8) pozwala wyprowadzić związek

$$\begin{aligned}\Delta h_k=&(H_{k1}F_{11}+H_{k2}F_{21}+...+H_{kn}F_{n1})\Delta\theta_1+...+\\&(H_{k1}F_{1k}+H_{k2}F_{2k}+...+H_{kn}F_{nk})\Delta\theta_k+...+,\\&(H_{k1}F_{1p}+H_{k2}F_{2p}+...+H_{kn}F_{np})\Delta\theta_p\quad,\end{aligned} \tag{10}$$

w którym

$$H_{ki}=\left.\frac{\partial h_k}{\partial y_i}\right|_{f^*},\qquad F_{ik}=\left.\frac{\partial f(t_i,\theta)}{\partial\theta_k}\right|_{\theta^*},\qquad k=1,2,...,p,\quad i=1,2,...,n. \tag{11}$$

Celem uczenia jest osiągnięcie sytuacji, w której funkcja $h_k(.)$ byłaby równa θ_k i jednocześnie niezależna od pozostałych parametrów θ_i, $i=1,2,...,p$, $i\neq k$, dla wszystkich $\theta\in\Theta$. Prowadzi to do następującego warunku:

$$h_k(f(t_1,\theta^*+\Delta\theta),...,f(t_n,\theta^*+\Delta\theta)=\theta_k^*+\Delta\theta_k. \tag{12}$$

Korzystając z (12) i (10) można sformułować układ (13) p równań liniowych z n niewiadomymi H_{k1}, H_{k2},...,H_{kn}:

$$\begin{bmatrix} F_{11} & F_{21} & \dots & F_{n1} \\ \cdot & \cdot & \dots & \cdot \\ F_{1k} & F_{2k} & \dots & F_{nk} \\ \cdot & \cdot & \dots & \cdot \\ F_{1p} & F_{2p} & \dots & F_{np} \end{bmatrix} \begin{bmatrix} H_{k1} \\ H_{k2} \\ \cdot \\ H_{kn} \end{bmatrix} = \begin{bmatrix} 0 \\ \cdot \\ 1 \\ \cdot \\ 0 \end{bmatrix}. \tag{13}$$

Układ równań (13) może być jednoznacznie rozwiązany dla $n=p$ (co określa liczbę wejść sieci neuronowej) pod warunkiem, że dla każdego punktu $\theta^* \in \Theta$ jakobian $| F(\theta^*) |$ jest niezerowy.

Prowadząc powyższą dyskusję p razy dla $k=1,2,...,p$ można otrzymać p równań podobnych do (13). Wszystkie te równania mogą zostać zapisane łącznie w postaci równania macierzowego

$$F^T H = I, \tag{14}$$

w którym

$$H = \begin{bmatrix} H_{11} & H_{21} & \dots & H_{p1} \\ H_{12} & H_{22} & \dots & H_{p2} \\ \cdot & \cdot & \cdot & \cdot \\ H_{1p} & H_{2p} & \dots & H_{pp} \end{bmatrix}, \quad I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix} \tag{15}$$

a F jest macierzą (4) wyznaczoną dla $n=p$. Jeśli założymy, że macierz odwrotna $\mathbf{F}^{-1}$ istnieje możemy pomnożyć obie strony zależności (14) przez F^{-1} aby otrzymać

$$H = (F^T)^{-1}. \tag{16}$$

Podsumowując, funkcja $h(y)$ realizowana za pomocą sieci neuronowej może być jednoznacznie określona jeśli F jest macierzą nieosobliwą w każdym punkcie $\theta^* \in \Theta$, przy czym liczba obserwacji n (wejść sieci) musi być równa liczbie nieznanych parametrów p.

Równanie (16) określa warunki jakie powinny być spełnione przez pochodne odwzorowania $h(.)$. To, czy sieć neuronowa o danej architekturze i liczbie neuronów jest w stanie zapewnić dostatecznie mały błąd aproksymacji odwzorowania $g(.)$ oraz jego pochodnych [dla danego układu dynamicznego opisanego zbiorem obserwacji (1)] jest oddzielną kwestią, do wyjaśnienia w dalszych badaniach.

4. WPŁYW ZAKŁÓCEŃ LOSOWYCH NA WŁAŚCIWOŚCI ESTYMATORÓW I NA PROCES UCZENIA SIECI NEURONOWEJ

W powyższych rozważaniach przyjęto, że sieć neuronowa jest uczona za pomocą równań modelu. Tym samym obserwacje układu użyte do uczenia sieci nie były zakłócone, $y \equiv f$. Jednakże w pewnych przypadkach funkcja $f(.)$ jest nieznana, ale są

dostępne zakłócone obserwacje (2) wraz z odpowiadającymi im wartościami wektora parametrów θ. Wykażemy, że uczony na takich przykładach estymator

$$\tilde{\theta} = \boldsymbol{h}(\boldsymbol{y}) \tag{17}$$

jest estymatorem obciążonym, nawet wtedy gdy sieć neuronowa może w ogólności przybliżać funkcję $g(.)$ z zerowym błędem. Wykażemy też, że z drugiej strony estymator ten, uczony na przykładach zakłóconych obserwacji, ma mniejszą wariancję od estymatora uczonego na przykładach bez zakłóceń [4, 6] a także od estymatora średniokwadratowego. Istnieje możliwość optymalizacji processu uczenia tak, aby otrzymać określony kompromis między wartościami obciążenia i wariancji, co jest oryginalną cechą proponowanej metody identyfikacji, bardzo ważną w zastosowaniach.

Przyjmijmy, że wyrazy ε_i reprezentujące zakłócenia w założeniu (a) mają wariancję różną od zera. Jeśli wariancja zakłóceń jest dostatecznie mała, $\sigma^2 \cong 0$, to estymator $\tilde{\theta}$ można przybliżyć wzorem

$$\tilde{\theta} \cong h(f^*) + H^T \varepsilon \tag{18}$$

przy czym $\varepsilon = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_n]^T$ oraz $f^* = f(\theta^*)$, $\theta^* \in \Theta$. Wartość oczekiwana tego estymatora jest równa

$$E[\tilde{\theta}] \cong h(f^*). \tag{19}$$

Tym samym proponowany estymator, wykorzystujący sieć neuronową, może być obciążony. Wektor obciążenia β jest po prostu błędem aproksymacji

$$\begin{aligned} \beta &= h(f^*) - g(f^*) \\ &= h(f^*) - \theta^*. \end{aligned} \tag{20}$$

Wartość oczekiwana wyrażenia (18) jest równa dokładnej wartość θ^* wektora parametrów jeśli błąd przybliżenia funkcji $g(.)$ za pomocą funkcji $h(.)$ jest zerowy. W realistycznym przypadku skończonej dokładności aproksymacji, estymator (17) jest estymatorem obciążonym. Z drugiej strony wiadomo, że sieć neuronowa typu feedforward ma zdolność przybliżania dowolnej funkcji ciągłej wielu zmiennych z dowolną dokładnością, pod warunkiem, że funkcja nieliniowa neuronów takiej sieci spełnia pewne łagodne warunki [16, 17]. Tym samym błąd obciążenia (20) można uczynić pomijalnie małym, co znajduje potwierdzenie w praktyce [3, 4, 6, 7, 8, 9, 18]. Poszukiwanie optymalnej architektury sieci neuronowej, najbardziej odpowiedniej do rozwiązywania problemu aproksymacji (w sensie małego błędu przybliżania funkcji i jej pierwszej pochodnej; małej liczby współczynników wagowych oraz krótkiego czasu uczenia) pozostaje nadal ważnym problemem badawczym. Chociaż wybór architektury do zadania identyfikacji może być w pewnym stopniu zależny od konkretnego układu dynamicznego, proponowane są sieci np. [19], mające przewagę nad popularnym perceptronem w omawianym sensie.

Przyjmijmy, że sieć neuronowa była uczona na przykładach obserwacji *niezakłóconych*, z wykorzystaniem wskaźnika jakości (7). Załóżmy, że po zakończeniu procesu uczenia osiągnięto zerowy błąd aproksymacji funkcji $g(.)$, co oznacza $\| \beta \| = 0$. Przyjmijmy też realistycznie, że nauczona sieć jest testowana za pomocą

obserwacji zakłóconych (1), przy czym zmienne losowe ε_i, $i=1,2,...,n$ mają rozkład normalny o zerowej wartości średniej i wariancji $\sigma^2 \cong 0$. W tych warunkach błąd identyfikacji wywołany zakłóceniami losowymi będzie opisany p-wymiarowym rozkładem normalnym. Przy wykorzystaniu zależności (18), macierz kowariancji estymatora $\tilde{\theta}$ może być przedstawiona w następującej postaci:

$$\begin{aligned} \operatorname{cov}[\tilde{\theta}] &= E\left[(\tilde{\theta}-\theta^*)(\tilde{\theta}-\theta^*)^T\right] \\ &\cong E\left[(\boldsymbol{H}^T\varepsilon)(\boldsymbol{H}^T\varepsilon)^T\right] \\ &= E\left[\boldsymbol{H}^T\varepsilon\varepsilon^T\boldsymbol{H}\right] \\ &= \sigma^2\boldsymbol{H}^T\boldsymbol{H}, \end{aligned} \tag{20}$$

gdzie założono, że $n=p$. Jeśli spełniona jest zależność (16), oznaczająca zerowy błąd aproksymacji pierwszych pochodnych funkcji $g(.)$ względem składowych wektora y w punkcie θ^*, to kowariancja estymatora może być zapisana jako

$$\operatorname{cov}[\tilde{\theta}] \cong \sigma^2(FF^T)^{-1}. \tag{21}$$

Im większa jest wartość wyznacznika $|F^TF|$, tym mniejsza jest uogólniona wariancja estymatora $\tilde{\theta}$ i mniejsza objętość elipsoid wiarygodności [13] które charakteryzują rozrzut wartości identyfikowanych parametrów wokół prawdziwej wartości θ^*. Rozrzut ten zależy od wyboru punktów obserwacji t_i, $i=1,2,...,n$, które decydują o właściwościach macierzy $\boldsymbol{F}$.

Można łatwo wykazać, że wzór (21) jest słuszny także dla estymatora średniokwadratowego, jeśli $n=p$. Tym samym, dla przyjętych założeń, proponowany estymator uczony na przykładach niezakłóconych jest równoważny standardowemu estymatorowi średniokwadratowemu. Z drugiej strony, w celu zmniejszenia wpływu zakłóceń, estymator średniokwadratowy jest zwykle wykorzystywany przy liczbie obserwacji większej od liczby nieznannych parametrów, $n>p$. Asymptotyczne właściewości tego estymatora, dla $n\to\infty$, były dyskutowane w literaturze, np. [10]. Chociaż można podejmować próby uczenia sieci neuronowej przy $n>p$, co może prowadzić do jednego z wielu rozwiązań, to nie jest to podejście zalecane. Czas trwania procesu uczenia wzrasta bardzo szybko ze wzrostem liczby współczynników wagowych, a tym samym wejść sieci. Wzrastają również szanse na przedwczesne zakończenie procesu uczenia w efekcie zwiększonej liczby minimów lokalnych. Tym samym, po długim procesie uczenia można otrzymać sieć, która nie wykorzystuje w pełni zjawiska redukcji wpływu zakłóceń poprzez ich uśrednienia. Proponuje się, dla przypadku $n>p$, wykorzystanie modułowej sieci neuronowej. W sieci takiej obserwacje są wpierw przekształcane do postaci p-elementowego wektora cech, za pomocą modułu wstępnego przetwarzania. W wielu przypadkach moduł taki może realizować rozkład sygnału na składowe kanoniczne (ang. principal component analysis [20]). Wektor reprezentowanych cech jest sygnałem wejściowym dla modułu aproksymacji. Słuszność proponowanego rozwiązania zostanie zilustrowana przykładem symulacji komputerowej w dalszym ciągu artykułu. Bardziej szczegółowa analiza właściwości sieci modułowej wykracza poza zakres niniejszego opracowania.

Załóżmy teraz, że obserwacje zakłócone (1) są użyte do uczenia sieci a uczenie polega na minimalizacji następującego wskaźnika jakości:

$$Q_k(w)=N^{-1}\sum_{l=1}^{N}(\theta_k^{(l)}-h_k^{(l)}(w,y(\theta^{(l)})))^2 \tag{22}$$

w przestrzeni współczynników wagowych, $w\in\mathcal{R}q$. Korzystając z rozwinięcia funkcji $\boldsymbol{h}(.)$ w szereg Taylora możemy przyjąć dla małej mocy zakłóceń, że

$$\begin{aligned}\boldsymbol{h}_k[w,y(\theta^{(l)})]&=h_k[w,f(\theta^{(e)}+\varepsilon]\\&\cong h_k[w,f(\theta^{(l)})]+\gamma_l^T\varepsilon,\end{aligned} \tag{23}$$

gdzie elementy wektora $\gamma_l=[H_{k1},\ldots,H_{kp}]$ są zdefiniowane zależnościami (11) dla $n=p$ oraz $\theta^*=\theta^{(l)}$. Podstawiając (23) do (22) można otrzymać

$$\begin{aligned}Q_k(w)&\cong N^{-1}\sum_{l=1}^{N}(\theta_k^{(l)}-g_k(w,y^{(l)})-\gamma_l^T\varepsilon)^2\\&=N^{-1}\sum_{l=1}^{N}(\delta_l^2-2\delta_l\gamma_l^T\varepsilon+\gamma_l^T\varepsilon\varepsilon^T\gamma_l),\end{aligned} \tag{24}$$

gdzie $\delta_l=\delta_l(w)=\theta_k^{(l)}$-$h_k(w, f^{(l)})$ jest błędem aproksymacji dla l-tego przykładu. Korzystając z metodologii zaproponowanej przez Widrowa w [21] rozważmy zespół identycznych sieci neuronowych, z których każda posiada ten sam wektor wag w. Wszystkie wejścia rozważanych sieci mają tę samą składową deterministyczną $f^{(l)}$. Addytywna składowa losowa wejść zespołu jest dana wektorem ε, który jest nieskorelowany z δ_l. Po uśrednieniu równania (24) w zespole sieci otrzymujemy

$$E[Q_k(w)\cong N^{-1}\sum_{h=1}^{N}(\delta_l^2+\sigma^2\gamma_l^T\gamma_l). \tag{25}$$

Każdy z wyrazów w nawiasach wyrażenia (25) jest sumą dwóch nieujemnych składowych. Każda z nich jest funkcją w. Pierwsza składowa, δ_l^2, reprezentuje obciążenie estymatora $\tilde{\theta}_k$, k-tego parametru θ_k, podczas gdy druga z nich, $\sigma^2\,\gamma_l^T\gamma_l$, jest jego wariancją. Minimalizacja drugiej składowej prowadzi do zmniejszenia normy wektora pochodnych $\|\gamma_l^T\gamma_l\|$ i w efekcie do uczynienia funkcji $h_k(\theta)$ stałą w przestrzeni obserwacji. Macierz $\boldsymbol{H}$ dla funkcji stałej miałaby liniowo zależne wiersze (kolumny). Z drugiej strony, według (16) macierz ta powinna być równa odwrotności transponowanej macierzy Jakobiego $\boldsymbol{F}$, której wiersze (kolumny) muszą być liniowo niezależne, aby problem identyfikacji miał rozwiązanie. Dowolna macierz nie może mieć swych wierszy (kolumn) jednocześnie liniowo zależnych i niezależnych. Nie jest zatem możliwa jednoczesna minimalizacja obu wyrazów w nawiasach sumy (25), to jest mocy błędu δ_l^2 i normy wektora pochodnych $\|\gamma_l^2\gamma_l\|$, dla wszystkich $\theta^*\in\Theta$. Poprzez minimalizację wskaźnika jakości (25) zamiast minimalizacji sumy

$$P_k(w) = N^{-1} \sum_{l=1}^{N} \delta_l^2 \tag{26}$$

[która jest w istocie wskaźnikiem jakości (7) dla uczenia bez zakłóceń] odwzorowanie $h(.)$ staje się gładsze w takim sensie, że jego pierwsze pochodne stają się mniejsze co do wartości bezwzględnej. Zwiększa to wartość błędu δ_l. Równowaga między tymi dwoma czynnikami, osiągnięto po zakończeniu procesu uczenia zależy od wariancji szumów, co wynika z (25).

Dyskutowana właściwość proponowanego estymatora, a mianowicie możliwość kontrolowania proporcji między błędem obciążenia i wariancją estymatora jest interesująca tak z teoretycznego jak i praktycznego punktu widzenia. Rozważmy na przykład zagadnienie identyfikacji parametrów rozkładu temperatury w tkankach organizmu człowieka za pomocą nieinwazyjnych pomiarów z użyciem termografu mikrofalowego [22]. W takim zastosowaniu, dla uzyskania dużej precyzji pomiarów, należy wydłużyć czas uśredniania mocy szumów mikrofalowych mierzonych przez termograf. Z drugiej strony, w wydłużonym czasie obserwacji parametry rozkładu mogą się zmieniać, np. w wyniku oddychania pacjenta. Użycie odpowiednio nauczonej sieci neuronowej do identyfikacji parametrów daje potencjalnie możliwość zachowania określonego rozrzutu parametrów przy zwiększonej wariancji zakłóceń, jaka towarzyszyłaby żądanemu skróceniu czasu pomiarów [23].

Warto również zauważyć, że w celu kontrolowania proporcji między obciążeniem i wariancją estymatora nie jest konieczne dodawanie zakłóceń do obserwacji obliczanych za pomocą modelu badanego układu dynamicznego w procesie uczenia. Jak wynika z (25), wariancję zakłóceń σ^2 można zastąpić odpowiednio dobranym współczynnikiem większym od zera. Wprowadzenie wyrazu proporcjonalnego do deterministycznych pochodnych $\gamma_l^T \gamma_l$ do funkcji wskaźnika jakości [por. (25)] jest równoznaczne z wymuszeniem pewnej gładkości funkcji $h(.)$, zależnej od wartości wprowadzonego współczynnika.

5. SYMULACJA KOMPUTEROWA

Jako przykład rozważmy 3-parametrowy układ, którego sygnał wyjściowy dla pewnego wymuszania jest opisany sumą funkcji eksponencjalnych

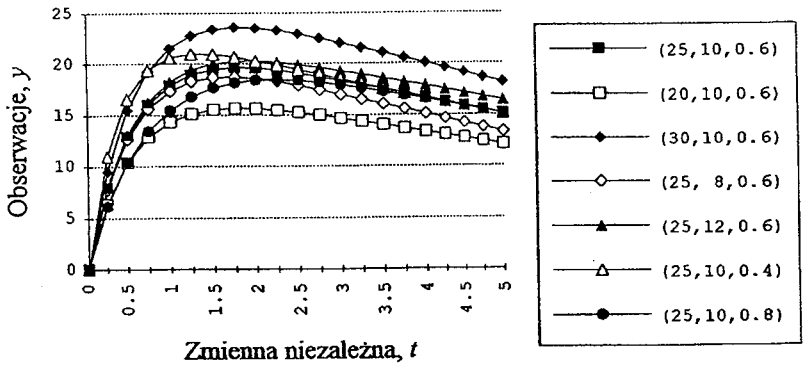
$$f(t,\theta) = \theta_1 \left[\exp\left(-\frac{t}{\theta_2}\right) - \exp\left(-\frac{t}{\theta_3}\right) \right] \tag{27}$$

w następującej przestrzeni parametrów Θ:

$$20.0 \leqslant \theta_1 \leqslant 30.0, \quad 8.0 \leqslant \theta_2 \leqslant 12.0, \quad 0.4 \leqslant \theta_3 \leqslant 0.8 \tag{28}$$

Model (27) może opisywać profil temperatury wewnątrz tkanek organizmu człowieka w badaniach nad metodami nieinwazyjnego pomiaru temperatury za pomocą termografii mikrofalowej [22]. Celem takich badań jest opracowanie narzędzi

wspomagających diagnostykę chorób nowotworowych. Model ten jest również specjalnym przypadkiem tzw. modeli kompartmentowych jakie są rozważane w zagadnieniach farmakokinetyki [24, 25]. Jest on także odpowiedzią układu elektronicznego drugiego rzędu na wymuszanie skokowe, co zostanie bliżej wyjaśnione w ramach niniejszej pracy.



Rys. 3. Wektory obserwacji dla układu opisanego modelem (27)

Rys. 3 przedstawia przykłady $n=20$-elementowego wektora obserwacji układu, obliczone na podstawie (27) dla kilku wybranych wartości wektora parametrów $\theta=[\theta_1, \theta_2, \theta_3]^T \in \Theta$. Trójwymiarowe wektory obserwacji są potrzebne do jednoznacznego zdefiniowania funkcji $\tilde{\theta}=h(y)$ dla modelu (27). Używając programu optymalizacyjnego ustalono, że następujące wartości zmiennej niezależnej

$$t_1=0{,}558, \quad t_2=2{,}613, \quad t_3=12{,}87 \tag{29}$$

maksymalizują wyznacznik $|F^TF|$ obliczony dla „nominalnego" wektora parametrów $\theta_0=[25{,}0\ 10{,}0\ 0{,}6]^T$. Wartości (29) pozostały niezmienne w procesie uczenia i testowania sieci neuronowej. W celu uproszczenia zadania uczenia sieci [26], odchylenia obserwacji od ich wartości nominalnych

$$\Delta y_i=y_i-f(t_i, \theta_0), \quad i=1, 2, 3 \tag{30}$$

tworzyły wektor wejściowy dla sieci neuronowej. Przykłady do uczenia i testowania wygenerowano rozpinając regularną sieć punktów w przestrzeni parametrów (28), przy K punktach na parametr:

$$\begin{aligned} \theta_{1i}&=20{,}0+10{,}0\ (i-1)/(K-1), \quad i=1, 2,...,K \\ \theta_{2j}&=8{,}0+4{,}0\ (j-1)/(K-1), \quad j=1,2,...,K \\ \theta_{3k}&=0{,}4+0{,}4\ (k-1)/(K-1), \quad k=1,2,...,K. \end{aligned} \tag{31}$$

Do uczenia przyjęto $K=8$, co daje $N=K^3=256$ przykładów wektora parametrów

$$\theta^{(l)}=[\theta_{1i}, \theta_{2j}, \theta_{3k}], \quad l=K^2\ (i-1)\ +\ K\ (j-1)\ +k. \tag{32}$$

Do testowania wykorzystano gęstszą siatkę, złożooną z $N=27000$ punktów a otrzymaną z (31) i (32) dla $K=30$.

Napisano program w języku Turbo Pascal® [27] do uczenia i testowania sieci perceptronowych dla zadań identyfikacji parametrów. Dla zredukowania czasu obliczeń zastosowano następującą funkcję sigmoidalną

$$\xi(\alpha)=\begin{cases} 0, & \alpha<-1 \\ \frac{1}{4}(2+3\alpha-\alpha^3), & -1\leqslant\alpha\leqslant 1 \\ 1, & \alpha>1 \end{cases} \tag{33}$$

zamiast tradycyjnie używanej funkcji $1/(1+exp(-\alpha))$. Jak można się było spodziewać [16], modyfikacja ta nie spowodowała, dla wielu układów testowych, żadnych istotnych różnic jakości aproksymacji. Czas obliczeń zależności (33) za pomocą komputera PC 386/33 jest w przybliżeniu trzy razy krótszy od czasu obliczeń funkcji tradycyjnej, co jest korzyścią z tej modyfikacji. Do wyznaczania początkowych wartości współczynników wagowych wykorzystano algorytm wstecznej propagacji błędu [14, 15]. Do dokładniejszego określenia wag wykorzystano program minimalizacji funkcji wielu zmiennych zmodyfikowaną metodą Daviesa-Swanna-Campeya (DSC) [28].

Przeprowadzono odpowiednie obliczenia za pomocą komputera PC 486/66 MHz, których wyniki zestawiono w Tabeli I, dla przypadku uczenia i testowania sieci przy zerowej wariancji zakłóceń. Pierwiastek kwadratowy z wyrażenia $P_k(\boldsymbol{w})$ (7), obliczony odpowiednio dla parametrów θ_1, θ_2, oraz θ_3 przy $N=27000$, posłużył do oceny błędu identyfikacji po etapie uczenia za pomocą algorytmu BP (rms_{BP}) oraz po etapie minimalizacji DSC (rms_{DSC}). Tabela zawiera również wartość maksymalną modułu błędu w zbiorze testowym, po zakończeniu procesu uczenia, max_{DSC}. Zwiększenie rozmiarów sieci perceptronowej, mierzonych liczbą jej neuronów m, prowadzi do zmniejszenia błędów identyfikacji, przy odpowiednio rosnącym czasie uczenia. Cztery neurony są potrzebne do otrzymania błędu $max_{DSC}=0.08$ dla parametru θ_1, co stanowi około 0.3% wartości nominalnej tego parametru, równej *25*. Siedem neuronów potrzeba do uzyskania tej samej wartości błędu względnego dla pozostałych dwóch parametrów modelu (27), (28). Zwiększona liczba potrzebnych neuronów może być uzasadniona faktem, że parametry te, w odróżnieniu od θ_1, w sposób nieliniowy zależą od obserwacji systemu.

Podczas gdy wyniki te dowodzą zasadności proponowanej metody identyfikacji, to trzeba jednak zaznaczyć, że architektura sieci perceptronowej nie jest najbardziej efektywna w tym zastosowaniu. Dla przykładu, tę samą wartość maksymalną błędu (*0.3%*) osiągnięto za pomocą sieci neuronowaj realizującej 13-parametrową funkcję wymierną [19] zamiast (6). Sieć taka może być uczona poprzez rozwiązanie układu równań liniowych, w czasie wielokrotnie (np. kilkusetkrotnie) krótszym od czasu uczenia sieci perceptronowej.

W celu zilustrowania wpływu zakłóceń na działanie estymatora (uczonego na przykładach bez zakłóceń), przeprowadzono symulację komputerową odpowiedniego eksperymentu. W ramach tej symulacji losowano wartości parametrów z przestrzeni (28) korzystając z generatora liczb pseudolosowych o rozkładzie równomiernym. Dla każdego wylosowanego wektora parametrów obliczono odpowiedź układu za pomocą wzoru (27) dla wartości zmiennej niezależnej danych przez (29). Następnie dodawano do każdej obliczonej obserwacji liczbę pseudolosową generowaną przez program symulujący rozkład Gaussa o zerowej wartości średniej i odchyleniu standardowym σ. W ten sposób wyznaczone obserwacje zakłócone stanowiły sygnał wejściowy dla sieci nuronowej uczonej na przykładach bez zakłóceń. W odpowiedzi na taki sygnał wejściowy sieć generowała wartość parametrów układu z pewnym błędem. Te same obserwacje służyły do wyznaczania estymat parametrów za pomocą metody najmniejszych kwadratów. Dla każdej wartości σ obliczono błąd średniokwadratowy dla populacji 500 losowych wektorów θ. Stwierdzono, że dla rozważonego zakresu σ wartości średnie błędów były około 10 razy mniejsze od odpowiednich wartości średniokwadratowych. Wyniki symulacji dla parametru θ_2 wykreślono na rys. 4. Dla małych poziomów szumu $\sigma < 0.01$, metoda najmniejszych kwadratów (KW3) daje mniejszy błąd od metody proponowanej (NN3). Wynika to ze skończonej dokładności odwzorowania funkcji $g(.)$ w rozważanym przykładzie za pomocą sieci neuronowej. Innymi słowy, efekt ten jest odwzorowaniem obciążenia estymatora neuronowego. Odpowiadająca temu wartość błędu pokrywa się z wartością $\text{rms}_{DSC} = 0.016$ w Tabe-

Tabela 1

Czas uczenia i błędy estymatorów z siecią perceptronową o trzech wejściach ($n = 3$) dla parametrów θ_1 (a) oraz θ_2 (b) układu (27)

(a)

m, q	rms_{BP}	czas uczenia [s]	rms_{DSC}	max_{DSC}
1,6	0,205	68	0,198	0,570
2,11	0,205	80	0,180	0,543
3,16	0,205	360	0,115	0,485
4,21	0,198	570	0,020	0,080

(b)

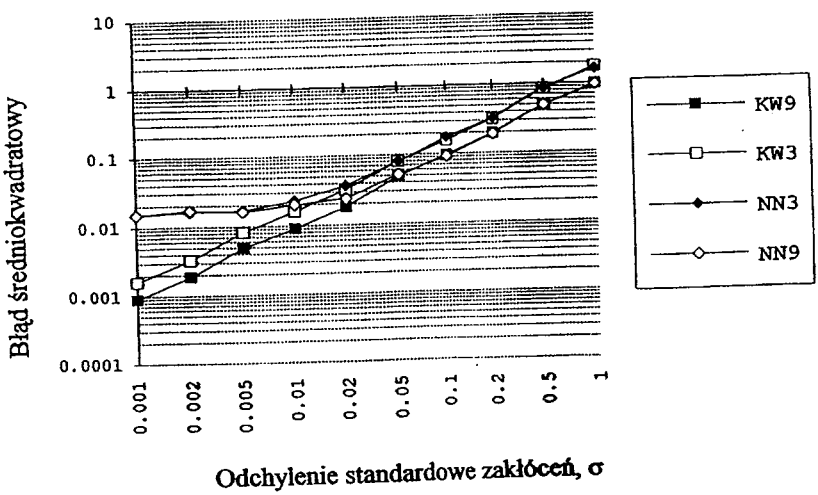
m, q	rms_{BP}	czas uczenia [s]	rms_{DSC}	max_{DSC}
1,6	0,162	68	0,142	0,443
2,11	0,097	127	0,069	0,247
3,16	0,157	351	0,055	0,157
4,21	0,087	610	0,029	0,142
5,26	0,057	1380	0,016	0,072
6,31	0,040	1740	0,012	0,062
7,36	0,033	1940	0,008	0,031

li 1(a) otrzymaną dla $(m, q)=(5,26)$. Dla większych poziomów szumu, $\sigma>0.02$, przy których błąd obciążenia jest pomijalny, oba estymatory oferują ten sam poziom błędu, w zgodzie z rozważaniami teoretycznymi.

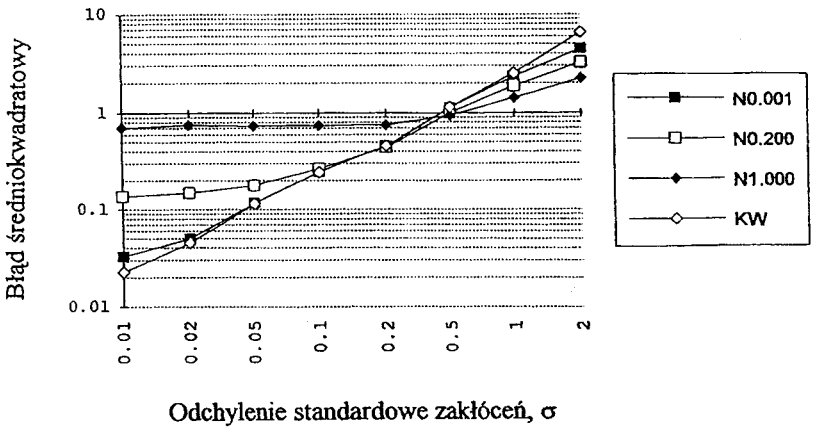
Liczba obserwacji n była do tej pory równa liczbie nieznanych parametrów p. W celu wykorzystania informacji zawartych w większej liczbie obserwacji, $n>p$, zastosowano modułową sieć dwustopniową. Pierwszy stopień realizuje rzutowanie wektora obserwacji na bazę p ortogonalnych wektorów za pomocą sieci liniowej (sumatora ważonego). Zadaniem tego stopnia jest redukcja wymiarowości problemu z zachowaniem istotnym informacji o parametrach układu. Do jego zaprojektowania wykorzystano zasady rozkładu sygnału na składowe kanoniczne (PCA) [20, 19]. Współczynniki rozkładu z wyjścia pierwszego stopnia sieci stanowiły sygnał wejściowy dla jej drugiego stopnia, który realizował zadanie aproksymacji funkcji odwrotnej $\boldsymbol{h}(.)$. Obserwacje układu, w liczbie $n=9$, pobrano w następujących punktach:

$$\begin{aligned} &t_1=0{,}358, \quad t_2=0{,}558, \quad t_3=0{,}758, \\ &t_4=2{,}413, \quad t_5=2{,}613, \quad t_6=2{,}813, \\ &t_7=12{,}67, \quad t_8=12{,}87, \quad t_9=13{,}07. \end{aligned} \tag{34}$$

Sieć dwustopniowa testowana obserwacjami niezakłóconymi wykazała ten sam poziom błędu co sieć scharakteryzowana wynikami z Tabeli I. Widać to na rys.4 (krzywa NN9 dla małych poziomów szumu). Jednakże dla dużych poziomów szumu, sieć dwustopniowa pozwoliła na zredukowanie błędów wywołanych szumami. Błąd średniokwadratowy był praktycznie taki sam jak błąd estymatora średniokwadratowego przy tej samej liczbie $n=9$ obserwacji (KW9), co również można dostrzec na rys. 4.



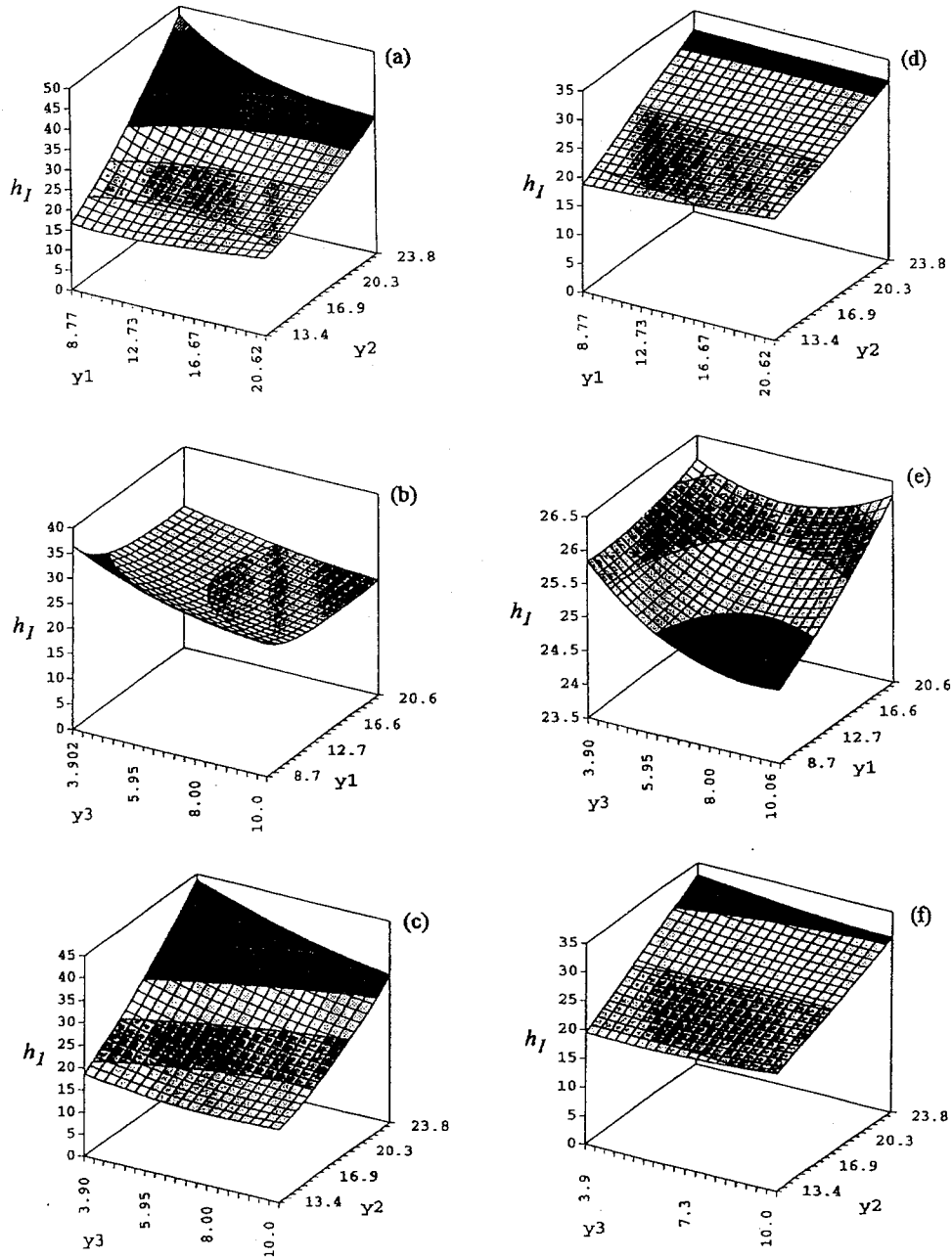
Rys. 4. Symulowany błąd średniokwadratowy identyfikacji parametru θ_2 układu (27) za pomocą sieci neuronowej uczonej bez zakłóceń przy $m=5$ (NN3, NN9) i za pomocą metody najmniejszych kwadratów (KW3, KW9). Estymacja z użyciem $n=3$ obserwacji (NN3, KW3) oraz $n=9$ obserwacji (NN9, KW9)



Rys. 5. Symulowany błąd średniokwadratowy identyfikacji parametru θ_2 układu (27) za pomocą sieci neuronowej uczonej z zakłóceniami

Kolejny eksperyment numeryczny przeprowadzono przy $n=3$ dla sieci neuronowej uczonej na przykładach obserwacji *zakłóconych*. Rys. 5 przedstawia wyniki otrzymane dla sieci perceptronowej o $m=4$ neuronach wraz z wynikami uzyskanymi metodą najmniejszych kwadratów, dla parametru θ_1. Proces uczenia powtórzono trzykrotnie, dla różnych wartości odchylenia standardowego zakłóceń, $\sigma=0.001$, $\sigma=0.2$ oraz $\sigma=1.0$ (krzywe odpowiednio N 0.001, N 0.200 and N 1.000 na rys.5). Jak widać, estymator neuronowy (o skończonych rozmiarach) produkuje większe błędy od estymatora średniokwadratowego dla małych poziomów szumu. Zostało to już objaśnione błędem obciążenia. Błędy te rosną ze wzrostem wariancji zakłóceń w procesie uczenia, jako że krzywa N 1.000 leży nad krzywą N 0.200, która z kolei znajduje się nad krzywą N o.001 na rys. 5, dla $\sigma<0.5$. Dla średnich poziomów szumu przy testowaniu i dla sieci neuronowych uczonych z mocą zakłóceń bliską zera, obie metody dają ten sam błąd, jak to przewidywano wcześniej. I rzeczywiście, krzywe N 0.001 oraz KW pokrywają się dla $0.02<\sigma<1.0$. Jednakże dla dużych poziomów szumu błąd estymatora neuronowego jest mniejszy od błędu metody najmniejszych kwadratów. Przykład ten ilustruje dyskutowaną wyżej tezę, że użycie sieci neuronowych do identyfikacji parametrów układów dynamicznych daje możliwość uzyskania zwiększonej odporności na zakłócenia i że odporność tę można kontrolować.

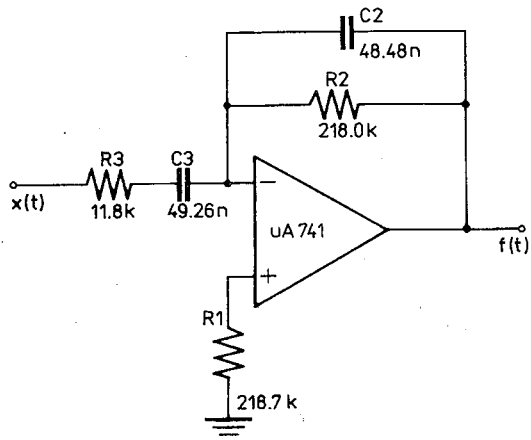
Na rys. 6 naszkicowano powierzchnie reprezentujące funkcję odwrotną $\boldsymbol{h}_1(y_1, y_2, y_3,)$ dla wybranych wartości obserwacji $(y_1, y_2, y_3,)$. Wykresy po lewej stronie rysunku otrzymano dla sieci neuronowej uczonej bez zakłóceń, a wykresy umieszczone po stronie prawej odpowiadają sieci uczonej na przykładach zakłóconych. Jest oczywiste, że powierzchnie w prawej kolumnie są „*gładsze*" od powierzchni z kolumny lewej (warto zwrócić uwagę na zmianę skali osi rzędnych). Tym samym odpowiedź sieci neuronowej uczonej z zakłóceniami jest mniej wrażliwa na zaburzenia obserwacji (szumy). Oczywiście jest to przyczyną zwiększonych błędów obciążenia estymatora.



Rys. 6. Powierzchnie $h_1(y_1y_2y_3)$ generowane dla układu (27) przez estymator z siecią neuronową uczoną na przykładach obserwacji bez zakłóceń (a) $y_3=7{,}15$, b) $y_2=18{,}9$, c) $y_1=15{,}0$ i z zakłóceniami d) $y_3=7{,}15$, e) $y_2=18{,}9$, f) $y_1=15{,}0$

6. EKSPERYMENT

W celu doświadczalnego zweryfikowania analizy teoretycznej zbudowano filtr pasmowoprzepustowy RC z wykorzystaniem wzmacniacza operacyjnego μA741. Rys. 7 przedstawia schemat ideowy filtru. Ten szczególny układ został wybrany ze względu na fakt, iż jego odpowiedź impulsowa jest także opisana multieksponencjalną zależnością (27). Aby lepiej uwidocznić potencjalne możliwości tkwiące w metodzie identyfikacji za pomocą sieci neuronowej, dodano jeszcze jeden parametr do modelu układu. Parametrem tym jest nieznany czas opóźnienia odpowiedzi skokowej.



Rys. 7. Schemat ideowy układu elektronicznego zbudowanego w celu przeprowadzenia doświadczeń z 4-parametrowym układem dynamicznym

Estymacja czasu opóźnienia jest ważna w praktyce i jest uważana za trudne i czasochłonne zadanie jeśli jest ono rozwiązywane metodą najmniejszych kwadratów [5]. Pokażemy, że estymacja czasu opóźnienia nie różni się w istocie od estymacji innych parametrów jeśli zostanie wykorzystana do tego celu sieć neuronowa. W odpowiednio zaplanowanym eksperymencie, opóźniony $\tau > 0$ jednoskokowy skok napięcia został podłączony do wejścia filtru z rys. 7:

$$x(t) = \begin{cases} 0, & t < \tau \\ -1, & t > \tau . \end{cases} \tag{35}$$

Niezakłócone obserwacje były próbkami następującej funkcji:

$$f(t, \theta) = \theta_1 \left[\exp\left(\frac{\theta_4 - t}{10\theta_2}\right) - \exp\left(\frac{\theta_4 - t}{\theta_3}\right) \right], \tag{36}$$

gdzie

$$\theta_1 = \frac{R_2 C_3}{R_2 C_2 - R_3 C_3}, \quad \theta_2 = 0.1 R_2 C_2 [\text{ms}], \quad \theta_3 = R_3 C_3 [\text{ms}], \quad \theta_4 = \tau [\text{ms}]. \tag{37}$$

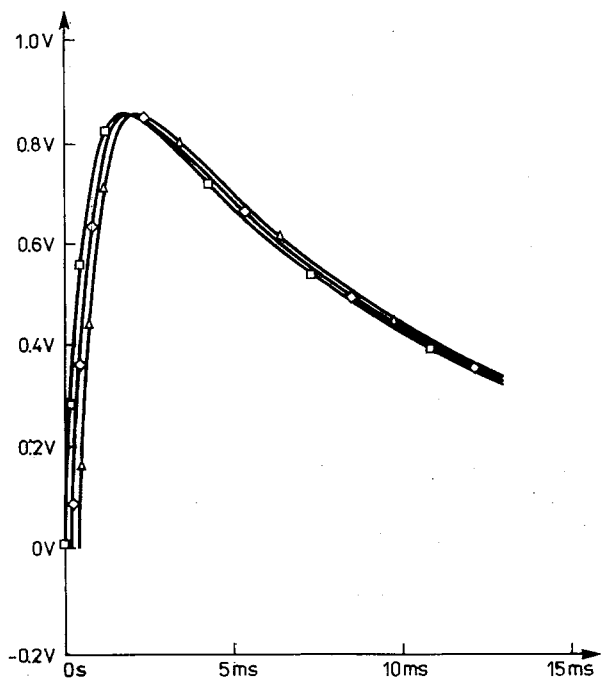
Przestrzeń parametrów Θ zdefiniowano nierównościami (38)

$$0.8 \leqslant \theta_1 \leqslant 1.2, \quad 0.8 \leqslant \theta_2 \leqslant 1.2, \quad 0.4 \leqslant \theta_3 \leqslant 0.8, \quad 0.0 \leqslant \theta_4 \leqslant 0.4. \tag{38}$$

Uczenie i testowanie przeprowadzono według tej samej procedury, którą opisano wyżej dla układu trójparametrowego. Przyjęto osiem wartości każdego z parametrów, równomiernie rozłożonych w przedziałach (38) dla generacji przykładów do uczenia, co odpowiada całkowitej liczbie przykładów $N=4096$. Do testowania wykorzystano $N=160{,}000$ przykładów równomiernie pokrywających przestrzeń (38). Do jednoznacznego zdefiniowania funkcji $\tilde{\theta}=\boldsymbol{h}(\boldsymbol{y})$ dla układu (36) potrzebne były czteroelementowe wektory obserwacji. Dla uczenia, symulacji i pomiarów ustalono następujące wartości zmiennej niezależnej t którą jest w przypadku badanego układu czas:

$$t_1=0.5 \text{ ms} \quad t_2=1.0 \text{ ms} \quad t_3=3.0 \text{ ms} \quad t_4=13.0 \text{ ms} \tag{39}$$

Przeprowadzono uczenie czterech różnych sieci perceptronowych, odpowiednio dla każdego z czterech parametrów układu (36). Stwierdzono, że dla sieci zawierających $m=10$ neuronów w warstwie wewnętrznej (czemu odpowiada $q=61$ nastawionych wag), wartość maksymalna błędu identyfikacji w zbiorze testowym nie przekracza arbitralnie wybranej małej liczby 0.01. Wagi czterech symulowanych sieci neuronowych zostały „zamrożone" po zakończeniu procesu uczenia.



Rys. 8. Odpowiedź skokowa $f(t,\theta)$ układu z rys. 7, symulowana za pomocą programu PSPICE®: □ $\tau=0$, ◇ $\tau=0{,}2$ ms, △ $\tau=0{,}4$ ms

Na czas testowania do wejścia sieci dołączono obserwacje układu. W celu wszechstronnego zbadania właściwości sieci neuronowych, posłużono się zarówno obserwacjami symulowanymi za pomocą programu PSpice [30], jak i obserwacjami zmierzonymi, które były generowane przez odpowiednio skonstruowany układ elektroniczny. Użyte wartości elementów RC filtru, zmierzone za pomocą mostka impedancji, zaznaczono na rys. 7. Rysunek 8 przedstawia symulowane odpowiedzi $f(t, \theta^*)$ układu. Obliczono je dla kilku ustalonych wartości wektora parametrów $\theta^* = [1{,}075, 1{,}057, 0{,}581, \tau]^T$, które odpowiadały zmierzonym wartościom RC oraz trzem arbitralnie wybranym wartościom czasu opóźnienia τ. Tabela 2 zawiera zestawienie błędów identyfikacji, których przyczyną były błędy zaokrągleń w obliczeniach za pomocą programu PSpice (składowe η) oraz skończona dokładność aproksymacji za pomocą sieci neuronowej (składowe β, reprezentujące obciążenie estymatorów). Obie porównywane metody identyfikacji wykazały podobny, raczej niski, poziom błędów.

Tabela 2

Błędy identyfikacji parametrów układu z rys. 7 otrzymane dla obserwacji symulowanych za pomocą programu PSPICE®

	SIEĆ NEURONOWA				MET. NAJMNIEJSZYCH KWADR.			
τ, ms	$\eta_1+\beta_1$	$\eta_2+\beta_2$	$\eta_3+\beta_3$	$\eta_4+\beta_4$	η_1	η_2	η_3	η_4
0,0	0,0003	−0,002	0,0009	−0,0005	0,0002	−0,0003	0,0004	0,0003
0,1	−0,0002	−0,0009	0,0004	0,0003	−0,0003	−0,0004	0,0005	−0,0002
0,2	−0,0004	0,0004	−0,0009	−0,0002	−0,0003	−0,0001	0,0006	0,0001
0,3	0,0006	0,0011	0,0016	−0,0022	0,0003	−0,0002	0,0005	0,0002
0,4	0,0011	0,0017	0,0009	0,0003	0,0001	0,0	0,0005	0,0002

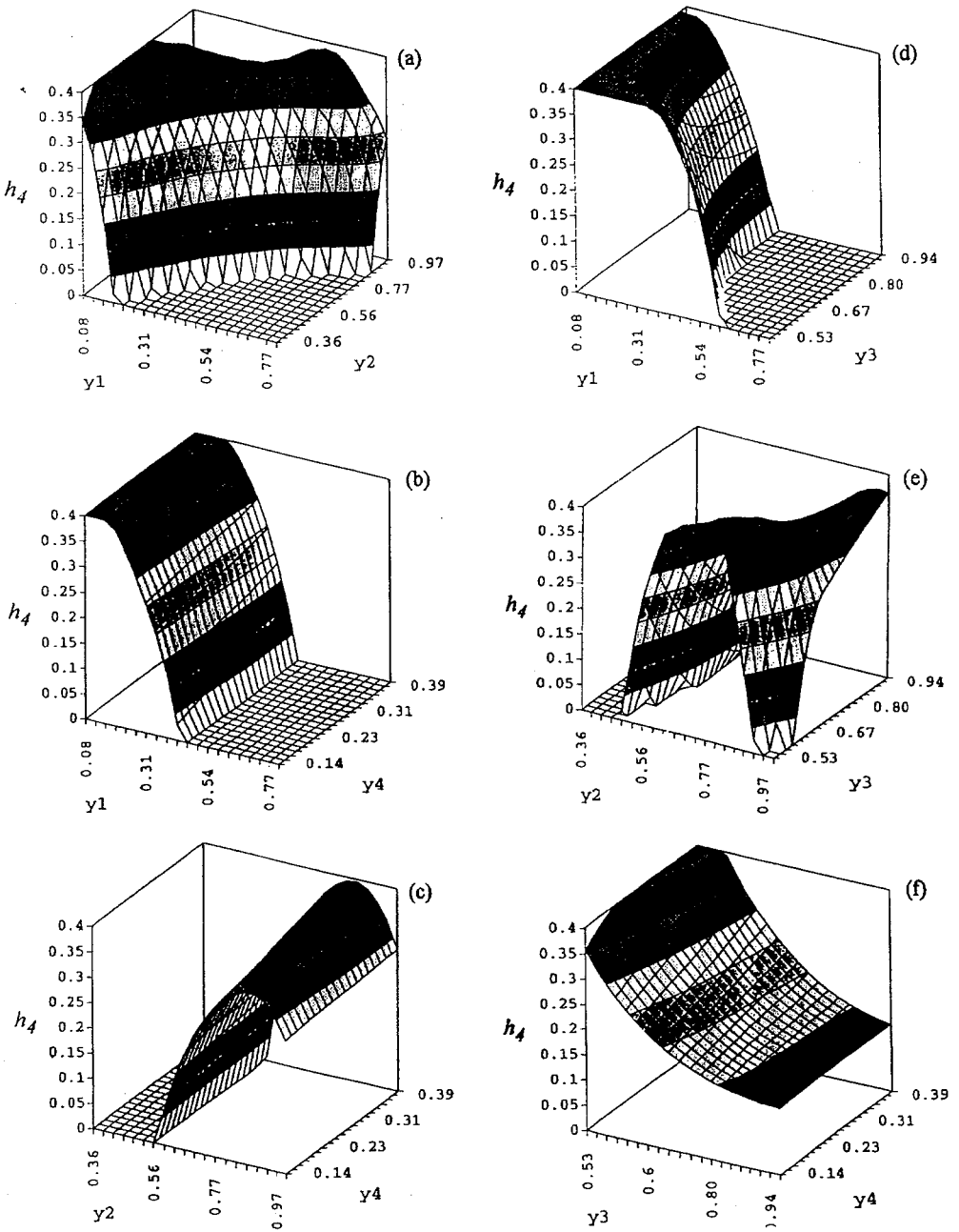
Następnie zmierzono odpowiedzi skokowe filtru posługując się kartą IBM PC z procesorem sygnałowym TMS320C25 (−5 *V* do 5 *V*, 12 bitów) [31], sterowaną programem napisanym w językach Assembler C25 i Turbo Pascal®. Program realizował generację wymuszania skokowego oraz akwizycję odpowiedzi i jej zapis na

Tabela 3

Błędy identyfikacji parametrów układu z rys. 7 otrzymane dla obserwacji zmierzonych w układzie ze wzmacniaczem operacyjnym μA741

	SIEĆ NEURONOWA				MET. NAJMNIEJSZYCH KWADR.			
τ, ms	$\eta_1+\beta_1$	$\eta_2+\beta_2$	$\eta_3+\beta_3$	$\eta_4+\beta_4$	η_1	η_2	η_3	η_4
0,0	0,0009	0,0071	0,0009	0,0022	0,0008	0,0089	0,0004	0,0029
0,1	0,0024	0,0058	0,004	0,0	0,0028	0,0065	0,0040	0,0
0,2	0,003	0,0051	0,0049	−0,0015	0,0037	0,0048	0,0064	−0,0011
0,3	0,0024	0,0097	0,006	−0,0031	0,002	0,0085	0,0051	−0,0006
0,4	0,0046	0,0076	0,0074	0,0005	0,0036	0,0059	0,0067	0,0004

dysku. Wyniki identyfikacji z użyciem zmierzonych odpowiedzi zestawione w Tabeli 3 dowodzą, że proponowana metoda daje dla rzeczywistych układów wyniki porównywalne z metodą najmniejszych kwadratów i może mieć zastosowanie



Rys. 9. Powierzchnie $h_4(y_1, y_2, y_3, y_4)$ generowane dla układu (36) przez estymator z siecią neuronową uczoną na przykładach obserwacji bez zakłóceń

w praktyce, na przykład do diagnostyki uszkodzeń parametrycznych w układach elektronicznych. Na rys. 9 wykreślono przykładowe przekroje hiperpowierzchni aproksymującej odwzorowanie obserwacji układu z rys. 7 na czas opóźnienia τ.

Średni czas obliczeń potrzebny do estymacji parametrów układu za pomocą metody najmniejszych kwadratów przy $n=4$ był równy 4.6 s (dla komputera PC386/33). Czas ten byłby wielokrotnie dłuższy w przypadku $n>p$ oraz przy zwiększonej mocy zakłóceń. Z drugiej strony, odpowiedzi wszystkich czterech sieci perceptronowych (o $n=4$ wejściach i $m=10$ neuronach każda) symulowanych za pomocą tego samego komputera szeregowego otrzymywano w czasie 8.6 ms od chwili dołączenia wektora obserwacji do wejścia sieci. Tym samym, proponowana metoda pozwala na ponad 500-krotne zmniejszenie czasu niezbędnego na identyfikację parametrów przykładowego układu. Czas ten byłby jeszcze krótszy w przypadku użycia specjalizowanych układów sieci neuronowych, o strukturze równoległej.

7. WNIOSKI

W artykule przedstawiono nową metodę identyfikacji parametrów układu dynamicznego wykorzystując sztuczne sieci typu „feedforward". W metodzie tej sieć neuronowa aproksymuje wielowymiarową funkcję próbek odpowiedzi układu na wybrany sygnał testowy. Sieć taka jest uczona na przykładach odpowiedzi obliczonych lub zmierzonych dla znanych wartości parametrów. Wykazano, że tak sformułowany problem identyfikacji ma jednoznaczne rozwiązanie jeśli (a) liczba próbek (obserwacji) jest równa liczbie nieznanych parametrów, $n=p$, oraz gdy (b) jakobian obserwacji względem parametrów jest niezerowy w każdym punkcie rozważanej podprzestrzeni parametrów. Sieć jest uczona poprzez takie dostrajanie wag jej wewnętrznych połączeń aby zminimalizować średniokwadratowy błąd aproksymacji w zbiorze przykładów. Jeśli błąd ten jest dostatecznie mały, to dla małej mocy zakłóceń macierz kowariancji estymatora wykorzystującego sieć neuronową jest praktycznie równa macierzy kowariancji estymatora średniokwadratowego zdefiniowanego na tych samych obserwacjach. Oba estymatory, proponowany i tradycyjny są w tych warunkach równoważne. Zaproponowano dwustopniową strukturę sieci neuronowej dla przypadku, w którym liczba obserwacji jest większa od liczby nieznanych parametrów, $n>p$.

Proces uczenia, chociaż czasochłonny, jest realizowany tylko raz dla danego układu i danego zakresu zmienności jego parametrów. Po zakończeniu uczenia, identyfikacja nie wymaga prowadzenia obliczeń iteracyjnych. Proponowany estymator jest zatem pozbawiony problemów ze zbieżnością, jaki są typowe dla tradycyjnej metody najmniejszych kwadratów [2].

Pokazano i uzasadniono analitycznie, że estymator neuronowy staje się obciążony, gdy obserwacje zakłócone są wykorzystane w procesie uczenia. Jednocześnie wariancja charakteryzująca rozrzut wyników identyfikacji zostaje zredukowana. Istnieje możliwość uzyskiwania kompromisu między wzrostem obciążenia i maleniem wariancji, poprzez odpowiednią optymalizację procesu uczenia. W wyniku tej

właściwości, wywołany zakłóceniami rozrzut parametrów może być mniejszy niż w przypadku estymatora tradycyjnego. Dotyczy to szczególnie ważnego dla praktyki dużego poziomu zakłóceń. Przedstawiono wyniki symulacji komputerowej i pomiarów potwierdzające zasadność rozważań teoretycznych.

Proponowana metoda identyfikacji stanowi podstawę nowej klasy przyrządów pomiarowych dla szybkich pomiarów pośrednich. Przed etapem wdrożenia metody konieczne są badania ukierunkowane na poszukiwanie nowych architektur sieci neuronowych do zadań aproksymacji funkcji wielu zmiennych. Sieci takie powinny zapewnić mały błąd aproksymacji funkcji (na ogół gładkiej dla większości układów dynamicznych) przy jak najmniejszej liczbie współczynników wagowych. Licza wag powinna wolno wzrastać z liczbą nieznanych parametrów układu oraz z zakresami zmienności tych parametrów. Uczenie takiej sieci nie powinno zajmować długiego czasu komputera. Powinny też one nadawać się do realizacji pod postacią układów elrktronicznych.

BIBLIOGRAFIA

1. K. Baszczyński, A. Materka: *Metoda komputerowaj identyfikacji dla wybranej klasy układów dynamicznych rzędu drugiego*. Kwartalnik Elektroniki i Telekomunikacji, t. *38*, z. *1*, 1992, ss. 17–33.
2. A. Cadzow: *Signalprocessing via least-squares modelling*. IEEE ASSP Magazine, Oct. 1990 s. 12–31.
3. A. Materka: *Application of neural networks to dynamic system parameter estimation*. I4th Annual International Conference IEEE Engineering in Medicine and Biology Society, *3*, Paris, France, 1992, ss. 1042–1045.
4. A. Materka, M. Telfer: *Model identification for time-invariant nonlinear system with delay using feedforward neural networks*. 2nd International Conference Modelling and Simulation, *1*, Melbourne, Australia, 1993, ss. 125–131.
5. A. Ubenhauen, G. P. Rao: *Identification of continuous systems*. Amsterdam: Noth Holland, 1987.
6. A. Materka: *Fast and robust edge evaluation using feedforward neural networks*. IEEE Workshop Visual Signal Processing and Communications, Melbourne, Australia, 1993, ss. 141–144.
7. A. Materka: *Analog circuit soft-fault diagnosis using neural networks*. 12th Australian Microelectronics Conference, Gold Coast, Australia, 1993, ss. 7–12.
8. A. Materka: *Fast and accurate identification of electronic circuit parameters using regularised feedforward neural networks*. International Joint Conference on Neural Networks, Nagoya, Japan, 1993, ss. 509–512.
9. A. Materka: *Parametric restoration of noisy blurred signal using feedforward neural networks*. Australasian Conference on Physical Sciences in Medicine and Biomedical Engineering, Melbourne, Australia, 1993, s. 149.
10. G. A. F. Seber and C. J. Wild: *Nonlinear regression*. New York: John Wiley & Sons, 1989.
11. F. Pukelsheim: *Optimal design of experiments*. New York: Wiley, 1993.
12. J. V. Beck and V. J. Arnold: *Parametr estimation in engineering and science*. New York: John Wiley & Sons, 1977.
13. A. C. Atkinson and A. N. Donev: *Optimum experimental design*. Oxford: Oxford Science Publications, 1992.
14. R. Hecht-Nielsen: *Neurocomputing*. Reading: Addison-Wesley, 1990.
15. R. P. Lippmann: *An introduction to computing with neural nets*. IEEE ASSP Magazine, April 1987, pp. 4–21.

16. K. Hornik: *Multilayer feedforward neural networks are universal approximators.* Neural Networks, 1989, **2**, pp. 359–366.
17. P. D. Wasserman: *Advanced methods in neural computing.* New York: Van Nostrad Reinhold, 1993.
18. A. Materka: *Application of artificial neural networks to parameter estimation of dynamical systems.* 10th IEEE International Conference on Instrumentation and Measueremnt Technology, Hamamatsu, Japan, 1994, pp. 123–126.
19. J. W. Ponton, J. Klemens: *Alternatives to neural networks for inferntial measurements.* Computers Chem. Engng., 1993, **17**, No, 10, pp. 991–1000.
20. E. Oja: *Principal components, minor components and linear neural networks.* Neural Networks, 1992, **5**, pp. 927–935.
21. B. Widrow, M. A. Lehr: *30 years of adaptive neural networks.* Perceptron, Madaline and Backpropagation. Proceadings IEEE, 78, pp. 1416–1442, Sept. 1990.
22. S. Mizushina, T. Shimizu, T. Sugiura: *Non-invasive thermometry with multiplefrequency microwave radiometry.* Frontiers Med. Biol. Engng., 1992, **4**, No. 2, pp. 129–133.
23. S. Mizushina, Konsultacja osobista.
24. S. Bielawski: *Modele farmakokinetyczne.* Warszawa: WKiŁ, 1989.
25. E. Małafiej, A. Materka, P. Koman, A. Denys: *Użyteczność programów komputerowych w badaniach aktywności antybiotyków oraz oznaczeniach poziomu leku w organiźmie. Druga Konferencja „Komputery w Medycynie"*, Łódź, 1991, s. 256–262.
26. R. Stein: *Selecting data for neural networks. AI Expert*, February 1993, pp. 42–47.
27. *Turbo Pascal for Windows. Programmer's Guide.* Borland International, Inc. 1800 Green Hills Rd., PO Box 660001, Scotts Valley, CA 95067-0001, USA, 1991.
28. A. Materka: *Model ładunkowy tranzystora bipolarnego do analizy zniekształceń nieliniowych w układach elektronicznych.* Rozprawa doktorska, t. II, Politechnika Łódzka, 1978.
29. W. J. Krzanowski: Principles of multivariate analysis. Oxford: Clarendon Press, 1988.
30. P. W. Tuinenga: *SPICE a guide to circuit simulation & analysis using* PSPICE® Sydney: Prentice-Hall International, 1992.
31. *Model 250 data acquisition and signal processing board for the* IBM PC AT *and* ISA *bus compatibles.* Dalanco Spry, 89 Westland Avenue, Rochester, NY 14618, USA, 1991.

A. MATERKA

NEW METHOD OF DYNAMIC SYSTEM PARAMETER IDENTIFICATION USING ARTIFICIAL NEURAL NETWORKS

Summary

The paper postulates that feedforward artificial neural networks can be used for dynamic system parameter identification. The method proposed does not require iterative calculations and is thus highly suitable real-time identification. Noise-induced error analysis is carried out. It is shown that error variance is not higher than that obtainable from a standard least-square-error technique applied to the identification problem. Moreover, the new method offers possibilities of variance reduction through proper modification of the neural network training algorithm. Results of computer simulation and measurements are presented to confirm the validity of the theoretical analysis.

Key words: neural networks, error analysis, computer simulation.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 1. ss. 69–87

Błąd przesterowania przetworników napięciowych sygnałów impulsowych

KRZYSZTOF PACHOLSKI

Instytut Podstaw Elektrotechniki, Politechnika Łódzka

Otrzymano 1994.08.11

Autoryzowano do druku 1995.01.20

Przedmiotem artykułu jest, zdefiniowana przez autora, nowa składowa błędu przetwarzania przetwornikow napięciowych wartości średniej i skutecznej występująca przy impulsowych sygnałach wejściowych. Składowej tej nie uwzględniano dotychczas w opisie właściwości metrologicznych przetworników napięciowych. Przedstawiono ponadto sygnał wzorcowy, za pomocą którego okreslić można w procesie oceny dokładności przetworników graniczną (maksymalną) wartość nowej składowej błędu przetwarzania.

Słowa kluczowe: przetwornik napięciowy, błąd przesterowania.

1. WPROWADZENIE

W trakcie eksploatacji elektroniczne przetworniki pomiarowe sygnałów napięciowych mogą być przesterowane.

Jedną z przyczyn przesterowania przetworników jest przekroczenie górnej granicy zakresu przetwarzania przez wartość szczytową sygnału pomiarowego [8]. Przesterowanie takie nazywane jest niekiedy przesterowaniem amplitudowym. Przesterowanie amplitudowe występuje najczęściej w przetwornikach wzorcowanych za pomocą sygnału sinusoidalnego [9, 10, 12] przetwarzających sygnały odkształcone o dużym współczynniku szczytu tzn. sygnały impulsowe. Nie uwzględnienie przewidywanej wartości szczytowej sygnału pomiarowego przy doborze górnej granicy zakresu przetwarzania jest przyczyną dużej wartości błędu przetwornika. Ze względu na błąd taki przypadek należy uznać za niedopuszczalny w praktyce pomiarowej.

Drugą przyczyną przesterowania jest skończona szybkość zmian sygnału wyjściowego wzmacniaczy operacyjnych zastosowanych w układzie przetwornika [3, 4, 9, 10, 12]. Tego rodzaju przesterowanie nazwano przesterowaniem szybkościowym [12].

Występuje ono przy impulsowych sygnałach wejściowych, niezależnie od wartości szczytowej i częstotliwości tych sygnałów, nawet w przetwornikach przystosowanych do przetwarzania tego rodzaju sygnałów.

Przesterowanie szybkościowe jest przyczyną występowania w błędzie przetwarzania przetworników nowej składowej, nazwanej przez autora błędem przesterowania [12]. Składowej tej nie uwzględniano dotychczas w ocenie właściwości metrologicznych przetworników napięciowych. Bląd przesterowania wykorzystać można do określania przydatnosci przetworników do przetwarzania sygnałów impulsowych.

W artykule przedstawiono definicję błędu przesterowania elektronicznych przetworników sygnałów napięciowych oraz syntezę kształtu sygnału wzorcowego, za pomocą którego okreslić można graniczną wartość tego błędu.

2. PRZESTEROWANIE SZYBKOŚCIOWE PRZETWORNIKÓW NAPIĘCIOWYCH

W przemysłowej praktyce pomiarowej wykorzystywane są najczęściej elektroniczne przetworniki napięciowe wartości średniej i skutecznej.

Pomiar wartości średniej i skutecznej sygnałów napięciowych realizowany może być zarówno w technice analogowych jak i cyfrowej. Produkowane obecnie przetworniki analogowe są zarówno pod względem kosztów oraz parametrów konkurencyjne w stosunku do układów zrealizowanych w technice cyfrowej [1]. Dlatego metody analogowe są nadal rozwijane i udoskonalane [2, 7, 14]. Produkowane obecnie analogowe przetworniki wartości średniej sygnałów napięciowych charakteryzują się strukturą przedstawioną na rys. 1a [1].

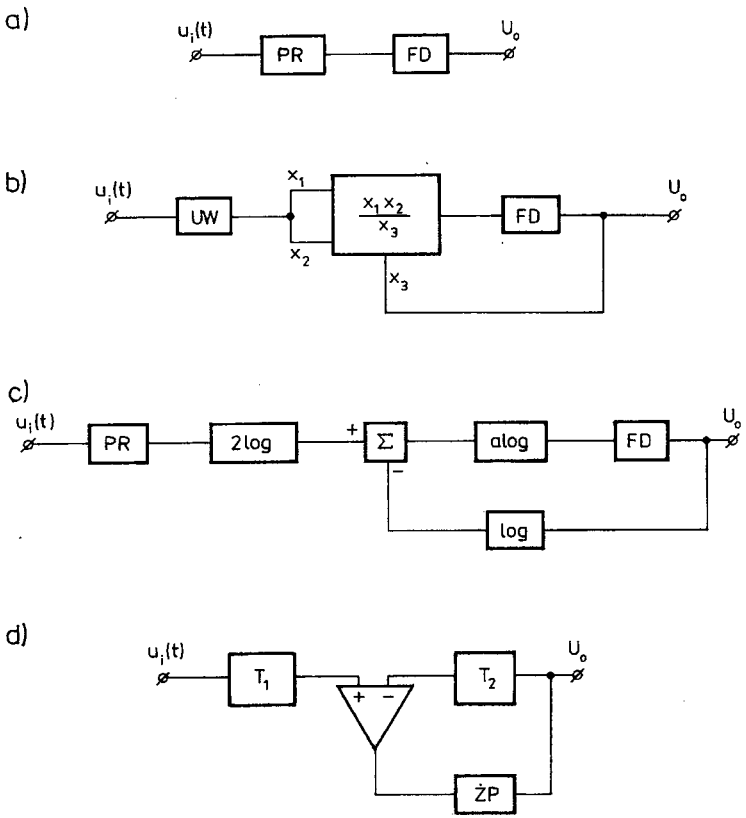
W układach tych przetworników do wyjścia dwupołówkowego prostownika operacyjnego dołączony jest aktywny filtr dolnoprzepustowy pełniący rolę układu uśredniającego. Współczesne analogowe przetworniki wartości skutecznej napięcia realizowane są w układach o strukturze zamkniętej [1, 7] (rys. 1b, c i d).

Zasadniczym elementem przetwornika z rys. 1b jest układ mnożąco-dzielący. Na wyjściu tego układu występuje sygnał proporcjonalny do kwadratu sygnału wejściowego przetwornika odniesiony do aktualnej wartości wielkości mierzonej tzn. do stałonapięciowego sygnału wyjściowego przetwornika. Takie rozwiązanie pozwala na realizację operacji pierwiastkowania w sposób pośredni. Zasadę działania przedstawionego układu opisuje następująca zależność [1]:

$$-\frac{1}{CU_0}\left[\frac{c^2}{T}\int_0^T u_i^2(t)dt\right]=CU_0 \tag{1}$$

gdzie: T — jest okresem sygnału wejściowego $u_i(t)$ przetwornika;
U_0, C — oznaczają odpowiednio sygnał wyjściowy oraz stałą przetwarzania przetwornika.

W zależności tej lewa strona równości reprezentuje uśredniony za pomocą filtru dolnoprzepustowego kwadrat sygnału wejściowego podzielony przez wartość skuteczną.



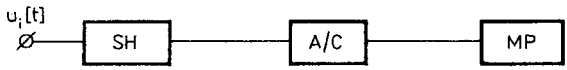
Rys. 1. Przetwornik wartości średniej (a) oraz przetworniki wartości skutecznej: z układem mnożąco-dzielącym b), log-alog z odejmowaniem logarytmów c), z przetwornikami termicznymi d) (FD — filtr dolnoprzepustowy; PR — dwupołówkowy prostownik operacyjny; T_1, T_2 — przetworniki termiczne, UW — układ wejściowy, ŹP — źródło prądowe)

Na rys. 1c przedstawiono wersję układu realizującego metodę mnożenia-dzielenia sygnałów [7]. Sygnał wejściowy przetwarzany jest początkowo przez dwupołówkowy prostownik operacyjny oraz w układzie logarytmującym dwukrotnie. Do wyjścia układu logarytmującego dołączony jest sumator, w którym realizowana jest operacja odejmowania: $2\log[u_i(t)] - \log U_0$. Sygnał wyjściowy sumatora przetwarzany jest następnie przez układ alogarytmujący i uśredniany w układzie filtru dolnoprzepustowego. Na wyjściu filtru występuje sygnał proporcjonalny do wartości skutecznej sygnału wejściowego przetwornika.

Najdokładniejszą aktualnie metodę pomiaru wartości skutecznej ilustruje rys. 1d [1]. Sygnał niesinusoidalny doprowadzany do przetwornika termicznego T_1 zostaje w nim przetworzony w odpowiedni sygnał stałoprądowy. Do drugiego przetwornika termicznego T_2 doprowadzony jest prąd stały ze źródła prądowego ŹP, które sterowane jest różnicą sygnałów wyjściowych obu przetworników. Jeżeli charakterystyki obu przetworników będą identyczne, to przy zerowej różnicy

sygnałów wyjściowych przetworników prąd stały źródła prądowego będzie dokładnie równy wartości skutecznej sygnału wejściowgo przetwornika.

Cyfrowe przetworniki wartości średniej i skutecznej sygnałów realizowane są w układzie przedstawionym na rys. 2 [1].

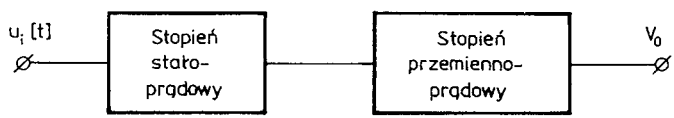


Rys. 2. Próbkujący przetwornik napięciowy (SH — układ próbkująco-pamiętający; A/C — przetwornik analogowo-cyfrowy; MP — system mikroprocesorowy)

W tego rodzaju przetwornikach sygnał pomiarowy doprowadzony jest do układu próbkująco-pamiętającego. Próbki napięcia za pośrednictwem przetwornika analogowo-cyfrowego przekształcane są w odpowiadający im sygnał cyfrowy. Sygnał ten doprowadzany jest do układu mikroprocesorowego, którego zadaniem jest obliczenie wartości średniej lub skutecznej sygnału pomiarowego.

Analizując sygnały wyjściowe poszczególnych bloków przedstawionych przetworników autor stwierdził, że szybkościowo mogą być przesterowane bloki wejściowe przetwarzające zmienny lub przemienny sygnał analogowy. Ze względu na przebieg przetwarzanego sygnału przesterowane szybkościowo bloki wejściowe przetwornika autor nazwał stopniem przemiennoprądowym [12].

Stopień przemiennoprądowy przetworników analogowych stanowią bloki przetwornika poprzedzające układ uśredniający. Wejściowym stopniem przemiennoprądowym przetworników próbkujących jest układ próbkująco-pamiętający poprzedzony w niektórych rozwiązaniach przetworników filtrem dolnoprzepustowym [12]. Dołączony do wyjścia stopnia przemiennoprądowego, w przetwornikach analogowych, układ uśredniający oraz bloki stałonapięciowej bądź stałoprądowej pętli sprzężenia zwrotnego autor określił mianem stopnia stałoprądowego [7] — rys. 3.



Rys. 3. Struktura analogowych przetworników napięciowych

Ze względu na zasadę działania w przetwornikach próbkujących stopień stałoprądowy nie występuje. Z prac [9, 10, 12, 13] autora wynika, że przesterowany szybkościowo stopień przemiennoprądowy przetworników, niezależnie od jego struktury, zachowuje się podobnie jak przesterowany szybkościowo wzmacniacz operacyjny [12]. Uwzględniając ten fakt w sygnale wyjściowym przesterowanego stopnia przemiennoprądowego $u_{op}(t)$ autor wyodrębnił dwie składowe [12]:

— składową $u_s(t)$ sterowaną sygnałem pomiarowym,

— składową $u_z(t)$ niesterowaną sygnałem pomiarowym:

$$u_{op}(t)=\begin{cases} u_s(t) & \text{dla} \quad t\in\langle 0,\ T\rangle\backslash\Delta_z \\ u_z(t) & \text{dla} \quad t\in\langle 0,\ T\rangle\backslash\Delta_s \end{cases}. \tag{2}$$

Składowe $u_s(t)$ i $u_z(t)$ w ciągu okresu T sygnału wejściowego przetwornika są określone w rozdzielnych zbiorach przedziałów czasowych Δ_s i Δ_z (tzn. $\Delta_s \cup \Delta_z = \phi_i$ ϕ — jest zbiorem pustym). Dla przetworników napięciowych stanowiących przedmiot rozważań składową $u_s(t)$ opisuje funkcjonał ciągły, którego postać określona jest strukturą stopnia przemiennoprądowego [12]:

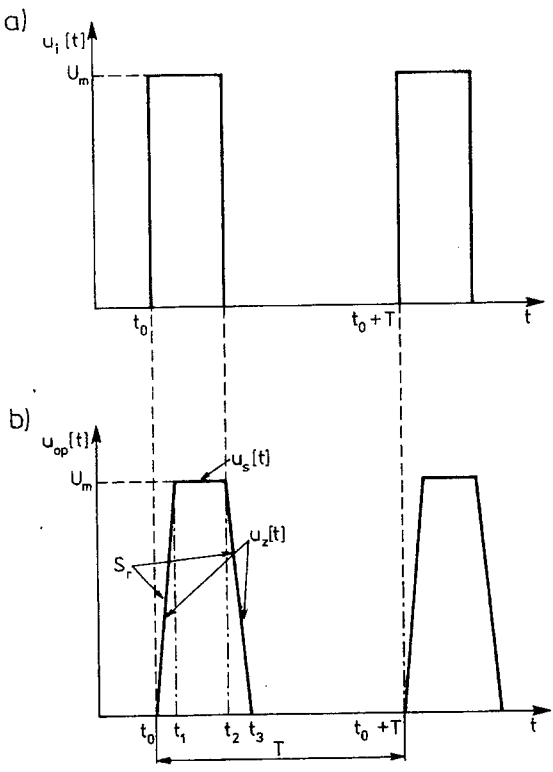
$$u_s(t)=F_s[u_i(t),\ k(t),\ t)], \tag{3}$$

gdzie: $k(t)$ — oznacza zastępczą małosygnałową impulsową funkcję przejścia stopnia przemiennoprądowego.

Składową $u_z(t)$ określa przesterowanie szybkościowe stopnia przemiennoprądowego i opisana jest zależnością [12]:

$$u_z(t)=u_s(t_1)\pm S_r(t-t_1). \tag{4}$$

W zależności tej t_1 oznacza początek 1-tego przedziału czasowego, w którym stopień przemiennoprądowy jest przesterowany ($\langle t_l,\ t_{l+1}\rangle \in \Delta_z$). S_r jest maksymalną szybkością zmian sygnału wyjściowego stopnia przemiennoprądowego. Parametr ten określa



Rys. 4. a) Prostokątny sygnał wejściowy przetwornika $u_i(t)$ oraz b) sygnał wyjściowy stopnia przemiennoprądowego $u_{op}(t)$ wraz ze składowymi $u_s(t)$ i $u_z(t)$

maksymalna szybkość zmian sygnału wyjściowego bloku stopnia przemiennoprądowego o najwęższym paśmie przetwarzania mocy [3, 4, 12].

Zależność (2) ilustruje rys. 4. Na rysunku tym przedstawiono przebieg sygnału wyjściowego $u_{op}(t)$ stopnia przemiennoprądowego z uwzględnieniem składowych $u_s(t)$ oraz $u_z(t)$ (rys. 4b) przy prostokątnym sygnale wyjściowym (rys. 4a).

Z pracy [12] autora wynika, że zał. (2) obowiązuje dla sygnałów wejściowych, których częstotliwość należy do zakresu niskich lub średnich częstotliwości pasma przetwarzania przetworników. Przy impulsowych sygnałach wejściowych o częstotliwości zbliżonej do górnej częstotliwości pasma przetwarzania przetworniki mogą być przesterowane przez cały czas trwania sygnału wejściowego [12]. Dla takich sygnałów w zał. (2) występuje jedynie niesterowana sygnałem pomiarowym składowa $u_z(t)$.

Aby opis sygnału wyjściowego $u_{op}(t)$ przesterowanego stopnia przemiennoprądowego za pomocą zał. (2) był kompletny należy określić granice przedziałów czasowych, w których stopień ten jest przesterowany. W tym celu należy w układzie przetwornika zlokalizować blok o najwęższym pasmie przetwarzania mocy. Następnie dla wyróżnionego bloku należy określić granice przedziałów czasowych, w których jest on przesterowany. Z prac [9, 10, 12] autora wynika, że jest to możliwe jedynie przy znajomości kształtu sygnału wejściowego oraz układu stopnia przemiennoprądowego przetwornika. W pracach tych przedstawiono metodykę wyznaczania ww granic poszukiwanych przedziałów czasowych dla przetworników napięciowych wartości średniej i skutecznej z prostownikiem operacyjnym na wejściu. Sposób ten może być stosowany dla przetworników z innymi układami operacyjnymi.

3. BŁĄD PRZESTEROWANIA PRZETWORNIKÓW NAPIĘCIOWYCH

Właściwości przyrządów i przetworników pomiarowych odwzorować można dwoma sposobami [6]. Pierwszy sposób polega na wykorzystaniu modelu matematycznego uwzględniającego interesujące nas parametry i właściwości przyrządu. Drugi ze sposobów polega na określeniu odpowiedniej miary błędu przetwarzania przetworników.

Opracowanie modelu matematycznego przesterowanego szybkościowo przetwornika napięciowego wymaga znajomość struktury jego układu oraz kształtu sygnału pomiarowego [12].

Produkowane obecnie przetworniki napięciowe wykonywane są najczęściej w postaci monolitycznych lub hybrydowych układów scalonych [12]. Ze względu na tajemnicę patentową uzyskanie szczegółowych informacji o układzie przetworników jest niemożliwe. Również nieznany jest najczęściej ksztalt sygnału pomiarowego. Z punktu widzenia praktyki pomiarowej bezpośrednie wykorzystanie modelu matematycznego do oceny właściwości metrologicznych przetworników napięciowych nie jest dogodne. Wynika to stąd, że model matematyczny nie dostarcza wprost informacji o błędzie przetwarzania. Parametry modelu wykorzystać można wprost jedynie do porównania właściwości przetworników pomiarowych.

Ze względu na wymienione czynniki właściwości przetworników napięciowych przy impulsowych sygnałach wejściowych charakteryzować należy za pomocą granicznej wartości odpowiedniej miary błędu przetwarzania obowiązującej dla określonej klasy sygnałów pomiarowych. Pod pojęciem *granicznej* wartości błędu należy rozumieć jego wartość maksymalną [6]. Konieczność wprowadzenia nowej miary błędu uzasadniona jest również tym, że nie znany jest dotychczas sposób uwzględniania wpływu przesterowania szybkościowego na właściwości metrologiczne przetworników.

Autor proponuje, aby wpływ przesterowania szybkościowego na dokładność przetworników uwzględnić za pomocą błędu zdefiniowanego zależnością (5) zwanego dalej błędem przesterowania [12]:

$$\varepsilon_p = F_N[S_r,\ u_i(t),\ k(t),\ t] - F_{Nw}[u_i(t),\ k_w(t),\ t]\,. \tag{5}$$

W zależności tej $F_N[-]$ oraz $F_{Nw}[-]$ są ogólnym opisem matematycznym przetworników: rzeczywistego i wzorcowego. Przetwornik rzeczywisty opisuje funkcjonał nieliniowy $F_N[-]$ zależny od parametru S_r stopnia przemiennoprądowego. Parametry przetwornika wzorcowego powinny być tak dobrane aby jego właściwości opisywał funkcjonał $F_{Nw}[-]$ zależny jedynie od zastępczej impulsowej funkcji przejścia $k_w(t)$ i niezależny od parametru S_r.

Zależnośc (5) obowiązuje dla sygnałów u_i(t), przy których w sygnale wyjściowym stopnia przemiennoprądowego $u_{op}(t)$ występują obie składowe $u_z(t)$ oraz $u_s(t)$ uwzględnione w zależności (2) [12].

Do celów porównawczych korzystniejsze jest przedstawienie błędu przesterowania przetworników (zał. 5) w postaci względnej odniesionej do rzeczywistej wartości wielkości mierzonej [12]:

$$\delta_p = \frac{F_N[S_r, u_i(t), k(t), t] - F_{Nw}[u_i(t), k_w(t), t]}{F_{Nw}[u_i(t), k_w(t), t]}\,. \tag{6}$$

Błąd przesterowania w postaci względnej pozwala na bezpośrednie porównanie właściwości metrologicznych, przy impulsowych sygnałach wejściowych, przetworników o różnych zakresach przetwarzania [6, 12].

4. OKREŚLANIE GRANICZNEGO BŁĘDU PRZESTEROWANIA

Wykorzystanie błędu przesterowania zdefiniowanego zależnością (5) w celu porównania przydatnosci przetworników do przetwarzania sygnałów impulsowych wymaga znajomości kształtu sygnału pomiarowego [12]. W praktyce dokladne określenie kształtu sygnału pomiarowego jest najczęściej niemożliwe.

W sytuacji zupełnej dowolności kształtu sygnału pomiarowego jedynie graniczna wartość błędu przesterowania stanowi podstawę do porównania właściwości metrologicznych przetworników przeznaczonych do przetwarzania sygnałów impulsowych.

Zagadnienie wyznaczenia granicznej wartości błędu przesterowania sprowadza się do określenia kształtu sygnału testowego ekstremalizującego miarę tego błędu [12]:

$$\varepsilon_{pgr} \overset{df}{=} \sup_{u_T(t)\in\mathcal{U}_T} |\varepsilon_p[u_T(t)]| \,. \tag{7}$$

W zależności tej ε_{pgr} oznacza graniczną wartość błędu przesterowania. $\mathcal{U}_T$ jest zbiorem sygnałów testowych $u_T(t)$ należącym do przestrzeni tzw. dopuszczalnych sygnałów pomiarowych $\mathcal{U}(\mathcal{U}_T \in \mathcal{U})$. W rozważaniach przyjęto, że przestrzeń $\mathcal{U}$ stanowią sygnały okresowe o okresie T całkowalne w kwadracie w sensie Lebesgue'a w przedziale $\langle 0, T\rangle$ $(\mathcal{U} \in L^2\langle 0, T\rangle)$ zmieniające w tym przedziale znak tylko jeden raz w punkcie t_p $(0 < t_p < T)$.

Poniżej przedstawiono sposób syntezy przebiegu sygnału testowego $u_T(t)$ spełniającego zależność (7). Aby definicja granicznego błędu przesterowania opisana zal. (7) miała sens, należy przyjąć dodatkowo założenie zwartości zbioru $\mathcal{U}_T$ oraz półciągłości z góry funkcjonału $\varepsilon_p[-]$ opisującego błąd przesterowania na zbiorze $\mathcal{U}_T$.

Nieliniowy charakter przesterowania szybkościowego przetworników nie pozwala na bezpośrednie zastosowanie do syntezy sygnału $u_T(t)$ znanych metod optymalizacji obowiązujących dla układów dynamicznych opisanych liniowymi równaniami różniczkowymi [5, 6]. Problem ten autor rozwiązał ustalając wstępnie kształt sygnałów stanowiących elementy zbioru $\mathcal{U}_T$ [12]. Następnie stosując zasady optymalizacji parametrycznej [15] określił ekstremalne wartości wybranych parametrów sygnału $u_T(t)$, dla których spełniony jest warunek kryterialny (7). Przy ustalonym kształcie sygnału $u_T(t)$, który charakteryzuje zbiór parametrów P $(u_T(t) = u_T(P, t))$ warunek ten przedstawić można w następującej postaci [12]:

$$\varepsilon_{pgr} = \max_{P} | \varepsilon_p[u_T(P, t)] | \,. \tag{8}$$

Sygnały reprezentujące zbiór $\mathcal{U}_T$ powinny, przede wszystkim, spełniać warunki określone twierdzeniem Riesza [5]. Z twierdzenia tego wynika, że warunkiem koniecznym i dostatecznym zwartości zbioru $\mathcal{U}_T$ w przestrzeni $L^2\langle 0, T\rangle$ jest istnienie stałej K takiej, że:

$$\sup_{u_T(t)\in\mathcal{U}_T} \int_0^T | u_T(t) |^2 dt \leqslant K^2 \tag{9}$$

i aby dla dowolnego $\xi > 0$ istniało $\eta > 0$ takie, że z warunku $| h | \leqslant \eta$ wynika:

$$\sup_{u_T(t)\in\mathcal{U}_T} \int_0^T | u_T(t+h) - u_T(t) |^2 \leqslant \xi^2. \tag{10}$$

Przy określeniu kształtu sygnałów $u_T(t)$ oprócz wymagań matematycznych należy uwzględnić fizyczne przyczyny przesterowania szybkościowego. Jako miarę przesterowania szybkościowego, przy ustalonej wartości parametru S_r, przetwornika autor przyjął czas tego przesterowania. W porównaniu z czasem przesterowania

przetworników przy sygnałach pomiarowych należących do przestrzeni dopuszczalnych sygnałów pomiarowych $\mathcal{U}$ czas ten dla sygnałów testowych $u_T(t)$ powinien mieć wartość maksymalną. Przy ustalaniu tego warunku wykluczono przypadek całkowitego przesterowania przetwornika. Przesterowanie takie występuje przy impulsowych sygnałach pomiarowych o małym wypełnieniu. Wówczas sygnałem wyjściowym stopnia przemiennoprądowego jest sygnał trójkątny, którego zbocza mają nachylenie równe parametrowi S_r przetwornika [3, 4, 12].

Przykładem sygnału spełniającego przedstawione wymagania jest unipolarny impulsowy sygnał prostokątny (rys. 4a) stosowany do oceny wpływu odkształcenia sygnału pomiarowego na właściwości metrologiczne przetworników napięciowych [12]. Wpływ ten charakteryzowany jest za pomocą dodatkowego błędu przetwarzania określanego przy wymaganej dla danego przetwornika wartości współczynnika szczytu ww sygnału.

Aby przy określaniu granicznej wartości błędu przesterowania można było uwzględnić, dopuszczalną dla danego przetwornika, wartość współczynnika szczytu odkształconego sygnału pomiarowego sygnał testowy $u_T(t)$ musi charakteryzować się takimi samymi właściwościami jak unipolarny sygnał prostokątny [12]. Biorąc to pod uwagę przyjęto, że zbiór $\mathcal{U}_T$ stanowią impulsowe unipolarne quasi-prostokątne sygnały trapezowe. Sygnały te w porównaniu z idealnym sygnałem prostokątnym (rys. 4a), różnią się jedynie nachyleniem zboczy [12], tzn.:

$$\mathcal{U}_T \equiv \{u_T(P, t): P \equiv [S_{tn}, S_{t0}];\ S_{tn} \geqslant S_r;\ |\ S_{t0}\ | \geqslant S_r\}, \tag{11}$$

gdzie: S_{tn}, S_{t0} — oznaczają nachylenie zboczy: narastającego i opadającego sygnałów należących do zbioru $\mathcal{U}_T$.

Parametrami sygnału $u_T(t)$ względem których określano maksymalną wartość błędu przesterowania przetworników są nachylenia zboczy narastającego S_{tn} i opadającego S_{t0} tego sygnału. Założono przy tym, że sygnał $u_T(t)$ charakteryzuje się częstotliwością, przy której wpływ skończonej szerokości pasma przetwarzania stopnia przemiennoprądowego na błąd przetwarzania przetworników jest do pominięcia [12].

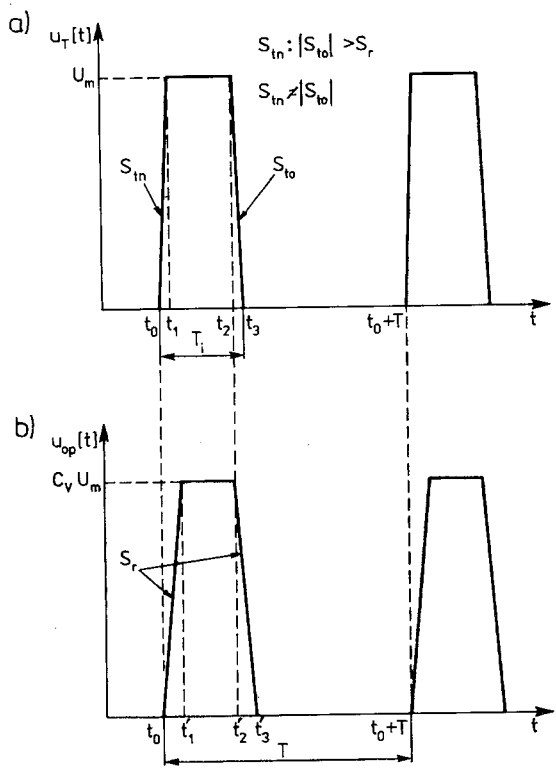
Sygnał wyjściowy $u_{op}(t)$ stopnia przemiennoprądowego rozpatrywanych układów przy trapezowym sygnale wejściowym $u_T(t)$ (rys. 5a) o wartości szczytowej U_m mniejszej lub równej parametrowi U_{mx} przetwornika ($U_m \leqslant U_{mx}$) ma kształt przedstawiony na rys. 5b. Wartość parametru U_{mx} określa górna granica zakresu przetwarzania przetwornika [12].

Uwzględniając założenia błąd przesterowania przetworników opisano odpowiednio wyrażeniami [12]:

— dla przetwornika wartości średniej

$$\delta_{pV} = \frac{U_{0V} - U_{TV}}{U_{TV}} = \frac{1}{2T} \frac{C_v U_m}{S_r} \left(\frac{S_r}{C_v S_{tn}} - \frac{S_r}{C_V S_{t0}} \right) \frac{1}{\eta}. \tag{12}$$

— dla przetwornika wartości skutecznej



Rys. 5. Sygnał wyjściowy stopnia przemiennoprądowego u_{op} (b) dla trapezowego sygnału wejściowego $u_T(t)$ (a)

$$\delta_{pR} = \frac{1}{2}\frac{U_{0R}^2 - U_{TR}^2}{U_{TR}^2} = \frac{1}{3T}\frac{C_v U_m}{S_r}\left(\frac{S_r}{C_v S_{tn}} - \frac{S_r}{C_V S_{t0}}\right)\frac{1}{\eta}. \tag{13}$$

W zależnościach (12) i (13):

C_v — jest stałą przetwarzania stopnia przemiennoprądowego przetworników.

U_{TV}, U_{TR} — oznaczają wartość średnią i skuteczną sygnału trapezowego $u_T(t)$ (rys. 5a):

$$U_{TV} = U_m\left[\eta - \frac{1}{2T}\frac{C_v U_m}{S_r}\left(\frac{S_r}{C_v S_{tn}} + \frac{S_r}{C_v S_{t0}}\right)\right],$$

$$U_{TR} = U_m\left[\eta - \frac{1}{3T}\frac{C_v U_m}{S_r}\left(2\frac{S_r}{C_v S_{tn}} + \frac{S_r}{C_v S_{t0}}\right)\right]^{\frac{1}{2}};$$

U_{0V}, U_{0R} — oznaczają sygnał wyjściowy przetwornika wartości średniej i skutecznej przy trapezowym sygnale wejściowym $u_T(t)$:

$$U_{0V} = U_m\left(\eta - \frac{1}{T}\frac{U_m}{S_r}\right),$$

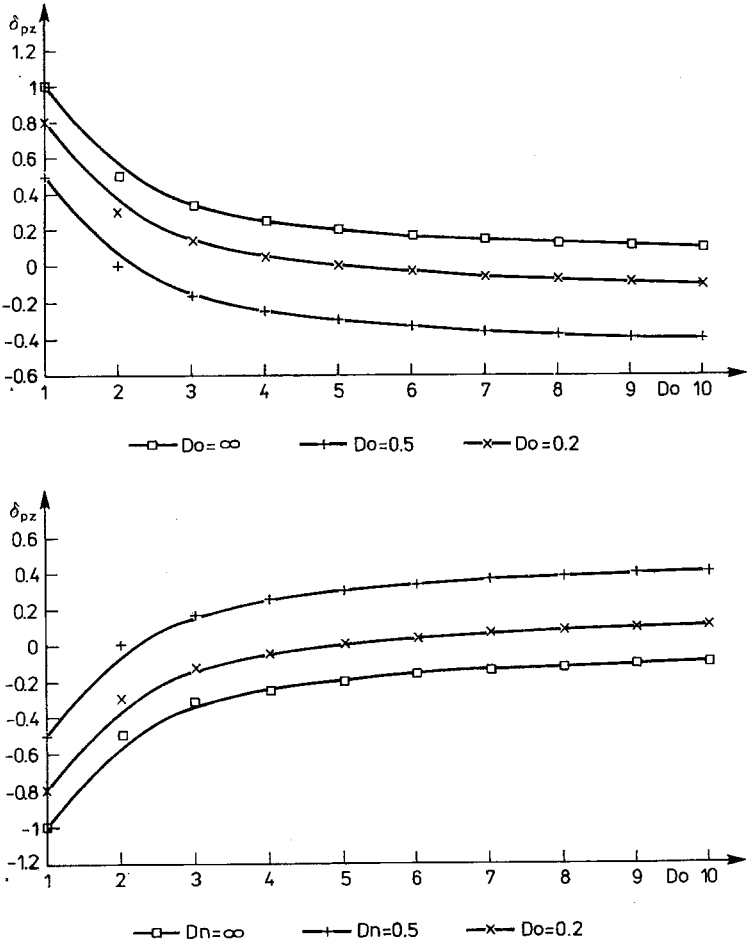
$$U_{0R} = U_m\left(\eta - \frac{1}{T}\frac{U_m}{S_r}\right)^{\frac{1}{2}}.$$

Aby wyniki analizy były niezależne od rodzaju przetwornika zależności (12) i (13) przekształcono do postaci:

$$\delta_{pz} = \frac{\delta_{pV}}{\frac{1}{2T}\frac{C_v U_m}{S_r \eta}} = \frac{\delta_{pR}}{\frac{1}{3T}\frac{C_v U_m}{S_r \eta}} = \left(\frac{S_r}{C_v S_{tn}} - \frac{S_r}{C_v S_{to}}\right) \tag{14}$$

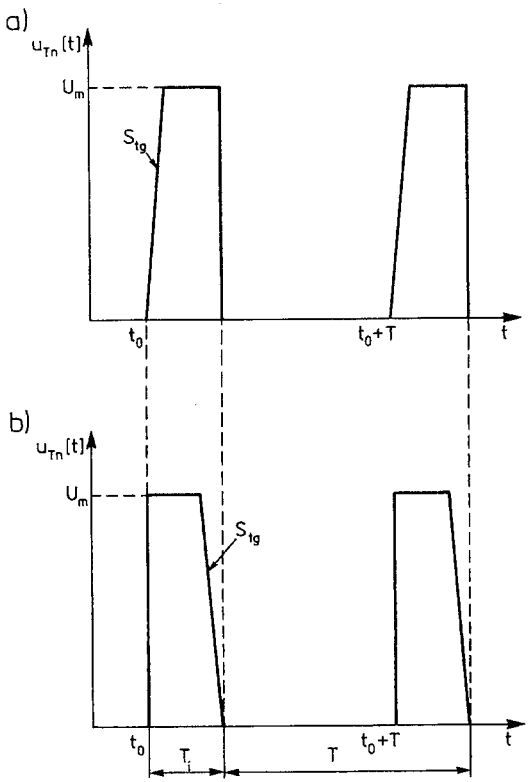
Wielkość δ_{pz} nazwano unormowanym błędem przesterowania przetworników napięciowych [12].

Z zał. (14) wynika, że dla sygnałów trapezowych, których zbocza mają nachylenie o takiej samej wartości bezwzględnej błąd przesterowania przetworników równy jest zeru niezależnie od wartości tego nachylenia. Z tego względu za pomocą trapezowych sygnałów wzorcowych stosowanych dotychczas do określania błędów dynamicznych przetwornikow nie można określać granicznego błędu przesterowania. Przebieg tego rodzaju sygnałów przedstawiono między innymi w pracy [6].



Rys. 6. a) Charakterystyka $\delta_{pz}=f(Dn)\left(Dn=\frac{C_v S_{tn}}{S_r}\right)$, b) $\delta_{pz}=f(D_0)\left(D_0=\frac{C_v S_{t0}}{S_r}\right)$

Interesujące rezultaty uzyskano analizując wpływ zmiany nachylenia tylko jednego zbocza sygnału trapezowego przy ustalonej stromości drugiego [12]. Wpływ zmian nachylenia jednego zbocza sygnału $u_T(t)$ na wartość błędu δ_{pz} ilustrują charakterystyki przedstawione na rys. 6.



Rys. 7. Asymetryczne trapezowe sygnały wzorcowe z deformacją kształtu: a) lewostronną $u_{Tn}(t)$, b) prawostronną $u_{T0}(t)$

Z przebiegu tych charakterystyk wynika, że błąd przesterowania przyjmuje wartości ekstremalne (tzn. $(\mid \delta_{pz} \mid = 1)$ dla trapezowych sygnałów wzorcowych z asymetryczną deformacją kształtu — rys. 7 [12].

Nachylenie jednego zbocza asymetrycznych sygnałów trapezowych równe jest wartości granicznej S_{tg}:

$$S_{tg} = \frac{S_r}{C_v}. \tag{15}$$

Zmniejszanie nachylenia tego zbocza poniżej wartości S_{tg} nie wpływa na wartość błędu δ_{pz}. Błąd δ_{pz} dla takich sygnałów ma wartość ekstremalną: $\mid d_{pz} \mid = 1$ (rys. 6).

Nachylenie drugiego zbocza sygnałów trapezowych z rys. 7 powinno charakteryzować się stromoscią zbliżoną, w idealnym przypadku, do nieskończonosci. W praktyce wystarczy aby nachylenie tego zbocza było co najmniej dziesięciokrotnie większe w porównaniu z nachyleniem zbocza o stromości granicznej [12]. Dokładność określania błędu przesterowania za pomocą takich sygnałów wzorcowych będzie mniejsza od 10%.

Ze względu na proporcjonalną zależność pomiędzy czasem przesterowania szybkościowego przetworników i wartością szczytową U_m trapezowych sygnałów wzorcowych (rys. 7) błąd przesterowania przyjmuje wartość graniczną wówczas, gdy $U_m = U_{mx}$ [12].

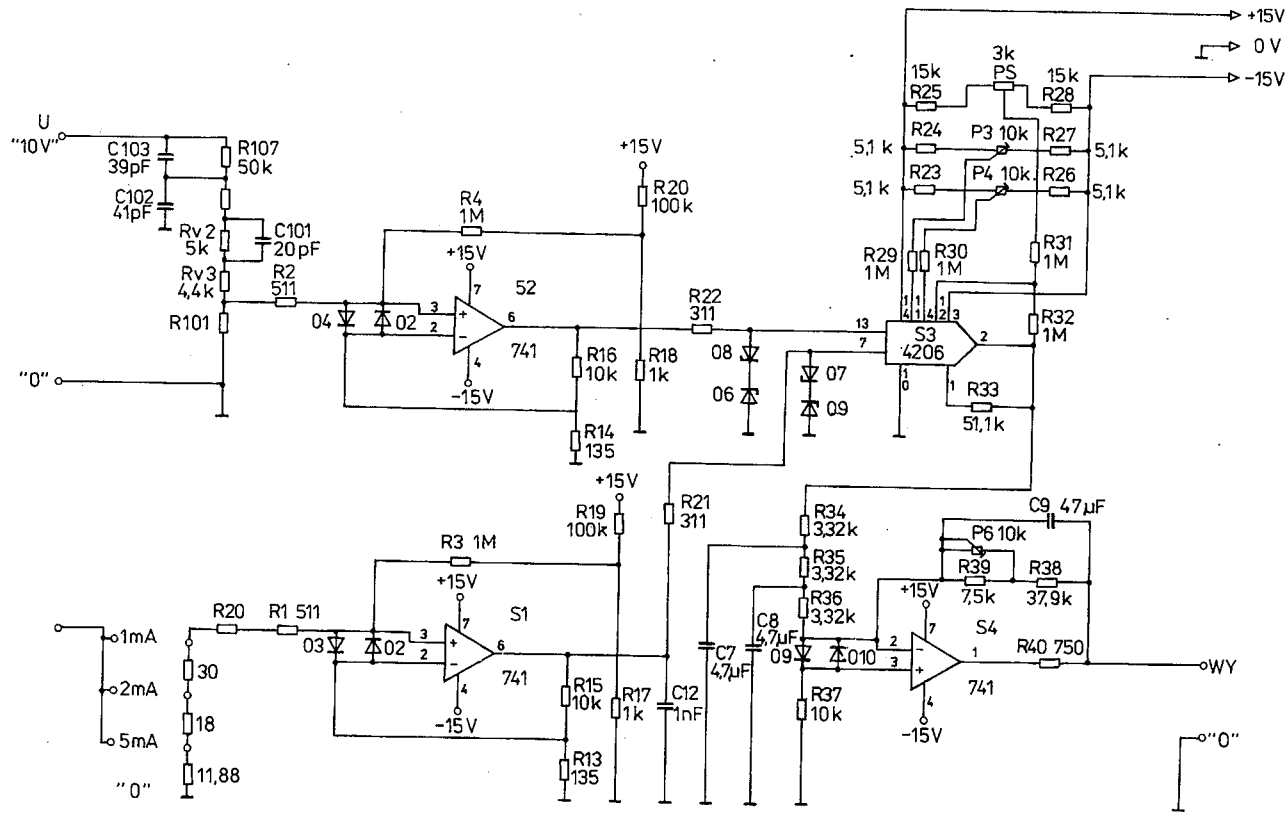
Użytkownicy przyrządów i przetworników pomiarowych nie znają najczęściej wartości współczynnika wzmocnienia C_v oraz parametru S_r stopnia przemienno-prądowego. W związku z tym autor proponuje aby przydatności przetworników do przetwarzania sygnałów impulsowych określać podając oprócz dopuszczalnej wartości współczynnika szczytu sygnałów pomiarowych również graniczną wartość nachylenia zboczy tych sygnałów. Ponadto przy charakterystyce właściwości metrologicznych tych przetworników należy uwzględniać, oprócz błędu dodatkowego wynikającego z odkształcenia sygnału pomiarowego, graniczną wartość błędu przesterowania.

Przebieg charakterystyk z rys. 6 wskazuje praktyczny sposób określania parametru S_{tg} przetwornika [11, 12]. Początkowo nachylenie zboczy sygnału trapezowego powinno być co najmniej dziesięciokrotnie większe w stosunku do przewidywanej wartości tego parametru. Nastepnie nachylenie jednego zbocza należy zmniejszać do wartości, przy której wartość bezwzględna różnicy sygnałów wyjściowych przetworników badanego i wzorcowego ma wartość maksymalna. Autor określił właściwości przetworników, które mogą być wykorzystane jako wzorcowe przy wyznaczaniu granicznego błędu przesterowania oraz parametru S_{tg} [11, 12]. Przetwornik taki w porównaniu z przyrządem badanym powinien, oprócz wymagań dokładnościowych obowiązujących dla porównawczej metody dokładności, charakteryzować się co najmniej dziesięciokrotnie szerszym pasmem przetwarzania mocy.

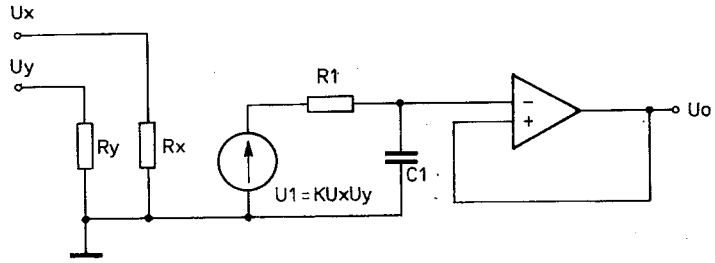
5. WYNIKI SYMULACJI KOMPUTEROWEJ ORAZ WYNIKI BADAŃ LABORATORYJNYCH

Rezultaty rozważań teoretycznych potwierdzają, przedstawione poniżej, wyniki symulacji komputerowej wpływu przesterowania na właściwości metrologiczne przetwornika pomiarowego typu WP-1 [2] wykorzystywanego jako przetwornik wartości skutecznej napięcia. Za pomocą tego przetwornika określać można również wartość skuteczną prądu oraz moc czynną odbiorników jednofazowych. Przetwornik przystosowany jest do przetwarzania sygnałów odkształconych o współczynniku szczytu mniejszym lub równym 3 w paśmie częstotliwości od 30 Hz do 5 kHz. Uproszczoną strukturę układu przedstawia rys. 8.

Układami wejściowymi torów prądowego i napięciowego przetwornika są układy zrealizowane na bazie wzmacniaczy operacyjnych S1 i S2 (typu μA 741). Sygnały wyjściowe tych układów doprowadzone są do logarytmującego układu mnożącego S3 typu 4203 firmy Burr – Brown. Układem wyjściowym przetwornika jest integrator wykorzystujący wzmacniacz operacyjny S4 (μA 741).



Rys. 8. Układ przetwornika pomiarowego typu WP-1



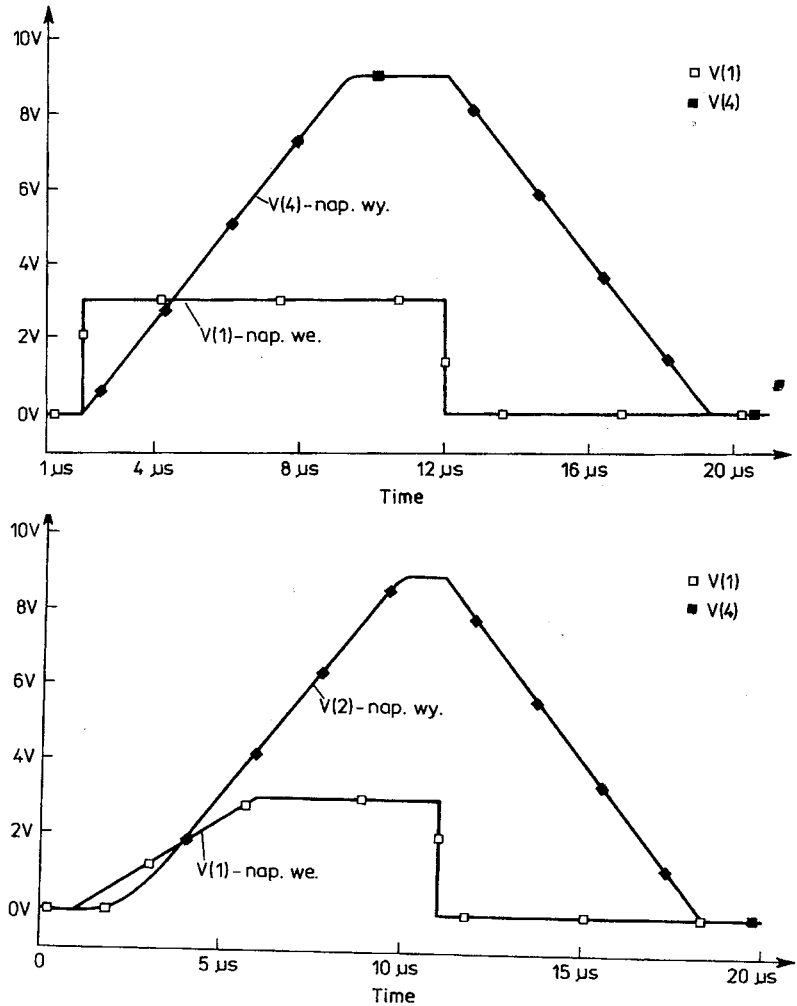
Rys. 9. Makromodel układu mnożącego

Badania symulacyjne przetwornika przeprowadzono za pomocą programu PSPACE. Do symulacji wzmacniaczy operacyjnych wykorzystano ogólnie znany makromodel, którego strukturę przedstawiono między innymi w pracy [4]. Ze względu na brak informacji o elementach logarytmującego układu mnożącego do symulacji tego układu zastosowano makromodel przedstawiony na rys. 9.

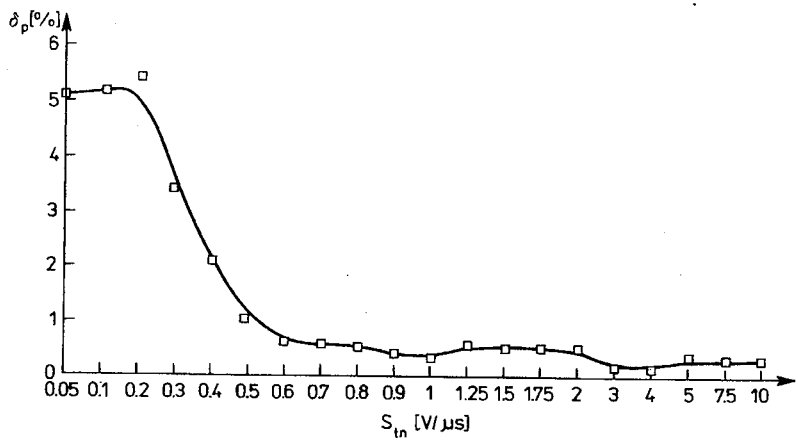
W makromodelu tym operacja mnożenia sygnałów wejściowych Ux i Uy realizowana jest przez sterowane źródło napięciowe $U1$ ($U1 = KUxUy$; $0 < K \leqslant 1$). Wartości rezystorów Rx i Ry odpowiadają rezystancji wejściowej torów X i Y układu rzeczywistego. Parametry obwodu R1C1 określają pulsację bieguna dominującego transmitancji napięciowej tego układu. W makromodelu maksymalną szybkość zmian sygnału wyjściowego ukladu rzeczywistego odwzorowuje parametr S_r wzmacniacza operacyjnego występującego w układzie wtórnika wyjściowego. O właściwym doborze struktury makromodelu układu mnożącego świadczą wyniki badań przetwornika. Przykład przebiegu sygnału wyjściowego układu mnożącego o stałej przetwarzania K równej 1 ($K=1$) w konfiguracji kwadratora dla prostokątnego oraz trapezowego sygnału wejściowego przedstawiono na rys. 10.

Badania symulacyjne przetwornika zrealizowano w dwóch etapach. W pierwszym etapie określono wpływ zmiany nachylenia zbocza narastającego S_{tn} trapezowego sygnału wzorcowego z lewostronną deformacją kształtu na wartość błędu przesterowania przetwornika. Przyjęto przy tym, że nachylenie drugiego zbocza sygnału wzorcowego równe jest nieskończoności. Wartość szczytowa, częstotliwość i wypełnienie tego sygnału były odpowiednio równe: 7 V, 2 kHz oraz 3. Wartość szczytowa sygnału trapezowego określona jest przez górną granicę badanego zakresu przetwarzania przetwornika tzn. zakresu 5 V RMS [2]. Błąd przesterowania przetwornika określono zakładając, że przetwornik wzorcowy jest idealnym przetwornikiem wartości skutecznej napięcia. Charakterystykę błędu przesterowania w funkcji parametru S_{tn} sygnału wzorcowego przedstawia rys. 11.

Z przebiegu tej charakterystyki wyznaczono graniczną wartość nachylenia zboczy S_{tg} sygnału wejściowego przetwornika. Przy tym nachyleniu błąd przesterowania przyjmuje wartośc maksymalną. Dla badanego modelu przetwornika parametr S_{tg} równy jest 0.2 V/μs.

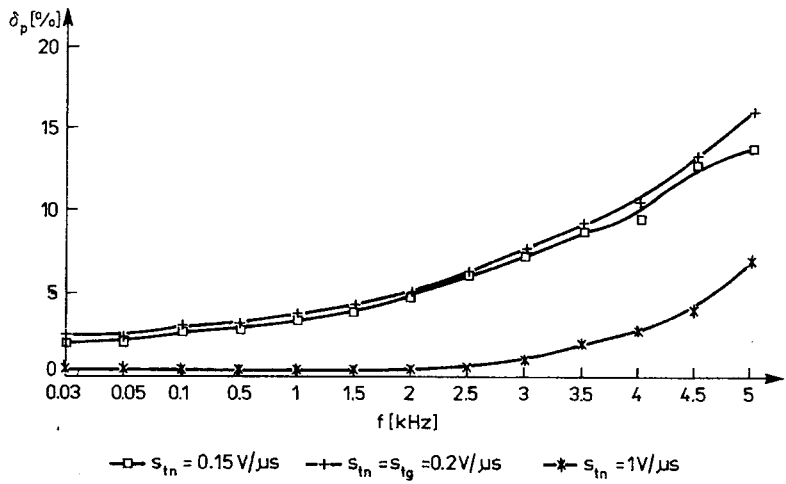


Rys. 10. Sygnał wyjściowy kwadratora dla prostokątnego a) i trapezowego b) sygnału wejściowego (V(1) jest sygnałem wejściowym, V(2) i V(4) oznaczają sygnał wyjściowy)



Rys. 11. Charakterystyka $\delta_p = f(S_{tn})$

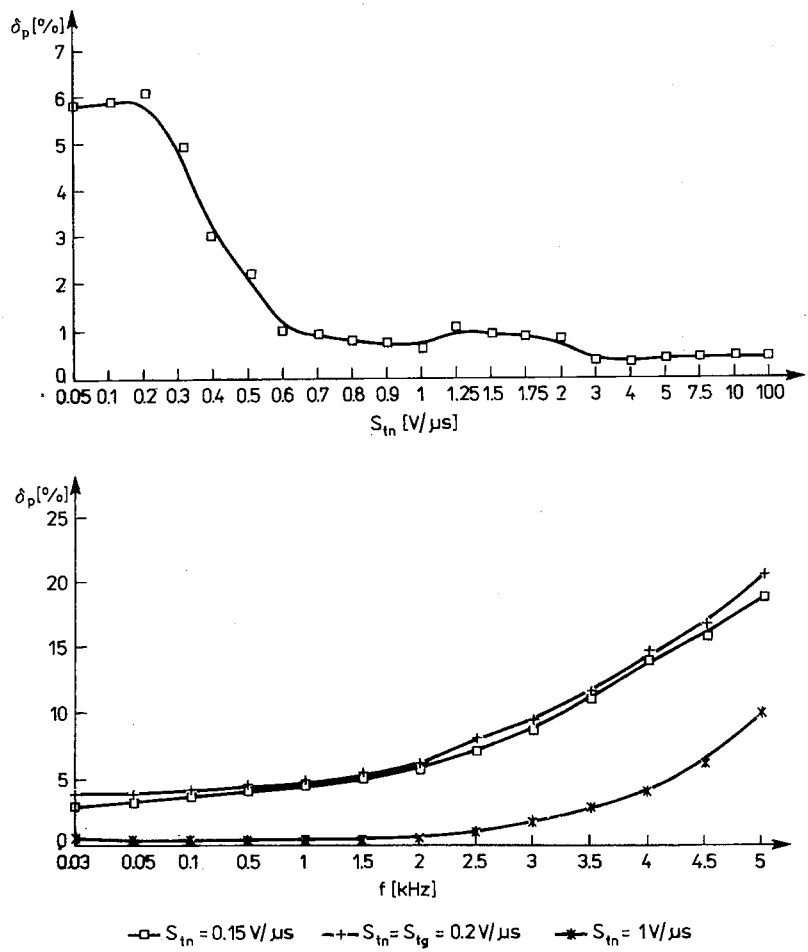
Kolejny etap badań symulacyjnych obejmował wyznaczenie charakterystyki błędu przesterowania w funkcji częstotliwości trapezowego sygnału wzorcowego z lewostronną deformacją kształtu. Nachylenie przedniego zbocza tego sygnału równe było wartości granicznej S_{tg} przy ustalonej wcześniej wartości szczytowej i współczynniku szczytu równym odpowiednio 7 V oraz 3. Wyniki badań w postaci graficznej przedstawiono na rys. 12.



Rys. 12. Charakterystyka $\delta_p = f(f)$

Na rys. 12 oprócz charakterystyki błędu przesterowania wyznaczonej za pomocą sygnału wzorcowego o nachyleniu zbocza narastającego równym S_{tg} ($S_{tg} = 0.2$ V/μs) przedstawiono charakterystyki błędu przesterowania wyznaczone za pomocą sygnału trapezowego, którego zbocze narastające ma nachylenie o wartości mniejszej oraz większej w porównaniu z S_{tg}.

Wyniki symulacji komputerowej potwierdziły badania laboratoryjne. Zakres tych badań był taki sam jak badań symulacyjnych. Trapezowy sygnał wzorcowy wykorzystywany w badaniach laboratoryjnych w porównaniu z sygnałem stosowanym w symulacji komputerowej różnił się jedynie nachyleniem zbocza opadającego. Sygnał ten charakteryzował się nachyleniem równym -0.05 V/μs. W badaniach laboratoryjnych rolę przetwornika wzorcowego pełnił oscyloskop cyfrowy firmy Tektronix typu 2221. Charakterystyki wyznaczone w trakcie badań laboratoryjnych przedstawiono na rys. 13.



Rys. 13. Empiryczne charakterystyki a) $\delta_p = f(S_{tn})$ oraz b) $\delta_p = f(f)$

PODSUMOWANIE

W artykule przedstawiono nowy sposób opisu właściwości metrologicznych przetworników napięciowych przeznaczonych do przetwarzania sygnałów impulsowych. Sposób ten uwzględnia wpływ tzw. przesterowania szybkościowego. Przyczyną przesterowania szybkościowego jest skończona szybkość zmian sygnału wyjściowego wzmacniaczy operacyjnych przetwornika.

Wpływ przesterowania szybkościowego na właściwości metrologiczne przetworników opisano ilościowo za pomocą zdefiniowanego przez autora tzw. błędu przesterowania. Przedstawiono również sygnał wzrocowy, za pomocą którego określić można w porównawczej metodzie oceny dokładności graniczną wartość tego błędu. Sygnałem tym jest unipolarny asymetryczny sygnał trapezowy.

BIBLIOGRAFIA

1. J. Bolikowski, L. Czarnecki, M. Miłek: *Pomiary wartości skutecznej i mocy w obwodach o przebiegach niesinusoidalnych.* PWN, Warszawa 1990
2. Era: *Instrukcja obsługi przetwornika pomiarowego typu WP-1*
3. W. Golde, L. Śliwa: *Wzmacniacze operacyjne i ich zastosowania — podstawy teoretyczne.* WNT, Warszawa 1982
4. M. Herpy: *Analog Integrated Circuits.* Akademiai Kiado, Budapest 1980
5. J. Kudrewicz: *Analiza funkcjonalna dla elektroników i automatyków.* PWN, Warszawa 1976
6. E. Layer, W. Gawędzki: *Dynamika aparatury pomiarowej. Badania i ocena.* WNT, Warszawa 1991
7. Materiały katalogowe firm Analog Device i Burr-Brown
8. E. Nowaczyk: *Dopuszczalny współczynnik szczytu sygnału jako kryterium oceny przyrządu pomiarowego.* Rozprawy Elektrotechniczne Nr 32, 1986, ss. 443–454
9. K. Pacholski: *Analiza dynamicznych właściwości amplitudowo-częstotliwościowych operacyjnych przetworników wartości średniej.* Rozprawy Elektrotechniczne Nr 2, 1989, ss. 295–321
10. K. Pacholski: *Właściwości metrologiczne dwupołówkowych prostowników operacyjnych dla odkształconych sygnałów napięciowych o dużym współczynniku szczytu.* Metrologia i Systemy Pomiarowe Nr 6, 1990, ss. 127–143
11. K. Pacholski: *Sposób oceny dokładności przetworników napięciowych.* Patent Nr P-2881189, 1989
12. K. Pacholski: *Modele metrologiczne napięciowych przetworników pomiarowych uwzględniające zjawisko przesterowania.* Zeszyty Nauk. Polit. Łódzkiej Nr 565, 1993
13. K. Pacholski: *The high-overloading error of voltage transducers.* Measurement (w druku)
14. Tabor Electronics, Ltd.: *Instrukcja obsługi przetwornika pomiarowego typu 4501B*
15. R. Zielonko: *Ogólne zasady projektowania kształtu sygnału pomiarowego.* Metrologia i Systemy Pomiarowe Nr 1, 1988, ss. 53–74

K. PACHOLSKI

THE OVERLOADING ERROR OF VOLTAGE PULSE SIGNALS TRANSDUCERS

Summary

The subject of a paper is a new component of the transduction error of mean-value and rms-value voltage transducers defined by the author. Its appear for impulse input signals. The new component of the error has not been included in the description of metrological properties of voltage transducers up to now. The standard signal which can be applied to determining the maximum value of the new component of the error, in accuracy assessment of voltage pulse signals transducers is presented too.

Key words: voltage transducer, overloading error.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 1. ss. 89–102

Komputerowa optymalizacja grubości słabo i silnie domieszkowanych warstw n-bazy w tyrystorze asymetrycznym

ZBIGNIEW LISIK

Instytut Elektroniki, Politechnika Łódzka

WOJCIECH CIEPLAK, MARIA LEŚKIEWICZ-KRAWCZYK, MAREK SMOLUCHOWSKI

Zakłady Elektronowe LAMINA

Otrzymano 1994.09.24

Autoryzowano do druku 1995.02.25

Omówiono metodę komputerowej optymalizacji stosowanej przy projektowaniu tyrystorów ASCR. Przedstawiony numeryczny algorytm wspomagania określa warunki optymanego doboru wymiarów warstw *n*-bazy silnie i słabo domieszkowanych. Artykuł zawiera opis proceu technologicznego do realizacji struktur testowych tyrystorów i wyniki ich badań.

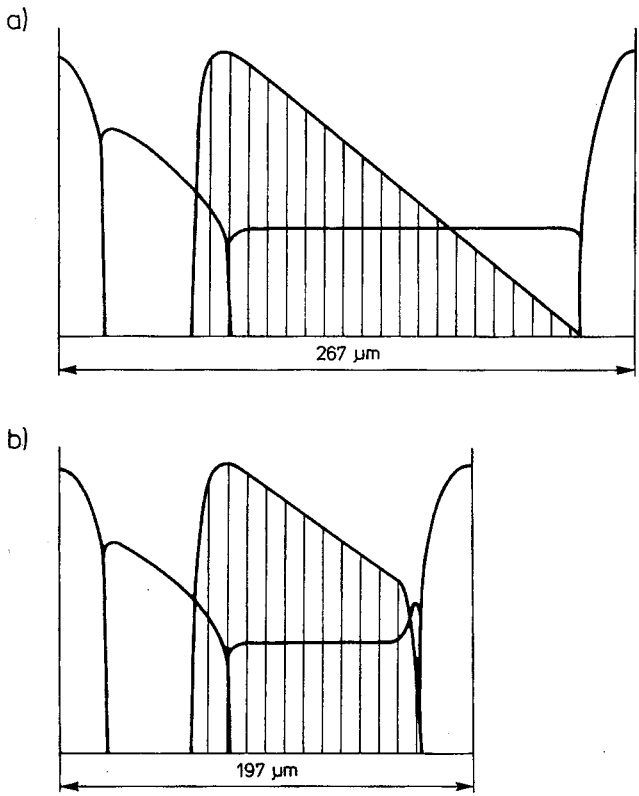
Słowa kluczowe: tyrystory ASCR, domieszkowanie bazy, komputerowe metody projektowania.

1. WSTĘP

Tyrystory asymetryczne (ASCR) należą do rodziny tyrystorów przeznaczonych do zastosowań w układach wysokiej częstotliwości (np. w elektrotermii, zasilaczach impulsowych lub generatorach ultradźwiękowych [1–4]). Tyrystory należące do tej rodziny charakteryzują się stosunkowo niewielkim czasem wyłączania uzyskiwanym przede wszystkim przez zmniejszenie czasu życia w wyniku wprowadzenia do struktury półprzewodnikowej dodatkowych centrów rekombinacyjnych metodą wdyfundowania do niej domieszek metalicznych (Pt, Au) lub poddania struktury wysokoenergetycznemu promieniowaniu (γ, β) generującemu defekty sieci krystalicznej [5–8].

Postępowanie takie pociąga za sobą pogorszenie innych parametrów tyrystora (np. wzrost napięcia przewodzenia i strat przy przełączaniu [4]) i z tego względu może być stosowane jedynie w ograniczonym zakresie. Dalsze zmniejszenie czasu wyłączania uzyskuje się w tych tyrystorach poprzez odpowiednie zmiany konstrukcyjne. Zmiany te mają na celu zmniejszenie grubości struktury półprzewodnikowej (tyrystory RCT i ASCR) lub też stworzenie warunków umożliwiających przyspieszenie procesu wyłączania w wyniku przepływu ujemnego impulsu prądu bramki (tyrystory GATT i GTO).

Istotą konstrukcji tyrystora ASCR jest wprowadzenie dodatkowej silnie domieszkowanej warstwy typu n do obszaru *n*-bazy tyrystora. Pozwala to na znaczne zredukowanie grubości słabo domieszkowanej warstwy *n*-bazy i w konsekwencji na istotne obniżenie napięcia przewodzenia oraz czasu wyłączania tyrystora, przy zachowaniu tego samego napięcia blokowania. Ilustruje to rys. 1, na którym porównano profile domieszkowania i wymiary poszczególnych warstw w tyrystorze normalnym i w tyrystorze ASCR o takiej samej maksymalnej wartości napięcia blokowania.



Rys. 1. Rozkład domieszek w tyrystorze normalnym a) i tyrystorze asymetrycznym b) — linia ograniczająca obszar zakreskowany reprezentuje zmiany natężenia pola elektrycznego odpowiadające napięciu 1200 V

Uzyskanie optymalnej relacji pomiędzy napięciem blokowania a czasem wyłączania i napięciem przewodzenia wymaga odpowiedniego doboru grubości obu warstw n-bazy tyrystora ASCR, tak aby ich łączna grubość była najmniejsza z możliwych do przyjęcia dla danego napięcia blokowania. Optymalizacji takiej nie można przeprowadzić z zadawalającą dokładnością korzystając ze znanych typowych wzorów analitycznych [10] z uwagi na niejednorodne domieszkowanie n-bazy. Dlatego, opracowując założenia konstrukcyjne dla nowego tyrystora ASCR, zdecydowano się na użycie do optymalizacji tego fragmentu tyrystora komputerowych metod wspomagania projektowania (CAD) bazujących na numerycznym modelowaniu własności przyrządów półprzewodnikowych.

W pracy opisano postępowanie przyjęte przy projektowaniu i wykonaniu prototypów tyrystorów ASCR. Ponieważ podstawowym problemem, który należało rozwiązać, był optymalny dobór wymiarów słabo i silnie domieszkowanych warstw n-bazy tyrystora, jest ona głównie poświęcona zaprezentowaniu opracowanego w tym celu oprogramowania wspomagającego prace projektowe. Obok algorytmu numerycznego przedstawiono wyniki symulacji dla kilku z rozważanych konstrukcji tyrystora ASCR. Dla jednej z optymalizowanych konstrukcji przedstawiono kolejne etapy jej powstawania — od przyjęcia założeń konstrukcyjnych poprzez kolejne fazy procesu technologicznego aż do wyników badań partii struktur prototypowych.

2. OPIS MODELU

Rozwiązanie postawionego problemu wymagało stworzenia modelu symulującego stan blokowania w strukturze tyrystora ASCR przedstawionej na rys. 1b. W ogólnym przypadku, model ten powinien być oparty o rozwiązanie pełnego układu równań struktury półprzewodnikowej obejmującego:

— równanie Poissona:

$$\nabla^2\Psi = -\operatorname{div} E = -\frac{q}{\varepsilon}(p - n + N_d - N_a) \tag{1}$$

— równania transportu:

$$J_e = q(n\mu_e E + D_e \operatorname{grad} n) \tag{2}$$

$$J_h = q(p\mu_h E - D_h \operatorname{grad} p) \tag{3}$$

— równania ciagłości prądu:

$$\frac{\partial n}{\partial t} = g - R + \frac{1}{q}\operatorname{div} J_e \tag{4}$$

$$\frac{\partial p}{\partial t} = g - R - \frac{1}{q}\operatorname{div} J_h \tag{5}$$

— równanie prądowe:

$$J = J_e + J_h + \varepsilon \frac{dE}{dt}. \tag{6}$$

Ponieważ przeznaczeniem modelu była jedynie symulacja stanu ustalonego dla polaryzacji zaporowej środkowego złącza J_2 tyrystora, zdecydowano się jednak na opracowanie modelu prostszego, bazującego na numerycznym rozwiązaniu równania Poissona (1).

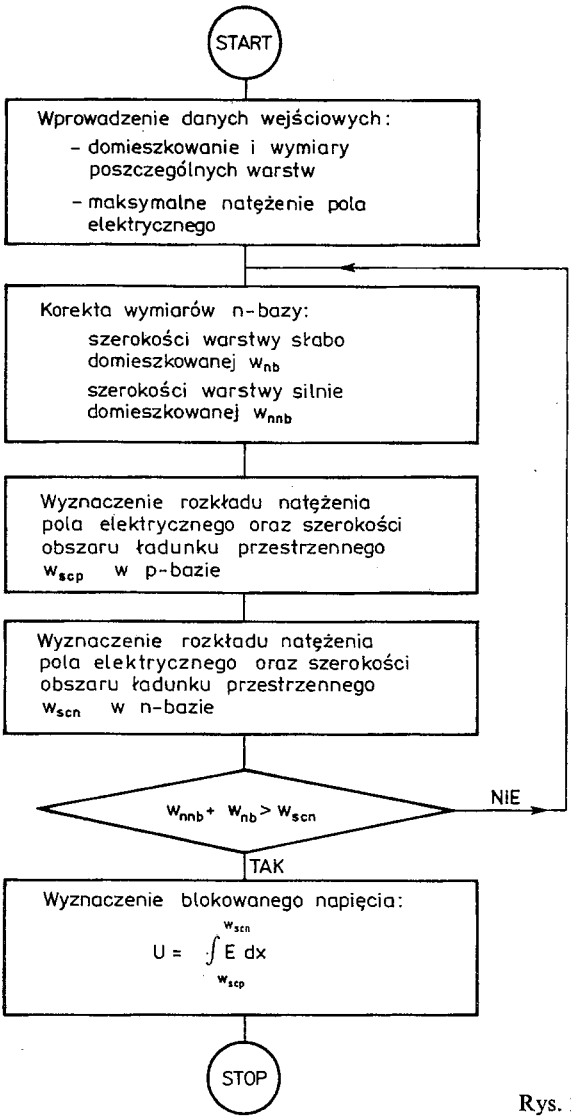
Przyjęto, że będzie to model jednowymiarowy, pozwalający na wyznaczenie:
— napięcia blokowanego przez tyrystor,
— maksymalnego natężenia pola elektrycznego w tyrystorze,
— szerokości obszaru zajętego przez ładunek przestrzenny,
dla zadanych parametrów zewnętrznych obejmujących wymiary warstw przedstawionych na rys. 1 oraz odpowiadające im powierzchniowe koncentracje domieszek. W modelu tym założono ponadto, że w obszarze ładunku przestrzennego gęstości nośników ładunku są pomijalnie małe, a ładunek przestrzenny jest utworzony jedynie przez zjonizowane atomy domieszek. W efekcie w procedurze numerycznej jest ostatecznie rozwiązywane jednowymiarowe równanie Poissona w jego dyskretnej postaci:

$$\frac{E_{i+1} - E_{i-1}}{\Delta x} = \frac{q}{\varepsilon}(N_d - N_a)_{x=x_i}, \tag{7}$$

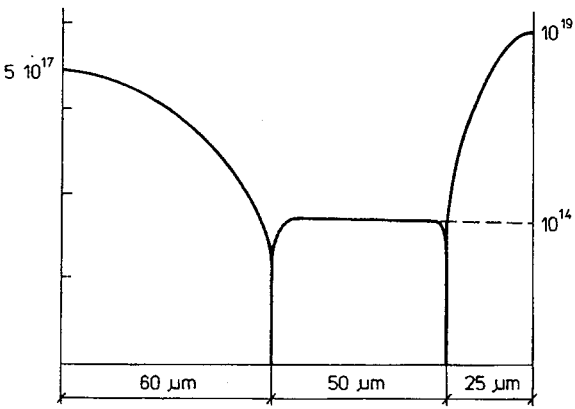
gdzie: Δx — krok dyskretyzacji przestrzeni, a i oraz x_i — dyskretna oraz przestrzenna współrzędna i-tego punktu.

Schemat blokowy procedury jest przedstawiony na rys. 2. Proces obliczeń rozpoczyna się od ustalenia parametrów opisujących strukturę (wymiary, koncentracje powierzchniowe domieszek) oraz maksymalnego natężenia pola elektrycznego, które może wystąpić na środkowym złączu. Następnie, kolejno dla n- i p-bazy, jest rozwiązywane równanie (7), w wyniku czego otrzymuje się rozkład natężenia pola elektrycznego oraz wymiary obszarów ładunku przestrzennego w_{scn} i w_{scp} po obu stronach złącza J_2. Ostatnim etapem obliczeń jest wyznaczenie na drodze całkowania numerycznego napięcia występującego na tyrystorze.

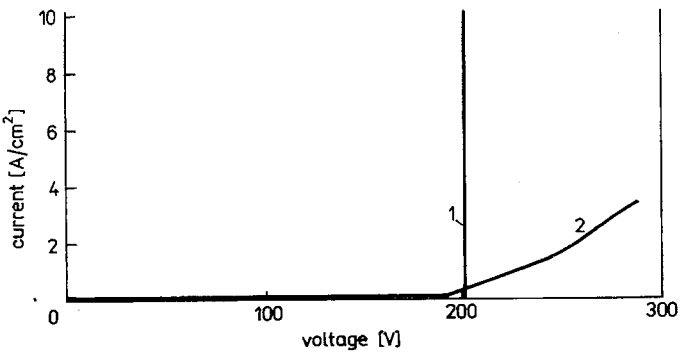
W celu zweryfikowania wpływu poczynionych założeń upraszczających na dokładność modelu przeprowadzono obliczenia testowe dla prostej struktury diaka p-n-p przedstawionej na rys. 3. Obliczenia te miały na celu określenie maksymalnego napięcia blokowania z użyciem opisanej wyżej procedury numerycznej oraz procedury dokładnej utworzonej w oparciu o numeryczne rozwiązanie pełnego układu równań struktury półprzewodnikowej (1) – (6). Zestawione na rys. 4 wyniki obliczeń wykazują, że zaproponowana procedura numeryczna pozwala na oszacowanie maksymalnego napięcia blokowania z zadawalającą dokładnością.



Rys. 2. Schemat blokowy procedury numerycznej



Rys. 3. Struktura testowa diaka *p-n-p*; (brzegowe koncentracje domieszek)



Rys. 4. Napięcia przebicia diaka wyznaczone przy pomocy procedury przybliżonej (1) oraz jego charakterystyka prądowo-napięciowa uzyskana przy pomocy dokładnego modelu numerycznego (2)

3. WYNIKI SYMULACJI KOMPUTEROWYCH

3.1. ZAŁOŻENIA DOTYCZĄCE ZAKRESU OBLICZEŃ

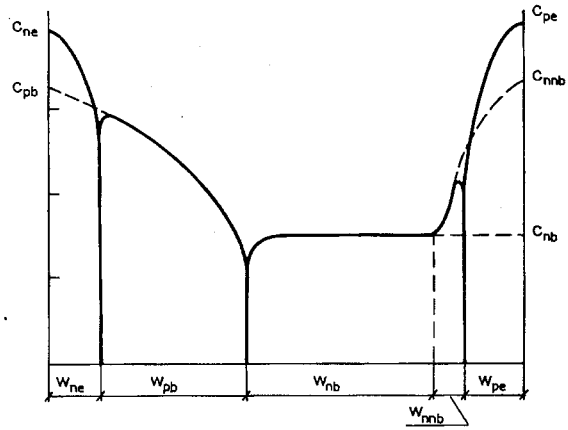
Obliczenia wykonano dla struktury tyrystora ASCR przedstawionej na rys. 5. Wiekszość parametrów opisujących tę strukturę została określona arbitralnie, a ich wartości liczbowe są zebrane w Tabeli 1. Parametrami zmiennymi podczas obliczeń były jedynie parametry opisujące *n*-bazę.

Tabela 1

Zestawienie stałych parametrów opisujących modelowaną strukturę tyrystora ASCR

Nazwa wielkości	Symbol	Wartość
powierzchniowa gęstość domieszkowania katody	C_{ne}	6.10^{19} cm^{-3}
powierzchniowa gęstość domieszkowania *n*-bazy	C_{pb}	5.10^{17} cm^{-3}
powierzchniowa gęstość domieszkowania anody	C_{pe}	1.10^{20} cm^{-3}
grubość warstwy katody	w_{ne}	20 μm
grubość warstwy *p*-bazy	w_{pb}	60 μm
grubość warstwy anody	w_{pe}	20 μm

Prezentowane wyniki dotyczą optymalizacji konstrukcji tyrystorów o napięciu blokowania 1500 i 2800 V. Rozważano wykorzystanie do ich realizacji trzech rodzajów materiału wyjściowego o rezystywności podanych w Tabeli 2. Dla tych



Rys. 5. Profil domieszkowania analizowanej struktury tyrystora ASCR

rezystywności wyznaczono domieszkowania C_{nb} słabo domieszkowanej warstwy n-bazy korzystając z danych doświadczalnych [9], a krytyczne napięcia pola elektrycznego E_{max}, odpowiadające wystąpieniu powielania lawinowego, wyznaczono z napięcia przebicia lawinowego U_B określonego wzorem [10]:

$$U_B = 5.34\ 10^{13} \cdot N^{-0.75}. \tag{8}$$

Powierzchniową gęstość domieszkowania C_{nnb}, odpowiadającą bogato domieszkowanej warstwie n-bazy, przyjęto w obliczeniach jako mieszczącą się w przedziale $7 \cdot 10^{18} - 1.5 \cdot 10^{19}$ cm^{-3}, co odpowiada otrzymywanym w praktyce rozrzutom tej koncentracji.

3.2. WYNIKI OBLICZEŃ

Celem obliczeń było uzyskanie kryterium optymalnego doboru wymiarów silnie i słabo domieszkowanych warstw n-bazy, który zapewni uzyskanie zadanego napięcia blokowania przy minimalnej łącznej grubości n-bazy. Jako takie kryterium wybrano w pracy zależność całkowitej grubości n-bazy $w_{nb} + w_{nnb}$ od grubości warstwy silnie domieszkowanej w_{nnb}, obliczoną dla zadanego napięcia blokowania. Zależność ta posiada wyraźne minimum określone przez punkt zejścia się dwóch linii, wyznaczonych, odpowiednio, przy założeniu, że w tyrystorze:

— maksymalne natężenie pola elektrycznego w n-bazie jest równe podanemu w Tabeli 2 maksymalnemu dopuszczalnemu natężeniu dla danego materiału wyjściowego,

— pole elektryczne w n-bazie wypełnia całą przestrzeń n-bazy, jak to ma miejsce na rys. 1b.

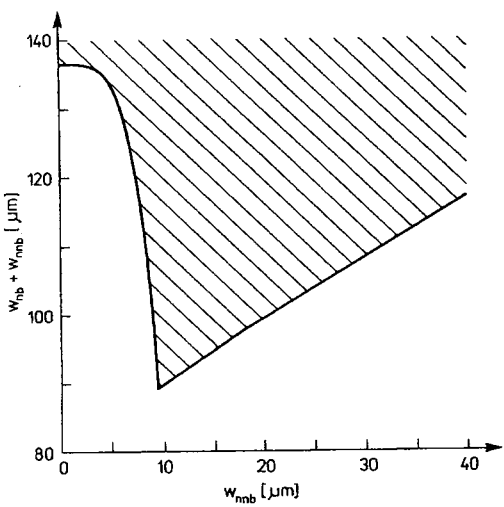
Punkt ten odpowiada poszukiwanej przez nas minimalnej łącznej grubości n-bazy niezbędnej do uzyskania zadanego napięcia blokowania.

Tabela 2

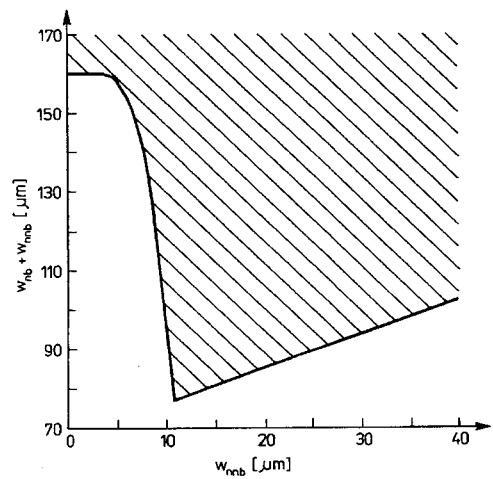
Parametry krzemu przyjętego jako materiał wyjściow

Lp.	Rezystywność ρ	Domieszkowanie C_{nb}	Natężenie pola E_{max}
1	62 Ωcm	8.5 10^{13} cm^{-3}	20 kV/cm
2	90 Ωcm	6.6 10^{13} cm^{-3}	198 kV/cm
3	120 Ωcm	4.4 10^{13} cm^{-3}	198 kV/cm

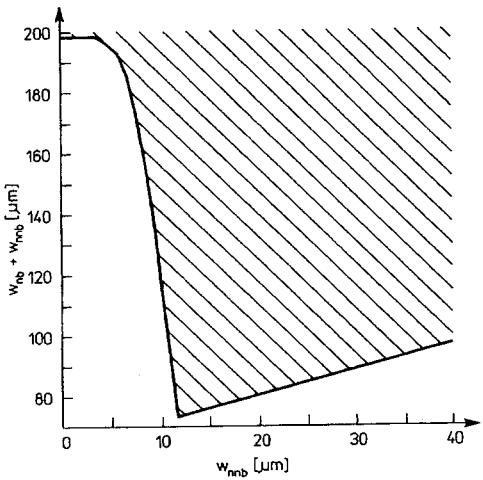
Całość uzyskanych wyników dla przyjętych napięć blokowania, 1500 i 2800 V, jest zebrana na rys. 6–9. Rys. 6–8 przedstawiają kolejno wyniki uzyskane, dla krzemu o rezystywności 50, 90 i 120 Ωcm, przy założeniu, że napięcie blokowania jest równe 1500 V, natomiast rys. 9 zawiera wyniki uzyskane dla krzemu o rezystywności 120 Ωcm, przy założeniu, że napięcie to jest równe 2800 V. Dla napięcia 2800 V nie przedstawiono wyników dla innych wartości rezystywności, ponieważ okazały się za małe dla uzyskania takiego napięcia blokowania. Na każdym z rysunków naniesiono linie odpowiadające dwom granicznym wartościowm powierzchniowej gęstości domieszkowania C_{nnb} oraz zakreskowano obszar odpowiadający napięciom blokowania nie mniejszym od przyjętego jako parametr procesu optymalizacji.



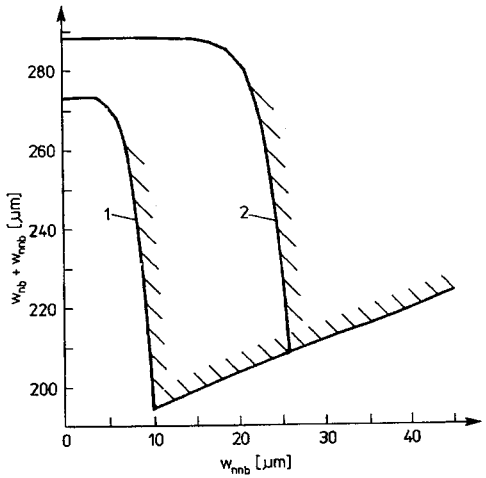
Rys. 6. Zależność grubości całkowitej *n*-bazy $w_{nnb}+w_{nb}$ od grubości warstwy silnie domieszkowanej w_{nnb} dla rezystywności materiału wyjściowego 50 Ωcm i założonym napięciu blokowania 1500 V



Rys. 7. Zależność grubości całkowitej *n*-bazy $w_{nnb}+w_{nb}$ od grubości warstwy silnie domieszkowanej w_{nnb} dla rezystywności materiału wyjściowego 120 Ωcm i założonym napięciu blokowania 1500 V



Rys. 8. Zależność grubości całkowitej *n*-bazy $w_{nnb} + w_{nb}$ od grubości warstwy silnie domieszkowanej w_{nnb} dla rezystywności materiału wyjściowego 120 Ωcm i założonym napięciu blokowania 1500 V



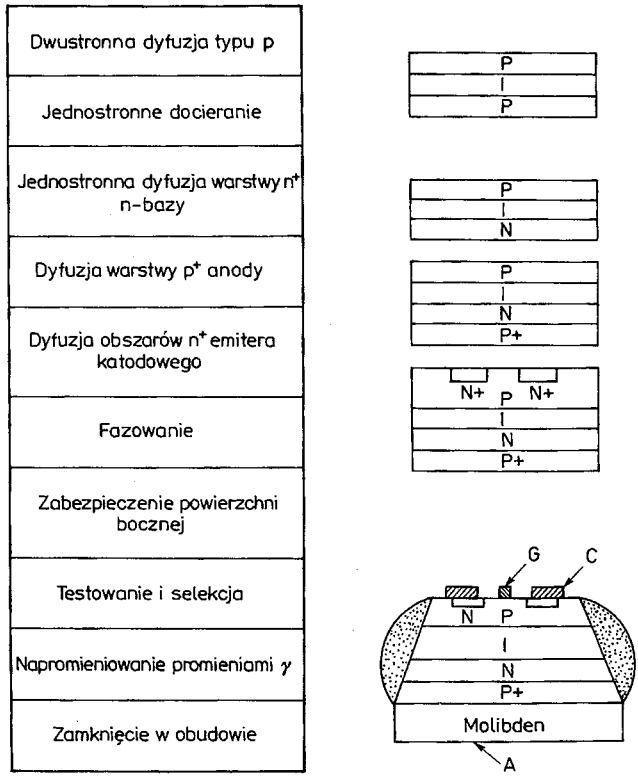
Rys. 9. Zależność grubości całkowitej *n*-bazy $w_{nnb} + w_{nb}$ od grubości warstwy silnie domieszkowanej w_{nnb} dla rezystywności materiału wyjściowego 120 Ωcm i założonym napięciu blokowania 2800 V: 1 — brak zmniejszenia efektywnej grubości warstwy w_{nnb}, 2 — efektywna grubość zmniejszona o 15 μm

4. REALIZACJA I BADANIE STRUKTUR TESTOWYCH

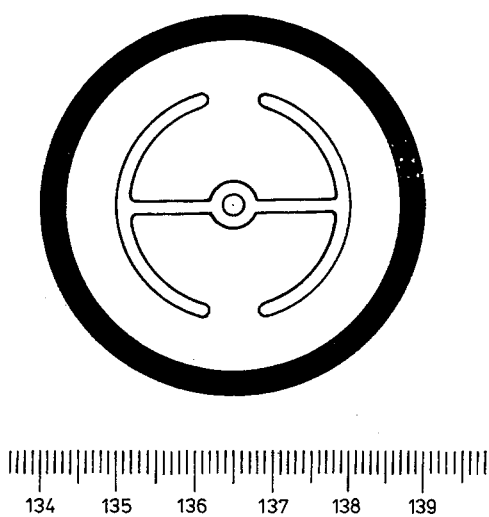
4.1. PRZYJĘTA TECHNOLOGIA REALIZACJI

Modelowe struktury tyrystora ASCR wykonano w technologii dyfuzyjnej, której podstawowe etapy są uwidocznione na schemacie przedstawionym na rys. 10.

Punktem wyjścia jest płytka krzemowa typu *n* o zadanej rezystywności i grubości. Po wstępnych operacjach przygotowawczych, w pierwszym etapie procesu technologicznego są tworzone warstwy typu *p* na drodze dwustronnej dyfuzji, kolejno, aluminium i galu. W efekcie uzyskuje się warstwy typu *p* grubości 30—40 μm przy koncentracji powierzchniowej akceptorów rzędu 10^{17} cm^{-3}. W kolejnych etapach jest usuwana warstwa *p* od strony anody w procesie docierania, a nastepnie płytka jest przygotowywana do jednostronnej dyfuzji silnie domieszkowanej warstwy *n*-bazy. Warstwa ta jest wykonywana na drodze wprowadzenia atomów fosforu z wykorzystaniem jako źródła domieszek $POCl_3$. Teraz, w procesie dyfuzji jednostronnej boru, wprowadza się warstwę p^+ emitera anodowego wykorzystując jako źródło domieszek kwas borowy. W ostatnim etapie jest wykonywana warstwa emitera katodowego. W tym celu, na powierzchnię od strony katody, naklada się w procesie fotolitografii maskę odpowiadającą kształtowi rozwiniętej bramki, a następnie wprowadza się



Rys. 10. Podstawowe etapy realizacji struktury tyrystora ASCR

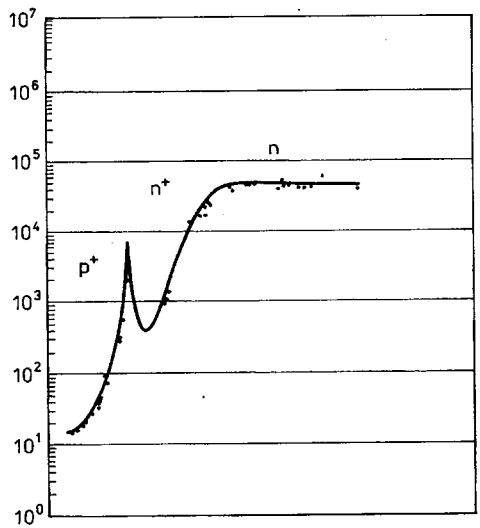


Rys. 11. Wykonana struktura tyrystora ASCR

w obszar odpowiadający emiterowi domieszki fosforu ze źródła $POCl_3$, tak aby otrzymać powierzchniową gęstość donorów 10^{20} cm^{-3} przy grubości warstwy emitera ok. 20 μm.

Końcowymi etapami wytwarzania tyrystora są: przylutowanie molibdenowej płytki kompensującej do powierzchni anodowej otrzymanych struktur, utworzenie na powierzchni katodowej kontaktów katody i bramki w procesie naparowania warstwy aluminium oraz sfazowanie i zabezpieczenie powierzchni bocznych struktury. Uzyskany w efekcie tyrystor ASCR jest pokazany na rys. 11.

Przedstawiona technologia w znacznym stopniu pokrywa się z technologią stosowaną przy wytwarzaniu standardowych tyrystorów mocy. Istotnie nowym elementem są w niej procesy związane z wprowadzeniem nowej, silnie domieszkowanej warstwy *n*-bazy. Podstawowym problemem jaki należało w tym przypadku rozwiązać był taki dobór parametrów tych procesów, a także modyfikacja pozostałych procesów, aby w ostatecznym efekcie nowowprowadzona warstwa miała założone parametry konstrukcyjne, a pozostałe parametry struktury nie uległy niekorzystnym zmianom. Problem ten został rozwiązany pomyślnie, a uzyskany profil domieszkowania fragmentu tyrystora obejmujący słabo domieszkowaną warstwę *n*-bazy, silnie domieszkowaną warstwę *n*-bazy i warstwę p^+ emitera anodowego przedstawia rys. 12.



Rys. 12. Profil domieszkowania fragmentu tyrystora ASCR obejmujący słabo i silnie domieszkowaną warstwę n-bazy oraz warstwę p^+ emitera anodowego

4.2. WYNIKI BADAŃ STRUKTUR TESTOWYCH

Do realizacji struktur o napięciu blokowania 1500 V wybrano jako materiał wyjściowy krzem o rezystywności 50 Ωcm. Przyjęto, że grubość silnie domieszkowanego obszaru *n*-bazy w_{nnb} będzie równa 20 μm, a następnie, korzystając z rys. 6,

ustalono całkowitą grubość n-bazy $w_{nb}+w_{nnb}$ jako równą 115 μm. Pozostałe parametry struktury zostały ustalone zgodnie z Tabelą 1.

Wykonano, zmontowano i przebadano partię próbną liczącą 12 szt. tyrystorów. Uzyskany rozkład napięć blokowania zmierzonych w temperaturze $T_j=25°C$, przedstawia Tabela 3. W większości tyrystorów napięcie to było rzędu 1000 V, a jedynie w jednym osiągnęło ono wartość 1400 V, a więc wartość bliską założonego napięcia

Tabela 3

Rozkład napięć blokowania w partii tyrystorów zaprojektowanych na napięcie 1500 V

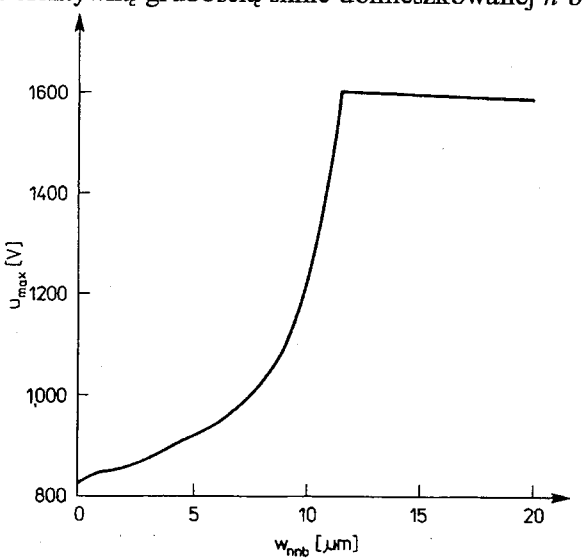
Napięcie	Ilość
1400 V	1
1200 V	3
1000 V	6
800 V	2

Tabela 4

Rozkład napięć blokowania w partii tyrystorów zaprojektowanych na napięcie 2800 V

Napięcie	Ilość
$\geqslant$2800 V	8
2600 V	5
2400 V	1
2200 V	3
2000 V	1
$\leqslant$1800 V	5

1500 V. Badania mikroskopowe wykazały, że przyczyną tak drastycznego spadku napięcia blokowania mogły być *„aluminiowe języki"* powstające podczas procesu lutowania. Wnikają one przez p-emiter do silnie domieszkowanej n-bazy redukując jej efektywną grubość i zmniejszając tym samym napięcie objętościowego przebicia n-bazy. Aby móc ocenić ten efekt przeprowadzono dla tej struktury tyrystorowej dodatkowe symulacje komputerowe z wykorzystaniem opracowanego modelu tyrystora. Ich efektem jest pokazana na rys. 13 zależność pomiędzy maksymalnym napięciem blokowania U_{max} a efektywną grubością silnie domieszkowanej n-bazy w_{nnb} przy stałej



Rys. 13. Zależność napięcia blokowania U_{max} od efektywnej grubości silnie domieszkowanej n-bazy w_{nnb}

wartości $w_{nnb}+w_{pe}$. Wykazuje ona, że uzyskane wartości napięć blokowania odpowiadają zmniejszeniu efektywnej grubości warstwy w_{nnb} około 10 – 15 μm.

Powyższe spostrzeżenia wykorzystano ustalając wymiary odpowiednich warstw n-bazy w partii tyrystorów o przewidywanym napięciu blokowania 2800 V. Jako materiał wyjściowy przyjęto dla nich krzem o rezystywności 120 Ωcm, a wymiary warstw n-bazy dobrano korzystając z krzywej 2 na rys. 9, która odpowiada zmniejszeniu efektywnej grubości warstwy w_{nnb} o 15 μm. Z obszaru dopuszczalnych wymiarów wybrano punkt, który odpowiada grubości silnie domieszkowanego obszaru n-bazy $w_{nnb}=40$ μm oraz całkowitej grubości n-bazy $w_{nb}+w_{nnb}$ równej 240 μm. Pozostałe parametry struktury przyjęto zgodnie z Tabelą 1.

Zgodnie z powyższymi założeniami wykonano, zmontowano i przebadano partię próbną liczącą 23 szt. Uzyskany rozkład napięć blokowania zmierzonych dla tej partii tyrystorów w temperaturze $T_j=25°C$, przedstawia Tabela 4. Założone lub większe napięcie blokowania uzyskano w 8 tyrystorach, co stanowi ponad 1/3 partii, a napięcie w przedziale 2800 – 2400 już w ponad 2/3 całej partii. Tak więc można stwierdzic, że cel optymalizacji został osiągnięty.

PODSUMOWANIE

W pracy przedstawiono numeryczny algorytm wspomagania projektowania tyrystorów ASCR, którego zadaniem jest określenie warunków optymalnego doboru wymiarów silnie i słabo domieszkowanych warstw n-bazy, tak aby zadane napięcie blokowania mogło być osiągnięte przy minimalnej łącznej grubości n-bazy. Jako kryterium optymalizacji wybrano zależność całkowitej grubości n-bazy od grubości warstwy silnie domieszkowanej, obliczoną dla zadanego napięcia blokowania. Posiada ona wyraźne minimum odpowiadające poszukiwanej, minimalnej łącznej grubości n-bazy, a linie ją tworzące wyznaczają obszar dopuszczalnych zmian grubości obu warstw n-bazy, dla zadanego napięcia blokowania.

W rozdziale 2 przedstawiono numeryczny model tyrystora ASCR, wykorzystany następnie w rozdziale 3 do wyznaczania zależności w_{nnb} od łącznej grubości n-bazy $w_{nnb}+w_{nb}$. Wyniki takich symulacji, przeprowadzonych dla tyrystorów o napięciu blokowania, odpowiednio 1500 V i 2800 V, wykorzystano do przygotowania założeń konstrukcyjnych struktur testowych takich tyrystorów.

Rozdział 4 zawiera opis procesu technologicznego wykorzystanego do realizacji struktur testowych oraz wyniki ich badań. W serii próbnej tyrystorów o założonym napięciu blokowania 1500 V stwierdzono znaczne zaniżenie faktycznie uzyskanych wartości tego napięcia. Ponieważ zarówno obserwacje mikroskopowe jak i przeprowadzone symulacje komputerowe wskazały jako jedną z możliwych przyczyn tego zjawiska zmniejszenie efektywnej grubości warstwy w_{nnb}, następną partię tyrystorów wykonano uwzględniając taką możliwość. Była to partia tyrystorów o założonym napięciu blokowania 2800 V, a uzyskane wyniki pomiarów napięć blokowania potwierdziły słuszność przyjętego postępowania.

BIBLIOGRAFIA

1. W. Bösterling, M. Frzöhlich: *Thyristorarten ASCR, RLT und GTO Technik und Grenzen ihrer Anwendung.* ETZ, 1983, vol. 104, nr 24, ss. 1246–1251
2. J. Vitins, P. Wetzel: *Rückwärrsletende Thyristoren für Leistungelektronik.* 1981, vol. 63, nr 2, ss. 74–83
3. Z.F. Chang: *Application of ASCR in 4o-kHz sine-wave converter.* Int. Telecomm. Energy Conf. INTELEC, November 1979, ss. 100–107
4. Z. Lisik: *Tyrystory do układów wielkiej częstotliwości — rozwój produkcji w krajach RWPG.* Elektronika, 1986, vol. 27, nr 1–2, ss. 29–34
5. A. Tada, T. Nakagawa, H. Hagino: *Improvement in turn-off time and other electrical characteristics of fast-switching thyristors.* Jap. J. Appl. Phys. 1982, vol. 21, nr 4, pp. 617–623
6. B.J. Baliga, S. Krishna: *Optimization of recombination levels and their capture cross section in power rectifiers and thyristors.* Solid-St. Electron. 1977, vol. 20, pp. 225–232
7. R.O. Carlson, Y.S. Sun, H.B. Assalit: *Lifetime control in silicon power devices by electron or gamma irradiation.* IEEE Trans. Electron Dev., 1977, vol. ED-24, pp. 1103–1108
8. V. Benda, P. Spur, B. Kojecky: *Electron irradiation influence on large-area GTO homogeneity.* ISPS'92, Praga 9–11 IX 1992, ss. 126–132
9. N. Sclar: *Resistivity and deep impurity levels in silicon at 300 K.* IEEE Trans. Electron Devices, 1977, vol. ED-24, nr 6, pp. 709–712
10. S.M. Sze: *Physics of semiconductor devices.* J. Willey, 1981

Z. LISIK, W. CIEPLAK, M. LEŚKIEWICZ-KRAWCZYK, M. SMOLUCHOWSKI

COMPUTER OPTIMISATION OF THICKNESS OF LOW-DOPED N-BASE LAYERS IN ASCR THYRISTOR

Summary

In the paper, the CAD method applied for optimisation of the design of an ASCR thyristor is presented. The worked out algorithm allows determination of optimal dimensions of low- and high-doped n-base layers of the thyristor for an assumed blocking voltage. Both the manufacturing process of the designed ASCR test structures and the results of their examinations are described ae well.

Key words: thyristors, computer optimisation.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 1. ss. 103–112

Uproszczona metoda określenia czułości heterodynowego odbioru sygnału optycznego o znacznej szerokości linii widmowej

JERZY SIUZDAK

Instytut Telekomunikacji, Politechnika Warszawska

Otrzymano 1995.10.24

Autoryzowano do druku 1994.12.16

Zaprezentowano przybliżoną metodę określenia czułości heterodynowego odbioru sygnału optycznego o znacznej szerokości widmowej za pomocą parametru będącego uogólnieniem stosunku mocy sygnału do mocy szumu. Metoda została zastosowana do analizy koherentnego demodulatora binarnej modulacji częstotliwości z jednym filtrem. Porównano czułość takiego demodulatora z czułością odbiornika o detekcji bezpośredniej (niekoherentnej).

Słowa kluczowe: sygnały optyczne, demodulator koherentny, rozkład gaussowski szumów.

1. WSTĘP

Teoretyczne określenie parametrów czułości koherentnego odbioru heterodynowego binarnego sygnału optycznego o szerokości linii widmowej porównywalnej z szybkością transmisji napotyka na znaczne trudności. Znalezienie potrzebnych dla wyznaczenia elementowej stopy błędów rozkładów prawdopodobieństw procesu stochastycznego na wyjściu demodulatora, kiedy transmitowane są symbole „1" i „0" jest bardzo trudne. Przyczyną jest złożona struktura odbiornika z nieliniową obróbką, stosowanie filtracji przed- i podetekcyjnej oraz skomplikowana postać sygnału użytecznego, który wskutek obecności szumu fazowego ma charakter niedeterministyczny. Powstające problemy próbowano rozwiązać różnymi metodami: metodą rozkładu na funkcje własne [1], rozwiązaniem równania Fokkera-Plancka dla procesów Markowa [2], zastosowaniem pochodnej Radona–Nikodyma [3] i innymi. Są to jednak metody złożone, wymagające skomplikowanych obliczeń numerycznych.

Dla stosunkowo niewielkich szerokości linii widmowych laserów przybliżona analiza była prowadzona w pracach [8, 10], nie obejmują one jednak interesującego nas przypadku, kiedy linie widmowe są szerokie w stosunku do szybkości transmisji.

Poniżej zostanie zaprezentowana metoda przybliżona oparta na koncepcji określenia czułości odbiornika poprzez parametr będący uogólnieniem dobrze znanego pojęcia stosunku mocy sygnału do mocy szumu (ang. SNR — signal to noise ratio). Metoda zaprezentowana zostanie na przykładzie koherentnego odbiornika z jednym filtrem przeznaczonego do demodulacji sygnału FSK.

2. PODSTAWA METODY

Rozpatrzmy system transmisji binarnej, w którym wysyłane sa dwa symbole „0" i „1". Klasyczne pojęcie stosunku mocy sygnału do mocy szumu dobrze określa parametry odbiornika w przypadku jeśli obróbka sygnału przed układem decyzyjnym jest liniowa. Z grubsza rzecz biorąc w takim układzie na wejściu układu decyzyjnego obydwu symbolom nadawanym odpowiadają określone poziomy sygnału użytecznego, natomiast moc i rozkład prawdopodobieństwa szumów są dla obydwu nadawanych symboli takie same. Obydwie wielkości, tzn. moc sygnału i moc szumu pozwalają wówczas wyznaczyć takie parametry odbiornika jak np. elementowa stopa błędów (ang. BER — bit error rate), oczywiście przy znanym, najczęściej gaussowskim, rozkładzie prawdopodobieństwa szumów.

Inna jest sytuacja w przypadku odbiornika z nieliniową obróbką sygnału przed układem decyzyjnym. Wówczas symbolom „0" i „1" mogą odpowiadać nie tylko różne moce szumów w odbiorniku, ale nawet inne ich rozkłady prawdopodobieństwa. Ponieważ na charakterystyki statystyczne układu decyzyjnego mają wpływ moce szumów w przypadku nadawania obydwu symboli, więc klasyczne pojęcie stosunku mocy sygnału do mocy szumu traci sens. Możliwe są różne uogólnienia tego pojęcia na układy nieliniowe. Dla przykładu w pracach [11, 12] zaproponowano użycie parametru

$$\gamma = \frac{m(1) - m(0)}{\sigma(1) + \sigma(0)}, \tag{1}$$

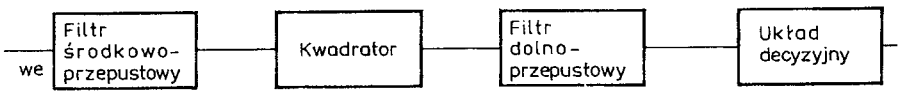
gdzie $m(\cdot)$ i $\sigma(\cdot)$ są odpowiednio średnim poziomem i odchyleniem standardowym sygnału przy nadawaniu odpowiedniego symbolu (0 lub 1).

W niniejszej pracy proponujemy zbliżony parametr w postaci

$$\gamma^2 = \frac{[m(1) - m(0)]^2}{\sigma^2(1) + \sigma^2(0)}. \tag{2}$$

W odróżnieniu od poprzedniej definicji tutaj dodają się moce szumów, a nie ich odchylenia standardowe, co zdaniem autora lepiej określa odpowiednik mocy szumów.

W dalszej części zajmiemy się wyznaczeniem czułości demodulatora sygnału o binarnej modulacji częstotliwości z jednym filtrem [8]. Schemat blokowy tego demodulatora przedstawiony jest na rys. 1. Przeznaczony jest on do odbioru sygnałów o dużej dewiacji częstotliwości, charakteryzujących się widmem skupionym w pobliżu dwóch częstotliwości. Filtr w odbiorniku jest dopasowany do jednej z tych dwóch częstotliwości. Czułość tego demodulatora jest równa czułości odbiornika ASK. Jego zaletą jest prostota i podobna do odbiornika z dwoma filtrami tolerancja na szum fazowy laserów, jak i na nieliniowość charakterystyki FM lasera nadawczego. Sumaryczna szerokość linii widmowej laserów może wynosić 5% szybkości transmisji przy stracie czułości równej 1 dB [7].



Rys. 1. Schemat blokowy odbiornika sygnału FM z jednym filtrem

Praktyczna możliwość tolerancji dużej szerokości widmowej laserów przez układ demodulatora z jednym filtrem jest znaczna. Konieczne jest jedynie zapewnienie dostatecznie dużej dewiacji częstotliwości tak, aby widma odpowiadające poszczególnym częstotliwościom się nie pokrywały, oraz rozszerzenie pasma każdego z filtrów tak, aby obejmował on całe widmo sygnału modulowanego odpowiadającego danej częstotliwości. Widmo to jest bowiem splotem widma niemodulowanego lasera (laserów) (9) i widma modulacji danego wzorem (4):

$$S_c(f) = S(f) * \Phi(f), \tag{3}$$

zaś widmo modulacji wyraża się wzorem [7]

$$\Phi(f) = A_0 \left[\frac{\sin \pi f T}{\pi f T} \right]^2 (1 + \cos 2\pi f T), \tag{4}$$

gdzie $1/T$ jest szybkością transmisji, zaś przy wyprowadzeniu wzoru założono brak pokrywania się widm odpowiadających dwóm częstotliwościom modulacji.

W dalszej części wyznaczona zostanie czułość omawianego demodulatora, a następnie dokonane zostanie porównanie tej czułości z czułością odbiornika o detekcji bezpośredniej.

3. MODEL MATEMATYCZNY LINII WIDMOWEJ LASERA

Wskutek fluktuacji gęstości nośników w rezonatorze i innych zewnętrznych wpływów, emitowana z lasera fala jest wąskopasmowym procesem przypadkowym. Zaniedbując fluktuacje amplitudy, pole elektryczne promieniowania laserowego można zapisać wzorem [5]:

$$A(t)=A_0 e^{j[\omega_0 t+\Theta(t)]}. \tag{5}$$

We wzorze tym ω_0 jest środkową pulsacja lasera, A_0 — amplitudą promieniowania, zaś $\theta(t)$ opisuje proces fluktuacji fazy fali. Dla znalezienia kształtu widma powyższego procesu należy obliczyć jego funkcję autokorelacji $C(t)$:

$$C(t)=E[A(t_0+t)A^*(t_0)]=A_0^2 e^{j\omega_0 t}E[e^{j[\Theta(t_0+t)-\Theta(t_0)]}]=A_0^2 e^{j\omega_0 t}\exp[-0.5E(\Delta\Theta^2(t))]. \tag{6}$$

W ostatnim wzorze symbol $E(\cdot)$ jest operatorem wartości oczekiwanej, zaś przy wyprowadzeniu przyjęto, że zmienna losowa

$$\Delta\Theta(t)=\Theta(t_0+t)-\Theta(t_0) \tag{7}$$

ma rozkład gaussowski. Gdy laser pracuje znacznie powyżej prądu progowego, wówczas [5, 6]

$$E[\Delta\Theta^2(t)]=2\pi B_L t. \tag{8}$$

Tutaj B_L jest szerokością widmową linii lasera (FWHM — pełna szerokość w połowie maksimum) w przypadku obserwacji jednej linii, bądź sumą szerokości linii widmowych lasera nadawczego i heterodyny w przypadku odbioru heterodynowego. Rozkład widmowy gęstości mocy wyznacza się jako transformatę Fouriera funkcji autokorelacji C(t) z uwzględnieniem powyższego wzoru. Ma on postać tzw. rozkładu Lorentza [5, 6] i jest dany przez

$$S(f)=\frac{2P_0}{\pi B_L}\frac{1}{1+\left(\dfrac{f-f_0}{B_L/2}\right)^2}. \tag{9}$$

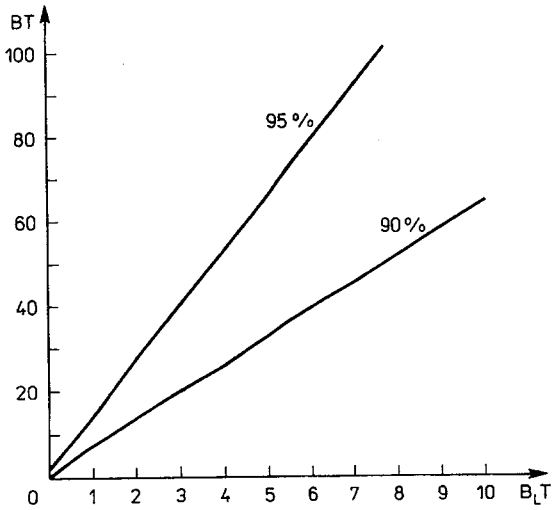
W powyższym wzorze P_0 jest mocą emitowaną przez laser (lasery).

4. WYBÓR PASMA PRZEPUSTOWEGO FILTRU

Filtr występujący w układzie demodulatora z jednym filtrem powinien mieć pasmo przenoszące większość mocy sygnału wejściowego (np. 90% lub 95%). Moc sygnału zawarta w paśmie (-B/2, B/2) jest równa [7]

$$P=\int_{-B/2}^{B/2} S_c(f)\,df=\frac{1}{\pi}\int_{-\infty}^{\infty}\Phi(f)\left[\operatorname{arctg}\frac{BT/2-f}{e}+\operatorname{arctg}\frac{BT/2+f}{e}\right]df. \tag{10}$$

Tutaj $\phi(f)$ jest widmem sygnału o modulacji FM danym przez (4), B jest szerokością pasma przepustowego filtru prostokątnego, zaś $e=B_L T/2$. Szerokość pasma filtru, który przenosi większość energii sygnału (90%, 95%) potrzebnej do niezniekształconego odbioru przedstawiona jest na rys. 2 w zależności od szerokości widma laserów. Wyniki te są zgodne z optymalnymi pasmami filtrów IF zamieszczonymi w pracy [9].



Rys. 2. Zależność pasma przepustowego filtru IF od szerokości widma laserów

Wybór nierealizowalnego filtru prostokątnego jest podyktowany stosunkowo łatwymi obliczeniami. Wyniki obliczeń są o tyle istotne, że własności energetyczne filtru zależą głownie od jego pasma przepustowego, zaś w mniejszym stopniu od kształtu transmitancji. Z obliczeń wynika, że przy dużych wartościach B_LT pasmo filtru przenoszącego 90% mocy sygnału jest około 6.3 krotnie większe od sumy szerokości widmowych laserów (FWHM) B_L, zaś pasmo filtru przenoszącego 95% mocy około 12.7 krotnie większe. Warto tutaj zauważyć, że w pasmie przepustowym o szerokości równej szerokości linii widmowej B_L zawiera się jedynie połowa mocy sygnału.

5. CZUŁOŚĆ ODBIORNIKA HETERODYNOWEGO

Założymy, że filtr przeddetekcyjny jest środkowoprzepustowym filtrem prostokątnym o paśmie B tak dobranym, żeby przepuszczać praktycznie cały sygnał bez zniekształceń. Badania wykazały, że jest to pasmo przenoszące około 95% mocy sygnału, przy czym jego szerokość jest zbliżona do szerokości pasma filtru optymalnego zamieszczonego w pracy [9]. Wybór szerokości pasma wydaje się nieco sztuczny, lecz takie podejście jest często stosowane [14], a jego poprawność wynika bezpośredno z równości Parsevala. Szerokość pasma B w zależności od wielkości B_LT przedstawiono na rys. 2, przy czym dla B_LT zbliżonych do 0 — $B=2/T$, a dla $B_LT \gg 1 - B = 12.7B_L$.

Zakładając, że sygnał użyteczny przechodzi przez filtr predetekcyjny bez zniekształceń, a jedynie ze zmniejszoną mocą, można sygnał na wyjściu tego filtru przedstawić w postaci

$$s(t)=2\mathrm{R}b\sqrt{z\mathrm{P}_0\mathrm{P}_\mathrm{L}}\sin[2\pi f_d t+\Theta(t)]+n(t). \tag{11}$$

R [A/W] jest tutaj czułością fotoodbiornika, b (0 lub 1) — bitem nadawanym, P_0 — mocą sygnału odbieranego, P_L — mocą heterodyny optycznej, f_d — częstotliwością różnicową, $\theta(t)$ — szumem fazowym, zaś $n(t)$ — szumem addytywnym. Jeżeli, tak jak powinno być, moc heterodyny optycznej jest dostatecznie duża, szum $n(t)$ jest gaussowskim szumem śrutowym o gęstości widmowej

$$\mathrm{N}_0=2q\mathrm{R}\mathrm{P}_\mathrm{L}. \tag{12}$$

gdzie q jest ładunkiem elektronu. Moc szumów i moc sygnału na wyjściu filtru przeddetekcyjnego są równe

$$\mathrm{E}(n^2)=\mathrm{B}\mathrm{N}_0;\qquad \mathrm{E}(s^2)=2bz\mathrm{R}^2\mathrm{P}_0\mathrm{P}_\mathrm{L}, \tag{13}$$

gdzie E(·) oznacza operator wartości oczekiwanej.

Całkowity sygnał na wyjściu układu podnoszącego do kwadratu po odfiltrowaniu składowych o podwójnej częstotliwości jest dany przez

$$\mathrm{S}_1=2bz\mathrm{R}^2\mathrm{P}_0\mathrm{P}_\mathrm{L}+n^2(t)+4bn(t)\mathrm{R}\sqrt{z\mathrm{P}_0\mathrm{P}_\mathrm{L}}\sin[2\pi f_d+\Theta(t)]. \tag{14}$$

Dla uproszczenia obliczeń filtr podetekcyjny zastępujemy dyskretnym integratorem sumującym w czasie T próbki pobierane z odstępem czasowym $t_d=1/\mathrm{B}$. Z własności funkcji autokorelacji procesu przypadkowego $n(t)$ o widmie prostokątnym wynika, że są one nieskorelowane, a ponieważ są również gaussowskie więc i niezależne. Po rozłożeniu szumu $n(t)$ na składowe dolnopasmowe zgodnie z zależnością $n(t)=n_s(t)\sin(\cdot)+n_c(t)\cos(\cdot)$ oraz po odfiltrowaniu składowych w.cz. i skorzystaniu z niezależności składowych dolnopasmowych szumu, otrzymujemy zależność na wartość sygnału na wyjściu integratora

$$\mathrm{S}_2=\frac{1}{\mathrm{L}}\sum_{i=1}^{\mathrm{L}}\left(2bz\mathrm{R}^2\mathrm{P}_0\mathrm{P}_\mathrm{L}+\frac{n_s^2(it_d)+n_c^2(it_d)}{2}+2bn_s(it_d)\mathrm{R}\sqrt{z\mathrm{P}_0\mathrm{P}_\mathrm{L}}\right), \tag{15}$$

gdzie $\mathrm{L}=\mathrm{T}/t_d=\mathrm{BT}$. Korzystając z niezależności próbek szumu pobranych w różnych momentach czasowych nietrudno jest policzyć odpowiednie momenty sygnału na wyjściu filtru podetekcyjnego

$$m(1)=2\mathrm{R}^2z\mathrm{P}_0\mathrm{P}_\mathrm{L}+\mathrm{E}(n^2),\qquad m(0)=\mathrm{E}(n^2), \tag{16}$$

$$\sigma^2(1)=\frac{4}{\mathrm{L}}\mathrm{E}(n^2)z\mathrm{R}^2\mathrm{P}_0\mathrm{P}_\mathrm{L}+\frac{1}{2\mathrm{L}}\mathrm{E}(n^2)^2,\qquad \sigma^2(0)=\frac{1}{2\mathrm{L}}\mathrm{E}(n^2)^2. \tag{17}$$

Korzystając z definicji wielkości γ^2, określenia wartości $\mathrm{E}(n^2)$ i powyższych wzorów, nietrudno jest wyliczyć wartość γ^2:

$$\gamma^2=\frac{z\mathrm{R}\mathrm{P}_0\mathrm{T}}{2q\left(1+\dfrac{1}{2\mathrm{SNR}_w}\right)}, \tag{18}$$

gdzie SNR_w jest stosunkiem mocy sygnału do mocy szumu na wyjściu filtru predekcyjnego. Porównajmy teraz otrzymaną powyżej wielkość z odpowiednią wartością SNR dla układu o detekcji bezpośredniej z fotodiodą lawinową [13]. W takim układzie jest również możliwe zastosowanie modulacji FM poprzez modulację prądu lasera nadawczego o niewielkiej głębokości i zamianę modulacji FM na AM w odbiorniku przez zastosowanie filtru optycznego przed fotodetektorem. W odbiorniku istotną rolę zaczynają wtedy odgrywać szumy termiczne. Po zaniedbaniu prądów ciemnych i tła, można na podstawie [4, 13] z łatwością obliczyć parametr γ^2. Wynosi on w tym przypadku

$$\gamma_0^2=\frac{R^2P_0^2M^2}{2qBRP_0M^{2+x}+\frac{8kTBF}{r}}. \tag{19}$$

W powyższym wzorze oprócz opisanych poprzednio oznaczeń wprowadzono następujące symbole: M — współczynnik powielenia lawinowego, x — współczynnik określający szumy nadmiarowe powielenia, k — stała Boltzmanna, T — temperatura bezwzględna, F — współczynnik szumów wzmacniacza, r — rezystancja obciążenia diody, B — pasmo przepustowe wzmacniacza. W naszym przypadku 95% pasmo wynosi $B=1/T$ (ze względu na pracę w pasmie podstawowym i brak wpływu szumu fazowego lasera). Należy jeszcze raz zaznaczyć, że powyższy współczynnik nie jest bezpośrednią definicją stosunku mocy sygnału do szumu, a jedynie jej uogólnieniem w przypadku różnych rozkładów prawdopodobieństw przy transmisji symbolu 0 i 1.

Istnieje optymalna wartość współczynnika powielenia M maksymalizująca wielkość γ_0^2. Łatwo ją otrzymać licząc odpowiednie pochodne

$$M_{opt}=\left(\frac{8kTF}{xqRP_0r}\right)^{\frac{1}{2+x}}. \tag{20}$$

Wstawiając tę wartość do zależności na γ_0^2 otrzymujemy ostatecznie

$$\gamma_{opt}^2=\frac{RP_0}{qB}\frac{1}{2+x}\left(\frac{4}{x}\right)^{-\frac{x}{2+x}}\beta^{-\frac{x}{2+x}}. \tag{21}$$

W powyższym wzorze wielkość β określa stosunek mocy szumu termicznego do mocy szumu śrutowego (ale bez powielenia lawinowego) w odbiorniku.

Porównajmy teraz wielkości γ^2 w przypadku detekcji heterodynowej i detekcji bezpośredniej. Mamy w tym przypadku

$$\kappa=\frac{\gamma^2}{\gamma_0^2}=\frac{z}{1+\frac{1}{2SNR_w}}\frac{2+x}{2}\left(\frac{4}{x}\right)^{\frac{x}{2+x}}\beta^{\frac{x}{2+x}}. \tag{22}$$

Parametr kappa (22) określa nam ile razy uogólniony stosunek mocy sygnału do mocy szumu jest większy w przypadku detekcji heterodynowej. Parametr ten zależy jedynie od 3 czynników: SNR_w — określającego stosunek sygnału do szumu na

wyjściu filtru predetekcyjnego przy detekcji heterodynowej, x — określającego własności szumowe fotodiody lawinowej oraz β — określającego stosunek mocy szumów termicznych do szumów śrutowych (bez powielenia lawinowego) w odbiorniku. Zaznaczmy, że w praktycznych układach wejściowych z detekcja bezpośrednią moc szumu termicznego zazwyczaj wielokrotnie przewyższa moc szumów śrutowych (bez powielenia lawinowego).

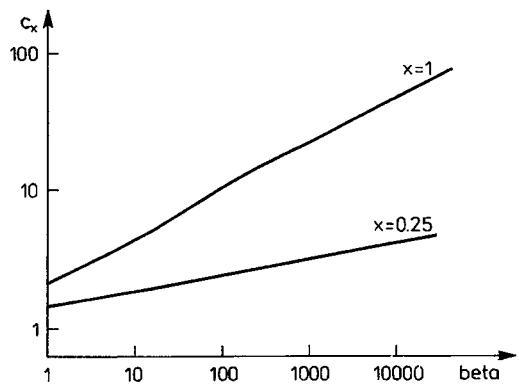
Rozpatrzmy dla przykładu dwa przypadki fotodiod lawinowych: krzemowej używanej w paśmie 850 nm i germanowej używanej dla dłuższych fal. Typowe wartości parametru x wynoszą [4, 13] x=0.25 (krzemowa) i x=1 (germanowa). Dla fotodiody krzemowej zależność (46) upraszcza się do

$$\kappa=\frac{1.46}{1+\dfrac{1}{2\mathrm{SNR}_w}}\beta^{\frac{1}{9}}=\frac{c_{0.25}}{1+0.5\mathrm{SNR}_w^{-1}}, \tag{23}$$

zaś dla fotodiody germanowej do

$$\kappa=\frac{2.26}{1+\dfrac{1}{2\mathrm{SNR}_w}}\beta^{\frac{1}{3}}=\frac{c_1}{1+0.5\mathrm{SNR}_w^{-1}}, \tag{24}$$

Wartości współczynnika c_x dla różnych wielkości β przedstawiono na rys. 3.



Rys. 3. Zależność współczynnika c_x od stosunku mocy szumów termicznych do szumów śrutowych w odbiorniku o detekcji bezpośredniej

WNIOSKI

Zaprezentowano metodę określania czułości odbioru heterodynowego i zastosowano ją w przypadku demodulatora binarnego sygnału FM z jednym filtrem.

Z przeprowadzonej analizy wynika, że o przydatności danego sposobu detekcji decydują głównie dwa czynniki:

— dioda lawinowa jaką dysponujemy (parametr x),

— stosunek szumów termicznych do szumów śrutowych (bez powielenia lawinowego) w odbiorniku (parametr β).

Przy dobrych lawinowych diodach krzemowych (zakres 850 nm) korzyści ze stosowania detekcji heterodynowej są niewielkie i nie usprawiedliwiają znacznej komplikacji i kosztów układu.

Przy lawinowych diodach germanowych (zakres 1300 i 1550 nm) korzyści ze stosowania detekcji heterodynowej stają się znaczne, zwłaszcza jeśli stosunek mocy szumów termicznych do mocy szumów śrutowych jest duży. W tym przypadku istotny staje się również wpływ prądów ciemnych. Warunkiem stosowania detekcji heterodynowej jest utrzymanie odpowiedniej wielkości SNR_w.

BIBLIOGRAFIA

1. J. Siuzdak, W. van Etten: *Heterodyne ASK Multiport optical receivers using postdetection filtering*. J. Lightwave Technol., 1990, vol. 8, no. 1, pp. 71–77
2. G. Jacobsen, I. Garrett: *Impact of phase noise on coherent systems — a Fokker–Planck approach*. Proc. Fourth Tirrenia International Workshop on Digital Communications, 1989, Elsevier, Amsterdam 1990, pp. 133–148
3. G. Vanucci, L.J. Greenstein, G.J. Foschini: *The detection and analysis of coherent lightwave signals corrupted by laser phase noise*. Proc. Fourth Tirrenia International Workshop on Digital Communications, 1989, Elsevier, Amsterdam, 1990, pp. 117–131
4. W. van Etten, J. van der Plaats: *Fundamentals of optical fiber communications*. Prentice Hall, 1991
5. K. Holejko: *Pomiary widma laserów półprzewodnikowych za pomocą interferometrów światłowodowych*. Referaty Instytutu Telekomunikacji PW, z. 140, Warszawa, marzec 1992
6. C.H. Henry: *Theory of the linewidth of semiconductor lasers*. IEEE J. Quantum Electron., 1982, vol. QE-18, no.2, pp. 259–264
7. J. Siuzdak, W. van Etten: *BER performance evaluation for CPFSK phase and polarisation diversity coherent optical receivers*. J. Lightwave Technol. vol. 9, no. 11, pp. 1583–1592, November 1991
8. G. Nicholson: *Transmission performance of an optical FSK heterodyne system with a single filter envelope detection receiver*. J. Lightwave Technol., 1987, vol. 5, no. 4, pp. 502–509
9. I. Garrett, G. Jacobsen: *Broad linewidth (DFB) lasers in coherent optical systems for local loop applications*. ECOC 87, pp. 78–82
10. R. Hofstetter: *Koherenten Detektion optischer ASK- und FSK-Signale mit einem Energiemessemempfanger*. AEU, 1991, vol. 45, no. 3, pp. 168–175
11. L.G. Kazovsky i in.: *ASK multiport optical homodyne receivers*. ECOC'86, pp. 395–398
12. S.D. Personick: *Receiver design for digital fiber optic communication systems*. The Bell System Technical Journal, 1973, vol. 52, no. 6, pp. 843–874
13. K. Holejko: *Precyzyjne elektroniczne pomiary odległości i kątów*. WNT, Warszawa 1981
14. K. Emura i in.: *Optimum system design for CPFSK heterodyne delay demodulation systems with DFB LD's*. J. Lightwave Technol., 1990, vol. LT-8, no. 2, pp. 251–258

J. SIUZDAK

A SIMPLIFIED METHOD FOR THE DETERMINATION OF SENSITIVITY OF RECEPTION OF HETERODYNE OPTICAL SIGNAL WITH SUBSTANTIAL LINEWIDTH

Summary

Presented is an approximate method for the determination of sensitivity of heterodyne reception of optical signal with substantial linewidth by means of a parameter which generalizes SNR. The method was applied for the analysis of a coherent one filter demodulator of binary FSK signals. Sensitivity of this demodulator was compared with the sensitivity of the direct detection (noncoherent) receiver.

Key words: optical signals, alemodulators.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 1. ss. 113–129

Eksperymentalny system światłowodowej transmisji koherentnej z wykorzystaniem modulacji częstotliwości lasera

JERZY SIUZDAK

Instytut Telekomunikacji, Politechnika Warszawska

Otrzymano 1994.10.24

Autoryzowano do druku 1994.12.16

Opisano wykonany w Instytucie Telekomunikacji PW model systemu światłowodowej transmisji koherentnej wykorzystujący półprzewodnikowe lasery jednomodowe i modulację częstotliwości. System składa się z układów stabilizacji temperatury i prądów laserów, modulatora, odbiornika i demodulatora, sprzęgacza kierunkowego oraz kontrolera polaryzacji. W tak skonstruowanym systemie przeprowadzono transmisję sygnału świetlnego modulowanego częstotliwościowo. Modulację tę uzyskano przez modulację prądu lasera o małej głębokości. Udało się również uzyskać demodulację koherentną poprzez zmieszanie sygnału przychodzącego ze światłem heterodyny optycznej, a następnie detekcję i obróbkę elektryczną tak otrzymanego sygnału. Niestety wskutek dużej szerokości linii widmowych laserów nie udało się uzyskać stabilnej pracy systemu.

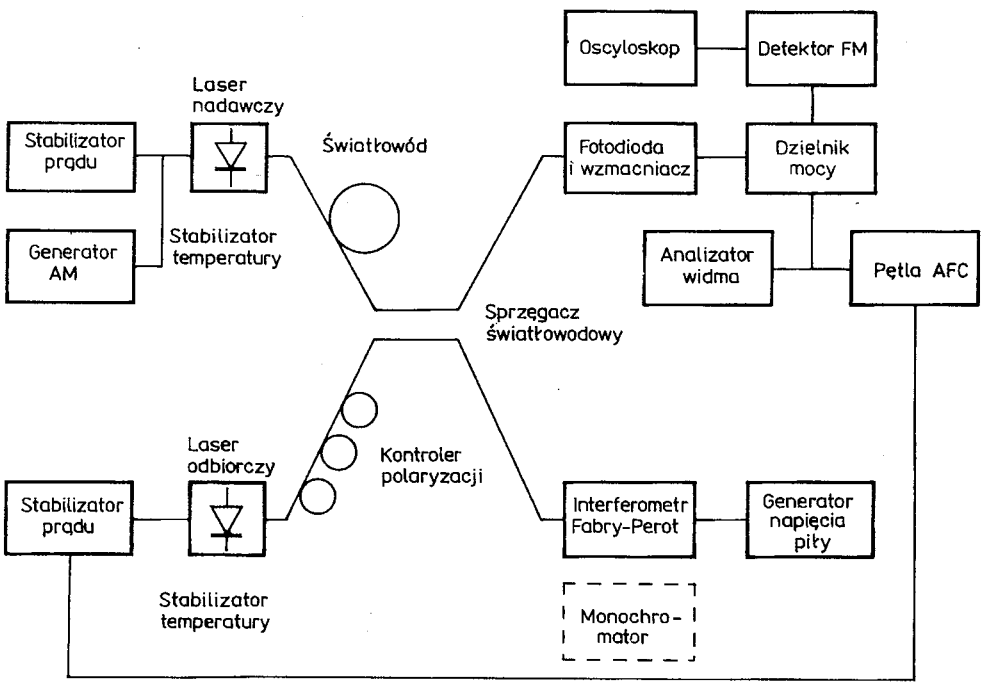
Słowa kluczowe: lasery jednomodowe, transmisja koherentna, modulacja częstotliwościowa, sygnały świetlne.

1. ZAŁOŻENIA I CELE PROJEKTU

Zgodnie z umową z KBN, projekt nr 802199101 obejmował wykonanie modelu systemu światłowodowej transmisji koherentnej z wykorzystaniem binarnej modulacji częstotliwości i użyciem jednomodowych laserów półprzewodnikowych. W skład systemu miały wchodzić następujące podukłady: układy stabilizacji temperatury i prądów laserów, modulator prądu lasera nadawczego, wzmacniacz i demodulator odbiorczy, pętla stabilizacji częstotliwości, generator danych pseudoprzypadkowych oraz kontroler polaryzacji. Spodziewane było również określenie optymalnych parametrów modulacji, kodowania oraz struktury demodulatora sygnału binarnego w przypadku, gdy szerokość widma linii emisyjnych laserów półprzewodnikowych jest porównywalna z szybkością transmisji, jak również wpływu zjawisk pasożytniczych na parametry transmisji.

2. OPIS OGÓLNY ZREALIZOWANEGO SYSTEMU

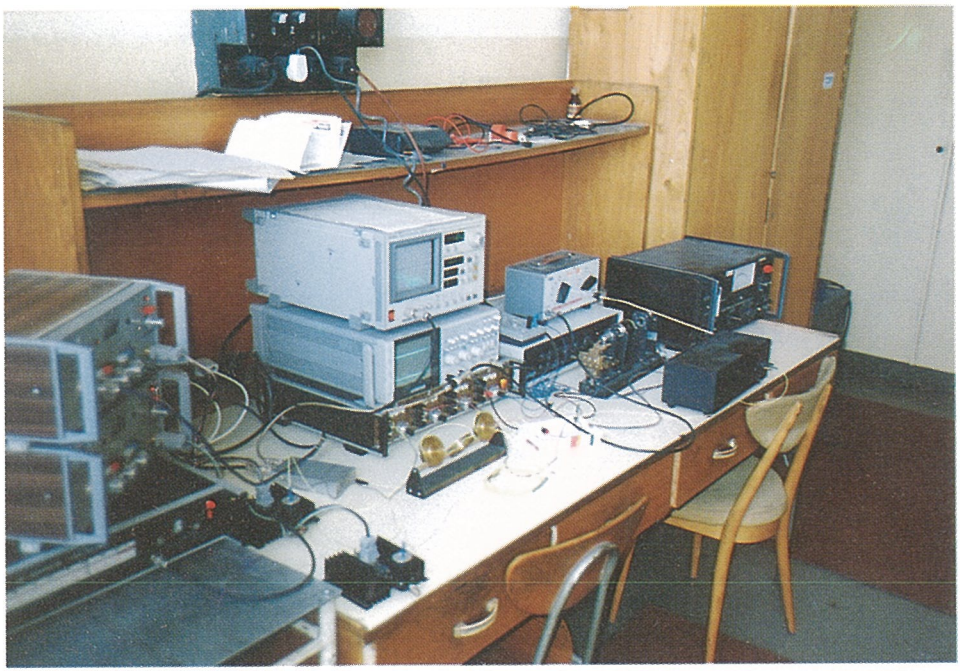
System pracował na długości fali 780 nm. Konieczna tutaj jest nastepująca uwaga. Otóż systemy koherentne budowane są zazwyczaj na długość fali 1550 nm i sporadycznie na 1300 nm. Wynika to z minimów tłumienia światłowodu na tych długościach fal. Wybór mniejszej długości fali w projekcie był podyktowany znacznie mniejszymi kosztami jednomodowych laserów półprzewodnikowych pracujących na tej długości fali (kilkaset wobec kilku – kilkunastu tysięcy dolarów).



Rys. 1. Schemat blokowy systemu

Schemat blokowy zrealizowanego systemu pokazano na rys. 1. Laser nadawczy (Seastar PT450H780), którego temperatura i prąd stały są stabilizowane, modulowany jest prądem o niewielkiej amplitudzie (kilka mA, przy prądzie stałym przekraczającym 50 mA). Wywołuje to zmiany długości emitowanego przez laser promieniowania w takt zmian prądu modulującego.

Po przejściu przez światłowód jednomodowy, sygnał świetlny dochodzi na wejście sprzęgacza światłowodowego o dwóch wejściach i dwóch wyjściach. Do drugiego wejścia doprowadzany jest sygnał świetlny z lasera lokalnego, który podobnie jak poprzedni laser ma stabilizowaną temperaturę i prąd. Dodatkowo objęty jest on pętlą sprzężenia zwrotnego, mającą stabilizować różnicę częstotliwości między obydwoma laserami. Pomiędzy laserem lokalnym, a sprzęgaczem światłowodowym umieszczono manualny kontroler polaryzacji, mający za zadanie dopasowanie stanu polaryzacji



Rys. 2. Wygląd zewnętrzny systemu transmisyjnego. Od lewej widać stabilizatory laserów wraz z głowicami (czarne radiatory), układ fotodetektora, wzmacniacza i demodulatora (srebrne pudełko), kontroler polaryzacji (złote krążki) oraz oscyloskop i analizator widma. Bardziej na prawo widać interferometr Fabry-Perot (w centrum), monochromator (dolna prawa strona), generator napięcia piłokształtnego (lewa strona). Widoczne są też przyrządy pomocnicze (miernik selektywny, miernik uniwersalny i tłumik dekadowy)

światła lasera lokalnego do odbieranego sygnału. Na obydwu wyjściach sprzęgacza światłowodowego otrzymuje się superpozycję sygnałów wejściowych.

Jedno z wyjść sprzęgacza wykorzystywane jest do wstępnego przyrównania długości fal obydwu laserów poprzez obserwację ich widm za pomocą monochromatora lub interferometru Fabry-Perot. Zmianę długości fali uzyskuje się poprzez zmiany temperatury i prądu laserów. Sygnał z drugiego wyjścia zamieniany jest w fotodiodzie pin (Seastar CP-120-20) na sygnał elektryczny. Prąd płynący przez fotodiodę jest proporcjonalny do mocy padającego nań promieniowania

$$I = RP = R(\sqrt{2P_0}\cos(2\pi f_1 t + \phi_1) + \sqrt{2P_L}\cos(2\pi f_2 t + \phi_2))^2, \tag{1}$$

gdzie R jest czułością fotodiody, P_0 i P_L są mocami promieniowania obydwu laserów, f_1 i f_2 ich częstotliwościami leżącymi w zakresie częstotliwości fal optycznych, zaś ϕ_1 i ϕ_2 ich fazami. Z wzoru (1) wynika, że prąd fotodiody oprócz sygnałów stałych i sygnałów leżących w zakresie częstotliwości optycznych, zawiera również składowe widmowe o częstotliwościach różnicowych $(f_1 - f_2)$:

$$i = 2R\sqrt{P_0 P_L}\cos[2\pi(f_1 - f_2)t + \phi_1 - \phi_2]. \tag{2}$$

Jeżeli polaryzacje sygnałów odbieranego i heterodyny optycznej są zbliżone, zaś różnica częstotliwości między nimi znajduje się w obrębie pasma przepustowego wzmacniacza, na ekranie analizatora widma można obserwować widmo sygnału elektrycznego odpowiadającego częstotliwości różnicowej między obydwoma laserami. Jest to oczywiście prawdziwe tylko wtedy, gdy podwójna szerokość linii widmowej jest mniejsza od pasma wzmacniacza i analizatora widma.

Sygnał wyjściowy wzmacniacza jest wykorzystywany również w pętli automatycznej regulacji częstotliwości, jak również dochodzi do detektora częstotliwości. Jeżeli częstotliwość różnicowa ulega zmianom w takt zmian prądu modulującego laser nadawczy, to zmiany te odzwierciedlone są na wyjściu detektora częstotliwości i uwidaczniane na oscyloskopie. Zdjęcie systemu przedstawione jest na rys. 2 wraz z objaśnieniami.

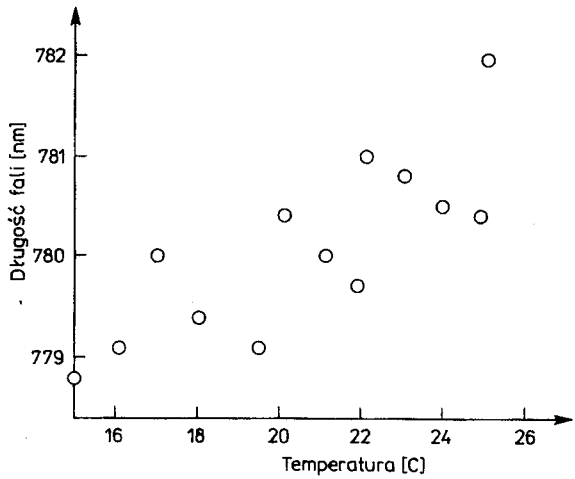
3. ELEMENTY SKŁADOWE SYSTEMU

Omówimy teraz dokładniej najważniejsze elementy zrealizowanego systemu transmisyjnego.

3.1. LASERY NADAWCZE I ODBIORCZE

Wpływ temperatury na długość fali emitowanego promieniowania laserowego

Pomierzona za pomocą monochromatora charakterystyka temperaturowa długości emitowanego promieniowania znacznie odbiega od charakterystyki typowej

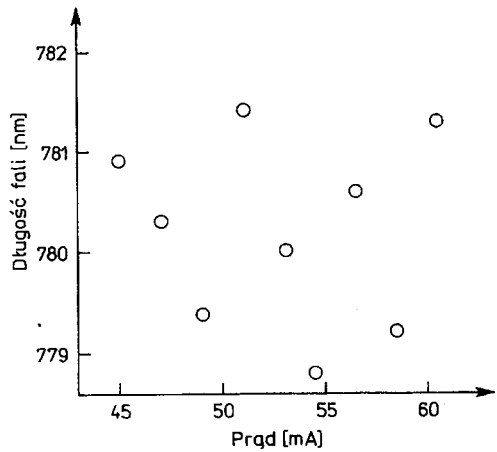


Rys. 3. Zależność długości emitowanego przez laser promieniowania od temperatury

[5, 14] i pokazana jest na rys. 3. Nie jest jasne, czy taki przebieg charakterystyki jest wynikiem celowego zabiegu zmierzającego do ograniczenia zalezności temperaturowej emitowanego promieniowania, czy efektem wywołanym odbiciami wstecznymi od układu optycznego sprzęgającego strukturę laserową ze światłowodem. Średnie nachylenie charakterystyki wynosi około 0.25 nm/K, co odpowiada 120 GHz/K i jest zgodne z danymi katalogowymi. Taki przebieg zależności wynika m.in. ze zmian temperaturowych wzmocnienia, absorpcji, dryftu nosników, współczynnika załamania oraz temperaturowych zmian długości rezonatora. Należy zaznaczyć, że te wyniki są wynikami średnimi i nie odzwierciedlają struktury modowej badanego lasera.

Zależność długości emitowanego promieniowania od prądu

Na rys. 4 pokazano zmierzoną zależność długości emitowanego przez laser promieniowania od prądu. Odbiega ona znacznie od typowej charakterystyki. Taki przebieg zależności wynika m.in. ze zmian współczynnika załamania wywoływanych zmianami koncentracji nośników. Ponownie nie jest jasne, czy jest to celowy zabieg zmierzający do ograniczenia zależności prądowej długości fali emitowanego promieniowania, czy efekt wywołany odbiciami wstecznymi od układu optycznego sprzęgającego strukturę laserową ze światłowodem. Należy znowu zaznaczyć, że te wyniki są wynikami średnimi i nie odzwierciedlają struktury modowej badanego lasera. Pomierzone interferometrem Fabry-Perot nachylenie charakterystyki w obrębie jednego modu wynosiło około 2 GHz/mA, co jest dość typową wartością.



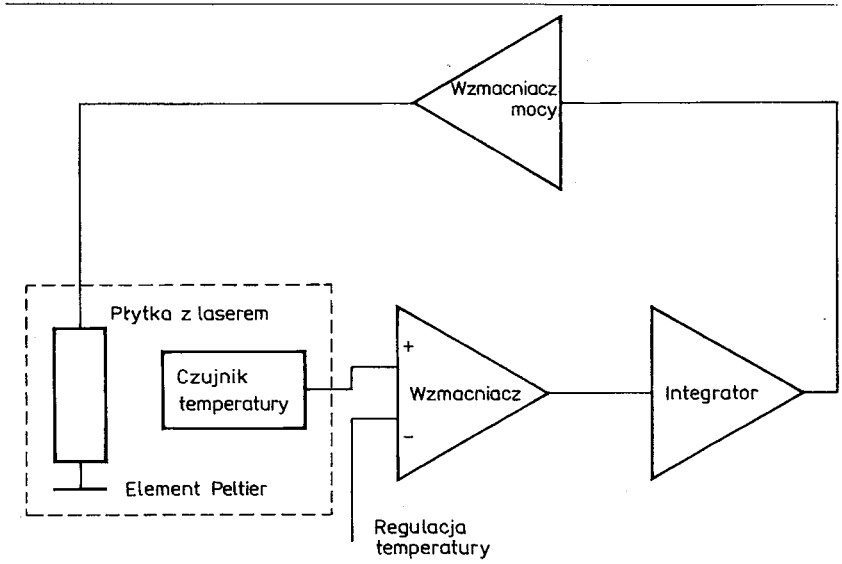
Rys. 4. Zależność długości emitowanego przez laser promieniowania od prądu stałego

3.2. STABILIZACJA DŁUGOŚCI FALI EMITOWANEGO PRZEZ LASERY PROMIENIOWANIA

Ażeby detekcja koherentna była możliwa konieczne jest, aby laser nadawczy i heterodyna optyczna pracowały stabilnie na jednakowych długościach fal. Ponieważ, jak omówiono poprzednio, długość emigowanego przez laser promieniowania zależy zarówno od jego temperatury jak i prądu, konieczna jest stabilizacja tych obydwu parametrów.

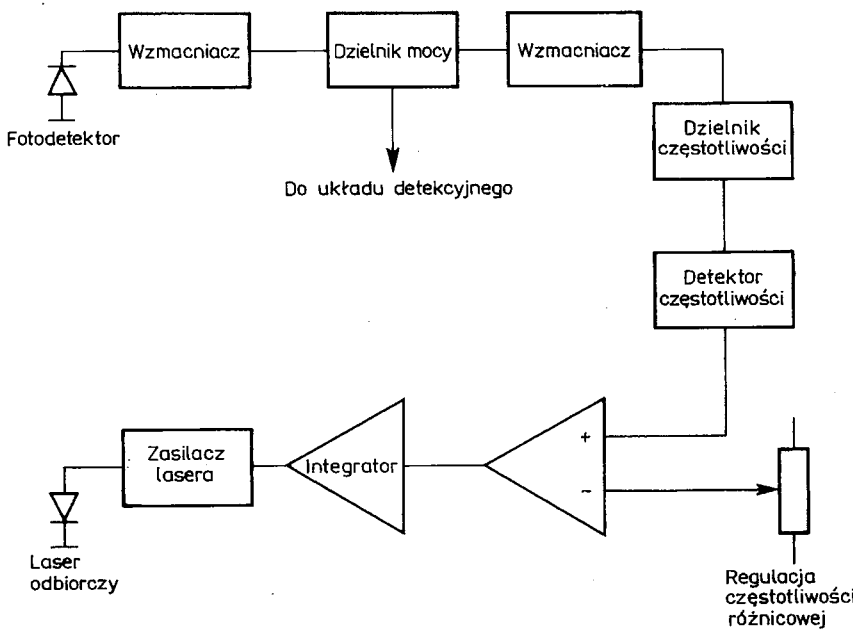
Zmiany częstotliwości promieniowania lasera wraz z jego prądem są rzędu 2 GHz/mA, w związku z tym zachowanie stabilizacji częstotliwości lasera w zakresie $+/-10$ MHz wymaga stabilizacji prądu z dokładnością kilku mikroamperów. Jest to możliwe do uzyskania przy użycu precyzyjnych i stabilnych elementów elektronicznych wysokiej klasy oraz układu sprzężenia zwrotnego. Zastosowany sprzęt umożliwiał stabilizację prądu z dokładnością pojedynczych mikroamperów, a więc zupełnie wystarczającą.

Schemat układu stabilizacji temperatury jest układem typowym [8, 26] i pokazano go na rys. 5. Laser półprzewodnikowy wraz z czujnikiem temperatury (układ LM35), którego napięcie wyjściowe jest proporcjonalne do temperatury, umieszczone są w jednej obudowie oddzielonej od radiatora elementem Peltier. W stanie równowagi, kiedy temperatura lasera (i obudowy) jest równa żądanej temperaturze, sygnał wyjściowy ze wzmacniacza rożnicowego jest zerowy i integrator utrzymuje kierunek i wielkość prądu płynącego przez element Peltier. Wszelke odchyłki od żądanej temperatury powodują powstanie napięcia niezrównoważenia sterującego integrator. Napięcie wyjściowe integratora, poprzez wzmacniacz mocy, steruje prądem elementu Peltier powodując zmianę temperatury między jego okładkami kompensujące początkową odchyłkę. Układ zapewniał stabilizację temperatury z dokładnością 0.1°C.



Rys. 5. Schemat blokowy układu stabilizacji temperatury

Dokładność stabilizacji temperatury odpowiadała dokładności stabilizacji częstotliwości różnicowej laserów w obrębie jednego modu rzędu pojedynczych gigaherców, co jest wielkością o wiele za dużą. W związku z tym w układzie zastosowano pętlę automatycznej stabilizacji częstotliwości różnicowej laserów typu prądowego [8, 25]. Przedstawiono ją na rys. 6. Zaznaczmy tutaj, że dokładność stabilizacji w obrębie



Rys. 6. Schemat układu automatycznej stabilizacji częstotliwości

jednego modu jest lepsza o rząd wielkości od wielkości wynikającej ze średniego nachylenia charakterystyki.

Sygnał elektryczny z fotodiody o częstotliwości równej różnicy między częstotliwościami obydwu laserów ulegał wzmocnieniu, a następnie dochodził do cyfrowego dzielnika częstotliwości i detektora częstotliwości. Sygnał wyjściowy detektora częstotliwości (proporcjonalny do chwilowej częstotliwości różnicowej) porównywany był z sygnałem odniesienia odpowiadającym żądanej częstotliwości różnicowej. Dopóki obydwie te częstotliwości były jednakowe, nie zachodziły żadne zmiany w prądzie stałym sterującym laser. Jeżeli jednak wystąpiła zmiana częstotliwości różnicowej, to wówczas ulegał zmianie prąd stały lasera tak, aby skompensować tę zmianę częstotliwości.

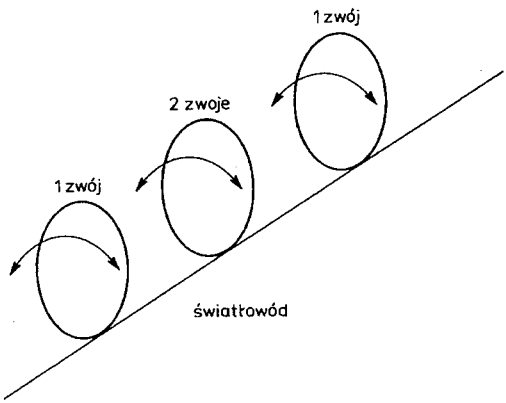
Przewidziany okres chwytania pętli jest równy pasmu kaskadowego połączenia fotodiody, wzmacniaczy i dzielnika częstotliwości i wynosi około 1 GHz. Konieczne jest tu wstępne sprowadzenie częstotliwości różnicowej do częstotliwości mniejszej od 1 GHz. Przewidziany zakres trzymania jest znacznie większy i praktycznie równy zakresowi pracy lasera w obrębie jednego modu.

Niestety w praktycznym układzie pętla nie pracowała prawidłowo. Przyczyną tego faktu było to, że sumaryczna szerokość linii widmowych laserów (powyżej 4 GHz) znacznie przekraczała pasmo pętli (1 GHz). W efekcie tego sygnał pochodzący od mieszania koherentnego był widziany przez pętlę jedynie jako wzrost poziomu szumów, a nie jako prążek o zdefiniowanej częstotliwości. Detektor częstotliwosci nie mógł pracować poprawnie w takich warunkach i normalna praca pętli nie była możliwa.

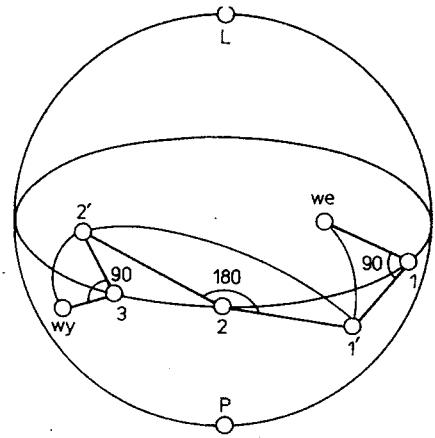
3.3. INNE PRZYRZĄDY

Dla umożliwienia detekcji heterodynowej na fotodiodzie, konieczne jest dodanie do siebie promieniowania odbieranego i promieniowania pochodzącego z lasera lokalnego. W układzie eksperymentalnym wykorzystano w tym celu światłowodowy sprzęgacz kierunkowy o dwóch wejściach i dwóch wyjściach.

Aby detekcja koherentna była możliwa, polaryzacja sygnału odbieranego musi być zgodna z polaryzacją światła z generatora lokalnego (heterodyny optycznej). Zatem konieczna jest możliwość regulacji stanu polaryzacji heterodyny optycznej tak, aby dopasować jej stan do dowolnej polaryzacji sygnał odbieranego. Ze względu na prostotę zdecydowano się na wykorzystanie stosunkowo najprostszego manualnego kontrolera polaryzacji typu opisanego w pracy [11]. Składał się z kolejno nawiniętych światłowodem na różnych rdzeniach o jednakowych średniach 1 zwoju (odpowiednik płytki ćwierć falowej), 2 zwojów (-pół falowej) i znowu 1 zwoju (-ćwierć falowej). Obrot głównych osi dwójłomnych cewek światłowodowych był dokonywany poprzez obrót płaszczyzny cewki światłowodowej (rys. 7), przez co uzyskano zachowanie analogiczne do klasycznych płytek ćwierć- i półfalowej. Można łatwo pokazać, że takie połączenie umożliwia zamianę dowolnej polaryzacji w żądaną polaryzację. Zilustrowano to na rys. 8 na sferze Poincare, będącej dogodnym zobrazowaniem stanu polaryzacji fali elektromagnetycznej [12].



Rys. 7. Kontroler polaryzacji



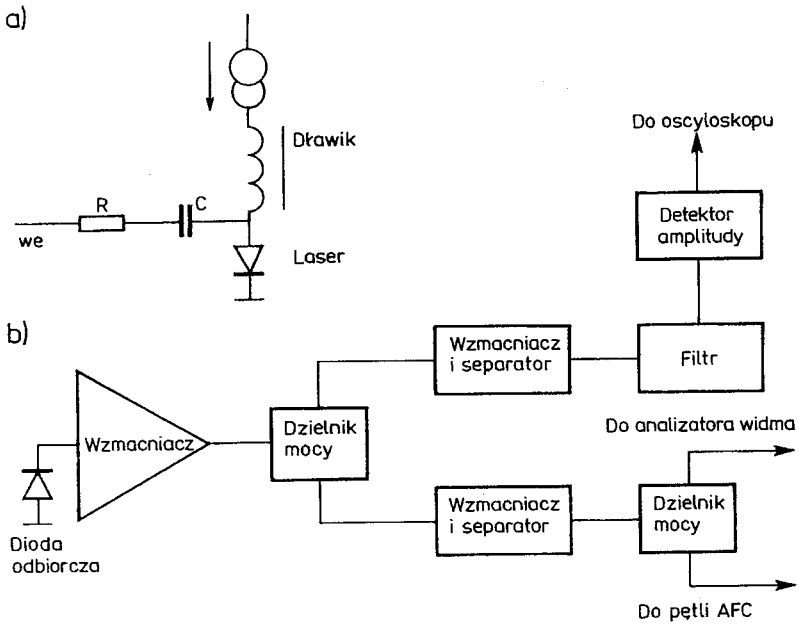
Rys. 8. Ilustracja działania kontrolera polaryzacji na sferze Poincare. Punkty 1, 2, 3 odpowiadają osiom optycznym odpowiednio: wejściowej cewki ćwierćfalowej, cewki półfalowej i wyjściowej cewki ćwierćfalowej

Do zgrubnego określenia długości fali lasera półprzewodnikowego służy monochromator. Do precyzyjniejszego wyznaczenia długości fali badanego promieniowania służy interferometr Fabry-Perot [28].

3.4. MODULATOR I DEMODULATOR

Układ modulatora przedstawiono na rys. 9 a. Był on włączony w pętlę stabilizacji prądu lasera i miał za zadanie umożliwienie modulacji prądu lasera o niewielkiej głębokości (rzędu kilku procent). Modulacja taka pozwalała zmieniać w pewnych granicach długość fali emitowanego przez laser promieniowania i nie wpływała na stały prąd lasera. Rezystor R = 50 omów zapewniał dopasowanie do linii, a jednocześnie linearyzował charakterystykę modulacji ze względu na znacznie większą wartość od rezystancji dynamicznej lasera (rzędu 2.5 ohma).

Ze względu na prostotę i małą wrażliwość na szumy fazowe laserów, w realizacji wybrano demodulator z jednym filtrem [15] pokazany na rys. 9 b). Przeznaczony jest do odbioru sygnałów o dużej dewiacji częstotliwości, charakteryzujących się widmem skupionym w pobliżu dwóch częstotliwości. Filtr w odbiorniku jest dopasowany do jednej z nich. Wadą tego typu demodulatora jest stosunkowo niewielka czułość, zaś jego zaletą — prostota i tolerancja szumu fazowego laserów i nieliniowości charakterystyki FM lasera nadawczego. Sumaryczna szerokość linii widmowej laserów może wynosić 5% szybkości transmisji przy stracie czułości równej 1 dB [29]. Jednak praktyczna możliwość tolerancji dużej szerokości widmowej laserów przez układ demodulatora z jednym filtrem jest znaczna. Konieczne jest jedynie zapewnienie dostatecznie dużej dewiacji częstotliwości tak, aby widma odpowiadające poszczególnym częstotliwościom się nie pokrywały, oraz rozszerzenie pasma każdego z filtrów



Rys. 9. Schematy układów: a) modulatora, b) demodulatora

tak, aby obejmował on całe widmo sygnału modulowanego odpowiadającego danej częstotliwości. Widmo to jest bowiem splotem widma niemodulowanego lasera (laserów) i widma modulacji

$$S_c(f)=S(f)*\Phi(f)\,. \tag{3}$$

gdzie

$$S(f)=\frac{2P_0}{\pi B_L}\,\frac{1}{1+\left(\dfrac{f-f_0}{B_L/2}\right)^2}\,. \tag{4}$$

W powyższym wzorze P_0 jest mocą emitowaną przez laser (lasery), B_L jest szerokością widmową linii lasera (FWHM — pełna szerokość w połowie maksimum) w przypadku obserwacji jednej linii, bądź sumą szerokości linii widmowych lasera nadawczego i heterodyny w przypadku odbioru heterodynowego.
Ponadto

$$\Phi(f)=A_0\left[\frac{\sin\pi fT}{\pi fT}\right]^2(1+\cos 2\pi fT)\,, \tag{5}$$

gdzie $1/T$ jest szybkością transmisji.

W praktycznej realizacji sygnał świetlny ze światłowodu po detekcji na diodzie *pin* ulegał wzmocnieniu w trzystopniowym wzmacniaczu w.cz. o minimalnym wzmocnieniu 42 dB. Pasmo przenoszenia odbiornika wynosiło około 1 GHz. Sygnał

wyjściowy ze wzmacniacza był dzielony pomiędzy analizator widma, pętlę AFC i demodulator FM. Demodulator FM składał się z filtru dolnopasmowego o paśmie przepustowym 50 MHz oraz detektora amplitudy (obwiedni), którego wyjście połączone było z oscyloskopem. Jeżeli bieżąca różnica częstotliwości między laserem nadawczym, a heterodyną optyczną, mieściła się w paśmie filtru, to wówczas sygnał na wyjściu detektora amplitudy wzrastał, co odpowiadało symbolowi „1", w przeciwnym przypadku sygnał był mały, co odpowiadało symbolowi „0". Ze względu na to, że szerokość pasma laserów była w rzeczywistym układzie znacznie większa od szybkości modulacji zastosowano w demodulatorze filtr dolnopasmowy. Umożliwiło to obserwację zdemodulowanego sygnału na ekranie oscyloskopu (rys. 10). Rozwiązanie takie zastosowano dla celów demonstracyjnych, przy braku mozliwości (jak to będzie później wyjaśnione) stabilizacji częstotliwości laserów. Nie ma ono jednak korzystnych własności, jeśli chodzi o elementową stopę błędów. Wynika to z tego, że wielkość mocy sygnału wewnątrz pasma przepustowego filtru jest wielkością statystyczną i podlega fluktuacjom, powodując zjawisko tzw. „BER floor" [6].

4. WPŁYW ZJAWISK PASOŻYTNICZYCH NA PRACĘ SYSTEMU I KOMPENSACJA TEGO WPŁYWU

W tej części zajmiemy się opisem niepożądanych zjawisk wystepujących w systemie transmisyjnym. W praktyce wyznaczyły one pracę systemu. Do tych zjawisk należa przede wszystkim odbicia wsteczne i nierównomierna charakterystyka modulacyjna lasera.

4.1. WPŁYW ODBIĆ WSTECZNYCH NA PRACĘ LASERÓW

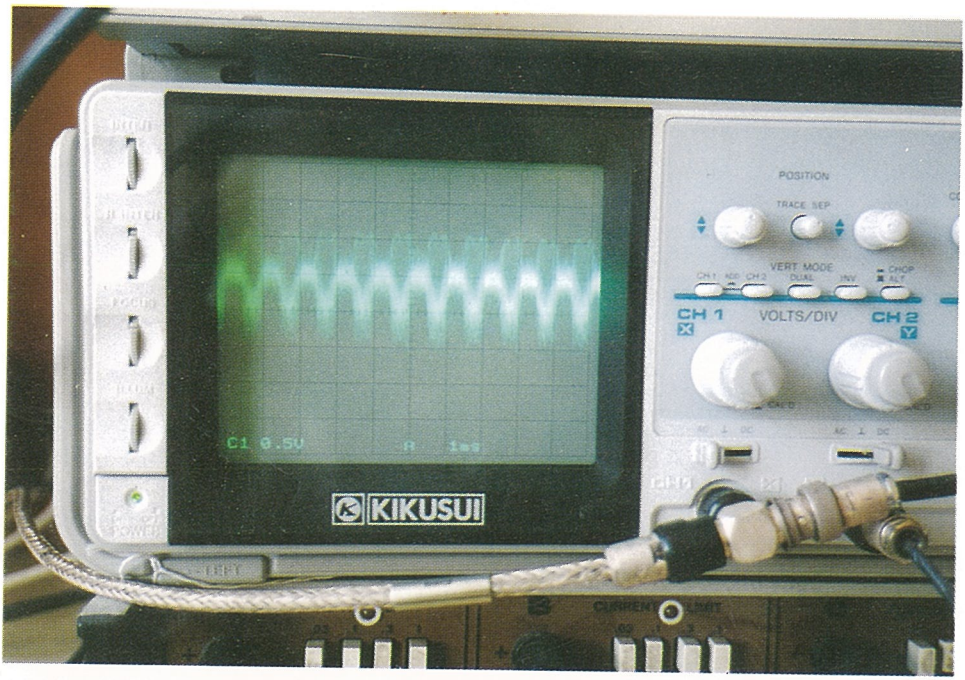
Jeżeli zapisze się równanie van der Pola dla pola elektrycznego w laserze z dodatkowym członem opisującym pole odbite, przyjmie ono postać [1]

$$\frac{dE}{dt}=\left(-j\omega_0+\frac{\Delta G}{2}(1-j\alpha)\right)E(t)+k_eE(t-t_e), \tag{6}$$

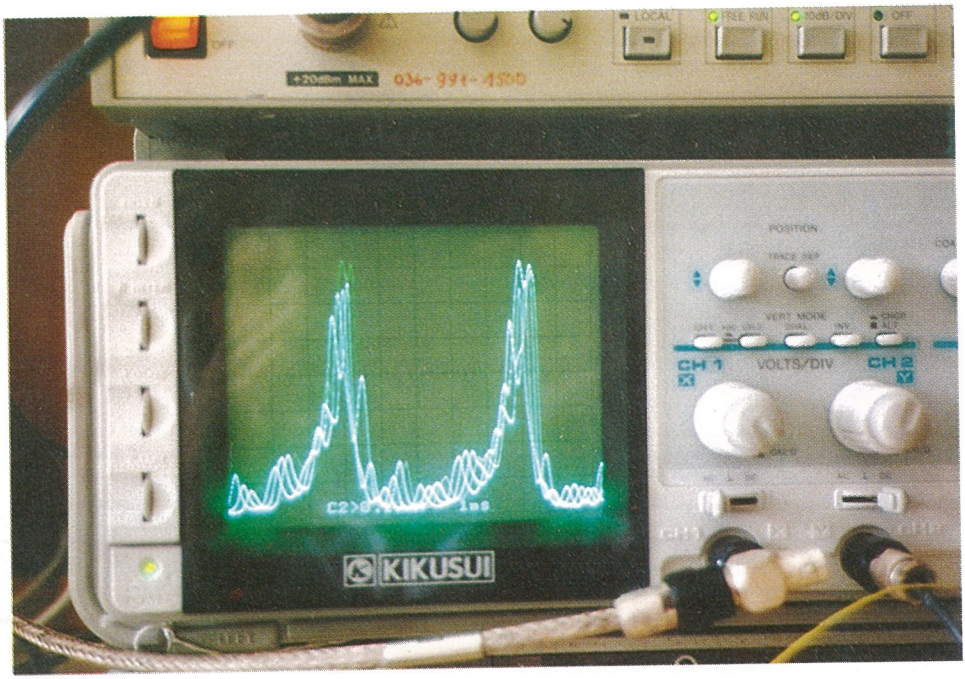
gdzie ω_0 jest pulsacją swobodnych drgań lasera, ΔG jest zmianą wzmocnienia spowodowaną odbiciami wstecznymi, α jest tzw. współczynnikiem wzmocnienia linii [2], t_e jest opóźnieniem odbitego pola, zaś k_e jest współczynnikiem wiążącym współczynniki odbicia wewnętrzny i zewnętrzny z czasem przejścia promieniowania przez strukturę lasera [1]. W powyższym równaniu pominięto składniki szumowe. Równania wyrażające zmiany wzmocnienia i pulsacji spowodowane odbiciami są następujące

$$\Delta G=-2k_e\cos(\omega t_e), \qquad \delta\omega=-k_e(\sin(\omega t_e)+\alpha\cos(\omega t_e)). \tag{7}$$

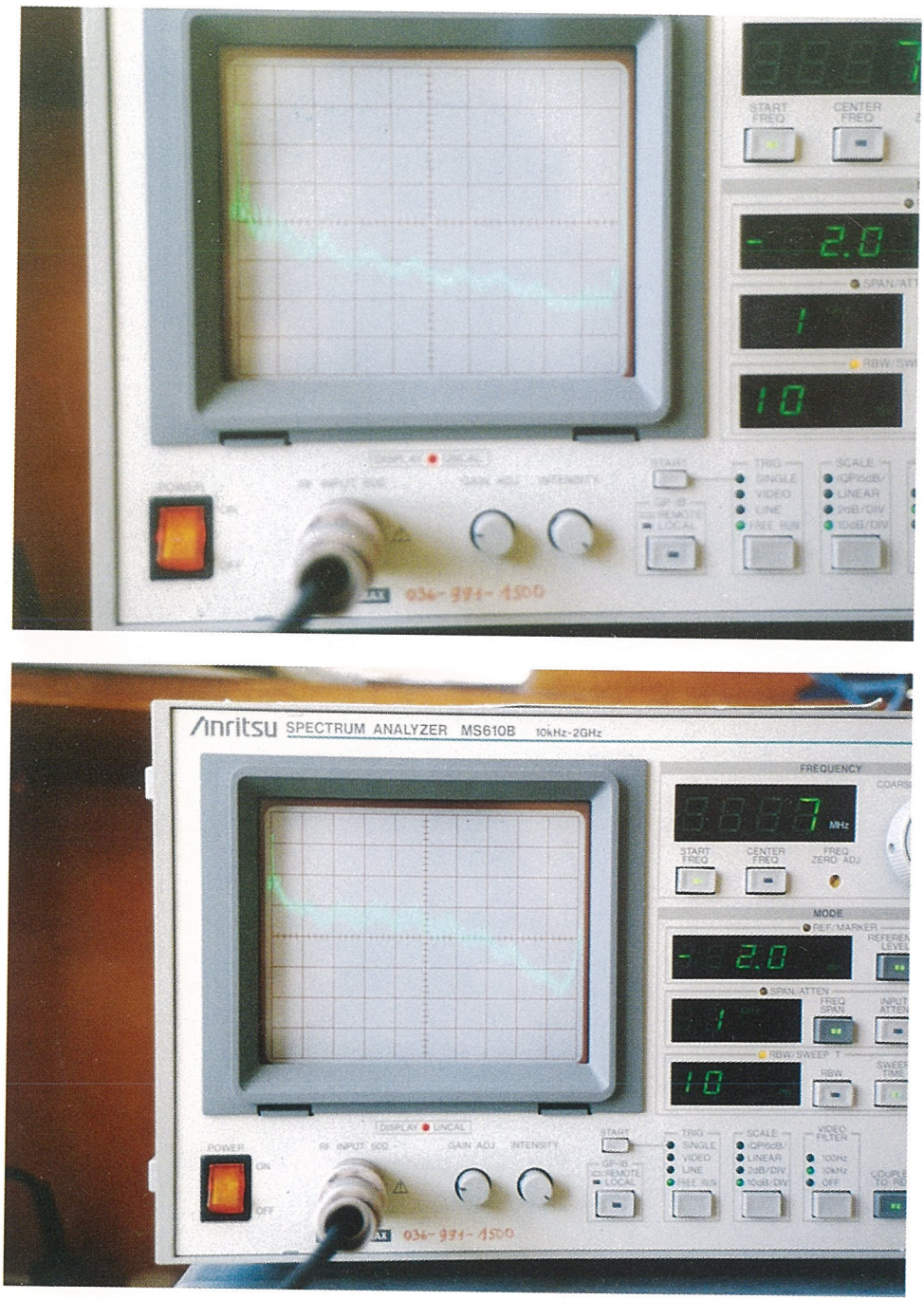
Jeżeli opóźnienie zewnętrzne t_e jest mniejsze od czasu koherencji promieniowania, można obliczyć szerokość linii widmowej z uwzględnieniem odbić [3]:


POSITION
TRACE SEP
VERT MODE
CH 1
VOLTS/DIV
CH 2
C1 0.5V
A 1ms
KIKUSUI

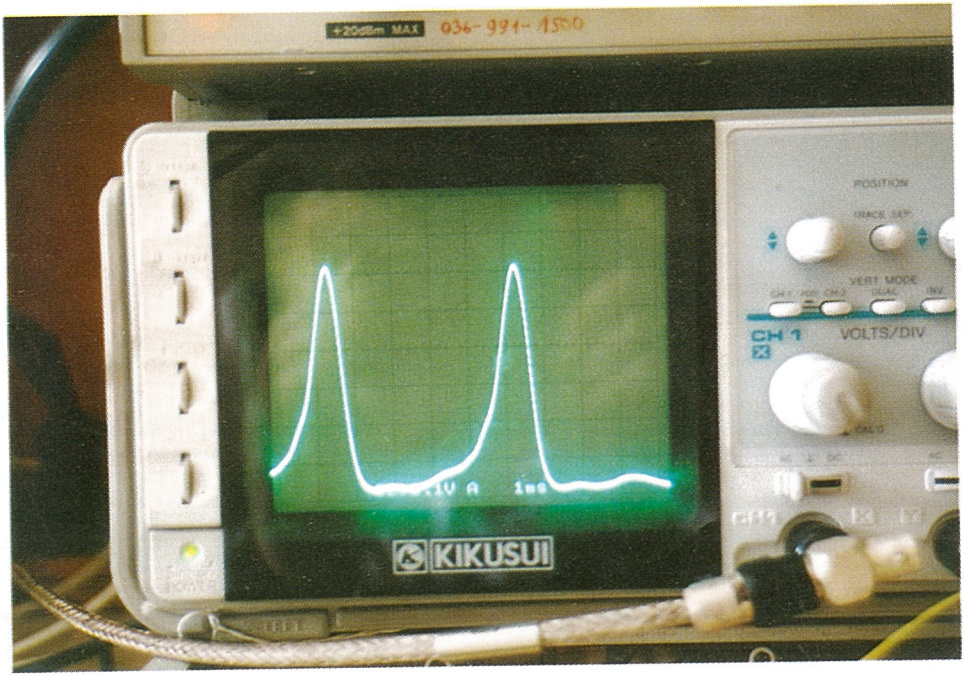

KIKUSUI



Rys. 10. Sygnał na wyjściu demodulatora obserwowany na ekranie oscyloskopu (zdjęcie górne $f_{mod} = 1$ kHz, $i_{mod} = 1$ mA). Zdjęcie środkowe odpowiada podobnie zmodulowanemu sygnałowi, ale bez detekcji heterodynowej. Zdjęcie dolne pokazuje otrzymane w interferometrze Fabry-Perot (FSR = 15 GHz) widmo zmodulowanego sygnału



Rys. 11. Wygląd linii widmowej użytego lasera, FSR = 15 GHz



Rys. 12. Ekran analizatora widma: a) same szumy, b) linie widmowe laserów (detekcja heterodynowa), zakres częstotliwości = 1 GHz, czułość = 10 dB/dz

$$\delta f = \frac{\delta f_0}{[1+\sqrt{1+\alpha^2}\, k_e t_e \cos(\omega t_e + \tan^{-1}\alpha)]^2}, \tag{8}$$

gdzie δf_0 jest szerokością widmową lasera bez odbić zewnętrznych. W zależności od siły sprzężenia zwrotnego (współczynniki odbicia), reprezentowanego przez k_e, oraz fazy odbitego promieniowania reprezentowanego przez $\omega_0 t_e$, można wyróżnić kilka rodzajów pracy lasera [1]:

I. Dla najniższych poziomów sprzężenia zwrotnego (rzędu -80 dB) obserwowane jest kilkudziesięcioprocentowe rozszerzenie lub zwężenie szerokości linii lasera, zależne od fazy odbitego promieniowania.

II. Przy zwiększaniu poziomu odbicia obserwuje się rozpad linii widmowej na kilka rywalizujących ze sobą linii (tzw. mode hopping).

III. Przy dalszym zwiększaniu poziomu odbicia, w zakresie -45 dB do -39 dB nie ma zjawiska mode hopping, obserwuje się stabilną jednomodową pracę lasera przy zwężeniu szerokości jego linii widmowej.

IV. W zakresie poziomu sprzężenia zwrotnego od około -40 dB do -10 dB pojawiają się mody satelitarne, odseparowane od głównego modu o częstotliwość równą częstotliwości relaksacji. Przy wzroście poziomu sprzężenia zwrotnego mody satelitarne rosną powodując zwiększenie szerokości linii nawet do kilkudziesięciu gigahertzów.

V. W zakresie poziomu sprzężenia zwrotnego powyżej -10 dB laser pracuje jako laser o długiej wnęce rezonansowej z krótkim obszarem aktywnym, co daje stabilną jednomodową pracę przy wielokrotnej redukcji szerokości linii widmowej lasera.
Dla użytych w eksperymencie diod laserowych długość koherencji nie przekraczała pojedynczych metrów. W związku z tym opisana powyżej teoria poprawnie opisuje jedynie najbliższe odbicia dla pojedynczego lasera. Na te odbicia nakładają się odbicia z dalszej odległości światła już niekoherentnego pochodzącego z danego lasera, jak również (poprzez sprzęgacz światłowodowy) od drugiego lasera.

Zjawiska zachodzące w laserach półprzewodnikowych wywołane odbiciem od odległych reflektorów były analizowane w pracy [4]. Chodzi tu o periodyczne niskoczęstotliwościowe fluktuacje mocy wyjściowej w obecności umiarkowanego sprzężenia zwrotnego i spowodowane tym zwiększenie szerokości widmowej linii lasera oraz wiążące się z tym zwiększenie szumu RIN.

Przyczyną tych fluktuacji są skończone zmiany amplitudy, fazy promieniowania oraz liczby nośników spowodowane emisją spontaniczną. Te fluktuacje zaburzają synchronizację pola wewnątrz rezonatora laserowego i pola odbitego, przy czym są one niewielkie dla prądów nieznacznie przekraczających prąd progowy i rosną wraz ze wzrostem prądu lasera.

Użyte w eksperymencie diody laserowe typu PT-450-780 H według danych katalogowych [5] były laserami jednomodowymi i miały szerokości linii widmowych rzędu

$$\Delta f = \frac{c\,[\mathrm{m/s}]}{0.8 P_{out}\,[\mathrm{mW}]}\,[\mathrm{Hz}] \tag{9}$$

co daje wynik rzędu kilkudziesięciu megahertzów przy typowych mocach wyjściowych.

W rzeczywistym eksperymencie zmierzona szerokość linii widmowej lasera (FWHM) wynosiła około 2 GHz przy prądzie rzędu 55 mA. Jej wygląd pokazano na rys. 11. Zdjęcia wykonano przy pomocy interferometru Fabry-Perot (długość wnęki rezonansowej 10 mm, co odpowiada wolnemu zakresowi spektralnemu FSR = 15 GHz).

Na ekranie analizatora widma linie widmowe były obserwowane jedynie jako zwiększenie poziomu szumów ze względu na to, że sumaryczna szerokość linii laserów była znacznie większa od pasma przepustowego układu pomiarowego. Pokazane jest to na rys. 12.

W badanym układzie można wyróżnić trzy źródła odbić wstecznych:

— optyka sprzęgająca laser ze światłowodem znajdująca się wewnątrz obudowy lasera,

— sprzęgacz światłowodowy,

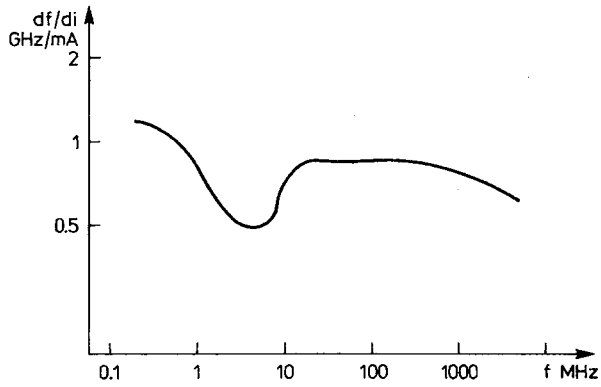
— koniec światłowodu.

Temu pierwszemu źródłu należy najprawdopodobniej przypisać wzmiankowane uprzednio rozszerzenie linii widmowej, przy czym była to wielkość stosunkowo stabilna i mało zależna od warunków pracy lasera (temperatura, prąd). Dwa następne źródła odbić znajdowały się w odległości przekraczającej drogę koherencji promieniowania laserowego i były najprawdopodobniej przyczyną obserwowanych czasami fluktuacji mocy promieniowania. Dała się ponadto zauważyć, zgodna z teorią [4], stabilniejsza praca laserów dla mniejszych prądów.

Jedynym rozwiązaniem pozwalającym wyeliminować niekorzystny wpływ odbić wstecznych wydaje się być użycie zintegrowanych z laserem izolatorów optycznych [5], ale jest to rozwiązanie bardzo kosztowne. Z braku środków finansowych nie można było go zastosować.

4.2. ZALEŻNOŚĆ DŁUGOŚCI FALI EMITOWANEJ PRZEZ LASER OD PRĄDU MODULUJĄCEGO

Jak już wspomniano w rozdziale dotyczącym stabilizacji temperatury i prądów laserów, długość fali promieniowania emitowanego przez laser zależy od prądu przezeń płynącego. Zaprezentowane we wspomnianym rozdziale zależności (rys. 4) są zależnościami statycznymi tzn. obowiązującymi dla prądów stałych. Długość fali lasera modulowanego prądem zmiennym zachowuje się w nieco inny sposób, okazuje się bowiem, że dewiacja częstotliwości dla bezpośredniej modulacji FM zmienia się wraz z częstotliwością modulującego prądu [7, 8]. Zazwyczaj jest ona największa dla prądu stałego, w badanym układzie wynosiła ona ok. 2 GHz/mA, a następnie maleje niemonotonicznie wraz z częstotliwością. Zmianie ulega również charakterystyka



Rys. 13. Przykładowa zależność czułości modulacji FM od częstotliwości

fazowa [7]. Przykład charakterystyki częstotliwościowej czułości modulacji podano na rys. 13. Można w niej wyróżnić dwa zakresy: nieregularny początkowy, wywołany modulacją temperatury lasera spowodowaną modulacją prądową, oraz drugi zbliżony do stałego w zakresie wyższych częstotliwości wywołany zmianą koncentracji nośników. W przeprowadzonym eksperymencie nie pomierzono charakterystyki modulacji z braku odpowiedniego sprzętu.

4.3. KOREKCJA NIERÓWNOMIERNEJ CHARAKTERYSTYKI MODULACJI FM LASERA

Nierównomierna charakterystyka modulacji FM lasera jest przyczyną wzrostu stopy błędów niezależnie od użytej metody demodulacji [31, 32]. Powodem tego faktu jest zmiana głębokości modulacji wraz ze zmianą częstotliwości sygnału modulującego. W związku z tym, dla dwóch sygnałów modulujących o jednakowych amplitudach, lecz różnych częstotliwościach, sygnały wyjściowe z demodulatora FM w odbiorniku będą miały różne amplitudy, w oczywisty sposób zmniejszając odporność odbiornika na zakłócenia. Iloraz amplitud na wyjściu demodulatora jest taki sam jak stosunek transmitancji FM lasera na odpowiednich częstotliwościach.

Dla kompensacji wpływu tego negatywnego zjawiska konieczne jest zachowanie określonego przebiegu charakterystyki FM lasera, jak również wybór optymalnego indeksu modulacji [31, 32]. W praktycznych układach spełnienie tych dwóch warunków jest bardzo utrudnione ze względu na to, że przebieg charakterystyki FM jest zależny od wielu czynników; m.in. od egzemplarza lasera i od prądu płynącego przez laser. Ta zmienność praktycznie uniemożliwia zwykłą korekcję omawianej charakterystyki. Istnieje kilka metod rozwiązania tego problemu:

1. Użycie wielosekcyjnych bądź sterowanych fazowo laserów DFB [30, 34], które mają płaską charakterystykę FM dla określonego zakresu polaryzacji prądem stałym.
2. Wykorzystanie adaptacyjnej korekcji charakterystyki FM lasera niezależnej od przebiegu charakterystyki FM danego lasera [33].

3. Kodowanie sygnału modulującego polegające na przesunięciu widma sygnału binarnego w zakres równomiernej charakterystyki FM lasera. Wykorzystywane tutaj jest to, że częstotliwość zagłębienia na charakterystyce FM nie przekracza zwykle 10 MHz, a dla wyższych częstotliwości charakterystyka jest w przybliżeniu równomierna aż do końca pasma.

Kod odpowiedni do celów modulacji lasera powinien spełniać następujące warunki:

a) brak składowej stałej,

b) maksymalnie skupione widmo,

c) dwupoziomowy alfabet wyjściowy (tzn. kod binarny).

Szczegółowa analiza widmowa dla szerokiej grupy kodów transmisyjnych została dokonana w pracach [16, 17]. Z porównania wynika, że najkorzystniejsze właściwości ma kod oznaczony jako AMI III. Należy on do grupy tzw. dwupoziomowych kodów AMI; symbol 0 kodowany jest jako 10 po 11 i po 01, a jako 01 po 00 i po 10, zaś symbol 1 kodowany jest na przemian jako 11 lub 00. Kod ten charakteryzuje się brakiem skladowej stałej, najwęższym zajmowanym pasmem częstotliwości i to pomimo tego, że częstotliwość zegara ciagu kodowego jest dwukrotnie większa niż ciągu kodowanego. Cechuje go mała zależność widma od prawdopodobieństwa q wystąpienia symbolu zero w ciągu kodowanym i dobre własności synchronizacyjne [16]. Ponadto w odróżnieniu od klasycznych kodów AMI, które są pseudotenarne, jest to kod binarny i nie ma niekorzystnych właściwości demodulacyjnych kodów ternarnych.

4.4. INNE EFEKTY

Na omówione powyżej efekty nakładają się jeszcze inne zjawiska, z których najważniejsze to szum mocy promieniowania (RIN — *radiation intensity noise*) oraz szum wywołany przeskakiwaniem między modami (mode hopping noise). Szum mocy promieniowania jest wywołany głównie przez emisję spontaniczną [21, 22] i znacznie rośnie pod wpływem odbić wstecznych dla pewnego zakresu współczynnika odbicia [20]. Ma on pomijalne znaczenie dla sygnału nadawanego, natomiast jest istotny w przypadku odbiornika, gdzie słaby sygnał odbierany może zostać silnie zakłócony przez fluktuacje mocy heterodyny optycznej. Dla eliminacji tego zjawiska stosuje się tzw. odbiorniki zrównoważone z dwoma detektorami [23], w których szumy RIN pochodzące od lasera lokalnego są od siebie odejmowane w dwóch gałęziach odbiornika. Szum przeskakiwania między modami można wyeliminować poprzez zapewnienie stabilnej pracy lasera w obrębie jednego modu.

5. WNIOSKI KOŃCOWE

Udało się zrealizować działający system koherentnej transmisji światłowodowej. Struktura systemu była zgodna z zamierzoną. Dokonano transmisji sygnału świetlnego o modulacji FM otrzymanej za pomocą modulacji prądu lasera nadawczego

o niewielkiej głębokości, jak również jego detekcji przy pomocy odbiornika z heterodyną optyczną i jednym filtrem. Zasadnicze cele projektu zostały więc osiągnięte. Niestety każdorazowo po zestrojeniu systemu, tzn. sprowadzeniu częstotliwości różnicowej laserów do żądanej częstotliwości leżącej w zakresie pasma elektrycznego układu, system pracował poprawnie jedynie kilka – kilkanaście sekund. Wskutek dryftów temperaturowych częstotliwość różnicowa podlegała wahaniom, wychodząc w końcu poza pasmo użyteczne układu. Niemożliwe było wykorzystanie do stabilizacji częstotliwości zaprojektowanej pętli sprzężenia zwrotnego. Było to spowodowane tym, że użyte lasery półprzewodnikowe miały widmo wielorotnie szersze niż podawały to dane katalogowe. Ewentualna lepsza stabilizacja temperatury nie rozwiązywałaby w dalszym ciągu problemu tak znacznego rozszerzenia linii widmowych. Wiadomo bowiem, że w przypadku transmisji koherentnej szerokość linii widmowych laserów nie może przekraczać szybkości transmisji, a w większości przypadków powinna być od niej znacznie mniejsza.

Zasadniczą przyczyną rozszerzenia linii widmowych laserów były, zdaniem autora, odbicia wsteczne promieniowania do laserów. W żadnym systemie transmisji światłowodowej nie da się ich uniknąć. W związku z tym powstaje konieczność zastosowania izolatorów optycznych, przy czym zazwyczaj potrzebne są dwa izolatory na jeden laser. Niestety zakup kilku sztuk izolatorów optycznych, bądź diod laserowych z wbudowanymi izolatorami (cena kilka tys. $ za sztukę) przekraczał możliwości finansowe pracy.

Konieczność stosowania izolatorów powoduje to, że zastosowanie systemów transmisji koherentnej w pierwszym oknie transmisyjnym mija się z celem ze względu na znaczne koszty i niewielką poprawę czułości, w porównaniu z detekcją bezpośrednią. Należy podkreślić przydatność modulacji FM lasera za pomocą modulacji jego prądu, gdyż możliwe jest jej zastosowanie również w systemach niekoherentnych wykorzystujących filtry optyczne i detekcję bezpośrednią.

BIBLIOGRAFIA

1. R.W. Tkach, A.R. Chraplyvy: *Regimes of feedback effects in 1.5 μm distributed feedback lasers.* J. Lightwave Tech., vol. LT-4, no. 11, pp. 1655–1661, November 1986
2. D. Lenstra, B.H. Verbeek, A.J. den Boef: *Coherence collapse in single mode semiconductor lasers due to optical feedback.* IEEE J. Quantum Electron. vol. QE-21, pp. 67–679, 1985
3. G.P. Agrawal: *Line narrowing in a single mode injection laser due to external optical feedback.* IEEE J. Quantum Electron., vol. QE-20, pp. 468–471, 1984
4. C. Henry, R.F. Kazarinov: *Instability of semiconductor lasers due to optical feedback from distant reflectors.* IEEE J. Quantum Electron., vol. QE-22, pp. 295–301, 1986
5. Seastar Optics Inc.: Optical devices & laser diode instrumentation product catalog, Canada 1993
6. I. Garrett, G. Jacobsen: *Theoretical analysis of heterodyne optical receivers for transmission systems using (semiconductor) lasers with non negligible linewidth.* J. Lighwave Technol., vol. LT-4, no. 3, pp. 323–334, 1986
7. R.S. Vodhanel, B. Enning, A.F. Elrefaie: *Bipolar optical FSK transmission experiments at 150 Mbit/s and 1 Gbit/s,* J. Lightwave Techn., vol. LT-6, no. 6, pp. 1549–1553, October 1988

8. S. Saito, Y. Yamamoto, T. Kimura: *Optical FSK heterodyne detection experiments using semiconductor laser transmitter and local oscillator.* IEEE J. Quantum Electron., vol. QE-17, no. 6, pp. 935–941, June 1981
9. W. van Etten, J. van der Plaats: *Fundamentals of optical fiber communications.* Prentice Hall, 1991
10. T. Okoshi: *Polarization-state control schemes for heterodyne or homodyne optical fiber communications.* J. Lightwave Techn., vol. LT-3, no. 6, December 1985, pp. 1232–1236
11. H.C. Lefevre: *Single mode fibre fractional wave devices and polarisation controllers.* Electron. Lett., vol. 16, no. 20, pp. 778–780, 1980
12. G.N. Ramachandra, G. Ramaseshan: *Encyclopedia of Physics.* Springer Verlag, Berlin 1962
13. R. Ulrich, A. Simon: *Polarization optics of twisted single mode fibers.* Appl. Opt. vol. 18, pp. 2241–2251, 1979
14. T. Okoshi, K. Kikuchi: *Heterodyne-type optical fiber communiations.* J. Optical Commun., vol. 2, no. 3, pp. 82–88, 1981
15. G. Nicholson: *Transmission performance of an optical FSK heterodyne system with a single filtre envelope detection receiver.* J. Lightwave Technol., vol. 5, no. 4, pp. 502–509, 1987
16. M. Rydel: *Charakterystyki kodów transmisyjnych dla światłowodowych systemów telekomunikacyjnych.* Rozprawy elektrotechniczne, vol. 26, no. 3, str. 747–766, 1980
17. M. Rydel, M. Słomiński: *An algorythmic method of analysis of coded signals.* Rozprawy elektrotechniczne, vol. 25, no. 3, str. 673–701, 1979
18. G. Georgiou, A.C. Boucouvalas: *Low-loss single-mode optical couplers.* IEE Proc. vol. Pt. J., 132, no. 5, pp. 297–302, 1985
19. T. Bricheno, A. Fielding: *Stable low-loss single-mode couplers.* Electron. Lett., vol. 20, no. 6, pp. 230–232, 1984
20. *Sharp LT022PD laser diode technical data*
21. Y. Yamamoto: *AM and FM quantum noise in semiconductor lasers — Part 1: theoretical analysis.* IEEE J. Quantum Electron., vol. QE-19, no. 1, pp. 34–46, 1983
22. Y. Yamamoto, S. Saito, T. Mukai: *AM and FM quantum noise in semiconductor lasers — part 2: comparison of theoretical and experimental results for AlGaAs lasers.* IEEE J. Quantum Electron., vol. QE-19, no. 1, pp. 47–58, 1983
23. G.L. Abbas, W.W.S. Chan, T.K. Yee: *A dual-detector optical heterodyne receiver for local oscillator noise suppression.* J. Lightwave Techn., vol. LT-3, no. 5, pp. 1110–1122, 1985
24. J. Siuzdak: *Optical couplers for coherent optical phase diversity systems.* EUT Report, Eindhoven 1988, ISBN 90-6144-190-0
25. R. Steele, S. Wright: *Electronic control of the difference frequency between two semiconductor lasers.* Ann. Telecommun., vol. 38, no. 1–2, pp. 53–57, 1983
26. J. Siuzdak: *Wybrane problemy koherentnej komunikacji światłowodowej.* KST 89, E-14, str. 112–116, Bydgoszcz 1989
27. Burleigh Instruments Inc.: *Laser diagnostic systems.* 1990
28. J. Siuzdak: *Zestaw przyrządów do badania charakterystyk widmowych laserów półprzewodnikowych dla koherentnej komunikacji światłowodowej.* KST 93, Bydgoszcz 1993, D-8.05 str. 261–267
29. J. Siuzdak, W. van Etten: *BER performance evaluation for CPFSK phase and polarisation diversity coherent optical receivers.* J. Lightwave Technol. vol. 9, no. 11, pp. 1583–1592, November 1991
30. T. Imaiin.: *Over 300 km CPFSK transmission experiment using 67 photon/bit sensitivity receiver at 2.5 Gbit/s.* Electronics Letters, pp. 357–358, 1990
31. K. Emuraiin.: *Requirements for high-frequency LD FM response in an optical CPFSK heterodyne delay demodulation system.* J. Lightwave Technol., vol. 9, no. 11, pp. 1602–1608, 1991
32. G. Jacobseni in.: *Requirements for LD FM characteristics in an optical CPFSK system.* J. Lightwave Technol., vol. 9, no. 9, pp. 1113–1123, 1991
33. B. Enning, G. Lange: *Adaptive quantized feedback equalization of laser FM-response for optical FSK transmission.* IEEE Photonics Technol. Letters, vol. 2, no. 11, pp. 838–839, 1990

34. D. F e l i c i o: *140 Mbit/s optical FSK heterodyne dual filter detection system with standard DFB lasers.* J. Optical Commun., vol. 11, no. 3, pp. 88 – 91, 1990

J. SIUZDAK

AN EXPERIMENTAL FIBER OPTICS COHERENT TRANSMISSION SYSTEM THAT USES LASERS FM

S u m m a r y

A model of the coherent optical fiber transmission system is described. The system uses monomode semiconductor lasers with frequency modulation and has been developed in the Telecommunications Institute of the Warsaw Technical University. It consists of lasers current and temperture stabilisation circuitry, a modulator, a receiver and demodulator, a fiber coupler and a polarisation controler. A light signal with frequency modulation has been transmitted in this system. FM has been obtained by low index AM modulation of the transmitting laser current. The coherent demodulation has been obtained by mixing the incoming light with an optical heterodyne and further detection and electronic processing of the resulting signal. Unfortunately, due to the broad linewidths of the lasers stable operation of the system failed.

Key words: optical fibers, lasers F11, modulators.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 1. ss. 131–141

Współczesne rozumienie pojęcia „Technika Obrazowa" oraz jego relacje do innych pojęć z tym związanych

HANNA GÓRKIEWICZ-GALWAS, JERZY M. WOŹNICKI

Instytut Mikroelektroniki i Optoelektroniki, Politechnika Warszawska

Otrzymano 1994.12.14

Autoryzowano do druku 1995.02.15

Artykuł poświęcony jest ogólnej charakterystyce techniki obrazowej jako wyodrębnionemu obszarowi wiedzy oraz uporządkowaniu wybranych pojęć z nim związanych. Przedstawia on także zarys tendencji rozwojowych oraz przegląd wybranych prac z tego obszaru prowadzonych w kraju.

Słowa kluczowe: informacje obrazowe, optoelektronika obrazowa, przetworniki obrazu.

1. WPROWADZENIE

Technika obrazowa, związana z akwizycją, przetwarzaniem, transmisją i wyświetlaniem informacji obrazowych [1], znajduje szerokie zastosowania we współczesnym świecie. Stosując techniki analogowe i cyfrowe dziedzina ta wykorzystuje osiągnięcia hardware'owe w zakresie układów przetwarzania informacji przestrzennej (głównie optoelektronicznych przetworników obrazu) oraz software'owe — dotyczące konstruowania algorytmów i procedur służących do przetwarzania i analizy oraz rozpoznawania i animacji obrazów.

Ogromny rozwój techniki obrazowej, charakteryzujący się wielkim postępem technologicznym, dokonał się w dekadzie lat 80-tych. W wyniku integrowania się różnych, nie związanych wcześniej ze sobą technik i technologii przetwarzania obrazu, pojawiły się nowe rozwiązania techniczne w dziedzinie optoelektroniki obrazowej. W ślad za tym nastąpił rozwój nowych gałęzi przemysłu i towarzyszący ich wyrobom sukces handlowy.

Lata 90-te przyniosły dalszy rozwój techniki obrazowej — szczególnie systemów wysokiej rozdzielczości — i to zarówno w odniesieniu do obrazów w zakresie

widzialnym widma promieniowania elektromagnetycznego jak i wizualizowanych obrazów z zakresów podczerwieni i ultrafioletu. Postęp dotyczy także obrazów o ekstremalnie krótkich czasach trwania i obrazów szybko-zmiennych. Wykorzystanie technik cyfrowego przetwarznia informacji pozwoliło na polepszenie jakości w procesach przetwarzania i rozpoznawania obrazu.

W zakres zainteresowań techniki obrazowej wchodzą zarówno obraz i jego nośnik jak i cały tor przenoszenia, analizy i syntezy obrazu, poczynając od detekcji, a kończąc na wyświetlaniu. Rozwijane dynamicznie metody analizy treści obrazu poszerzają stale zakres zastosowań techniki obrazowej.

Interdyscyplinarny charakter techniki obrazowej sprawia, że wiąże ona ze sobą różne dziedziny wiedzy, wykorzystując prace specjalistów reprezentujacych wiele dyscyplin naukowych. Rozwiązywanie problemów związanych z przetwarzaniem i analizą obrazów jest wobec tego często zawężane do zakresu tematycznego związanego ze specyfiką danej specjalności. W tej sytuacji celowe jest ogólne scharakteryzowanie obszaru wiedzy określanego jako technika obrazowa, uporządkowanie pojęć z nim związanych oraz zarysowanie powiązań między nimi.

Zamiarem autorów artykułu jest przedstawienie współczesnego rozumienia pojęcia technika obrazowa na tle jej tendencji rozwojowych oraz dokonanie przeglądu prac prowadzonych w Polsce w najbardziej liczących się w tej dziedzinie ośrodkach.

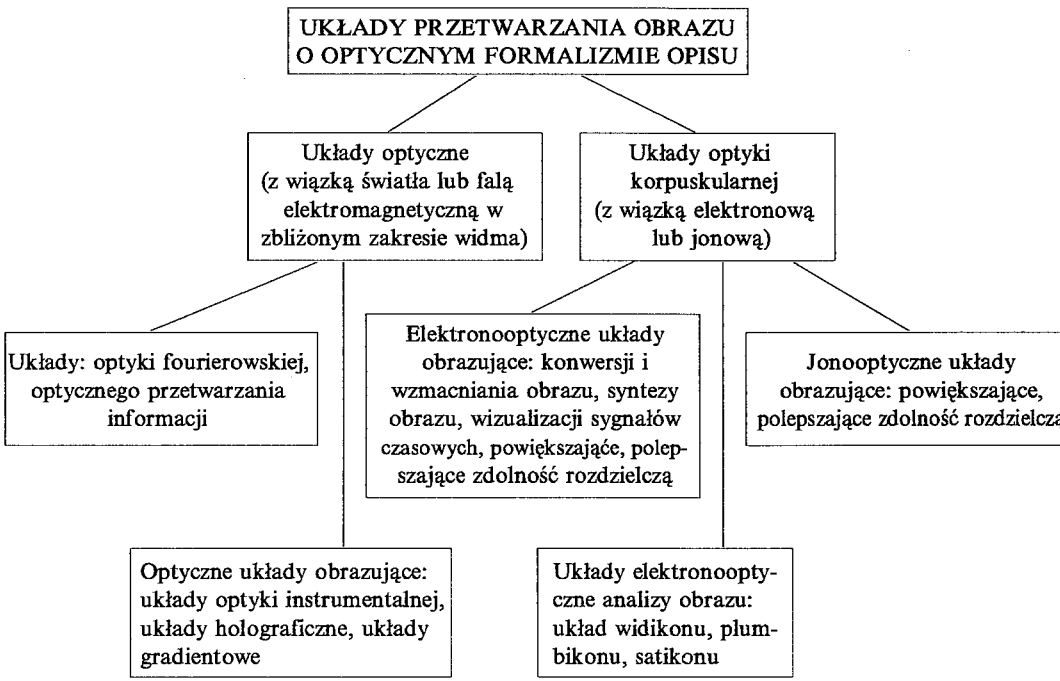
2. TECHNIKA OBRAZOWA NA TLE INNYCH ZWIĄZANYCH Z NIĄ OBSZARÓW TECHNIKI

Technika obrazowa, bazująca na osiągnięciach z pogranicza optyki, elektroniki i informatyki, wykorzystuje uogólnione pojęcie obrazu jako przestrzennego rozkładu informacji, niezależnie od fizycznej natury jej nośnika. Nośnikiem informacji obrazowych może być nie tylko fala elektromagnetyczna, ale także wiązka cząstek naładowanych, potencjał elektrostatyczny, potencjał chemiczny itp. Obraz jest natomiast reprezentowany przez dwuwymiarową funkcję zmiennych przestrzennych opisujących powierzchniowe współrzędne poszczególnych punktów obrazu. Wartości funkcji są określone przez poziomy — rozumianej w sensie uogólnionym — intensywności nośnika obrazu.

W technice obrazowej najistotniejszą rolę odgrywają nośniki obrazu w postaci wiązki fal elektromagnetycznych oraz wiązka elektronowa lub jonowa, umożliwiające tzw. optyczne przetwarzanie informacji. Do ich opisu i analizy stosowane są metody związane z obserwacją i wykorzystaniem światła. Wygodnie jest więc określać je wspólną nazwą nośników informacji przestrzennej o optycznym formalizmie opisu [2] i konsekwentnie, bazujące na ich wykorzystaniu układy przetwarzania obrazu — układami o optycznym formaliźmie opisu.

W technice współczesnej występuje zróżnicowane rozumienie pojęcia światła — węższe, ograniczające się do zakresu widzialnego promieniowania elektromagnetycznego oraz szersze, odwołujące się do szerszego lub znacznie szerszego zakresu

widma. Autorzy niniejszego artykułu używają pojęcia światła w węższym, tradycyjnym sensie jako zakresu widzialnego, akceptując jednocześnie pogląd, że tzw. optyczny zakres promieniowania elektromagnetycznego obejmuje przedział długości fali od 0,01 μm do 1 mm. Stwierdzenie to odwołuje się do przyjęcia umownego rozróżnienia optyki i elektroniki, jako dyscyplin naukowych, wg kryterium zdolności do rejestracji i pomiaru fazy fali elektromagnetycznej.

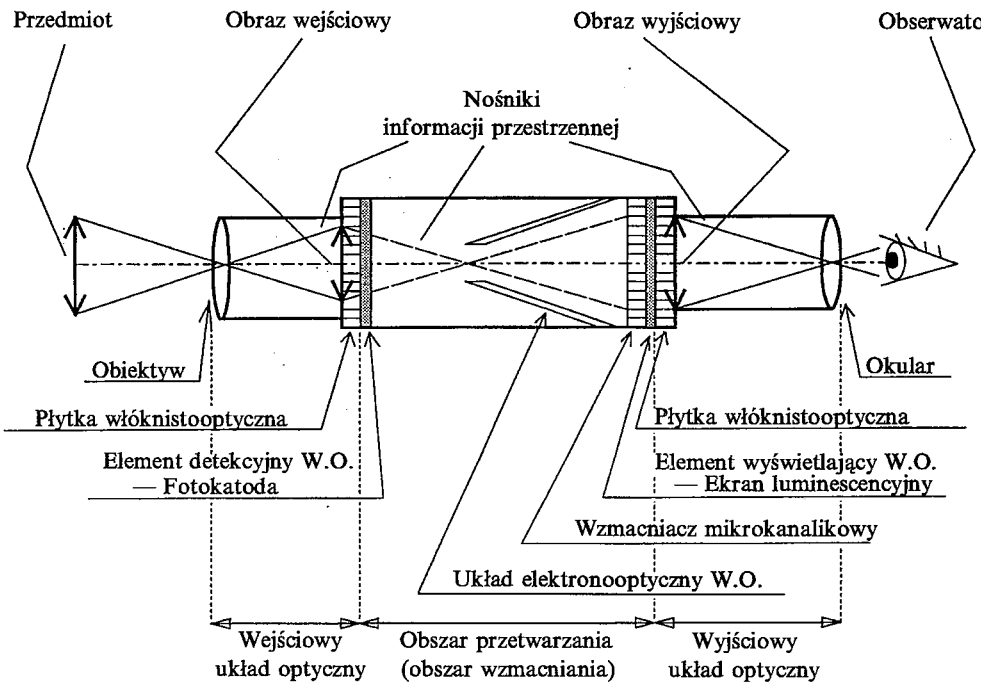


Rys. 1. Podstawowe rodzaje układów przetwarzania obrazu o optycznym formaliźmie opisu

Na rys. 1 dokonano klasyfikacji układów przetwarzania obrazu o optycznym formaliźmie opisu. Układy optyki światła obejmują zarówno klasyczne niekoherentne układy obrazujące, jak i koherentne i niekoherentne układy optycznego przetwarzania informacji przestrzennej [3]. Optyczne uklady obrazujące reprezentują zarówno układy konwencjonalne optyki instrumentalnej, układy holograficzne, jak i układy gradientowe napotykające jednak ciągle na problemy technologiczne. Wśród układów przetwarzania obrazu najistotniejsze znaczenie mają koherentne i niekoherentne procesory optyczne. Stosowana w nich przestrzenna modulacja wiązki/fali pozwala analizować obraz w oparciu o klasyczne w optyce metody analizy fourierowskiej. Uzupełnienie metod przestrzennych (rozumianych jako bezpośrednie działanie na obrazie) przez szeroko stosowane w optyce i elektronice metody częstotliwościowe (działanie na widmie częstości przestrzennych) jest podstawą opracowywania wielu algorytmów i procedur przetwarzania obrazu, służących do poprawy jakości, kodowania, kompresji i rekonstrukcji obrazu, a także wyszukiwania elementów obrazu i rozpoznawania ich cech.

W grupie układów optyki korpuskularnej dominują elektronooptyczne układy obrazujące w tym układy konwersji, np. układ wzmacniacza obrazu, układy syntezy obrazu, np. układ kineskopu i układy wizualizacji sygnałów czasowych, np. układ lampy oscyloskopowej. Obrazujące układy o zdolnościach powiększających i poprawiających zdolność rozdzielczą realizowane są w formie układów elektronooptycznych lub jednooptycznych np. w mikroskopach elektronowych lub jonowych. Układy elektronooptyczne analizy obrazu związane są z rodziną tzw. lamp analizujących. Ich przedstawicielami są, ciągle stosowane m.in. w technice studyjnej, układy superortikonu, widikonu, plumbikonu, satikonu i inne.

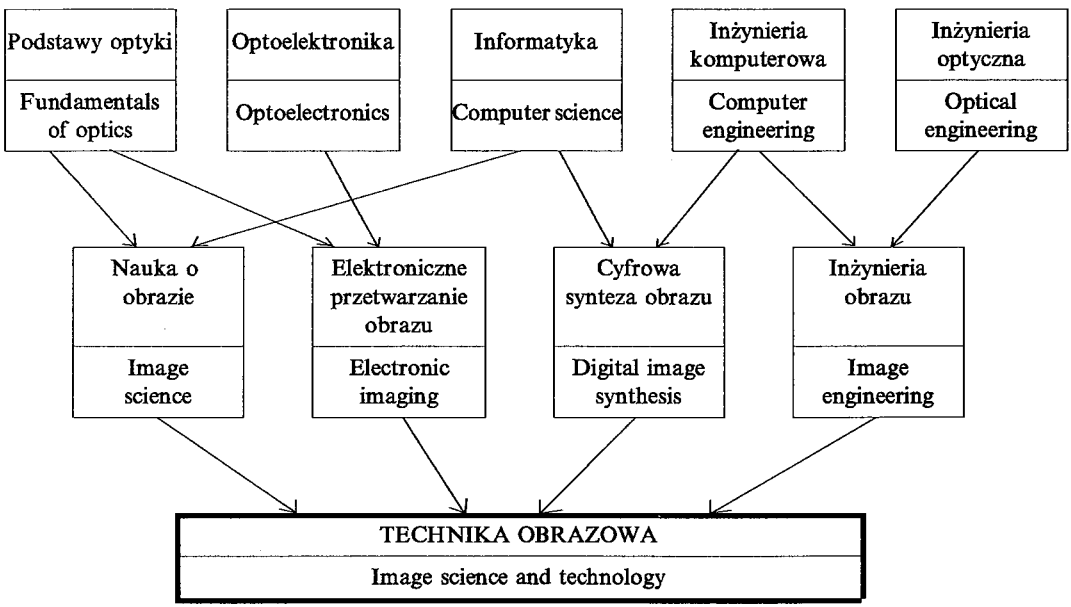
Efekt przetwarzania obrazu uwarunkowany jest, niezależnie od rodzaju nośnika inforamcji obrazowej, właściwościami całego systemu przetwarzania obrazu [4]. W jego skład mogą wchodzić układy optyczne przetwarzające i obrazujące oraz układy optyki korpuskularnej wzbogacone o układy elektroniczne sterowania. Przykładowy analogowy system przetwarzania obrazu przedstawiono na rys. 2. Jest



Rys. 2. Przykład analogowego systemu przetwarzania obrazu ze wzmacniaczem obrazu (W.O.)

to system bezpośredniej obserwacji sceny w warunkach braku możliwości postrzegania wzrokowego. Dotyczy to przypadków niedopasowania widmowego detektora do nośnika obrazu, niedostatecznej intensywności promieniowania lub zbyt krótkiego czasu ekspozycji. W torze przetwarzania obrazu znajdują się wzmacniacz obrazu oraz wejściowy i wyjściowy układy optyczne. Prezentowany system może stanowić wejściowy, analogowy stopień wprowadzający informację przestrzenną do

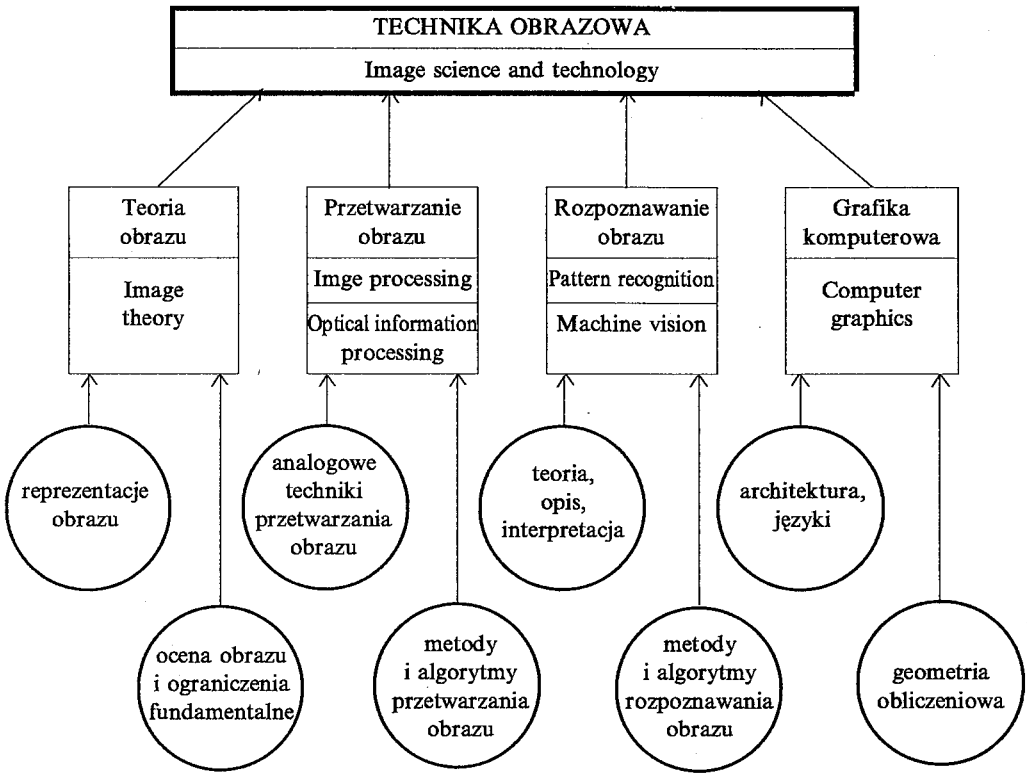
toru cyfrowego przetwarzania obrazu np. jako układ wejściowy kamery CCD i przetwornika AC dyskretyzujących obraz. Polepszenie parametrów takiego systemu uzyskuje się dzięki wprowadzeniu rozwiązań technologicznych takich jak mikrokanalikowe powielacze elektronów, płytki włóknistooptyczne itp., a także konstrukcyjnych dzięki optymalizacji konstrukcji przy wykorzystaniu technik CAD. W ostatnich kilkunastu latach obserwuje się w tej dziedzinie znaczące osiągnięcia związane z rozwojem układów komputerowego wspomagania projektowania optycznego (CAOD) oraz metod modelowania matematycznego zjawisk fizycznych warunkujących działanie elektronooptycznych układów przetwarzania obrazu (z uwzględnieniem zmian parametrów geometrycznych i materiałowych ich elementów).



Rys. 3. Powiązania techniki obrazowej z innymi obszarami nauki i techniki

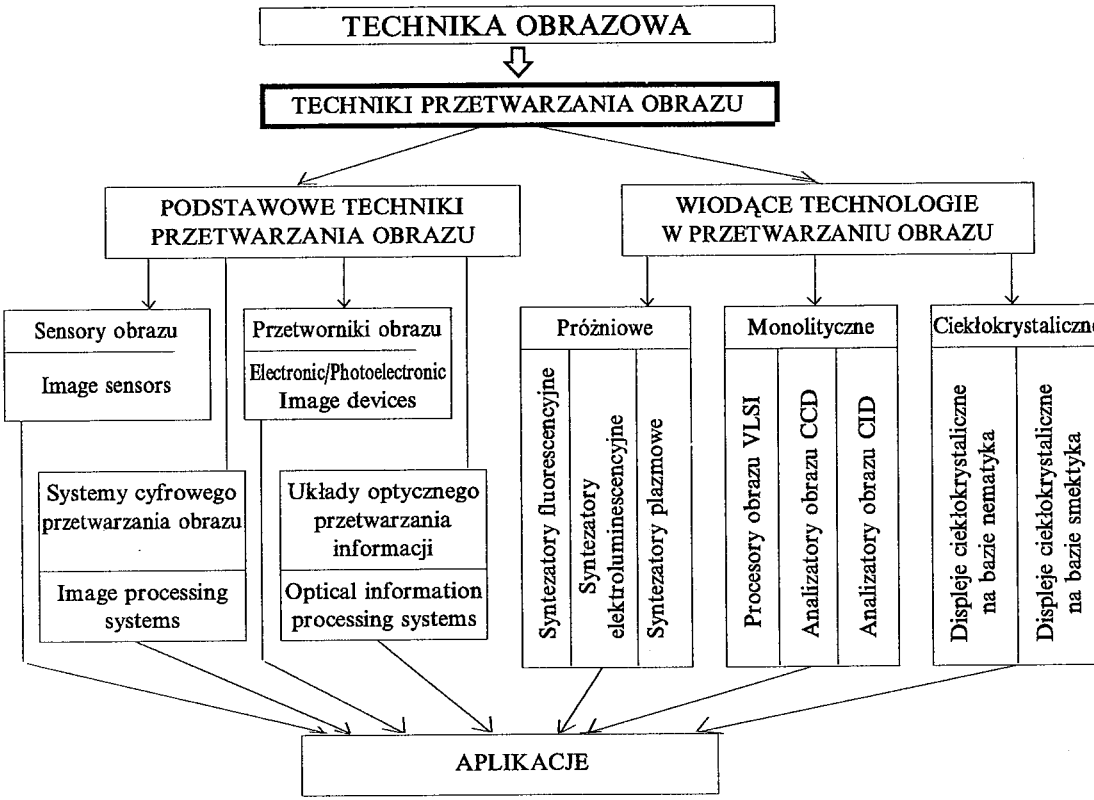
Na rys. 3 pokazano schematycznie powiązania dziedziny „technika obrazowa" z innymi dziedzinami nauki i techniki. Są to: optyka jako dział fizyki, optoelektronika oraz informatyka a także inżynieria optyczna i inżynieria komputerowa. Technika obrazowa wykorzystuje osiągnięcia z zakresu nauki o obrazie [5, 6] oraz jego elektronicznym przetwarzaniu, a także cyfrowej syntezy obrazu polegającej na generacji scen na podstawie danych innych niż dane obrazowe [7, 8, 9, 10] oraz inżynierii obrazu [11] zajmującej się przetwarzaniem istniejącego obrazu i rozpoznawaniem jego cech technikami cyfrowymi i analogowymi. Rysunek charakteryzuje więc wkład klasycznych dyscyplin w rozwój techniki obrazowej.

Na rysunku 4 w każdym z wymienionych działów techniki obrazowej wydzielono różne podobszary związane z podstawami fizycznymi i technicznymi uwarunkowaniami jakości układów, matematycznymi metodami zapisu i analizy danych obrazowych oraz ich interpretacji.



Rys. 4. Podstawowe podobszary techniki obrazowej

Rysunek 5 określa pola zastosowań aktualnie stosowanych technik przetwarzania obrazu. Są to sensory obrazu wykorzystywane w procesach detekcji i akwizycji obrazu, przetworniki obrazu z grupy analizatorów, syntezatorów i konweterów (w tym wzmacniaczy) obrazu oraz systemy cyfrowego przetwarzania obrazu i układy optycznego przetwarzania informacji obrazowych. Rozwój zaawansowanych technologii pozwolił na opracowanie procesorów obrazu VLSI, monolitycznych analizatorów obrazu typu CCD oraz CID w wersji matrycowej oraz syntezatorów obrazu nowej generacji w technice próżniowej (displeje plazmowe, fluorescencyjne, elektroluminescencyjne) i ciekłokrystalicznej (na bazie nematyków, superskręconych nematyków oraz smektyków ferroelektrycznych). Rozszerzony zakres związanych z nimi zastosowań obejmuje systemy medyczne przetwarzania obrazu, miernictwo optoelektroniczne oraz telewizję cyfrową.



Rys. 5. Przykłady zastosowań technik przetwarzania obrazu

3. KIERUNKI ROZWOJOWE TECHNIKI OBRAZOWEJ

Rozwój techniki obrazowej postępuje od szeregu lat w dwóch głownych kierunkach — układów przetwarzania obrazu oraz metod numerycznych jego przetwarzania i analizy. Tendencje rozwojowe, przedstawione na rys. 5, wskazują na znaczenie uwarunkowań technologicznych [12] dla nowych opracowań hardware'owych jakich można oczekiwać do końca lat 90-tych. W zakresie techniki konwersji obrazu dominować będą próżniowe przetworniki obrazu trzeciej generacji z wysokowydajnymi fotokatodami, współpracujące z powielaczami mikrokanalikowymi. Rozwój systemów wizualizujących uzależniony będzie od pojawienia się nowych generacji kamer smugowych i kadrowych wykorzystujących elektronooptyczne przetworniki obrazu. W grupie urządzeń powszechnego użytku analizatory obrazu już zostały zdominowane przez technologie monolityczne typu CCD i CID. Następował będzie jednak dalszy postęp w zakresie rozdzielczości tych przyrządów.

W zakresie zastosowań w badaniach naukowych i astronomii postępuje rozwój przetworników umożliwiających identyfikację mikroobiektów, działających w zakresie promieniowania rentgenowskiego, ultrafioletu, podczerwieni i promieniowania widzialnego. Ta technika, ciagle opanowana przez rozwiązania elektronooptyczne z płytkami mikrokanalikowymi, ma szanse w coraz większym zakresie wykorzystywać także monolityczne mozaiki CCD o odpowiednim formacie i dostatecznie małych wymiarach pojedynczych pixeli. Wskazują na to osiągnięcia technologiczne dotyczące analizatorów obrazowych dla zakresu widzialnego. Przykładem jest analizator o rozdzielczości 4096 × 4096 pixeli, z wymiarami pojedynczego pixela 7,5 μm × 7,5 μm. Podobne rozwiązania mogą stanowić podstawę nie tylko do rozwoju kamer i systemów bezpośredniej obserwacji, ale także kamer pracujących w warunkach niedostatecznego oświetlenia i przy niedopasowaniu widmowym. W warunkach ekstremalnie niskich poziomów oświetlenia (10^{-8}—10^{-6} lx) stosowane są jednak i w przewidywalnej perspektywie stosowane będą jedynie przetworniki elektronooptyczne pracujące w modzie zliczania pojedynczych fotonów.

Wśród syntezatorów obrazów wiodącą pozycję w najbliższych latach utrzyma nadal ciągle doskonalony kineskop elektronowiązkowy. Intensywne prace związane z rozwojem techniki mikrokomputerowej oraz z planami wprowadzania systemu telewizji HDTV są jednak silnym motorem postępu w dziedzinie syntezatorów obrazu nowych generacji: ciekłokrystalicznych, plazmowych, fluorescencyjnych, elektroluminescencyjnych i matryc zwierciadeł adaptacyjnych. Zastosowanie systemów projekcyjnych do syntezy obrazów HDTV daje szczególne szanse na nowe operacowania displeji ciekłokrystalicznych.

Cyfrowe przetwarzanie obrazu stanowi osobny, rozległy dział techniki obrazowej. Charakteryzuje się ogromną dynamiką rozwojową, co jest efektem postępu technologicznego w sprzęcie komputerowym w ostatnich kilku latach. Nowoczesne karty przetwarzania obrazu typu Frame Grabber i procesory obrazu, wykorzystujące rozwinięte biblioteki standardowych funkcji i procedur, pozwalają na bardzo szybkie hardware'owe wykonywanie złożonych operacji matematycznych na obrazie, co umożliwia realizację systemów przetwarzania obrazu czasu rzeczywistego.

Główne kierunki prac w zakresie cyfrowego przetwarzania obrazu dotyczą widzenia maszynowego (machine and computer vision) [13], analizy i rozpoznawania obrazu (pattern analysis and recognition) oraz grafiki komputerowej (computer graphics) [14]. Badania prowadzone są w zakresie [15]:

— przetwarzania obrazu w tym poprawy jego jakości i wydzielania cech
— rozpoznawania obiektów i analizy scen
— rekonstrukcji kształtów obiektów trójwymiarowych
— modelowania geometrycznego i animacji komputerowej.

Prace rozwojowe mają za zadanie kojarzenie technik analogowych i cyfrowych w celu opracowania:

— systemów widzenia maszynowego
— układów symulowania rzeczywistości wirtualnej (pozornej)
— zastosowań sieci neuronowych i logiki rozproszonej do identyfikacji, klasyfikacji i rozpoznawania obrazów, a także sterowania i optymalizacji procesów w oparciu o dane obrazowe.

Na bazie tych opracowań możliwe będzie dalsze poszerzanie zakresu aplikacji cyfrowego przetwarzania obrazu w medycynie, systemach wizyjnych, kartografii, miernictwie optoelektronicznym, telekomunikacji, systemach naprowadzania na cel i sterowania ruchem oraz w układach oceny i kontroli jakości materiałów i procesów w wielu zastosowaniach przemysłowych.

4. WYBRANE KIERUNKI PRAC PROWADZONYCH W KRAJU

W badaniach i opracowaniach w dziedzinie techniki obrazowej w Polsce przeważają w ostatnich latach prace o charakterze software'owym. Uczestniczą w nich zwłaszcza Instytuty Polskiej Akademii Nauk: Podstaw Informatyki, Podstawowych Problemów Techniki, Biocybernetyki i Inżynierii Biomedycznej oraz Instytuty wyższych uczelni technicznych — w Politechnice Warszawskiej są to Instytuty: Informatyki, Mikroelektroniki i Optoelektroniki, Radioelektroniki, Konstrukcji Przyrządów Precyzyjnych i Optycznych, Sterowania i Elektroniki Przemysłowej; Instytut Informatyki Uniwersytetu Jagiellońskiego, Instytut Elektroniki Politechniki Łódzkiej i in. [16, 17]. Prowadzone prace związane są m.in. z opracowywaniem metod poprawy jakości obrazów cyfrowych oraz kompresji i segmentacji danych obrazowych. Przykładami mogą być metody korekcji obrazów telewizyjnych i obrazow prążkowych polegające na zastosowaniu filtrów medianowego, Markowa i membranowego. Znaczące są osiągnięcia w zakresie kompresji obrazów telewizyjnych, medycznych i prążkowych z wykorzystaniem transformacji wavelet i kosinusowej oraz technik fraktalnych [18]. W zakresie metod analizy treści obrazu intensywnie rozwijane są metody fazowe. Opracowywanie nowych metod fazowych wiąże się z poszukiwaniem możliwości prowadzenia analizy obrazu w czasie rzeczywistym. Podejmowane są próby wykorzystania do tych celów technik sieci neuronowych komórkowych i logiki rozproszonej. Na bazie tych opracowań tworzone są nowoczense systemy analizy obrazów. Dotyczą one w szczególności obrazów prążkowych [19], które pozwalają na dokładną i jednoznaczną interpretację inforamcji zawartych w obrazie, a zatem nadają się do wykorzystania w ukladach kontrolnych i pomiarowych. Powstające systemy ekspertowe mogą mieć zastosowanie w automatyce i robotyce, telekomunikacji, diagnostyce technicznej i medycznej i in.

Oddzielny obszar prowadzonych prac rozwojowych dotyczy grafiki komputerowej [20]. Opracowywane algorytmy modelowania obrazów cyfrowych pozwalają na generowanie obrazów realistycznych, możliwie wiernie oddających rzeczywistość, z uwzględnieniem zarówno barw jak i ruchu w elementach obrazu. Stwarza to szanse na wykorystywanie grafiki komputerowej do wizualizacji danych, symulacji zjawisk naturalnych, do prezentowania sztucznej rzeczywistości itp.

Prowadzone w nielicznych ośrodkach krajowych prace hardware'owe związane są w znacznym stopniu z nowoczesnymi technologiami. Poważne osiągnięcia w technologiach monolitycznych dotyczą w szczególności detektorów podczerwieni. W wyniku optymalizacji profilu i składu domieszkowania materiałów oraz nierównomiernego

zubożania w nośniki ładunku stosowanych struktur półprzewodnikowych, w Instytucie Fizyki Technicznej Wojskowej Akademii Technicznej (WAT) opracowano detektory średniej i dalekiej podczerwieni o obniżonym poziomie szumów. W tym samym ośrodku technologie ciekłokrystaliczne rozwijane są pod kątem opracowania displejów video i graficznych do obserwacji bezpośredniej oraz zastosowania w projektorach obrazów barwnych dla telewizji HDTV. Badane są efekty dwójłomności strumienia świetlnego w substancjach ciekłokrystalicznych o układzie molekuł superskręconego nematyka, w smektycznych ferroelektrykach i in. [21, 22, 23]. Znaczne perspektywiczne możliwości tkwią w badaniach prowadzonych w Instytucie Techniki Próżni OBREP, związanych z rozwojem technologii MBE (epitaksja z wiązek molekularnych) oraz CVD (osadzanie chemiczne w próżni). Istnieją szanse wykorzystania wyników prac do wytwarzania płytek CCD analizatorow monolitycznych oraz do wytwarzania fotokatod dla potrzeb produkcji konwerterów obrazu i pewnych rodzajów sensorów.

Efekty prac software'owych i hardware'owych są wykorzystywane do tworzenia specjalistycznych systemów przetwarzania obrazu w tym m.in. kamer elektrooptycznych smugowych i kadrowych (Instytut Fizyki Plazmy i Laserowej Mikrosyntezy), systemów bezpośredniej obserwacji w warunkach niedostatecznego oświetlenia (Przemysłowe Centrum Optyki — PCO), systemów termowizyjnych (Instytut Optoelektroniki WAT), systemów o przeznaczniu militarnym (PCO, WAT, Polskie Zakłady Optyczne — PZO), systemów diagnostycznych przemysłowych i medycznych (Fabryka Aparatury Rentgenowskiej i Urządzeń Medycznych FARUM, Warszawskie Zakłady Telewizyjne ELEMIS, Instytut Technologii Elektronowej) i in. Ośrodki krajowe mają też pewne osiągnięcia w opracowaniu kamer elektrooptycznych do specjalistycznych zastosowań np. w robotyce oraz systemów do badania mikroobiektów przeznaczonych do wykorzystania w miernictwie optoelektronicznym.

PODSUMOWANIE

W artykule przedstawiono podstawowe pojęcia związane z techniką przetwarzania obrazu. Na tle współczesnego rozumienia „techniki obrazowej", jako nabierającego coraz większego znaczenia obszaru nauki i techniki, wskazano pewne związki pojęciowe i powiązania merytoryczne podkreślające interdyscyplinarny charakter tego obszaru wiedzy. Ilustrują to przeprowadzone w sposób konsekwentny klasyfikacje i schematy zamieszczone w artykule. Jest oczywiste, że jak wszystkie klasyfikacje, mają one charakter umowny i mogłyby być uzupełniane o nowe, inne zestawienia. Podstawą wszystkich klasyfikacji są bowiem kryteria mające charakter subiektywny.

W artykule zarysowano także aktualny stan prac w dziedzinie techniki obrazowej prowadzonych w kraju. W tym kontekście intencją autorów było zwrócenie uwagi, że obszar cyfrowego przetwarzania obrazu nie jest tożsamy z pojęciem techniki obrazowej, która obejmuje także problematykę układów analogowych i zagadnienia technologiczne.

BIBLIOGRAFIA

1. M. Ostrowski (koordynator): *Informacja obrazowa.* WNT, 1992
2. J.M. Woźnicki: *Analiza i projektowanie konwerterów i wzmacniaczy obrazów.* Wyd. P.W, Elektronika z. 82, 1988
3. R. Jóźwicki: *Teoria odwzorowania optycznego.* PWN, 1988
4. W.F. Schreiber: *Fundamentals of Electronic Imaging Systems, Some Aspects of Image Processing.* Springer-Verlag 1986
5. J.C. Dainty, R. Shaw: *Image Science, Principles, Analysis and Evaluation of Photographic — Type Imaging Processes.* Academic Press 1974
6. O.J. Braddich, A.C. Sleigh: *Physical and Biological Processing of Images.* Springer-Verlag, 1983
7. C.M. Brown, H. Ballard: *Computer vision,* Prentice Hall, 1962
8. L.P. Yaroslavsky: *Digital Picture Processing — an Introduction.* Springer Verlag, 1985
9. R.C. Gonzales, P. Winz: *Digital Image Processing.* Addison—Wesley, 1987
10. T. Pavlidis: *Grafika i przetwarzanie obrazów.* WNT, 1987
11. King-sun Fu, T.L. Kunii: *Picture Engineering.* Springer-Verlag, 1982
12. *Photoelectronic Image Devices 1991.* Institute of Physics, Conference Series No 121, 1991
13. E.R. Davis: *Machine Vision: Theory, Algorithme, Practicalities.* Academic Press, 1990
14. J.C. Russ: *The Image Processing Handbook.* Academic Press, 1992.
15. Proceedings of SPIE, vol.vol.: 1246, 1260—1990; 1470, 1472 — 1991; 1769 — 1992; 1904, 1908 — 1993
16. *Techniki Przetwarzania Obrazu.* II Sympozjum Naukowe, Materiały konferencyjne, Oficyna Wydawnicza P.W., 1993
17. *Elementy Cyfrowego Przetwarzania i Analizy Obrazów.* Praca pod redakcją A. Materki, PWN 1991
18. W. Skarbek: *Metody reprezentacji obrazów cyfrowych.* Akademicka Oficyna Wydawnicza PLJ, 1993
19. M. Kujawińska: *Fringe Analysis: Anything New?* International Trends in Optics, C. Dainty (ed.), Academic Press, 1994
20. *Machine Graphics and Vision.* International Journal, 1994, vol. 3, no. 1/2
21. J. Żmija, J. Zieliński, J. Parka, E. Nowinowski-Kruszelnicki: *Displeje ciekłokrystaliczne.* Fizyka — Technologia — Zastosowanie, PWN, 1993
22. *11th Conference on solid and liquid crystals — material science and applications.* Materiały konferencyjne, 23—27 Oct. 1994
23. *Two Day's Display Seminar.* Materiały konferencyjne, 21—22 Oct. 1994

H. GÓRKIEWICZ-GALWAS, J.M. WOŹNICKI

CURRENT UNDERSTANDING OF "IMAGE SCIENCE AND TECHNOLOGY" AND RELATED CONCEPTS

Summary

This paper presents general characteristics of image science and technology as a field of knowledge. The authors attempt to present and classify systematically basic concepts in this area and related fields. The paper gives also an outline of the development trends and a brief review of selected research projects in the domain of image science and technology in Poland.

Key words: image science, digital image processing.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 1.

SPIS TREŚCI

CONTENTS

Subscription for external subscribers:

The promotional subscription price in 1995 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 5 zł / 50 000 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., urzędy pocztowe oddawcze 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 prenumeraty rocznej krajowej i na zagranicę przez RUCH S.A. i Zakład Kolportażu Wydawnictwa SIGMA-NOT
do 20.XI. roku poprzedzającego

Termin przyjmowania prenumeraty rocznej krajowej przez Pocztę Polską
do 25.XI. 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 370015-1573-139-11
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 Warszawa, 00-958 Warszawa, ul. Towarowa 28
Konto PBK XIII Oddział Warszawa nr 370044-1195-139-11
Dostawa odbywa się przesyłką 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 Warszawskim RUCH na konto: PBK XIII Oddział Warszawa, nr 370044-1195-139-11.

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 1995	
przewidywana cena 1 egz. wyniesie	5,00 zł (50.000 zł)
prenumerata roczna w kraju	20,00 zł (200.000 zł)
prenumerata roczna za granicę	90 $ USA

POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

Indeks 363189
ISSN 0867-6747

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

TOM 41 — ZESZYT 2

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1995

Kwartalnik jest wydawany z funduszu Polskiej Akademii Nauk

i dofinansowany z instytucji:

- Komitet Badań Naukowych
- Politechnika Warszawska, Wydział Elektroniki i Technik Informacyjnych
- Politechnika Gdańska, Wydział Elektryczny
- Politechnika Szczecińska, Instytut Elektroniki i Informatyki
- Ministerstwo Łączności
- 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 41 — ZESZYT 2

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1995

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. 660 77 37; 628 89 81; 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. 11,75 Ark. druk. 9.5	Podpisano do druku we wrześniu 1995 r.
Papier offsetowy kl. III 70 g. B-1	Druk ukończono w październiku 1995 r.

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

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 2, ss. 145–155

Projektowanie cyfrowego filtru SOI o zadanej charakterystyce częstotliwościowej i równomiernie falistym przebiegu funkcji błędu

FELICJA WYSOCKA-SCHILLAK

Wyższa Szkoła Pedagogiczna, Bydgoszcz

Otrzymano 1995.04.12

Autoryzowano do druku 1995.08.16

W pracy przedstawiono metodę projektowania cyfrowego dolnoprzepustowego filtru SOI o zadanej charakterystyce częstotliwościowej. W metodzie tej rozpatrywane zadanie projektowania filtru jest przekształcane w prawie równoważne rzeczywiste zagadnienie uzyskania równomiernie falistej funkcji błędu. Zagadnienie to jest z kolei przeformułowywane w odpowiednie zadane optymalizacji, do rozwiązania którego jest stosowana metoda gradientu sprzężonego. Praca zawiera ponadto opis opracowanego programu AMFA oraz przykład obliczeniowy ilustrujący działanie tego programu.

Słowa kluczowe: filtry cyfrowe, optymalizacja

1. WSTĘP

Przy projektowaniu cyfrowych filtrów o skończonej odpowiedzi impulsowej (w skrócie SOI) najczęściej zakłada się, że charakterystyka fazowa projektowanego filtru jest liniowa. W takich przypadkach filtry SOI o równomiernie falistych charakterystykach amplitudowych są na ogół projektowane metodami wykorzystującymi algorytm Remeza [1, 2]. Jeżeli oprócz wymagania dotyczącego uzyskania równomiernie falistej charakterystyki amplitudowej, niezbędne jest uwzględnienie innych, dodatkowych ograniczeń dotyczących np. odpowiedzi impulsowej czy jednostkowej filtru, do projektowania filtrów wykorzystuje się zwykle metody programowania liniowego lub nieliniowego [3–7]. We wszystkich tych przypadkach, przy założeniu liniowej charakterystyki fazowej, zadanie aproksymacji charakterystyki amplitudowej jest rozwiązywane w dziedzinie liczb rzeczywistych.

W bardziej ogólnych przypadkach, takich jak np. projektowanie układów wszechprzepustowych (korektorów fazy) czy filtrów SOI posiadających w przybliżeniu

liniową charakterystykę fazową jedynie w pasmie przepustowym uzyskanie charakterystyki fazowej o przebiegu różnym od liniowego. Przy takich założeniach, zadanie aproksymacji charakterystyki częstotliwościowej (amplitudowo-fazowej) filtru jest formułowane w dziedzinie zespolonej. Istnieją różne podejścia do rozwiązywania tego zadania, a tym samym do projektowania tego typu filtrów [8–14]. W części z nich, projektowanie przeprowadza się w drodze rozwiązania zadania aproksymacji charakterystyki częstotliwościowej filtru bezpośrednio w dziedzinie zespolonej. W innych natomiast, zadanie projektowania filtru sformułowane w dziedzinie zespolonej przekształca się w odpowiednie, prawie równoważne, zagadnienie rzeczywiste i następnie rozwiązuje się je w dziedzinie rzeczywistej.

W niniejszej pracy zaproponowano metodę projektowania dolnoprzepustowego filtru SOI, w której stosuje się drugie z wymienionych podejść. Zadanie zaprojektowania filtru o zadanej charakterystyce częstotliwościowej przy założeniu równomiernie falistego przebiegu funkcji błędu jest przekształcane w odpowiednie, rzeczywiste zadanie optymalizacji, które jest następnie rozwiązywane metodą gradientu sprzężonego. Praca zawiera również opis opracowanego programu AMFA oraz przykład obliczeniowy.

2. SFORMUŁOWANIE PROBLEMU

Rozpatrzmy filtr cyfrowy SOI. Charakterystyka częstotliwościowa $H(e^{j\omega})$ takiego filtru jest wyrażona następująco [15, 16]:

$$H(e^{j\omega}) = A(\omega)e^{j\theta(\omega)} = \sum_{n=0}^{N-1} h(n)e^{-j\omega n}, \tag{1}$$

gdzie: $A(\omega) = |H(e^{j\omega})|$ — charakterystyka amplitudowa,
$\theta(\omega) = \arg H(e^{j\omega})$ — charakterystyka fazowa,
$h(n)$ — odpowiedź impulsowa filtru,
N — długość odpowiedzi impulsowej,
ω — pulsacja unormowana względem częstotliwości próbkowania F_p ($\omega \in [-\pi, \pi]$).

Zadanie zaprojektowania filtru o określonej charakterystyce częstotliwościowej sprowadza się do wyznaczenia odpowiednich wartości współczynników $h(n)$, $n = 0, 1, ..., N-1$. W najbardziej ogólnym przypadku współczynniki $h(n)$ są liczbami zespolonymi. Aby współczynniki te były liczbami rzeczywistymi, muszą być spełnione następujące warunki [8]:

$$A(\omega) = A(-\omega) \text{ i } \theta(\omega) = -\theta(-\omega). \tag{2}$$

W dalszej części ograniczymy się do rozpatrywania filtrów o współczynnikach rzeczywistych.

Niech $\tilde{H}(e^{j\omega})$ będzie zadaną charakterystyką częstotliwościową projektowanego filtru. Oznaczmy ponadto przez $\mathbf{Y} = [y_1, y_2, ..., y_N]^T$ wektor o współrzędnych zdefiniowanych następująco:

$$y_i = h(i-1),\ i=1,\ 2,...,N; \tag{3}$$

a przez $H(e^{j\omega}, \mathbf{Y})$ charakterystykę częstotliwościową otrzymaną przy przyjęciu we wzorze (1) współczynników o wartościach określonych przez wektor $\mathbf{Y}$. Funkcja błędu $E(\omega, \mathbf{Y})$ może być wyrażona w postaci:

$$E(\omega, \mathbf{Y}) = W(\omega) \mid [\tilde{H}(e^{j\omega}) - H(e^{j\omega}, \mathbf{Y})] \mid . \tag{4}$$

gdzie:

$$W(\omega) = \begin{cases} 1 \text{ dla } \omega \in P = [0, \omega_p], \\ w \text{ dla } \omega \in S = [\omega_s, \pi], \end{cases} \tag{5}$$

przy czym $w > 0$ jest współczynnikiem wagowym, a ω_p i ω_s są unormowanymi pulsacjami krańcowymi odpowiednio pasma przepustowego i zaporowego.

W rozpatrywanym przez nas przypadku żądamy, aby funkcja błędu miała przebieg równomiernie falisty zarówno w pasmie przepustowym jak i w zaporowym. Zadanie projektowania filtru sformułujemy w sposób następujący: Dla zadanych wartości N, w, ω_p, ω_s, oraz zadanej charakterystyki częstotliwościowej $\tilde{H}(e^{j\omega})$ znaleźć wektor $\mathbf{Y} = [y_1, y_2,,...,y_N]^T$, który minimalizuje

$$\max_{\omega \in P \cup S} E(\omega, \mathbf{Y}). \tag{6}$$

3. PRZEKSZTAŁCENIE PROBLEMU W ZADANIE OPTYMALIZACJI

Rozpatrywane zadanie projektowania filtru, sformułowane zależnością (6), można przekształcić w odpowiednie zadanie optymalizacji. W tym celu przy założonych odpowiednich wartościach początkowych współrzędnych wektora $\mathbf{Y}^{(1)}$, podzielmy najpierw pasmo przepustowe $P = [0, \omega_p]$ i zaporowe $S = [\omega_S, \pi]$ na podprzedziały θ_j, $j = 1, 2,...,J+K$, zdefiniowane następująco:

$$\theta_j = [\omega_{2j-1}; \omega_{2j}], \quad j = 1, 2,..., J+K, \tag{7}$$

gdzie: J i K są liczbami przedziałów θ_i występujących odpowiednio w pasmie przepustowym i zaporowym, a ω_r, $r = 1, 2,..., 2(J+K)$, $(\omega_1 < \omega_2 < ... < \omega_{2(J+K)})$ są unormowanymi pulsacjami, dla których spełnione są warunki:

$$\frac{dE(\omega, \mathbf{Y})}{d\omega} \mid_{\omega = \omega_{2j-1}} > 0, \quad j = 1, 2,..., J+K, \tag{8}$$

$$\frac{dE(\omega, \mathbf{Y})}{d\omega} \mid_{\omega = \omega_{2j}} < 0, \quad j = 1, 2,..., J+K, \tag{9}$$

(1) Sposób wyznaczania tych współrzędnych jest szczegółowo omówiony w dalszej części pracy.

W każdym z tak zdefiniowanych podprzedziałów θ_i, funkcja $E(\omega, \mathbf{Y})$ ma dokładnie jedno maksimum lokalne.

Określmy następnie wartości największych odległości $\Delta H_i(\mathbf{Y})$, pomiędzy charakterystykami $\tilde{H}(e^{j\omega})$ i $H(e^{j\omega}, \mathbf{Y})$. Częstotliwości, przy których te odległości są największe, nazywane są w literaturze [11] częstotliwościami krytycznymi. Częstotliwościami krytycznymi są te, przy których występują lokalne maksima w kolejnych podprzedziałach θ_i, $i=1, 2,..., J+K$.

Częstotliwościami krytycznymi są również częstotliwości ω_p i ω_s. Ponadto, częstotliwościami krytycznymi mogą być też częstotliwości 0 oraz π, jeżeli przy nich występują lokalne maksima funkcji $E(\omega, \mathbf{Y})$. Tak więc wartości największych odległości $\Delta H_i(\mathbf{Y})$ wyznaczymy w sposób następujący:

— w pasmie przepustowym

$$\Delta H_k(\mathbf{Y}) = \max_{\omega \in \theta_k} | \tilde{H}(e^{j\omega}) - H(e^{j\omega}, \mathbf{Y}) | , \quad k=1, 2,..., J \tag{10}$$

$$\Delta H_{J+1}(\mathbf{Y}) = | \tilde{H}(e^{j\omega_p}) - H(e^{j\omega_p}, \mathbf{Y}) | ; \tag{11}$$

— w pasmie zaporowym

$$\Delta H_{J+2}(\mathbf{Y}) = | \tilde{H}(e^{j\omega_s}) - H(e^{j\omega_s}, \mathbf{Y}) | \tag{12}$$

$$\Delta H_k(\mathbf{Y}) = \max_{\omega \in \theta_{k-2}} | \tilde{H}(e^{j\omega}) - H(e^{j\omega}, \mathbf{Y}) | , \quad k=J+3, J+4,..., J+K+2. \tag{13}$$

Ponadto jeżeli $\omega=0$ lub $\omega=\pi$ są częstotliwościami krytycznymi, należy dodatkowo uwzględnić

$$\Delta H_{J+K+3}(\mathbf{Y}) = | \tilde{H}(e^{j0}) - H(e^{j0}, \mathbf{Y}) | \tag{14}$$

$$\Delta H_{J+K+4}(\mathbf{Y}) = | \tilde{H}(e^{j\pi}) - H(e^{j\pi}, \mathbf{Y}) | . \tag{15}$$

Jeżeli tylko jedna z wymienionych częstotliwości jest częstotliwością krytyczną, należy odpowiednio uwzględnić tylko wyrażenie (14) lub (15).

Liczba częstotliwości krytycznych jest zależna od wartości N. Bardziej szczegółowe rozważania na ten temat oraz tablice dla poszczególnych typów filtrów można znaleźć w [11].

Zadanie projektowania filtru o równomiernie falistej funkcji błędu sprowadza się do znalezienia takiego wektora $\mathbf{Y}=[y_1, y_2,..., y_N]^T$, dla którego spełniony jest warunek

$$\Delta E_k(\mathbf{Y}) = \Delta E_l(\mathbf{Y}), \quad k, l=1, 2,...,M, \tag{16}$$

gdzie M jest liczbą częstotliwości krytycznych, a maksima funkcji błędu $\Delta E_k(\mathbf{Y})$ są określone następująco:

$$\Delta E_k(\mathbf{Y}) = \Delta H_k(\mathbf{Y}), \quad k = 1, 2, \ldots, J+1 \quad \text{i w razie potrzeby dodatkowo } k = J+K+3 \tag{17}$$

$$\Delta E_k(\mathbf{Y}) = w\Delta H_k(\mathbf{Y}), \quad k = J+2, J+3, \ldots, J+K+2 \quad \text{i w razie potrzeby dodatkowo } k = J+K+4 \tag{18}$$

Wprowadźmy teraz funkcję $X(\mathbf{Y})$ wyrażoną w sposób następujący:

$$X(\mathbf{Y}) = \sum_{i=1}^{M} (\Delta E_i(\mathbf{Y}) - \tilde{S})^2, \tag{19}$$

przy czym

$$\tilde{S} = \frac{1}{M} \sum_{i=1}^{M} \Delta E_i(\mathbf{Y}) . \tag{20}$$

Przy spełnieniu warunku (16) tak określona funkcja $X(\mathbf{Y})$ osiąga minimum globalne równe zeru.

Rozpatrywane zadanie projektowania filtru, określone zależnością (6), można przekształcić w następujące równoważne zadanie optymalizacji: Znaleźć wektor $\mathbf{Y} = [y_1, y_2, \ldots, y_N]^T$, który minimalizuje funkcję celu $X(\mathbf{Y})$ określoną wzorem (19).

Tak sformułowane zadanie optymalizacji można rozwiązać jedną z metod poszukiwania minimum bez ograniczeń. Z podanych w literaturze metod, do rozwiązywania rozpatrywanego zadania wybrano metodę gradientu sprzężonego Polaka-Ribiery [17, str. 223–227 18, str. 93–96]. W metodzie tej pierwszy wyznaczony kierunek $\mathbf{d}_1$ jest kierunkiem minus gradientu

$$\mathbf{d}_1 = -\nabla f(\mathbf{x}_1), \tag{21}$$

a kolejne kierunki obliczane są nastepująco

$$\mathbf{d}_k = -\nabla f(\mathbf{x}_k) + \beta_{k-1}\mathbf{d}_{k-1}, \quad k = 2, 3, 4, \ldots \tag{22}$$

gdzie

$$\beta_{k-1} = \frac{\langle \nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}_{k-1}), \nabla f(\mathbf{x}_k) \rangle}{\langle \nabla f(\mathbf{x}_{k-1}), \nabla f(\mathbf{x}_k) \rangle} \tag{23}$$

przy czym $\mathbf{x}_k$ i $\nabla f(\mathbf{x}_k)$ oznaczają bieżący punkt i gradient, a symbol $\langle a, b \rangle$ określa iloczyn skalarny wektorów a i b. W rozpatrywanym przy nas zadaniu optymalizacji minimalizowaną funkcją f jest funkcja $x(\mathbf{Y})$ określona zależnością (19), a poszukiwanym wektorem minimalizującym funkcję f jest wektor $\mathbf{Y}$.

Przy rozwiązywaniu zadań optymalizacji, podobnie jak przy rozwiązywaniu innych zagadnień metodami numerycznymi, istotnym problemem jest wybóodpowiedniego punktu startowego. Punkt ten powinien być niezbyt odległy od poszukiwanego rozwiązania, gdyż przyjęcie zbyt odległego punktu startowego może w wielu przypadkach powodować numeryczną rozbieżność algorytmu.

Jako punkt startowy przyjęto wektor $\tilde{\mathbf{Y}}$ będący rozwiązaniem zadania minimalizacji błędu średniokwadratowego $E_2(\mathbf{Y})$. Błąd ten jest zdefiniowany zależnością:

$$E_2(\mathbf{Y}) = \int_{-\pi}^{\pi} | \tilde{H}(e^{j\omega}) - H(e^{j\omega}, \mathbf{Y}) |^2 \, d\omega \, . \tag{24}$$

Rozwiązaniem zadania minimalizacji błędu średniokwadratowego jest suma częściowa szeregu Fouriera. Tak przyjęty punkt startowy jest niezbyt odległy od poszukiwanego rozwiązania zadania równomiernie falistej aproksymacji funkcji błędu $E(\omega, \mathbf{Y})$.

Do określenia wartości funkcji $X(\mathbf{Y})$ niezbędna jest znajomość maksimów funkcji błędu $\Delta E_i(\mathbf{Y})$. Wartości tych maksimów w przedziałach θ_l, $l=1, 2,...,J$ oraz θ_k, $k=J+1, J+3,..., J+K$, wyznaczano posługując się metodą złotego podziału [17, 18]. Metoda ta umożliwia znalezienie odciętej $\tilde{a}$ oraz rzędnej $f(\tilde{a})$ minimum[2] rozpatrywanej funkcji $f(a)$ w określonym przedziale, przy założeniu, że w tym przedziale funkcja $f(a)$ jest ciągła i ma dokładnie jedno minimum.

4. OPIS PROGRAMU AMFA

Jako przykład zastosowania zaproponowanej metody rozpatrzymy dość często spotykane w literaturze zadanie zaprojektowania filtru SOI posiadającego w przybliżeniu liniową charakterystykę fazową jedynie w pasmie przepustowym [8, 11–13]. W przypadku tego rodzaju filtrów można uzyskać opóźnienie grupowe mniejsze niż w przypadku filtrów o takiej samej długości i dokładnie liniowej charakterystyce fazowej zarówno w pasmie przepustowym jak i zaporowym. Zadana (aproksymowana) charakterystyka częstotliwościowa dana jest zależnością [11, 13]:

$$\tilde{H}(e^{j\omega}) = \begin{cases} d e^{-j\omega\tau} & \text{dla} \quad \omega \in P, \\ 0 & \text{dla} \quad \omega \in S, \end{cases} \tag{25}$$

gdzie d jest liczbą rzeczywistą, a $0 \leqslant \tau \leqslant N-1$ jest liczbą całkowitą. Dla tak określonej charakterystyki częstotliwościowej, punkt startowy wyznaczono rozwijając w szereg Fouriera funkcję $\tilde{H}_0(e^{j\omega})$ o postaci:

[2] Należy zauważyć, że każde zadanie maksymalizacji może być sprowadzone do zadania minimalizacji, gdyż $\max f(x) = -\min [-f(x)]$.

$$\tilde{H}_0(e^{j\omega}) = \begin{cases} d\, e^{-j\omega\tau}, & 0 \leqslant \omega_p \leqslant \omega, \\ \dfrac{\omega - \omega_s}{\omega_p - \omega_s} d\, e^{-j\omega\tau}, & \omega_p < \omega < \omega_S, \\ 0, & \omega_s \leqslant \omega \leqslant \pi. \end{cases} \tag{26}$$

Współczynniki tego szeregu są wyrażone następująco:

$$h(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{H}_0(e^{j\omega}) e^{j\omega k} d\omega = \tag{27}$$

$$= \frac{2d}{\pi(\omega_s - \omega_p)(k-\tau)^2} \sin\left[\frac{(\omega_s + \omega_p)(k-\tau)}{2}\right] \sin\left[\frac{(\omega_s - \omega_p)(k-\tau)}{2}\right]$$

$$k = 0, 1, 2, \ldots$$

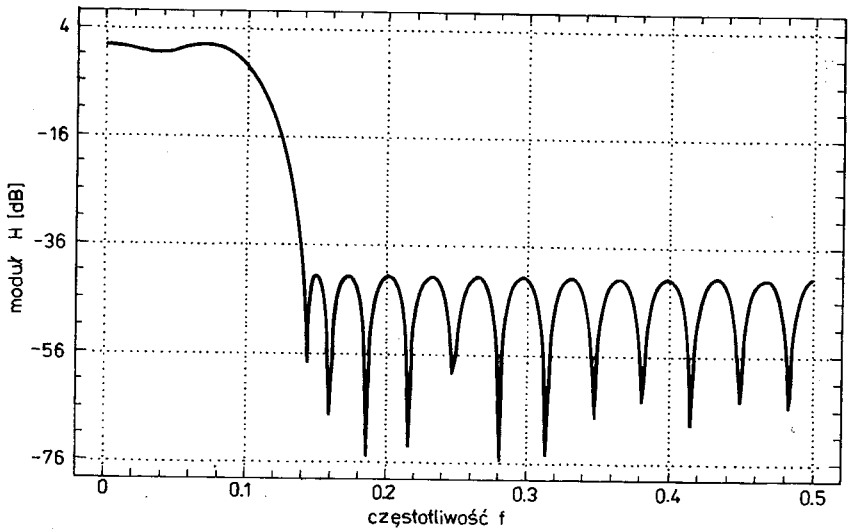
Dla zadanej charakterystyki częstotliwościowej określonej zależnością (25), został opracowany program AMFA. Program ten umożliwia rozwiązanie zadania zaprojektowania filtru o zadanej charakterystyce częstotliwościowej przy założeniu równomiernie falistego przebiegu funkcji błędu. Został on napisany w języku Fortran 77, przy czym do kompilacji wykorzystano kompilator Fortranu firmy Microsoft (wersja 5.1).

Danymi wejściowymi dla programu są:

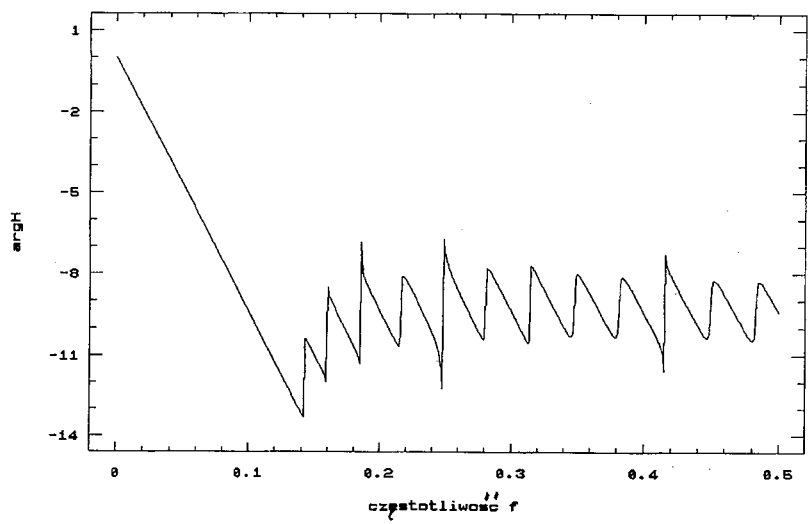
— liczba N,
— wartość w,
— wartości unormowanych częstotliwości $f_p = \omega_p/2\pi$ oraz $f_s = \omega_s/2\pi$,
— wartość parametru IW informującego, czy chcemy otrzymać wydruki kolejnych współrzędnych wyznaczonych charakterystyk amplitudowej $|H(e^{j\omega}, \mathbf{Y})|$ i fazowej $\arg H(e^{j\omega}, \mathbf{Y})$ oraz funkcji błędu $E(\omega, \mathbf{Y})$ (IW = 1), czy też nie (IW = 0),
— wartość parametru IDB informującego, czy wartości funkcji $|H(e^{j\omega}, \mathbf{Y})|$ mają być wyrażone w decybelach (IDB = 1), czy też nie (IDB = 0),
— wartość parametru EPS, zwanego kryterium zakończenia procedury minimalizacji bez ograniczeń [18] (procedura ta kończy swoje działanie, gdy $\| \nabla X(\mathbf{Y}) \| \leqslant$ EPS).

W programie AMFA wykorzystano odpowiednio zmodyfikowaną procedurę GSPR1E opracowaną w Instytucie Automatyki Politechniki Warszawskiej [18]. Procedura ta umożliwia rozwiązywanie zadań programowania nieliniowego bez ograniczeń metodą gradientu sprzężonego Polaka-Ribiery, przy czym gradient funkcji celu jest estymowany numerycznie.

Po przeprowadzeniu obliczeń przy użyciu programu AMFA jako wyniki końcowe otrzymujemy wartości współczynników $h(i)$, $i = 0, 1, \ldots, (N-1)$, występujących we wzorze (1).



Rys. 1. Przebieg charakterystyki amplitudowej $|H(e^{j\omega}), \mathbf{Y})|$ otrzymanej w wyniku obliczeń dla danych: $N=31$, $w=10$, $f_p=0.09$, $f_s=0.14$, $\tau=15$ i $d=1$

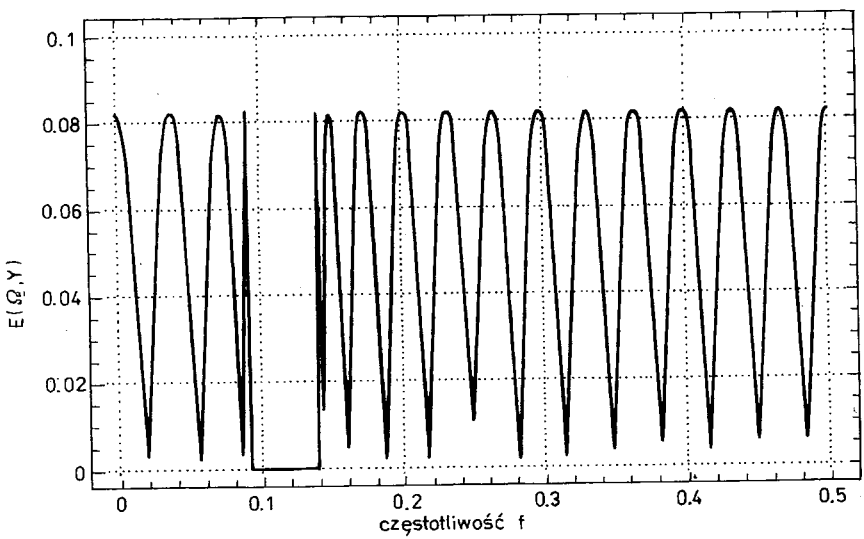


Rys. 2. Przebieg charakterystyki fazowej $\arg H(e^{j\omega}, \mathbf{Y})$ otrzymanej dla danych: $N=31$, $w=10$, $f_p=0.09$, $f_s=0.14$, $\tau=15$ i $d=1$

Istnieje ponadto możliwość drukowania:
— kolejnych wartości współrzędnych wyznaczonych charakterystyk amplitudowej i fazowej oraz funkcji błędu $E(\omega, \mathbf{Y})$ w pasmach przepustowym i zaporowym,
— wartości współrzędnych maksimów wyznaczonej funkcji $E(\omega, \mathbf{Y})$.

Przy użyciu programu AMFA przeprowadzono projektowanie szeregu dolnoprzepustowych filtrów SOI o zadanych charakterystykach częstotliwościowych przy jednoczesnym założeniu równomiernie falistego przebiegu funkcji błędu. Obliczenia wykonywano dla $N \leqslant 35$ przy użyciu mikrokomputera IBM PC 486 z zegarem 33 MHz. Czas wykonywania obliczeń wahał się od kilkunastu sekund do kilku minut, przy czym czas ten wyraźnie zwiększał się w miarę wzrostu liczby N oraz w miarę zmniejszania się wartości parametru EPS.

Przykładowe wyniki projektowania filtru dolnoprzepustowego przy użyciu programu AMFA dla danych $N=31$, $w=10, f_p=0{,}09, f_s=0{,}14$, $\tau=15$ i $d=1$ są podane w tablicy 1. Do obliczeń przyjęto wartość parametru EPS$=10^{-6}$. Przebieg wyznaczonych charaktrystyk amplitudowej i fazowej oraz funkcji błędu dla powyższych danych są pokazane odpowiednio na rysunkach 1, 2 i 3. Należy podkreślić, że przebieg wyznaczonej charakterystyki fazowej w pasmie przepustowym jest praktycznie liniowy. Przebieg tej charakterystyki w pasmie zaporowym nie jest uwzględniany w procesie aproksymacji, co wynika z postaci zadawanej charakterystyki częstotliwościowej danej wzorem (25).



Rys. 3. Przebieg funkcji błędu $E(\omega, \mathbf{Y})$ otrzymanej dla danych: $N=31$, $w=10, f_p=0.09, f_s=0.14$, $\tau=15$ i $d=1$

Tablica 1

Dolnoprzepustowy filtr cyfrowy SOI o zadanej charakterystyce częstotliwościowej

Wyniki obliczeń dla danych: $N=31$, w=10, $f_p=0.09$, $f_s=0.14$, $\tau=15$, d=1

DANE:

liczba N=31
waga w=10
$\tau=15.000$

pasmo przepustowe	.000000	.090000
pasmo zaporowe	.140000	.500000

Wartości współczynników h(i), i=0, 1, ... $(N-1)$

H(0)= −.14557510E−02	H(16)= .20166160E+00
H(1)= .79493810E−02	H(17)= .15405280E+00
H(2)= .14133990E−01	H(18)= .90337770E−01
H(3)= .20632540E−01	H(19)= .28438080E−01
H(4)= .21908810E−01	H(20)= −.17402550E−01
H(5)= .14432220E−01	H(21)= −.39397180E−01
H(6)= −.20129150E−02	H(22)= −.38566630E−01
H(7)= −.22479120E−01	H(23)= −.22680440E−01
H(8)= −.38521140E−01	H(24)= −.20230960E−02
H(9)= −.39421860E−01	H(25)= .14065300E−01
H(10)= −.17406180E−01	H(26)= .21388030E−01
H(11)= .28551980E−01	H(27)= .20168390E−01
H(12)= .90911840E−01	H(28)= .13993700E−01
H(13)= .15424590E+00	H(29)= .74420950E−02
H(14)= .20168890E+00	H(30)= −.15891320E−02
H(15)= .21928130E+00	

UWAGI KOŃCOWE

W pracy przedstawiono metodę projektowania cyfrowego dolnoprzepustowego filtru SOI o zadanej charakterystyce częstotliwościowej, umożliwiającą uzyskanie równomiernie falistej funkcji błędu. Zaletą zaproponowanej metody projektowania jest jej elastyczność. Zmiana postaci zadanej charakterystyki częstotliwościowej wymaga jedynie wprowadzenia niewielkiej modyfikacji w programie. Ponadto w przypadku zaproponowanej metody możliwe jest stosunkowo łatwe uwzględnienie ewentualnych dodatkwych wymagań, również nieliniowych, poprzez wprowadzenie odpowiednich ograniczeń w zadaniu optymalizacji. W przypadku innych metod uwzględnienie dodatkowych wymagań często nie jest możliwe. Ma to miejsce np. w przypadku metod opartych na algorytmie Remeza [1, 2, 11, 12].

BIBLIOGRAFIA

1. D.J. Shpak, A. Antoniou: *A generalized Remez metod for the design of FIR digital filters.* IEEE Trans. Circuits and Syst., 1990, vol. CAS-37, No. 2, p. 161 – 174
2. L.R. Rabiner, J.H. McClellan, T.W. Park: *FIR digital filters design techniques using weighted Chebyshew approximation.* Proc. IEEE, 1975, vol. 64, No 4, p. 596 – 610
3. K. Steiglitz, T.W. Parks, J.F. Kaiser: METEOR: *A constraint-based FIR filter design program.* IEEE Trans. Signal Processing, 1992, vol. 40, No. 8, p. 1901 – 1909
4. A.J. Johnson: *Optimal linear phase digital filter design by one-phase linear programming.* IEEE Trans. Circuits and Syst., 1990, vol. CAS-37, No. 4, p. 554 – 558
5. H. Samueli: *On the design of optimal equiripple FIR digital filters for data transmission applications.* IEEE Trans. Circuits and Syst., 1988, vol. CAS-35, No. 12, Dec. 1988, p. 1542 – 1546
6. F. Wysocka-Schillak: *Projektowanie cyfrowego filru SOI o liniowej fazie z uwzględnieniem wymagań dotyczących jego odpowiedzi jednostkowej.* Kwartalnik Elektroniki i Telekomunikacji, t. 40, z. 2, 1994, ss. 185 – 199
7. F. Wysocka-Schillak, T. Wysocki: *Design of FIR digital filters with monotone pass-band and equiripple stop-band frequency response.* Proceedings of International Symposium on Information Theory & Its Applications, Sydney, Australia, 20 – 24 November 1994, Vol. 1, p. 331 – 334
8. N.N. Chit, J.S. Mason: *Complex Chebyshev approximation for FIR digital filters.* IEEE Trans. Signal Processing, 1991, vol. 39, No. 1, p. 49 – 54
9. T.Q. Nguyen: *The design of arbitrary FIR digital filters using the eigenfilter method.* IEEE Trans. Signal Processing, 1993, vol. 41, No. 3, p. 1128 – 1139
10. C.-Y. Tseng, L.J. Griffiths: *Are equiripple digital filters always optimal with minimax error criterion?* IEEE Signal Processing Letters, 1994, vol. 1, No. 1, p. 5 – 8
11. A.S. Alkhairy, K.G. Christian, J.S. Lim: *Design and characterization of optimal FIR filters with arbitrary phase.* IEEE Trans. Signal Processing, 1993, vol. 41, No. 2, p. 559 – 572
12. K. Preuss: *On the design of FIR filters by complex Chebyshev approximation.* IEEE Trns. Acoust., Speech, Signal Processing, 1989, vol. ASSP-37, No. 5, p. 702 – 712
13. X. Chen, T.W. Parks: *Design of FIR filters in the complex domain.* IEEE Trans. Acoust., Speech, Signal Processing, 1987, vol. ASSP-35, No. 2, p. 144 – 152
14. M. Lang, J. Bamberger: *Nonlinear phase FIR filter design according to the L_2 norm with constraints for the complex error.* Signal Processing, 1994, vol. 36, p. 31 – 40
15. A.V. Oppenheim, R.W. Schaffer: *Cyfrowe przetwarzanie sygnałów,* WKŁ, Warszawa 1979
16. A. Wojtkiewicz: *Elementy syntezy filtrów cyfrowych.* WNT, Warszawa 1984
17. W. Findeisen, J. Szymanowski, A. Wierzbicki: *Teoria i metody obliczeniowe optymalizacji.* WNT, Warszawa 1980
18. T. Kręglewski, T. Rogowski, A. Ruszczyński, J. Szymanowski: *Metody optymalizacji w języku* FORTRAN. PWN, Warszawa 1984

F. WYSOCKA-SCHILLAK

DESIGN OF A FIR DIGITAL FILTER WITH DESIRED FREQUENCY RESPONSE AND EQUIRIPPLE ERROR FUNCTION

Summary

A method is presented for designing FIR low-pass digital filters with desired frequency response using the Chebyshev error. In the method the frequency response approximation problem is converted into a nearly equivalent real error function approximation problem. Then the obtained approximation problem is reformulated and solved using a conjugate gradient method. A description of the computer program AMFA is included. An illustrative numerical design example is also provided.

Key words: digital filters, optimization.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 2, ss. 157–167

Szerokopasmowe, trójwrotowe dzielniki mocy typu Wilkinson'a obciążone zespolonymi impedancjami

STANISŁAW ROSŁONIEC

Instytut Radioelektroniki, Politechnika Warszawska

Otrzymano 1995.05.19

Autoryzowano do druku 1995.

W artykule przedstawiono nową optymalizcyjną metodę projektowania szerokopasmowych dzielników (sumatorów) mocy typu Wilkinson'a obciążonych zespolonymi, zależnymi od częstotliwości, impedancjami. Dzielniki projektowane tą metodą składają się z rezystorów separujących o parametrach skupionych i niewspółmiernych odcinków linii, których impedancje charakterystyczne przyjmują jedną z zadanych wcześniej wartości. Dzięki temu dzielniki te mogą być łatwo realizowalne w technice linii mikropaskowych. Prezentowane wyniki obliczeń i eksperymentu przeprowadzonego w pasmie VHF w pełni potwierdzają słuszność proponowanej metody projektowania. Ponadto, wskazują one na przydatność tego typu dzielników dla praktyki, głównie do urządzeń telekomunikacyjnych pracujących w zakresach VHF i mikrofalowym.

Słowa kluczowe: sumatory, rezystory, dzielniki mocy, linie mikropaskowe.

1. WSTĘP

Trójwrotowe, hybrydowe dzielniki mocy typu Wilkinson'a są szeroko wykorzystywane w różnego rodzaju urządzeniach telekomunikacyjnych pracujących w zakresach VHF i mikrofalowym. Z reguły są one obciążone jednakowymi, niezależnymi od częstotliwości, rezystancjami równymi 50 Ω. Metody projektowania i przykłady konstrukcyjnych rozwiązań tego typu dzielników są opisane w obszernej na ten temat literaturze, dla której publikacjami źródłowymi są [1]–[4]. W późniejszych pracach [5]–[7] przedstawiono algorytmy projektowania bardziej ogólnych wersji omawianych dzielników, których wrota wejściowe i wyjściowe są obciążone różnymi, ale również niezależnymi od częstotliwości rezystancjami. W praktyce jednakże, często zachodzi potrzeba zastosowania szerokopasmowych, hybrydowych dzielników mocy (równodzielących i z izolowanymi wrotami wyjściowymi), których wrota wejściowe

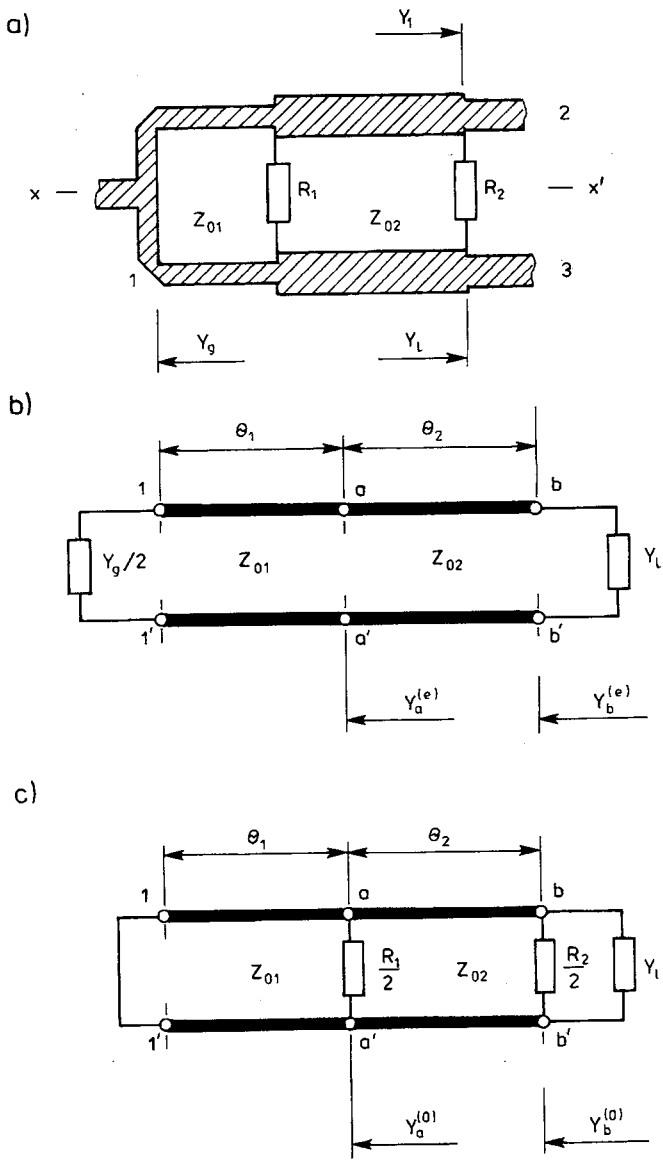
i wyjściowe mogły by być obciążone różnymi i zależnymi od częstotliwości zespolonymi impedancjami. Niestety, ze względu na złożoność problemu zagadnienie projektowania tego typu „uniwersalnych" dzielników nie zostało dotychczas w dostatecznym stopniu omówione w ogólnie dostępnej literaturze. Zatem, celem niniejszego artykułu jest zaprezentowanie nieskomplikowanej metody rozwiązywania tego, ważnego dla praktyki, zagadnienia. Zaproponowano algorytm komputerowego projektowania dwu- i czterosekcyjnych dzielników złożonych z niewspółmiernych odcinków linii TEM i rezystorów separujących o parametrach skupionych. Impedancje charakterystyczne tworzących te dzielniki odcinków linii przyjmują wartości Z_{0min} lub Z_{0max}, które są zadawane przez projektanta. Dzięki temu omawiane dzielniki mogą być łatwo realizowalne w technice linii mikropasmowych. Słuszność proponowanego algorytmu projektowania potwierdzono zarówno wynikami obliczeń jak i badań eksperymentalnych przeprowadzonych w pasmie od 0.8 do 1.2 GHz.

2. ALGORYTM PROJEKTOWANIA

Na rys. 1(a) i 2(a) przedstawiono kolejno zarysy konstrukcji dwu- i czterosekcyjnego dzielników mocy zrealizowanych w technice linii mikropaskowych. Oba przedstawione dzielniki charakteryzują się symetrią lustrzaną względem płaszczyzny x−x', zatem do ich analizy może być stosowana metoda pobudzeń w fazie i w przeciwfazie [8] [3]. Schematy elektryczne uzyskiwanych w procesie analizy dwuwrotników charakterystycznych (+ +) i (+ −) przedstawiono odpowiednio na rys. 1(b, c) i 2(b,c). Zgodnie z [3] [5] [9] falowa macierz rozproszenia [S(f)] całego dzielnika jest związana z macierzami rozproszenia $[S^{++}(f)]$ i $[S^{+-}(f)]$ reprezentujących go dwuwrotników charakterystycznych natępującymi zależnościami:

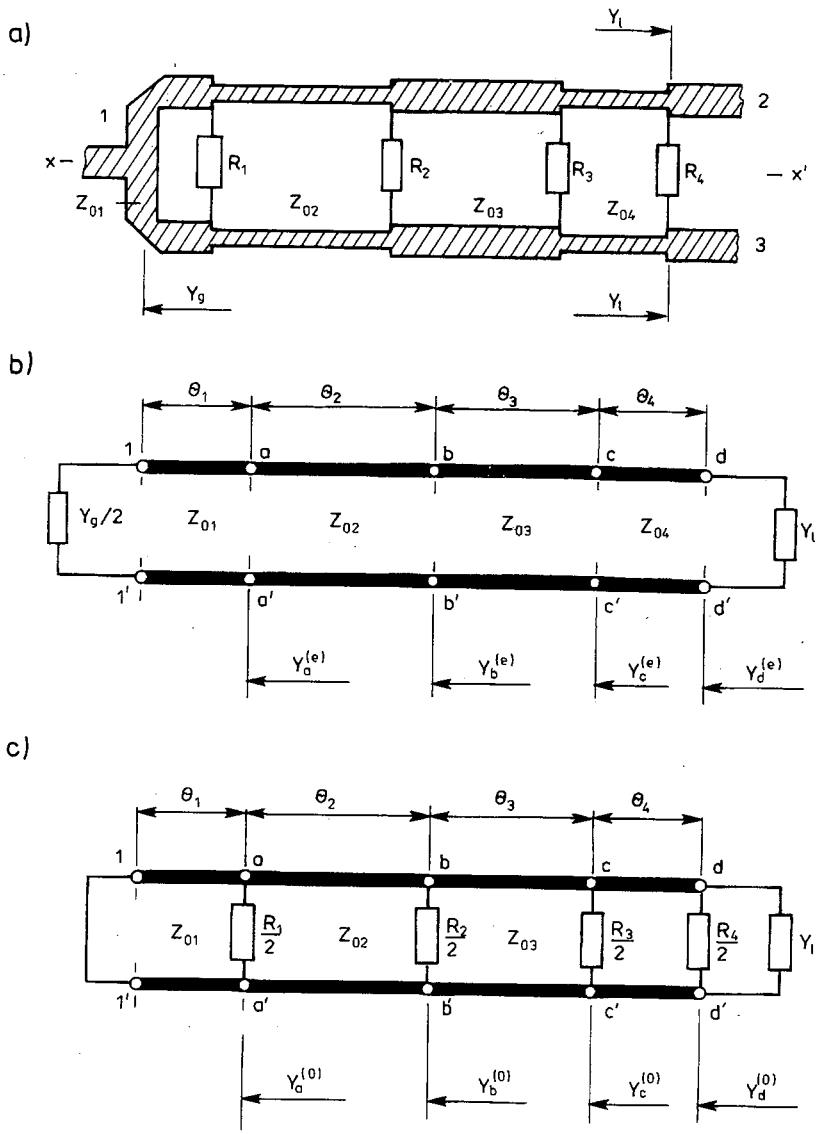
$$\begin{aligned} S_{11}(f) &= S^{++}_{11}(f) \\ S_{12}(f) &= S_{21}(f) = S_{13}(f) = S_{31}(f) = S^{++}_{12}(f)/\sqrt{2} \\ S_{22}(f) &= S_{33}(f) = [S^{++}_{22}(f) + S^{+-}_{22}(f)]/2 \\ S_{23}(f) &= S_{32}(f) = [S^{++}_{22}(f) + S^{+-}_{22}(f)]/2 \,. \end{aligned} \tag{1}$$

Dwuwrotniki charakterystyczne (+ +) pokazane na rys. 1(b) i 2(b) mogą być rozważane jako szerokopasmowe obwody dopasowujące dwie zespolone admitancje, a mianowicie $Y_g(f)/2$ i $Y_1(f)$. Do ich projektowania zaleca sie stosować metodę optymalizacyjną opisaną w [10], ponieważ umożliwia ona uwzględnienie ograniczeń konstrukcyjnych określonych poprzez minimalną Z_{0min} i maksymalną Z_{0max} wartości impedancji charakterystycznej tworzących te obwody odcinków linii. Z podanych wyżej zależności (1) wynika, że parametry rozproszenia $S_{11}(f)$ rozważanych dzielników są tożsame z parametrami rozproszenia $S^{++}_{11}(f)$ wyznaczonymi dla ich dwuwrotników charakterystycznych (+ +). Oznacza to, że są one zależne tylko od admitancji obciążających $Y_g(f)/2$, $Y_1(f)$ i parametrów dopasowującego je obwodu.



Rys. 1. Dwusekcyjny dzielnik mocy (a) struktura dzielnika, (b) schemat dwuwrotnika charakterystycznego (++), (c) schemat dwuwrotnika charakterystycznego (+−).

Innymi słowami, charakterystyka $S_{11}(f)$ jest niezależna od wykorzystywanych w dzielnikach rezystorów separujących. Właściwość ta pozwala na kształtowanie charakterystyki izolacji dzielnika $I(f)[dB]=20\log \mid 1/S_{23}(f) \mid$ niezależnie od charakterystyki strat odbiciowych (ang. *return loss*) $S_{11}(f)[dB]=20\log \mid S_{11}(f) \mid$. Załóżmy zatem, że impedancje charakterystyczne i długości elektryczne odcinków linii tworzących projektowany dzielnik zostały wyznaczone według metody opisanej w [10]. Gdy wymienione



Rys. 2. Czterosekcyjny dzielnik mocy (a) struktura dzielnika, (b) schemat dwuwrotnika charakterystycznego (++), (c) schemat dwuwrotnika charakterystycznego (+−).

wyżej wielkości są znane, najtrudniejszym zadaniem procesu projektowania dzielnika pozostaje wyznaczenie takiego rozkładu rezystorów separujących, dla którego charakterystyki izolacji pomiędzy wrotami wyjściowymi $I(f)[dB]=20\log \mid 1/S_{23}(f) \mid$ i strat odbiciowych we wrotach wyjściowych $S_{22}(f)[dB]=S_{33}(f)[dB]=20\log \mid S_{22}(f) \mid$ będą mieć optymalne przebiegi w sensie kryterium minimax [3][11]. W tym przypadku przez rozkład rezystorów separujących rozumie się ich ilość, wartości rezystancji

i sposób rozmieszczenia w strukturze dzielnika, patrz rys. 1 i 2. Powyższe zadanie można efektywnie rozwiązywać stosując przedstawioną poniżej dwuetapową metodę optymalizacyjną. Dla konkretności dalszych rozważań przyjmijmy, że projektowany jest dzielnik czterosekcyjny z czteroma rezystorami separującymi rozmieszczonymi jak na rys. 2.

Przyjęte założenie nie ogranicza ogólności prezentowanej metody, służącej do minimalizacji fukcji

$$| S_{23}(f) | = | [S^{++}_{22}(f) - S^{+-}_{22}(f)] | /2 \tag{2}$$

gdzie:

$$S^{++}_{22}(f) = \exp(j2\phi)[Y^{*}_{1}(f) - Y^{(e)}_{d}(f)]/[Y_{1}(f) + Y^{(e)}_{d}(f)]$$

$$S^{+-}_{22}(f) = \exp(j2\phi)[Y^{*}_{1}(f) - Y^{(0)}_{d}(f)]/[Y_{1}(f) + Y^{(0)}_{d}(f)].$$

Admitancja $Y^{*}_{1}(f)$ występująca w powyższej zależności jest wielkością zespoloną, sprzężoną względem admitancji obciążenia $Y_1(f) = | Y_1(f) | \exp(j\phi)$. Z zależności (2) wynika wniosek, że parametr rozproszenia $| S_{23}(f) |$ osiąga minimalne, zerowe wartości gdy admitancje $Y^{(e)}_{d}(f) = G^{(0)}_{d}(f) + jB^{(e)}_{d}(f)$ i $Y^{(0)}_{d}(f) = G^{(e)}_{d}(f) + jB^{(0)}_{d}(f)$ są sobie równe. Zatem, w celu minimalizacji maksymalnej wartości funkcji $| S_{23}(f) |$ należy dążyć do zbliżenia wymienionych wyżej admitancji w wymaganym pasmie częstotliwości $f_1 \div f_2$. Zadanie to można zapisać następująco:

$$\min_{G} \max_{f_1 \leqslant f \leqslant f_2} \frac{[G^{(e)}_{d}(f) - G^{(0)}_{d}(f)]^2 + [B^{(e)}_{d}(f) - B^{(0)}_{d}(f)]^2}{[G^{(e)}_{d}(f)]^2 + [B^{(e)}_{d}(f)]^2} \tag{3}$$

gdzie $G = [1/R_1, 1/R_2, 1/R_3, 1/R_4]$ jest czterowymiarowym wektorem, którego elementami składowymi są poszukiwane konduktancje (rezystancje R_1, R_2, R_3 i R_4). Pierwszym krokiem pierwszego etapu proponowanej metody rozwiązywania zadania minimax (3) jest wyznaczenie takiej wartości konduktancji $G_1 = 1/R_1$ dla której admitancje $Y^{(e)}_{a}(f) = G^{(e)}_{a}(f) + jB^{(e)}_{a}(f)$ i $Y^{(0)}_{a}(f) = G^{(0)}_{a}(f) + jB^{(0)}_{a}(f)$ będą maksymalnie zbliżone do siebie w sensie następującego kryterium:

$$\min_{G_1} \max_{f_1 \leqslant f \leqslant f_2} \frac{[G^{(e)}_{a}(f) - G^{(0)}_{a}(f)]^2 + [B^{(e)}_{a}(f) - B^{(0)}_{a}(f)]^2}{[G^{(e)}_{a}(f)]^2 + [B^{(e)}_{a}(f)]^2} \tag{4}$$

Po wyznaczeniu konduktancji $G^{(1)}_{1}$, stanowiącej rozwiązanie zadania (4), w podobny sposób wyznacza się pierwsze przybliżenie konduktancji $G_2 = 1/R_2$. Naturalnie, w tym przypadku należy rozwiązać następujący, jednowymiarowy problem:

$$\min_{G_2} \max_{f_1 \leqslant f \leqslant f_2} \frac{[G_b^{(e)}(f) - G_b^{(0)}(f)]^2 + [B_b^{(e)}(f) - B_b^{(0)}(f)]^2}{[G_b^{(e)}(f)]^2 + [B_b^{(e)}(f)]^2} \tag{5}$$

gdzie $G_b^{(e)}(f)$, $B_b^{(e)}(f)$, $G_b^{(0)}(f)$ i $B_b^{(0)}(f)$ są odpowiednio rzeczywistymi i urojonymi składowymi admitancji $Y_b^{(e)}(f)$ i $Y_b^{(0)}(f)$. W wyniku rozwiązania zadania (5) otrzymuje się konduktancję $G_2^{(1)}$, co stanowi wynik końcowy drugiego kroku pierwszego etapu. Powtarzając powyższą procedurę dla rezystorów (konduktancji) rozmieszczonych w płaszczyznach c – c' i d – d', patrz rys. 2(c), wyznacza się trzecią i czwartą składowe wektora $G^{(1)} = [G_1^{(1)}, G_2^{(1)}, G_3^{(1)}, G_4^{(1)}]$. Wyznaczone w ten sposób pierwsze przybliżenie wketora G jest wykorzystywane jako punkt startowy dla drugiego etapu procesu optymalizacji. Dalszą minimalizację funkcji $| S_{23}(f, G) |$ w zadanym pasmie częstotliwości $f_1 \div f_2$ kontynuuje się metodą ε — najszybszego spadku [11]. Istota tej metody polega na minimalizacji funkcji celu

$$F = \sum_{f_{ei} \subset Q} [\, | S_{23}(f_{ei}, G) | \,]^2, \tag{6}$$

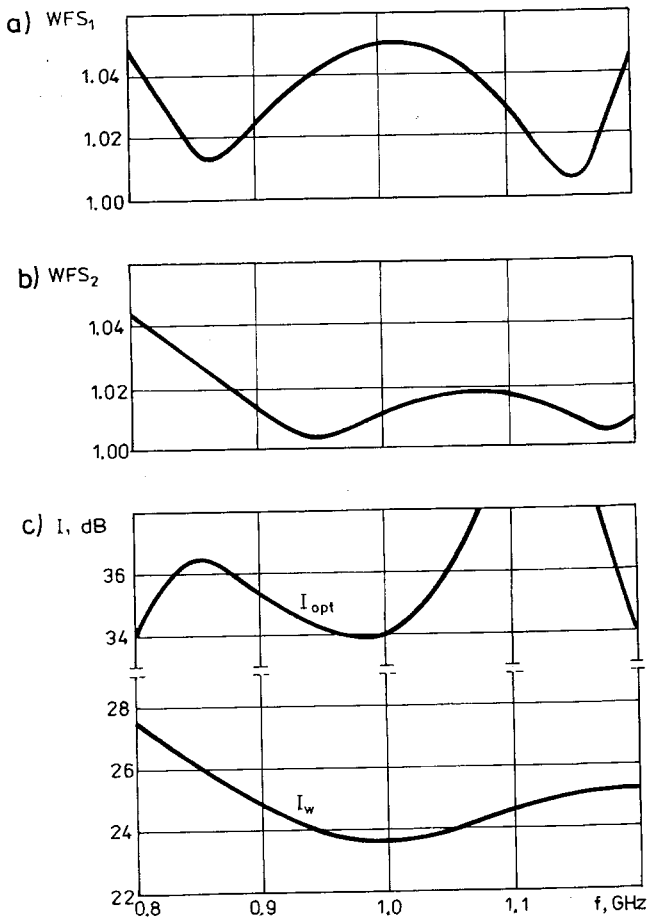
gdzie Q jest zbiorem dyskretnych częstotliwości f_{ei}, dla których spełnione są następujące nierówności:

$$0.97 \, | S_{23}(f,G) |_{\max} \leqslant | S_{23}(f_{ei},G) | \leqslant | S_{23}(f,G) |_{\max} \tag{7}$$
$$f_1 \leqslant f \leqslant f_2.$$

Do minimalizacji funkcji (6), względem konduktancji G_1, G_2, G_3 i G_4 można wykorzystywać dowolną z metod gradientowych, na przykład metodę Fletchera – Reeves'a lub Davidona – Fletchera – Powella, [12] – [14]. Podobnie jak w poprzednim etapie konduktancje G_1, G_2, G_3 i G_4 muszą być wielkościami nieujemnymi i niezależnymi od częstotliwości.

3. WYNIKI BADAŃ

W celu potwierdzenia słuszności prezentowanej metody i jej przydatności dla praktyki, zaprojektowano czterosekcyjny dzielnik mocy dla następujących danych: $Y_g(f) = (1/50)$ S, $Y_1(f) = 1/50 + j(1/75)\tan(0.4f/f_0)$ S, $f_1 = 0.8$ GHz i $f_2 = 1.2$ GHz. Przyjęto ponadto, że projektowany dzielnik jest złożony z odcinków linii o impedancjach charakterystycznych $Z_{0\min} = 40\ \Omega$ i $Z_{0\max} = 80\ \Omega$. Wyznaczone według metody opisanej w [10] impedancje charakterystyczne i długości elektryczne tworzących dzielnik odcinków linii są równe: $Z_{01} = Z_{04} = 80\ \Omega$, $Z_{02} = Z_{03} = 40\ \Omega$, $\theta_1(f_0) = 1.8757$ rad, $\theta_2(f_0) = \theta_3(f_0) = 0.2646$ rad i $\theta_4(f_0) = 0.6424$ rad, gdzie $f_0 = 1$ GHz. Obliczoną dla tych danych charakterystykę współczynnika fali stojącej we wrotach wejściowych dzielnika $WFS_1(f) = [1 + | S_{11}(f) |]/[1 - | S_{11}(f) |]$ pokazano na rys. 3(a). W trakcie projektowania przyjęto, że pięć rezystorów separujących zostanie rozmieszczonych w płaszczyznach a – a', b-b', c-c', d-d' i e-e', których położenie określają długości



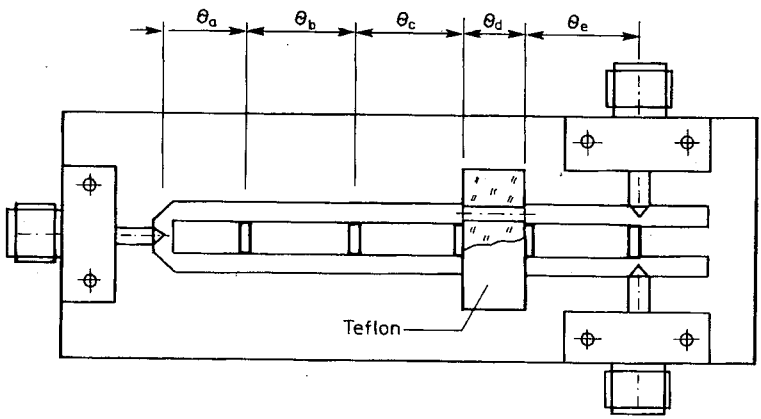
Rys. 3 Charakterystyki współczynnika fali stojącej $WFS_1(f)$, $WFS_2(f)$ i izolacji $I(f)$[dB] obliczone dla projektowanego dzielnika

elektryczne $\theta_a(f_0)=0.6250$ rad, $\theta_b(f_0)=0.6250$ rad, $\theta_c(f_0)=0.6257$ rad, $\theta_d(f_0)=0.5293$ rad i $\theta_e(f_0)=0.6424$ rad, patrz rys. 4(a). Wartości rezystancji $R_1 \div R_5$ wyznaczone w pierwszym etapie procesu optymalizacji są równe: $R_1=100\ \Omega$, $R_2=142.84\ \Omega$, $R_3=181.80\ \Omega$, $R_4=1998.90\ \Omega$ i $R_5=166.65\ \Omega$. Odpowiadającą im „wstępną" charakterystykę izolacji dzielnika $I_w(f)$ [dB] przedstawia rys. 3(c). W drugim etapie procesu optymalizacji rezystancje te ulegają istotnym zmianom i przyjmują następujące, optymalne wartości: $R_1=100.02\ \Omega$, $R_2=144.67\ \Omega$, $R_3=171.19\ \Omega$, $R_4=1190.23\ \Omega$ i $R_5=256.31\ \Omega$. W następstwie tych zmian uzyskuje się charakterystykę izolacji o optymalnym przebiegu $I_{opt}(f)$ [dB], przedstawioną także na rys. 3(c). Wyznaczone powyżej optymalne przebiegi parametrów rozproszenia $S_{22}^{++}(f)$ i $S_{22}^{+-}(f)$ umożliwiają również obliczenie charakterystyki współczynnika fali stojącej we wrotach wyjściowych $WFS_2(f)=WFS_3(f)=[1+|S_{22}(f)|]/[1-|S_{22}(f)|]$. W tym celu wykorzystuje się podaną wcześniej zależność $S_{22}(f)=S_{33}(f)=$

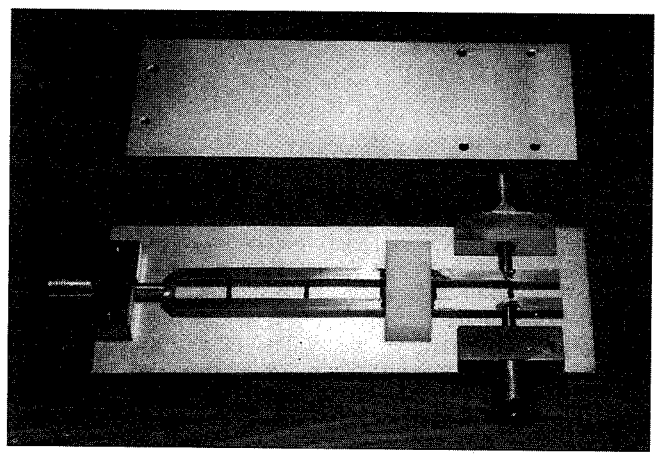
$=[S^{++}_{22}(f)+S^{+-}_{22}(f)]/2$. Przebieg obliczonej w ten sposób charakterystyki $WFS_2(f)$ przedstawiono na rys. 3(b).

Zaprojektowany powyżej dzielnik zrealizowano w technice symetrycznej, powietrznej linii paskowej o odstępie pomiędzy płaszczyznami ekranującymi $b=9.1$ mm i grubości przewodnika wewnętrznego $t=0.8$ mm [14] [15]. Wrota wyjściowe dzielnika są obciążone admitancjami $Y_1(f)$ utworzonymi poprzez równoległe dołączenie do stałej rezystancji $Z_0=50\ \Omega$ rozwartego na końcu odcinka linii o impedancji charakterystycznej $Z_{0s}=75\ \Omega$ i długości elektrycznej $\theta_s(f_0)=0.4$ rad. Zarys konstrukcji tego dzielnika pokazano na rys. 4. Natomiast na rys. 5 przedstawiono wyznaczone eksperymentalnie charakterystyki strat odbiciowych $S_{11}(f)$ [dB] $=20\log|S_{11}(f)|$, $S_{22}(f)[dB]=S_{33}(f)[dB]=20\log|S_{22}(f)|$ i izolacji $I(f)[dB]=20\log|S_{23}(f)|$. Na podkreślenie zasługuje duża zgodność prezentowanych wyników eksperymentalnych z odpowiednimi charakterystykami teoretycznymi, pokazanymi na rys. 3.

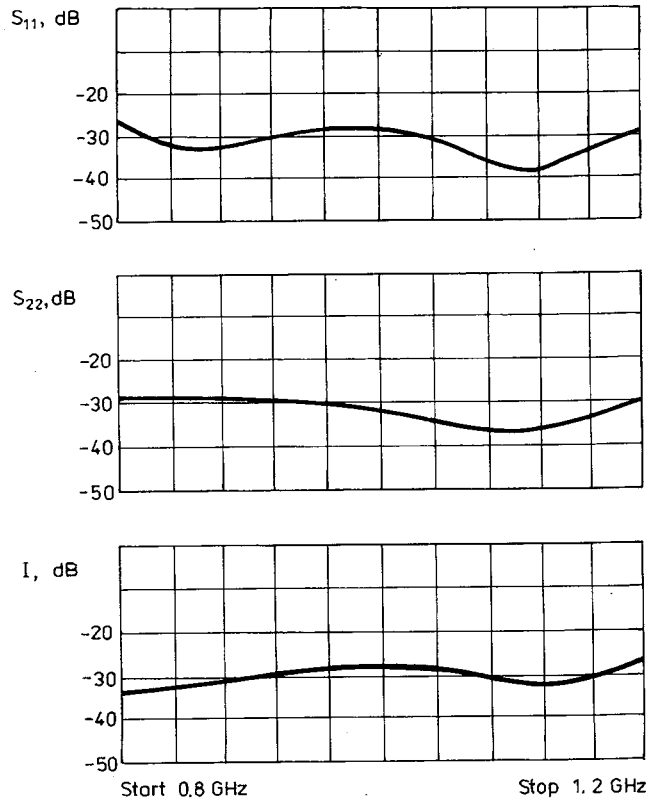
a)



b)



Rys. 4 Czterosekcyjny dzielnik mocy badany eksperymentalnie a) zakres konstrukcji b) widok dzielnika po zdjęciu górnej płyty ekranującej.



Rys. 5 Charakterystyki strat odbiciowych $S_{11}(f)$[dB], $S_{22}(f)$[dB] i izolacji $I(f)$[dB] wyznaczone eksperymentalnie dla dzielnika pokazanego na rys. 4

PODSUMOWANIE

W artykule przedstawiono nową optymalizacyjną metodę projektowania szerokopasmowych dzielników (sumatorów) mocy typu Wilkinson'a obciążonych zespolonymi, zależnymi od częstotliwości, impedancjami. Projektowane tą metodą dzielniki składają się z rezystorów separujących o parametrach skupionych i niewspółmiernych odcinków linii, których impedancje charakterystyczne przyjmują jedną z zadanych wartości Z_{0min} lub Z_{0max} Dzięki temu dzielniki te mogą być łatwo realizowalne w technice linii mikropaskowych. Danymi wejściowymi dla procesu projektowania są impedancje obciążające wrota wejściowe $Z_g(f)$, wrota wyjściowe $Z_1(f)$, szerokość pasma $\Delta f=(f_2-f_1)$ oraz wymienione wyżej wartości impedancji charakterystycznej Z_{0min} i Z_{0max}. W zależności od wymagań stawianych charakterystykom dopasowania w poszczególnych wrotach wybiera się liczbę sekcji dzielnika, tj. projektuje dzielnik dwusekcyjny, lub bardziej szerokopasmowy, czterosekcyjny. Wymaganą charakterys-

tykę izolacji pomiędzy wrotami wyjściowymi projektowanego dzielnika uzyskuje się poprzez zapewnienie odpowiedniego rozkładu rezystorów separujących, których ilość i sposób rozmieszczenia należy dobrać metodą modelowania numerycznego, indywidualnie dla każdego zadania. Optymalne wartości rezystancji tych rezystorów wyznacza się stosując opisaną w artykule dwuetapową metodę minimalizacji funkcji $|S_{23}(f)|$. Słuszność proponowanej metody projektowania potwierdzają przedstawione wyniki obliczeń i eksperymentu przeprowadzonego w pasmie $0.8 \div 1.2$ GHz.

BIBLIOGRAFIA

1. E.J. Wilkinson: *An n-way hybrid power divider*. IRE Trans. MTT-8, Microwave Theory and Techniques, January 1960, pp. 116–118
2. L. Parad and R.L. Moynihan: *Split-tee power divider*. IEEE Trans. MTT-13, Microwave Theory and Techniques, January 1965, pp. 91–95.
3. S.B. Cohn: *A class of broadband three-port TEM-mode hybrids*. IEEE Trans., MTT-16, Microwave Theory and Techniques, February 1968, pp 110–116
4. N. Nagai, E. Maekawa, K. Ono: *New n-way hybrid Power dividers*. IEEE Trans., MTT-25, Microwave Theory and Techniques, December 1977, pp. 1008–1012
5. R.B. Ekinge: *A new method of synthesizing matched broad-band TEM-mode three ports*. IEEE Trans., MTT-19, Microwave Theory and Techniques, January 1971, pp. 81–88
6. S. Rosłoniec: *An impedance transforming hybrid power divider and its miniaturized structures*. Proc. of the 37 Wiessen. Kolloquium, Ilmenau 1992, Band 2, s. 841–848
7. S. Rosłoniec, J. Kleinknecht: *Impedance transforming hybrid power dividers*. AEU-47, Archiv für Elektronik und Übertragungstechnik, July 1993, pp. 270–272
8. J. Reed, G. Wheeler: *A method of analysis of symmetrical four-port networks*. IRE Trans. MTT-4, Microwave Theory and Techniques, October 1956, pp. 246–252
9. P.A. Rizzi: *Microwave engineering, passive circits*, Prentice-Hall, Englewood Cliffs, NJ, 1988
10. S. Rosłoniec: *Design of stepped transmission line matching circuits by optimization methods*. IEEE Trans., MTT-41, Microwave Theory and Techniques, December 1994, pp. 2255–2260
11. W.F. Diemianov, W.T. Małoziemov: *Introduction into the minimax problems* (in Russian), Nauka, Moscow 1972
12. D.M. Himmelblau: *Applied nonlinear programming*. McGraw-Hill, New York 1972
13. M.S. Bazaraa, C.M. Shetty: *Nonlinear programming, theory and algorithms*, John Wiley and sons, New York 1979
14. S. Rosłoniec: *Algorithms for computer-aided design of linear microwave circuits*, Artech House, Boston–London 1990
15. B.C. Wadell: *Transmission line design handbook*, Artech House, Boston–London 1991

BROADBAND HYBRID POWER DIVIDERS TERMINATED IN COMPLEX FREQUENCY DEPENDENT IMPEDANCES

Summary

A new CAD algorithm for design of two- and four-section threeport hybrid power dividers terminated with complex frequency dependent impedances is described. The dividers under consideration are composed of lumped element resistors and noncommensurate transmission line sections whose characteristic impedances take extreme, practically realizable, values. These values are assumed freely at the beginning of a design process. It has been confirmed numerically and experimentally that dividers of theat type are suitable for many telecommunication applications.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 2, ss. 169—184

Synteza struktur kanonicznych czwórników i trójników zawierających wzmacniacz operacyjny

ZYGMUNT WRÓBEL

Zakład Elektrotechniki i Automatyki, Uniwersytet Śląski

Otrzymano 1994.12.07

Autoryzowano do druku 1995.03.02

W pracy przedstawiono procedurę syntezy topologicznej, pozwalającą wygenerować wszystkie możliwe struktury kanoniczne układów czwórników i trójników zawierających idealny wzmacniacz operacyjny. Podano warunki realizacji minimalno-elementowych struktur kanonicznych czwórników i trójników aktywnych.

Słowa kluczowe: czwórniki, trójniki, synteza topologiczna, struktura kanoniczna, wzmacniacz operacyjny

1. WPROWADZENIE

Procedurę syntezy układów można w zasadzie podzielić na dwie odrębne grupy zagadnień [10]:

— syntezę topologiczną układów,

— syntezę parametryczną układów.

Pod pojęciem syntezy topologicznej układów będziemy rozumieć generację wszystkich możliwych struktur równoważnych. Struktura kanoniczna to reprezentant danej klasy struktur równoważnych. Pod pojęciem syntezy parametrycznej układów, będziemy rozumieć opracowanie algorytmów umożliwiających takie wyznaczenie rodzaju i wartości poszczególnych elementów tworzących wygenerowane struktury topologiczne, aby skonstruowany w efekcie układ posiadał pożądane własności zewnętrzne.

W pracach dotyczących syntezy układów elektronicznych [1—4] pierwszemu etapowi syntezy (czyli generacji struktur kanonicznych) poświęca się niewiele uwagi. Związane jest to przede wszystkim z faktem, że dotychczas nie opracowano formalnych metod syntezy kanonicznych struktur układów. Strukturę układu (w pracach dotyczących ich syntezy) zadaje się zwykle *a priori* (najczęściej znane

z literatury typy struktur np.: drabinkowe, mostkowe) koncentrując uwagę przede wszystkim na drugim etapie syntezy, poszukując zwykle rozwiązań optymalnych ze względu na zadany parametr funkcji celu procesu syntezy. Generowane w procesie syntezy układy, mogą być między sobą równoważne. Zagadnienie rówoważności układów pasywnych przedstawiono w pracy [9].

Jak wynika z rozważań pracy [11] włączenie idealnego wzmacniacza operacyjnego do układu pasywnego — powoduje, że opisująca własności powstałego układu aktywnego sumy H^v_{k-1} dendrytów $(k-1)$-drzewowych T^v_{k-1} są tworzone z sum H^v_k dendrytów k-drzewowych T^v_k charakteryzujących część pasywną układu. Opracowany w pracy [10] algorytm syntezy struktur kanonicznych układów pasywnych jest bazą dla opracowania algorytmu syntezy struktur kanonicznych układów aktywnych z jednym wzmacniaczem operacyjnym.

Celem niniejszej pracy jest przedstawienie zagadnienia syntezy struktur kanonicznych układów czwórników i trójników zawierających idealny wzmacniacz operacyjny. Przedstawiono warunki realizacji minimalno-elementowych struktur kanonicznych czwórników i trójników aktywnych. W pracy, dla rozróżnienia układów pasywnych i aktywnych, ich parametry będziemy oznaczać odpowiednio indeksami p i a. Sposób numeracji wierzchołków i gałęzi zgodny z pracą [6]. Podstawowe zagadnienia konieczne do zrozumienia rozważań niniejszej pracy przedstawiono w pracach [6–10].

2. D-RÓWNOWAŻNOŚĆ STRUKTUR UKŁADÓW ZAWIERAJĄCYCH IDEALNY WZMACNIACZ OPERACYJNY

Generowane w procesie syntezy struktury układów zawierających wzmacniacz operacyjny, należy podzielić na klasy struktur D-równoważnych.

Korzystając z rozważań pracy [9], strukturalną macierz dendrytów 1-drzewowych $\mathbf{D}\{T^v_1(t_1,t_2)\}|^a$ bi-struktury $G^v(t_1,t_2)|^a$ zapiszemy w postać:

$$\mathbf{D}\{T^v_1(t_1,t_2)\}|^a = \begin{vmatrix} G^v(t_1,t_2)|^a \\ \mathbf{D}\{T^v_1(t2)\}|^p \end{vmatrix}. \tag{1}$$

Korzystając z rozważań pracy [9] oraz uwzględniając ograniczenia dotyczące włączenia wzmacniacza operacyjnego do układu [11], D-równoważność struktur aktywnych można również zdefiniować następująco:

Dwie bi-struktury $G^v(t_1,t_2)|^a_{II}$ układów aktywnych są D-równoważne, jeśli istnieje funkcja różnowartościowa ciągła i mająca odwrócenie ciągłe, która odwzorowuje macierz dendrytów 1-drzewowych $\mathbf{D}\{T^v_1(t_1,t_2)|^a_I$ bi-struktury $G^v(t_1,t_2)|^a_I$, w macierz dendrytów 1-drzewowych $\mathbf{D}\{T^v_1(t_1,t_2)\}|^a_{II}$ bi-struktury $G^v(t_1,t_2)|^a_{II}$, czyli:

$$f: \mathbf{D}\{T^v_1(t_1,t_2)\}|^a_I \Rightarrow \mathbf{D}\{T^v_1(t_1,t_2)\}|^a_{II}. \tag{2}$$

Analogicznie można zdefiniować D-równoważność multi-struktur $G^v(tm)$ i uni-struktur $G^v(t)$ układów aktywnych, co w skrócie zapisujemy:

Dla multi-struktur układów aktywnych:

$$f: \mathbf{D}\{T_1^v(t2)\}|_I^a \Rightarrow \mathbf{D}\{T_1^v(t2)\}|_{II}^a . \tag{3}$$

Dla uni-struktur układów aktywnych:

$$f: \mathbf{D}\{T_1^v(t)\}|_I^a \Rightarrow \mathbf{D}\{T_1^v(t)\}|_{II}^a . \tag{4}$$

Z powyższych rozważań wynika, że przedstawiony w pracy [9], algorytm D-równoważności struktur układów pasywnych można po niewielkiej modyfikacj zastosować rownież do stwierdzenia D-równoważności struktur układów aktywnych.

Modyfikacja algorytmu D-równoważności dotyczy przede wszystkim ograniczenia związanego z permutacją poszczególnych elementów struktury układu. Jak wynika z pracy [11] norator jest zawsze dołączony do wierzchołka odniesienia $\{w_0\}$. W związku z tym permutacje w pierwszej kolejności muszą dotyczyć gałęzi, w które włączono norator do struktury układu.

3. SYNTEZA STRUKTUR KANONICZNYCH UKŁADÓW ZAWIERAJĄCYCH IDEALNY WZMACNIACZ OPERACYJNY

Jak wynika z prac [2, 11] „włączenie" idealnego wzmacniacza operacyjngo do układu pasywnego, nie tylko nie komplikuje obliczeń poszczególnych parametrów powstałego układu aktywnego, ale nawet je nieco upraszcza w stosunku do bazowego układu pasywnego. Układ aktywny zawierający idealny wzmacniacz operacyjny będzie w procesie syntezy topologicznej, tworzony jako suma mnogościowa układu pasywnego oraz nulatora i noratora modelujących idealny wzmacniacz operacyjny.

Korzystając z rozważań prac [8, 10], algorytm syntezy topologicznej struktur kanonicznych trójników i czwórników aktywnych metodą przeglądu hierarchicznego można w skrócie zapisać w postaci dwóch etapów:

1. Generacji pełnego zbioru uni-struktur $G^v(t)|^a$ układów aktywnych i ich podziału na klasy struktur kanonicznych $\underline{G}^v(t)|^a$.

2. Tworzenie bi-struktur w poszczególnych klasach uni-struktur kanonicznych $\underline{G}^v(t)|^a$ oraz podział wygenerowanych bi-struktur $G^v(t_1,t_0)|^a$ na klasy ich struktur kanonicznych $\underline{G}^v(t_1,t_2)|^a$.

Niżej krótko omówiono poszczególne etapy syntezy struktur kanonicznych układów aktywnych metodą przeglądu hierarchicznego.

Ad. 1. Generacji pełnego zbioru uni-struktur $G^v(t)|^a$ układów aktywnych i ich podziału na klasy struktur kanonicznych $\underline{G}^v(t_1,t_2)|^a$.

Przystępując do generacji wszystkich możliwych uni-struktur $G^v(t)|^a$ układów, należy zwrócić uwagę, że na kolejnym etapie syntezy topologicznej będziemy tworzyć minimalno-elementowe bi-struktury układów jako sumy mnogościowe $G^v(t_1,t_2)|^a = T_{k-1}^v(t_1) \cup T_{k-1}^v(t_2)$. W związku z tym należy rozpatrywać tylko te uni-struktury $G^v(t)|^a$, których liczba gałęzi l_0 spełnia zależność:

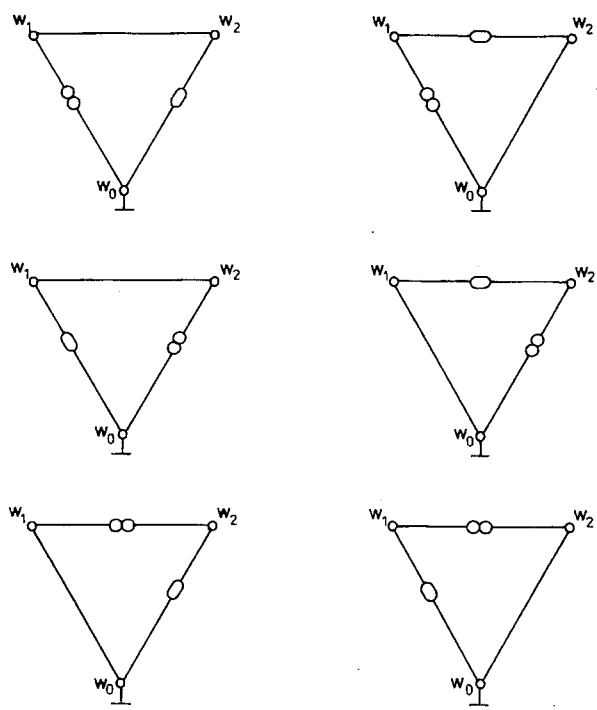
$$v \leqslant l_e \leqslant 2(v-2)+2. \tag{5}$$

Ograniczenie od dołu liczby gałęzi w uni-strukturze $G^v(t)|^a$ układu wynika z faktu, że uni-struktura $G^v(t)|^a$ czwórnika i trójnika może być cyklicznie spójna względem obu par wierzchołków brzegowych (wejściowych i wyjściowych) gdy ma co najmniej v gałęzi.

Ograniczenie od góry liczby gałęzi w uni-strukturze $G^v(t)|^a$ trójnika lub czwórnika wynika z faktu, że parametry z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij} mają postać funkcji wymiernej, której licznik i mianownik są sumami iloczynów dendrytów $(k-1)$-drzewowych [11]. Minimalno-elementowe układy trójników i czwórników realizując funkcję wymierną o maksymalnym stopniu licznika i mianownika, są w ogólnym przypadku tworzone jako sumy mnogościowe $G^v(t_1,t_2)|^a = T^v_{k-1}(t_1) \cup T^v_{k-1}(t_2)$ dwóch dendrytów 2-drzewowych zawierających po $(v-2)$ elementów. Ponadto uni-struktura układu aktywnego posiada dwie gałęzie „zarezerwowane" dla nulatora i noratora. Jeśli gałęzie dendrytów 2-drzewowych się nie pokrywają w bi-strukturze układu to mamy ograniczenie od góry liczby gałęzi w układzie aktywnym. Z powyższych rozważań wynika, że uni-struktura układu aktywnego $G^v(t)|^a$ zawiera tyle samo gałęzi jak uni-struktura układu pasywnego $G^v(t)|^p$.

Uni-struktury $G^v(t)|^a$ układów aktywnych będziemy tworzyli na bazie uni-struktur kanonicznych $\underline{G^v(t)}|^p$ układów pasywnych rozpatrując wszystkie możliwe permutacje włączenia nulatora i noratora do danej uni-struktury.

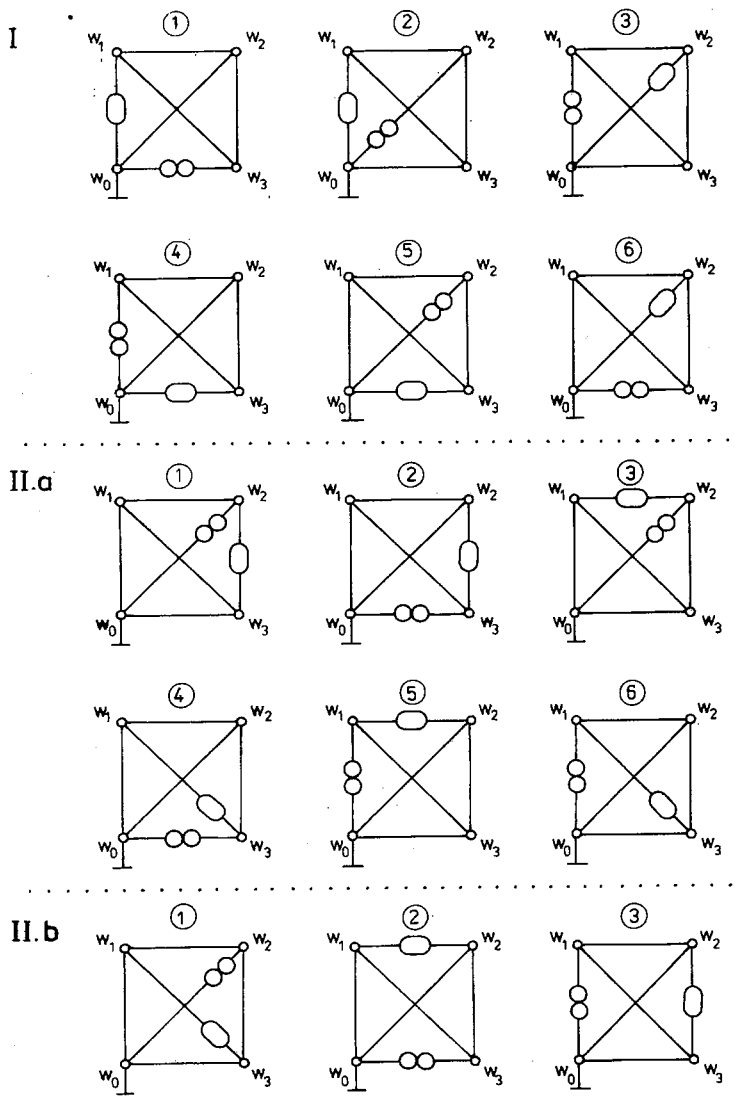
Na rysunku 1 zilustrowano proces włączania nulatora i noratora do tzw. pełnej uni-struktury kanonicznej $\underline{G^3(t)}|^p$ układu pasywnego. Moc otrzymanego zbioru



Rys. 1. Wariancje włączania nulatora i noratora do pełnej uni-struktury $G^3(t)|^p$ układu pasywnego

układów pretendentów wynosi $M = l_{emax} \cdot (l_{emax} - 1)$. Taka metoda włączania nulatora i noratora do układu pasywnego jest mało efektywna, ponieważ generuje wiele układów, które nie mogą być fizycznie zrealizowane. Uwzględniając ograniczenia wynikające z rozważań poprzedniego rozdziału, można zaproponować nieco inne podejście do zagadnienia syntezy uni-struktur kanonicznych $\underline{G^v(t)}|^a$.

Nulator zgodnie z rozważaniami pracy [11] jest zawsze włączany między wierzchołkiem odniesienia $\{w_0\}$ a dowolnym wierzchołkiem uni-struktury $\{w_h\}$, czyli na $(v-1)$ sposobów. Norator natomiast może być włączany jako dowolna gałąź uni-struktury układu, czyli na $(l_{emax}-1)$ sposobów. W tym przypadku



Rys. 2. Wariancje włączania nulatora i noratora do pełnej uni-struktury $G^4(t)|^p$ układu pasywnego z ograniczeniami sformułowanymi w pracy [11]

moc otrzymanego zbioru układów pretendentów wynosi $M=(v-1)\cdot(l_{emax}-1)$ (gdzie: v — liczba wierzchołków uni-struktury, l_{emax} — liczba gałęzi w pełnej uni-strukturze układu). Na rysunku 2 zilustrowano proces włączania nulatora i noratora do tzw. pełnej uni-struktury kanonicznej $\underline{G^4(t)}|^p$ czwórnika pasywnego z zastrzeżeniami sformułowanymi w pracy [11].

Ad. 2. Tworzenie bi-struktur $G^v(t_1,t_0)|^a$ w poszczególnych klasach uni-struktur kanonicznych $\underline{G^v(t)}|^a$ oraz ich podział na klasy struktur kanonicznych $\underline{G^v(t_1,t_0)}|^a$.

Jedną z funkcji celu w procesie syntezy, jest generacja struktur kanonicznych układów, które realizowałaby parametry z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij} w postaci funkcji wymiernej o maksymalnym stopniu licznika i mianownika przy minimalnej liczbie elementów tzw. struktury minimalno-elementowe.

W pracy [10] wykazano, że minimalno-elementowe struktury czwórników i trójnikóow pasywnych można utworzyć jako sumy mnogościowe dwóch rodzajów dendrytów $T^v_{k1}(t_1)$ i $T^v_{k2}(t_0)$. Ponieważ parametry F_{ij} czwórników i trójników aktywnych są opisywane poprzez sumy $H^v_{k-1}(t_1,t_2)$ dendrytów $T^v_{(k-1)1}(t_1)$ i $T^v_{(k-1)0}(t_2)$ co można zapisać w postaci:

$$T^v_{(k-1)i}(t_i)\,\underline{G^v(t)}|^a = T^v_{(k-1)i}(t_i) \tag{6}$$

$$\underline{G^v(t_1,t_0)}|^a = T^v_{(k-1)1}(t_1)\cup T^v_{(k-1)0}(t_0)\,. \tag{7}$$

Równanie (6) daje odpowiedź na pytanie, czy dendryty $T^v_{(k-1)i}(t_i)$ zawierają te same gałęzie co uni-struktura kanoniczna $\underline{G^v(t)}|^a$ układu pasywnego.

Równanie (7) daje odpowiedź na pytanie czy utworzona bi-struktura $G^v(t_1,t_2)|^p$ zawiera wszystkie gałęzie rozpatrywanej ui-struktury kanonicznej $\underline{G^v(t)}|^a$.

4. PRZYKŁAD SYNTEZY STRUKTUR KANONICZNYCH UKŁADÓW ZAWIERAJĄCYCH IDEALNY WZMACNIACZ OPERACYJNY

Syntezę bi-struktur kanonicznych układów aktywnych zilustrujemy na przykładzie układów zawierających dwa rodzaje elementów $\{g_i,\ G_jx\}$ i realizujacych parametr $z_{21}(x)$.

Procedurę „wypełnienia" jednego przedstawiciela z każdej klasy uni-struktur kanonicznych trójników dendrytami 2-drzewowymi $T^4_2(G_jx)$ i $T^4_2(g_i)$ zilustrujemy na przykładzie uni-struktur $\mathbf{G}\{G^4(t)\}=[\infty,1,1,1,o,1]$ i $\mathbf{G}\{G^4(t)\}=[1,o,1,\infty,1,1]$. Parametr $z_{21}(x)$ układu pasywnego z włączonym według drugiego sposobu (rys. 1e z pracy [11]) idealnym wzmacniaczem operacyjnym, (zgodnie z tab. 4 z pracy [11]), można zapisać w postaci:

$$z_{21}(x)=\frac{H^v_{hj,1,2,0}(x)-H^v_{hi,1,2,0}(x)}{H^v_{hj,i\,0}(x)-H^v_{hi\,,j\,0}(x)}\,. \tag{9}$$

Dla uni-struktury trójnika pokazanego w tab. 1 ($h=3$, $i=2$, $j=1$) strukturalna macierz ma postać:

$$\mathbf{Q}=\begin{vmatrix} e_1+e_3 & -e_1 & -e_3 & 0 \\ -e_1 & e_1+e_5 & 0 & -e_5 \\ -e_3 & 0 & e_3+e_4 & -e_4 \\ 0 & -e_5 & -e_4 & e_4+e_5 \end{vmatrix}. \tag{10}$$

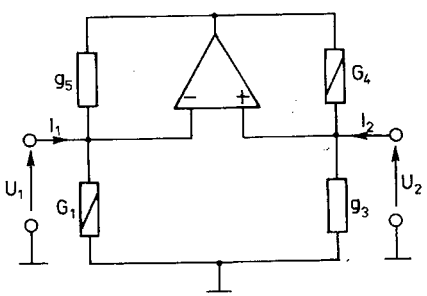
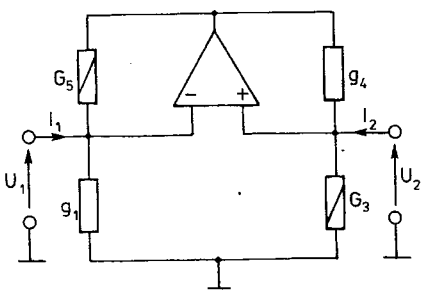
Dla rozpatrywanej uni-struktury $G^4(t) \mid^a$ strukturalny parametr z_{21} można zapisać w postaci:

$$z_{21}=\frac{e_4}{e_3 \cdot e_5 - e_1 \cdot e_4}. \tag{11}$$

W ogólnym przypadku element $\{e_l\}$ może przyjmować wartości $e_l=\{0 \vee g_l \vee G_l x \vee (g_l+G_l x)\}$. Minimalno-elementowe bi-struktury $G^4(t_1,t_2) \mid^a$ trójników, zrealizowane jako sumy mnogościowe dendrytów 2-drzewowych $T^4_{32,10}(G_j x)$ i $T^4_{32,10}(g_i)$ w uni-strukturze typu $\mathbf{G}\{G^4(t)\}=[\infty,1,1,1,o,1]$, przedstawiono w tab. 1.

Tabela 1

Minimalno-elementowe bi-struktury kanoniczne trójnikow aktywnych utworzone w uni-strukturze $\mathbf{G}\{G^4(t)\}=[\infty,1,1,1,o,1]$ i realizujące parametr $z_{21}(x)$

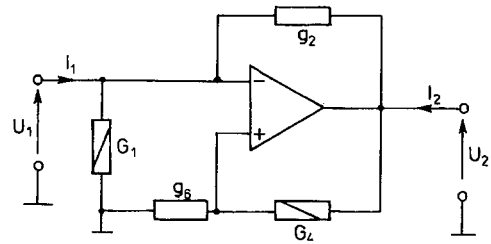
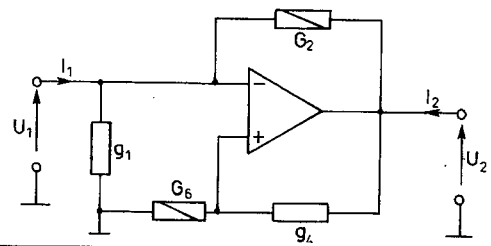
Lp	Realizowana funkcja	Struktura układu
1	$z_{21}(x)=\dfrac{G_4 x}{-G_1 G_4 x^2+g_3 g_5}$	
2	$z_{21}(x)=\dfrac{g_4}{G_3 G_5 x^2-g_1 g_4}$	

Dla uni-struktury trójnika pokazanego w tab. 2 ($h=2$, $i=3$, $j=1$) strukturalna macierz ma postać:

$$\mathbf{Q}=\begin{vmatrix} e_1+e_6 & -e_1 & 0 & -e_6 \\ -e_1 & e_1+e_2 & -e_2 & 0 \\ 0 & -e_2 & e_2+e_4 & -e_4 \\ -e_6 & 0 & -e_4 & e_4+e_6 \end{vmatrix}. \tag{12}$$

Tabela 2

Minimalno-elementowe bi-struktury kanoniczne trójników aktywnych utworzone w uni-strukturze $\mathbf{G}\{G^4(t)\}=[1,o,1,\infty,1,1]$ i realizujące parametr $z_{21}(x)$

Lp	Realizowana funkcja	Struktura układu
1	$z_{21}(x)=\dfrac{G_4x+g_6}{-G_1G_4x^2+g_2g_6}$	
2	$z_{21}(x)=\dfrac{G_6x+g_4}{G_2G_6x^2-g_1g_4}$	

W tym przypadku strukturalny parametr z_{21} można zapisać następująco:

$$z_{21}=\frac{e_4+e_6}{e_1\cdot e_4-e_2\cdot e_6}, \tag{13}$$

gdzie:

$$e_l=\{0\vee g_l\vee G_lx\vee(g_l+G_lx)\}.$$

Minimalno-elementowe bi-struktury $G^4(t_1,t_2)|^a$ trójników, zrealizowane jako sumy mnogościowe dendrytów 2-drzewowych $T^4_{23,10}(G_jx)$ i $T^4_{23,10}(g_i)$ w uni-strukturze typu $\mathbf{G}\{G^4(t)\}=[1,o,1,\infty,1,1]$, prezentuje tab. 2.

Tworząc bi-struktury układów jako sumy mnogościowe dendrytów 1- i 2-drzewowych, można uzyskać o wiele większą różnorodność funkcji wymiernych opisujących dany parametr układu. Możliwe klasy funkcji wymiernych w postaci $z_{21}(x)$, realizowane przez bi-struktury $G^4(x)=T^4_{ki}(g_i)\cup T^4_{kj}(G_jx)$ (gdzie: $k_i=1\vee 2$ oraz $k_j=1\vee 2$) utworzone w uni-strukturach $\mathbf{G}\{G^4(t)\}=[\infty,1,1,1,o,1]$ i $\mathbf{G}\{G^4(t)\}=[1,o,1,\infty,1,1]$, zestawiono odpowiednio w tab. 3 i 4. Do danej klasy struktur kanonicznych będziemy zaliczać te struktury, których parametr $z_{21}(x)$ realizuje dany typ funkcji wymiernej.

Sposób zapisu parametru $z_{21}(x)$ w postaci funkcji wymiernej przedstawimy na przykładzie bi-struktury o numerze 16 z tab. 3, dla której parametr $z_{21}(x)$ ma postać:

$$z_{21}(x)=\frac{G_4x+g_4}{G_3G_5x^2+(g_3G_5-g_1G_4)x-g_1g_4},$$

Przyklady struktur kanonicznych trójników aktywnych utworzone w uni-strukturze $\mathbf{G}\{G^4(t)\}=[o,1,1,1,\infty,1]$ i realizujące parametr $z_{21}(x)$

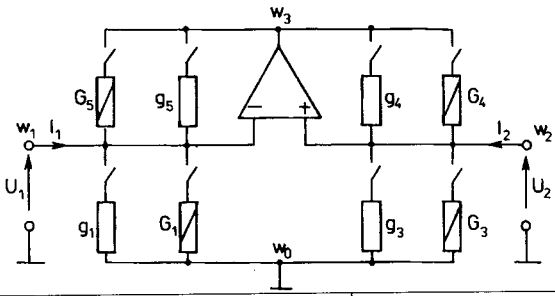
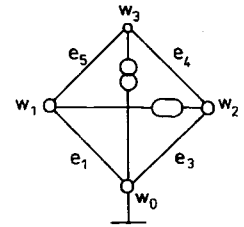
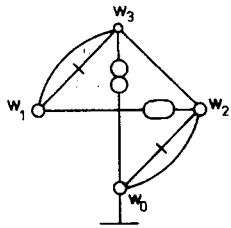
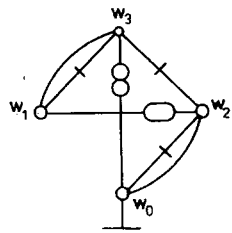
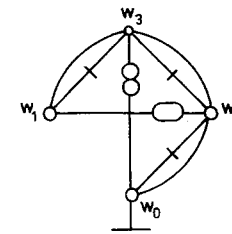
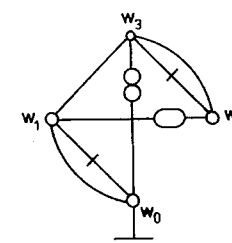
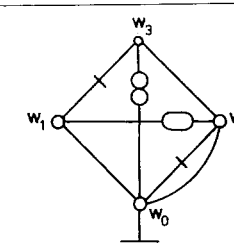
Lp		
		
1	$z_{21}(x)=\dfrac{a_0}{b_2x^2+b_1x+b_0}$	
2	$z_{21}(x)=\dfrac{a_1x}{b_2x^2+b_1x+b_0}$	
3	$z_{21}(x)=\dfrac{a_1x+a_0}{b_2x^2+b_1x+b_0}$	
4	$z_{21}(x)=-\dfrac{a_1x+a_0}{b_2x^2+b_1x+b_0}$	
5	$z_{21}(x)=\dfrac{a_0}{b_2x^2+b_1x-b_0}$	

Tabela 3 c.d.

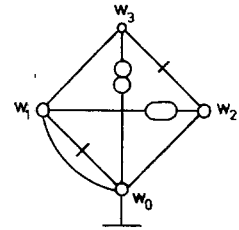
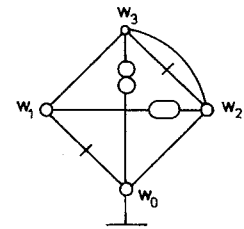
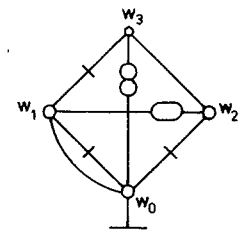
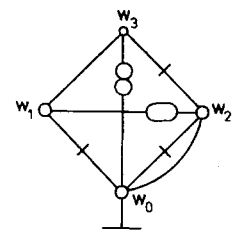
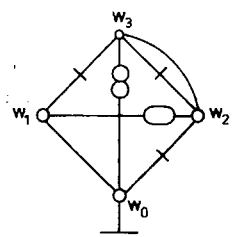
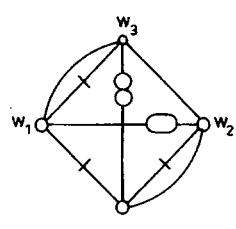
6	$z_{21}(x)=\dfrac{a_1x}{b_2x^2+b_1x-b_0}$	
7	$z_{21}(x)=\dfrac{a_1x+a_0}{b_2x^2+b_1x-b_0}$	
8	$z_{21}(x)=\dfrac{a_0}{b_2x^2-b_1x-b_0}$	
9	$z_{21}(x)=-\dfrac{a_1x}{b_2x^2-b_1x-b_0}$	
10	$z_{21}(x)=-\dfrac{a_1x+a_0}{b_2x^2-b_1x-b_0}$	
11	$z_{21}(x)=\dfrac{a_0}{b_2x^2+(b_1^+-b_1^-)x+b_0}$	

Tabela 3 c.d.

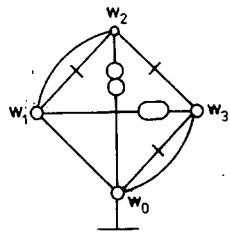
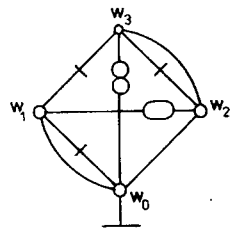
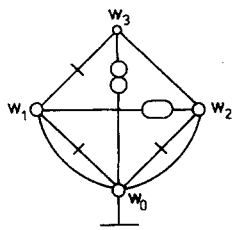
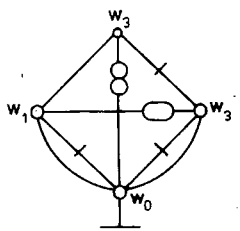
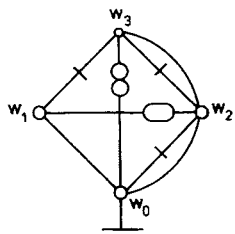
12	$z_{21}(x)=\dfrac{a_1x}{b_2x^2+(b_1^+-b_1^-)x+b_0}$	
13	$z_{21}(x)=\dfrac{a_1x+a_0}{b_2x^2+(b_1^+-b_1^-)x+b_0}$	
14	$z_{21}(x)=\dfrac{a_0}{b_2x^2+(b_1^+-b_1^-)x-b_0}$	
15	$z_{21}(x)=\dfrac{a_1x}{b_2x^2+(b_1^+-b_1^-)x-b_0}$	
16	$z_{21}(x)=\dfrac{a_1x+a_0}{b_2x^2+(b_1^+-b_1^-)x-b_0}$	

Tabela 4

Przykłady struktur kanonicznych trójników aktywnych utworzone w uni-strukturze $\mathbf{G}\{G^4(t)\} = [1,o,1,\infty,1,1]$ i realizujące parametr $z_{21}(x)$

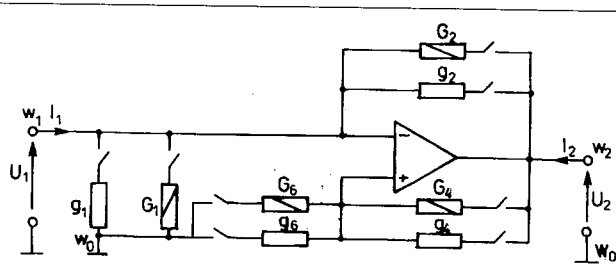
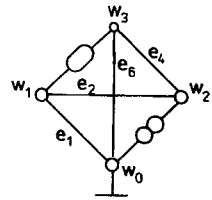
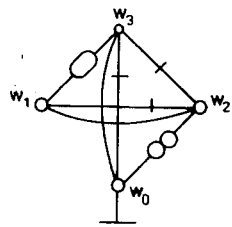
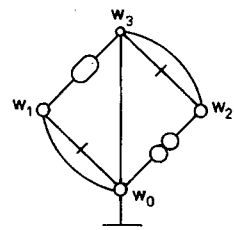
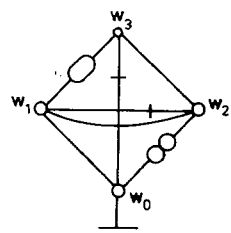
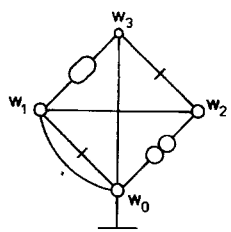
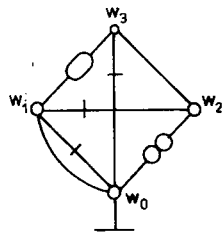
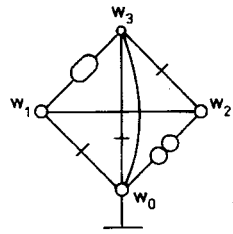
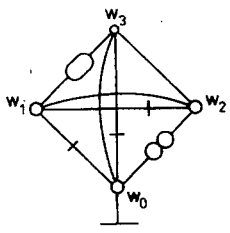
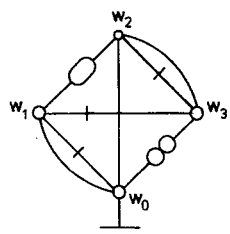
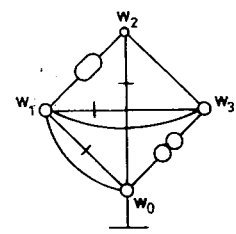
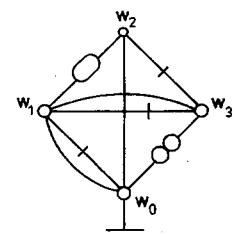
Lp		
		
1	$z_{21}(x) = \dfrac{a_1x + a_0}{b_2x^2 + b_1x + b_0}$	
2	$z_{21}(x) = -\dfrac{a_1x + a_0}{b_2x^2 + b_1x + b_0}$	
3	$z_{21}(x) = \dfrac{a_1x + a_0}{b_2x^2 + b_1x - b_0}$	
4	$z_{21}(x) = -\dfrac{a_1x + a_0}{b_2x^2 + b_1x - b_0}$	

Tabela 4 c.d.

5	$z_{21}(x)=\dfrac{a_1x+a_0}{b_2x^2-b_1x-b_0}$	
6	$z_{21}(x)=-\dfrac{a_1x+a_0}{b_2x^2-b_1x-b_0}$	
7	$z_{21}(x)=\dfrac{a_1x+a_0}{b_2x^2+(b_1^+-b_1^-)x+b_0}$	
8	$z_{21}(x)=-\dfrac{a_1x+a_0}{b_2x^2+(b_1^+-b_1^-)x+b_0}$	
9	$z_{21}(x)=\dfrac{a_1x+a_0}{b_2x^2+(b_1^+-b_1^-)x-b_0}$	
10	$z_{21}(x)=-\dfrac{a_1x+a_0}{b_2x^2+(b_1^+-b_1^-)x-b_0}$	

co w ogólnym przypadku można zapisać jako:

$$z_{21}(x)=\frac{a_1x+a_0}{b_2x^2+(b_1^+-b_1^-)x-b_0}.$$

Należy zwrócić uwagę, że dla takiego sposobu włączenia wzmacniacza operacyjnego w macierzy $\mathbf{Y}_w^a$ pojawią się iloczyny admitancji z różnymi znakami. Zastosowanie takich struktur układów aktywnych pozwala nie tylko na rozszerzenie funkcjonalnych możliwości generowanych struktur kanonicznych układów aktywnych, ale również na osiągnięcie zupełnego zbioru struktur z punktu widzenia położenia zer i biegunów realizowanych funkcji wymiernych F_{ij}.

5. ALGORYTM GENERACJI STRUKTUR KANONICZNYCH UKŁADÓW ZAWIERAJĄCYCH IDEALNY WZMACNIACZ OPERACYJNY

Można zaproponować nieco inne podejście do zagadnienia sposobów „rozmieszczenia" nulatora i noratora w uni-strukturze kanonicznej $\underline{G^v(t)}|^p$. Dla układów o $v>4$ (dla czwórników) i $v>3$ (dla trójników) wierzchołkach moce generowanych zbiorów uni-struktur układów pretendentów $G^v(t)|^a$, po uwzględnieniu ograniczeń omówionych w pracy [11], można jeszcze dodatkowo ograniczyć biorąc pod uwagę następujące fakty.

Ogólnie wierzchołki rozpatrywanych struktur układów można podzielić na dwie grupy:

1. Wierzchołki brzegowe układów (dla czwórników $\{w_1,w_0\}$ i $\{w_2,w_3\}$, dla trójników $\{w_1,w_0\}$ i $\{w_2,w_0\}$).

2. Wierzchołki wewnętrzne układów $\{w_l\}$ (dla czwórników $l>4$, dla trójników $l>3$). Nulator zgodnie z rozważaniami rozdziału szóstego powinien być zawsze włączany między wierzchołkiem odniesienia $\{w_0\}$ a dowolnym wierzchołkiem uni-struktury $\{w_h\}$, czyli na $(v-1)$ sposobów.

Liczbę sposobów włączenia nulatora do układu można dodatkowo ograniczyć biorąc pod uwagę fakt, że przenumerowanie wierzchołków wewnętrznych układu nie zmienia parametrów rozpatrywanych układów.

Dwie uni-struktury $G^v(t)|_I^a$ i $G^v(t)|_{II}^a$ z dołączonym nulatorem do dowolnego wierzchołka wewnętrznego (w_h) pełnej uni-struktury są D-równoważne, ponieważ występujące w poszczególnych macierzach $\mathbf{D}\{T_{k-1}^v(t)\}|_I^a$ i $\mathbf{D}\{T_{k-1}^v(t)\}|_{II}^a$ dendryty $(k-1)$-drzewowe są określane tylko przez wierzchołki brzegowe i wierzchołek $\{w_h\}$ (tabele 3 i 4 z pracy [11]). Wynika stąd, że nulator można włączyć do układu czwórnika na cztery sposoby, natomiast do układu trójnika na trzy sposoby. Norator zgodnie z rozważaniami pracy [11], może być włączany jako dowolna gałąź uni-struktury układu, czyli na $(l_{\max}-1)$ sposobów. Liczbę sposobów włączenia noratora do układu można dodatkowo ograniczyć biorąc pod uwagę fakt, że przenumeracja wierzchołków wewnętrznych układu nie zmienia parametrów rozpatrywanych układów.

Mając włączony nulator do struktury, norator według pierwszego sposobu (rys. 1 d z pracy [11]) można włączyć na trzy sposoby do układu czwórnika, natomiast na dwa sposoby do układu trójnika.

Mając włączony nulator do struktury, norator według drugiego sposobu (rys. 1 e z pracy [11]) można włączyć na trzy sposoby do układu czwórnika, natomiast na dwa sposoby do układu trójnika.

Mając włączony nulator do struktury, norator według trzeciego sposobu (rys. 1 f z pracy [11]) można włączyć na trzy sposoby do układu czwórnika, natomiast na jeden sposób do układu trójnika.

Ogólnie mając włączony nulator do struktury, norator można włączyć na dziewięć sposobów do układu czwórnika, natomiast na pięć sposobów do układu trójnika. Tak więc w ogólnym przypadku dla czwórników o $v>4$ wierzchołkach, nulator i norator do pełnej uni-struktury czwórnika można włączyć na $4*9=36$ sposobów. Natomiast w ogólnym przypadku dla trójnikow o $v>3$ wierzchołkach, nulator i norator do pełnej uni-struktury trójnika można włączyć na $3*5=15$ sposobów.

6. PODSUMOWANIE

W pracy przedstawiono syntezę struktur kanonicznych układów pasywnych zawierających idealny wzmacniacz operacyjny. Do najważniejszych rezultatów pracy można zaliczyć:

1. Opracowanie niezmienników D-równoważności struktur układów

Wykazano, że niezmiennikami D-równoważnści struktur układów elektronicznych są macierze dendrytów k-drzewowych zdefiniowane w pracy [9]. Dwie uni-struktury $G^v(t)|_I^a$ i $G^v(t)|_{II}^a$, multi-struktury $G^v(tm)|_I^a$ i $G^v(tm)|_{II}^a$ lub bi-struktury $G^v(t_1,t_2)|_I^a$ i $G^v(t_1,t_2)|_{II}^a$ zawierające idealny wzmacniacz operacyjny, są D-równoważne wtedy, gdy:

$$\mathbf{D}\{T_1^v(t)\}|_I^a \Leftrightarrow \mathbf{D}\{T_1^v(t)\}|_{II}^a,$$

$$\mathbf{D}\{T_1^v(tm)\}|_I^a \Leftrightarrow \mathbf{D}\{T_1^v(tm)\}|_{II}^a,$$

$$\mathbf{D}\{T_1^v(t_1,t_2)\}|_I^a \Leftrightarrow \mathbf{D}\{T_1^v(t_1,t_2)\}|_{II}^a.$$

2. Opracowanie algorytmu syntezy struktur kanonicznych układów zawierających idealny wzmacniacz operacyjny.

Wykazano, że układ aktywny zawierający idealny wzmacniacz operacyjny można traktować jako układ pasywny z ograniczeniami związanymi z włączeniem noratora i nulatora, modelujących idealny wzmacniacz operacyjny [11].

Opracowany w pracy [10] algorytm syntezy struktur kanonicznych układów pasywnych stanowi podstawę algorytmu syntezy struktur kanonicznych układów aktywnych z jednym idealnym wzmacniaczem operacyjnym. Można wykazać, że algorytm syntezy struktur kanonicznych układów z jednym wzmacniaczem operacyjnym pozwala na opracowanie algorytmu syntezy struktur kanonicznych układów aktywnych z dwoma wzmacniaczami operacyjnymi. Korzystając z zasad indukcji matematycznej, dalej można generować kolejne struktury kanoniczne układów zawierających $(w+1)$ wzmacniaczy operacyjnych.

BIBLIOGRAFIA

1. M. Białko, A. Guziński, W. Sienko, J. Zaruda: *Filtry aktywne RC.* WNT, Warszawa 1979
2. L. Lasek: *Analiza i synteza układów generacyjnych RC.* Zeszyty Naukowe, Pol. Śląskiej, 1988, Automatyka Nr 91, Gliwice 1988
3. S.K. Mitra: *Analiza i synteza układów aktywnych liniowych.* WNT, Warszawa 1974
4. S. Osowski: *Zagadnienia minimalizacji wzmacniaczy operacyjnych w syntezie układów wielowejściowych.* Prace naukowe Politechniki Warszawskiej, Elektryka, z. 58, Warszawa 1978
5. A.G. Ostapienko: *Analiz i sintez liniejnych radioelektronnych cepiej s pomoszcziu grafów.* Radio i Swiaz, Moskwa 1985
6. Z. Wróbel: *Funkcje strukturalne grafów układów dwójników i czwórników składających się z dwóch rodzajów elementów.* Kwart. Elektr. i Telekom., 1991, t. 37, z. 1–2, s. 35–52
7. Z. Wróbel: *Synteza struktur kanonicznych bi-grafów układów dwójnikow składających się z dwóch rodzajów elementów. Część I. Metoda przeglądu bezpośredniego.* Kwart. Elektr. i Telekom., 1991, t. 36, z. 1–2, s. 67–83
8. Z. Wróbel: *Synteza struktur kanonicznych bi-grafów układów dwójników składających się z dwóch rodzajów elementów. Część II. Metoda przeglądu hierarchicznego.* Kwart. Elektr. i Telekom., 1991, t. 37, z. 1–2, s. 85–97
9. Z. Wróbel: *D-równoważność struktur układów realizujących funkcje wymierne.* Kwart. Elektr. i Telekom., 1993, t. 39, z. 3, ss. 523–539
10. Z. Wróbel: *Synteza struktur kanonicznych układów czwórników składających się z dwóch rodzajów elementów.* Kwart. Elektr. i Telekom., 1993, t. 39, z. 3, ss. 541–559
11. Z. Wróbel: *Warianty włączenia wzmacniacza operacyjnego do układu pasywnego.* Kwart. Elektr. i Telekom., 1994, t. 40, z. 4, ss. 511–528

Z. WRÓBEL

SYNTHESIS OF CANONICAL STRUCTURE OF FOUR- AND THEREE TERMINAL NETWORKS CONTAINING AN IDEAL OPERATIONAL AMPLIFIER

Summary

Procedure of topological synthesis, allowing to generate all possible of canonical structure of three- and four-terminal networks containing an ideal operational amplifier was shown in the paper. Condition of realizing minimal-elements canonical structure of three- and four-terminal networks containing an ideal operational amplifier.

Key words: three- and four-terminal networks, canonical structure, operational amplifier.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1995. 41, z. 2, ss. 185—201

Właściwości optyczne dielektrycznych struktur stożkowych

KRZYSZTOF PERLICKI

Instytut Telekomunikacji, Politechnika Warszawska

Otrzymano 1995.05.05

Autoryzowano do druku 1995.06.14

W artykule zaprezentowano analizę propagacji światła o różnej długości fali w stożkowych strukturach dielektrycznych w ujęciu teorii sprzężeń modowych. Przeanalizowano przepływ mocy zachodzący między modem podstawowym HE_{11} i modem HE_{12} w wielomodowych, bezstratnych strukturach. Przedstawiono właściwości złącz optycznych o konstrukcji opartej na strukturach stożkowych.

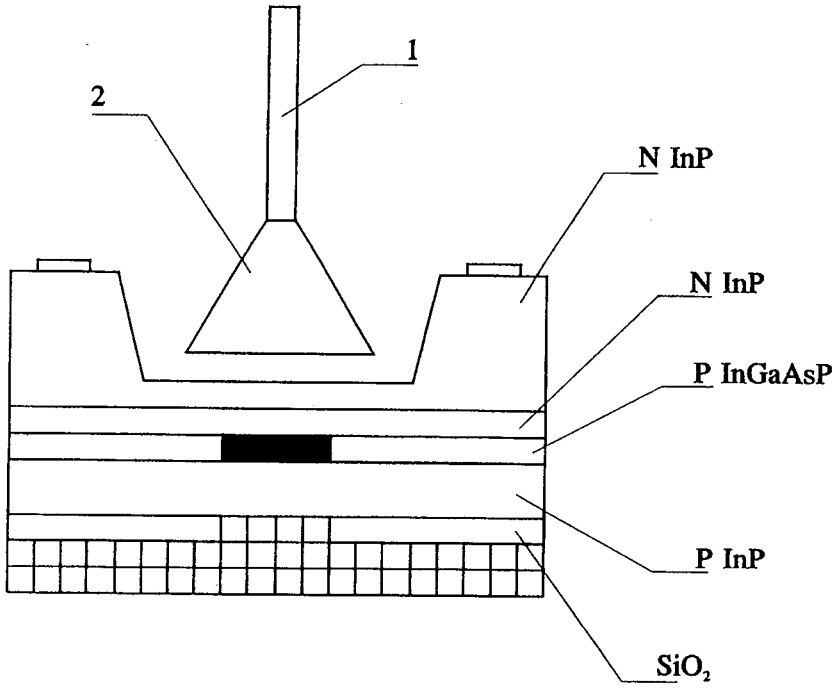
Słowa kluczowe: światłowody, stożki szklane, złącza optyczne.

1. WPROWADZENIE

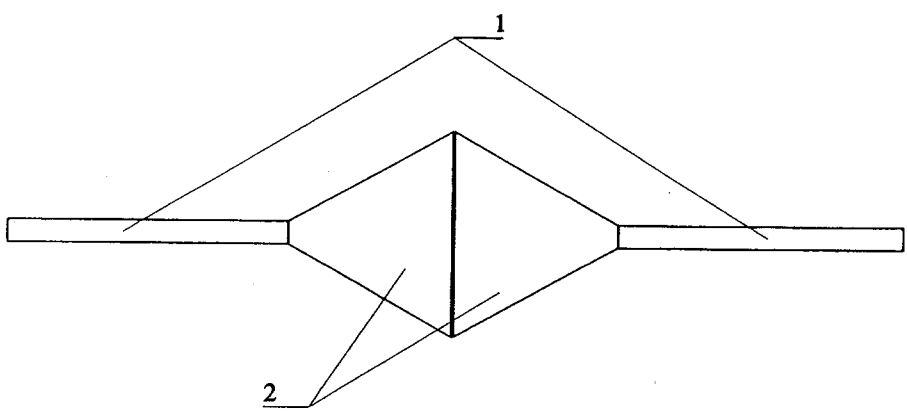
Cylindryczne, stożkowe struktury dielektryczne (światłowodowe) są elementami, które znalazły szerokie zastosowanie w technice światłowodowej. Za ich pomocą można na przykład realizować wygodne, o korzystnych parametrach połączenia pomiędzy źródłem światła (na przykład diodą LED) a światłowodem. Ze względu na dość dużą kątową rozbieżność wiązki wychodzącej z diody, struktury stożkowe pozwalają na poprawę sprawności sprzężenia powierzchni świecącej z włóknem optycznym (rysunek 1).

Stożki szklane stosuje się również w niektórych konstrukcjach sprzęgaczy typu gwiazda z użyciem mieszaczy optycznych. Spełniają one w nich rolę elementu wyrównującego poprzeczny rozkład mocy optycznej. Tego typu struktury pozwalają także na realizację wygodnych złącz optycznych (rysunek 2) o mniejszych stratach powodowanych przez niedokładne wzajemne ustawienie osi łączonych światłowodów i powierzchni czołowych. Szczególnie ma to znaczenie przy złączach dla światłowodów jednomodowych, gdzie ze względu na niewielkie geometryczne rozmiary rdzenia ($5-10$ μm), wszelkie niedokładności we wzajemnym ustawieniu czół włókien powodują znaczne straty energii.

Analiza sprzężeń modowych w stożkach dielektrycznych umożliwia dokładne określenie ich własności optycznych i odpowiedni dobór kształtu tego typu struktur



Rys. 1. Przykładowy sposób sprzężenia powierzchniowej diody elektroluminescencyjnej z włóknem optycznym z wykorzystaniem stożka dielektrycznego. 1 — światłowód, 2 — dielektryczna strutkura stożkowa



Rys. 2. Złącze o konstrukcji opartej na strukturach stożkowych. 1 — światłowód, 2 — stożek dielektryczny

pod kątem konkretnego ich zastosowania. Analiza propagacji światła o różnej długości fali pozwala na określenie zachowania się tych elementów w systemach transmisji z podziałem według długości fali (WDM).

Badanie własności optycznych stożków dielektrycznych ma również znaczenie w zrozumieniu mechanizmu widzenia tak u człowieka jak i u innych wyższych

organizmów. Fotoreceptory tworzące siatkówkę oka składają się bowiem ze struktur stożkowych mających własności optyczne. Występują w nich sprzężenia międzymodowe; co, jak się uważa, może odgrywać rolę w mechanizmie widzenia barwnego [2, 3].

2. OPIS SPRZĘŻEŃ MODOWYCH

Przedmiotem analizy są sprzężenia międzymodowe zachodzące w dielektrycznych, bezstratnych strukturach stożkowych. W badaniach skoncentrowano się na zjawisku sprzężeń zachodzących w obszarze rdzenia. Założono, że nie występuje przekazywanie mocy modom o kierunku przeciwnym oraz założono brak sprzężeń pomiędzy modami rdzeniowymi a modami radiacyjnymi. Ostatnie z wymienionych założeń jest słuszne pod warunkiem, że nie poddajemy analizie modu dla częstotliwości bliskiej jego częstotliwości odcięcia. W takich bowiem warunkach ma miejsce silny transfer mocy z modu odciętego do modów radiacyjnych przy równoczesnym bardzo małym przepływie mocy do innych modów rdzeniowych. Jeżeli analizę ograniczymy do przypadku gdy $\delta < 0.1$, gdzie:

$$\delta = 1 - \frac{n_2^2}{n_1^2} \tag{1}$$

n_1 — współczynnik załamania materiału stożka, n_2 — współczynnik załamania otaczającego ośrodka, to sprzężenie (przepływ mocy) zachodzi jedynie między modem podstawowym HE_{11} i modem HE_{12}. Sprzężenie dla kolejnego modu jest rzędu δ (dla standardowych światłowodów telekomunikacyjnych $\delta \approx 0.01$). Równanie opisujące sprzęganie się N-modów dla q-tego modu ma następującą postać [4, 9]:

$$\frac{dA_q}{dz} + j\beta_q^* \cdot A_q = 0.5 \cdot \sum_{p=1}^{N \neq q} A_p \cdot (K_{pq} - K_{qp}^*) \tag{2}$$

$$K_{pq} = \int e_p \times \frac{\partial h_q}{\partial z} da, \tag{3}$$

gdzie:
$q = 1,...,N$-ty mod,
A_p, A_q — amplituda odpowiednio modu p i q,
β — stała propagacji,
e_p, h_q — pola modowe,
K_{pq} — współczynnik sprzężenia.

Stwierdzono, że przepływ mocy następuje tylko między modami z identyczną liczbą azymutalną oraz, że między modami typu TM a TE sprzężenia nie zachodzą [4, 10]. Ograniczając się do dwóch modów (HE_{11} i HE_{12}) równanie (2) możemy przekształcić do układu dwóch równań [5]:

$$\frac{dA_q}{dV}+j\gamma_q \cdot A_q = C_{pq} \cdot A_p \tag{4}$$

$$\frac{dA_p}{dV}+j\gamma_p \cdot A_p = -C_{pq} \cdot A_q, \tag{5}$$

gdzie:
V — częstotliwość znormalizowana:

$$V=\frac{2 \cdot \Pi}{\lambda} \cdot a \cdot \sqrt{n_1^2-n_2^2} \tag{6}$$

C_{pq} — współczynnik sprzężenia:

$$C_{pq}=(a/V \cdot \Omega) \cdot K_{pq} \tag{7}$$

$$\gamma_{pq}=(1/\Omega \cdot \sqrt{\delta})-\sqrt{\delta} \cdot U_q^2(\infty) \cdot e^{-2/V}/2 \cdot \Omega \cdot V^2 \tag{8}$$

przy czym:
a — promień stożka,
Ω — kąt nachylenia stożka,
λ — długość fali,
$U_q(\infty)=2.405$ (dla modu HE_{11}),
$U_q(\infty)=5.520$ (dla modu HE_{12}),
δ, n_1, n_2 — jak we wzorze (1),
K_{pq} — jak we wzorze (3).

Za [4, 5] możemy przyjąć, żę współczynnik sprzężenia C_{pq} jest równy 1/V. Daje to nam dobrą dokładność określenia jego wartości. Porównując z dokładnym rozwiązaniem numerycznym zamieszczonym w pracy [5] błąd nie przekracza, poza częstotliwością odcięcia, kilku procent.

3. WYNIKI

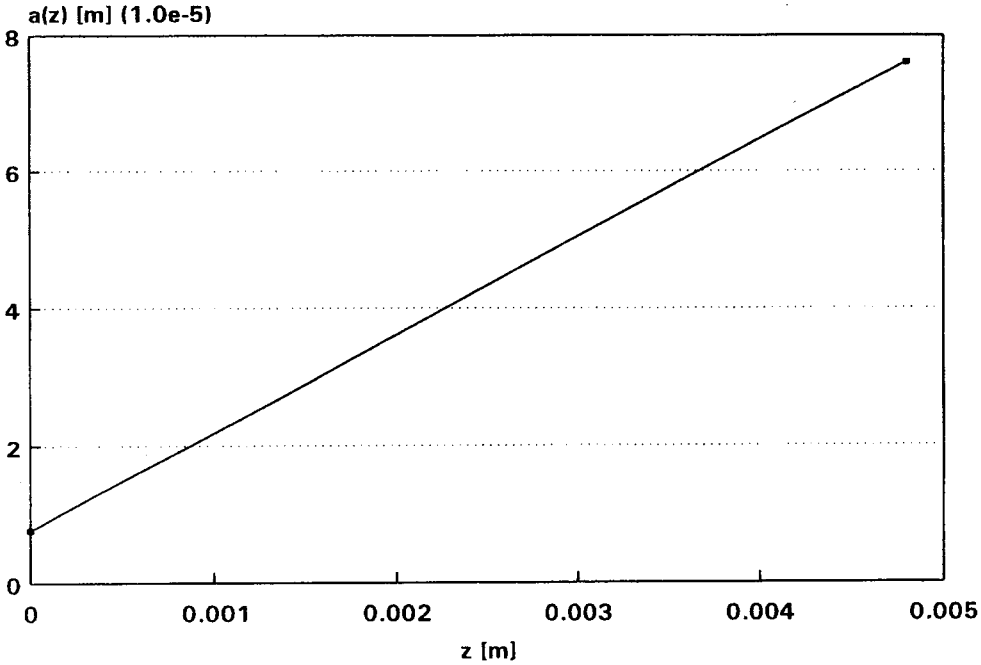
Przeanalizowano zjawisko sprzężenia modu podstawowego HE_{11} z modem HE_{12} dla trzech długości fal: 1300 nm, 1500 nm, i 1550 nm. Przebadano struktury stożkowe liniowe (rysunek 3) o następującej zależności promienia od długości:

$$a(z)=z \cdot tg\Omega + a_{\min} \tag{9}$$

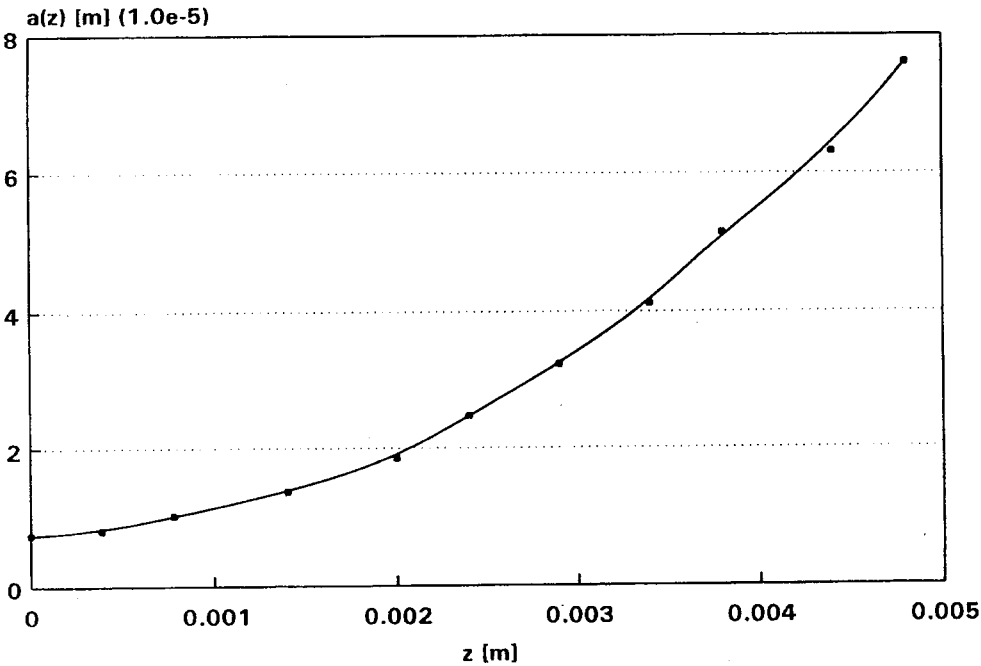
z — długość, $a_{\min}$ — promień wejściowy stożka, oraz struktury o nieliniowej zmianie promienia w funkcji długości, opisanej zaleznością:

$$a(z)=d \cdot z^{\alpha} + a_{\min} \tag{10}$$

gdzie:
d — stała,
$\alpha=2$ lub 0.5.

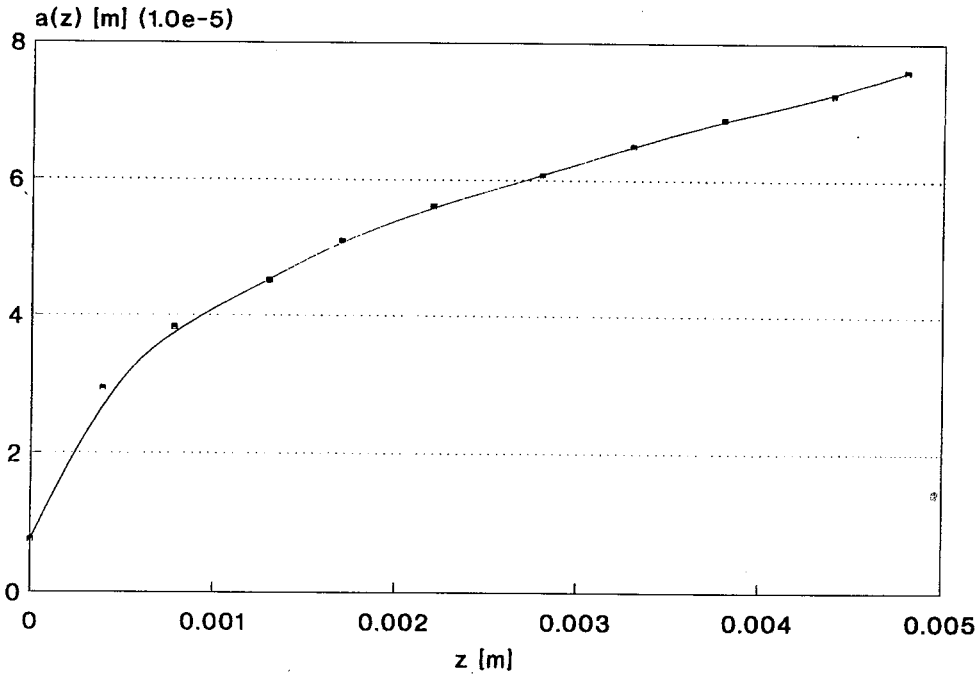


Rys. 3. Zmiana promienia stożka w funkcji jego długości dla struktury o kształcie opisanym zależnością (9)



Rys. 4. Zmiana promienia stożka w funkcji jego długości dla struktury o kształcie opisanym zależnością (10), przy $\alpha=2$

Na rysunku 4 przedstawiono zmianę promienia stożka w funkcji długości dla $\alpha=2$, a na rysunku 5 dla $\alpha=0.5$. Badane stożki miały maksymalny promień równy 76156 nm (odpowiada temu częstotliwość znormalizowana $V=40$ przy długości fali 1550 nm), a minimalny wynosił 7616 nm (odpowiada temu częstotliwość znormalizowana $V=4$ przy $\lambda=1550$ nm). Przyjęto, że do struktury wprowadzono mod HE_{11} o mocy jednostkowej.



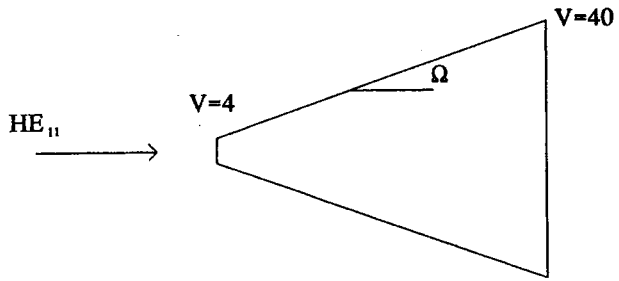
Rys. 5. Zmiana promienia stożka w funkcji jego długości dla struktury o kształcie opisanym zależnością (10), przy $\alpha=0.5$

3.1. LINIOWE STRUKTURY STOŻKOWE

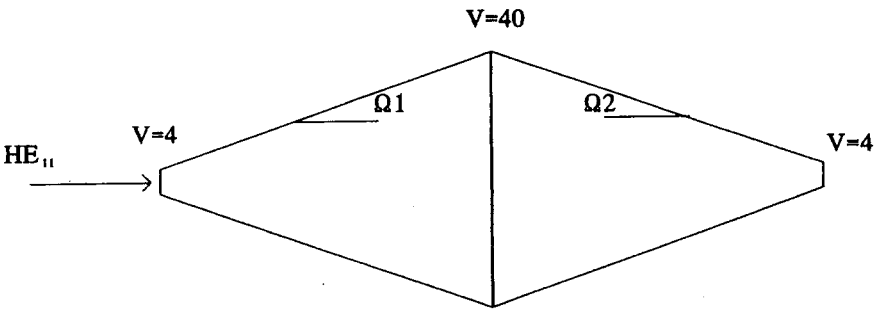
Badania przeprowadzono dla struktury rozszerzającej się (rysunek 6) i konstrukcji złączowej (rysunek 7).

3.1.1. Stożek rozszerzający się

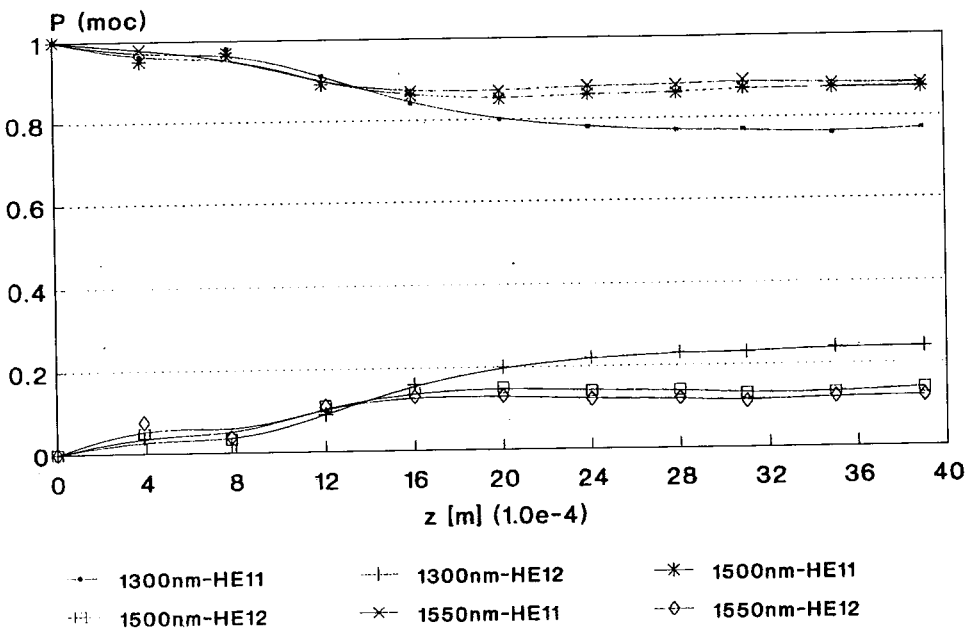
Przeanalizowano zjawisko sprzężeń międzymodowych w stożkach rozszerzających się dla czterech wartości kąta nachylenia, to jest dla kąta: 1°, 3°, 5° i 15°. Uzyskane wyniki dla różnych kątów pokazano odpowiednio na rysunku 8, 9, 10, 11.
Rysunek 12 przedstawia zmianę mocy niesionej przez mod HE_{12} w funkcji zmiany promienia stożka dla różnych kątów nachylenia (przy $\lambda=1550$ nm).



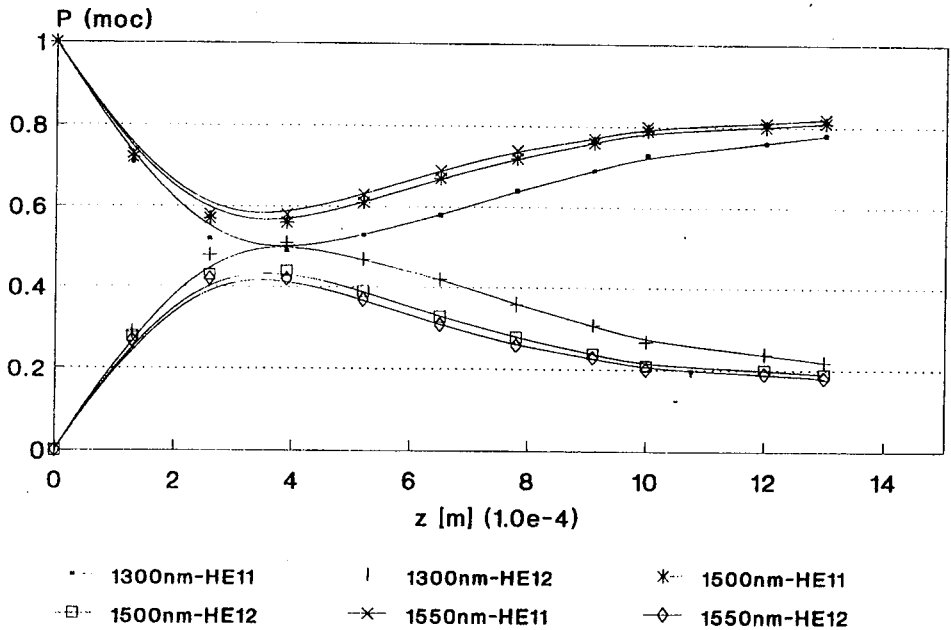
Rys. 6. Badana rozszerzająca się struktura stożkowa



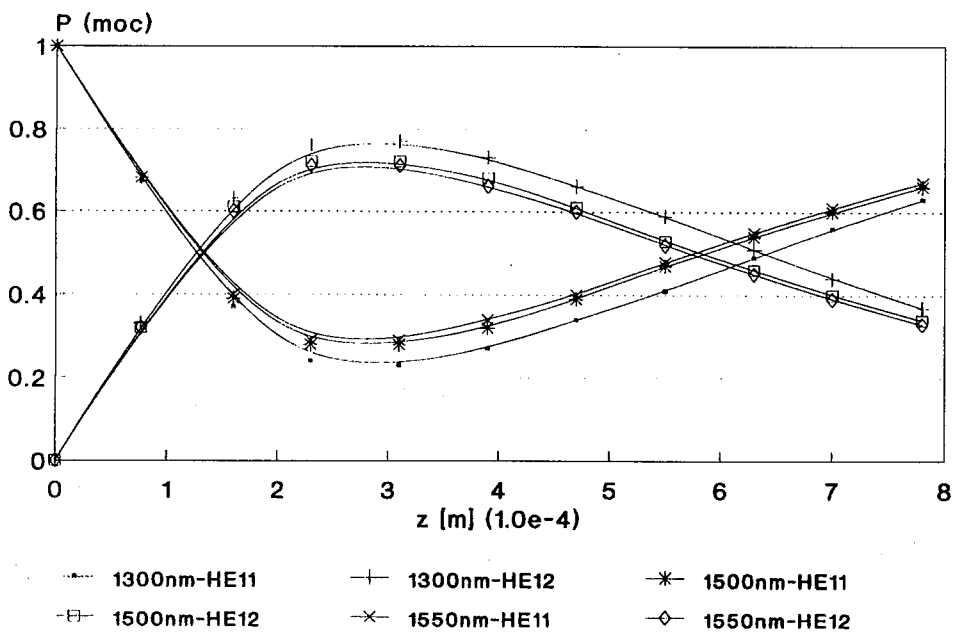
Rys. 7. Badana konstrukcja złączowa



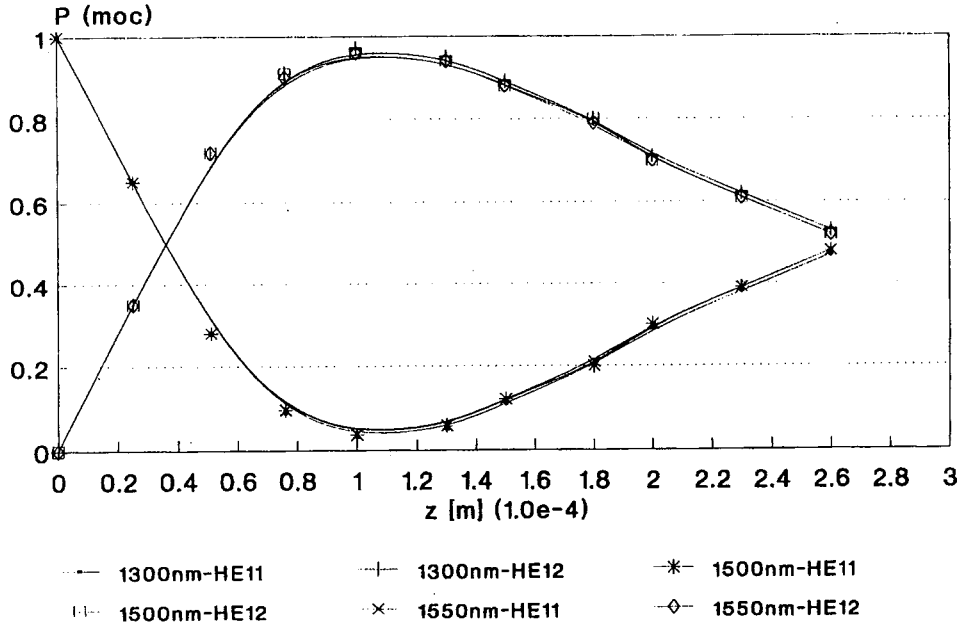
Rys. 8. Przebieg mocy dla modu HE_{11} i HE_{12} w funkcji długości z dla rozszerzającej się o $\Omega = 1^\circ$



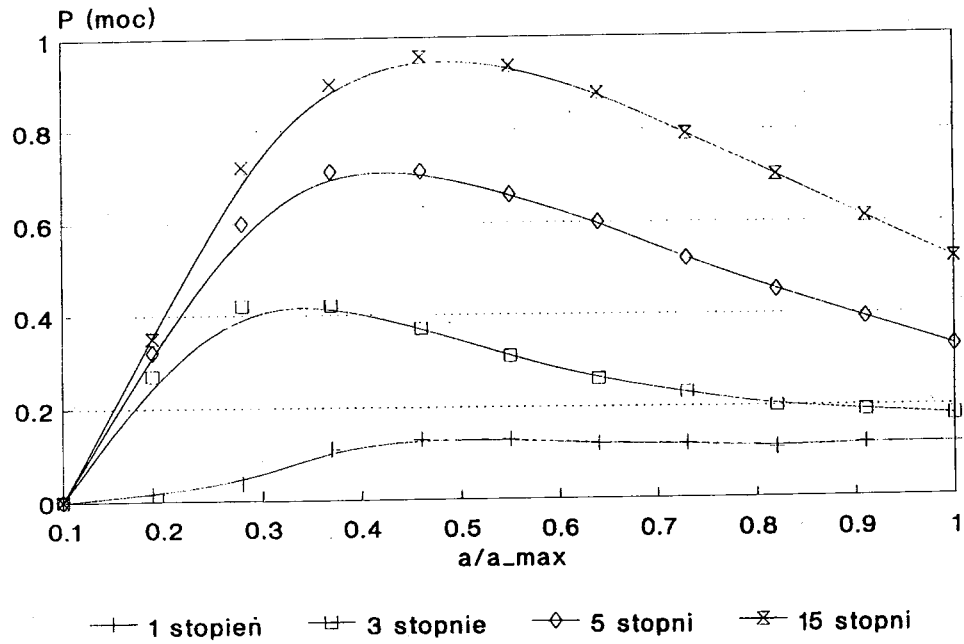
Rys. 9. Przebieg mocy dla modu HE_{11} i HE_{12} w funkcji długości z dla rozszerzającej się o $\Omega=3°$



Rys. 10. Przebieg mocy dla modu HE_{11} i HE_{12} w funkcji długości z dla rozszerzającej się o $\Omega=5°$



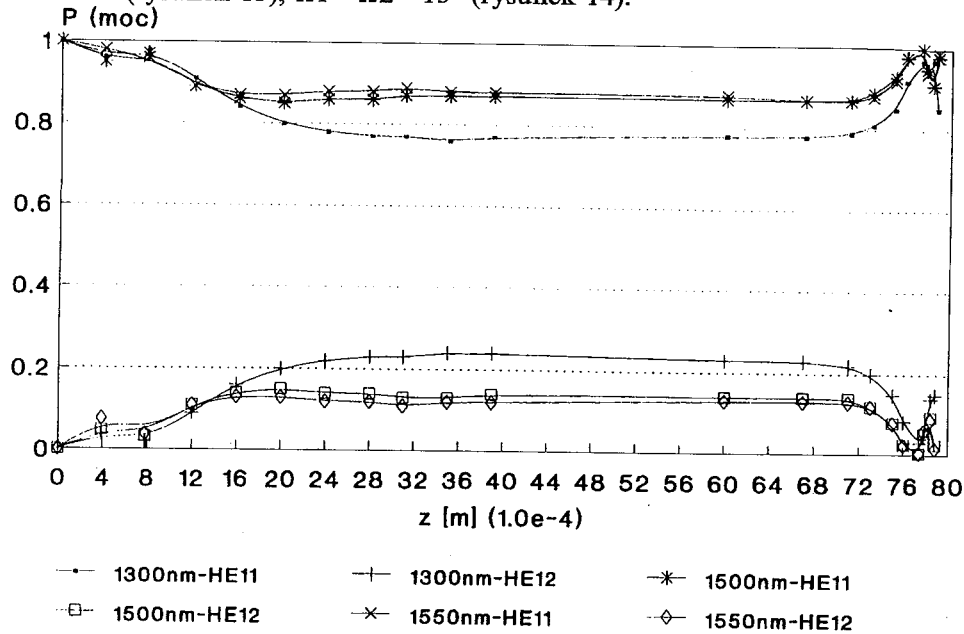
Rys. 11. Przebieg mocy dla modu HE_{11} i HE_{12} w funkcji długości z dla rozszerzającej się o $\Omega = 15°$



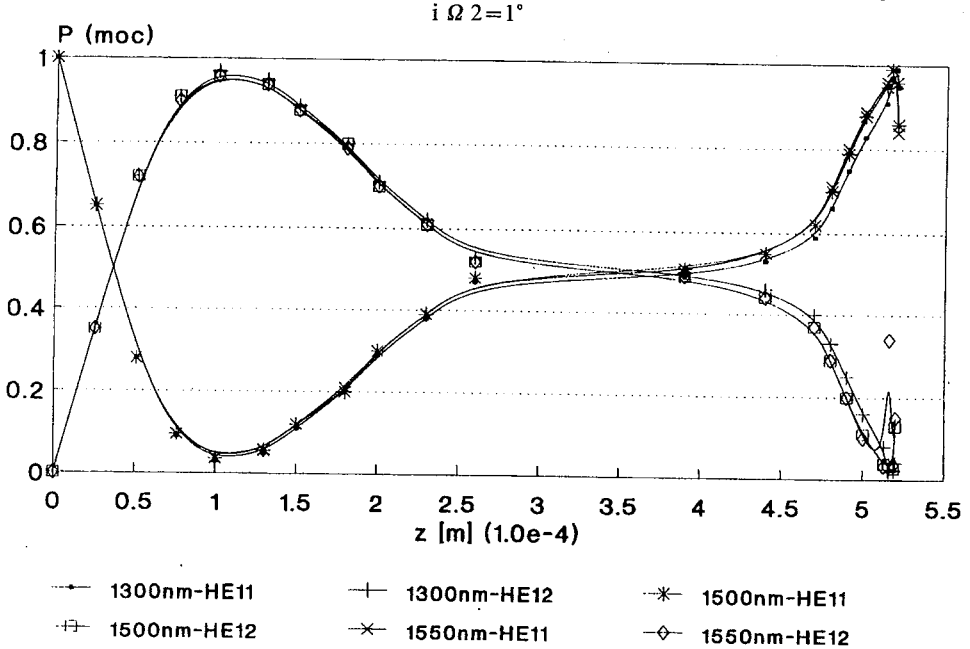
Rys. 12. Zmiana mocy prowadzonej w modzie HE_{12} w funkcji zmiany promienia dla struktury rozszerzającej się o różnej wartości kąta Ω

3.1.2. Konstrukcja złączowa

Przepływ mocy zachodzący między modem HE_{11} a HE_{12} przebadano dla konstrukcji złączowej złożonej ze stożka zwężającego się i rozszerzającego o następujących kątach nachylenia: $\Omega 1 = \Omega 2 = 1°$ (rysunek 13), $\Omega 1 = \Omega 2 = 15°$ (rysunek 14).



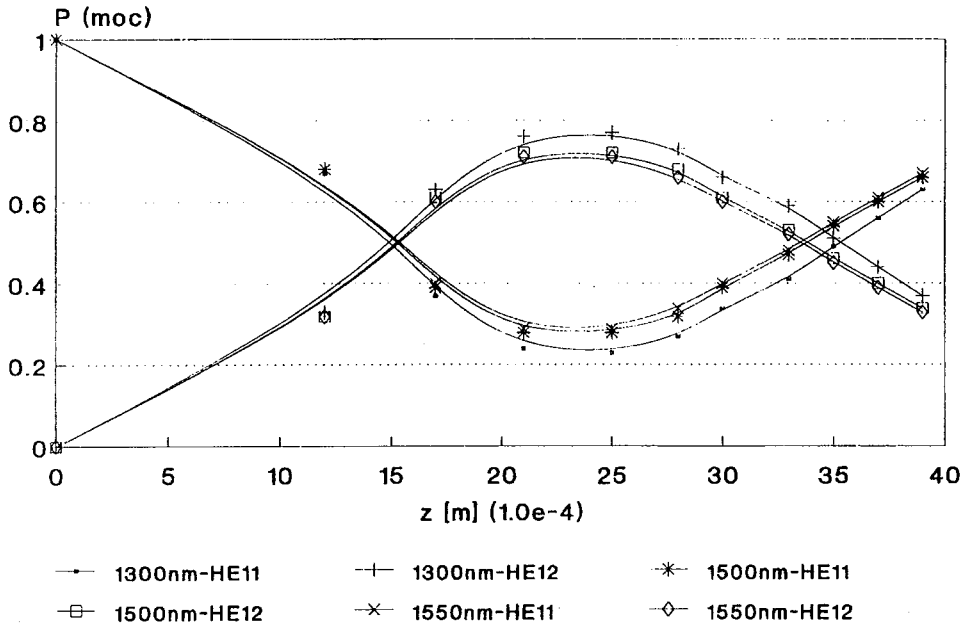
Rys. 13. Przebieg mocy dla modu HE_{11} i HE_{12} w funkcji długości z dla konstrukcji złączowej o $\Omega 1 = 1°$ i $\Omega 2 = 1°$



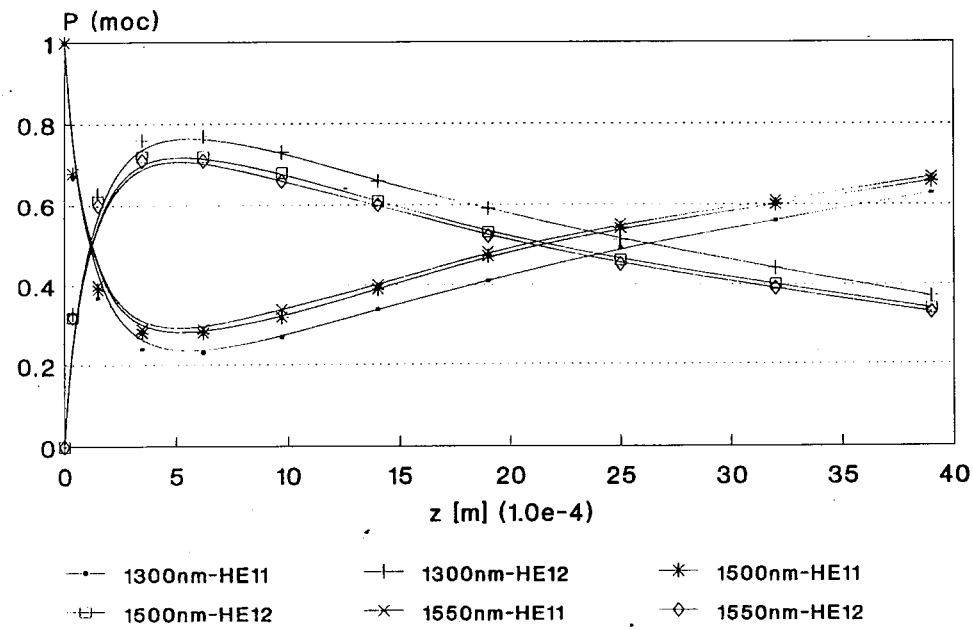
Rys. 14. Przebieg mocy dla modu HE_{11} i HE_{12} w funkcji długości z dla konstrukcji złączowej o $\Omega 1 = 15°$ i $\Omega 2 = 15°$

3.2. NIELINIOWE STRUKTURY STOŻKOWE

W przypadku tego typu struktur badania ograniczono do stożków rozszerzających się. Transfer mocy pomiędzy modami w strukturach o kształcie opisanym zależnością (10) przedstawiono na rysunku 15 (dla $\alpha=2$) i na rysunku 16 (dla $\alpha=0.5$).



Rys. 15. Przebieg mocy dla modu HE_{11} i HE_{12} w funkcji długości z dla nielinowej strukturze rozszerzającej się o $\alpha=2$



Rys. 16. Przebieg mocy dla modu HE_{11} i HE_{12} w funkcji długości z dla nielinowej strukturze rozszerzającej się o $\alpha=0.5$

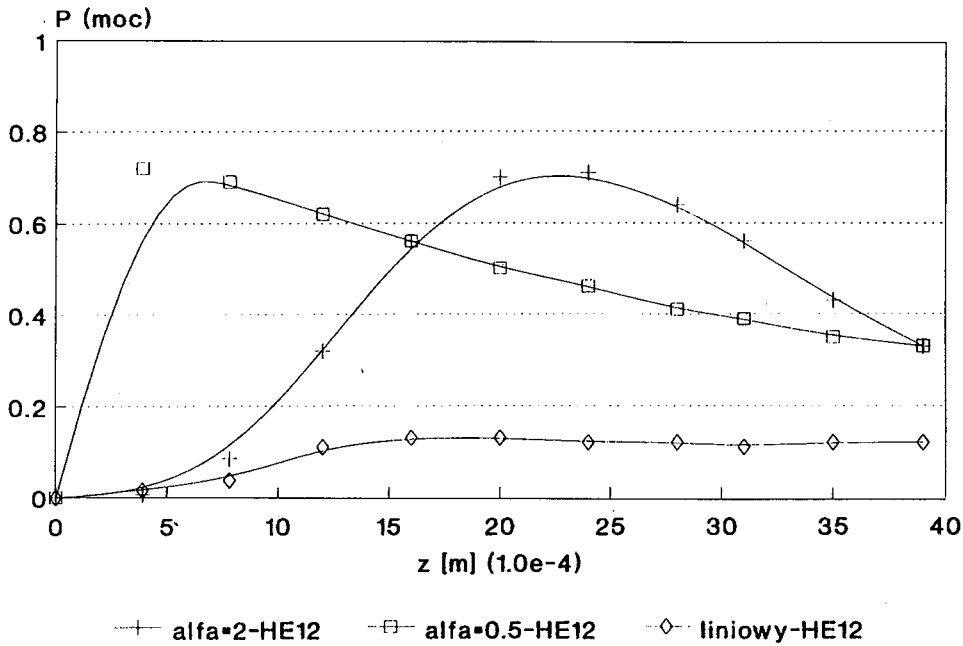
4. ANALIZA WYNIKÓW

Analizując przepływ mocy pomiędzy modem podstawowym a modem HE_{12} zauważalne jest, że wraz ze wzrostem kąta nachylenia Ω następuje silniejszy transfer mocy. Dla małych kątów, na przykład dla $\Omega = 1°$, sprzężenie jest bardzo małe, rzędu kilku, kilkunastu procent (rysunek 8); podczas gdy dla większych, na przykład dla $\Omega = 15°$, występuje silny, pełen przepływ mocy z modu HE_{11} do modu HE_{12} (rysunek 11). Okazuje się, że silny transfer mocy ma miejsce gdy spełniony jest warunek: $(\Gamma/V) \ll 1$, a mały gdy $(\Gamma/V) \gg 1$ [5], przy czym:

$$\Gamma = \frac{\sqrt{\delta}}{2 \cdot \Omega} \cdot \{U_q^2(\infty) - U_p^2(\infty)\} \tag{11}$$

Charakter sprzężeń modowych dla struktur stożkowych, w przypadku dużych kątów Ω, jest niezależny od długości fali (rysunek 11). Dla małych kątów wraz ze zmianą długości fali pojawiają się pewne różnice w przekazywaniu mocy między modami (rysunek 9). Różnice te są tym większe im kąt Ω jest mniejszy i różnica między długościami fal większa. Na rysunku 17 pokazano zmiany mocy prowadzonej modem HE_{12} w funkcji długości dla liniowej struktury stożkowej o $\Omega = 1°$ oraz dla nieliniowej o $\alpha = 2$ i $\alpha = 0.5$ (przy $\lambda = 1550$ nm).

Okazuje się, że szybszy na początku wzrost promienia stożka, jak to ma miejsce w strukturze nieliniowej o $\alpha = 0.5$, powoduje gwałtowniejszy przyrost mocy w modzie HE_{12} w porównaniu ze stożkiem liniowym czy też nieliniowym o $\alpha = 2$.



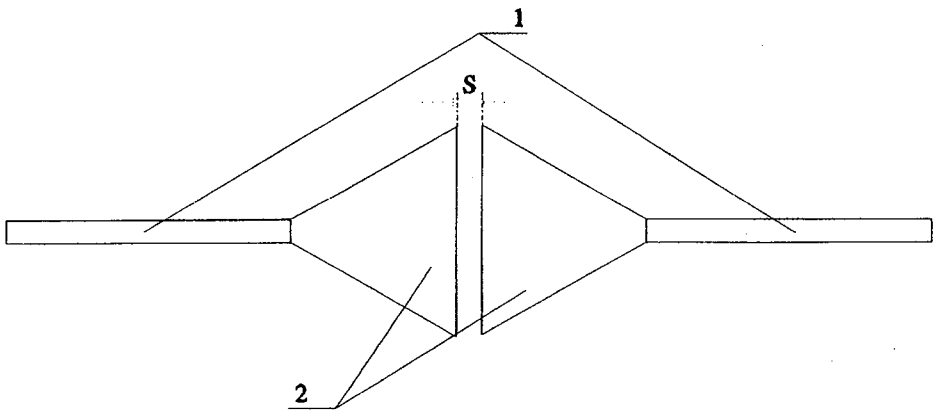
Rys. 17. Zmiana mocy prowadzonej w modzie HE_{12} w funkcji długości z dla struktury liniowej o $\Omega = 1°$ i nielinowej o $\alpha = 2$ i 0.5

5. ANALIZA ZŁĄCZ STOŻKOWYCH

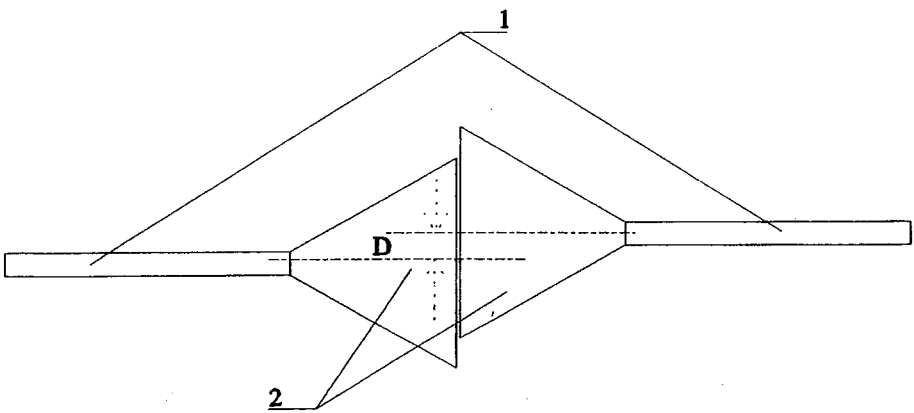
Jednym z zastosowań dielektrycznych struktur stożkowych są połączenia światłowodowe rozłączne (rysunek 2). Sprzężenie optyczne dwóch odcinków światłowodów wykonuje się zbliżając czołowo końce włókien do uzyskania kontaktu optycznego. Transmisja mocy optycznej odbywa się przede wszystkim poprzez sprzężenie rdzeni światłowodów.

Występujące straty energii promieniowania wynikające z łączenia włókien spowodowane są wieloma przyczynami. Do głównych przyczyn strat zaliczamy: przesunięcie poosiowe (rysunek 18), przesunięcie radialne (rysunek 19) i niedopasowanie kątowe (rysunek 20). Dla dwóch identycznych struktur stożkowych straty sprzężenia mogą być opisane następującymi zależnościami [11]:

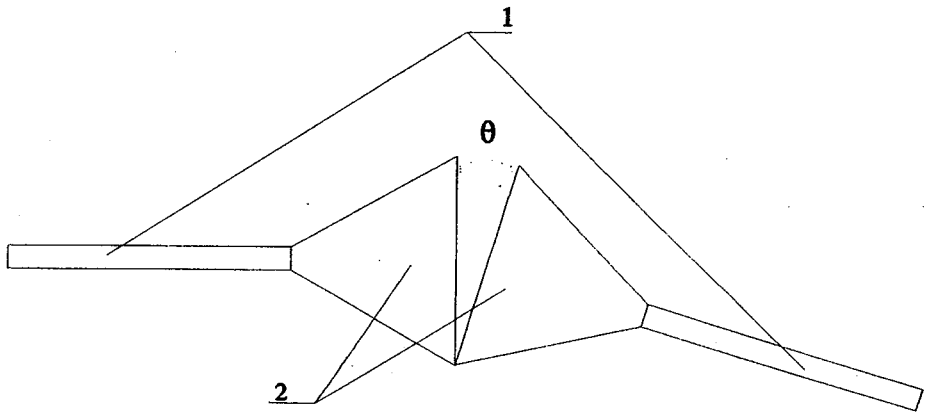
$$Ts = \frac{4}{4 + (S/\lambda)^2/\Pi^2 \cdot (w/\lambda)^4} \tag{12}$$



Rys. 18. Przesunięcie poosiowe S łączonych światłowodów zakończonych strukturami stożkowymi
1 – światłowód, 2 – stożek dielektryczny



Rys. 19. Przesunięcie radialne D osi łączonych światłowodów zakończonych strukturami stożkowymi
1 – światłowód, 2 – stożek dielektryczny



Rys. 20. Niedopasowanie kątowe θ osi łączonych światłowodów zakończonych strukturami stożkowymi 1 — światłowód, 2 — stożek dielektryczny

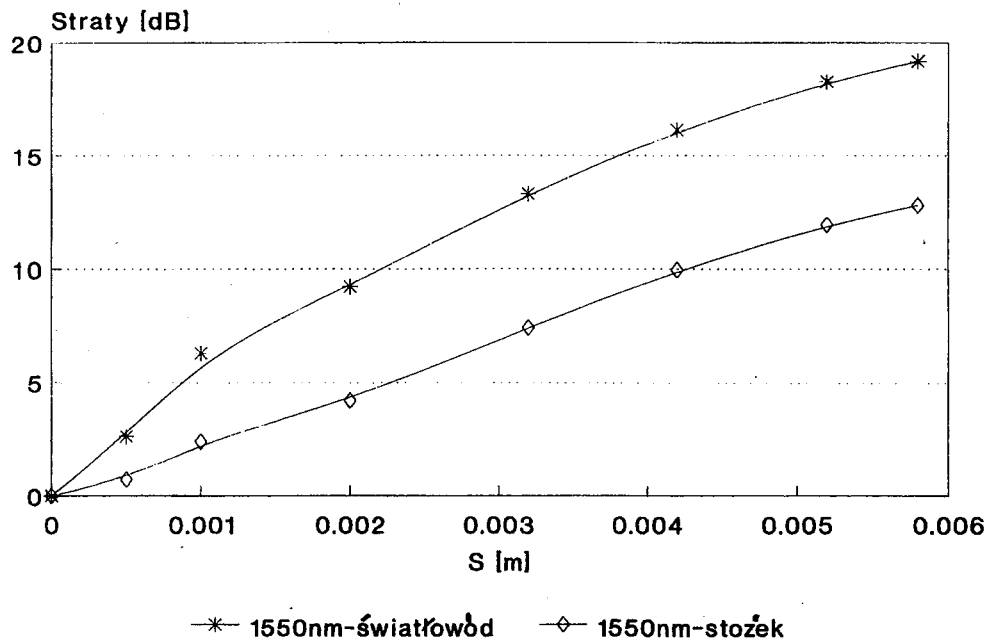
dla przesunięcia poosiowego,

$$Td_{Ts=1} = \exp[-(D/w)^2] \tag{13}$$

dla przesunięcia radialnego,

$$T\theta_{Ts=1} = \exp[-(\Pi \cdot w \cdot \theta/\lambda)^2] \tag{14}$$

dla niedopasowania kątowego.

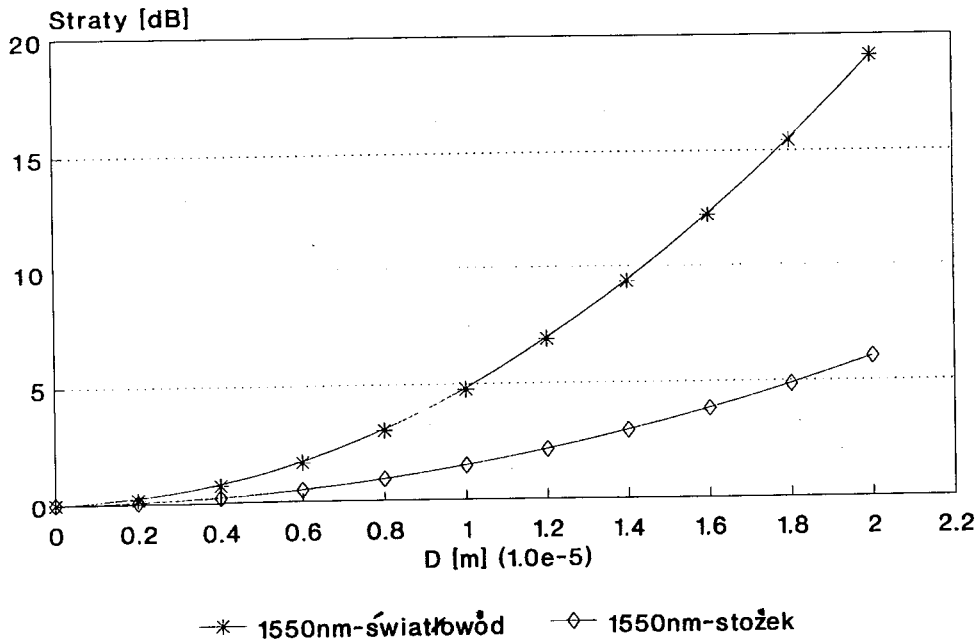


Rys. 21. Zależność strat od wartości przesunięcia poosiowego S dla złącza stożkowego i światłowodu jednomodowego

Przez w oznaczono średnicę pola modowego. Na rysunku 21 przedstawiono zależność strat związanych z przesunięciem poosiowym dla złącza stożkowego o maksymalnym promieniu równym 76156 nm i światłowodu jednomodowego o $w=9.87$ μm (przy $\lambda=1550$ nm).

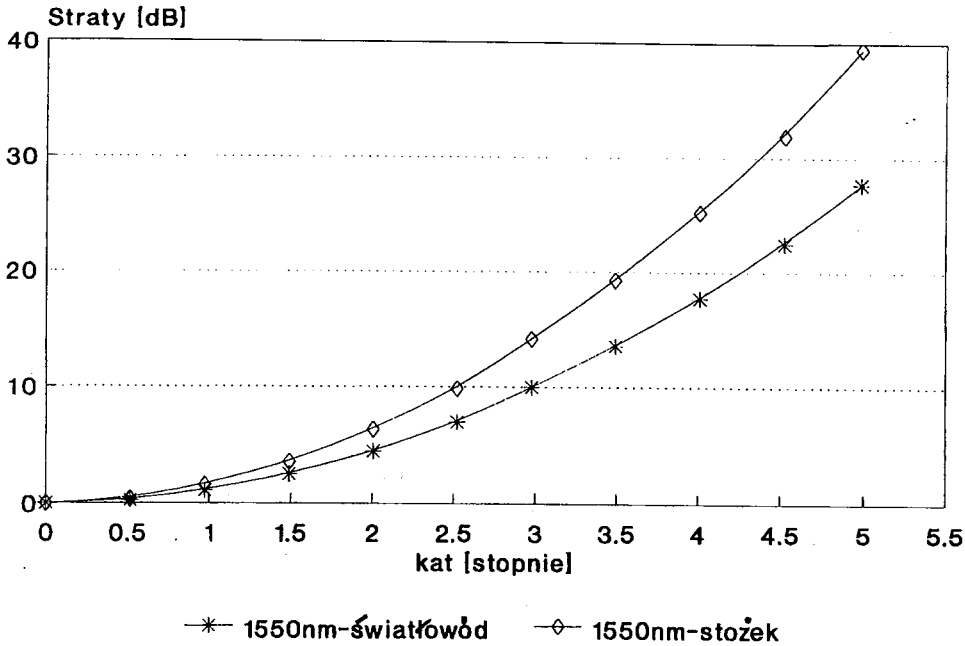
Straty wnoszone przez tę samą wartość przesunięcie poosiowego S są wyraźnie mniejsze dla złącz stożkowych niż dla złącz typu włókno jednomodowe — włókno jednomodowe. Dla przesunięcia $S=2$ mm w pierwszym przypadku mamy straty około 4.20 dB a w drugim 8.97 dB.

Na rysunku 22 pokazano zależność strat od wielkości przesunięcia radialnego D dla struktury stożkowej i włókna jednomodowego. Podobnie jak w poprzednim przypadku straty są mniejsze dla złącz o konstrukcji opartej na strukturze stożkowej. Dla $D=6$ μm straty wynoszą 0.54 dB dla stożka i 1.61 dB dla światłowodu jednomodowego.



Rys. 22. Zależność strat od wartości przesunięcia poosiowego D dla złącza stożkowgo i światłowodu jednomodowego

W przypadku strat spowodowanych niedopasowaniem kątowym (rysunek 23) wiekszą wrażliwość na ten typ niedokładności wykazują struktury stożkowe niż włókna jednomodowe. Dla kąta $\Omega=1.5°$ w przypadku struktury stożkowej mamy straty 3.54 dB a dla światłowodu 2.50 dB.



Rys. 23. Zależność strat od wartości niedopasowania kątowego θ dla złącza stożkowego i światłowodu jednomodowego

PODSUMOWANIE

W pracy poddano analizie sprzężenia międzymodowe zachodzące w dielektrycznych strukturach stożkowych przy różnych kątach nachylenia Ω i różnej długości fali λ. Na podstawie przeprowadzonych badań stwierdzono:

a) dużą wrażliwość przebiegu transferu mocy między modami na niewielkie zmiany kąta Ω, szczególnie w zakresie małych wartości tego kąta,

b) niewielkie zmiany charakteru sprzężeń międzymodowych przy zmianie długości fali; znaczne zmiany uwidaczniają się dopiero w konstrukcji złożonej z kilku struktur stożkowych, a szczególnie przy naprzemiennym ustawieniu struktur rozszerzających sie i zwężających,

c) dla typowych struktur stożkowych istotną rolę odgrywa jedynie sprzężenie pomiędzy modem HE_{11} i modem HE_{12} dopiero gdy δ ma wartość znacznie większą od 0.1 konieczne staje się uwzględnienie we wzajemnym transferze mocy kolejnego modu.

Porównując struktury nieliniowe o różnej wartości α i liniowe widzimy, że dany stan transferu mocy między modami wcześniej jest uzyskiwany dla stożków o szybszym początkowym wzroście promienia a w funkcji długości z.

Przeanalizowano również właściwości złącz optycznych o konstrukcji opartej na stożkach. Tego typu konstrukcje wykazują znacznie mniejsze straty niż klasyczne złącza dla światłowodów jednomodowych.