POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

TOM 43 — ZESZYT 3

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1997

ADRES REDAKCJI

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

Dyżury Redakcji: środy i piątki, godz. 14—16
tel/fax (0-22) 660 77 37, 25 29 18+aut. sekr.

Telefony domowe: Redaktora Naczelnego: 12 17 65
Zastępcy Red. Naczelnego: 826 83 41
Sekretarza Odpowiedzialnego: 25 29 18

WYDAWNICTWO NAUKOWE PWN
Warszawa, ul. Miodowa 10

Ark. wyd. 10,75 Ark. druk. 9,00	Podpisano do druku w październiku 1997 r.
Papier offsetowy kl. III 70 g. B-1	Druk ukończono listopad 1997 r.

Skład: *Amel*, Warszawa
Druk i oprawa: Drukarnia Braci Grodzickich, Żabieniec, ul. Przelotowa 7

Szanowni Autorzy

„Kwartalnik Elektroniki i Telekomunikacji — Electronics and Telecommunication Quarterly" jest kontynuatorem tradycji powstałego 43 lata temu kwartalnika pt. „Rozprawy Elektrotechniczne". „Kwartalnik Elektroniki i Telekomunikacji — czasopismo Komitetu Elektroniki i Telekomunikacji Polskiej Akademii Nauk — wydawany jest przez Wydawnictwo Naukowe PWN.

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

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

Czasopismo dociera do wszystkich zajmujących się elektroniką i telekomunikacją krajowych ośrodków naukowych oraz technicznych i do wielu instytucji zagranicznych. jest ponadto prenumerowane przez liczne grono specjalistów.

Spełniając życzenia Autorów i Czytelników Redakcja stara się o zwiększenie prenumeraty zagranicznej i preferuje artykuły pisane w języku angielskim, co zwiększy ich odbiór w środowiskach naukowych na całym świecie. W związku z tym, streszczenia w językach polskim i angielskim powinny być obszerne — do 60 wierszy — i dawać jasny pogląd na tezy i wyniki prac opisanych w artykule.

Każdy Autor otrzymuje bezpłatnie 20 egzemplarzy nadbitek swojego artykułu, co ułatwia przesłanie tej publikacji do indywidualnie wybranych przez Autora osób i instytucji w kraju lub za granicą.

Nadsyłane do redakcji artykuły są publikowane w terminie około pół roku, w przypadku sprawnej współpracy Autora z redakcją. Wytyczne dla Autorów dotyczące formy publikacji są zamieszczone w zeszytach Kwartalnika. Można je także otrzymać w siedzibie redakcji.

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

Redakcja

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1997, 43, z. 3,ss. 301–312

Identyfikacja metodą momentów i metodą powierzchni — wnioski z badań numerycznych

HENRYK KUNERT

Instytut Cybernetyki Technicznej, Politechnika Wrocławska

Otrzymano 1996.12.23

Autoryzowano 1997.04.20

W pracy porównano dokładność badań symulacyjnych działania dwóch metod identyfikacji liniowych obiektów dynamicznych — metody momentów i metody powierzchni. Porównania dokonano metodą komputerowej symulacji obiektu o zadanej transmitancji, którego parametry identyfikowano przy użyciu badanych metod. Wyniki badań przedstawiono w postaci wykresów, zestawienia tabelarycznego i omówienia.

Słowa kluczowe: obiekty dynamiczne, metody identyfikacji, komputerowa symulacja, zakłócenia addytywne.

W literaturze specjalistycznej z zakresu identyfikacji prezentowane są różne metody identyfikacyjne, jednakże w niewielu publikacjach podane są wyniki badań dotyczące zakresu stosowalności, zalet i wad prezentowanych metod.
Omawiane niżej metody zostały wybrane ze względu na ich prostotę, co decyduje o łatwości ich aplikacji w nieskomplikowanych systemach mikroprocesorowych identyfikujących układy rzeczywiste.
Badaniom poddano następujące metody:
1. Metoda momentów [1, 2, 3];
2. Metoda powierzchni [2].

Jako błąd aproksymacji odpowiedzi układu stanowiacy kryterium jakości identyfikacji przyjęto

$$\Delta = \sqrt{\int_0^{t_c} (y_r(t) - y_{obl}(t))^2 \, dt}, \tag{1}$$

gdzie: $y_r(t)$ — wartość odpowiedzi ukladu wyjściowego,
$y_{obl}(t)$ — wartość odpowiedzi modelu,
t_c — przedział czasu, w którym sprawdzana jest zgodność modelu z układem wyjściowym.

1. METODA MOMENTÓW

Metoda ta zaliczana jest do grupy metod identyfikacji parametrycznej. Identyfikacja liniowego stacjonarnego układu sterowania przy użyciu tej metody polega na wyznaczeniu współczynników transmitancji identyfikowanego układu na podstawie momentów jego odpowiedzi impulsowej $g(t)$, które opisane są zależnością (moment i-tego rzędu):

$$m_i = \int_0^\infty t^i g(t)\,\mathrm{d}t, \tag{2}$$

gdzie: $g(t)$ — odpowiedź impulsowa układu.
Transmitancję identyfikowanego układu można przedstawić w postaci:

$$G(s) = \int_0^\infty g(t)e^{-st}\,\mathrm{d}t \tag{3}$$

lub

$$G(s) = \frac{b_m s^m + b_{m-1}s^{m-1} + \ldots — b_{1s} + b_0}{a_n s^n + a_{n-1}s^{n-1} + \ldots + a_{1s} + 1} \tag{4}$$

gdzie: $n > m$.

Poprzez rozwinięcie we wzorze (3) członu e^{-st} w szereg Taylora w otoczeniu punktu st = 0, otrzymuje się:

$$G(s) = \int_0^\infty \left[1 - st + \frac{st^2}{2!} - \frac{st^3}{3!} + \ldots\right] g(t)\,\mathrm{d}t \tag{5}$$

nastepnie korzystając z zależności (2) można wyrażenie (5) zapisać w postaci:

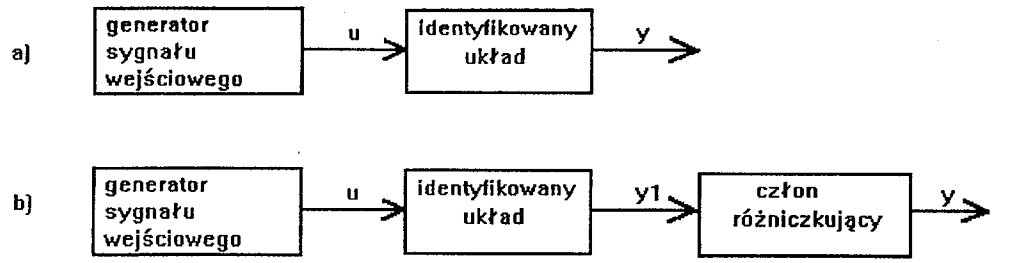
$$G(s) = m_0 - sm_1 + \frac{s^2}{2!}m_2 - \frac{s^3}{3!}m_3 + \ldots, \tag{6}$$

skąd po porównaniu prawych stron w zależnościach (4) i (6) opisujących transmitancję identyfikowanego układu oraz po wymnożeniu otrzymuje się:

$$\left(m_0 - sm_1 + \frac{s^2}{2!}m_2 - \frac{s^3}{3!}m_3 + \ldots\right)(1 + a_1 s + a_2 s^2 + \ldots + a_n s^n) = \\ = b_0 + b_1 s + b_2 s^2 = \ldots + b_m s^m. \tag{7}$$

Nastepnie poprzez porównanie współczynników po obu stronach równości (7) otrzymuje się układ równań, z którego wyznacza się nieznane wartości współczynników $a_1,\ldots, a_n$ oraz $b_0,\ldots, b_m$. Szczegółowy opis metody przedstawiony jest w pracach [1, 2, 3].

Wspomniana wyżej metoda poddana została badaniom numerycznym przy użyciu programu symulacyjnego TUTSIM wersja 7.0 w układach identyfikacyjnych przedstawionych na rysunku 1.



Rys. 1. Schematy blokowe układów identyfikacji stosowanych w metodzie momentów: a) dla syugnału wejsciowego $u(t) \cong \delta(t)$ b) dla sygnału wejściowego $u(t) \equiv 1(t)$

W pierwszym układzie identyfikacyjnym (rys. 1a) sygnałem wejściowym był impuls o parametrach zbliżonych do impulsu Dirac'a, którego amplituda i czas trwania dobierane były odpowiednio do dynamiki badanych układów przy jednoczesnym zachowaniu warunku $\int_{-\infty}^{+\infty} \delta(t)\,dt = 1$. Na rys. 1b przedstawiony jest układ, w którym sygnałem wejściowym był skok jednostkowy ($u(t) \equiv 1(t)0$, a odpowiedź skokowa identyfikowanego układu po zróżniczkowaniu stanowiła jego odpowiedź impulsową — w ten sposób wyeliminowano wpływ parametrów sygnału wejściowego na własności ilościowe i jakosciowe odpowiedzi impulsowej układu. Przedstawione sposoby symulacji działania metody dały identyczne wyniki.

1.1. IDENTYFIKACJA BEZ ZAKŁÓCEŃ SYGNAŁÓW MIERZONYCH

1. Identyfikację prowadzono do chwili zaniku procesu przejściowego w badanym układzie, co odpowiada w przybliżeniu chwili, w której wartość ustalona odpowiedzi impulsowej stanowi około 2% jej wartości maksymalnej.

2. Identyfikowano układy opisane transmitancjami 2-go do 5-go rzędu modelami układów 2-go i 3-go rzędu.

3. Po czasie t_k (ustalonym w punkcie 1.1.) wyniki identyfikacji były zadowalające (tzn. występowała wizualna zgodność odpowiedzi czasowych układu wyjściowego i jego modelu, czego potwierdzenie stanowiła mała wartość błędu wyliczona zgodnie z (1) dla czasu t_c) dla identyfikowanych układów nie zawierajacych członków różniczkujących, czyli takich, w których współczynniki $b_1,\ldots, b_m$ w transmitancji (4) były równe zeru. Ustalenie czasu t_c jest arbitralne i służy jedynie do wyznaczenia błędu aproksymacji (1) odpowiedzi czasowych układu wyjściowego przez odpowiednie odpowiedzi modelu w celu dokonania porównania jakości identyfikacji tego samego układu przy użyciu różnych metod.

4. Identyfikacja układów zawierajacych człony różniczkujące jest bardzo wrażliwa na dokładność wyznaczenia momentów odpowiedzi impulsowej identyfikowanego

układu. Do identyfikacji takich układów nie można stosować kryterium na czas identyfikacji t_k przedstawionego w punkcie 1.1, w takich przypadkach identyfikację należałoby prowadzić do chwili, w której zwiększanie czasu identyfikacji t_k nie będzie wprowadzało istotnych zmian odpowiedzi czasowych uzyskanego modelu. Rozpatrywany przypadek przedstawiono w przykładzie 1.

Przykład 1

Dla układu o transmitancji

$$G(s) = \frac{7s^2 + 5s + 6}{62.5s^4 + 112.5s^3 + 62.5s^2 + 13.5s + 1} \tag{8}$$

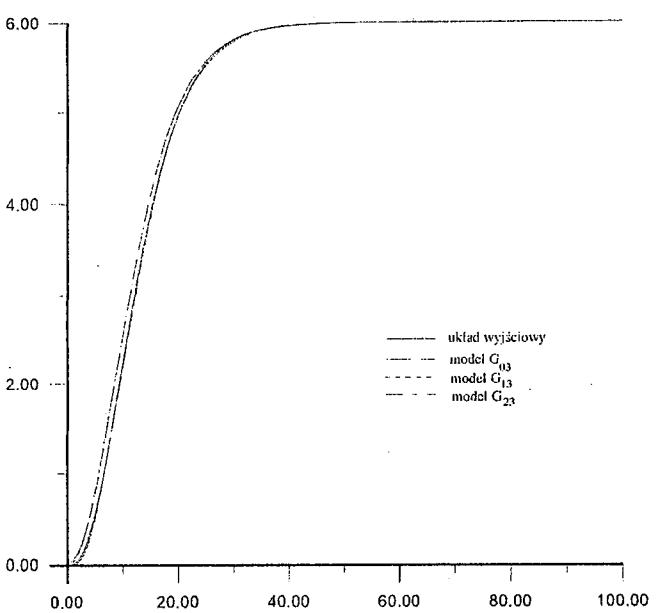
w wyniku przeprowadzonej identyfikacji otrzymano modele 3-go rzędu o transmitancjach:

$$G_{03}(s) = \frac{6}{56.56s^3 + 50.99s^2 + 12.68s + 1}, \tag{9}$$

$$G_{13}(s) = \frac{4.13s + 6}{92.06s^3 + 59.82s^2 + 13.37s + 1}, \tag{10}$$

$$G_{23}(s) = \frac{3.68s^2 + 1.01s + 6}{72.66s^3 + 53.74s^2 + 12.85s + 1}. \tag{11}$$

Odpowiedzi skokowe identyfikowanego układu o transmitancji (8) oraz modeli o transmitancjach (9), (10), (11) przedstawiono na rysunku 2.



Rys. 2. Wykresy odpowiedzi skokowych układu wyjściowego oraz modeli z przykładu 1

5. Podczas identyfikacji układów nie zawierajacych członów różniczkujących, przy zastosowaniu modelu identyfikowanego układu wyższego rzędu niż rząd układu wyjściowego wartości wszystkich współczynników mianownika transmitancji modelu mających indeks większy niż rząd układu wyjściowego zdążały do zera przy wzroście czasu identyfikacji t_k, co sprawia, że identyfikacja przy użyciu tej metody prowadzi w opisanym przypadku do uzyskania modeli niższego bądź tego samego rzędu co rząd układu wyjściowego. Jednak zastosowanie modeli zawierających różniczkowanie do identyfikacji układow nie zawierających różniczkowania prowadzi nieraz do uzyskania modeli niestabilnych lub takich, których odpowiedzi czasowe znacnzie odbiegaja od odpowiedzi układu wyjściowego. Powyższa sytuacja została przedstawiona w przykładzie 2.

Przykład 2

Układ wyjściowy opisany jest transmitancją

$$G(s) = \frac{6}{3s^2 + 4s + 1} \tag{12}$$

identyfikacja prowadzi do otrzymania następujących modeli 3-go rzędu:

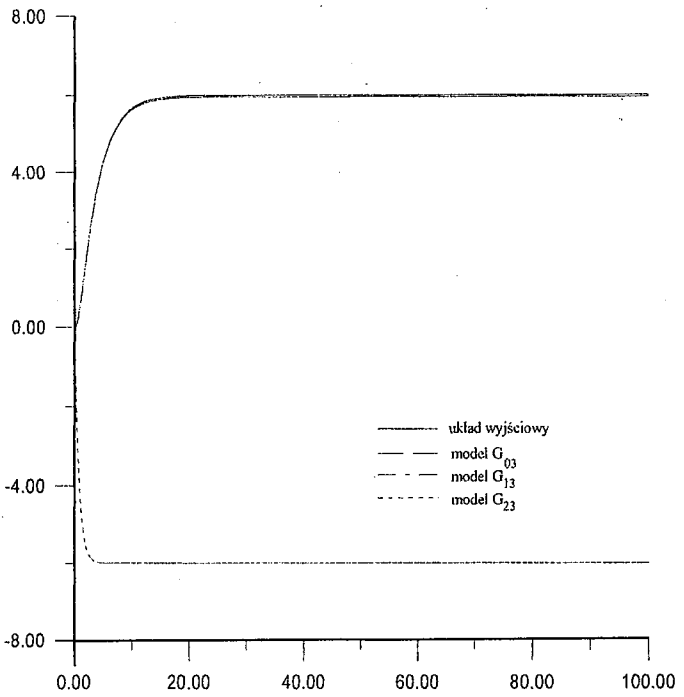
$$G_{03}(s) = \frac{6}{0.05s^3 + 3.08s^2 + 4.02s + 1}, \tag{13}$$

$$G_{13}(s) = \frac{11.46s + 6}{5.92s^3 + 10.74s^2 + 5.93s + 1}, \tag{14}$$

$$G_{23}(s) = \frac{-9.42 \cdot 10^7 s^2 - 0.6s + 6}{1.18 \cdot 10^7 s^3 + 1.57 \cdot 10^7 s^2 + 3.92s + 1}. \tag{15}$$

Wykresy odpowiedzi skokowych układu wyjściowego o transmitancji (12) oraz modeli o transmitancjach (13), (14), (15) przedstawione są na rysunku 3.

6. W wyniku identyfikacji układów nie zawierających różniczkowania, których mianownik transmitancji posiada odległy pierwiastek rzeczywisty uzyskano stabilne modele niższego rzędu doskonale aproksymujące (w sensie kryterium (1)) odpowiedzi czasowe układów wyjściowych, przy czym wraz ze wzrostem odległości tego pierwiastka od położenia pierwiastków dominujących malał jego wpływ na własności dynamiczne układu wyjściowego oraz na jakość aproksymacji przez model. W tym przypadku również można zauważyć pewną rozbieżność między odpowiedziami skokowymi identyfikowanego układu i jego modelu w sytuacji, gdy model zawiera różniczkowanie, które nie występuje w układzie wyjściowym. Przypadek ten rozpatrywany jest w przykładzie 3, gdzie model o transmitancji $G_{12}(s)$ jest niestabilny mimo stabilności układu wyjściowego. Na fakt uzyskiwania niestabil-



Rys. 3. Wykresy odpowiedzi skokowych układów z przykładu 2

nych modeli w wyniku identyfikacji układów stabilnych zwrócono również uwagę w pracy [3].

Przykład 3

W wyniku identyfikacji układu opisanego transmitancją

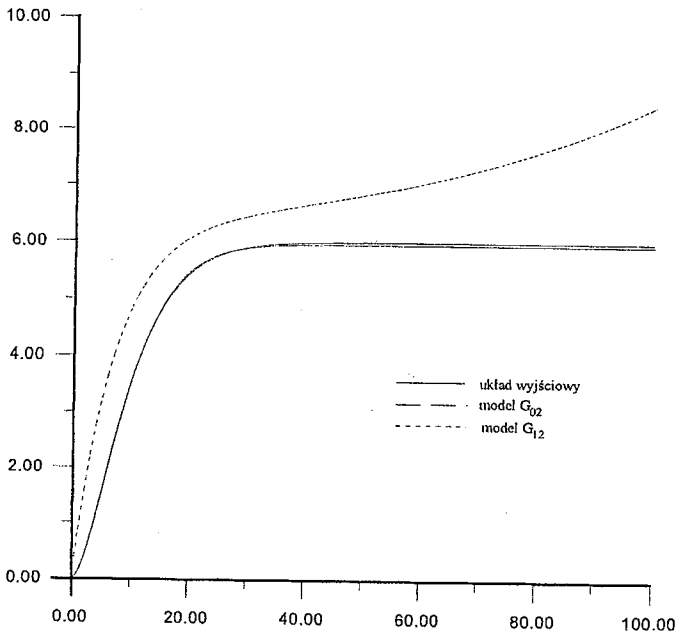
$$G(s) = \frac{6}{(6s+1)^2(0.2s+1)} = \frac{6}{0.5s^3 + 25.2s^2 + 10.02s + 1} \tag{16}$$

uzyskano następujące modele 2-go rzędu:

$$G_{02}(s) = \frac{5.95}{26.61s^2 + 9.79s + 1}, \tag{17}$$

$$G_{12}(s) = \frac{-301.52s + 5.95}{-330.98s^2 - 40.89s + 1}. \tag{18}$$

Odpowiedzi skokowe układu o transmitancji (16) oraz modeli o transmitancjach (17), (18) przedstawiono na rysunku 4.



Rys. 4. Wykresy odpowiedzi skokowych układow badanych w przykladzie 3

1.2. IDENTYFIKACJA W WARUNKACH ZAKŁÓCEŃ

1. Podczas identyfikacji układów w warunkach addytywnych zakłóceń losowych (przypadek najczęściej występujący w rzeczywistości) sumowanych z sygnałem odpowiedzi impulsowej badanego ukladu można zauważyć duże ograniczenia w stosowalności tej metody albo stwierdzić wręcz jej nieprzydatność.

Jako zakłócenie addytywne wykorzystano:

a) sygnał pseudolosowy o rozkładzie jednostajnym na odcinku $x \in (-1;1)$ z wartością oczekiwaną $E(X) = 0$, przy czym wartości realizacji zmiennej losowej X były kształtowane tak, aby stanowiły kilka do kilkunastu procent maksymalnej wartości odpowiedzi impulsowej badanego układu;

b) sygnał sinusoidalny o różnych częstotliwościach i amplitudzie kształtowanej jak dla sygnału pseudolosowego opisanego w punkcie a).

2. Niemożliwym wręcz do ustalenia jest kryterium czasu identyfikacji w warunkach zakłóceń, gdyż jest to zależne w dużym stopniu od przyjętego modelu identyfikowanego układu oraz od parametrow tego układu.

Zgodnie z założeniami badanej metody [1, 2, 3], czas identyfikacji powinien dążyć do nieskonczoności ($t_k \to \infty$), jednakże im większy czas identyfikacji, tym dłużej całkowany jest zakłócony sygnał odpoiwezi impulsowej, co z kolei prowadzi do zwiększenia błędów identyfikacji:

a) przy $t_k \to \infty$ wzrasta czynnik t^i w zależności (2);
b) przy $t_k \to \infty$ maleje stosunek sygnału do szumu $S/N \to 0$,
stąd w dwojaki sposób zwiększa się udział zakłóceń w wyliczonych wartościach momentów, ponieważ:

$$m_i = \int_0^\infty t^i [g(t) + \varepsilon(t)]\,\mathrm{d}t = \underbrace{\int_0^\infty t^i g(t)\,\mathrm{d}t}_{całka\ 1} + \underbrace{\int_0^\infty t^i \varepsilon(t)\,\mathrm{d}t}_{całka\ 2}, \tag{19}$$

gdzie: $\varepsilon(t)$ — wartość sygnału zakłócającego;
Przy $t \to \infty \Rightarrow g(t) \to 0$ stąd całka 1 dąży do wartości ustalonej natomiast zwiększa się udział całki 2 w określeniu wartości m_i. Teoretycznie wartość całki 2 winna zdążać do zera lecz w praktyce czas identyfikacji jakkolwiek nie byłby on długi ma wartość skończoną, toteż założenie metody nigdy nie będzie spełnione. Z tego względu proponuje się ustalenie czasu identyfikacji indywidualnie na podstawie obserwacji wartości wyliczonych momentów przy założonym modelu identyfikowanego układu.

3. Niemożliwe staje się również ustalenie poziomu zakłóceń, który nie będzie miał istotnego wpływu na poprawność wyników identyfikacji w zależnosci od parametrow odpowiedzi impulsowej.
Na podstawie analizy procesu identyfikacji w warunkach zakłóceń przy użyciu metody momentów [1, 2, 3], nasuwa się wniosek, iż należy korzystać z modeli możliwie najniższego stopnia, wtedy wpływ zakłóceń będzie najmniejszy, chociaż przybliżenie odpowiedzi układu wyjściowego przy użyciu takiego modelu może okazać się niewystarczające.

2. METODA POWIERZCHNI

Metoda ta również, jak wyżej prezentowana metoda momentów, należy do metod identyfikacji parametrycznej. Szczegółowy opis metody zamieszczono w pracy [2]. Metoda powierzchni oparta jest również o całkowanie sygnału wyjściowego identyfikowanego układu, przy czym sygnałem tym jest odpowiedź skokowa.

Korzystając z twierdzenia o przejściu granicznym z rachunku operatorowego, dla układu opisanego transmitancją (4), można napisać następujące zależności:

$$\begin{aligned}
K_0 &= \lim_{s\to 0} G(s) = b_0 \\
K_1 &= \lim_{s\to 0} G_1(s) = \lim_{s\to 0} \frac{1}{s}[K_0 - G(s)] = -b_1 + K_0 s_1 \\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
K_l &= \lim_{s\to 0} G_l(s) = \lim_{s\to 0} \frac{1}{s}[K_{l-1} = G_{l-1}(s)] = (-1)^l b_l + K_{l-1} a_1 - K_{l-2} a_2 + \ldots + (-1)^{l-1} K_0 a_l
\end{aligned} \tag{20}$$

przy czym $b_l = 0$ dla $l > m$ oraz $a_l = b_l = 0$ dla $l > n$.

Następnie nieznane wartości współczynników $K_0,\ldots, K_l$ wyznacza się na podstawie następujących związków:

$$K_0 = \lim_{s\to 0} G(s) = \lim_{t\to\infty} h(t),$$

$$K_1 = \lim_{s\to 0} G_1(s) = \lim_{t\to\infty} h_1(t) = \lim_{t\to\infty} \int_0^t [K_0 - h(\tau)]\,d\tau,$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$K_l = \lim_{s\to 0} G_l(s) = \lim_{t\to\infty} h_l(t) = \lim_{t\to\infty} \int_0^t [K_{l-1} - h_{l-1}(\tau)]\,d\tau, \tag{21}$$

gdzie: $h(t)$ — odpowiedź skokowa identyfikowanego układu.
Na podstawie zależności (20) i (21) otrzymuje się układ $l+1$ równań, skąd wyznacza się szukane wartości współczynników ($a_1,\ldots, a_n$ oraz $b_0,\ldots, b_m$) transmitancji identyfikowanego układu.

3. WŁASNOŚCI METODY POWIERZCHNI

1. Metoda posiada zalety metody momentów przedstawione w punktach I.5 i I.6, jednak wyznaczone przy jej zastosowaniu modele identyfikowanych układow lepiej aproksymują układy wyjściowe w porównaniu z modelami uzyskanymi przy użyciu metody momentów. Porównanie jakości działania obu metod dla układów przedstawionych w przykładach 1 ÷ 3 przedstawiono w tabeli 1.
Przy stosowaniu metody powierzchni proponuje się przyjęcie minimalnego czasu identyfikacji jak dla metody momentów (punkt 1.1).

Tabela 1

Błędy aproksymacji identyfikacji układów przez ich modele dla przykładów 1 ÷ 3, dla $t_c = 100$

	transmitancja modelu	błąd aproksymacji	
		metoda momentów	metoda powierzchni
przykład 1	G_{03}	0.1382	0.1388
	G_{13}	0.0483	0.0602
	G_{23}	0.0239	model niestabilny
przykład 2	G_{03}	0.3968	0.0376
	G_{13}	0.3899	0.0366
	G_{23}	117.7210	1.7653
przykład 3	G_{02}	0.1174	0.3978
	G_{12}	model niestabilny	0.3907

2. Addytywne zakłócenia odpowiedzi skokowej układu mają znacznie mniejszy wpływ na jakość procesu identyfikacji przy użyciu tej metody ze względu na to, że:
a) tylko raz całkowany jest zakłócony sygnał wyjściowy identyfikowanego układu;
b) zakłócenia w całkowanych sygnałach nie są zwielokrotniane poprzez mnożenie przez jakiekolwiek współczynniki, jak to miało miejsce w metodzie momentów, co sprawia, że następuje ich uśrednianie;
c) występuje znacznie mniejszy udział zakłóceń w sygnale wyjściowym identyfikowanego układu (większy stosunek sygnału do szumu S/N).

3. Metoda powierzchni posiada również istotną wadę mającą duże znaczenie podczas identyfikacji w warunkach zakłóceń polegającą na trudności wyznaczenia wartości ustalonej odpowiedzi skokowej identyfikowanego układu $h_\infty = \lim_{t \to \infty} h(t)$, która to wielkość ma istotny wpływ na wyznaczenie pośrednich parametrów K_r dla $r = 0,..., l$ wyznaczanych przy pomocy zależności (21).
Dlatego też proponuje sie uśrednienie wyników pomiarow wyjścia układ uw otoczeniu punktu $t = t_k$, gdzie odpowiedź skokową układu przyjmuje się za ustaloną.

4. Metoda nie nadaje się do identyfikacji *on-line*, gdyż wymagane jest kilkakrotne (zależnie od stopnia modelu układu) przeglądanie tablicy danych z wartościami kolejnych funkcji $h_r(t)$ oraz przechowywanie tej tablicy w pamięci urządzenia identyfikacyjnego, gdzie:

$$h_r(t) = \lim_{t \to \infty} \int_0^t [K_{r-1} - h_{r-1}(\tau)]\,d\tau, \qquad \text{dla} \qquad r = 1,..., l$$

oraz

$$h_0(t) = h(t)$$

WNIOSKI

W wyniku przeprowadzonych badań symulacyjnych metody momentów stwierdzono silną zależność między jakością uzyskanych modeli, a dokładnością wyznaczonych momentów odpowiedzi impulsowej identyfikowanego układu. Biorąc pod uwagę powyższy wniosek nie można w sposób jednoznaczny określić kryterium na minimalny czas identyfikacji t_k. Czas t_k należy dobierać każdorazowo do dynamiki badanego obiektu.
Stosując metodę momentów występują przypadki, kiedy aproksymacja odpowiedzi czasowych identyfikowanego układu przez odpowiedzi uzyskanego modelu nie jest zadowalająca, może to być spowodowane przyjęciem modelu o transmitancji, w której liczniku występuje nie tylko wyraz wolny, w sytuacji gdy w liczniku transmitancji układu wyjściowego wyraz wolny stanowił jego jedyny element.
Korzystając z metody momentów należy porównać wyniki identyfikacji dla kilku modeli parametrycznych, następnie wybrać model zapewniający najlepszą aproksymację.

Metoda powierzchni wykazuje wrażliwość na dokładność wyznaczenia parametrow K_r, które nie zależą jednak tak silnie od rzędu zastosowanego modelu jak momenty odpowiedzi impulsowej identyfikowanego układu, które należy wyznaczyć w rozpatrywanej wyżej metodzie momentów. Między innymi z tego powodu identyfikacja przy użyciu metody powierzchni daje lepszą aproksymację układu wyjściowego niż metoda momentów w jednakowych warunkach przeprowadzanego eksperymentu.

W konkluzji można również stwierdzić, iż wspomniane wyżej metody nadają się do identyfikacji układów, w których nie występują znaczące zakłócenia sygnałów wyjściowych. W sytuacji, gdy niemożliwe jest spełnienie tego warunku zastosować należy wstępnie filtrację danych pomiarowych w celu otrzymania niezakłóconych sygnałów odpowiedzi skokowej bądź impulsowej badanego układu.

BIBLIOGRAFIA

1. J.M. Douglas: *Analiza układów dynamicznych*. WNT, Warszawa 1976
2. H. Górecki: *Analiza i synteza układów regulacji z opóźnieniem*. WNT, Warszawa 1971
3. J. Halawa: *Metody wyznaczania transmitancji uproszczonych i ich zastosowanie w automatyce i elektroenergetyce*. Prace Naukowe Instytutu Cybernetyki Technicznej Politechniki Wrocławskiej, Monografie nr 88, Wydawnictwo Politechniki Wrocławskiej, Wrocław 1991
4. J. Halawa: *Determination of simplified transfer functions by means of the moment method*. Trans. S. C. Sim. (USA), 1989, vol. 6, nr 1, ss. 17—19

H. KUNERT

IDENTIFICATION OF THE TWO METHODS OF MOMENTUM AND SURFACE — SUMMARY OF NUMERIC EXAMINATION

Summary

Accuracy of two identification methods for linear dynamic systems, namely the method of moments and the method of surfaces, have been compared. The comparison has been made using computer simulation for preset transfer function system which parameters have been identified with the methods under examination. The results of investigations have been provided as diagrams, tabularised summary and discussion.

Key words: Identification methods, computer simulation, additive disturbances.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1997, 43, z. 3, ss. 313–320

On sensitivity calculations based on symbolic network functions having a nested form

CZESŁAW STEFAŃSKI

Katedra Teorii Obwodów i Układów, Politechnika Gdańska

Received 1996.12.23

Authorized 1997.08.26

New formulas for the sensitivities of network function are presented in the paper. The formulas are especially useful when the investigated network function is given in a symbolic form. A new algorithm of sorting for biquadratic function is also presented. The paper gives a complete tool of using symbolic network functions in sensitivity analysis.

Key words: Sensitivity analysis, symbolic network functions, LLS network.

INTRODUCTION

There are many methods of symbolic analysis of LLS networks (newer of them are mentioned in [5]). The usefulness of the methods is discussed in [1]. Results of the symbolic analysis can have a developed or nested form. The number of components in the methods giving developed results increases exponentially with the complexity of the analysed network, so methods giving nested results are preferred as more effective [4, 5]. Nested results cannot be interpreted at a glance, so special techniques for different purposes are needed. Such techniques were found for the purpose of sensitivity analysis. Their presentation here is limited to a typical situation in which the investigated network functions of interesting variables are either bilinear or biquadratic. Sensitivity formulas are presented and algorithms are given which can be used to extract the needed subexpressions from expressions having a nested form. Assume that the nested expressions of numerator and denominator of the investigated function are known as a result of symbolic analysis. Some subexpressions of the numerator and denominator are components of the sensitivity formulas given below. How to find the subexpressions is also described here. Thus the paper gives a complete tool of using symbolic network functions in sensitivity analysis. Nota

bene to use the proper formulas we should know the degrees of numerator and denominator of the investigated network function. How to calculate the degrees knowing only kinds of elements creating a LLS network is described in [1–3]. A paper containing proofs, omitted here, and presenting the problem for the general situation of the function being a ratio of polynomials is in preparation.

NOTATION REMARKS

To avoid repetitions in the text we introduce special notations of assumptions which will appear more than once:
I. $F = F(x, y, z, \xi,...)$ is a function of variables $x, y, z, \xi,...$ and $F = L/M$, where L and M are also functions of the varialbes $x, y, z, \xi,...$; II. $L = L(x, y, z, \xi,...) = \hat{l}_0 + \hat{l}_1 x + \hat{l}_2 x^2$, $M = M(x, y, z, ...) = \hat{m}_0 + \hat{m}_1 x + \hat{m}_2 x^2$, where $\hat{l}_i = \hat{l}_i(y, z, \xi,...) = \text{const}(x)$, $\hat{m}_i = \hat{m}_i(y, z, \xi,...) = \text{const}(x)$ and some $\hat{l}_i$, $\hat{m}_i$ can be equal to zero; III. $L = L(x, y, z, \xi,...) = \check{l}_0 + \check{l}_1 y + \check{l}_2 y^2$, $M = M(x, y, z, \xi,...) = \check{m}_0 + \check{m}_1 y + \check{m}_2 y^2$, where $\check{l}_i = \check{l}_i(x, z, \xi,...) = \text{const}(y)$, $\check{m}_i = \check{m}_i(x, z, \xi,...) = \text{const}(y)$ and some $\check{l}_i$, $\check{m}_i$ can be equal to zero; IV. $L = \Sigma_{i=0}^{2}\Sigma_{j=0}^{2} l_{ij} x^i \cdot y^j$, $M = \Sigma_{i=0}^{2}\Sigma_{j=0}^{2} m_{ij} x^i y^j$, where $l_{ij} = l_{ij}(z, \xi,...) = \text{const}(x, y)$, $m_{ij}(z, \xi,...) = \text{const}(x, y)$ and some l_{ij}, m_{ij} can be equal to zero.

FORMULAS

For completion we recollect the well known definitions of sensitivities of different kinds and some relations among them (a collection of definitions and relations is given by (1)–(14)). Further we will give formulas only for S_x^F, $s_{x,x}^F$, $s_{x,y}^F$ and SW_x^F.

$S_x^F \triangleq \frac{x}{F}\frac{\partial}{\partial x}F$ (1)	$s_x^F \triangleq \frac{\partial}{\partial x}F = \frac{F}{x}S_x^F$ (2)
$\tilde{Q}_x^F \triangleq \frac{1}{F}\frac{\partial}{\partial x}F = \frac{1}{x}S_x^F$ (3)	$Q_x^F \triangleq x\frac{\partial}{\partial x}F = FS_x^F$ (4)
$s_{x,x}^F \triangleq \frac{\partial^2}{\partial x^2}F$ (5)	$S_{x,x}^F \triangleq \frac{x^2}{F}\frac{\partial^2}{\partial x^2}F = \frac{x^2}{F}s_{x,x}^F$ (6)
$\tilde{Q}_{x,x}^F \triangleq \frac{1}{F}\frac{\partial^2}{\partial x^2}F = \frac{1}{F}s_{x,x}^F$ (7)	$Q_{x,x}^F \triangleq x^2\frac{\partial^2}{\partial x^2}F = x^2 s_{x,x}^F$ (8)
$s_{x,y}^F \triangleq \frac{\partial^2}{\partial x \partial y}F = s_{y,x}^F$ (9)	$S_{x,y}^F \triangleq \frac{xy}{F}\frac{\partial^2}{\partial x \partial y}F = S_{x,y}^F = \frac{xy}{F}s_{x,y}^F$ (10)
$s_{x,y}^F \triangleq \frac{\partial^2}{\partial x\,\partial y}F = = \tilde{Q}_{y,x}^F = \frac{1}{F}s_{x,y}^F$ (11)	$Q_{x,y}^F \triangleq xy\frac{\partial^2}{\partial x\,\partial y}F = = Q_{y,x}^F = xy \cdot s_{x,y}^F$ (12)
$SW_x^F \triangleq \frac{x}{F}\frac{\Delta F}{\Delta x}$ (13)	$\delta_x^F \triangleq \frac{S_x^F - SW_x^F}{SW_x^F}$ (14)

Appropriate formulas for the rest of sensitivities mentioned in the collection can be constructed easily on the base of the formulas given later and the relations shown in the collection.

FORMULAS FOR S_x^F AND $s_{x,x}^F$

Formulas (1a) and (5b) are true when the function F of variable x is bilinear (i.e. assumptions I and II are satisfied with $\hat{l}_2 = \hat{m}_2 = 0$). Formulas (1b) and (5b) are true when the function F of variable x is biquadratic (i.e. assumptions I and II are satisfied).

$S_x^F = \frac{\hat{m}_0}{M} - \frac{\hat{l}_0}{L}$ (1a)	$s_{x,x}^F = \frac{2F}{x^2} \cdot S_x^F \cdot \left(\frac{\hat{m}_0}{M} - 1\right)$ (5a)
$S_x^F = \frac{\hat{m}_0 - x^2 \cdot \hat{m}_2}{M} - \frac{\hat{l}_0 - x^2 \cdot \hat{l}_2}{L}$ (1b)	$s_{x,x}^F = \frac{2F}{x^2}\left[\frac{x^2 \hat{l}_2}{L} - \frac{x^2 \hat{m}_2}{M} + S_x^F\left(\frac{\hat{m}_0 - x^2 \hat{m}_2}{M} - 1\right)\right]$ (5b)

FORMULAS FOR $s_{x,y}^F$

Formula (9a) is true when the function F of variables x and y is bilinear (i.e. assumptions I, II, III and IV are satisfied with $\hat{l}_2 = \hat{m}_2 = \check{l}_2 = \check{m}_2 = 0$). Formula (9b) is true when the function F of variables x, y is biquadratic with respect to x and bilinear with respect to y (i.e. assumptions I, II, III and IV are satisfied with $\check{l}_2 = \check{m}_2 = 0$). Formula (9c) is true when the function F of variables x and y is biquadratic (i.e. assumptions I, II, III and IV are satisfied).

$$s_{x,y}^F = \frac{F}{xy}\left(S_x^F \frac{\check{m}_0}{M} + S_y^F \frac{\hat{m}_0}{M} + \frac{l_{00}}{L} - \frac{m_{00}}{M}\right) \tag{9a}$$

$$s_{x,y}^F = \frac{F}{xy}\left(S_x^F \frac{\check{m}_0}{M} + S_y^F \frac{\hat{m}_0 - \hat{m}_2 x^2}{M} + \frac{l_{00} - l_{20} x^2}{L} - \frac{m_{00} - m_{20} x^2}{M}\right) \tag{9b}$$

$$s_{x,y}^F = \frac{F}{xy}\left(S_x^F \frac{\check{m}_0 - \check{m}_2 y^2}{M} + S_y^F \frac{\hat{m}_0 - \hat{m}_2 x^2}{M} + \right.$$
$$\left. + \frac{l_{00} - l_{20} x^2 - l_{02} y^2 + l_{22} x^2 y^2}{L} - \frac{m_{00} - m_{20} x^2 - m_{02} y^2 + m_{22} x^2 y^2}{M}\right) \tag{9c}$$

FORMULAS FOR SW_x^F

Formula (13a) is true when the function F of variable x is bilinear (i.e. assumptions I and II are satisfied with $\hat{l}_2 = \hat{m}_2 = 0$). Formula (13b) is true when the function F of variable x is biquadratic (i.e. assumptions I and II are satisfied).

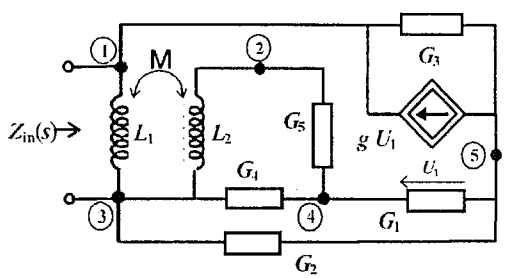
$$SW_x^F = \left(\frac{\hat{m}_0}{M} - \frac{\hat{l}_0}{L}\right)\left[1 + \frac{\Delta x}{x}\left(1 - \frac{\hat{m}_0}{M}\right)\right]^{-1} \tag{13a}$$

$$SW_x^F = \frac{\dfrac{\hat{m}_0}{M} - \dfrac{\hat{l}_0}{L} - \left(1 + \dfrac{\Delta x}{x}\right)\left(\dfrac{\hat{m}_2 x^2}{M} - \dfrac{\hat{l}_2 x^2}{L}\right)}{1 + \dfrac{\Delta x}{x}\left(1 - \dfrac{\hat{m}_0 - \hat{m}_2 x^2}{M}\right) + \left(\dfrac{\Delta x}{x}\right)^2 \dfrac{\hat{m}_2 x^2}{M}} \tag{13b}$$

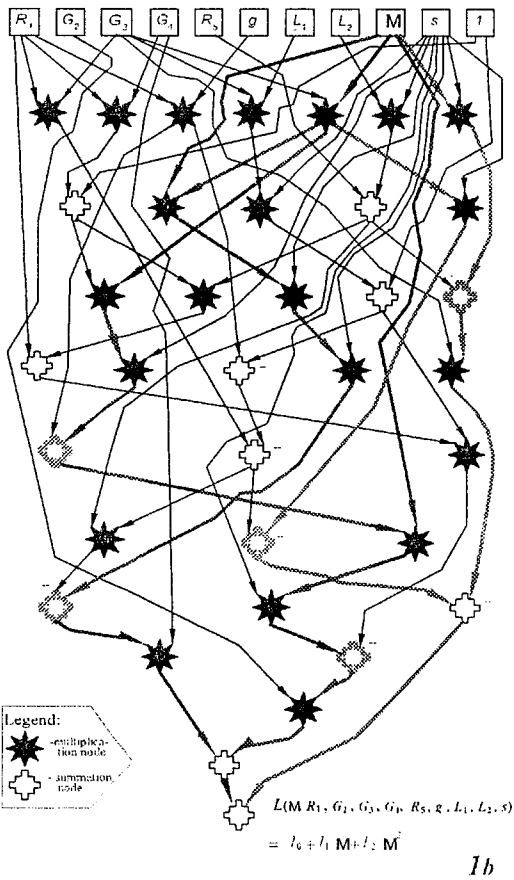
ALGORITHMS

To use effectively the formulas given above we need to know the expressions for the numerator L and the denominator M of the analysed function F. Of course, when we find F by symbolic analysis we obtain the expressions for L and M. In some methods (for example [4] and other mentioned in [5]) the results (i.e. L and M) have the form of nested expressions. To find $\hat{l}_0$, $\hat{m}_0$, $\check{l}_0$, $\check{m}_0$ and $\hat{l}_2 x^2$, $\hat{m}_2 x^2$, $\check{l}_2 x^2$, $\check{m}_2 x^2$ from M and L, two new algorithms were worked out. These algorithms are given below.

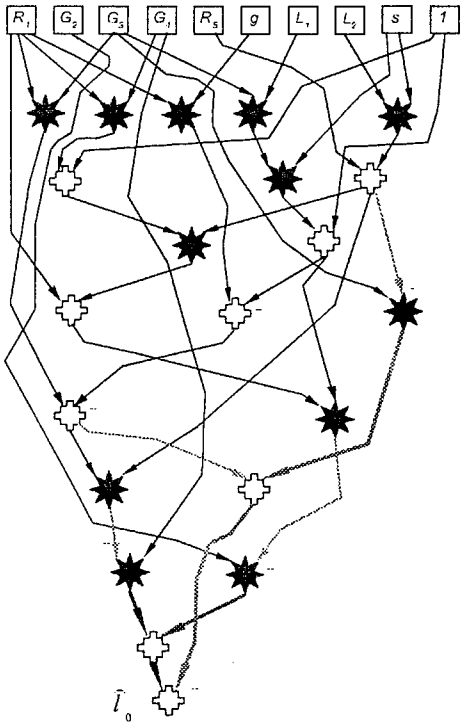
Algorithm 1: Suppose that P is a polynomial of variable q (i.e. $P = P(q) = \Sigma_{i=0}^{d(P,q)} p_i q^i$) and a function of other variables (P can be given in a disordered form). In order to create a graph of operations for $p_0 = P(0)$ the following operations and transformations in the graph for $P(q)$ have to be made: 1° determine all paths from q to P, 2° mark first summation nodes of these paths, 3° eliminate all nodes and branches which precede the marked nodes in the paths, 4° short-circuit branches entering these marked nodes that have input degrees equal to one and rename node P into $P(0)$, 5° eliminate nodes with zero output degree except node $P(0)$. *Algorithm 2:* Suppose that P is as above but $d(P, q) \leqslant 2$. In order to create a graph of operations for $p_2 q^2$ the following operations and transformations in the graph for $P(q)$ have to be made: 1° determine the set Ψ of all paths from q to P, 2° find the subset ψ of Ψ having the following proprieties: a) each path from ψ touches with another one from ψ in a multiplication node, b) no one path from $\Psi - \psi$ touches with another one from Ψ in a multiplication node, 3° eliminate each branch belonging to a path of Ψ but not belonging to any path of ψ, 4° eliminate each branch incoming to a summation node of a path from ψ and not belonging to any path of ψ, 5° for all nodes having a zero output degree (except node P) eliminate branches incoming to these nodes and then these nodes; repeat the operation if possible, 6° short-circuit branches entering summation or multiplication nodes that have their input degrees equal to one (move the weight of each such the node onto the preceding node), 7° rename the node P into $p_2 q^2$.



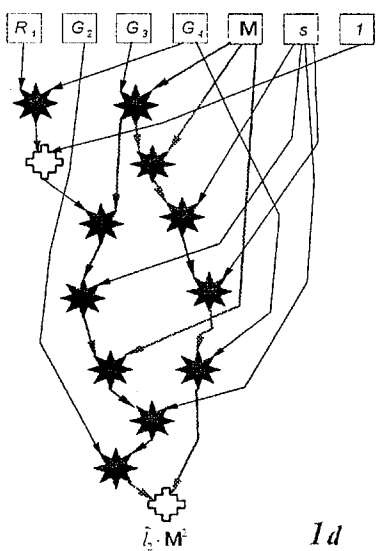
1a



1b



1c



1d

Fig. 1.

Example. For the network in Figure 1a the input impedance $Z_{in}(s)$ was found with the symbolic analysis program described in [5]. Let us denote the mutual inductance **M** (see Fig. 1a) with x. Then $Z_{in}(s) = L(x,...)/M(x,...)$ is the biquadratic function of x. The graph of operations for the numerator L is shown in Figure 1b. To find various sensitivity functions for $Z_{in}(s)$ we need, among others, subexpressions l_0 and $l_2 x^2$ of L. These subexpressions can be found with the algorithms 1 and 2 (we should use them with $P = L$ and $q = x$). The results of applying the algorithms to the graph of L are shown in Figure 1c (graph for l_0) and 1d (graph for $l_2 x^2$).

The graphs presented in the example are quite complicated (because $L(x,...)$ is also not simple) but in fact the graphs of operations are only used for internal computer manipulations, for computers storage of the nested results and for clear and easy description of various algorithms applied to the group of symbolic methods giving such results.

Use of the formulas: The algorithms 1 and 2 allow to determine all the parts of polynomials that are needed in sensitivity formulas in the case of bilinear and biquadratic functions. Figure 2 explains how to do it for the polynomial of denominator M. For example to find $y^2 m_{02}$ we should use the Algorithm 2 with $P = \hat{m}_0$ and $q = y$, but to find $\hat{m}_0$ we should use the Algorithm 1 with $P = M$ and $q = x$. Exchanging M with L and letter m with letter l we obtain the graph explaining the same for the polynomial of numerator L. These graphs and formulas given before allow us to realise (manually or by means of a computer) a very effective path of calculating different kinds of sensitivity, when the symbolic form of the numerator and denominator of the analysed function is known.

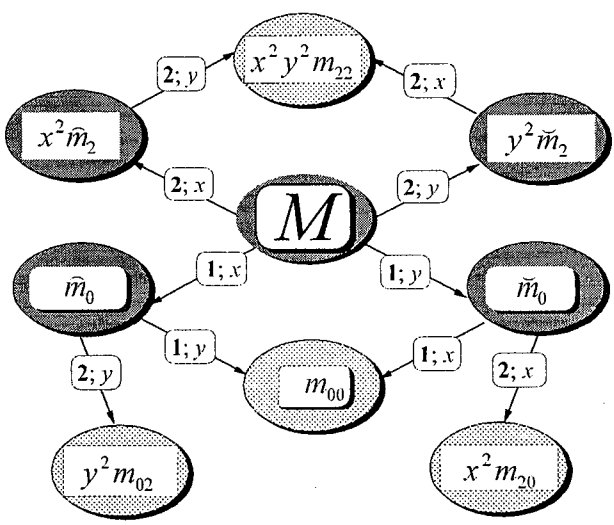


Fig. 2

REFERENCES

1. P.M. Lin: *A survey of applications of symbolic networks functions.* IEEE Trans. on CAS, 1973, CAS-20, pp. 732 – 737
2. K. Géher: *Theory of network tolerances.* Akadémiai Kiadó, Budapest 1971
3. C. Stefański: *A generalisation of Lin's theorem on bilinear form of network functions.* Proc. IEEE ISCAS, 1984, pp. 887 – 890
4. M.M. Hassoun and K.S. McCarville: *Symbolic analysis of large scale networks using a hierarchical signal flowgraph approach.* Analog. Integr. Circ. Signal Process (Netherlands), 1993, vol. 3, pp. 31 – 42
5. C. Stefański: *On a self-optimising algebraic approach to symbolic analysis.* Intern. J. of Circuit Theory and Applications, 1994, vol. 22, pp. 447 – 466

CZ. STEFAŃSKI

O OBLICZENIU WRAŻLIWOŚCI Z WYKORZYSTANIEM SYMBOLICZNYCH FUNKCJI UKŁADOWYCH O POSTACI ZAGNIEŻDŻONEJ

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

W artykule zestawiono znane i dodano nowe wzory do obliczania wrażliwości funkcji układowych. Są one przydatne w sytuacji, gdy badana funkcja układowa jest podana w postaci symbolicznej, a nieodzowne, gdy funkcja ta ma postać wielokrotnie zagnieżdżonych wyrażeń symbolicznych. Podano także nowy algorytm sortowania wyrażeń — dla funkcji bikwadratowych. Artykuł daje brakujące narzędzia do wykorzystania wyników analizy symbolicznej w obliczaniu wrażliwości.

Słowa kluczowe: analiza symboliczna, analiza wrażliwości, symboliczne funkcje układowe, układy SLS.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1997, 43, z. 3,ss. 321–332

Nonlinear negative capacitance in oscillator circuit

LECH TOMAWSKI, MAREK MAŃKA

Instytut Fizyki, Uniwersytet Śląski

JERZY DĄBROWSKI

Instytut Problemów Techniki, Uniwersytet Śląski

Received 1997.03.15

Authorized 1997.05.15

Properties of RLC and DCG resonant circuits are compared. Nonlinear negative capacitance is described and its usage in oscillator with DCG resonant circuit is proposed. This oscillator converts a capacitance into AC voltage. Losses of the converted capacitance can be very big.

Key words: resonant circuits, oscillators circuits.

1. INTRODUCTION

In conventional oscillators oscillations occur when resistance R of the resonant circuit RLC is compensated by a negative resistance realised by an amplifier unit. Voltage of oscillations is determined by the natural nonlinearity of the amplifier.

In oscillators with DCG resonant circuits [1, 2] (where D denotes FDNC for which $Y_{(S)} = s^2 D$, C — capacitance, G — conductance), oscillations occur when capacitance C is compensated by negative capacitance realised at the input terminals of an amplifier unit with capacitance feedback. This negative capacitance must be treated as nonlinear if the amplitude of oscillations is assumed as determined. In order to obtain a strong dependence between voltage and capacitance C, strong nonlinearity should be introduced into the oscillator circuit using a special nonlinear amplifier unit.

In our paper an oscillator of this kind, acting as converter of capacitance into AC voltage is proposed. It is shown that this converter operates in the region of very high losses of the converted capacitance.

2. COMPARISON RLC AND DCG CIRCUITS

Fig. 1a shows an RLC resonant circuit supplied by the resistor R_g and Fig. 1b shows a DCG circuit supplied by capacitor C_g.

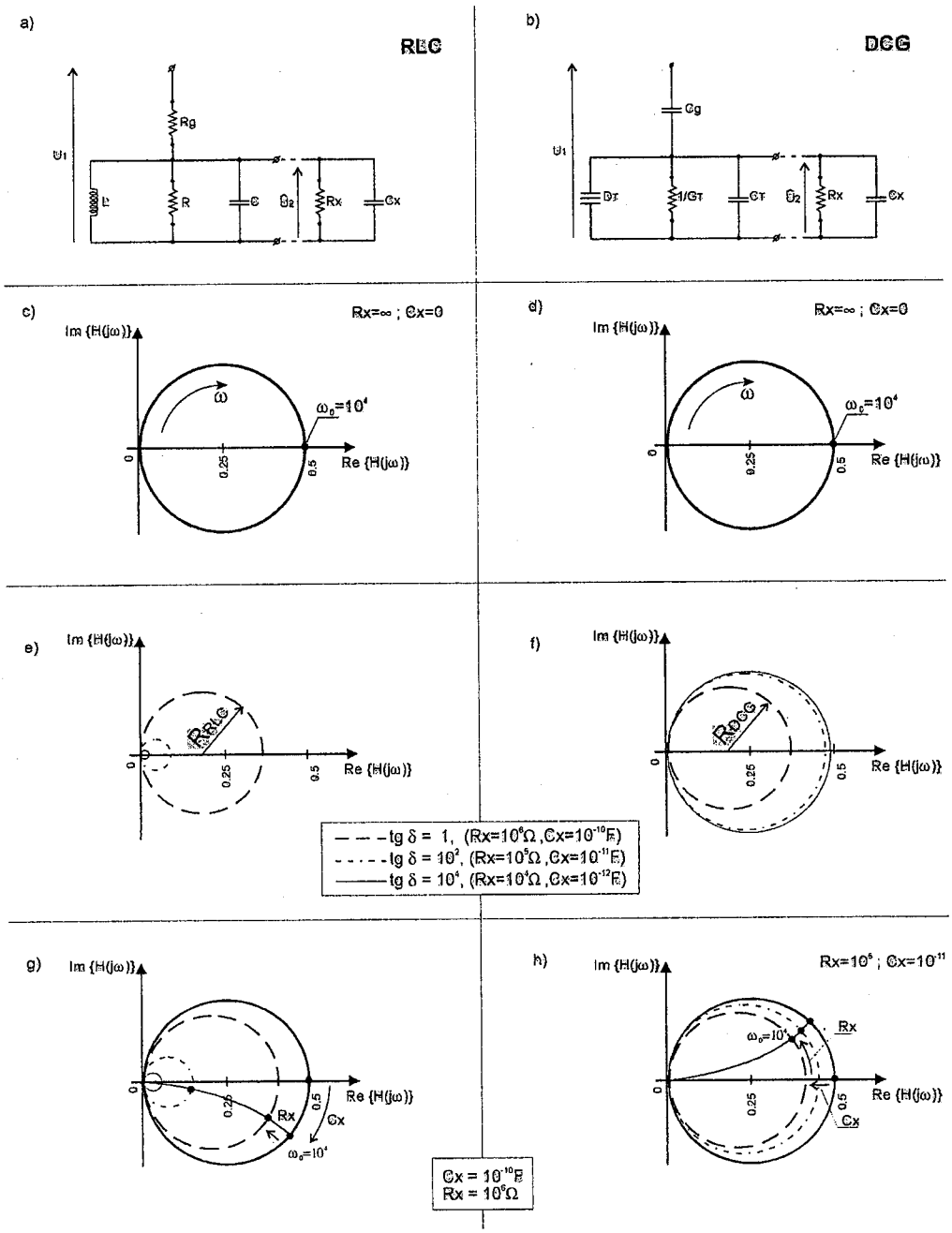


Fig. 1. Polar plots of RLC and DCG circuits

To compare these circuits it is also assumed that the values of elements C_g, C_T, G_T, D_T are calculated from the following formulas:

$$C_g = \frac{1}{\omega_0 R_g}, \quad C_T = \frac{1}{\omega_0 R}, \quad G_T = \frac{1}{\omega_0 L}, \quad D_T = \frac{C}{\omega_0}. \tag{1}$$

resulting from Bruton's transformation [3], and that resonant frequencies of both circuits are the same: $\omega_{0\,\mathrm{RLC}} = \omega_{0\,\mathrm{DCG}} = \omega_0 = 10^4$ rad/s. Figures 1c and 1d illustrate this case for the following values of elements: $R = R_g = 10^6\ \Omega$, $C = 10^{-8}$ F, $L = 1$H, $C_T = C_g = 10^{-10}$ F, $1/G_T = 10^4\ \Omega$, $D_T = 10^{-12}$ As^2/V (C_x and R_x are not connected). Then polar plots of both of the transfer function for circuits have the same radii and ω_0 lies on the real axes of $\underline{H(j\omega)}$.

C_x denotes the unknown capacitance whose value is converted to AC voltage if the resonant circuit is connected to the oscillator circuit R_x represents the losses of capacitance C_x. It is assumed that the dissipation factor $\tan\delta = \dfrac{1}{\omega_0 C_x R_x}$ has a high value, since the resistance R_x has a suitably low value.

If the capacitance C_x and the resistance R_x are connected to the RLC circuit then R_x causes a decrease in voltage and C_x causes a decrease in the resonant frequency of the circuit. If the same elements C_x and R_x are connected to the DCG resonant circuit then voltage $\hat{U}_2$ will decrease due to C_x and the resonant frequency will increase due to R_x.

Transfer functions $\hat{H}(j\omega) = \dfrac{\hat{U}_2}{U_1}$ for both resonant circuits, when unknown elements C_x and R_x (unknown sample $C_x - R_x$) are connected to them, can be described as:

$$\underline{\hat{H}(j\omega)}_{\mathrm{RCL}} = \frac{\dfrac{1}{R_g}\left(\dfrac{1}{R_g}+\dfrac{1}{R}+\dfrac{1}{R_x}\right) - j\dfrac{\omega}{R_g}\left(C + C_x - \dfrac{1}{\omega^2 L}\right)}{\left(\dfrac{1}{R_g}+\dfrac{1}{R}+\dfrac{1}{R_x}\right)^2 + \omega^2\left(C + C_x - \dfrac{1}{\omega^2 L}\right)^2} \tag{2}$$

and

$$\underline{\hat{H}(j\omega)}_{\mathrm{DCG}} = \frac{\omega^2 C_g(C_g + C_T + C_x) + j\omega C_g\left(\dfrac{1}{R_T}+\dfrac{1}{R_x} - \omega^2 D_T\right)}{\left(\dfrac{1}{R_T}+\dfrac{1}{R_x} - \omega^2 D_T\right)^2 + \omega^2(C_g + C_T + C_x)^2}. \tag{3}$$

It is easy to show that the radii of the polar plots of $\underline{\hat{H}(j\omega)}$ for both circuits are equal to half of the real part of the coirresponding transmittance $\hat{H}(j\omega)$ for the resonant angular frequency $\omega_{0\,\mathrm{RLC}}$ or $\omega_{0\,\mathrm{DCG}}$ respectively. By connecting elements $C_x - R_x$ a slight shift of the resonant angular frequency occurs, which for the RLC circuit is:

$$\omega_{0\,\mathrm{RLC}} = \frac{1}{\sqrt{L(C + C_x)}}, \tag{4}$$

and for the DCG circuit:

$$\omega_{0\,\mathrm{DCG}} = \frac{1}{\sqrt{D_T \dfrac{R_T R_x}{R_T + R_x}}}. \tag{5}$$

Assuming that $\omega_{0\,\mathrm{RLC}} \approx \omega_{0\,\mathrm{DCG}} \approx \omega_0 = 10^4$ rad./s, the radius of the polar plot for circuit RLC can be written as:

$$R_{\mathrm{RLC}} = \frac{\dfrac{1}{R_g}}{2\left(\dfrac{1}{R_g} + \dfrac{1}{R} + \dfrac{1}{R_x}\right)} \tag{6}$$

and for circuit DCG as:

$$R_{\mathrm{DCG}} = \frac{C_g}{2(C_g + C_T + C_x)}. \tag{7}$$

To compare these radii, formulas (1) should be inserted to (6), giving:

$$R_{\mathrm{RLC}} = \frac{\omega C_g}{2\left(\omega C_g + \omega C_T + \dfrac{1}{R_x}\right)}. \tag{8}$$

Capacitance C_x is not transformed to resistance, since the unknown sample $C_x - R_x$ is also connected to another circuit. Formula (8) will be equal to (7) if $\dfrac{1}{R_x} = \omega C_x$ which is only possible if the dissipation factor $\tan\delta = \dfrac{1}{\omega_0 R_x C_x}$ has the value 1, i.e. for $\tan\delta = 1$ the radii of $\hat{H}(j\omega)$ for polar plots of both circuits have the same value, (which can be easily calculated as 0.166 for $C_x = 10^{-10}$ F and $R_x = 10^6\ \Omega$). The approximation made above yields the resonant angular frequency calculated from formula (4) is $0.995 * 10^4$ rad./s and from formula (5) is $1.005 * 10^4$ rad./s.

It may also be shown that for $\tan\delta > 1$ the radii of the polar plots of $\hat{H}(j\omega)$ for circuit RLC are smaller, and for the circuit DCG they are greater than the radius for $\tan\delta = 1$. This is shown in Fig. 1e and 1f, when $C_x = 10^{-11}$ F and $R_x = 10^5\ \Omega$ for $\tan\delta = 10^2$, and $C_x = 10^{-12}$ F and $R_x = 10^4\ \Omega$ for $\tan\delta = 10^4$. Science the radii of the polar plots the $\hat{H}(j\omega)_{\mathrm{RLC}}$ (Fig. 1e) decrease to very small values when $\tan\delta$ increases, the use of RLC resonant circuit in such a case is impossible, because then voltage $\hat{U}_2$ also decreases to very small values. The changes occurring when capacitance C_x and resistance R_x are connected to the RLC or DCG circuits are shown in Fig. 1g and 1h. All polar plots of $\hat{H}(j\omega)$ shown in Fig. 1 were calculated from formulas (2) and (3).

It is clear that there is an important difference between RLC and DCG circuits i.e. RLC resonant circuits can be used in the region of low dissipation factor and DCG resonant circuits in the region of high dissipation factor.

3. NONLINEAR NEGATIVE CAPACITANCE

The nonlinear amplifier unit which realises the nonlinear negative capacitance is shown on the right side of Fig. 2. The output voltage U_{2-2} of the amplifier is determined by the FET BF245 acting as a variable resistor. The voltage, taken as the difference between the rectified output voltage and the reference voltage U_0 and amplified by OA_2 is fed to the gate of the FET. The phase shift of the amplifier is O^0. Adjusting of the phase shift is achieved by changing the setting of capacitor C_s. The transfer characteristic of the amplifier is shown on Fig. 3a. As the amplifier total harmonic distortions are low and the output voltage is fed to the resonant circuit, hence the nonlinear properties of the amplifier can be analysed by the method of describing functions [4]. The describing function $g(u_{1-1})$ for nonlinear amplifier is shown on Fig. 3b. Since the feedback of the amplifier is capacitive (C_g on the Fig. 2) then at the input terminals 1 – 1 a negative nonlinear capacitance occurs, as in the formula:

$$C_n = [1 - g(U_{1-1})]\, C_g. \tag{9}$$

The values of C_n, both measured and calculated from (9), are shown on Fig. 3c.

4. OSCILLATOR WITH RESONANT CIRCUIT DCG

The resonant circuit DCG is connected to the input terminals 1 – 1 of the nonlinear amplifier, as shown on the left side of Fig. 2. This circuit consists of FDNC circuit, capacitance C_c, conductance $1/R_T$ and an unknown elements $C_x - R_x$. This FDNC circuit was proposed by Horn and Moschytz in [5]. The equivalent diagram of the Horn and Moschytz circuit contains ideal FDNC, capacitance and conductance. They are connected in parallel and their values are described by the following formulas:

$$D_T = \frac{C_1 C_3 g_0}{a(g_1 + g_0)}, \tag{10}$$

$$C_{\mathrm{FDNC}} = \frac{C_1 g_0 h + a g_0 (C_1 + C_3) - C_3 g_1 d_0}{a(g_1 + g_0)} \tag{11}$$

$$G_{\mathrm{FDNC}} = \frac{a g_0 (g_1 + h) - g_1 h d_0}{a(g_1 + g_0)}. \tag{12}$$

Symbols a, g_0, g_1, d_0, h denote conductances. The value of D_T was 10^{-11} As^2/V and does not change with frequency. The capacitance C_{FDNC} was negative having a value of -150 pF and was compensated by the capacitance C_c. The value of conductance G_{FDNC} was very small and was hence neglected. The resonance phenomenon occurs between D_T and resultant resistance of parallel connection R_T and R_x. The resonant frequency was 1620 Hz (see Fig. 5b).

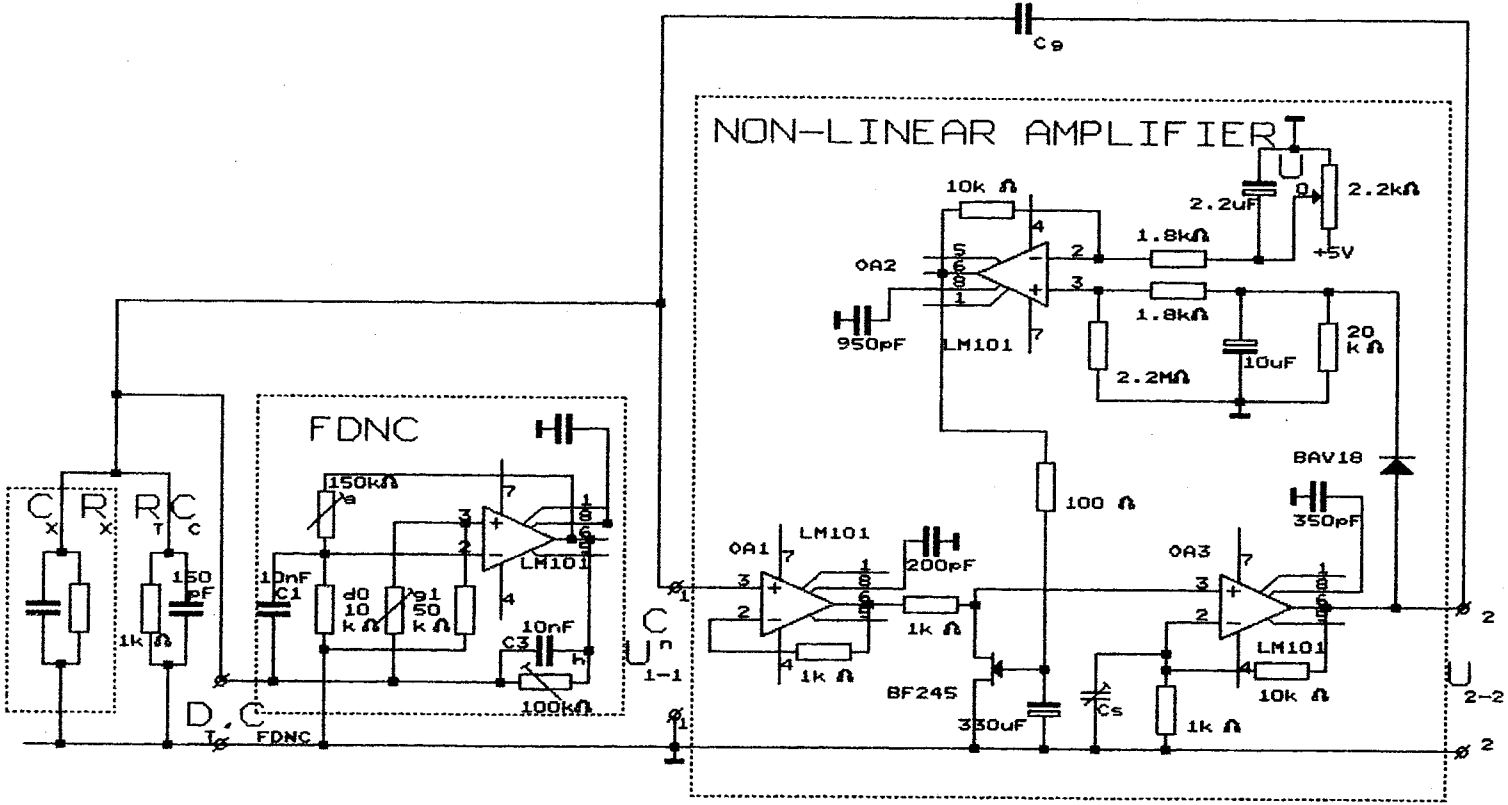


Fig. 2. Scheme of an oscillator with DCG resonant circuit and nonlinear amplifier unit

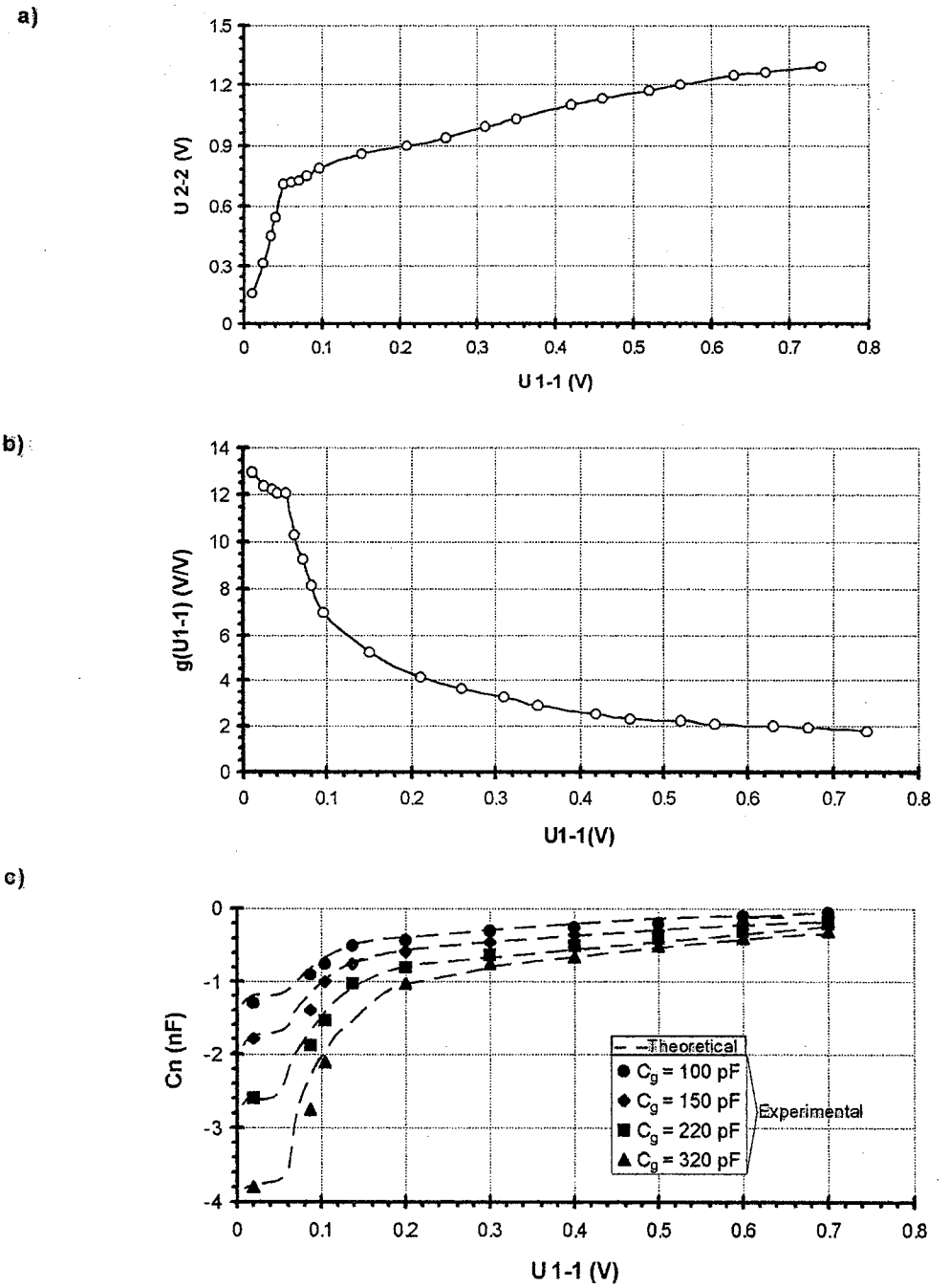


Fig. 3. Transfer characteristic of the nonlinear amplifier unit (a), its describing function (b) and nonlinear negative capacitance (c) vs. input voltage U_{1-1}

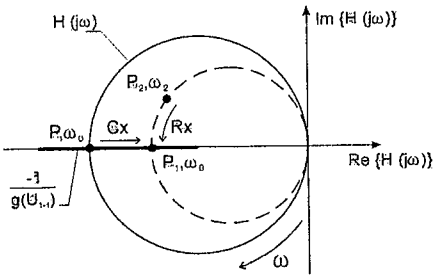


Fig. 4. Influence of C_x and R_x on voltage U_{1-1} and frequency of oscillation f

Other FDNC circuits appropriate for this application can be realised with the use of operational amplifiers or current conveyors CCII. The former, acting in general as a suitable applications of the Generalised Immitance Converters is reported by Antoniou [6], Riordan [7], Bui and Huynh [8]. A good solution is also reported by Genin [9]. The latter concerning current conveyors are widely described in literature [10, 11] however not enough technical parameters or results of measurements have been reported. Practically verified stable circuits realising FDNC elements, using current conveyors can be found in [12, 13] and [14].

The point P (on Fig. 4) where the polar plot of the transfer function of linear part of the circuit intersects the function $-1/g(u_{1-1})$ determines the frequency and voltage of oscillations. Connecting the capacitance C_x and resistance R_x to the resonant circuit causes shifting of point P to P_1 due to C_x and displacement of point P_2 to P_1 due to R_x (angular frequency ω_0 then increases to ω_2).

In practice, resistance R_x does not influence the frequency and voltage of oscillations if the resistance R_T is many times smaller then R_x. In the case described in this paper the resistance R_T was 1 kΩ (see Fig. 2), but if necessary the resonance phenomenon in a circuit containing the Horn and Moschytz FDNC can also be achieved with R_T equal to 50 Ω.

5. PRACTICAL RESULTS

The circuit described above was tested as a converter of capacitance into AC voltage. The AC voltage from terminals 1−1 as a function of capacitance C_x is shown in Fig. 5a. Four curves were measured for various values of feedback capacitance C_g. Theoretical curves were calculated as follows:

1) Value of the nonlinear negative capacitance C_n was found from formula (9).
2) Total capacitance occuring at terminals 1−1 of the nonlinear amplifier was: $C_c + C_{\text{FDNC}} + C_x = C_x$ (since $C_c = 150$ pF and $C_{\text{FDNC}} = -150$ pF).
3) Capacitance C_x was compared with nonlinear negative capacitance C_n in Fig. 3c and from this voltage u_{1-1} was determined.

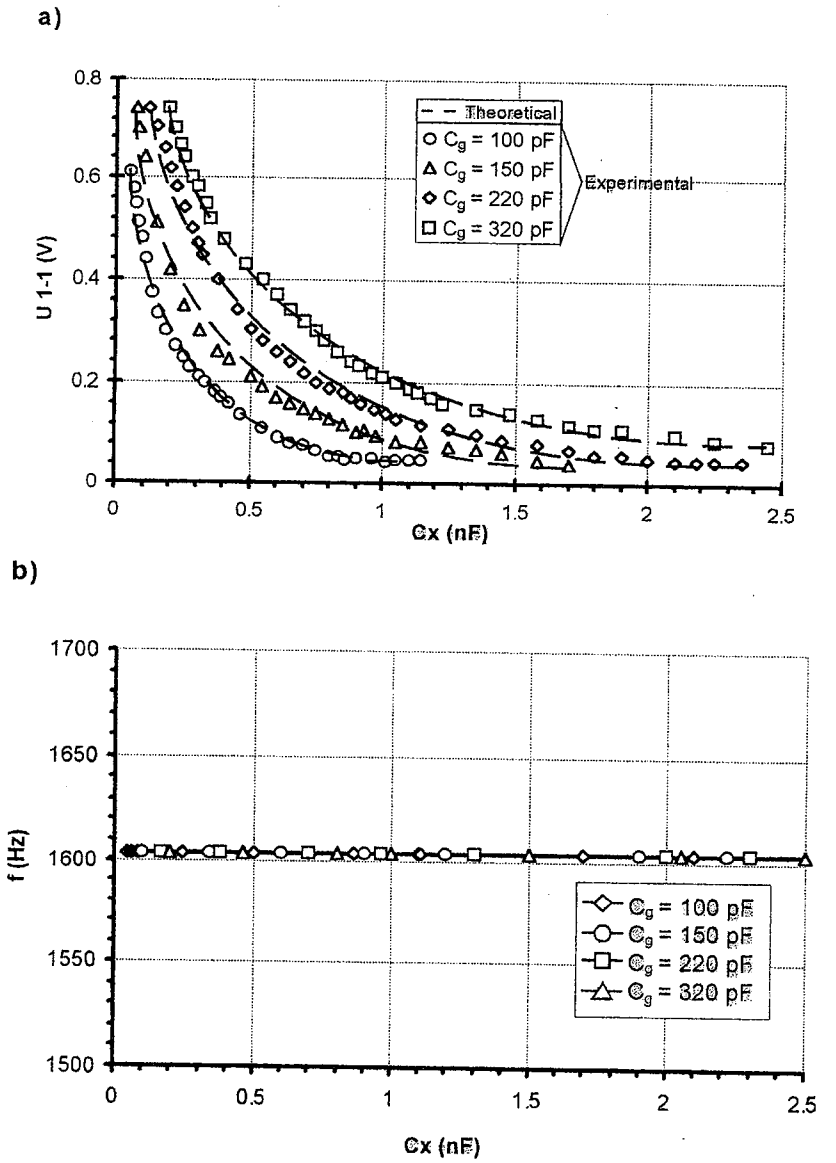
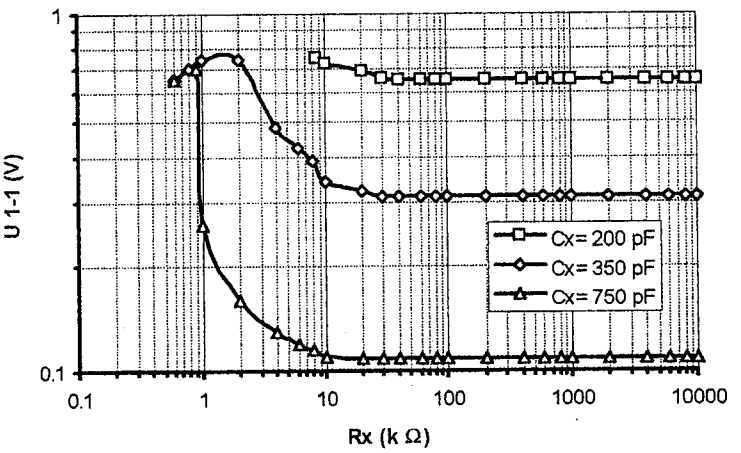


Fig. 5. Voltage U_{1-1} and frequency f vs. capacitance C_x

No influence of capacitance C_x on the frequency of oscillations was found. Relevant curves are shown in Fig. 5b. The influence of resistance R_x on voltage U_{1-1} and on frequency is shown in Figures 6a and 6b. It is clear that this influence can be neglected since resistance R_x is greater then 20 kΩ. Taking into account that

a)



b)

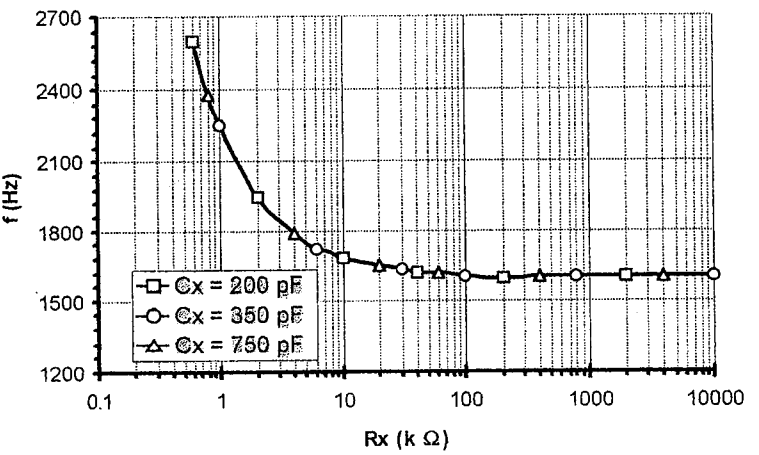


Fig. 6. Voltage U_{1-1} and frequency f vs. resistance R_x

the range of converted capacitance C_x was from about 100 pF to about 1 nF (see Fig. 5a, curve for $C_g = 100$ pF) and assuming the lowest value of resistance R_x as 20 kΩ (see Fig. 6a, b) and taking value of angular frequency ω about 10^4 (1620 Hz on the Fig. 5b), it can be calculated that value of $\tan\delta = 1/(\omega C_x R_x)$ of the converted capacitance C_x should be less than 500 for $C_x = 100$ pF, or 50 for $C_x = 1$ nF.

CONCLUSION

The oscillator described above makes it possible to convert capacitance into AC voltage. If the dissipation factor of the capacitance, $\tan\delta$, is less then 500, the losses do not affect the conversion accuracy. Then, this circuit can be used for a humidity measurements of the paper or fabric. There is a possibility to spread the range of dissipation factor up to a value of few thousand by tuning the resonance frequency at resistance R_T about $50\,\Omega$.

REFERENCES

1. Fr. Molo: *Parallel resonator with a resistance and a frequency-dependent negative resistance realised with a single operational amplifier.* IEEE Trans. Circuit and Systems, 1974, vol. 21, no. 1, pp. 783 – 788
2. A.M. Soliman, S.S. Awad: *Canonical high-selectivity parallel resonator, using a single operational amplifier, and its applications in filters.* Electronic Circuits and Systems, 1977 vol. 1 no 4, pp. 145 – 148
3. L.T. Bruton: *Network transfer functions using the concept of frequency-dependent negative resistance.* IEEE Trans. Circuit Theory, 1969 vol. 16, no 8 pp. 406 – 408
4. F.H. Raven: *Automatic Control Engineering.* McGraw-Hill Book Company, 1968
5. P. Horn, G.S. Moschytz: *Active RC single-opamp design of driving-point impedances.* IEEE Trans. Circuit and Systems, 1979, vol. 26, no. 1, pp. 22 – 29
6. A. Antoniou: *Realisation of gyrators using operational amplifiers.* IEE Proc., 1969, vol. 116, no. 11, pp. 1838 – 1850
7. R.H.S. Riordan: *Simulated inductors using differential amplifiers.* Electron. Lett., 1967, vol. 3, no. 2, pp. 50 – 51
8. N.C. Bui, L.T. Huynh: *Realization de FDNR et FDNC avec des convertisseurs de courant.* Agen Mitteilungen, 1976, vol. 20, pp. 29 – 40
9. R. Genin: *A sine wave generator using a frequency-dependent negative conductance.* Proc. IEEE, 1975, vol. 63, pp. 1611 – 1612
10. U. Kumar, S.K. Shukla: *Recent development in current conveyors and their applications.* Microelectronics Journal, 1985, vol. 16, no. 1, pp. 47 – 52
11. B. Wilson: *Recent development in current conveyors and current-mode circuits.* IEE Proc. G, 1990, vol. 137, no. 2, pp. 63 – 77
12. A.M. Soliman: *New generalized-immitance converter circuits obtained by using the current conveyor.* Int. J. Electronics, 1972, vol. 32, no. 6, pp. 673 – 679
13. Ibid: *Active RC realization of current transfer functions using voltage generalized-immitance converters.* Int. J. Electronics, 1972, vol. 33, no. 3, pp. 273 – 280
14. M. Higashimura, Y. Fukui: *Type 1 mutator using current conveyor and its application to immitance simulation.* Int. J. Electronics, 1988, vol. 64, no. 3, pp. 377 – 383

L. TOMAWSKI, M. MAŃKA, J. DĄBROWSKI

NIELINIOWA POJEMNOŚĆ UJEMNA W UKŁADZIE OSCYLATOROWYM

Streszczenie

Przedstawiono porównanie własności układów rezonansowych RLC i DCG. Wskazano zastosowanie nieliniowej pojemności ujemnej w oscylatorze z rezonansowym układem DCG. Praktyczne wyniki prac zilustrowano pięcioma charakterystykami uzyskanymi przy różnych wartościach pojemności.

Słowa kluczowe: układy rezonansowe, układy oscylatorowe.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1997, **43**, z. 3, ss. 333–352

Key-controlled bit permutations in substitution-permutation networks

ALBERT SADOWSKI

Instytut Podstaw Elektroniki, Politechnika Warszawska

Otrzymano 1997.02.27

Autoryzowano 1997.05.08

In the early 90th E. Biham and A. Shamir for the first time presented new, effective metod of the cryptanalysis of the Data Encryption Standard (DES) algorithm — a differential cryptanalysis. This method can be applyed to many types of the algorithms based on substitutions and permutations called the substitution-permutation networks (SPNs).

Together with introducing the differential cryptanalysis appeared a problem of improving resistance of the ciphers against this method of attac. The differential cryptanalysis is based on existance of the differential characteristics. Designing the differential characteristics is a fundamental of the differential cryptanalysis. In this paper we present a kind of extention of the substitution-permutation networks called *a position permutation.* Applying the position permutations in SPN disables creating the differential characteristics like presented by Biham and Shamir. It is necessary to emphasize that applying the position permutations in the SPN does not change the type of algorithm; it is still the substitution-permutation network.

Differential cryptanalysis of the SPNs with the position permutations can be made with use of so called *variant characteristics.* In this paper we show that these characteristics are inefficient and the differential cryptanalysis of the networks with the position permutations is much more difficult than the cryptanalysisof the network without these permutations.

Key words: Subsitution-permutation network, product cipher, differential cryptanalysis.

1. INTRODUCTION

Authors of the papers which concern improving resistance of the iterated algorithms against the differential cryptanalysis usually concentrate on the modifica-

tions of the substitution boxes (S-boxes). This kind of solutions is widely discussed a.o. in [1], [6], [9], [10], [14]. In [9] authors present an extension of the substitution-permutation networks created by introducing the linear transformations between rounds. According to the authors of that paper this transformations increase the resistance of the algorithms against a linear cryptanalysis [13] and against the differential cryptanalysis, but only for algorithms with the S-boxes satisfying a diffusion order [9] equall to zero. For algorithms with the S-boxes satisfying higher diffusion orders (like DES — the diffusion order equall to 1, see [9]) authors claim that the linear transformations do not improve the resistance against this line of the attack.

In this paper we present a modification of the substitution-permutation network. This modification strengthens the SPN in respect to the differential cryptanalysis attacks. We extend the permutations used in the network by applying the transformations called *the position permutations*. The differential cryptanalysis is made with use of the differential characteristics. For the substitution-permutation network in which the position permutations were appled creating the differential characteristics as described in [5] is impossible.

It is important that introducing the position permutations in the SPN does not change the type of the algorithm; we still have the number of substitution rounds connected by the permutations of the special kind.

2. BACKGROUD

In this paper we discuss an N-bit SPN consisting of R rounds of S-boxes connected by permutations P (Fig. 1) .The S-boxes are m x m bijective mappings and P permutation belongs to a special set of permutations for which no two outputs of an S-box are connected to one S-box in the next round. The round (or iteration) keys are in genral independed on each other (or generated by a key scheduling algorithm) and are used to select which mapping from the set of possible mappings is to be used for a particular S-box. The keying can be done by xoring the iteration keys with the inputs to the rows of the S-boxes in each round and by xoring result of the last round with the last iteration key (note that the result of the last round of substitution is not permuted by P). In our modified permutation block we moreover use keying the permutation.

Applying the differential cryptanalysis to the described substitution-permutation network allows us to determine the iteration key xored to the output of the last round of S-boxes (the last iteration key) by using the knowledge of the two ciphertext values (together with their difference) and the input xor of the last round (Fig. 2). We know the xor value of the ciphertext pair. The same value has the output xor of the last round of substitutions. We also have — determined with the characteristic's probability

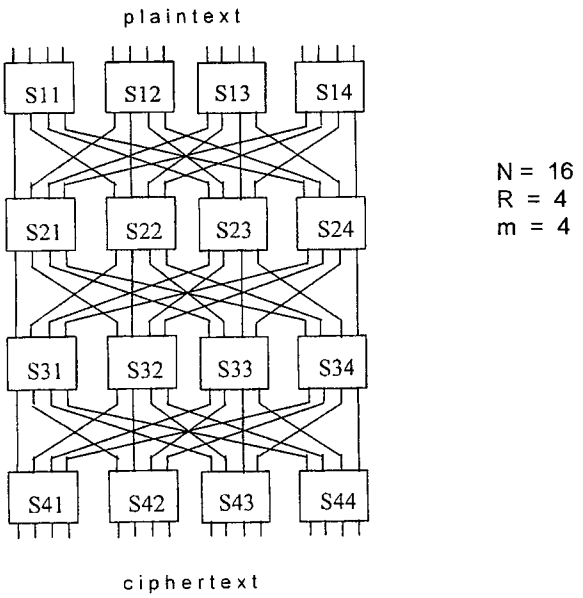


Fig. 1. A substitution-permutation network

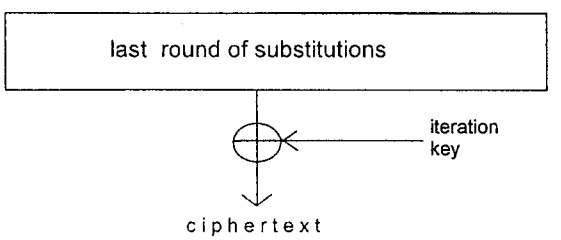


Fig. 2. The last round of the substitution-permutation network

— the input xor of the last round. The pair of input and output xors of the last round of substitutions gives us a set of possible input pairs. From this pairs we calculate the S-boxes output pairs. These pairs xored to the ciphertext pair give us a set of possible last iteration keys. If a rigth pair (a plaintext pair, for which the xor values at the inputs of the rounds assumed in characteristic occured) was used to generate given ciphertext values we have in the set of possible keys the correct one.

This way we can find the last iteration key. Then by "working backwards" and deciphering the last rounds of the SPN we can determine the remaining iteration keys. Success of the differential attack depends a.o. on existance of highly probable differential characteristic for the first R-1 rounds.

3. POSITION PERMUTATIONS

We will use a following defition of the differential characteristic:

Definition 1 (informal)/ *An r-round differential characteristic* is a sequence of r + 1 N-bit numbers, which are assigned to the inputs of the first r rounds and the output of the r-th round of the substitution-permutation network and N is the block size of the network .The numbers in the differential characteristic represent the xor sume of the corresponding values of the two executions of the algorithm [5].

Definition 2. *An S-box position* is a subset of these bits of particular word of the SPN, which come from — for S-boxes output word — or go to — for S-boxes input word — one S-box.

Therefore in particular word bits 1..m (counting from the left) create first S-box position, bits m+1,..., 2m — second one, etc.. We will use the natural numbers (1, 2,...) to denote the S-box positions.

Next we describe a position permutation. Position permutation is a transformation which moves every position of the input word onto new location at the output (in particular onto the same location). Therefore, it is a transformation which changes the arrangement of the positions in the word (in particular to the same arrangement). Its example is presented on Fig. 1. We can think of the position permutation as a function, which assigns a new position number to each S-box position. We can say that this transformation mixes whole S-box positions of the word.

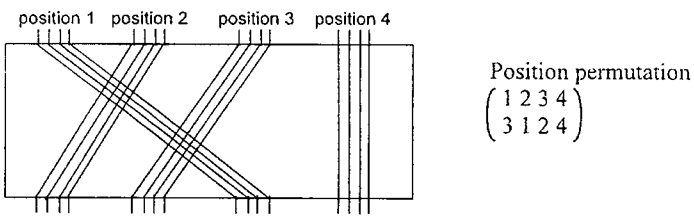


Fig. 3. Example of the position pemtation

Let Π represent the set of the bit permutations for which no two outputs of an S-box are connected to one S-box in the next round (this set is not empty if the number of bits of the S-box output is not greater than the number of S-boxes in one round). The permutation P used in substitution-permutation networks belongs to the set Π.

Let P_p represent the position permutation of the S-box positions of one word. We will apply the permutation P_p to the output of the row of S-boxes in each round of the SPN. Consider one round. We can concatenate the permutations P_p and P and consider them together. Let P' represent a superposition of the position permutation P_p and the bit permutation P. It will be easy to prove the following fact:

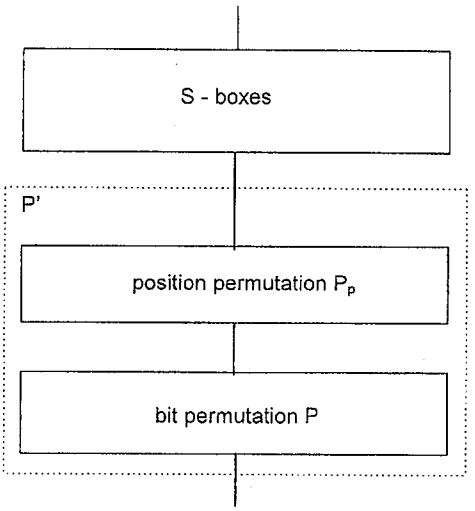


Fig. 4. One round of the SPN with position permutation

Fact 1. The superposition of the position permutation and the bit permutation, $P' = P \circ P_p$ is a permutation from the set $\Pi(\Pi$ — defined as above).

Proof. Proof is obvious; in the SPN without P_p, for each S=box of the i+1 st round we had each bit comming from different S-box of the i-th iteration. Introducing the position permutation changes the numbers of the S-boxes of i-th iteration from which come particular bits or the numbers (indexes) of the output bits, but we still have each bit coming from different S-box. Therefore no two outputs of one S-box in the i-th round are connected to one S-box in the i+1st round.

□

From Fact 1 follows an important corallary: applying the position permutations to the output of S-boxes does not change the type of the algorithm; we still have the substitution-permutation network.

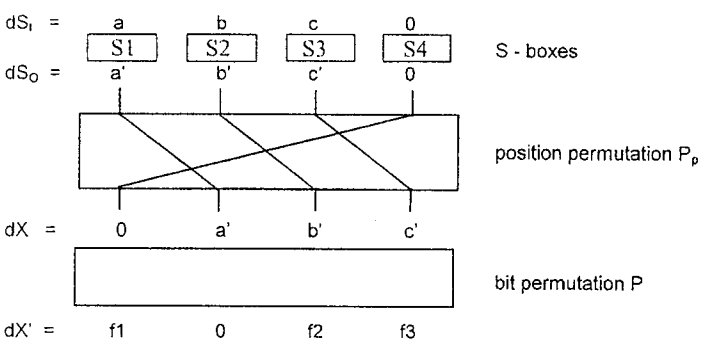


Fig. 5. The example of one round of the SPN

We will introduce a dependence of used in particular iteration position permutation on the iteration key. Used in the round key will be determining which permutation from the set of n! possible permutations is to be used (n represents the number of the S-boxes in one round). We can implement this solution on many ways. For example, if its exist $g \in \mathbf{N}$, such that $n! = 2^g$ and g|N, we can do it following way; we group the iteration key bits in g sets, we calculate xors of the bits in each set and concatenate the results in a g-bit word, value of which we use to determine which permutation from the set of 2^g possible permutations is to be used in given round. In such situation each permutation is determined by the same number of keys. If such g does not exist it may be necessary to use more complex solution. In further part of this paper we assume that each permutation is determined by the same number of keys (in practice it may be aproximatively the same number of keys).

Consider one round of the substitution-permutation network with the position permutations. The example xor values of the words in such round are presented on Fig. 5.

In the round:

dS_I represents the xor of the input values of the S-boxes — shortly: the S-boxes intput xor; in the considered example dS_I = a b c 0 (one sign represents one position),

dS_0 represents the X-boxes output xor; dS_0 = a′ b′ c′ 0,

dX represents the position permutation P_p output xor; dX = 0 a′ b′ c′

dX' xrepresents the permutation P output xor (and the whole round output xor); dX' = f1 0 f2 f3.

Note that for given dX', dX is determined unambiguous, because

$$dX = P^{-1}(dX'), \text{ where } P^{-1} \text{ is an inverse of the permution } P. \tag{1}$$

Assuming particular value of dX' is equivalent assuming corresponding value of dX, namely value resulting from (1).

Let $\Psi(A)$ represent the set of non-zero positions of the word A (the positions with at least one non-zero bit). In the example we have:

$\Psi(dS_1) = \{1, 2, 3\}$,

$\Psi(dS_0) = \{1, 2, 3\}$,

$\Psi(dX) = \{2, 3, 4\}$,

$\Psi(dX') = [1, 3, 4\}$.

We will name the non-zero positions *the involved positions.*

Assume, that we want to create a differential charteristic, in which at the input and at the output of the particular round we have particular xor values. Assume, that as in the considered example we want to have dS_I = a b c 0 and dX' = f1 0 f2 f3. the last condition is equivalent to the following: $dX = PI-1(dX') = 0\,a'\,b'\,c'$. These values at the input and at the output of the round are possible in the following cases:

— if for S1: a->a′ ("−>" means "may cause", see [5]) & for S2: b->b′ & for S3: c->c′ and the iteration key determining permutation $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$ has been used,

— if for S1: a->a' & S2: b->c' & for S3: c->b' and the iteration key determining permutation $\begin{pmatrix}1&2&3&4\\2&4&3&1\end{pmatrix}$ has been used,

— if for S1: a->b' & for S2: b->a' & for S3: c->c' and the iteration key determining permutation $\begin{pmatrix}1&2&3&4\\3&2&4&1\end{pmatrix}$ has been used,

— if for S1: a->b' & for S2: b->c' & S3: c->a' and the iteration key determining permutation $\begin{pmatrix}1&2&3&4\\3&4&2&1\end{pmatrix}$ has been used,

— if for S1: a->c' & for S2: b->a' & for S3: c->b' and the iteration key determining permutation $\begin{pmatrix}1&2&3&4\\4&2&3&1\end{pmatrix}$ has been used,

— if for S1: a->c' & for S2: b->b' & for S3: c->a' and the iteration key detrmining permutation $\begin{pmatrix}1&2&3&4\\4&3&2&1\end{pmatrix}$ has been used.

The above-mentioned cases have in general different probabilities. For each one of them we have $\Psi(dS_0) = \{1, 2, 3\}$ and $\Psi(dX) = \{2, 3, 4\}$. The set of the which we have the particular sets $\Psi(dS_0)$ and $\Psi(dX)$ we will name *a set of enabling keys* K_E. For given choice of the values of dS_1 and dX (or dX') we have the particular set K_e. these values determine the elements of the set K_E. In our example K_e is the sum of the six sets of the keys determining the permutations (1 2 3 4), (1 2 4), (1 3 4), (14) and (14) ∘ (23).

Theorem 1. For given xor values at the input and at the output of the round, the iteration keys which belong to determined by these values set of enabling keys K_E make a following fraction of all possible iteration keys:

$$\frac{k!\,(n-k)!}{n!},$$

where k is the number of involved in the round S-boxes (the S-boxes with non-zero input xors) and n is the total number of the S-boxes in one round. (We remind that we assumed that each position permutation is determined by the same number of keys.)

Proof. The xor values at the input and output of the round determine the sets of involved positions at the input and at the output of the position permutation: $\Psi(dS_0)$ and $\Psi(dX)$. The iteration key belongs to the set of enabling keys K_e if determined by it position permutation, for the set $\Psi(dS_0)$ of involved positions in dS_0, enables obtaining the set of the involved positions at the output of the position permutation equal to $\Psi(dX)$. Because we assumed, that each permutation is determined by the same number of keys the fraction of keys for which the above-mentioned condition is satisfied is equal to the fraction of possible position permutations for which it is satisfied.

The number of all possible permutations is equal to $n!$. We have $k!$ possible arrangements of the elements of the set $\Psi(dS_0^9)$. For each of these arrangements we have $(n-k)!$ arrangements of not invl;oved positions. So the number of the permutations for which we have particular values $\Psi(dS_0)$ and $\Psi(dX)$ is equal to $k!\,(n-k)!$. (Note that determining this fraction is possible even if we know only one of the values dS_0 and dX, because $|\Psi(dS_0)| = |\Psi(dX)|$.)

□

If, for assumed xor values at the input and at the output of the round, the used iteration key K_u does not belong to the set of enabling keys K_e. It is impossible to obtain these values at the input and the output of the round. Having K_u in the set is a necessary condition of obtaining given input-output xor combination.

Fact 2. The iteration key K_u belongs to the determined by round input and output xor values set of enabling keys K_E if, and only if, for the detrmined by it position prmutation a following condition is satisfied:

$$P_p(\Psi(dS_0)) = \Psi(dX). \tag{2}$$

Proof. The set K_E is determined by the combination dS_1, dX'. For given values dS_1, dX' we have the particular sets $\Psi(dS_0)$ and $\Psi(dX)$. The iteration key K_u belongs to the set K_E if and only if for the set of involved positions at the input of P_p equal to $\Psi(dS_0)$ the set of involved positions at the output of the P_p is equal to $\Psi(dX)$. We have this situation if, and only if the set $\Psi(dX)$ is an image of the set $\Psi(dS_0)$ in the transformation P_p:

$$P_p(\Psi(dS_0)) = \Psi(dX).$$

□

Note, that now while creating the differential characteristic we must take care not only of the fact whether particular xor combinations at the input and at the output of the S-boxes are possible to obtain:

Corollary 1. It is possible to obtain paricular xor values combination at the input and the output of the round if relevant pairs of the input and output xors of the S-boxes are possible and used iteration key K_u belongs to the set of enabling keys K_E detrmined by given combination dS_1 and $dX)\,dX')$.

The differential charactyeristics are constucted to make the cryptanalysis of the algorithm, it means — to find the iteration keys. But to construct the characteristic we need to know the position permutations used in the rounds, so we need to know the used keys. From this follows an important theorem.

Theorem 2. For any iteration in which the position permutation has been applied to the outputs of the S-boxes, for given S-boxes output xor (dS_0) it is impossible to determine unambiguous the round output xor if the used iteration key is unknown. For given S-boxes output xor the number of possible round output xor values (for all possible keys) is equall to N_{poss} where

$$N_{\text{poss}} = \frac{n!}{(n-k)!}, \tag{3}$$

where n is the number of S-boxes in one round and k is the number of the S-boxes involved in considered round ($k = |\Psi(dS_1)|$).

Proof. If the used iteration key is unknown we do not know used in the round position permutation, so it is impossible to determine unambiguous the xor value at the output of the block P_p. For given xor value dS_0 at the output of the S-boxes we can determine the set $\Psi(dS_0)$. If the number of the involved positions in dS_0 is equal to k ($|\Psi(dS_0)| = k$), the number of the involved positions at the output of the position permutation is also equal to k. The number of possible sets of involved positions in dX is equal to the number of k-element subsets of the n-element set:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

For each of these subsets $k!$ arrangementys of its elements are possible. Therefore the number of possible xor values at the output of P_p (and at the outpuit of the round) is equal to

$$\binom{n}{k} \cdot k! = \frac{n!}{(n-k)!}.$$

□

Theoren 3. For given xor value dS_1 at the input of the round the probability of obtaining a particular xor value A at the output of the position permutation P_p (or the xor value $A' = P(A)$ at the output of the round) is

$$Pr(dX = A) = \begin{cases} 0, & \text{if } K_u \notin K_E \text{ is the set of the enabling keys} \\ & \text{corresponding to given choice of } dS_1 \text{ and } dX, \\ Pr(S(dS_1) = P_p^{-1}(A), & \text{if } K_u \in K_E (S() \text{ represents S-boxes} \\ & \text{substitution}). \end{cases}$$

Proof. For given xor values dS_1 and dX' at the input and at the output of the round we have the particular sets $\Psi(dS_0)$ and $\Psi(dX)$. If used iteration key K_u does not belong to the set K_E corresponding to the values dS_1 and dX' from the Fact 2 we have:

$$P_p(\Psi(dS_0)) \neq \Psi(dX).$$

From this follows, that if the key K_u does not belong to the set K_E corresponding to the values dS_1 and A, for the xor dS_1 at the input of the round it is impossible to obtain the xor A at the output of the P_p.

If $K_u \in K_e$ equality (2) is satisified. For the position permutation output xor equal

to A, at the output of the S-boxes we have $dS_0 = P_p^{-1}(A)$ (P_p^{-1} is determined by the used key) and dS_0 is such that $\Psi(dS_0) = \Psi(dS_1)$. The probability of obtaining at the output of the block P_p the xor value equal to A is equal to the probability of obtaining at the output of the S-boxes the xor value $P_p^{-1}(A)$. □

Corollary 2. For given iteration key some combinations of dS_1 and dX are possible to occure, namely those, for which

$$Pr(S(dS_1) = P_p^{-1}(A)) > 0,$$

but it is impossible to determine them without the knowledge of the key.
From Theorem 2 and Corollary 2 follows the final conclusion.

Corollary 3. It is impossible to construct the differential characteristic as defined in Definition 1 and determine the probability of its occurance for the substitution-permutation network in which the position permutations have been used.

4. INVOLVED POSITIONS AT OUTPUTS OF ROUNDS WITH POSITION PERMUTATIONS

In this section we examine how, for the established xor value at the output of the S-boxes, the set of involved positions at the output of the round depends on the used position permutation. For different position permutations the distribution of the involved positions at the input of the bit permutation P will be different (the values of the involved positions will be determined, because we assumed, that the S-boxes output xor is established). Therefore the sets of involved positions at the output of the round will be different too.

We remind that used in the SPN permutation P belongs to the special set of permutations Π described in previous section. A kind of used permutation P and a relation between the number of the S-boxes in one round and the number of S-box outputs have influence on the fact whether, after changing the position permutation to the other, the set of involved positions (in the output xor) changes.

For the SPN with the number of the S-boxes in one round equal to the number of S-box outputs exist the permutations P for which independently on used position permutation P_p the set of involved at the output positions is fixed. An example of such permutation is the one for which i-th bit of the j-th position of the input goes to i-th position of the output as the j-th bit of this position. The permutation of this kind is presented on Fig. 1. From the description of this permutation follows that the numbers (indexes) of the involved positions at its output depend on the numbers (indexes) of the non-zero bits in involved positions at the input. Because the change of the used position permutation does not influence the numbers of non-zero bits in the involved positions at the input of P, it does not influence numbers of the positions involved at the output of the P.

From this follows, that for the described above permutation, for given xor value at the output of the S-boxes the change of P_p does not cause the change of the set of involved positions in output xor. Only the values of this positions will change.

But if for the same kind of network (the number of the S-boxes in one round equal to the number of S-box outputs) we use the bit permutation not satisfying described above condition, the change of the P_p may cause the change of the set of involved positions. For example, if in the permutation presented on Fig. 1. we change bits 1 and 4 at tyhe output of the third S-box we have (a.o) following relations:

— the bit 1 of the position 1 at the output of P is equal to the bit 1 of the position 1 at the input,
— the bit 4 of the position 4 at the output of P is equal to the bit 4 of the position 4 at the input,
— the bit 3 of the position 4 at the output of P is equal to the bit 1 of the position 3 at the input.

Of course, this permutation still belongs to the set Π. If now, for particular position permutation we have non-zero the bit 1 of the position 1 and the bit 4 of the position 4 at the output (Fig. 6a), then for the position permutation with the positions 1 and 3 exchanged — in relation to the previous permutation — at the output we have non-zero the bits 3 and 4 of the position 4 (Fig. 6b). In the first case we have two positions involved at the output, in the second one — only one.

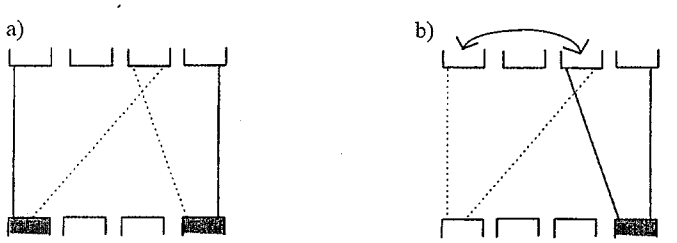


Fig. 6. The example of the bit permutation P

As well for the networks in which the number of the S-boxes in one round is greater than the number of the S-box outputs, for any permuatation P from the set Π it is possible to have a situation, in which we have different sets of involved positions at the output of the round for different position permutations. The examples of this situation are presented on Fig. 7

A fact, whether the kind of used position permutation may influence the set of involved positions at the output of the round depends on the structure of SPN and on used bit permutation P. There are networks, for which changing the position permutation causes the changes of the considered set and there are ones for wich this set is established.

Unlike the networks without position permutations the knowledge of the permutation P does not guarantee, for given S-boxes output xor, the knowledge of the set of involved positions at the output of the round.

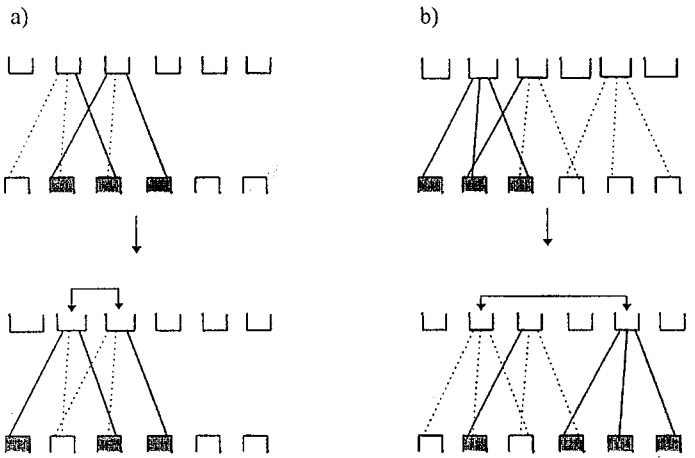


Fig. 7. The SPN with the number of the S-boxes in one round greater than the number of the S-box outputs

5. VARIANT CHARACTERISTICS

While creating the differential characteristic for the substitution permutation network without the position permutations for each round we assumed the pair of the xors: the round input values xor and the round output values xor. This assumption was equivalent to the assumption of the pair of the xors at the input and at the output of the S-boxes. The round input and output xor values were chosen to satisfy the following conditions:

— the probability of obtaining correpsonding xor values combination at the input adn at the output of the S-boxes should be as high as possible,

— the round input and output xor values should be "convenient", it means they should enable designing effective characteristics (for example: using fixpoints [3] [11], twinxors [12], zero xor values at the output of S-boxes [4] [12] a.o.).

In the following considerations we omit the bit permutation P, because for given xor value at the output of the round the xor value at the input of the P permutation is determined unambiguous. Therefore, for the particular round we consider the pair of the xors at its input and at the input of the P permutation.

While designing the differential characteristic for the SPN with the position permutations for the particular round we cannot assume one xor value at the input of the bit permutation P, because we do not know used in the round position permutation P_p. We can only assume a k-element set of the position xor values, where k is the number of involved positions in the round input xor. However we do

not know at which positions which values are to be placed. We assume that assumed position xor values are possible to obtain at the outputs of the corresponding involved S-boxes. In the other case the designed characteristic would be useless. While making the differential cryptanalysis we will have to consider all possible distributions of the assumed values, but only some of them will be possible to obtain.

It is obvious, that we will not obtain these distributions for which we have $P_p(\Psi(dS_0)) \diamond \Psi(dX)$ (we use denotations as presented on the Fig. 5). We will obtain the remaining distributions if the corresponding combinations of the xors at the inputs and outputs of corresponding S-boxes are possible. If, for example, we consider the case presented on the Fig. 2 — to which position permutation (1 2 3 4) has been applyed — and we assume that following conditions are satisfyed: for S1 a->a', a->b', a->c', for S2 b->a', b->b', b->c', for S3 c->a', c->b', c->c' then it is possible to obtain at the output of the position permutation following distributions: 0a'b'c, 0a'c'b', 0b'a'c', 0b'c'a', 0c'a'b', 0c'b'a'. In general, the probabilities of this distributions are different. While examining given distribution we cannot determine its probability; we even cannot determine wether it is possible to obtain it (we do not know used position permutation). Therefore, we must assume that each considered distribution is possible. There are $C_n^k P_k$ distributions to be examined, where n is the number of the S-boxes in one round and k is the number of the involved S-boxes in considered round and

$$C_n^k = \frac{n!}{k!(n-k)!} \quad \text{and} \quad P_k = k!.$$

Among these distributions, P_k of them are possible to obtain (if corresponding xor values at the input and the output of the S-boxes are possible).

One-round characteristic

We consider a theoretical case: a one-round differential characteristic. To fix our attention we assume that each one of the assumed xor values of the positions at the output of the position permutation is possible to obtain at the output of each one of the involved S-boxes (for given round input xor). The consequences following from this assumption will be discussed later.

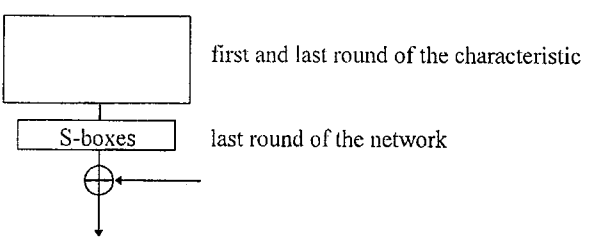


Fig. 8. One round differential characteristic

We must consider $C_n^k P_p$ cases of the output of the position permutation (k, n — as above). For each considered case we assume that it is possible to obtain. To have real chances for having the case, which is truly possible generate A times the correct key we apply $A^* 1/p_v$ pairs with given input xor to the input of the network, where p_v is the probability of the case. However, as we noted earlier while considering given case of the output of the position permutation (or the output of the round) we do not know its probability (we even do not know whether it is the possible case). One of the possible cases has the least probability of all — we represent this probability by $p_{\min}$. If we want to have realizable chances for having each possible case generate A (or more) occurrances of the correct key we must consider each of them as if it had the probability $p_{\min}$. It is not the only solution — we may not assume such probability. However, in that case we must remember that it is possible to have the situation, in which not all possible cases generate at least A occurances of the correct key. If A were rather small it might happen that some possible cases would not generate even one occurrance of the possible key. In the following part of this section we consider each case assuming that its probability is equal to $p_{\min}$.

Totaly we must apply $C_n^k P_k A/p_{\min}$ pairs to the input and these pairs will generate (with high probability) at least $A P_k$ occurrances of the correct key. Therefore the avarange number of the pairs per one occurance of the correct key we can aproximate by κ, where:

$$\kappa = \frac{C_n^k P_k A/p_{\min}}{A P_k} = \frac{1}{p_{\min}} C_n^k.$$

For the network without the position permutation for the round with the probability $p_{\min}$ we would have $C_n^k P_k A$ occurances of the correct key for $C_n^k P_k a/p_{\min}$ pairs applyed to the input of the round. In this situation we would have:

$$\kappa = \frac{1}{p_{\min}}.$$

If the probability were greater than $p_{\min}$ (remember that it is the probability of the least probable case) the ratio κ would be even smaller.

At the beginning of this section we established that each assumed xor value of the positions at the input of the position permutation is possible to obtain at the output of each one of the involved S-boxes. What about not assuming this situation? In that case some of the cases for wich $P_p(\Psi(dS_0)) = \Psi(dX)$ have zero probabilities — these for which corresponding xor values combinations at the inputs and outputs of the particular S-boxes are impossible. Therefore the number of the cases, which generate the correct key is smaller anf from this follows that the number of the occurrances of the correct key (after considering all cases) is smaller. The value of the ratio κ increases.

Many-rounds characteristics — variant characteristics

Because for given xor at the input of the round we assumed the set of the values of involved positions at the output of the position permutation, but we do not determine the distribution of these values at particular positions, we have to consider all possible distributions (the cases of the position permutation outputs). Considering particular cases in successive rounds — starting from the first one — we create a tree. A node of the tree represents the case of the xor value of the input of the particular iteration. For this case of input we must consider as many cases of the output as the number of cases of position permutation, it means $C_n^k P_k$. Some of these cases are impossible ones — we will not obtain them for any of the input pairs (Fig. 9). They are marked on the Fig. 9 with dashed lines. The remaining cases are

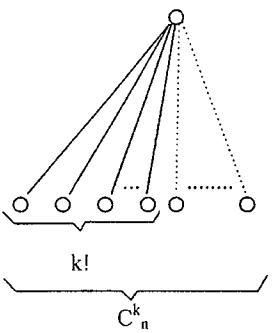


Fig. 9. The possible and impossible variants

these for which $P_p(\Psi(dS_0)) = \Psi(dX)$. If the corresponding xor values combinations at the inputs and outputs of the S=boxes are possible it is possible to obtain these variants. On the Fig. 9 they are marked with solid lines. For each node among $C_n^k P_k$ variants there are $k!$ possible ones. The tree created described way we name *a variant differential characteristic.*

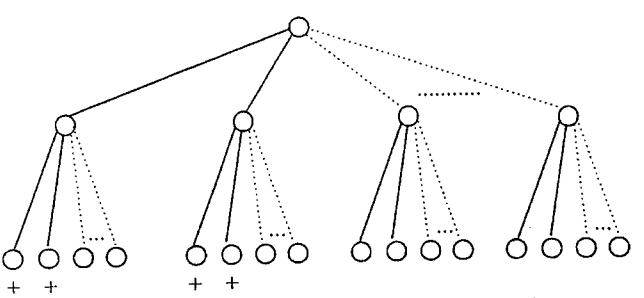


Fig. 10. The possible variants at the bottom of the characteristic

While designing the variant characteristic in described way at the bottom of it we obtain the nodes, to which we can move from the root of the tree passing only solid lines (see Fig. 10). These are the nodes possible to obtain, which represent the xors we can obtain at the output of the last round of the characteristic. They may be used to generate the correct key. Of course, while making the cryptanalysis of the algorithm we do not know the key and we do not know the possible variants. Therefore we must generate the keys for each node at the bottom of the tree.

For any bit permutation P from the set Π various distributions of the involved positions at the input of P may result in different sets of involved positions at its output (see section 4) — the number of involved positions may differ, too (Fig. 11).

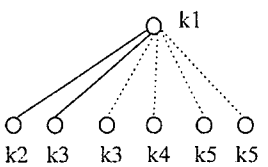


Fig. 11. The number of involved positions in nodes

The number of the nodes at the bottom of the tree depends on the actual xor values of the variants of the outputs of particular rounds. We create the variant characteristic successively considering rounds and all possible variants of their outputs. After we design the characteristic we can determine how many possible variants of the output of the last round we have.

Note, that it is possible to create the tree, but it is impossible to determine the kind of line connecting the nodes. For each node we can determine how many child-nodes it has, how many child-nodes which are possible to obtain it has, but we cannot determine which child-nodes are these possible ones and from this follows that we cannot determine possible nodes at the bottom of the tree. Also we cannot

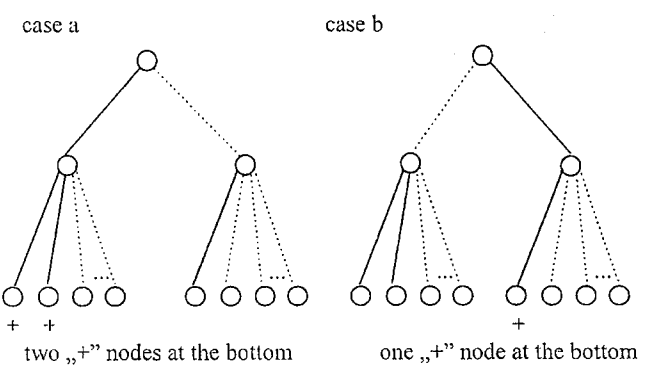


Fig. 12. Two cases of the same variant characteristic

determine the number of possible to obtain child-nodes at the bottom of the tree because the knowledge of the number of solid lines starting from particular node is not enough here; we should also know which lines are solid. The examples presented on the Fig. 12 ilustrate described situation.

From the presented above considerations follow three important corrollaries:

- After we generate the keys for all possible nodes at the bottom of the characteristic we have a number of occurances of the correct key. If for the network without the position permutations we generated the keys for the same number of characteristics as the number of the nodes at the bottom of the tree and the probabilities of these characteristics had the same values as the probabilities of the possible bottom nodes and the fraction of the characteristics with given probability were the same as the fraction of nodes with this probability amog possible ones then for the network without the position permutation we would obtain the correct key each time and the total number of its occurrances would be greater.
- Assuming the minimal probability for each of the variants of the output of the last round we must remeber that this probability may be relatively small. On the one hand it gives the greater number of occurances of the correct key (because for the possible variants with greater probabilities it occures more times), but on the other hand it increases the number of the pairs applyed to the input while considering each variant.

  We can choose another solution, it means we can assume greater probability. Consequences following from such choice were described earlier.
- If we did not want to check all nodes and decided to check only a part of them we could select these ones which are impossible to obtain. In general we have to check all nodes because it is possible to have a situation in which after checking the part of nodes — which are impossible — one of the keys occures significant more often than the others being a false clue.

SUMMARY

In this paper we presented a kind of extension of the substitution-permutation network — the position permutation. We showed, that applying this transformation to the output of the row of S-boxes does not change the type of the algorithm.

We showed, that designing the differential characteristics like presented in papers concerning the differential cryptanalysis of various block ciphers ([2], [3], [4], [5], [11] and others) is impossible. This follows from the fact, that used in particular round position permutation depends on the used iteration key, which is unknown. For given key, some combinations of the xor values at the input and at the output of the round are impossible, some are possible, but we cannot identify them without the knowledge of the key.

We also showed that in general for the substitution-permutation network the knowledge of the S-boxes output xor not only does not guarantee the knowledge of the xor at the output of the round, but also does not guarantee the knowledge of the

set of involved positions, as it was in the case of SPN without the position permutations.

Finally we presented a kind of a generalization of the differential characteristics for the SPN with the position permutations — the variant characteristics. We showed that these characteristics are inefficient, it means the cryptanalysis made with use of them is much more difficult than the one made with use of the ordinary differential characteristics. Because for the networks with the position permutations it is impossible to use the ordinary differential characteristics, these networks are more resistant against the differential cryptanalysis than the networks without the position permutations.

REFERENCES

1. C.M. Adams: *On immunity against Biham and Shamir's differential cryptanalysis.* Information Processing Letters, 1992, vol. 41, No. 2, pp. 77–80
2. I. Ben-Aroya, E. Biham: *Differential Cryptanalysis of Lucifer.* Journal of Cryptology, 1996, Vol. 9, No. 1, pp. 21–34
3. E. Biham, A. Shamir: *Cryptanalysis of Snefru, Khafre, REDOC-II, LOKI and Lucifer.* Advances in Cryptology — CRYPTO'91, Springer-Verlag, 1992, pp. 156–171
4. E. Biham, A. Shamir: *Differential cryptanalysis of the Full 16-Round DES.* Advances in Cryptology — CRYPTO'92 Springer-Verlag, 1993, pp. 487–496
5. E. Biham, A. Shamir: *Differential Cryptanalysis of DES-like Cryptosystems.* Journal of Cryptology 1991, Vol. 4, No. 1, pp. 3–72
6. L. Brown, M. Kwan, J. Pieprzyk, J. Sebbery: *Improving Resistance to Differential Cryptoanalysis and the Redesign of LOKI* Advances in Cryptology — ASIACRYPT'91, Springer-Verlag, 1992, pp. 36–50
7. L. Brown, J. Seberry: *On the design of permutation P in DES type cryptosystems.* Advances in Cryptology — EUROCRYPT'89. Springer-Verlag, 1990, pp. 696–705
8. C. Charnes, J. Pieprzyk: *Linear Nonequivalence versus Nonlinearity.* Advances in Cryptology — AUSCRYPT'92. Springer-Verlag, 1993, pp. 156–164
9. H.M. Heys, S.E. Tavares: *Substitution-Permutation Networks Resistant to Differential and Linear Cryptanalysis.* Journal of Crypology, 1996, Vol. 9, No. 1, p. 1–20
10. K. Kim: *Construction of DES-like S-boxes Based on Boolean Functions Satisfying the SAC.* Advances in Cryptology — ASIACRYPT'91. Springer-Verlag, 1992, pp. 59–72
11. L.R. Knudsen: *Cryptanalysis of LOKI 91.* Advances in Cryptology — AISCRYPT'92. Springer-Verlag, 1993, pp. 196–208
12. L.R. Knudsen: *Iterative Characteristics of DES and s^2 DES.* Advances in Cryptology CRYPTO'92. Springer-Verlag, 1993, pp. 497–511
13. M. Matsui: *Linear cryptanslysis method for DES cipher.* Advances in Cryptology — EUROCRYPT'93. Springer-Verlag, 1994, pp. 386–397
14. K. Nyberg, L.R. Knudsen: *Provable Security Against Differential Cryptanalysis.* Advances in Cryptology — CRYPTO'92. Springer-Verlag, 1993, pp. 566–574
15. M. Sivabalan, S.E. Tavares, L.E. Peppard: *On the3 Design of SP Networks from an Information Theoretic Point of View.* Advances in Cryptology — CRYPTO'92. Springer-Verlag, 1993, pp. 260–279

A. SADOWSKI

ZALEŻNE OD KLUCZA PERMUTACJE BITÓW W SIECIACH PODSTAWIENIOWO-PERMUTACYJNYCH

Streszczenie

Na początku lat 90-tych E. Biham i A. Shamir po raz pierwszy przedstawili nową, efektywną metodę krytpoanalizy algorytmu DES (ang. Data Encryption Standard) — kryptoanalizę różnicową. Metoda ta może być stosowana do wielu typów algorytmów opartych na podstawieniach i permutacjach zwanych sieciami podstawieniowo-permutacyjnymi. Wraz z wprowadzeniem kryptoanalizy różnicowej pojawił się problem zwiększania odporności szyfrów na tę linię ataku. Kryptoanaliza różnicowa oparta jest na istnieniu charakterystyk różnicowych. Projektowanie charakterystyk różnicowych jest podstawą kryptoanalizy różnicowej. W artykule przedstawiono rodzaj rozszerzenia sieci podstawieniowo-permutacyjnych zwany *permutacją pozycji*. Zastosowanie permutacji pozycji w sieciach podstawieniowo-permutacyjnych uniemozliwia tworzenie charakterystyk różnicowych takiego rodzaju jakie zaprezentowali Biham i Shamir. Należy podkreślić, że zastosowanie permutacji pozycji nie zmienia typu algorytmu; nadal mamy do czynienia z siecią podstawieniowo-permutacyjną. Kryptoanaliza różnicowa sieci podstawieniowo-permutacyjnych może być wykonywana z zastosowaniem tak zwanych *charakterystyk wariantowych*. W artykule pokazano, iż charakterystyki te są nieefektywne i kryptoanaliza różnicowa sieci podstawieniowo-permutacyjnych z permutacjami pozycji jest daleko bardziej złożona niż sieci bez takich permutacji.

Słowa kluczowe: Sieci podstawieniowo-permutacyjne, szyfry blokowe, kryptoanaliza różnicowa.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1997, 43, z. 3, ss. 353–372

Wykorzystanie standardowych kompilatorów FPLD do syntezy sterowników logicznych

MAREK WĘGRZYN, MARIAN ADAMSKI

Instytut Informatyki i Elektroniki, Politechnika Zielonogórska

Otrzymano 1997.02.03

Autoryzowano 1997.04.17

Metody projektowania kontrolerów współbieżnych odgrywają dużą rolę w projektowaniu złożonych układów cyfrowych, a zwłaszcza sterowników logicznych. Funkcjonowanie kontrolera może być opisane za pomocą kolorowanych i bezpiecznych sieci Petriego. Ze względu na efektywność implementacji sieć opisująca kontroler współbieżny może być podzielona (zdekomponowana) na części o charakterze sekwencyjnym, nazywane sieciami automatowymi, a następnie pokolorowana automatowo. SM-kolorowanie rozróżnia w tym przypadku procesy sekwencyjne, które są przewidziane przez projektanta, albo wykryte wtórnie metodami formalnymi. W poprzednich pracach autorów rozpatrywany rodzaj kolorowania sieci Petriego został wprowadzony głównie w celu dekompozycji równoległej. Kolorowanie przedstawione w artykule jest częścią nowej metodologii projektowania.

W pracy przedstawiono język pośredniczący typu *front-end*, zgodny z wprowadzonym uprzednio modelem koncepcyjnym. Jest on tłumaczony automatycznie na wybrane, standardowe lub specjalizowane języki HDL.

W artykule omówiono jedynie niektóre techniki przydatne w projektowaniu kontrolera modelowanego jako współbieżny automat cyfrowy. Szczególny nacisk położono na syntezę z wykorzystaniem programowanych struktur logicznych PLD i języków HDL. W celu porównania załączono trzy szczegółowe przykłady syntezy tego samego kontrolera z wykorzystaniem standardowych kompilatorów PLD: CUPL, PALASM/MachXL oraz VHDL. Przedstawiona metodologia projektowania może być łatwo dostosowana do innych specjalizowanych języków HDL, na przykład: DSL (MINC), ABEL (DATA/IO) i AHDL (ALTERA).

Słowa kluczowe: sterowniki logiczne, kontrolery współbieżne, sieci Petriego.

1. WSTĘP

Programowana struktura logiczna PLD (ang. *Programmable Logic Device*) jest scalonym układem cyfrowym, który może być zaprogramowany przez użytkownika w celu realizacji zespołu funkcji logicznych, albo współpracujących ze sobą sekwen-

cyjnych automatów cyfrowych. W tym drugim przypadku najczęstszą formą opisu jest specjalizowany język HDL (ang. *Hardware Description Language*), podający warunki przejścia z rozpatrywanego stanu do wszystkich możliwych stanów następnych. Znane powszechnie języki kompilatorów PLD nie przewidują bezpośredniej formy opisu automatów współbieżnych, które mogą być zadane interpretowanymi sieciami Petriego [17]. Aby wykorzystać istniejące już możliwości, sieć Petriego można pokryć składowymi automatowymi, rozróżnić je kolorami i traktować sieć jako superpozycję automatów sekwencyjnych. Tego rodzaju kolorowana sieć jest dogodną formą przedstawiania kontrolerów współbieżnych modelujących sterowniki logiczne. W artykule rozszerzono wyniki przedstawione w pracy [21] w kierunku konkretnych implementacji, kładąc nacisk na sposoby kodowania i efektywność realizacji w strukturach PLD. Z braku miejsca nie zamieszczono syntetycznych wyników eksperymentów. Ponadto, wybierając przykładową sieć Petriego starano się pokazać przydatność prezentowanej metody w syntezie reprogramowalnych sterowników logicznych [5, 7, 11, 21, 23] do zastosowań przemysłowych na podstawie normy IEC-1131-3 [7, 23].

W celu specyfikacji formalnej sieci wykorzystano postać regułową (rozdz. 2), ze szczególnym uwzględnieniem rozszerzonego formatu PNSF2 (opis formatu PNSF został przedstawiony w [16]). Automatowa dekompozycja sieci Petriego może być dokonana jednym z kilku sposobów kolorowania sieci (rozdz. 3). Poprzez kolorowanie wprowadza się do sieci kolorowe znaczniki, pozwalające w procesie animacji w dogodny sposób śledzić przebieg nakładających się procesów. Z drugiej strony kolorowanie pozwala odrzucić znaczną część wadliwie skonstruowanych sieci [3, 9]. Kolorowanie sieci jest szczególnie przydatne w kodowaniu stanów lokalnych miejsc sieci (rozdz. 4). Przykłady opisu w wybranych językach HDL demonstrują tylko kilka z możliwych stylów specyfikacji (rozdz. 5). Praca kończy się krótkim podsumowaniem (rozdz. 6).

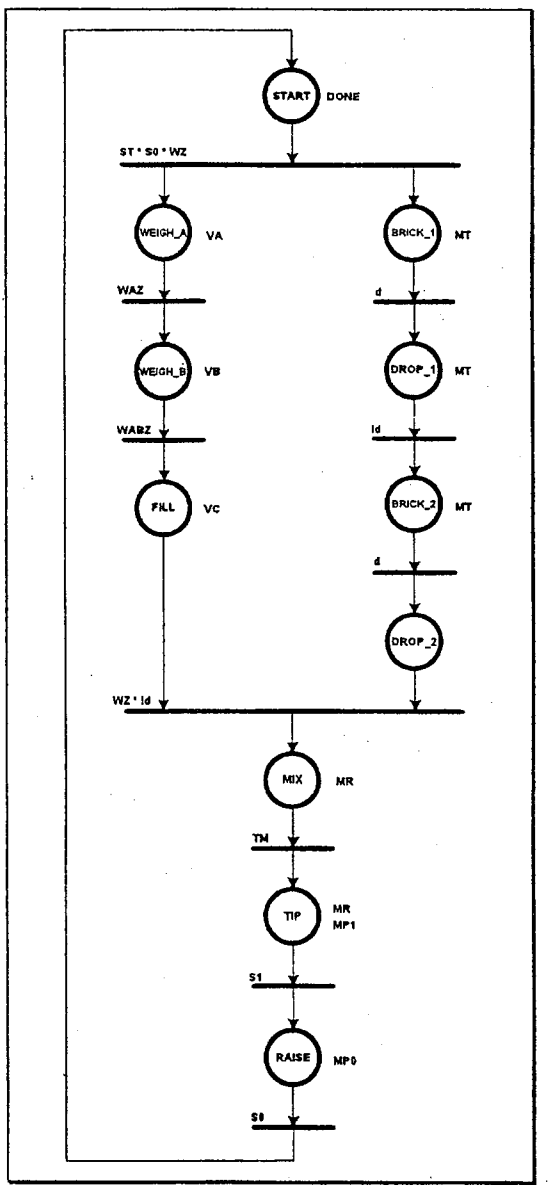
2. SYNTEZA KONTROLERÓW WSPÓŁBIEŻNYCH

2.1. REGULOWA SPECYFIKACJA SIECI PETRIEGO

Opisując sieci Petriego w sposób tekstowy można wykorzystywać język regułowy, oparty na logice sekwentów Gentzena, z elementami logiki temporalnej [4]. Uproszczoną postać regułową ,będącą pierwowzorem języka LOGICIAN można znaleźć między innymi w pracy [1]. Zaproponowana forma regułowa została zaadaptowana między innymi jako język specjalizowanych mikroprocesorowych sterowników logicznych [2, 12] oraz język eksperymentalnego systemu do projektowania układów cyfrowych z programowanymi strukturami logicznymi [3, 5] dla potrzeb metrologii.

Opis regułowy w języku sekwentów Gentzena [1] stał się podstawą uproszczonej, tekstowej reprezentacji interpretowanych sieci Petriego i jest nazywany formatem

PNSF (ang. *Petri Net Specification Format*) [16]. Język opracowany przez Kozłowskiego był wzorowany na formacie BLIF, stosowanym w popularnym, uniwersyteckim systemie CAD — SIS [19]. Należy tu nadmienić znaczne podobieństwo formatu PNSF do reguł produkcji, stosowanych do reprezentacji wiedzy i będących stwier-



Rys. 1. Zmodyfikowana interpretowana sieć Petriego

dzeniami o postaci typu IF ... THEN [4]. Nieco rozszerzonym dialektem języka PNSF jest ConPar [10].

Zalety języków LOGICIAN, PNSF i ConPar połączono w nowym, uzupełnionym formacie PNSF2, zorientowanym na formalną specyfikację sieci hierarchicznych. Do języka wprowadzono elementy formatu BLIF-MV [10]. Inna zasadnicza różnica między nową wersją a jej poprzednikami polega na możliwości przedstawiania w PNSF2 interpretowanych, kolorowanych sieci P/T [6].

Kolorowana i interpretowana P/T sieć jest użytecznym podzbiorem sieci CPN, analizowanych i symulowanych (animowanych graficznie) w systemie Design/CPN [13]. Kolorowana P/T sieć została zaadaptowana do potrzeb syntezy sterowników logicznych typu ASLC (ang. *Application Specific Logic Controller*) [22, 23]. Jest ona też dobrym narzędziem specyfikacji i syntezy układów cyfrowych.

Język PNSF2 daje umożliwia łatwą specyfikację układu cyfrowego z wyjściami Moore'a i Mealy'ego, zarówno typu rejestrowego, jak i kombinacyjnego. Poprzednia wersja formatu PNSF oraz format ConPar nie pozwalają na równoczesną deklarację wyjść kombinacyjnych, jak i rejestrowych w tym samym układzie FPGA lub CPLD. Format PNSF2 rozróżnia wyjścia kombinacyjne (*.comb_outputs*) i wyjścia rejestrowe (*.reg_outputs*). Zachowano ogólną deklarację wyjść (*.outputs*) w celu utrzymania zgodności ze starszą wersją, dzięki czemu nowa wersja akceptuje poprzednio wykonane projekty oraz przykładowe programy (ang. *benchmarks*). Przedstawione dalej układy automatycznej syntezy zorientowane na elementy PLD akceptują starszy format specyfikacji.

Na rys. 1 została podana modyfikacja sieci z prac [7, 23], opisującej sterownik logiczny z wyjściami typu Moore'a i Mealy'ego. Przekształcenie sieci polega na zamianie wszystkich wyjść typu Mealy'ego na wyjścia kombinacyjne typu Moore'a. Zaletą tego rodzaju transformacji jest możliwość kodowania miejsc z wykorzystaniem sygnałów wyjściowych (rozdz. 4).

2.2. JĘZYK OPISU SIECI PNSF2

Każda tranzycja sieci jest przedstawiona jako reguła przejść typu IF ... THEN ... i zapisana jako sekwent logiczny [4, 21].

$$Ti:\ pre\ (Ti)\ *\ warunek\ (Ti)\ |-\ post\ (Ti),$$

Po lewej stronie znaku wynikania |– podane są warunki realizacji (zapalenia) danej tranzycji *Ti* w postaci koniunkcji jej miejsc wejściowych *pre (Ti)*, natomiast po prawej stronie jej znaku wynikania — skutki jej zapalenia w postaci koniunkcji jej miejsc wyjściowych *post (Ti)*. Dodatkowo lewa strona sekwentu zawiera koniunkcje sygnałów wejściowych *warunek (Ti)* zezwalajacych na realizację tranzycji. Wyjścia typu Moore'a przedstawiane są regułą wyjść:

$$Pj\ |-\ Y(Pj);$$

w którym $Y(Pj)$ jest koniunkcją wyjść aktywnych w oznakowanym miejscu *Pj*.

W opisie wykorzystuje się symbole logiczne: + (or), * (and), ! (not).
Zapis:

$$T1: START * ST * SO * WZ \mid\!- WEIGH\_A * BRICK\_1,$$

oznacza, że warunkiem realizacji tranzycji *T1* jest oznakowanie miejsca *START* oraz pojawienie się sygnałów wejściowych *ST, SO, WZ* równych jeden, natomiast skutkiem realizacji tej tranzycji jest oznakowanie miejsc *WEIGH_A* i *BRICK_1*. Natomiast reguła wyjściowa:

$$START \mid\!- DONE$$

opisuje wyjście *DONE* aktywne w stanie lokalnym *START*.

Opis automatu współbieżnego zawiera listy wszystkich sygnałów wejściowych (.inputs), sygnałów wyjściowych (.outputs), miejsc sieci (.places) oraz miejsc w oznakowaniu początkowym (.marking). Pełny opis regułowy sieci składający się z reguł przejść (.transitions) i reguł wyjść (.MooreOutputs) zamieszczono na rys. 2.

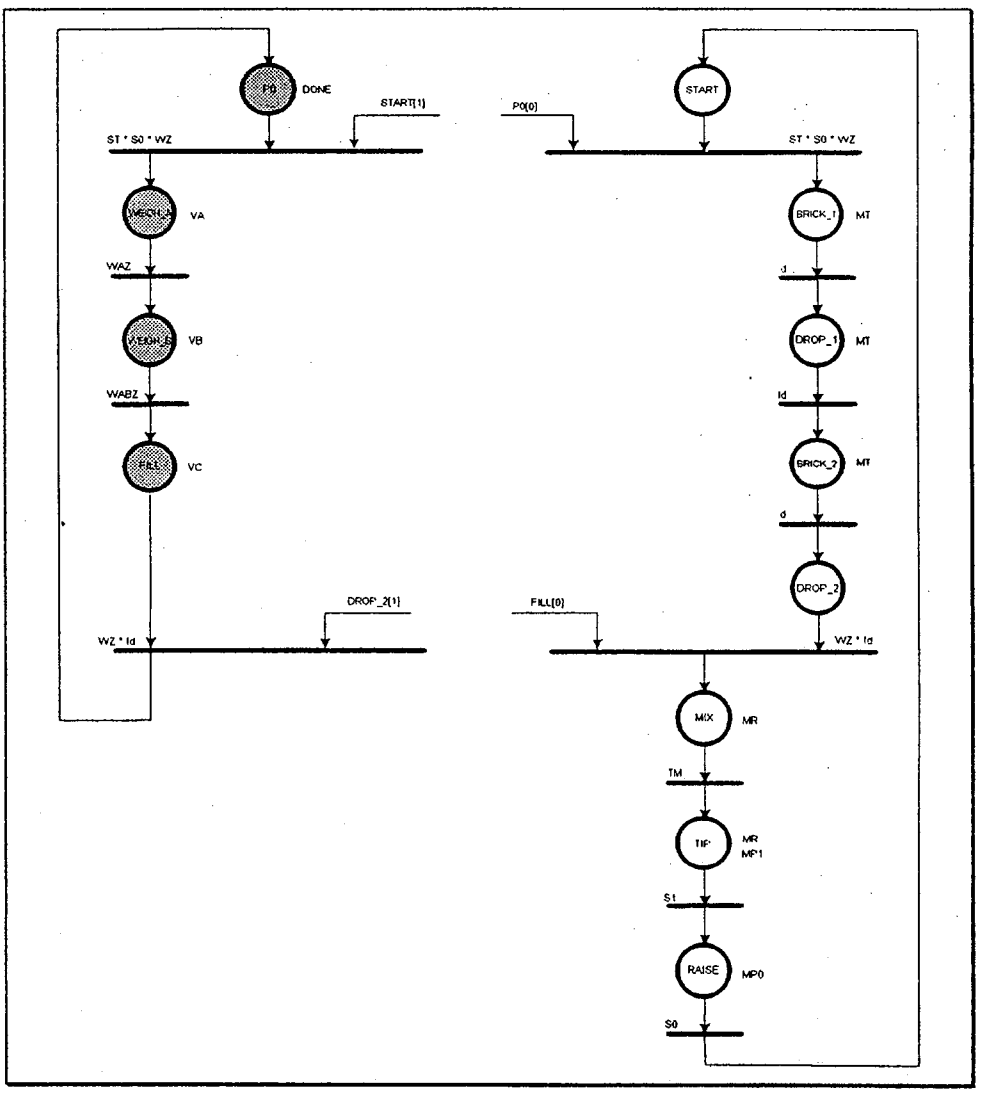
```
.clock CLK
.places START WEIGH_A WEIGH_B FILL BRICK_1 DROP_1
        BRICK_2 DROP_2 MIX TIP RAISE
.inputs ST S1 S0 WZ WAZ WABZ d TM
.outputs DONE VA VB VC MP1 MP0 MT MR
.transitions T1 T2 T3 T4 T5 T6 T7 T8 T9 T10
.net
 T1:  START * ST * S0 * WZ |- WEIGH_A * BRICK_1;
 T2:  WEIGH_A * WAZ |- WEIGH_B;
 T3:  WEIGH_B * WABZ |- FILL;
 T4:  BRICK_1 * d |- DROP_1;
 T5:  DROP_1 * !d |- BRICK_2;
 T6:  BRICK_2 * d |- DROP_2;
 T7:  FILL * DROP_2 * WZ * !d |- MIX;
 T8:  MIX * TM |- TIP;
 T9:  TIP * S1 |- RAISE;
 T10: RAISE * S0 |- START;
.MooreOutputs
 START |- DONE;
 WEIGH_A |- VA;
 WEIGH_B |- VB;
 FILL |- VC;
 BRICK_1 |- MT;
 DROP_1 |- MT;
 BRICK_2 |- MT;
 MIX |- MR;
 TIP |- MR * MP1;
 RAISE |- MP0;
.marking START
.e
```

Rys. 2. Opis sieci Petriego w formacie PNSF2

3. KOLOROWANIE AUTOMATOWE SIECI PETRIEGO

W pracy [21] omówiono zagadnienie dekompozycji sieci Petriego na składowe automatowe metodą topologiczno-algebraiczną (kolorowania). Na rys. 3 i 5 podano dwie wersje zdekomponowanej sieci Petriego z rys. 1. Sposób komunikacji i strukturę podsieci z rys. 3 przedstawiono w formacie PNSF2 (rys. 4).

Rezultaty kolorowania, oprócz dekompozycji, mogą być również wykorzystane do efektywnego kodowania lokalnych stanów wewnętrznych, czyli miejsc sieci

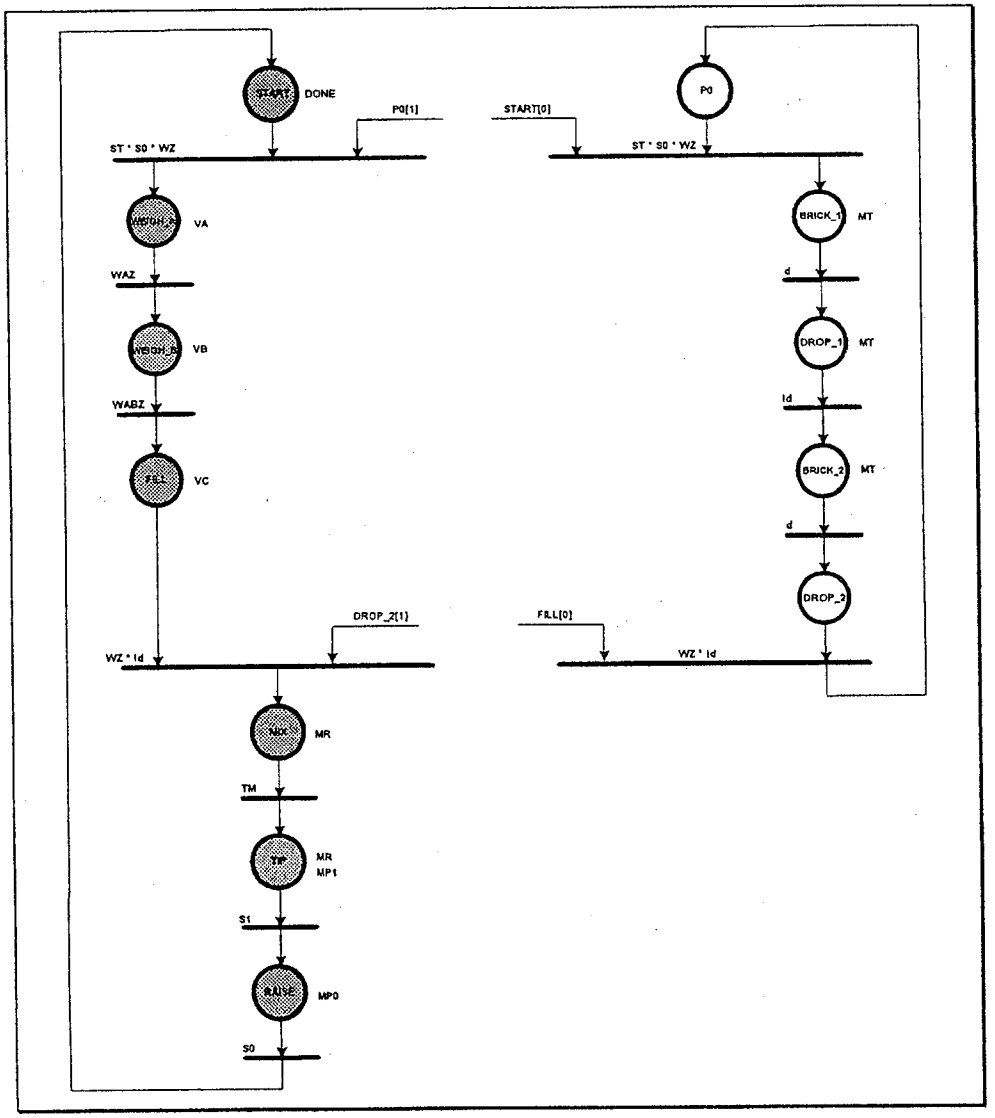


Rys. 3. Procesy w sieci Petriego (pierwsza wersja)

```
.clock CLK
.inputs ST S1 S0 WZ WAZ WABZ d TM
.part Automat0
.places P0 WEIGH_A WEIGH_B FILL
.outputs VA VB VC
.transitions T0_1 T0_2 T0_3 T0_7
.predicates Warunek1 Warunek2
.net
 T0_1:  P0 * Warunek1 |- WEIGH_A;
 T0_2:  WEIGH_A * WAZ |- WEIGH_B;
 T0_3:  WEIGH_B * WABZ |- FILL;
 T0_7:  FILL * Warunek2 |- P0;
.MooreOutputs
 WEIGH_A |- VA;
 WEIGH_B |- VB;
 FILL |- VC;
.marking P0
.part Automat_1
.places START BRICK_1 DROP_1 BRICK_2 DROP_2 MIX TIP RAISE
.outputs DONE MP1 MP0 MT MR
.transitions T1_1 T1_4 T1_5 T1_6 T1_8 T1_9 T1_10
.predicates Warunek3 Warunek4
.net
 T1_1:  START *  Warunek3 |- BRICK_1;
 T1_4:  BRICK_1 * d |- DROP_1;
 T1_5:  DROP_1 * !d |- BRICK_2;
 T1_6:  BRICK_2 * d |- DROP_2;
 T1_8:  DROP_2 * Warunek4 |- MIX;
 T1_9:  MIX * TM |- TIP;
 T1_10: TIP * S1 |- RAISE;
 T1_11: RAISE * S0 |- START;
.MooreOutputs
 START |- DONE;
 BRICK_1 |- MT;
 DROP_1 |- MT;
 BRICK_2 |- MT;
 MIX |- MR;
 TIP |- MR * MP1;
 RAISE |- MP0;
.marking START
.predicateDescriptions
 Warunek1 = START * ST * S0 * WZ;
 Warunek2 = DROP_2 * WZ * !d;
 Warunek3 = P0 * ST * S0 * WZ;
 Warunek4 = FILL * WZ * !d;
.e
```

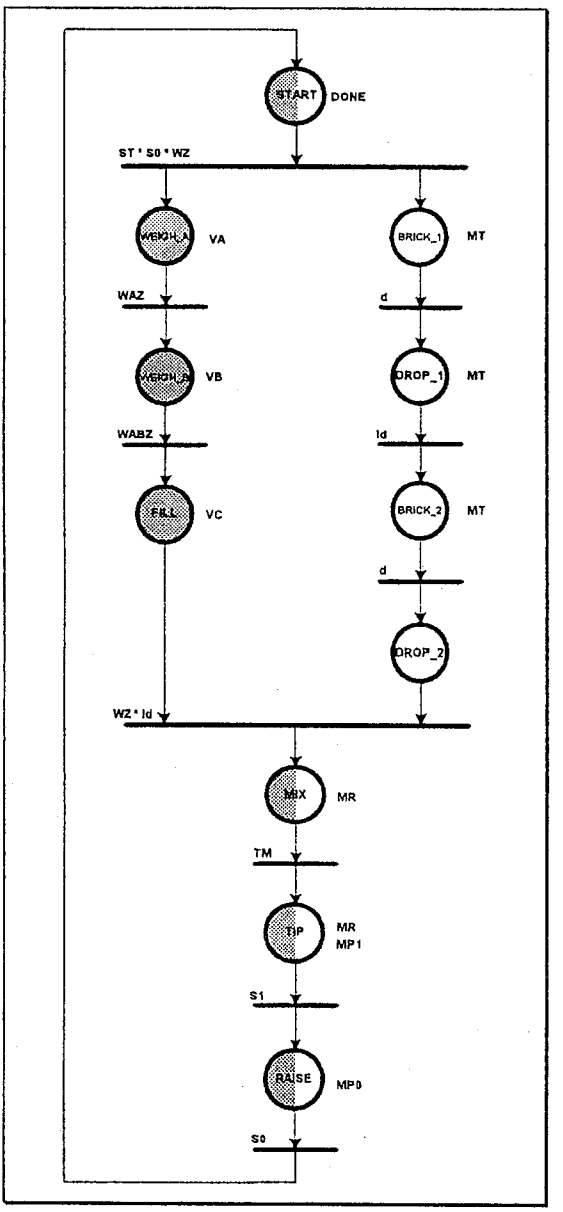
Rys. 4. Opis procesów automatowych w PNSF2

Petriego (rozdz. 4). Kolorowana sieć może być łatwo opisana nawet przy użyciu popularnych języków automatowych i łatwo syntetyzowana z wykorzystywaniem standardowych systemów syntezy układów PLD (rozdz. 5). Istotnym argumentem za kolorowaniem jest również łatwość modelowania sieci z wykorzystaniem procesów komunikujących się za pomocą sygnałów lub zmiennych globalnych, np. w języku VHDL.



Rys. 5. Procesy w sieci Petriego (druga wersja)

Dla odróżnienia procesów sekwencyjnych w tej samej sieci Petriego i celów modelowania w środowisku Design/CPN wprowadzono kolorowe znaczniki. Kolorowaną interpretowaną sieć Petriego zamieszczono na rys. 6.



Rys. 6. Kolorowana sieć Petriego

4. KODOWANIE AUTOMATOWE MIEJSC SIECI PETRIEGO

W dalszej części pracy będzie wykorzystane kodowanie metodą numerowania stanów lokalnych (minimalizując długość kodu dla obydwu automatów łącznie), przedstawione w tabeli 1. Kodowanie odbywa się w sposób standardowy przez przypisanie miejscom, interpretowanym jako stany automatów sekwencyjnych, kodu o minimalnej długości. Minimalizacja łącznej długości kodu dla wszystkich automatów składowych połączonych równolegle odbywa się w sposób heurystyczny. Polega on na wyborze najbardziej korzystnej dekompozycji sieci Petriego na składowe automatowe, minimalizującej długość kodu (liczbę przerzutników), bez konieczności wyznaczania wszystkich wariantów pokrycia sieci sieciami automatowymi (grafami stanów).

Tabela 1

Kodowanie miejsc sieci (wersja 1)

Miejsce	$Q4*Q3*Q2*Q1*Q0$
START	000 – –
BRICK 1	001 – –
DROP 1	010 – –
BRICK 2	011 – –
DROP 2	100 – –
MIX	101 – –
TIP	110 – –
RAISE	111 – –
PO	– – – 00
WEIGH A	– – – 01
WEIGH B	– – – 10
FILL	– – – 11

Realizując układ cyfrowy w elementach PLD w większości przypadków nie stosuje się kodowania typu *one-hot* [23] ze względu na ograniczoną liczbę przerzutników w układzie scalonym. Przerzutniki w PLD zazwyczaj połączone są bezpośrednio z końcówkami wyjściwymi układu scalonego, natomiast liczba przerzutnikow wewnętrznych (ang. *burried*) jest stosunkowo mała.

Opisując układ w postaci automatu współbieżnego Moore'a można potraktować sygnały wyjściowe jako rejestrowe zmienne stanu (pamięć stanu automatu współbieżnego) [7], jak podano w tabeli 2.

Tabela 2

Kodowanie miejsc sieci (wersja 2)

Miejsce	*DONE*MR*MP1*MP0*Q1*Q0 * VA*VB*VC*
START	1000– – – – –
BRICK 1	000000 – – –
DROP 1	000001 – – –
BRICK 2	000010 – – –
DROP 2	000011 – – –
MIX	0100– – – – –
TIP	0110– – – – –
RAISE	0001– – – – –
PO	– – – – – – 000
WEIGH A	– – – – – – 100
WEIGH B	– – – – – – 010
FILL	– – – – – – 001

MT = !(Q1 * Q0)

5. PRZYKŁADY MODELOWANIA SIECI KOLOROWANYCH AUTOMATOWO

W dalszej części pracy, w celu zaprezentowania kilku sposobów opisu układów współbieżnych, podano różne reprezentacje kolorowanych automatowo sieci Petriego w wybranych językach wykorzystywanych w systemach syntezy układów PLD. Z braku miejsca ograniczono się jedynie do podania istotnych cech sposobów modelowania, nie uwzględniajac szczegółów architektury wybranego elementu PLD. Zrezygnowano także z deklaracji dodatkowych, koniecznych do implementacji. W odróżnieniu do prac [21, 22], modelując sieć potraktowano ją bardziej jako jednorodną strukturę z kolorowanymi znacznikami, niż jako powiązane współbieżne automaty składowe o charakterze sekwencyjnym.

Pierwszy z przykładów (rozdz. 5.1) ilustruje sposób specyfikacji w specjalizowanym języku automatowym, na przykładzie kompilatora CUPL. Powiazane automaty sekwencyjne [21], odwzorowujące funkcjonalnie sieć Petriego, albo równoważny automat współbieżny [5] opisywane są ogólną formą regułową języka MachXL 2.0 (rozdz. 5.2) będącego nowszą wersją języka PALASM dedykowaną dla układów CPLD firmy AMD. Istnieje możliwość opisu pojedynczego procesu sekwencyjnego (automatu) w segmencie dedykowanym dla automatu (STATE) [17]. Ta forma jednak nie jest już zalecana. Ponadto próba deklaracji jednocześnie kilku automatów sygnalizowana jest jako błąd. Warto tu nadmienić, że opis regułowy możliwy jest również w języku CUPL.

Sposób modelowania w języku VHDL, przedstawiony w rozdz. 5.3 odzwierciedla swoim stylem opis sieci Petriego zamieszczony w rozdz. 5.1. Mimo, że VHDL nie zawiera deklaracji automatu sekwencyjnego, automaty można modelować różnymi sposobami [22, 24]. W ramach systemu PeNCAD zbudowano, między innymi, translator specyfikacji regułowej automatu na język VHDL [24].

5.1. PRZYKŁAD MODELOWANIA W JĘZYKU CUPL

Opis automatowy języka CUPL, w odróżnieniu od np. specjalizowanego (segment STATE) języka automatowego w kompilatore PALASM/MachXL, umożliwia łaczną syntezę kilku współpracujących automatów w jednym elemencie scalonym [17, 21]. Automaty te implementują odpowiednie podsieci zdekomponowanej sieci Petriego. W języku CUPL składnię opisu pojedynczego automatu przedstawią się jako:

SEQUENCE *Wektor_stanu* {*Opis_stanu* [...]}.

Wektor_stanu jest nazwą wektora stanów wykorzystywanego do kodowania poszczególnych stanów. Wektor ten musi być zdefiniowany przy użyciu odpowiedniej deklaracji, tzn.

Field *Automat0* = [*QA1..0*],

Opis_stanu opisuje zmiany poszczególnych stanów występujących w danym grafie wraz ze specyfikacją wyjść.

Każda zmiana stanu jest opisana w postaci segmentu PRESENT:

PRESENT *Stan_aktualny Opis_przejść_wyjść*

Stan_aktualny jest to zakodowana wartosć wektora *Wektor_stanu*, wymienionego w nagłówku opisu za deklaracją SEQUENCE, i identyfikująca rozpatrywany stan. Do czytelnego zakodowania stanów można użyć dyrektywy SDEFINE. Podana wartość kodu musi być unikalna dla każdej deklaracji PRESENT w rozpatrywanym automacie. *Opis_przejść_wyjść* odnosi się do pojedynczego miejsca (stanu). Każdy taki opis jest przedstawiony w następujący sposób:

IF *Warunek* NEXT *Stan_następny* OUT Wyjście.

Warunek reprezentuje warunek logiczny, który musi być spełniony, aby przejście automatu do nowego stanu mogło być wykonane. *Stan_następny* jest etykietą następnego stanu (miejsca w sieci Petriego). Każdej tranzycji odpowiada pojedynczy moduł NEXT. W prezentowanej metodzie stosowane jest również przejście DEFAULT wykorzystywane do utrzymania automatu w aktualnym stanie lokalnym (utrzymania znakowania miejsca).

Wyjście opisuje zadeklarowane wyjście automatu, które jest powiązane z aktualnym lokalnym stanem wewnętrznym lub zmianą takiego stanu. Jeżeli wyjście jest

```
Field Automat0 = [QA1..0];
$define  P0       'b'00
$define  WEIGH_A  'b'01
$define  WEIGH_B  'b'10
$define  FILL     'b'11
Field Automat1 = [QB2..0];
$define  START    'b'000
$define  BRICK_1  'b'001
$define  DROP_1   'b'010
$define  BRICK_2  'b'011
$define  DROP_2   'b'100
$define  MIX      'b'101
$define  TIP      'b'110
$define  RAISE    'b'111
Sequence Automat0  {
 Present  P0
        if Automat1:START & ST & S0 & WZ Next WEIGH_A;
        default Next P0;
 Present  WEIGH_A
        if WAZ Next WEIGH_B;
        default Next WEIGH_A;
        Out VA;
 Present  WEIGH_B
        if WABZ Next FILL;
        default Next WEIGH_B;
        Out VB;
 Present  FILL
        if Automat1:DROP_2 & WZ & !d Next P0;
        default Next FILL;
        Out VC;
}
Sequence Automat1  {
 Present  START
        if Automat0:P0 & ST & S0 & WZ Next BRICK_1;
        default Next START;
        Out DONE;
 Present  BRICK_1
        if d Next DROP_1;
        default Next BRICK_1;
        Out MT;
 Present  DROP_1
        if !d Next BRICK_2;
        default DROP_1;
        Out MT;
 Present  BRICK_2
        if d Next DROP_2;
        default BRICK_2;
        Out MT;
 Present  DROP_2
        if Automat0:FILL & WZ & !d Next MIX;
        default Next DROP_2;
 Present  MIX
        if TM Next TIP;
        default MIX;
        Out MR;
 Present  TIP
        if S1 Next RAISE;
        default Next TIP;
        Out MP1;
        Out MR;
 Present  RAISE
        if S0 Next START;
        default Next RAISE;
        Out MP0;
}
```

Rys. 7. Realizacja automatu w systemie CUPL

warunkowe i zależy od wejść układu, to jest zapisane po warunku IF i interpretowane jako wyjście typu Mealy'ego. W przeciwnym przypadku wyjście zależy tylko od stanu wewnętrznego i jest typu Moore'a. Definiując funkcjonowanie układu, dodatkowo należy uwzględnić, że jeżeli wyjście występuje po słowie kluczowym NEXT, to jest ono realizowane jako wyjście rejestrowe.

Opis sieci Petriego z rys. 6, zgodnie z podanym wyżej formatem, przedstawiono na rys. 7.

5.2. PRZYKŁAD MODELOWANIA W JĘZYKU MachXL

Język PALASM, a w szczególnosci jego wariant dla elementów CPLD typu AMD MACH umożliwia specyfikację o charakterze nieproceduralnym [4, 7, 21]. Regułowy opis w języku PALASM/MachXL z wykorzystaniem instrukcji warunkowej IF ... THEN ... ELSE ... szczególnie dobrze odzwierciedla i realizuje opis w formalnej logice sekwentów [7].

Opis regułowy przedstawia sieć Petriego jako superpozycję kilku współpracujących procesów realizowanych w tym samym układzie scalonym CPLD. Współpraca między procesami odbywa się w sposób naturalny, za pomocą pełnych wektorów stanów (tabela 1) dla miejsc skojarzonych z tą samą tranzycją (rys. 8).

W prezentowanym opisie zgodnym ze składnią języka PALASM dla CPLD (wersja MachXL 2.0) pominięto deklarację wyprowadzeń oraz sposób synchronizacji i zerowania. Sieć Petriego opisana jest regułami typu:

IF *Wyrażenie_logiczne* THEN BEGIN *Równanie_logiczne* [...]
END
ELSE BEGIN *Równanie_logiczne* [...]
END

Wszystkie reguły realizowane są współbieżnie, w związku z czym kolejność ich zapisu nie wpływa na strukturę układu. W celu zwiększenia czytelności opisu wprowadzono nazwy symboliczne wykorzystując konstrukcję STRING. Reprezentują one kody poszczególnych miejsc sieci (stanów lokalnych układu). Ewentualna superpozycja (iloczyn logiczny) kodów miejsc jednocześnie oznakowanych w jawny sposób określiłaby kod wierzchołka grafu znakowań sieci (stanu globalnego układu). Opis sposobów kodowania miejsc sieci wykracza poza założone ramy artykułu [7]. Kilka różnych wersji kodowania omawianego układu przedstawiono w rozdz. 4 (tabele 1 i 2). Po lewej stronie reguły IF ... THEN ... ELSE występują wyrażenia logiczne zbudowane z wykorzystaniem spójników logicznych (/ – NOT, * – AND, + – OR). Natomiast po prawej stronie podaje się zespół równań logicznych (przypisania) $Qi := 0$ lub $Qi := 1$, gdzie i — numer przerzutnika. Z tego względu symboliczny opis stanu aktualnego *STAN* i stanu następnego @ *STAN* różnia się formą, mimo jednakowego kodu. Oznaczenie @ (*next*) wybrano przez analogię do odpowiedniego operatora w logice temporalnej. Zmiany globalnych stanów układu przedstawiane są pośrednio

```
;Kodowanie stanów - miejsc dla automatu 0
;kod aktualnych stanów lokalnych
STRING P0        '(/QA1*/QA0)'
STRING WEIGH_A   '(/QA1* QA0)'
STRING WEIGH_B   '( QA1*/QA0)'
STRING FILL      '( QA1* QA0)'

;kod następnych stanów lokalnych
STRING @P0        ' QA1:=0 QA0:=0 '
STRING @WEIGH_A   ' QA1:=0 QA0:=1 '
STRING @WEIGH_B   ' QA1:=1 QA0:=0 '
STRING @FILL      ' QA1:=1 QA0:=1 '

;Kodowanie stanów - miejsc dla automatu 1
;kod aktualnych stanów lokalnych
STRING START      '(/QB2*/QB1*/QB0)'
STRING BRICK_1    '(/QB2*/QB1* QB0)'
STRING DROP_1     '(/QB2* QB1*/QB0)'
STRING BRICK_2    '(/QB2* QB1* QB0)'
STRING DROP_2     '( QB2*/QB1*/QB0)'
STRING MIX        '( QB2*/QB1* QB0)'
STRING TIP        '( QB2* QB1*/QB0)'
STRING RAISE      '( QB2* QB1* QB0)'
;kod stanów lokalnych nastepnych
STRING @START     ' QB2:=0 QB1:=0 QB0:=0 '
STRING @BRICK_1   ' QB2:=0 QB1:=0 QB0:=1 '
STRING @DROP_1    ' QB2:=0 QB1:=1 QB0:=0 '
STRING @BRICK_2   ' QB2:=0 QB1:=1 QB0:=1 '
STRING @DROP_2    ' QB2:=1 QB1:=0 QB0:=0 '
STRING @MIX       ' QB2:=1 QB1:=0 QB0:=1 '
STRING @TIP       ' QB2:=1 QB1:=1 QB0:=0 '
STRING @RAISE     ' QB2:=1 QB1:=1 QB0:=1 '

;Reguły przejść dla automatu nr 0
IF P0 THEN BEGIN
  IF START * ST * S0 * WZ    THEN BEGIN  @WEIGH_A  END
                             ELSE BEGIN  @P0       END
END

IF WEIGH_A THEN BEGIN
  IF  WAZ        THEN BEGIN  @WEIGH_B  END
                 ELSE BEGIN  @WEIGH_A  END
END

IF WEIGH_B THEN BEGIN
  IF WABZ        THEN BEGIN  @FILL     END
                 ELSE BEGIN  @WEIGH_B  END
END

IF FILL THEN BEGIN
  IF DROP_2 * WZ * /d        THEN BEGIN  @P0    END
                             ELSE BEGIN  @FILL  END
END

;Reguły przejść dla automatu nr 1
IF START THEN BEGIN
  IF P0 * ST * S0 * WZ       THEN BEGIN  @BRICK_1  END
                             ELSE BEGIN  @START    END
END

IF BRICK_1 THEN BEGIN
  IF  d          THEN BEGIN  @DROP_1   END
                 ELSE BEGIN  @BRICK_1  END
END

IF DROP_1 THEN BEGIN
  IF /d          THEN BEGIN  @BRICK_2  END
                 ELSE BEGIN  @DROP_1   END
END

IF BRICK_2 THEN BEGIN
  IF  d          THEN BEGIN  @DROP_2   END
                 ELSE BEGIN  @BRICK_2  END
END

IF DROP_2 THEN BEGIN
```

Rys. 8

w postaci pojedynczych zmian stanow lokalnych. W przedstawionym sposobie specyfikacji przejście między dwoma stanami lokalnymi *STAN1* i *STAN2* ma postać:

```
IF STAN1 THEN BEGIN
                IF Warunek THEN BEGIN STAN2 END
                           ELSE BEGIN @ STAN1 END
              END
```

Warunek jest iloczynem logicznym etykiety tranzycji, czyli wyrażenia logicznego przypisanego do tranzycji oraz symboli (iloczynów kodowych) wszystkich miejsc wejściowych tranzycji, oprócz rozpatrywanego miejsca początkowego *STAN1*. Inne, bardziej zwarte sposoby opisu sieci wykorzystujące formę regułową IF ... THEN ... ELSE podano między innymi w pracy [7].

5.3. PRZYKŁAD MODELOWANIA W JĘZYKU VHDL

Duża część systemów automatycznej syntezy układów PLD akceptuje podzbiór języka VHDL do specyfikacji projektów. Do najpopularniejszych należą: Xilinx (Foundation Series), Aldec (Active CAD), Synopsys (FPGA Express), Altera (MAX Plus), Cypress (Warp). Sygnały wejściowe i wyjściowe są przedstawione jako wektory binarne o odpowiedniej długości. Zastosowano przemianowywania nazwy ALIAS, w celu łatwiejszej symulacji wykorzystując znaczące nazwy poszczególnych sygnałów podanych w projekcie. Dodatkowo, aby uprościć automatyczne wygenerowanie modelu przygotowywany jest moduł biblioteki (PACKAGE), w którym zdefiniowane są niezbędne typy i stałe.

Wyjścia Moore'a są wyrażone jako przypisanie równoległe do danego sygnału wartości '1', gdy aktualny stan równy jest temu, w którym dane wyjście jest ustawione, w przeciwnym wypadku wartość '0'.

Każdy automat może być przedstawiony jako oddzielny proces z wykorzystaniem niezależnych zmiennych i sygnałów. Opisując kolejne automaty można zdefiniować własne typy wyliczeniowe, które symbolicznie przedstawiają poszczególne stany automatu. W rozpatrywanym przypadku są to symboliczne nazwy miejsc.

Każdy proces reprezentuje oddzielny automat. W procesie wykorzystywana jest instrukcja warunkowa (CASE), gdzie zmienną warunkową jest zmienna określająca aktualny stan danego automatu. Dalsze warunki przejść opisane są wyrażeniem logicznym, będącym warunkiem w instrukcji IF ... THEN, sygnałów wejściowych oraz miejsc pochodzących z komunikujących się automatów. W danym opisie specyfikuje się jedynie zmianę stanu. W odróżnienie np. od języka CUPL nie jest wymagane podtrzymywanie stanu (ang. *hold*). Przy zakończeniu każdego procesu następuje przypisanie do sygnału reprezentującego aktualny stan automatu nowo wyznaczonej albo poprzedniej wartości zapamiętywanej pod odpowiednią zmienną zadeklarowaną w procesie. Dodatkową zaletą opisu automatu współbieżnego w języku VHDL jest możliwość łącznej symulacji całego urządzenia cyfrowego [20].

Rys. 9 przedstawia realizację procesów z wykorzystaniem instrukcji warunkowych języka VHDL.

```
use std.std_logic_1164.all;
PACKAGE Example_Package is
  constant NUMBER_INPUTS       :       integer :=8;
  constant NUMBER_OUTPUTS      :       integer :=8;
  subtype T_INPUTS is std_logic_vector(NUMBER_INPUTS-1 downto 0);
  subtype T_OUTPUTS is std_logic_vector(NUMBER_OUTPUTS-1 downto 0);
  type T_STATES0 is (P0, WEIGH_A, WEIGH_B, FILL);
  type T_STATES1 is (START,BRICK_1,DROP_1,BRICK_2,DROP_2,MIX,TIP,
                     RAISE);
end Example_Package;

use std.std_logic_1164.all;
use work. Example_Package.all;
entity example is
    port ( CLK  :  in std_logic;
           X    :  in  T_INPUTS  := (others => '0');
           Y    :  out T_OUTPUTS := (others => '0')  );
end Example;

architecture ARCH1 of Example is
signal CURRENT_STATE0 : T_STATES0;
signal CURRENT_STATE1 : T_STATES1;

  alias ST   : std_logic is X(7);
  alias S1   : std_logic is X(6);
  alias S0   : std_logic is X(5);
  alias WZ   : std_logic is X(4);
  alias WAZ  : std_logic is X(3);
  alias WABZ : std_logic is X(2);
  alias d    : std_logic is X(1);
  alias TM   : std_logic is X(0);

  alias MT   : std_logic is Y(7);
  alias MR   : std_logic is Y(6);
..alias DONE : std_logic is Y(5);
  alias VA   : std_logic is Y(4);
  alias VB   : std_logic is Y(3);
  alias VC   : std_logic is Y(2);
  alias MP1  : std_logic is Y(1);
  alias MP0  : std_logic is Y(0);

begin
SET_STATE0 : process (CLK)
  variable NEXT_STATE0 : T_STATES0;
  begin
     case CURRENT_STATE0 is
      when P0 => if (ST='1') and (S0='1') and (WZ='1') and
                    (CURRENT_STATE1=START)
                    then NEXT_STATE0:=WEIGH_A;
                 end if;
      when WEIGH_A => if (WAZ='1') then NEXT_STATE0:=WEIGH_B;  end if;
      when WEIGH_B => if (WABZ='1') then NEXT_STATE0:=FILL;    end if;
      when FILL => if (WZ='1') and (d='0') and (CURRENT_STATE1=DROP_2)
                    then NEXT_STATE0:=P0;
                   end if;
     end case;
     CURRENT_STATE0 <= NEXT_STATE0;
  end process SET_STATE0;

SET_STATE1 : process (CLK)
  variable NEXT_STATE1 : T_STATES1;
  begin
     case CURRENT_STATE1 is
      when START   => if (ST='1') and (S0='1') and (WZ='1') and
                         (CURRENT_STATE0=P0)
                      then NEXT_STATE1:=BRICK_1;
                     end if;
      when BRICK_1 => if (d='1') then NEXT_STATE1:=DROP_1;  end if;
      when DROP_1  => if (d='0') then NEXT_STATE1:=BRICK_2; end if;
      when BRICK_2 => if (d='1') then NEXT_STATE1:=DROP_2;  end if;
      when DROP_2  => if (WZ='1') and (d='0') and (CURRENT_STATE0=FILL)
                      then NEXT_STATE1:=MIX;
```

Rys. 9.

```
                    end if;
      when MIX      => if (TM='1') then NEXT_STATE1:=TIP;    end if;
      when TIP      => if (S1='1') then NEXT_STATE1:=RAISE; end if;
      when RAISE    => if (S0='1') then NEXT_STATE1:=START;  end if;
     end case;
     CURRENT_STATE0 <= NEXT_STATE0;
  end process SET_STATE1;
 VA <= '1' when CURRENT_STATE0=WEIGH_A else '0';
   VB <= '1' when CURRENT_STATE0=WEIGH_B else '0';
   VC <= '1' when CURRENT_STATE0=FILL    else '0';
 DONE <= '1' when CURRENT_STATE1=START else '0';
   MT <= '1' when CURRENT_STATE1=BRICK_1
              or CURRENT_STATE1=DROP_1
              or CURRENT_STATE1=BRICK_2  else '0';
   MR <= '1' when CURRENT_STATE1=MIX
              or CURRENT_STATE1=TIP else '0';
   MP1  <= '1' when CURRENT_STATE1=TIP   else '0';
   MP0  <= '1' when CURRENT_STATE1=RAISE else '0';

end ARCH1;
```

Rys. 9. Realizacja procesów w języku VHDL

PODSUMOWANIE

W artykule rozszerzono wyniki dotychczasowych prac autorów, opublikowanych m.in. w [7, 21], przedstawiając efektywniejszy i łatwiejszy sposób realizacji kontrolerów współbieżnych przy zastosowaniu standardowych systemów syntezy układów programowalnych. Podano przykłady demonstrujące przydatność proponowanej metody w projektowaniu sterowników logicznych. Zamieszczono szczegółowe przykłady syntezy tego samego sterownika w systemach CUPL, PALASM i VHDL. Zademonstrowano różne sposoby kodowania kolorowanej P/T sieci. Krótko przedstawiono program (typu *front-end*) przygotowujący dane dla popularnych kompilatorów PLD i FPGA. Omówiono koncepcję rozszerzonego języka regułowego PNSF2 jako nakładki do kilku typowych kompilatorów. Stanowi ona dogodny pomost między specyfikacją, a metodami formalnymi bazującymi na logice matematycznej [4].

Dalsze prace autorów idą w kierunku implementacji sieci hierarchicznych bez konieczności ich redukowania do sieci płaskich. Szczegółowego przebadania w formie *bechmarków* wymaga wpływ rodzaju kodowania na złożoność układu. Wydaje się jednak, że najkorzystniejszym sposobem kodowania stanów lokalnych w syntezie układów PLD jest wykorzystanie do tego celu sygnałów wyjściowych w postaci rejestrowej.

Praca powstała w ramach grantu KBN nr 8 T11C 034 12.

BIBLIOGRAFIA

1. M. Adamski: *Realizacja sieci Petri z wykorzystaniem PLA*. IV Krajowa Konferencja „Teoria obwodów i układy elektroniczne", Drzonków k/Zielonej Góry, ss. 455–459, 28–30.10.1981
2. M. Adamski: *Microcomputer Implementation of Concurrent Control Algoritm*. Proceedings of the 2nd Conference on Microcomputer Systems MICROSYSTEM'84, Bratysława, Czechosłowacja, vol. 1, ss. 117–121, 23–25.10.1984
3. M. Adamski: *Koncepcja systemu CAD-LOGIK i jego wykorzystanie w projektowaniu specjalizowanych układów cyfrowych*. Metrologia i Systemy Pomiarowe, Warszawa, 1991, ss. 65–74
4. M. Adamski: *Projektowanie układów cyfrowych systematyczną metodą strukturalną*. Monografia, Wydawnictwo Wyższej Szkoły Inżynierskiej w Zielonej Górze, Zielona Góra, 1990
5. M. Adamski: *Parallel Controller Implementation using Standard PLD Software*. W.R. Moore, W. Luk [ed.], *FPGAs*, Abingdon EE&CS Books, Abingdon, England, 1991, ss. 296–304
6. M. Adamski, M. Węgrzyn: *Hierarchically structured coloured Petri net specification and validation of concurrent controllers*. Proceedings of the 39. Internationales Wissenschaftliches Kolloquium IWK'94, Technische Universität Ilmenau, Germany, Band 1, ss. 517–522, 27–30.09.1994
7. M.A. Adamski, J.L. Monteiro: *Declarative specification of System Independent Logic Controller Programs*. Proceedings of the IEEE International Symposium on Industrial Electronics ISIE'96, Warszawa 1996, ss. 305–310
8. M. Auguin, F. Boeri, C. Andre: *Systematic method of realisation of interpreted Petri nets*. Digital Processes, 1980, vol. 6, no. 1
9. K. Biliński, M. Adamski, J.M. Saul, E.L. Dagless: *Petri net based algorithms for parallel controllr synthesis*. IEE Proceedings — E, Computers and Digital Techniques, 1994, vol. 141, no. 6, ss. 405–412
10. R.K. Brayton [i inni]: *VIS: A System for Verification and Synthesis*. Proc. of the 8th Intenational Conference on Application and Theory of Petri Nets, Lecture Notes in Computer Science, Springer-Verlag, ss. 332–334, July 1996
11. J.M. Fernandes, M. Adamski, A. Proença: *VHDL Generation from Hierarchical Petri Net Specifications of Paralell Controllers*. IEE Proceedings — E, Computer and Digital Techniques, 1997, vol. 144, No. 2, ss. 127–137, march 1997
12. D. Hladik, K. Jelemenska, S. Młynska: *Design of Sequantial Control for parallel process*. Proceedings of the 33. Internationales Wissenschaftliches Kolloquium IWK'88, Technische Universität Ilmenau, Germany, 1988, Heft 4, ss. 183–186
13. K. Jensen: *Coloured Petri Nets. Basic Concept, Analysis Methods and Practical Use*. Vol. 1, Springer-Verlag, Berlin Heideldelberg, 1992
14. J. Kalinowski, K. Jasiński: *Synteza układów cyfrowych z wykorzystaniem sieci Petriego*. Rozparwy Elektrotechniczne, 1986, t. 32, z. 2, ss. 339–351
15. J. Kalinowski, T. Łuba: *Metoda syntezy logicznej układów cyfrowych opisywanych sieciami Petriego*. Rozprawy Elektrotechniczne 1986, t. 32, z. 4, ss. 1253–1263
16. T. Kozłowski, E.L. Dagless, J.M. Saul, M. Adamski, J. Szajna: *Parallel controller synthesis using Petri nets*. IEE Proceedings — E, Computers and Digital Techniques, July 1995, vol. 142, no. 4, ss. 263–271
17. T. Łuba, M.A. Markowski, B. Zbierzchowski: *Kompilatory układów logicznych*. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 1995
18. J. Pardey, A. Amroun, M. Bolton, M. Adamski: *Parallel Controller Synthesis for Programmable Logic Devices*. Microprocessors and Microsystems, 1994, vol. 18, no. 8, ss. 451–458
19. E.M. Sentovich [i inni]: *A System for Sequential Circuit Synthesis*. University of California, Berkeley, Memorandum No. UCB/ERL M92/41, May 1992
20. M. Węgrzyn, M. Oleszczuk: *Zastosowanie jezyka VHDL do modelowania i testowania programowalnych struktur logicznych PLD*. 16 Miedzynarodowe Sympozjum Naukowe, t. 2, Wyższa Szkoła Inżynierska w Zielonej Górze, ss. 75–78, Zielona Góra, 1994

21. M. Węgrzyn: *Implementacja współbieżnych kontrolerów cyfrowych z wykorzystaniem PLD.* Kwart. Elektr. i Telekom., 1996, t. 42, z. 2, ss. 235–251
22. M. Węgrzyn, P. Wolański, M. Adamski, J. Monteiro: *Field Programmable Device as a Logic Controller.* Proceedings of the 2nd International Automatic Control Conference CONTROL'96, Oporto, ss. 715–720, 11–13.09.1996
23. M. Węgrzyn, M.A. Adamski, J.L. Monteiro: *Reconfigurable Logic Controller with FPGA.* Proceedings of the 4th IFAC Workshop on Algorithms and Architectures for Real-Time Control-AARTC'97, Vilamoura, Algarve, Portugal, 9–11.04.1997, ss. 247–252
24. P. Wolański: *Modelowanie i symulacja ASM w podzbiorze języka VHDL.* Kwart. Elektr. i Telekom. 1996, t. 42, z. 2, ss. 129–144

M. WĘGRZYN, M. ADAMSKI

SNTHESIS OF LOGIC CONTROLLER USING STANDARD FPLD COMPILERS

Summary

Design methods for concurrent controllers play a very important role in advanced digital system design, especially of Logic Controllers. Behaviour of the controller can be described by means of using coloured, safe, interpreted Petri nets. For the reason of effective implementation, the net that describes concurrent controller, may be divided (decomposed) into the separate sequential parts, called State Machines, and be SM-coloured. The colours distinguish the intended or deduced sequential processes. In the previous authors' papers such kind of Petri net colouring has been presented for the Petri net parallel decomposition. It is demonstrated in this paper in the new context as a part of the design methodology for direct mapping of Petri net into LPLD. The front-end language, which is intended for the proposed conceptual model, is introduced. It is automatically translated to standard or proprietary HDLs. The paper presents only some selected techniques needed to design concurrent state machine-based controllers, with a special emphasis on synthesis using programmable logic devices and related hardware description languages. All of them are used as inputs to the popular LPDL compilers. The descriptions in three different styles of the same controller in CUPL, PALASM/MachXL, and VHDL are included. The presented design methodology may be easily adapted to other FPLD-oriented HDLs, for example for DSL (MINC), ABEL (DATA I/O) and AHDL ALTERA.

Key words: Logic controllers, FPLD compilers, Petri nets.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1997, 43, z. 3, ss. 373–396

Kształtowanie charakterystyk częstotliwościowych wzmacniaczy z prądowym sprzężeniem zwrotnym w technice CMOS

STANISŁAW KUTA, GRZEGORZ KRAJEWSKI, JACEK JASIELSKI

Katedra Elektroniki, Akademia Górniczo Hutnicza, Kraków

Otrzymano 1997.03.14

Autoryzowano 1997.05.06

W pracy dokonano porównania i oceny wpływu nieidealności wzmacniaczy CMOS z prądowym sprzężeniem zwrotnym (CFOAs) oraz impedancji obciążenia i impedancji w torze wzmocnienia zwrotnego na pole wzmacniacza. Przeprowadzono analizę transmitancji wzmacniaczy oraz określono parametry zewnętrznego obwodu sprzężenia zwrotnego, przy których następuje kompensacja dominującego bieguna charakterystyk częstotliwościowych i zwiększenie marginesów stabilności wzmacniaczy. Rozważono wzmacniacze CFOAs ze stopniem transimpedancyjnym i wyjściowym buforem napięciowym oraz wzmacniacz CFOA ze sterowanym źródłem prądu, zrealizowanym w postaci kaskadowego połączenia stopnia transimpedancyjnego sterującego stopniem transkonduktancyjnym. Przedstawiono projekty reprezentujące obydwie grupy wzmacniaczy CFOAs oraz wyniki ich symulacji w programie HSPICE dla technologii CMOS 1.2 μm z firmy AMS.

Słowa kluczowe: Wzmacniacze CMOS, CFOAs i CFOA, projektowanie symulacyjne.

1. WPROWADZENIE

Wzmacniacze operacyjne z prądowym sprzężeniem zwrotnym (ang. *Current — Feedback Operational Amplifiers* — CFOA's), nazywane również wzmacniaczami w trybie prądowym, wykazują kilka zalet w porównaniu z konwencjonalnymi wzmacniaczami operacyjnymi, spośrod których do najważniejszych należy zaliczyć większą maksymalną prędkość zmian napięcia wyjściowego SR (ang. *slew — rate*) oraz możliwość zapewnienia w przybliżeniu tej samej wartości górnej częstotliwości granicznej wzmacniacza przy zmianie wzmocnienia układu z zamkniętą pętlą sprzężenia zwrotnego.

Najbardziej popularne rozwiązanie układowe wzmacniacza z prądowym sprzężeniem zwrotnym wykorzystuje komplementarną technologię bipolarną o symetrycznych tranzystorach pnp i npn. Rozwiązanie to nie może być w prosty sposób zaadaptowane dla realizacji w technologii CMOS oraz niekomplementarnej technologii BiCMS.

Istnieje kilka typowych rozwiązań układowych CFOA's [1 ÷ 5] przy czym każde z nich może być zakwalifikowane do jednej z dwóch grup:

1) CFOA's ze stopniem transimpedancyjnym i wyjściowym buforem napięciowym;
2) CFOA's z prądowym źródłem sterowanym (zrealizowanym w postaci stopnia transimpedancyjnego sterującego stopień transkonduktancyjny).

W obu rozwiązaniach istnieją ograniczenia odnośnie wartości impedancji sprzężenia zwrotnego oraz obciążenia, szczególnie jeśli uwzględnić nieidealne parametry stosowanych wzmacniaczy CFOA's.

W rozdziale 2 dokonano porównania oraz oceny wpływu nieidealności wzmacniaczy CFOA's oraz impedancji obciążenia i impedancji w torze sprzężenia zwrotnego na pole wzmocnienia w obu grupach wzmacniaczy. W rozdziale 3 przeprowadzono analizę transmitancji nieodwracających wzmacniaczy napięciowych z prądowym sprzężeniem zwrotnym oraz wyprowadzono zależności pozwalające na dobór takich parametrów zewnętrznego obwodu sprzężenia zwrotnego, przy których następuje kompensacja dominującego bieguna charakterystyk częstotliwościowych wzmacniaczy i zwiększenie ich marginesów stabilności. W rozdziale 4 przedstawiono projekty CFOA's w technologii CMOS, reprezentujące obydwie grupy wzmacniaczy. W projektach tych zawarte są różne oryginalne ulepszenia wybranych bloków funkcjonalnych wykorzystanych do budowy wzmacniaczy, w celu zapewnienia lepszych parametrów wyjściowych. Ostatni rozdział pracy przedstawia wyniki symulacji zaprojektowanych wzmacniaczy w technologii 1.2 μm CMOS z firmy AMS, przy wykorzystaniu programu HSPICE.

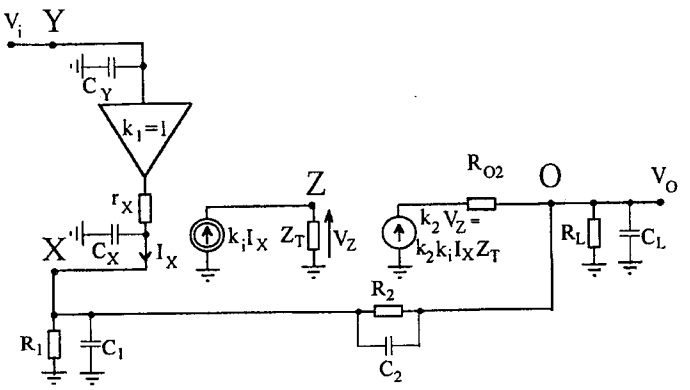
2. POLA WZMOCNIEŃ WZMACNIACZY Z PRĄDOWYM SPRZĘŻENIEM ZWROTNYM

Typowy wzmacniacz operacyjny ze sprzężeniem prądowym składa się z wejściowego wtórnika napięciowego i stopnia transimpedancyjnego. W celu zapewnienia małej rezystancji wyjściowej wzmacniacza operacyjnego, na wyjściu stopnia transimpedancyjnego stosuje się bufor napięciowy. Struktura opisywanego wzmacniacza ze sprzężeniem prądowym odpowiada strukturze konwejora prądowego drugiej generacji CCII, w którym napięcie z wyjścia Z podane jest na wyjście wzmacniacza operacyjnego O poprzez bufor napięciowy. Taki wzmacniacz został pokazany w formie schematu zastępczego na rys. 1.

W schemacie tym uwzględniono rzeczywiste parametry poszczególnych stopni. Wejściowy wtórnik napięciowy jest reprezentowany przez:

— wzmocnienie napięciowe $k_1 = 1$,

— pojemność wejściową C_y, która przy sterowaniu prawie napięciowym może być pominięta,



Rys. 1. Schemat zastępczy wzmacniacza CFOA ze stopniem transimpedancyjnym

— rezystancję wyjściową r_x i pojemnosć wyjściową C_x, które stanowią zarazem rezystancję wejściową i pojemność wejściową niskoimpedancyjnego wejścia prądowego X wzmacniacza operacyjnego.

Prąd wejściowy I_x wejścia X steruje stopniem transimpedancyjnym o transimpedancji Z_T, a napięcie wyjściowe tego stopnia poprzez wtórnik napięciowy o wzmocnieniu k_2 (w przybliżeniu równym jedności) podane jest na wyjście wzmacniacza. Rezystancja wyjściowa wtórnika napięciowego R_{02} stanowi rezystancję wyjściową wzmacniacza operacyjnego. Na rys. 1 uwzględniono także elementy obwodu sprzężenia zwrotnego R_1, R_2, C_1, C_2 w konfiguracji wzmacniacza nieodwracajacego oraz impedancję obciążenia R_L, C_L.

Przyjmiemy, że transimpedancja wzmacniacza ma jeden dominujący biegun, tzn. może być wyrażona wzorem:

$$Z_T(j\omega) = \frac{R_T}{1 + \dfrac{j\omega}{\omega_a}}. \tag{2.1}$$

Nie uwzględniając wpływu pojemności oraz stosując zasadę superpozycji, zgodnie z oznaczeniami na rys. 1, możemy zapisać równania

$$\begin{cases} I_x = \dfrac{k_1 V_i}{r_x + R_1 \parallel R_2} - \dfrac{V_0}{R_2 + r_x \parallel R_1} \dfrac{R_1}{r_x + R_1} \\ V_0 = k_i k_2 Z_T(j\omega) I_x - R_{02}\left(\dfrac{V_0}{R_2 + r_x \parallel R_1} - \dfrac{k_1 V_i}{r_x + R_1 \parallel R_2} \dfrac{R_1}{R_1 + R_2} + \dfrac{V_0}{R_L} \right). \end{cases} \tag{2.2}$$

Dla uproszczenia zapisu wprowadzono znak $\parallel$, oznaczający równoległe połączenie elementów.

Rozwiązując układ równań (2.2) oraz wykorzystując zależność (2.1), otrzymujemy

$$A(j\omega) = \frac{V_0}{V_i} = A_{0x}\frac{1 + \dfrac{j\omega}{\omega_{0x}}}{1 + \dfrac{j\omega}{\omega_{ax}}}, \tag{2.3}$$

gdzie:

$$A_{0x} = \frac{A_0 + \dfrac{R_{02}}{k_i k_2 R_T}}{\dfrac{1}{k_1} + \dfrac{r_x}{k_1 k_i k_2 R_T}\left[A_0\left(1 + \dfrac{R_{02}}{R_L}\right) + \dfrac{R_{02}}{R_1}\right] + \dfrac{R_2}{k_1 k_i k_2 R_T}\left(1 + \dfrac{R_{02}}{R_L}\right) + \dfrac{R_{02}}{k_1 k_i k_2 R_T}} \tag{2.4}$$

$$A_0 = 1 + \frac{R_2}{R_1} \tag{2.5}$$

$$\omega_{ax} = \omega_a \frac{k_1 k_i k_2 R_T}{r_x\left[\left(1 + \dfrac{R_{02}}{R_L}\right)A_0 + \dfrac{R_{02}}{R_1}\right] + \left(1 + \dfrac{R_{02}}{R_L}\right)R_2 + R_{02}} \cdot$$

$$\cdot\left\{\frac{1}{k_1} + \frac{r_x}{k_1 k_i k_2 R_T}\left[A_0\left(1 + \frac{R_{02}}{R_L}\right) + \frac{R_{02}}{R_1}\right] + \frac{R_2}{k_1 k_i k_2 R_T}\left(1 + \frac{R_{02}}{R_L}\right) + \frac{R_{02}}{k_1 k_i k_2 R_T}\right\} \tag{2.6}$$

$$\omega_{0x} = \omega_a\left(1 + A_0\frac{k_i k_2 R_T}{R_{02}}\right). \tag{2.7}$$

Zakładając, że biegun i zero transmitancji (2.3) spełniają warunek

$$\omega_{0x} \gg \omega_{ax}, \tag{2.8}$$

szerokość pasma wzmacniacza z zamkniętą pętlą wynosi

$$GW_x = \omega_{ax}, \tag{2.9}$$

a pole wzmocnienia wzmacniacza z zamkniętą pętlą jest określone zależnością

$$GB_x = A_{0x} GW_x =$$

$$= \left(A_0 + \frac{R_{02}}{k_i k_2 R_T}\right)\omega_a \frac{k_1 k_i k_2 R_T}{r_x\left[\left(1 + \dfrac{R_{02}}{R_L}\right)A_0 + \dfrac{R_{02}}{R_L}\right] + \left(1 + \dfrac{R_{02}}{R_L}\right)R_2 + R_{02}}. \tag{2.10}$$

Przedstawiając zależności (2.4, 2.6, 2.7, 2.9, 2.10) dla układu idealnego, w którym

$$k_1 = 1, \quad k_i = 1, \quad k_2 = 1, \quad r_x = 0, \quad R_{02} = 0 \tag{2.11}$$

otrzymujemy

$$A_{00} = \frac{A_0}{1 + \frac{R_2}{R_T}}. \tag{2.12}$$

Biegun transmitancji wynosi

$$\omega_{a0} = \omega_a \frac{R_T}{R_2}\left(1 + \frac{R_2}{R_T}\right), \tag{2.13}$$

zaś zero transmitancji dąży do nieskończoności

$$\omega_{00} \to \infty. \tag{2.14}$$

Szerokość pasma wynosi

$$GW_0 = \omega_{a0}, \tag{2.15}$$

natomiast pole wzmocnienia

$$GB_0 = A_{00} GW_0 = A_0 \omega_a \frac{R_T}{R_2}. \tag{2.16}$$

Gdy $R_T \to \infty$ to $A_{00} \to A_0$.

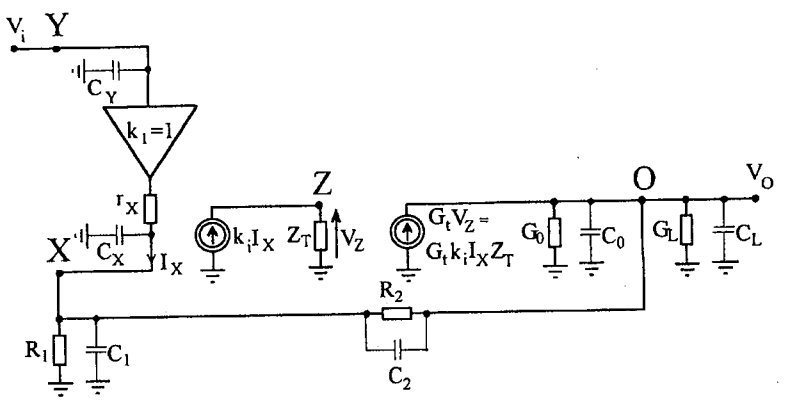
Na podstawie zależności (2.10) i (2.16) można obliczyć stosunek pola wzmocnienia wzmacniacza rzeczywistego do pola wzmocnienia wzmacniacza idealnego

$$\frac{GB_x}{GB_0} = \left(1 + \frac{R_{02}}{A_0 k_i k_2 R_T}\right) \frac{k_1 k_i k_2 R_2}{r_x\left[\left(1 + \frac{R_{02}}{R_L}\right)A_0 + \frac{R_{02}}{R_1}\right] + \left(1 + \frac{R_{02}}{R_1}\right)R_2 + R_{02}}. \tag{2.17}$$

Jak wynika z zależności (2.4) i (2.12) wzmocnienie wzmacniacza z zamkniętą pętlą sprzężenia zwrotnego zależy głównie od wartości rezystancji R_1 i R_2 oraz w niewielkim stopniu od wartości rezystancji r_x i R_{02}. Z równań (2.6) i (2.9) oraz (2.134) i (2.15) wynika, że szerokość pasma wzmacniacza zależy głównie od wartości rezystancji R_2, jednak wpływ pozostałych parametrów: r_x i R_{02} nie jst pomijalny. W rzeczywistym wzmacniaczu, przy niezerowych wartościach r_x i R_{02}, szerokość pasma zależy również od rezystancji R_1, przez co pasmo wzmacniacza maleje ze wzrostem wzmocnienia. Nie ma również pełnej swobody kształtowania pasma wzmacniacza niezależnie od wzmocnienia. Jednak w pewnym ograniczonym zakresie, zależnym od wartości parametrów r_x i R_{02}, istnieje możliwość doboru wartości rezystancji R_1 i R_2 zapewniających w przybliżeniu to samo pasmo dla różnych wartości wzmocnienia. W tym celu rezystancje te należy dobrać na podstawie zależności (2.4), (2.6) metodą kolejnych przybliżeń, przy czym dla zapewnienia stałego pasma, ze wzrostem wzmocnienia należy nieco zmniejszyć wartość rezystancji R_2, a następnie rezystancję R_1 należy dobrać dla zapewnienia wymaganego wzmocnienia.

Równania (2.4), (2.6) nie pozwalają na wyprowadzenie prostych zależności, określajacych bezpośrednio wartości rezystancji R_1, R_2, zapewniających wymagane pasmo i wzmocnienie wzmacniacza.

Innym sposobem budowy wzmacniacza operacyjnego z prądowym sprzężeniem zwrotnym jest wykorzystanie sterowanego źródła prądowego zrealizowanego w postaci stopnia transimpedancyjnego sterującego stopniem transkonduktancyjnym. Schemat zastępczy takiego układu przedstawiono na rys. 2.



Rys. 2. Schemat zastępczy wzmacniacza CFOA ze sterowanym źródłem prądowym

Oznaczenia zastosowane na rys. 2 są takie same jak na rys. 1. Końcowy stopień transkonduktancyjny posiada konduktancję wyjściową G_0 i pojemność wyjściową C_0. Wzmocnienie wzmacniacza, wyznaczone na podstawie schematu zastępczego z rys. 2, określone jest wzorem (2.3), w którym:

$$A_{0x} = \frac{A_0 + \dfrac{1}{k_i G_t R_T}}{\dfrac{1}{k_1} + \dfrac{r_x}{k_1 k_i G_t R_T}\left[A_0(G_0 + G_L) + \dfrac{1}{R_1}\right] + \dfrac{R_2}{k_1 k_i G_t R_T}(G_0 + G_L) + \dfrac{1}{k_1 k_i G_t R_T}} \tag{2.18}$$

$$\omega_{ax} = \omega_a \frac{k_1 k_i G_t R_T}{1 + r_x\left[(G_0 + G_L)A_0 + \dfrac{1}{R_1}\right] + (G_0 + G_L)R_2} \cdot$$

$$\cdot\left\{\frac{1}{k_i} + \frac{r_x}{k_1 k_i G_t R_T}\left[A_0(G_0 + G_L) + \frac{1}{R_1}\right] + \frac{R_2}{k_1 k_i G_t R_T}(G_0 + G_L) + \frac{1}{k_1 k_i G_t R_T}\right\}, \tag{2.19}$$

$$\omega_{0x} = \omega_a(1 + A_0 k_i G_t R_T), \tag{2.20}$$

$$GB_x = \left(A_0 + \frac{1}{k_i G_t R_T}\right)\omega_a \frac{k_1 k_i G_t R_T}{1 + r_x\left[(G_0 + G_L)A_0 + \frac{1}{R_1}\right] + (G_0 + G_L)R_2}. \quad (2.21)$$

Dla układu idealnego, w którym

$$k_1 = 1, \quad k_i = 1, \quad r_x = 0, \quad G_0 = 0 \quad (2.22)$$

otrzymujemy

$$A_{00} = \frac{A_0 + \frac{1}{G_t R_T}}{1 + \frac{R_2}{G_t R_T} + \frac{1}{G_t R_T}}, \quad (2.23)$$

gdzie: A_0 określone jest zależnoscią (2.5),

$$\omega_{a0} = \omega_a \frac{G_t R_T}{1 + G_L R_2}\left(1 + \frac{R_2}{G_t R_T} G_L + \frac{1}{G_t R_T}\right) \quad (2.24)$$

$$\omega_{00} = \omega_a (1 + A_0 G_t R_T). \quad (2.25)$$

Gdy $A_0 \gg 1$, to $\omega_{a0} \ll \omega_{00}$ i wtedy,

$$GW_0 = \omega_{a0} \quad (2.26)$$

$$GB_0 = A_{00} GW_0 = \left(A_0 + \frac{1}{G_t R_T}\right)\omega_a \frac{G_t R_T}{1 + G_L R_2}. \quad (2.27)$$

Stosunek pola wzmocnienia wzmacniacza rzeczywistego (2.21) do pola wzmocnienia wzmacniacza idealnego (2.27) określa zależność

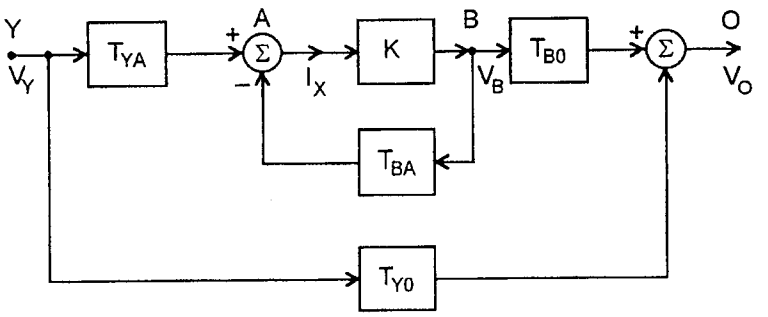
$$\frac{GB_x}{GB_0} = \frac{A_0 + \frac{1}{k_i G_t R_T}}{A_0 + \frac{1}{G_t R_T}} \frac{k_1 k_i (1 + G_L R_2)}{1 + r_x\left[(G_0 + G_L)A_0 + \frac{1}{R_1}\right] + (G_0 + G_L)R_2}. \quad (2.28)$$

Wzory (2.18) i (2.23) wskazują, że wzmocnienie wzmacniacza, podobnie jak poprzednio, zależy głównie od wartości rezystancji R_1, R_2, ale również zależy od wartości rezystancji r_x i konduktancji G_0 i G_L. Duże wartości konduktancji G_0 i G_L wpływają znacząco zarówn ona wzmocnienie, jak i na szerokość pasma wzmacniacza. Nie ma również pełnej swobody kształtowania pasma wzmacniacza niezależnie od wzmocnienia, ponieważ szerokość pasma zależy od wartości obu rezystancji R_1 i R_2 (głównie od R_2).

W podobny sposób do opisanego wcześniej, przy wykorzystaniu zależności (2.18) i (2.19), metodą kolejnych przybliżeń można dobrać wartosć rezystancji R_1, R_2 zapewniających w pewnym zakresie w przybliżeniu stałe pasmo przy zmianie wzmocnienia.

3. ANALIZA TRANSMITANCJI NIEODWRACAJĄCEGO WZMACNIACZA NAPIĘCIOWEGO Z PRĄDOWYM SPRZĘŻENIEM ZWROTNYM

Wzmacniacz napięciowy zbudowany w oparciu o układ ze sprzężeniem prądowym można przedstawić w postaci schematu blokowego z jednopętlowym sprzężeniem zwrotnym (rys. 3).



Rys. 3. Schemat blokowy nieodwracającego wzmacniacza napięciowego z prądowym sprzężeniem zwrotnym

Transmitancje występujące na rys. 3 reprezentują:

$$K = Z_T \tag{3.1}$$

jest transimpedancją wzmacniacza objętego pętlą sprzężenia zwrotnego.

$$T_{BA} = -\left.\frac{I_x}{V_B}\right|_{V_Y = 0} \tag{3.2}$$

określa wpływ napięcia wyjściowego pętli sprzężenia zwrotnego $V_B = I_X Z_T$ na prąd wejściowy pętli.

$$T_{YA} = \left.\frac{I_X}{V_Y}\right|_{K=0} \tag{3.3}$$

opisuje wpływ napięcia wejściowego V_Y na prąd I_X przy rozwartej pętli sprzężenia zwrotnego.

$$T_{B0} = \left.\frac{V_0}{V_B}\right|_{V_Y = 0} \tag{3.4}$$

określa wpływ napięcia wyjściowego pętli sprzężenia zwrotnego $V_B = Z_T I_X$ na napięcie wyjsciowe V_0 przy braku oddziaływania napięcia V_Y na napięcie V_0 przez układ pasywny.

$$T_{Y0} = \left.\frac{V_0}{V_Y}\right|_{K=0} \tag{3.5}$$

wyraża bezpośredni wpływ napięcia wejściowego V_Y, występujący poza pętlą sprzężenia zwrotnego, przez układ pasywny na napięcie wyjściowe V_0.

Wzmocnienie pętli otwartej wynosi

$$LG = KT_{BA}. \tag{3.6}$$

Wzmocnienie napięciowe A układu ze sprzężeniem zwrotnym, wyznaczone na podstawie schematu blokowego z rys. 3, wynosi

$$A = \frac{V_0}{V_Y} = T_{YA}\frac{K}{1 + KT_{BA}}T_{B0} + T_{Y0}. \tag{3.7}$$

3.1. WZMACNIACZ IDEALNY ZE STOPNIEM TRANSIMPEDANCYJNYM

Na rys. 4 przedstawiono schemat blokowy idealnego wzmacniacza CFOA nieobciążonego, otrzymany na podstawie schematu zastępczego układu ze stopniem transimpedancyjnym z rys. 1, przy założeniu: $r_x = 0$, $R_{02} = 0$, $k_1 = k_2 = k_i = 1$.

Wzmocnienie napięciowe analizowanego układu wynosi

$$A = \left(1 + \frac{R_2}{R_1}\right)\frac{1}{1 + \dfrac{1}{LG}} = \left(1 + \frac{R_2}{R_1}\right)\frac{1}{1 + \dfrac{R_2}{Z_T}}. \tag{3.8}$$

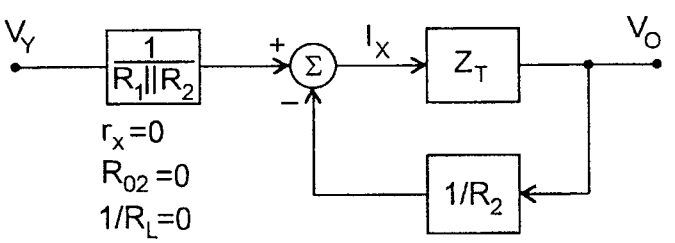
Można przyjąć, że transimpedancja Z_T jest funkcją dwubiegunową

$$Z_T(s) = \frac{R_T}{(1 + s\tau_T)\ (1 + s\tau_{cm})} \tag{3.9}$$

o biegunie dominującym $1/\tau_T$ $(\tau_T \gg \tau_{cm})$; $\tau_T = R_T C_T$. (3.10)

Lustra prądowe lub układ kaskadowy w stopniu transimpedancyjnym są przyczyną istnienia drugiego bieguna o pulsacji $1/\tau_{cm}$. Wykorzystując zależności (3.9, 3.10) oraz zakładając dużą wartość transrezystancji $R_T (R_T \gg R_2)$, otrzymujemy

$$A = \left(1 + \frac{R_2}{R_1}\right)\frac{\dfrac{R_T}{\tau_T \tau_{cm} R_2}}{s^2 + s\dfrac{\tau_T + \tau_{cm}}{\tau_t \tau_{cm}} + \dfrac{R_T}{R_2 \tau_T \tau_{cm}}}. \tag{3.11}$$



Rys. 4. Schemat blokowy idealnego wzmacniacza CFOA ze stopniem transimpedancyjnym

Pomijając wpływ bieguna $1/\tau_{cm}$ ($\tau_{cm} = 0$), wzmocnienie napięciowe i 3dB szerokość pasma określają zależości:

$$A = \left(1 + \frac{R_2}{R_1}\right) \frac{1}{1 + s\tau_T \frac{R_2}{R_T}} = \left(1 + \frac{R_2}{R_1}\right) \frac{1}{1 + sR_2 C_T} \tag{3.12}$$

$$\omega_{3dB} = \frac{1}{R_2 C_T}. \tag{3.13}$$

W przypadku, gdy biegun o pulsacji $1/\tau_T$ jest dominujący ($\tau_T \gg \tau_{cm}$), to:

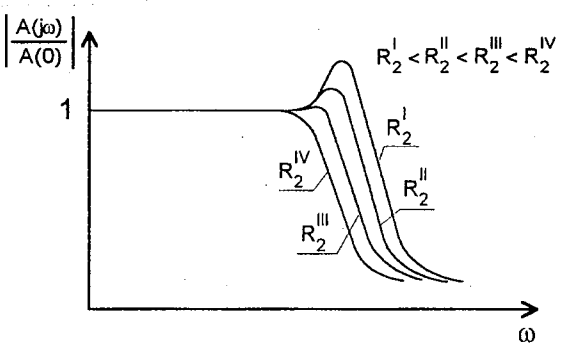
$$A = \left(1 + \frac{R_2}{R_1}\right) \frac{\omega_p^2}{s^2 + \frac{\omega_p}{Q} s + \omega_p^2}, \tag{3.14}$$

gdzie

$$\omega_p = \frac{1}{\sqrt{C_T R_2 \tau_{cm}}}, \tag{3.15}$$

$$Q = \sqrt{\frac{\tau_{cm}}{C_T R_2}}. \tag{3.16}$$

Znormalizowaną charakterystykę amplitudową wzmacniacza, wyznaczoną na podstawie zależności (3.14), przedstawiono na rys. 5.



Rys. 5. Znormalizowana charakterystyka amplitudowa wzmocnienia napięciowego

$$\left|\frac{A(j\omega)}{A(0)}\right| = \left|\frac{\omega_p^2}{(j\omega)^2 + \frac{\omega_p}{Q} j\omega + \omega_p^2}\right|$$

Dla rozważanego przypadku 3dB szerokość pasma wynosi

$$\omega_{3dB} = \omega_p \sqrt{\frac{\left(2 - \frac{1}{Q^2}\right) + \sqrt{\left(2 - \frac{1}{Q^2}\right)^2 + 4}}{2}}. \tag{3.17}$$

Zmniejszenie rezystancji R_2 pozwala zwiększyć pasmo wzmacniacza, jednak może to spowodować pojawienie się niepożądanego podbicia charakterystyki. Podbicie to nie występuje, gdy

$$Q \leqslant \frac{1}{\sqrt{2}}. \tag{3.18}$$

Zatem minimalna wartość rezystancji R_2, nie powodująca podbicia charakterystyki i pozwalająca uzyskać maksymalne pasmo, wynosi

$$R_2 = \frac{2\tau_{cm}}{C_T}. \tag{3.19}$$

Wtedy

$$\omega_{3\mathrm{dB}}\left(R_2 = \frac{2\tau_{cm}}{C_T}\right) = \omega_p\left(R_2 = \frac{2\tau_{cm}}{C_T}\right) = \frac{1}{\sqrt{2}\,\tau_{cm}}. \tag{3.20}$$

Margines fazy można określić badając charakterystykę fazową wzmocnienia pętli otwartej $LG = Z_T/R_2$. Korzystając z zależności (3.9) oraz poprzednio stosowanych przybliżeń ($\tau_T \gg \tau_{cm}$, $R_T \gg R_2$) możemy obliczyć pulsację, przy której moduł wzmocnienia pętli otwartej $|LG|$ jest równy jedności

$$\omega|_{|LG|=1} = \frac{1}{\tau_{cm}} \sqrt{\frac{-1 + \sqrt{1 + \left(\frac{2\tau_{cm}}{C_T R_2}\right)^2}}{2}} \tag{3.21}$$

a następnie margines fazy

$$\varphi_M = arc\,tg \frac{1}{\omega|_{|LG|=1}\,\tau_{cm}}. \tag{3.22}$$

Dla R_2 określonego zależnością (3.19), margines fazy $\varphi_M = 66°$.

3.1.1. Kompensacja bieguna dominującego charakterystyki amplitudowej

Zwiększenie szerokości pasma wzmacniacza można uzyskać kompensując biegun dominujący charakterystyki amplitudowej. Kompensację taką można zrealizować przez wprowadzenie w obwodzie sprzężenia zwrotnego pojemności C_1, C_2, połączonych równolegle odpowiednio do rezystancji R_1, R_2. Wtedy

$$Z_1 = \frac{R_1}{1 + s\tau_1}, \qquad \tau_1 = R_1 C_1 \tag{3.23}$$

$$Z_2 = \frac{R_2}{1 + s\tau_2}, \qquad \tau_2 = R_2 C_2. \tag{3.24}$$

Zakładając, że spełniony jest warunek

$$\tau_1 = \tau_2 = \tau_T \quad \left(C_1 = \frac{R_T}{R_1} C_T,\ C_2 = \frac{R_T}{R_2} C_T\right) \tag{3.25}$$

wzmocnienie napięciowe określone jest zależnością

$$A = \left(1 + \frac{Z_2}{Z_1}\right) \frac{1}{1 + \frac{1}{LG}} = \left(1 + \frac{Z_2}{Z_1}\right) \frac{1}{1 + \frac{Z_2}{Z_T}} \tag{3.26}$$

i zakładając $R_T \gg R_2$, otrzymujemy

$$A = \left(1 + \frac{R_2}{R_1}\right) \frac{1}{1 + s\frac{R_2}{R_T}\tau_{cm}} \tag{3.27}$$

$$\omega_{3dB} = \frac{R_T}{R_2 \tau_{cm}}. \tag{3.28}$$

Pulsacja ω_{3dB} w układzie z kompensacją (3.28) jest większa aniżeli w układzie o płaskiej charakterystyce bez kompensacji (3.20). Ponadto w układzie z kompensacja istnieje możliwość doboru pasma przez zmianę rezystancji R_2, gdy spełniony jest warunek $R_2 \ll R_T$.

Margines fazy wzmacniacza o charakterystyce danej zależnością (3.27) wynosi

$$\varnothing_M = \pi + arctg\,(-\omega|_{|LG|=1}\,\tau_{cm}), \tag{3.29}$$

gdzie:

$$\omega|_{|LG|=1} = \frac{R_T}{R_2 \tau_{cm}}. \tag{3.30}$$

Dla $R_T \gg R_2$ margines fazy wynosi w przybliżeniu 90°.

Margines fazy można zwiększyć przez zwiększenie rezystancji $R_2 (R_2 \ll R_T)$ jednak kosztem pasma wzmacniacza.

3.2. UKŁAD NIEIDEALNY

W tej części pracy przeanalizujemy wpływ rezystancji r_x i R_{02} oraz obciążenia R_L wzmacniacza z rys. 1 na transmitancję układu. Wykorzystując schemat zastępczy z rys. 1 oraz zależności (3.1 ÷ 3.5), poszczególne transmitancje schematu blokowego z rys. 3, określają zależności:

$$K = Z_T$$

$$T_{BA} = \frac{[R_2 + (r_x \| R_1)] \| R_L}{R_{02} + \{[R_2 + (r_x \| R_1)] \| R_L\}} \frac{r_x \| R_1}{R_2 + (r_x \| R_1)} \frac{1}{r_x} = \frac{R_L R_1}{M}, \tag{3.31}$$

gdzie

$$M = (R_{02} + R_L)[R_2(r_x + R_1) + r_x r_1] + R_{02} R_L (r_x + R_1) \tag{3.32}$$

$$T_{YA} = \frac{1}{r_x + \{R_1 \parallel [R_2 + (R_{02} \parallel R_L)]\}} = \frac{(R_1 + R_2)(R_{02} + R_L) + R_{02} R_L}{M} \tag{3.33}$$

$$T_{B0} = \frac{[R_2 + (r_x \parallel R_1)] \parallel R_L}{R_{02} + \{[R_2 + (r_x \parallel R_1)] \parallel R_L\}} = \frac{[R_2 + (r_x + R_1) + r_x R_1] R_L}{M} \tag{3.34}$$

$$T_{Y0} = \frac{R_1 \parallel [R_2 + (R_{02} \parallel R_L)]}{r_x + \{R_1 \parallel [R_2 + (R_{02} \parallel R_L)]\}} \frac{R_{02} \parallel R_L}{R_2 + (R_{02} \parallel R_L)} = \frac{R_1 R_{02} R_L}{M}. \tag{3.35}$$

Wykorzystując zależności (3.6, 3.7, 3.31÷3.35), otrzymujemy wzmocnienie napięciowe

$$A = \frac{Z_T(R_1 + R_2) + R_{02} R_1}{\left(1 + \frac{R_{02}}{R_L}\right)[R_2(r_x + R_1) + r_x R_1] + R_{02}(r_x + R_1) + Z_T R_1} = \tag{3.36}$$

$$= \frac{\left(1 + \frac{R_2}{R_1}\right) + \frac{R_{02}}{Z_T}}{1 + \frac{1}{LG}},$$

gdzie

$$LG = \frac{Z_T}{\left(1 + \frac{R_{02}}{R_L}\right)\left[R_2\left(1 + \frac{r_x}{R_1}\right) + r_x\right] + R_{02}\left(1 + \frac{r_x}{R_1}\right)} = \frac{Z_T}{R_2'} \tag{3.37}$$

$$R_2' = \left(1 + \frac{R_{02}}{R_L}\right)\left[R_2\left(1 + \frac{r_x}{R_1}\right) + r_x\right] + R_{02}\left(1 + \frac{r_x}{R_1}\right). \tag{3.38}$$

Zakładając $R_{02} \ll R_T$, zależność (3.36) można przedstawić w postaci

$$A = \left(1 + \frac{R_2}{R_1}\right)\frac{1}{1 + \frac{1}{LG}}, \quad LG = \frac{Z_T}{R_2'}. \tag{3.39}$$

Wzmocnienie układu rzeczywistego (3.39) i wzmocnienie układu idealnego (3.8) opisują funkcje o tej samej postaci. Różnica polega na zastąpieniu rezystancji R_2 rezystancją R_2' w członie częstotliwościwym funkcji. Zatem maksymalna pulsację ω_{3dB} wzmacniacza o charakterystyce amplitudowej bez podbicia określa zależność

$$\omega_{3dB}\left(R_2' = \frac{2\tau_{cm}}{C_T}\right) = \frac{1}{\sqrt{2}\,\tau_{cm}}, \tag{3.40}$$

wtedy rezystancja R_2 wynosi

$$R_2 = \left(\frac{2\frac{\tau_{cm}}{\tau_T} R_T - R_{02}\left(1 + \frac{r_x}{R_1}\right)}{1 + \frac{R_{02}}{R_L}} - r_x \right) \frac{1}{1 + \frac{r_x}{R_1}}. \tag{3.41}$$

W przypadku, gdy wzmacniacz jest obciążony tylko pojemnością C_L wzmocnienie układu dane jest dalej zależnością (3.36), a wzmocnienie pętli określa wzór

$$LG = \frac{Z_T}{(1 + sR_{02}C_L)\left[R_2\left(1 + \frac{r_x}{R_1}\right) + r_x\right] + R_{02}\left(1 + \frac{r_x}{R_1}\right)}. \tag{3.42}$$

Korzystając z zależności (3.9) i (3.42) oraz zakładając, $r_x \ll R_1$, R_2; $R_{02} \ll R_2$, $\tau^{cm} \ll \tau_T$, otrzymujemy przybliżoną zależność wzmocnienia pętli

$$LG = \frac{R_T}{(1 + sR_{02}C_L)(1 + s\tau_T)R_2}. \tag{3.43}$$

Załóżmy, że biegun o pulsacji $\frac{1}{R_{02}C_L}$ jest dominujący, tzn. $R_{02}C_L \gg \tau_T$. Wtedy pulsacja ω_{3dB} transmitancji wzmacniacza wynosi

$$\omega_{3dB} = \frac{R_T}{R_2 R_{02} C_L}. \tag{3.44}$$

Korzystając z zależności (3.36) i (3.43) ($R_{02} \ll R_2 \ll R_T$) wyznaczymy margines fazy

$$\omega|_{|LG|=1} = \frac{1}{\tau_T}\sqrt{\frac{-1 + \sqrt{1 + \left(\frac{2\tau_T}{R_{02}C_L}\frac{R_T}{R_2}\right)^2}}{2}} \tag{3.45}$$

$$\varphi_M = arctg \frac{1}{\omega|_{|LG|=1}\tau_T}. \tag{3.46}$$

Kompensację bieguna dominującego można przeprowadzić poprzez odpowiedni dobór stałych czasowych τ_1, τ_2

$$\tau_1 = \tau_2 = R_{02}C_L, \left(C_1 = \frac{R_{02}}{R_1}C_L,\ C_2 = \frac{R_{02}}{R_2}C_L\right). \tag{3.47}$$

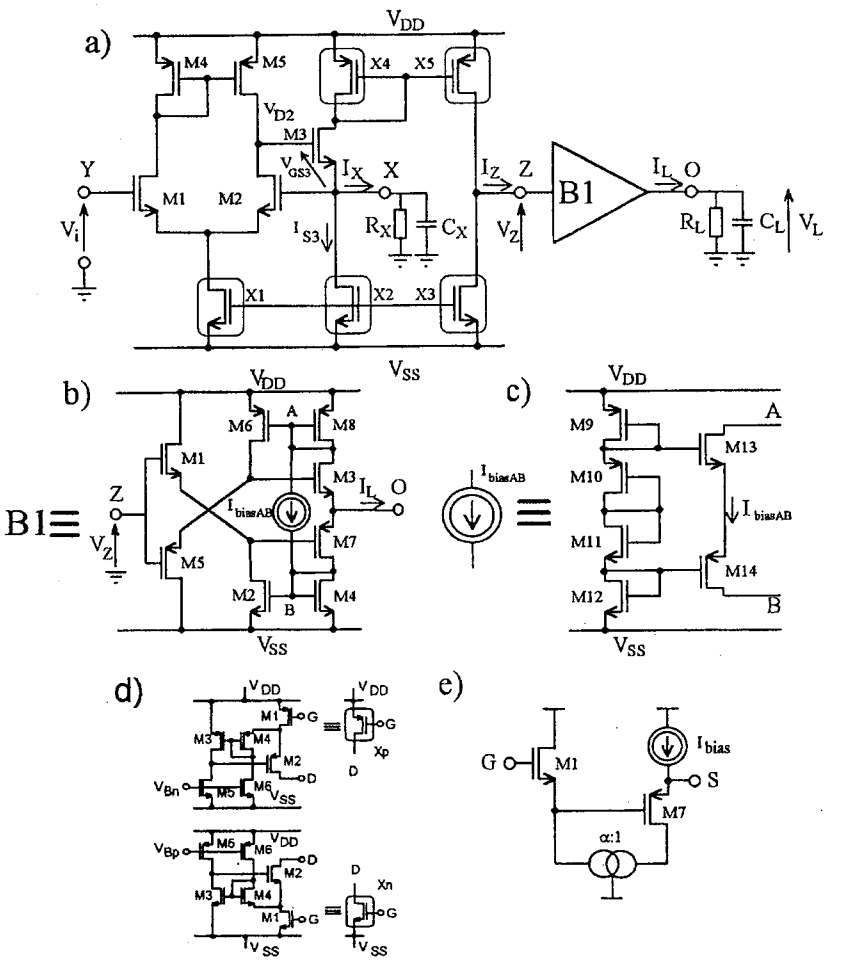
Pulsacja ω_{3dB} wzmacniacza ze skompensowanym wpływem bieguna dominującego wynosi

$$\omega_{3dB} = \frac{R_T}{R_2}\frac{1}{\tau_T}, \tag{3.48}$$

co wobec warunku $R_{02} C_L \gg \tau_T$, pozwala na znaczne rozszerzenie pasma z możliwością jego zmiany przez dobór rezystancji R_2. Kompensacja zapewnia zwiększenie marginesu fazy w przybliżeniu do 90°.

4. OPIS ROZWIĄZAŃ UKŁADOWYCH WZMACNIACZY CFOA's W TECHNICE CMOS

Schemat ideowy wzmacniacza CFOA ze stopniem transimpedancyjnym przedstawiono na rys. 6a. Stopień wejściowy wzmacniacza wraz z układem transimpedan-



Rys. 6. Schemat ideowy wzmacniacza CFOA z nieodwracającym stopniem transimpedancyjnym (a); schemat ideowy nieodwracającego wyjściowego bufora napięciowego (b); realizacja źródła polaryzacji wstępnej I_{biasAB} (c); schematy ideowe tranzystorów złożonych NMOS i PMOS (d); schemat ideowy złożonego wtórnika źródłowego z lokalnym prądowym sprzężeniem zwrotnym (e)

cyjnym jest zrealizowany przy wykorzystaniu dodatniego konwejora prądowego CCII$^+$. Wtórnik napięciowy, powtarzający napięcie z wejścia Y na niskoimpedancyjne wejście X, tworzy stopień różnicowy z tranzystorami M1, M2 wraz z wtórnikiem źródłowym na tranzystorze M3, zamknięty w pętlę stuprocentowego ujemnego sprzężenia zwrotnego. W celu zapewnienia bardzo małego napięcia niezrównoważenia wejściowego wtórnika napięciowego prąd I_{s3} polaryzujący wtórnik źródłowy na tranzystorze M3 należy tak dobrać, aby przy otwartej pętli sprzężenia zwrotnego (bramka tranzystora M2 uziemiona) napięcie na wyjściu X był orówne zeru: $V_x = V_{D2} - V_{GS3}(I_{S3}) \cong 0$.

Stopień transimpedancyjny jest zrealizowany przy użyciu kaskadowych regulowanych luster prądowych X2, X3 oraz X4, X5. Bardzo dogodną formą przedstawiania schematów ideowych kaskadowych regulowanych luster prądowych jest użycie symboli tranzystorów złożonych NMOS i PMOS (ang. *composite transistors* [6]), przedstawionych na rys. 6d. Przy użyciu tranzystorów złożonych są zrealizowane również wszystkie źródła prądowe w układzie na rys. 6a (X1, X2). Nieodwracający, wyjściowy bufor napięciowy B1 (rys. 6b) jest zrealizowany jako przeciwsobny układ w klasie AB o małej impedancji wyjściowej. Jego struktura oparta jest o znaną z literatury koncepcję [7] kaskadowego połączenia dwóch wtórników źródłowych z komplementarnymi tranzystorami MOS i z lokalnym prądowym sprzężeniem zwrotnym. Jak w uproszczeniu pokazano na rys. 6e, wtórnik źródłowy z tranzystorem M1 typu NMOS steruje wtórnikiem źródłowym z tranzystorem M7 typu PMOS, a prądy ich polaryzacji są sprzężone zwrotnie za pośrednictwem lustra prądowego o przekładni $\alpha:1$. Przy założeniu najprostszych modeli tranzystorów i pominięciu ich konduktancji wyjściowych, układowi z rys. 6e odpowiada równoważny wtórnik źródłowy z tranzystorem „złożonym" o zastępczej konduktancji bramkowej

$$g_m = \frac{g_{m1}\, g_{m7}}{g_{m1} - \alpha g_{m7}}, \tag{4.1}$$

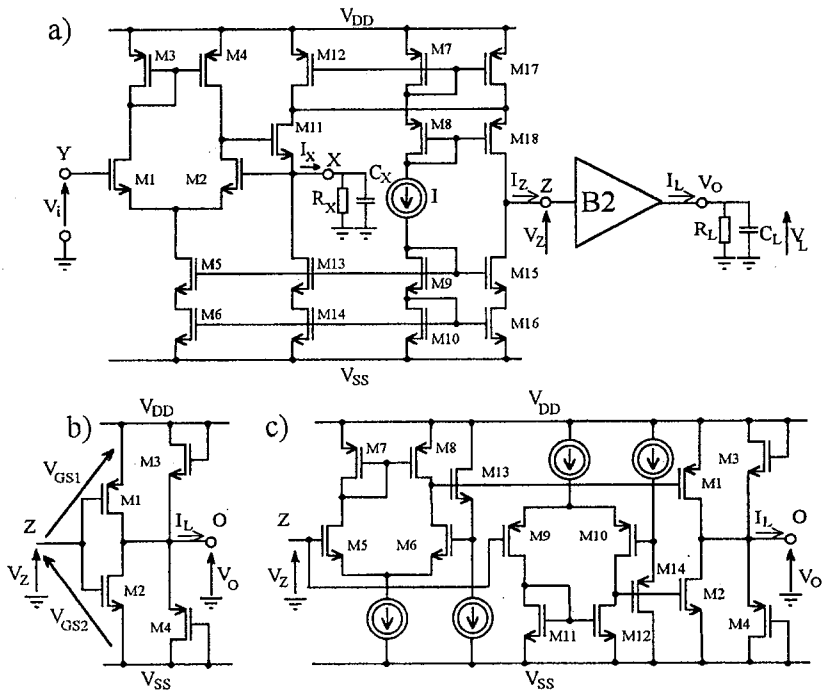
gdzie

$$\alpha = \frac{g_{m1}}{g_{m2}}. \tag{4.2}$$

Duża wartość konduktancji zastępczej g_m pozwala na zapewnienie małej wartości rezystancji wyjściowej bufora B1, a praca w konfiguracji OD zapewnia szerokie pasmo przenoszonych częstotliwości.

Wstępną polaryzację tranzystorów M1, M5 oraz M3, M7 do pracy w klasie AB zapewnia źródło prądowe I_{biasAB}, którego schemat ideowy przedstawia rys. 6c.

Alternatywnym rozwiązaniem układowym wzmacniacza CFOA ze stopniem transimpedancyjnym jest użycie odwracającego konwejora CCII$^-$ i odwracającego wyjściowego bufora napięciowego. Schemat ideowy takiego rozwiązania wzmacniacza jest przedstawiony na rys. 7.



Rys. 7. Schemat ideowy wzmacniacza CFOA z odwracającym stopniem transimpedancyjnym (a); schemat ideowy odwracającego wyjściowego bufora napięciowego w klasie A (b); schemat ideowy odwracającego wyjściowego bufora napięciowego w klasie B (c)

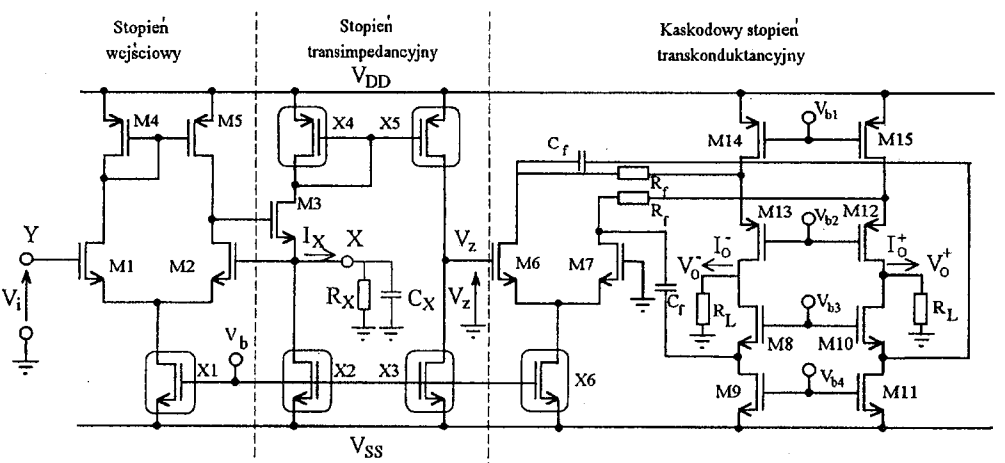
W układzie tym stopień wejściowy wzmacniacza zbudowany jest podobnie jak w układzie na rys. 6a. Odwracający stopień transimpedancyjny zrealizowany jest w oparciu o wzmacniacz kaskodowy, przy wykorzystaniu tranzystorów M11÷M16. Źródło prądowe I_{bias} oraz tranzystory M7÷M10 w połączeniu diodowym wytwarzaja potrzebne napięcia odniesienia do polaryzacji bramek tranzystorow wzmacniacza kaskodowego. Zastosowanie wzmacniacza kaskodowego jako transimpedancyjnego stopnia odwracającego jest lepszym rozwiązaniem niż stopień zbudowany z samych luster prądowych (z dodatkową parą skrzyżowanych luster prądowych, w porównaniu do rys. 6a, w celu zmiany znaku współczynnika transmisji prądowej), ponieważ zapewnia szersze pasmo częstotliwości transmisji prądowej z wejścia X na wyjście Z, porównywalne do układu z rys. 6a.

Odwracający, wyjściowy bufor napięciowy może być zrealizowany jak na rys. 7b w postaci komplementarnego układu ze wspólnymi źródłami na tranzystorach M1, M2 i aktywnym obciążeniem parą komplementarnych tranzystorów M3, M4. Układ zapewnia małą rezystancję wyjściową

$$r_0 = (g_{m3} + g_{bs3} + g_{ds3} + g_{m4} + g_{bs4} + g_{ds4} + g_{ds1} + g_{ds2})^{-1} \cong$$
$$\cong (g_{m3} + g_{bs3} + g_{m4} + g_{bs4})^{-1} \tag{4.3}$$

oraz dużą liniowość charakterystyki przejściowej $V_0 = f(V_z)$.
Układ z rys. 7b charakteryzuje się dużym poborem prądu, bowiem tranzystory M1, M2 pracują w klasie A przy dużych spoczynkowych napięciach V_{GS}: $V_{GS2} = V_{DD}$, $V_{GS1} = V_{SS}$ (przy $V_z = 0$) i dlatego ma ograniczone zastosowanie. Sterowanie tranzystorów M1, M2 za pośrednictwem komplementarnych wzmacniaczy różnicowych o niewielkich wzmocnieniach, z tranzystorami M5, M6 i M9, M10 (rys. 7c), pozwala na ich pracę w klasie B lub AB, zapewniając mniejszy pobór prądu.

Schemat ideowy wzmacniacza CFOA z prądowym źródłem sterowanym, zrealizowanym w postaci dwustopniowego układu jako transimpedancyjny stopień sterujący stopniem transkonduktancyjnym, przedstawiono na rys. 8.
Stopień wejściowy wzmacniacza i stopień transimpedancyjny zrealizowany jest w postaci dodatniego konwejora prądowego CCII$^+$, podobnie jak opisany wcześniej układ na rys. 6a.



Rys. 8. Schemat ideowy wzmacniacza CFOA ze sterowanym źródłem prądowym

Stopień różnicowy na tranzystorach M6, M7, wraz ze wzmacniaczem kaskodowym na tranzystorach M8÷M16 fromują kaskodowy stopień transkonduktancyjny o różnicowych wyjściach. Zastosowanie na wyjściu różnicowo sterowanych struktur kaskodowych zapewnia duże rezystancje wyjściowe, bez konieczności stosowania luster prądowych. W układzie zastosowano kompensację charakterystyk częstotliwościowych (ang. *feedforward compensation*) za pomocą elementów R_f, C_f.

Na podstawie małosygnałowego schematu zastępczego stopnia transkonduktancyjnego możemy wyznaczyć transkonduktancję układu:

$$G_t = \frac{-I_0^+}{V_z} = \frac{g_{m6}}{2} \frac{G_L}{\frac{(g_{ds12} + G_L)(g_{ds6} + g_{ds15})}{g_{m12} + g_{mb12}} + g_{c10,11} + G_L}, \tag{4.4}$$

gdzie konduktancja wyjściowa kaskadowego połączenia tranzystorów M10, M11, wynosi

$$g_{c10,11} = \frac{g_{ds10} g_{ds11}}{g_{ds10} + g_{m10} + g_{mb10} + g_{ds11}} \cong \frac{g_{ds10} g_{ds11}}{g_{m10} + g_{mb10}}. \tag{4.5}$$

Normalnie spełnione są warunki

$$G_L \gg g_{ds12},\ G_L \gg g_{c10,11},\ g_{ds6} + g_{ds15} \ll g_{m12} + g_{mb12}. \tag{4.6}$$

Dlatego

$$G_t = \frac{-I_0^+}{V_z} \cong \frac{g_{m6}}{2}\left(\frac{g_{ds} + g_{ds15}}{g_{m12} + g_{mb12}} + 1\right)^{-1} \cong \frac{g_{m6}}{2}. \tag{4.7}$$

Konduktancja wyjściowa każdego z wyjść różnicowych stopni kaskodowych wynosi:

$$G_0 = -\left.\frac{I_0^+}{V_z^+}\right|_{V_z = 0} = \frac{(g_{ds6} + g_{ds15}) g_{ds12}}{g_{ds12} + g_{m12} + g_{mb12} + g_{ds6} + g_{ds15}} \cong \frac{(g_{ds6} + g_{ds15}) g_{ds12}}{g_{m12} + g_{mb12}}. \tag{4.8}$$

Jak wynika ze wzorów (4.4 ÷ 4.8) transkonduktancja G_t jest prawie stała, niezależna od konduktancji obciążenia, a stopnie wyjściowe zachowują się jak sterowane źródła prądu o bardzo małej konduktancji wyjściowej.

5. WYNIKI SYMULACJI I WNIOSKI KOŃCOWE

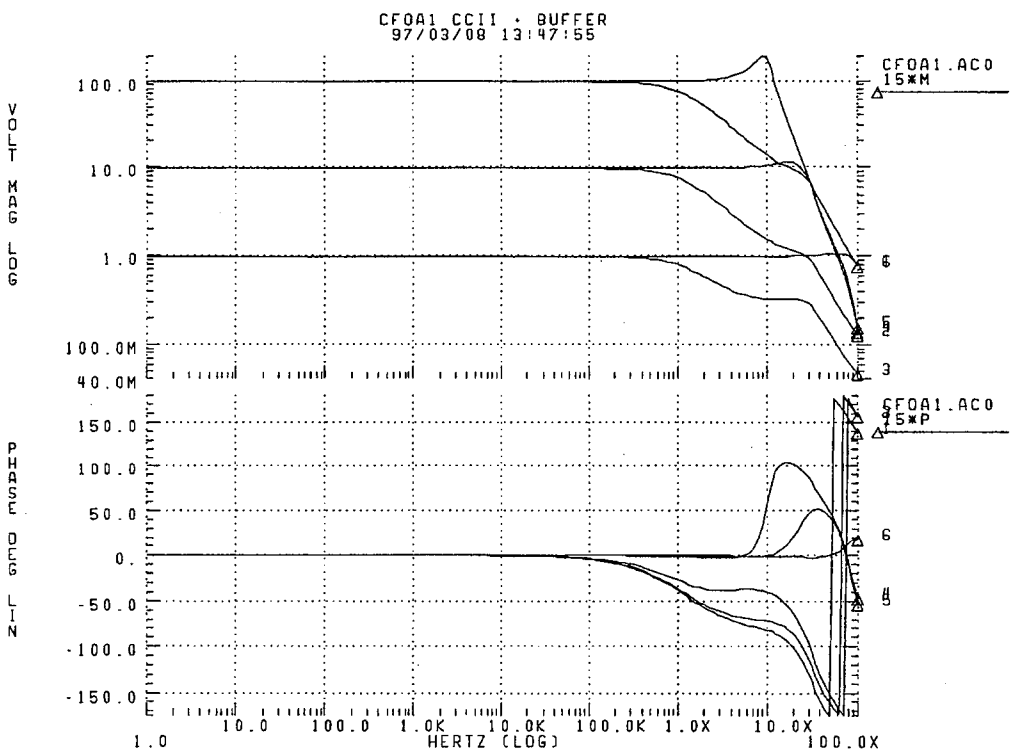
Wyprowadzone zależności w punktach 2 i 3 pozwalają na dobór parametrów zewnętrznego obwodu sprzężenia zwrotnego rzeczywistego zmacniacza CFOA, zapewniających jego maksymalne pasmo. Bardzo efektywnym sposobem poszerzenia pasma wzmacniacza CFOA jest metoda kompensacji dominującego bieguna charakterystyki amplitudowej za pomocą pojemnosci C_1, C_2 połączonych rownolegle do rezystancji R_1, R_2, jednak tylko w ogranicoznym zakresie wielkości wzmocnienia i pasma.

Kompensacja charakterystyki amplitudowej o dużym wzmocnieniu i szerokim pasmie wymaga, zgodnie z warunkiem kompensacji (3,25), stosowania dość dużych wartości pojemnosci C_1, C_2. W następstwie dość znacznego pojemnościowego obciążenia wzmacniacza elementami w torze sprzężenia zwrotnego pole wzmocnienia układu otwartego maleje i przy zamknięciu pętli sprzężenia zwrotnego z kompensacją górna częstotliwość graniczna wzmacniacza może być mniejsza niż układu bez kompensacji. Niżej zostaną przedstawione przykłady kształtowania charakterystyk częstotliwościowych wzmacniaczy CFOAs opisanych w punkcie 4

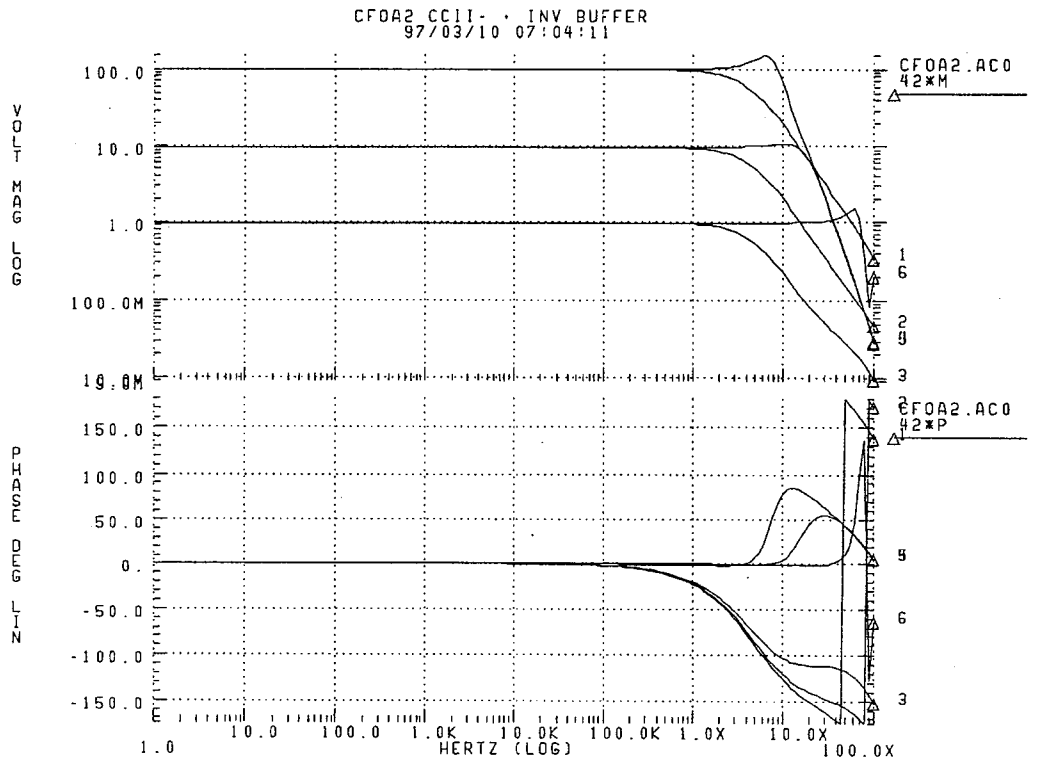
(rys. 6÷8), zaprojektowanych w technologii CMOS 1.2 μA z firmy AMS. Symulacje zostały przeprowadzone w programie HSPICE i modelach tranzystorów na poziomie LEVEL=2.

Na rys. 9 przedstawiono charakterystyki amplitudowe i fazowe bez kompensacji i z kompensacją wzmacniacza ze stopniem transrezystancyjnym o strukturze układowej przedstawionej na rys. 6. W układzie bez kompensacji dla wszystkich trzech wartości wzmocnienia $A_0 = 1$, 10, 100 przez odpowiedni dobór wartości rezystancji R_2, zapewniono prawie tę samą górną częstotliwość graniczną ok 1MHz. W układzie z kompensacją następuje znaczne rozszerzenie pasma (100MHz dla $A_0 = 1$, 30MHz dla $A_0 = 10$, 20MHz dla $A_0 = 100$), jednak z powodów opisanych wyżej, pasmo maleje ze wzrostem wzmocnienia, a oprócz tego, przy dużej wartości wzmocnienia występuje podbicie charakterystyki amplitudowej.

Charakterystyki częstotliwościowe dwóch pozostałych wzmacniaczy, opisanych w punkcie 4 (układy z rys. 7 i 8) przedstawiono na rys. 10 i 11. Z porównania tych



Rys. 9. Charakterystyki amplitudowe i fazowe bez kompensacji $A_0=1$ ($R_1\to\infty$, $R_2=55\,k\Omega$), $A_0=10$ ($R_1=6{,}1\,k\Omega$, $R_2=55\,k\Omega$), $A_0=100$ ($R_1=550\,k\Omega$, $R_2=55\,k\Omega$) i z kompensacją $A_0=1$($C_1=0$, $C_2=110\,pF$), $A_0=10$ ($C_1=1\,nF$, $C_2=110\,pF$), $A_0=100$ ($C_1=11\,nF$, $C_2=110\,pF$) wzmacniacza z rys. 6



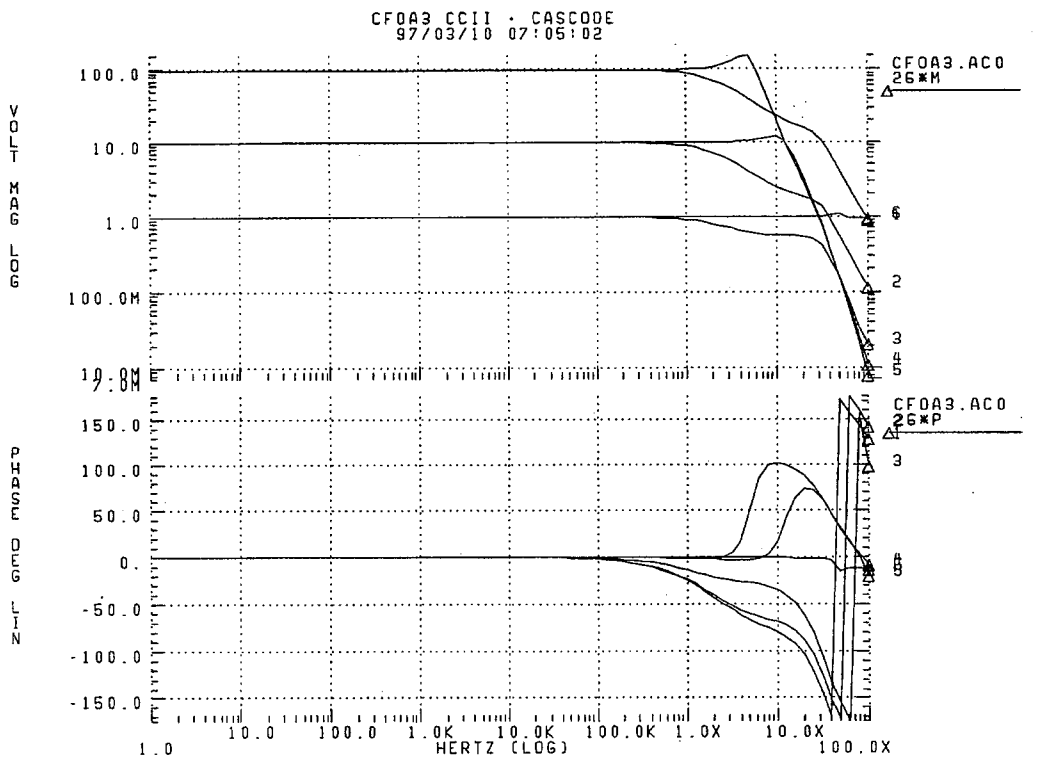
Rys. 10. Charakterystyki amplitudowe i fazowe bez kompensacji i z kompensacją wzmacniacza (wartości elementów jak na rys. 9)

charakterystyk wynika, że prezentowane układy charakteryzują się podobnymi właściwościami częstotliwościowymi.

Możemy również stwierdzić, że kompensacja dominującego bieguna charakterystyki układu otwartego ma ograniczone zastosowanie ze względu na konieczność stosowania zbyt dużych wartości pojemnosci C_1, C_2. Przeprowadzona analiza w punkcie 3 pozwala ocenić przydatność nieidealnych wzmacniaczy CFOAs do kształtowania dowolnych transmitancji (np. realizacji filtrów ze wzmacniaczami CFOAs). Jeżeli w obwodzie sprzężenia zwrotnego zastosujemy $C_2 \neq 0$, zaś $C_1 = 0$, to podstawiając w równaniu (3.37) w miejsce R_2 impedancję Z_2, (określoną równaniem (3.24)) możemy stwierdzić, że transmitancja układu otwartego posiada dodatkowe zero przy pulsacji $\omega_z = 1/R_2 C_2$ oraz dodatkowy biegun przy pulsacji $\omega_p = 1/\tau_p$, gdzie

$$\tau_p = C_2 \frac{R_2}{R_2'}\left[\left(1 + \frac{R_{02}}{R_L}\right) r_x + \left(1 + \frac{r_x}{R_1}\right) R_{02}\right] \tag{5.1}$$

zaś R_2' określone jest wzorem (3.38).

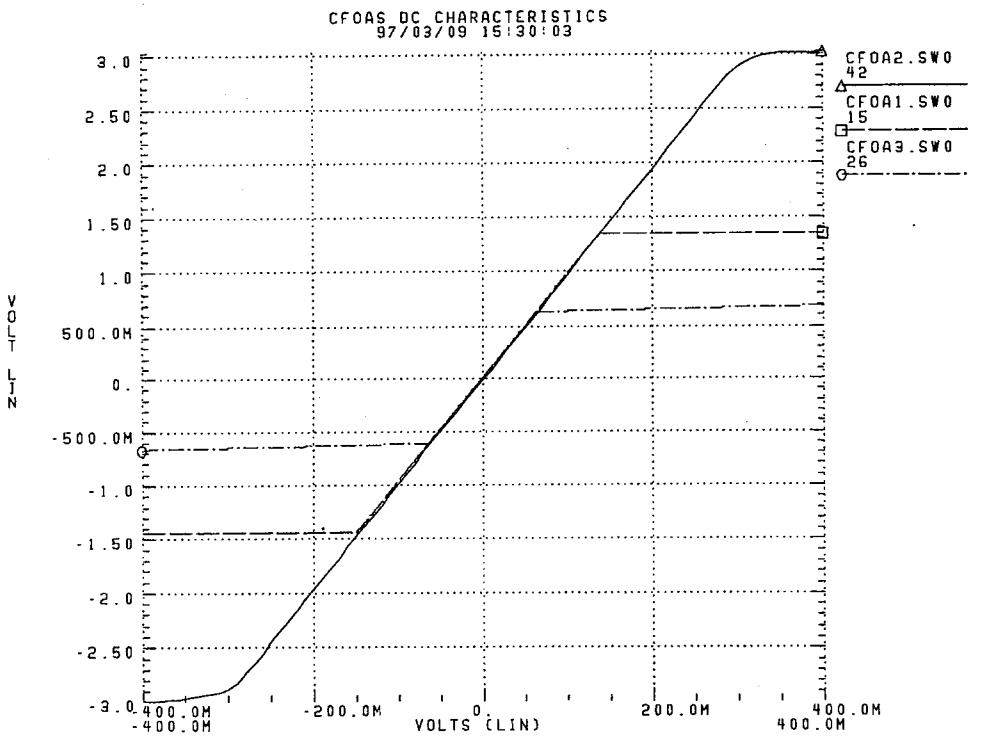


Rys. 11. Charakterystyki amplitudowe i fazowe bez kompensacji i z kompensacją wzmacniacza z rys. 8 (wartości elementów jak na rys. 9)

Ponieważ $\omega_p > \omega_z$, to pole wzmocnienia układu otwartego rośnie, jednak pulsacja ω_c przy której $|LG| = 1$ zostaje przesunięta w obszar znacznych przesunięć fazy spowodowanych biegunami wyższego rzędu transimpedancji $Z_T(s)$. Może to prowadzić nawet do utraty stabilności wzmacniacza.

Przy zastosowaniu $C_1 \neq 0$ oraz $C_2 = 0$, podobne rozwiązanie (w równaniu 3.37) w miejsce R_1 należy podstawić impedancję Z_1 określoną równaniem (3.23)) prowadzi do stwierdzenia, że transmitancja układu otwartego posiada dodatkowy biegun przy pulsacji $\omega_p = 1/\tau_p$, gdzie $\tau_p = C_1[R_1 \| r_x \| (R_2 + R_{02})]$. Jeżeli pojemność C_1 jest na tyle duża, że dodatkowy biegun wnosi przesunięcie fazowe w obszarze pulsacji przy której $|LG| > 1$, to zmniejsza to margines fazy i może być przyczyną niestabilności lub znacznego podbicia charakterystyki amplitudowej układu z zamkniętą pętlą sprzężenia zwrotnego.

Na rys. 12 przedstawiono charakterystyki przejściowe $V_0 = f(V_{in})$ wszystkich trzech wzmacniaczy, przy czym we wzmacniaczu z odwracającym stopniem transimpedancyjnym zastosowano odwracający bufor wyjściowy w klasie B z rys. 7c.



Rys. 12. Charakterystyki DC analizowanych wzmacniaczy dla wzmocnienia $k_u = 10$

Charakterystyki te zostały wyznaczone przy dużych wartościach rezystancji obciążenia, aby pokazać maksymalne zakresy liniowych części tych charakterystyk.

BIBLIOGRAFIA

1. K. Koli, K. Halonen: *A BiCMOS current — feedback operational amplifier with a 60 db constant bandwidth range*. Proc. of IEEE ISCAS, London, 30 May—2 June, 1994, pp. 525—528
2. Eyad Abou—Allam, Ezz I. El-Masry: *Design of Wineband CMOS Current — Mode Operational Amplifiers*. Proc. of IEEE ISCAS, Seatle, May, 1995, pp. 1556—1559
3. E. Bruun: *CMOS Technology and Current — Feedback Op — Amps*. Proc. of IEEE ISCAS, Chicago, May, 1993, pp. 1062—1065
4. St. Kuta, W. Machowski, J. Jasielski: *CMOS operational amplifier with current feedback*. Proc. of XVII Nat. Conf. Circuit Theory and Electronic Networks, September, 1994, pp. 113—118
5. M. Ehsanian, B. Kaminska: *BiCMOS Wideband Operational Amplifier With 900 MHz Gain — Bandwidth and 90 dB DC Gain*. Analog Integrated Circuits and Signal Processing, V. 11, No. 1, September 1996, pp. 63—71
6. K. Manetakis, C. Toumazou: *Current — feedback opamp suitable for CMOS VLSI technology*. Electronic Letters, 6th June, 1996, Vol. 32, No. 12, pp. 1090—1092

7. A.L. Coban, P.E. Allen: *A 1.75 V rail — to — rail CMOS opamp.* Proc. of IEEE ISCAS, London, 30 May–2 June, 1994, pp. 497–500
8. J. Mahattanakul, C. Toumazou: *A Theoretical Study of the Stability of High Frequency Current Feedback Op — Amp Integrators.* IEEE Trans. Circuits Syst. –1, Vol. 43, No. 1, January, 1996, pp. 2–12
9. J. Zhu, M. Sawan, K. Arabi: *An Ofset Compensated CMOS Current — Feedback Operational — Amplifier.* Proc. of IEEE ISCAS, Seatle, May, 1995, pp. 1552–1554
10. S. Franco: *Analytical foundations of current — feedback amplifiers.* Proc IEEE ISCAS, Chicago, May, 1993, pp. 1050–1053
11. Analog Devices Inc, *Amplifier Reference Manual,* 1992, USA

S. KUTA, C. KRAJEWSKI, J. JASIELSKI

CMOS CURRENT FEEDBACK OPERATIONAL AMPLIFIER'S FREQUENCY CHARACTERISTICS FORMING

Summary

The paper brings a review, comparison and estimation of influence of CMOS current feedback amplifiers (CFOAs), nonidealities as well as load- and feedback impedance on gain-bandwidth product. The transfer function analysis of CFOA was presented for resistive and resistive — capacitive feedback. Two types of CFOAs were investigated: one with a transimpedance gain stage and output voltage buffer and the second one with current-amplifier gain stage. Circuit implementations of both types of CFOAs and SPICE simulation results are also presented.

Key words: Operational amplifiers CMOS, CFOA, circuit implementations, simulation results.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1997, 43, z. 3, ss. 397–406

Inteligentne moduły mocy z tranzystorami IGBT

WITOLD PAWELSKI

Katedra Elektroniki, Instytut Elektroniki, Politechnika Łódzka

Otrzymano 1997.02.20

Autoryzowano 1997.04.25

W pracy przedstawiono jedną z najbardziej aktywnych gałęzi rozwojowych współczesnej elektroniki mocy, która wykorzystuje hybrydową integrację układów łącznikowych mocy i układów sterowania w jednym module elektroizolowanym. Opisano właściwości funkcjonalne podstawowych rodzajów inteligentnych modułów mocy (IPM) z tranzystorami IGBT. Podano też niektóre szczegóły budowy oraz scharakteryzowano parametry wybranych miarodajnych rozwiązań tych modułów.

Słowa kluczowe: układy łącznikowe, układy sterowania, moduły IPM, tranzystory IGBT.

1. WPROWADZENIE

W ostatnich dwóch dekadach lat o rozwoju elektronicznych urządzeń do przetwarzania mocy w znacznym stopniu decyduje postęp w dziedzinie łączników półprzewodnikowych. Rozmaitość odmian modyfikowanych lub opracowywanych obecnie w skali laboratoryjnej łączników mocy jest bardzo duża. Jednak tylko nieliczne spośród nich są kwalifikowane do stadium produkcji seryjnej. W tej grupie przyrządów, w klasie tzw. średnich mocy przetwarzanych, od ok. 1 kW do ok. 10 MW, największy postęp ilościowy i jakościowy obserwuje się w ostatnich latach w dziedzinie rozwoju elektroizolowanych modułów z tranzystorami IGBT [2–3] [8]. Naturalną tendencją rozwojową tych modułów jest hybrydowa integracja łączników IGBT ora szybkich diod mocy z układami sterowania i zabezpieczeń. Ideę tę w ostatnich pięciu latach udało się zrealizować w skali produkcji seryjnej zaledwie kilku wytwórcom na świecie, w postaci tzw. inteligentnych modułów mocy, zwanych dalej IPM (Intelligent Power Modules) [4], [6], [8].

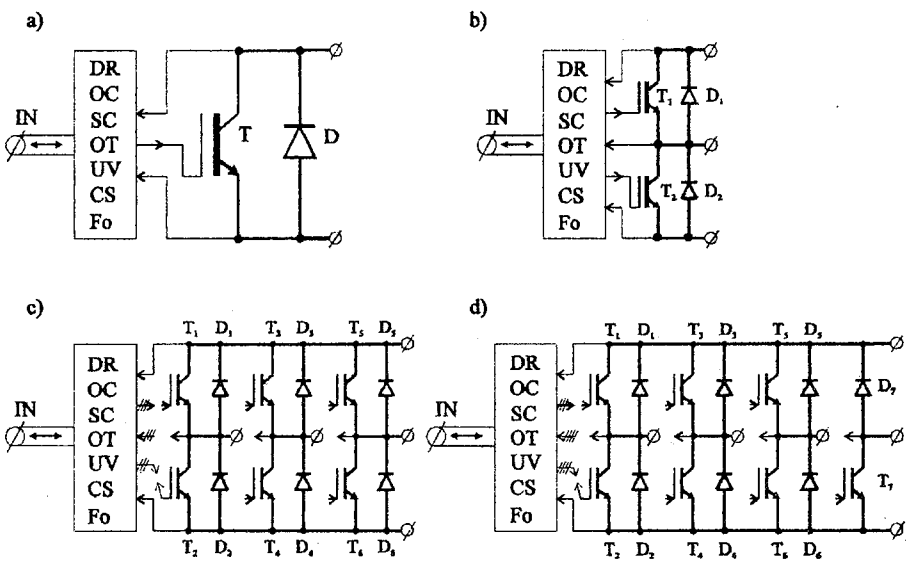
Moduły elektroizolowane jako wygodne w aplikacjach konfiguracje łącznikowe, zawierające diody mocy, tyrystory, tranzystory MOSFET mocy, IGBT oraz dwu- lub trój-elementowe układy Darlingtora tranzystorów bipolarnych, były znane i doskonalone od szeregu lat przez licznych producentów. Dopiero jednak IPM,

zawierające w jednej obudowie układy scalone, zabezpieczajace bezawaryjną pracę zarówno poszczególnych łączników jak i prawidłowe działanie tworzonego przez nie przekształtnika, stanowi dosyć wyraźny przełom w tej dziedzinie. Do sukcesu IPM przyczynił się niewątpliwie postęp w obniżeniu komutacyjnych i statycznych strat mocy w tranzystorach IGBT drugiej i trzeciej generacji. Sprzyjające temu okazały się również osiągniecia w technologii wytwarzania modułów elektroizolowanych, wykazujących współcześnie zdolność odprowadzania strat mocy rzędu 2000 W, rezystancję termiczną ($R_{th\,(j\text{-}c)}$) rzędu 0.03 °C/W oraz wytrzymałość izolacji rzędu 3 kV_{RMS}.

Rozwój inteligentnych modułów mocy, wykorzystujących technologię hybrydową, znacznie wyprzedził prace nad wprowadzeniem do produkcji masowej struktur „*smart power*", integrujących w jednej bryle półprzewodnikowej łączniki mocy oraz układy sterowania i zabezpieczeń [1], [5]. Powodem tego są między innymi nie opanowane w dostatecznym stopniu zagadnienia izolacyjne, termiczne i interakcyjne. Prace badawcze nad rozwiązaniem tych problemów od szeregu lat są prowadzone w Japonii, Europie i USA. Jednak, jak wszystko na to wskazuje, w najbliższych latach nie należy oczekiwać radykalnej dominacji tych struktur nad rozwiązaniami hybrydowymi stosowanymi w IPM.

2. KONFIGURACJE ŁĄCZNIKOWE OBWODÓW MOCY I BUDOWA IPM

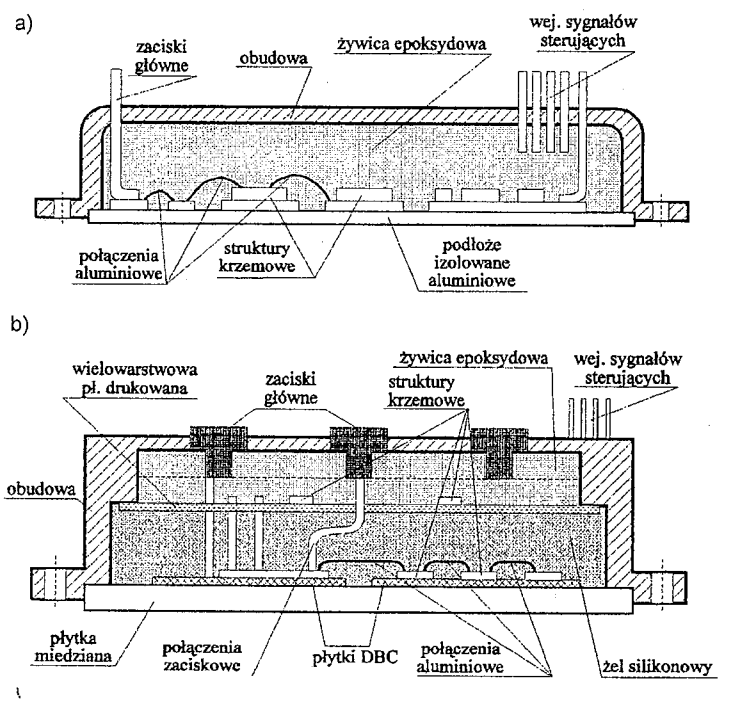
Na rys. 1 przedstawiono typowe rodzaje inteligentnych modułów mocy, zwracając szczególną uwagę na konfiguracje łącznikowe w obwodach wyjściowych, które mogą być traktowane jako standardowe. Rys. 1a pokazuje pojedynczy łącznik



Rys. 1. Podstawowe konfiguracje IGBT i diod w IPM: a — łącznik uniwersalny, b — układ półmostkowy, c — układ mostkowy trójfazowy, d — układ trójfazowy z przerywaczem

w układzie antyrównoległym IGBT z diodą mocy, który może być stosowany w wielu układach elektronicznych jako najbardziej uniwersalny ze wszystkich modułów. Na rys. 1b pokazana została wersja „półmostkowa". Bardziej złożonymi konfiguracjami obwodów mocy są układ mostkowy 3-fazowy (rys. 1c) oraz analogiczny do niego układ wyposażony dodatkowo w przerywacz (tzw. „dolny czoper"). Układ ten jest zazwyczaj wykorzystywany w falownikowych układach napędowych do rozpraszania w rezystorze nadmiarowej energii podczas procesu hamowania. Wszystkie IGBT układów mocy pokazane na rys. 1 wyposażone są w systemy sterowania, określone na tym rysunku w uproszczonej formie blokowej.

W chwili obecnej znane są dwa podstawowe rodzaje konstrukcji IPM. Pierwsza z nich, pokazana na rys. 2a, jest stosowana przy budowie modułów mniejszych mocy. Konstrukcja ta wykorzystuje jeden poziom montażowo-łączeniowy, oparty na wielowarstwowym obwodzie drukowanym z izolacją epoksydową, którego elementy są osadzone na aluminiowej płycie chłodzącej. Tak więc, zarówno krzemowe struktury IGBT i diod mocy jak również układy scalone systemu sterowania, umieszczane są w tej samej płaszczyźnie. Sygnałowe ścieżki obwodów drukowanych są przytym stosunkowo krótkie w celu zminimalizowania indukcyjnosci połączeń oraz są ekranowane elektrycznie izolowanymi warstwami miedzi. Przestrzenie wewnętrzne obudowy wypełnione są żywicą epoksydową.



Rys. 2. Budowa IPM: a — konstrukcja jednopoziomowa małej mocy, b — konstrukcja dwupoziomowa modułów większych mocy

Drugi rodzaj konstrukcji, stosowany przy realizacji modułów największych mocy, pokazany został na rys. 2b. Wykorzystuje on dwa poziomy montażowe. Elementy obwodów mocy, tj. diody mocy i IGBT, są rozmieszczane na dolnym poziomie. Ich struktury krzemowe są montowane na ceramicznym podłożu DBC (Direct Bonded Copper) lutowanym od dołu do miedzianej płaszczyzny chłodzącej. Izolacyjne podłoże ceramiczne DBC, zawierające zwykle azotek glinu, posiada bardzo wysoką przewodność cieplną i jest obecnie szeroko rozpowszechnione również przy konstrukcji zwykłych modułów elektroizolowanych mocy [8].

Jak widać z rys. 2b, płytki układów scalonych i inne elementy systemu sterowania w tej odmianie IPM są lokalizowane na górnym poziomie montażowym ekranowanym elektrycznie. Przestrzeń wewnątrz obudowy pomiędzy warstwami wypełnia żel silikonowy, a górna przestrzeń powyżej warstwy układów sterowania oraz warstwa zacisków zewnętrznych wypełnione są żywicą epoksydową.

W obu rozwiązaniach połączenia struktur półprzewodnikowych mocy oraz połączenia z zaciskami wewnętrznymi odgrywają istotną rolę, gdyż od ich długosci i kształtu decydująco zależy indukcyjność pasożytnicza w obwodach wyjściowych. Energia zgromadzona w tych indukcyjnościach w chwili wyłączenia nie może być rozładowana zewnętrznymi diodami zwrotnymi, staje się więc przyczyną przepięć. Powoduje to w efekcie ograniczenie obszaru FBSOA tranzystorów IGBT oraz RBSOA szybkich diod mocy od strony wysokich napięć. Ostatnio firma Mitsubishi, nie kwestionowany światowy lider w dziedzinie modułów mocy oraz IPM [6], opracowała tzw. serię U w ramach trzeciej generacji modułów. Wykazują one między innymi indukcyjność własną wyprowadzeń obwodów głównych zmniejszoną o 30% (na poziomie 30 nH) w stosunku do modułów poprzedniej serii H. Uzyskany rezultat jest o tyle spektakularny, że pozwala na rezygnację ze stosowania dodatkowych zewnętrznych tłumików przepięć w szerokim spektrum obciążeń, spotykanych w rzeczywistych aplikacjach takiego modułu.

3. WŁAŚCIWOŚCI FUNKCJONALNE INTELIGENTNYCH MODUŁÓW MOCY

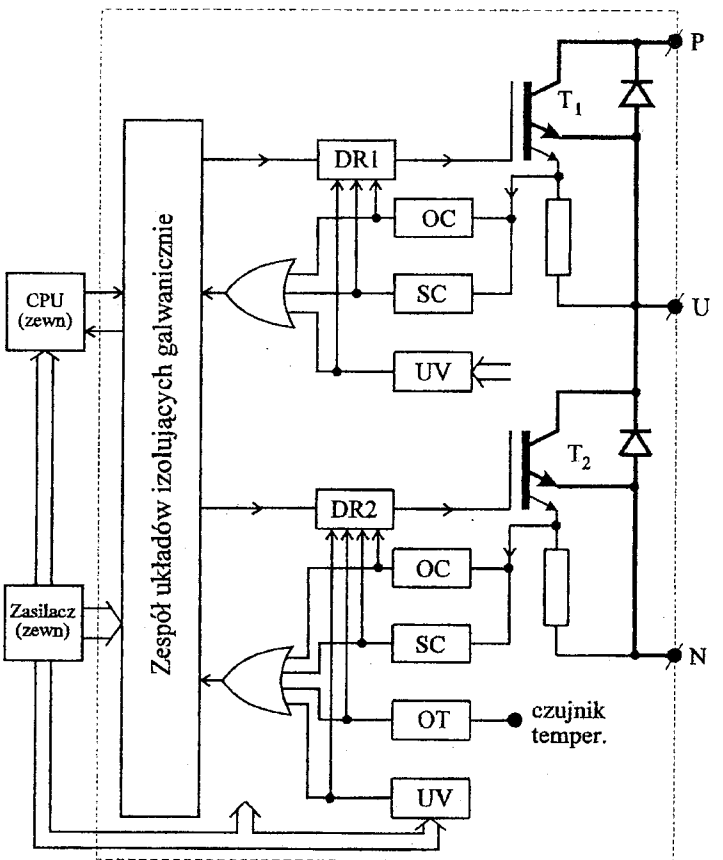
Systemy sterowania IPM, których typowe konfiguracje wewnętrzne obwodów mocy pokazane zostały na rys. 1, realizują naogół następujące programy funkcyjne:

- DR generowanie optymalnych i ściśle kontrolowanych impulsów bramkowych dla IGBT zastosowanych w danym IPM (*drivery*),
- OC detekcję przeciążenia w obwodzie wyjściowym oraz ograniczenie prądu wyjściowego wg odpowiednich algorytmów,
- SC detekcję stanu zwarcia w obwodzie wyjściowym oraz ograniczenie prądu kolektora IGBT wg odpowiednich algorytmów,
- UV detekcję stanu obniżonego napięcia zasilania układu sterującego,
- OT detekcję stanu nadmiernej temperatury wewnętrznej powierzchni montażowej modułu (jeden czujnik temperatury w module),
- CS programowe wyłączanie dużych prądów np. zwarciowych,

Fo generowanie sygnału monitorowania nieprawidłowych stanów łączeniowych modułu

Na rys. 3 przedstawiony został przykładowy uproszczony układ logiczny sterowania modułu odmiany półmostkowej z rys. 1b, który może być również fragmentem modułu IPM z rys. 1c lub 1d. Zarówno wejścia sygnałowe driverów DR1 i DR2, sygnałów blokad, jak również wyjścia sygnałów monitorowania nieprawidłowej pracy są odizolowane galwanicznie od obwodów tranzystorów T1 i T2. Rodzaj izolacji w różnych wykonaniach jest zróżnicowany. O ile firma Mitsubishi z reguły stosuje izolację transoptorową, to np. firma IXYS w swoich IPM systemu ISOSMART lansuje jako bardziej niezawodne izolowanie przy wykorzystaniu miniaturowych transformatorów, umieszczanych wewnątrz obudowy modułu.

Drobne różnice pomiędzy poszczególnymi IPM występują rownież w realizacji systemu zasilania wewnętrznych układów sterowania. Przeważają tu dwa rozwiązania. Pierwsze klasyczne, zapewniające napięcia zasilające dla obwodów tran-

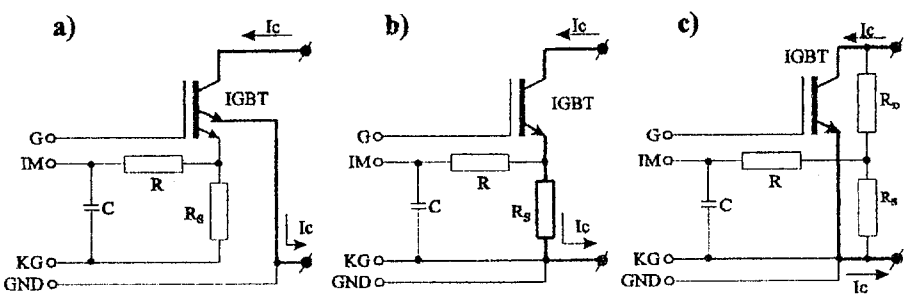


Rys. 3. Uproszczony układ logiczny sterowania jednej sekcji IPM

zystorów wysokich potencjałów (np. T1 na rys. 3), których obwody są odizolowane galwanicznie oraz jednego napięcia zasilania dla obwodów wszystkich tranzystorów niskiego potencjału (np. T2, T4, T6 i T7 na rys. 1d). Napięcia te pochodzą z oddzielnych uzwojeń zewnętrznego transformatora sieciowego. Inne rozwiązanie polega na dostarczaniu jednego napięcia zasilającego do obwodów wewnętrznych modułu. Zasilanie to jest następnie dystrybuowane do poszczególnych odizolowanych sekcji przy zastosowaniu przetwornicy podwyższonej częstotliwości o szeregu wyjściowych uzwojeniach małego transformatora z ferrytowym rdzeniem toroidalnym.

Niezwykle istotnym zagadnieniem IPM jest ochrona przeciwprzeciążeniowa (*OC-Over Current*) oraz przeciwzwarciowa (*SC-Short Circuit*). Na rys. 4 przedstawiono trzy podstawowe rodzaje detekcji przeciążeniowych i zwarciowych prądów kolektora, możliwe do zastosowania w przypadku IGBT. Pierwszy z nich, pokazany na rys. 4a wykorzystuje właściwości zwierciadła prądowego, tworzonego przez dodatkowy emiter[1] IGBT, wiodący niewielką (chociaż ściśle określoną) 1/n-tą część prądu kolektora. Typowa wartość współczynnika podziałowego n wynosi przy tym około 1500, co przy napięciowym poziomie detekcji rzędu 300 mV umożliwia zastosowanie bocznika prądowego małej mocy.

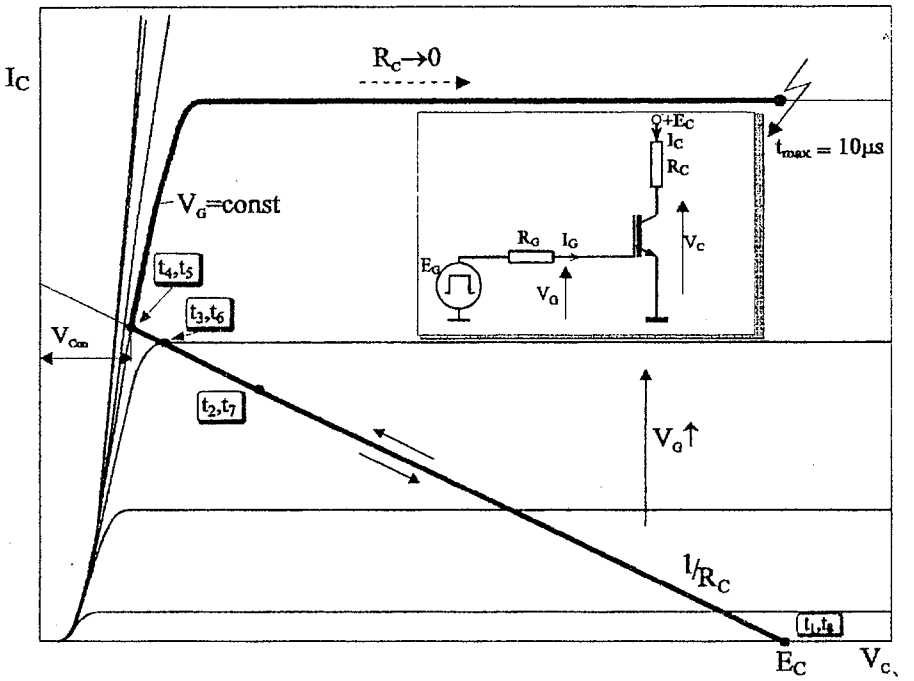
Drugim rozwiązaniem pomiaru prądu kolektora, które pokazane zostało na rys. 4b, jest zastosowanie klasycznego bocznika prądowego w emiterze jednoemiterowego IGBT. Wymagane rozmiary bocznika często wykluczają tę koncpecję z zastosowaniem w IPM, zwłaszcza przy większych mocach przetwarzanych przez moduł.



Rys. 4. Sposoby detekcji zwarć i przeciążeń: a — zwierciadło prądowe, b — bocznik w obwodzie wyjściowym, c — układ detekcji stanu aktywnego IGBT

Często spotykanym sposobem ograniczania przeciążeniowych i zwarciowych prądów wyjściowych IGBT jest również tzw. system *self-limiting*, którego działanie zostało zilustrowane na rys. 5. Jeśli w chwili t_4 podczas stanu przewodzenia IGBT, w sytuacji wykazanej na tym rysunku, wystąpi zwarcie ($R_c \to 0$) przy niezmiennym

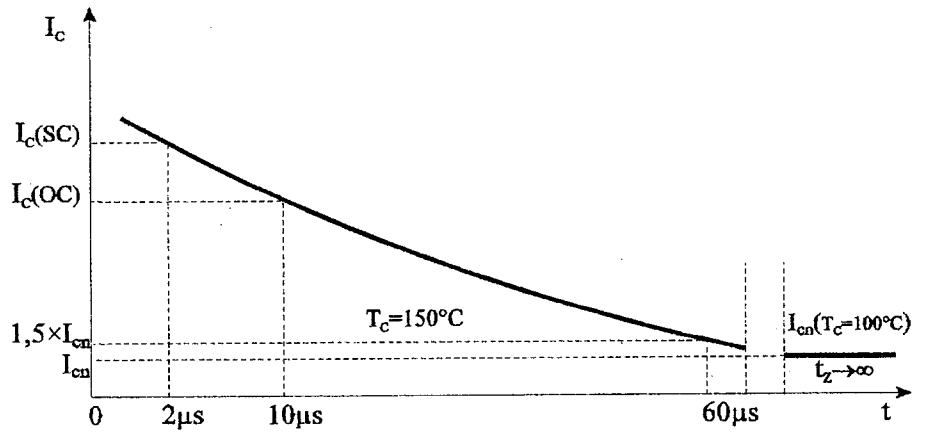
[1] Tranzystory IGBT pokazane na rys. 1 najczęściej są również tranzystorami dwuemiterowymi.



Rys. 5. Ilustracja zjawiska samoograniczania prądu kolektora (*self-limiting*) w IGBT podczas przeciążeń i zwarć

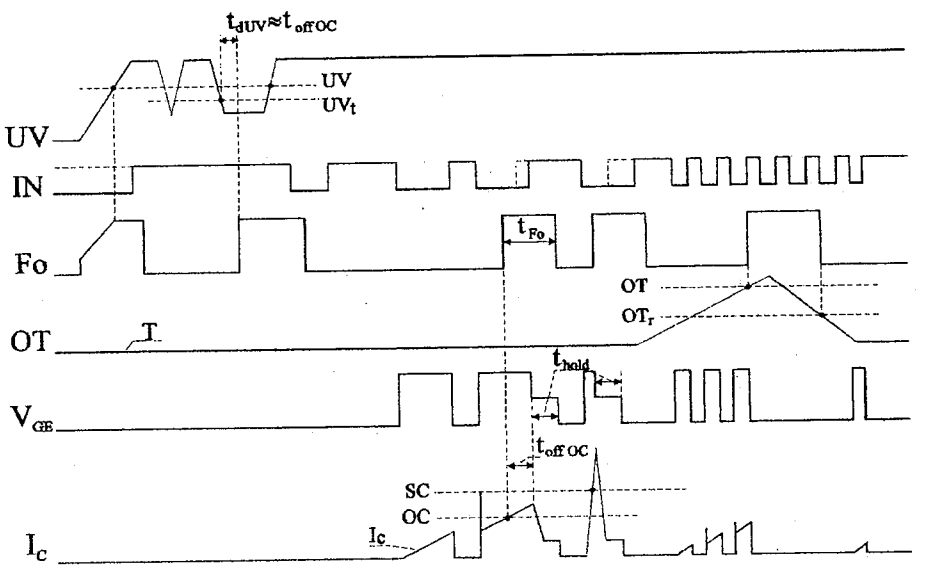
napięciu bramki V_G, wówczas punkt pracy przesuwa się wzdłuż statycznej charakterystyki wyjściowej w obszarze pentodowym. Tranzystor łącznikowy staje się w ten sposób źródłem prądowym, czemu towarzyszy radykalny wzrost napięcia kolektor-emiter. Tak więc objawy wzrostu napięcia kolektora IGBT przy niezmiennym napięciu bramki i prawidłowym jego doborze względem wymaganego poziomu prądu obciążenia staje się podstawą trzeciej metody detekcji prądu zwarciowego lub przeciążeniowego. Stosowny układ pomiarowy z dzielnikiem R_D i R_S pokazany został na rys. 4c. We wszystkich trzech układach z rys. 4 zastosowano również filtr dolnoprzepustowy RC, którego stała czasowa odgrywa istotną rolę w procesie ograniczania prądu i tolerowania krótkotrwałych stanów przeciążeniowych.

Porównując powyższe trzy sposoby detekcji przeciążeń i zwarć, zaprezentowane na rys. 4 kazuje się, że zdecydowanie najkorzystniejszym do stosowania w IPM jest pierwszy z nich (rys. 4a). Zapewnia największą dokładność pomiaru, nie wywołując przy tym stresu termicznego, jaki może wystąpić np. w konsekwencji gwałtownego wzrostu strat mocy w warunkach stosowania trzeciej metody (rys. 4c). Na rys. 6 określone zostały typowe poziomy progowe układów zabezpieczenia przeciwzwarciowego (SC) i przeciążeniowego (OC) oraz odpowiadające im maksymalne czasy trwania w określonej temperaturze płyty chłodzącej modułu T_c.



Rys. 6. Poziomy krytyczne prądu kolektora IGBT układów zabezpieczających OC i SC

Na rys. 7 przedstawione zostały przykładowe przebiegi czasowe ważniejszych wielkości elektrycznych i termicznych w jednym sektorze IPM. na rysunku uwzględniono poziomy dyskryminacji detektorów OC, SC, OT i UV oraz niektóre istotniejsze opóźnienia i czasy własne, wnoszone przez te układu. Wyjściowy sygnał monitorowania nieprawidłowości F_0, pojawiający się w postaci jedynki logicznej ma określony minimalny czas trwania i blokuje wówczas impulsy wyjściowe V_{GE} drivera mimo istnienia wejściowych impulsów sterujących IN. Jak widać z rys. 7, sygnał ten



Rys. 7. Przykładowe ważniejsze przebiegi czasowe w układach zabezpieczających IPM

jest generowany przez układ zabezpieczający UV (*Under Voltage*) w chwili, gdy napięcie zasilania zmniejszy się poniżej poziomu UV_t lub przez układ zabezpieczający (*OT-Over Temperature*), gdy temperatura płyty radiatorowej modułu przekroczy poziom OT. Ponadto pojawia się on wówczas, gdy poziom prądu kolektora I_c przekroczy wartość przeciążeniową OC lub poziom zwarciowy SC. W obu powyższych przypadkach czas reakcji na to sygnału bramkowego V_{GE} jest różny i zgodny z krzywą koordynacji zabezpieczeń pokazaną na rys. 6.

Niektóre IPM są wyposażone w układ CS (*Controlled Shutdown*) automatycznej redukcji napięcia bramki (najczęściej o 50%) w przypadku, gdy sygnał F_0 jest generowany w następstwie zadziałania układu OC lu SC. Efekt działania tego układu zilustrowano na przebiegach czasowych $V_{GE}(t)$ oraz $I_c(t)$ na rys. 7. Powodem stosowania tego układu jest dążenie do względnie łagodnego ograniczania prądu kolektora IGBT o stosunkowo dużych wartościach w celu minimalizacji przepięć generowanych na skutek istnienia indukcyjności pasożytniczych obwodu mocy.

UWAGI KOŃCOWE

Wyniki studiów światowego stanu rozwoju inteligentnych modułów mocy z łącznikami IGBT wykazują, że układy te stanowią nową jakość w dziedzinie układów elektronicznych średniej i dużej mocy. Takie ich cechy jak: izolacja galwaniczna między obwodami sterowania, obwodami mocy i radiatorem chłodzącym oraz rozbudowany „inteligentny" system zabezpieczeń, a także wzrastająca stale moc przetwarzana na wyjściu odgrywają tu istotną rolę. Podkreślić również trzeba znaczne ułatwienia w projektowaniu docelowych urządzeń z IPM [7]. Spośród walorów użytkowych na uwagę zasługuje możliwość monitorowania pracy IPM, jak rownież znacznie zwiększona odporność na zakłócenia zewnętrzne i interakcyjne. Wynikają one z faktu integracji, sprzyjających warunków filtracji, odpowiedniej konstrukcji oraz możliwości ekranowania układow i podzespołów.

Bardzo interesująco rysują się również perspektywy rozwojowe inteligentnych modułów mocy. Zapowiadana jest bowiem czwarta generacja realizacji tych modułów z tranzystorami TIGBT (Trench IGBT) [9] o znacznie zwiększonej szybkości przełączania oraz radykalnie zmniejszonych statycznych stratach mocy.

BIBLIOGRAFIA

1. B.J. Baliga: *An Overview of Smart Power Technology.* IEEE Transactions on Electron Devices, 1991, vol. 38, No. 7, pp. 1568 – 1575
2. B.J. Baliga: *Power ICs in the Saddle.* IEEE Spectrum, 1995, July, pp. 34 – 49
3. B.J. Baliga: *Power Semiconductor Devices for the 1990's* EPE'91 Proceedings pp. 001 – 002
4. R. Frank, P. Aloisi: *Power Devices with Integrated Protection* EPE'91 Procedings pp. 110 – 113
5. K. Gabriel: *An Intrinsic Save Smart Power IGBT* EPE'91 Proceedings pp. 179 – 182
6. G. Majumdar, D. Medaule: *A New Generation of Intelligent Power Devices for Motor Drive Applications* PEVSD Conference Publication, 1994, No. 399, pp. 35 – 41

7. T. Orłowska-Kowalska, T. Dudek, M. Kuśnierz: *Zastosowanie inteligentnych modułów mocy i mikrokontrolerów do sterowania silników elektrycznych prądu przemiennego*. Materiały V Sympozjum Podstawowe Problemy Energoelektroniki Gliwice 1993, pp. 327–331
8. W. Tursky: *Power Modules for Compact Inverters*. PCIM Europe, 1996, No. 6, pp. 380–384
9. F. Udrea, G.A.J. Amaratunga, Q. Huang: *The Effect of the Hole Current on the Channel Inversion in Trench Insulated Gate Bipolar Transistor*. (TIGBT), 1994, vol. 37, No. 3, pp. 507–514

W. PAWELSKI

THE INTELLIGENT POWER MODULES WITH IGBT SWITCHES

Summary

In the paper the intelligent power modules state of the art is presented. The basic kinds modules including both the IGBT power switches and control system are shown. IPM construction details as well as its functional features are discussed. Very miscellaneous protection system as important for module reliability is especially analyzed. The improvement methods of the 3-rd generation IPM proprieties and the development ways are also described.

Key words: power switches, control system, IPM modules, IGBT transistors.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1997, 43, z. 3, ss. 407–420

Korekcja interferencji metodą przewidywania nadchodzących danych

JERZY KISILEWICZ

Instytut Sterowania i Techniki Systemów, Politechnika Wrocławska

Otrzymano 1997.03.17

Autoryzowano 1997.07.28

W pracy opisano metodę korekcji interferencji międzysymbolowej przy użyciu korektora z decyzyjnym sprzężeniem zwrotnym bez korektora wstępnego. Korektor wstępny zastąpiono przewidywaniem przyszłych danych. Na podstawie odebranego sygnału określa się prawdopodobieństwa aktualnie odbieranych danych binarnych i danych, które będą odbierane w następnych taktach. Decyzje podejmuje się tak, aby wypadkowe prawdopodobieństwo związane z odbieranymi danymi było największe. Przeprowadzono symulację transmisji przy użyciu opisanego korektora oraz tradycyjnego korektora z decyzyjnym sprzężeniem zwrotnym i korektorem wstępnym i korektora transwersalnego prostego. Porównano częstości występowania błędów dla czterech wybranych kanałów w obecności białego szumu gaussowskiego. Eksperyment prowadzono dla różnej liczby przewidywanych decyzji.

Słowa kluczowe: transmisja danych, korektor transwersalny, decyzyjne sprzężenie zwrotne.

WSTĘP

Rozważmy kanał transmisji danych złożony z analogowego kanału podstawowego i korektora z detektorem. Na wejście kanału jest podawany z okresem T ciąg impulsów. Każdy impuls jest elementem sygnału o amplitudzie równej nadawanym danym a_k. Niech $y(t)$ będzie odpowiedzią kanału podstawowego na ten impuls. Odpowiedź kanału podstawowego jest próbkowana w chwilach $t = kT$. Oznaczmy $y_i = y(iT)$, gdzie i jest liczbą całkowitą. Przyjmijmy, że y_i dla $i < 0$ oraz $i > g$ są na tyle bliskie zeru, że można je pominąć. Próbka odpowiedzi kanału podstawowego na sygnał danych w chwili $t = kT$ będzie równa

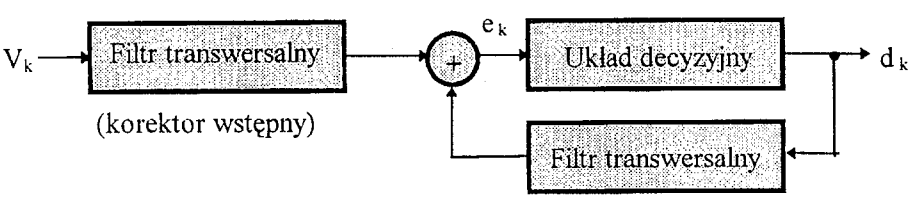
$$v_k = \sum_{i=0}^{g} a_{k-i} y_i \tag{1}$$

Jeśli kanał podstawowy wnosi opóźnienie $t = hT$ $(0<h<g)$, to próbka v_k reprezentuje dane (symbol) a_{k-h}. Wartość v_k zależy też od wartości innych symboli a_{k-i} $(i \neq h)$. Mówimy, że symbole te interferują z symbolem a_{k-h}. Występujące we wzorze (1) składniki $a_{k-i}y_i$ dla $0<i<g$ oraz $i \neq h$ są komponentami interferencji międzysymbolowej.

Zasadniczymi przyczynami powstawania błędów transmisji są różnego typu zakłócenia oraz zniekształcenia charakterystyk kanałów analogowych, a zwłaszcza związane z tymi zniekształceniami interferencje międzysymbolowe [1, 2, 3, 4, 5, 6, 9, 11, 14].

Dołączenie korektora do kanału podstawowego ma na celu zmniejszenie prawdopodobieństwa błędnego odbioru poprzez redukcję interferencji miedzysymbolowej.

W pracach [2, 3, 4, 5, 6, 9, 13] opisano typowe korektory. W [4] opisano algorytmy obliczania korektorów transwersalnych. Jednym z najczęściej stosowanych korektorów jest korektor z decyzyjnym sprzężeniem zwrotnym. Korektory te mają w pętli sprzężenia zwrotnego nieliniowy układ decyzyjny oraz filtr transwersalny zbudowany na cyfrowej linii opóźniającej. Filtr transwersalny posiada cyfrową linię opóźniającą, w której pamiętane są kolejne decyzje detektora. Korektory te wymagają wstępnej korekcji kanału mającej na celu wyzerowanie wartości prążków $y_0, y_1,..., y_{h-1}$, które występują przed prążkiem głównym. Rolę korektora wstępnego pełni z reguły prosty filtr transwersalny, który często znacznie pogarsza stosunek mocy sygnału do mocy szumu. Schemat blokowy korektora z decyzyjnym sprzężeniem zwrotnym i korektorem wstępnym pokazano na rys. 1.



Rys. 1. Schemat korektora z decyzyjnmym sprzężeniem zwrotnym

Podstawowym celem tej pracy jest wyeliminowanie korektora wstępnego. W pracy [8] opisano metodę z przewidywaniem jednej przyszłej decyzji. W tej pracy uogólnia się tę metodę uwzględniając większą liczbę przyszłych deyczji co pozwala korygować większą liczbę prążków $y_0, y_1,..., y_{h-1}$ występujących przed prążkiem głównym y_h.

REGUŁA DECYZYJNA

Przyjmijmy, że system transmisyjny zawiera korektor z decyzyjnym sprzężeniem zwrotnym oraz, że korektor ten nie posiada filtru wstępnego. Koryguje on zatem tylko te składowe interferencji, które są związane z danymi rozpoznanymi wcześniej.

W pracy [8] przyjęto taką interferencję kanału, która zawiera jeden istotny składnik nie korygowany przez decyzyjne sprzężenie zwrotne. W tej pracy odrzucono to ograniczenie dopuszczając większą liczbę nieskorygowanych składników. Jeśli wzmocnienia na kolejnych odczepach cyfrowej linii opóźniającej wynoszą y_{h+1}, $y_{h+2},\ldots, y_g$, to wartość próbki e_k na wejściu układu decyzyjnego przy braku zakłóceń wynosi

$$e_k = v_k - \sum_{i=h+1}^{g} d_{k-i} y_i = \sum_{i=0}^{h} a_{k-i} y_i + \sum_{i=h+1}^{g} (a_{k-i} - d_{k-i}) y_i. \tag{2}$$

Zakładając poprawność podejmowanych decyzji mamy $d_{k-i} = a_{k-i}$ oraz

$$e_k = \sum_{i=0}^{h} a_{k-i} y_i. \tag{3}$$

Zakładając transmisję binarnych danych otrzymujemy ze wzoru (3) 2^{h+1} nominalnych wartości próbek e_k. Tę liczbę próbek można zredukować rozpatrując r początkowych mniej znaczących próbek interferencyjnych $y_0, y_1,\ldots, y_{r-1}$ jako szum o mocy $\sigma_y^2 = y_0^2 + y_1^2 + \ldots + y_{r-1}^2$.
Oznaczamy przez D_k ciąg $a_{k-h}, a_{k-h+1},\ldots, a_{k-r}$

$$D_k = \{a_{k-h}, a_{k-h+1},\ldots, a_{k-r}\}.$$

Numerycznie utożsamimy D_k z wartością liczby binarnej utworzonej przez odpowiedni ciąg danych a_{k-i}. Tak więc D_k przyjmuje wartość z przedziału od 0 do $2^{h-r+1} - 1$, a wzór (3) przyjmie postać

$$e_k = \sum_{i=r}^{h} a_{k-i} y_i. \tag{4}$$

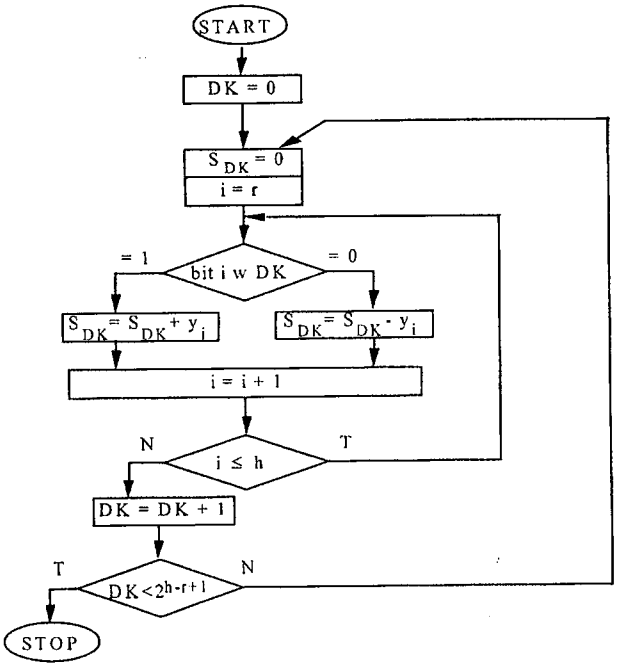
Oznaczamy przez S_{D_k} obliczoną z (4) wartość e_k dla ciągu D_k. Schemat blokowy algorytmu obliczania progów S_{D_k} pokazano na rys. 2. W algorytmie założono symetrię sygnału, czyli, że bitowi 0 odpowiada poziom -1 a bitowi 1 poziom sygnału równy 1.
Jeżeli rzeczywista wartość próbki na wejściu układu decyzyjnego wynosi x_k, a transmitowane są dane D_k, to wartość zakłócenia wynosi

$$z_k = x_k - S_{D_k}.$$

Zakładając dokładność pomiaru Δ oraz rozkład gęstości prawdopodobieństwa zakłóceń $f(z)$ obliczamy, że dla danych D_k oraz $\Delta \to 0$ zostanie zaobserwowana wartość x_k z prawdopodobieństwem

$$p(D_k) = \Delta f(x_k - S_{D_k}). \tag{5}$$

Ponieważ x_k jest wynikiem pomiaru, $p(D_k)$ jest prawdopodobieństwem nadania ciągu D_k dla zmierzonego x_k.



Rys. 2. Algorytm obliczania progów S_{D_k}

Aby wyznaczyć wartość Δ posługujemy się wzorem na prawdopodobieństwo całkowite (suma po wszystkich możliwych D_k)

$$1 = \sum_{D_k} p(D_k) = \Delta \sum_{D_k} f(x_k - S_{D_k}). \tag{6}$$

Dzieląc stronami (5) przez (6) otrzymujemy ostatecznie

$$p(D_k) = \frac{f(x_k - S_{D_k})}{\sum\limits_{D_k} f(x_k - S_{D_k})}. \tag{7}$$

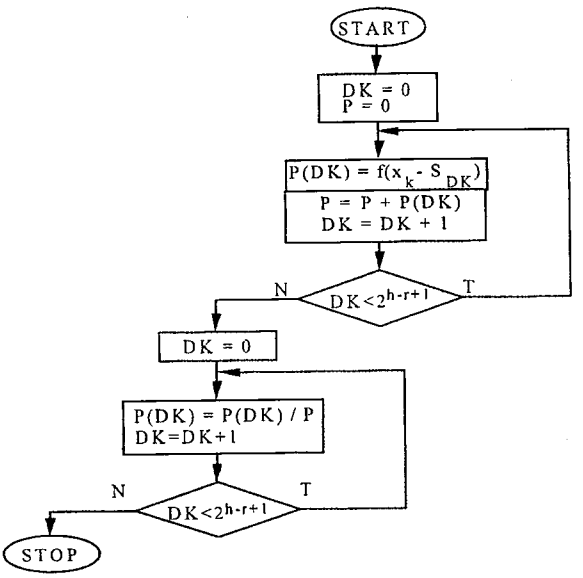
Oznaczamy przez A_i zbiór wszystkich ciagów D_k, w których $a_{k-i} = 1$, dla $i = r$, $r + 1,..., h$. Wyznaczone z próbki x_k prawdopodobieństwo $P_1^k(i)$ tego, że $a_{k-i} = 1$ wynosi

$$P_1^k(i) = \sum_{D_k \in A_i} p(D_k). \tag{8}$$

Analogicznie z x_k wyznaczamy prawdopodobieństwo $P_0^k(i)$ tego, że $a_{k-i} = 0$ jako

$$P_0^k(i) = \sum_{D_k \notin A_i} p(D_k) = 1 - P_1^k(i). \tag{9}$$

Algorytm obliczania $p(D_k)$ pokazano na rys. 3 a algorytm obliczania $P_1^k(i)$ oraz $P_0^k(i)$ pokazano na rys. 4.



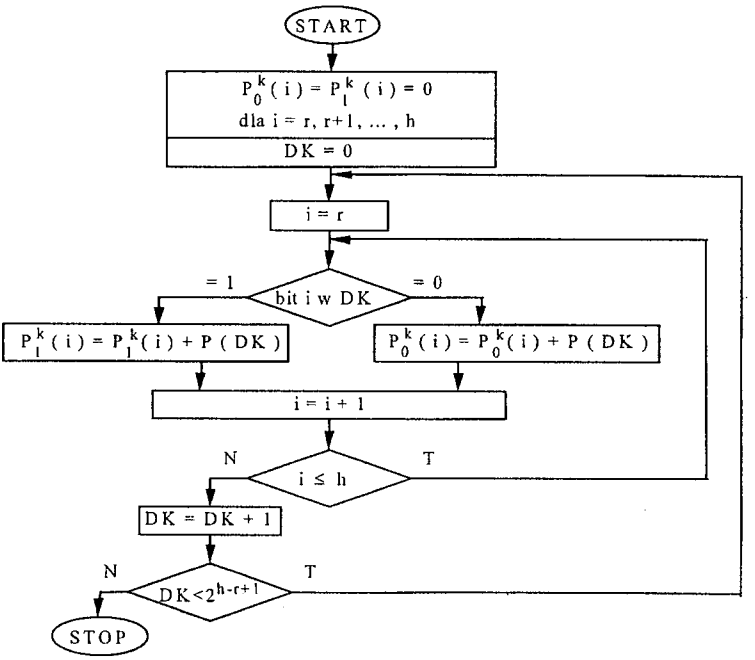
Rys. 3. Algorytm obliczania $P(D_k)$

Po odebraniu próbki x_k należy podjąć decyzję d_{k-h} taką, aby prawdopodobieństwo tego, że $d_{k-h} = a_{k-h}$ było największe.
Z próbki x_k wyznaczono prawdopodobieństwa $P_1^k(h)$ oraz $P_0^k(h)$ tego, że $a_{k-h} = 1$ oraz, że $a_{k-h} = 0$.
Zauważmy, że

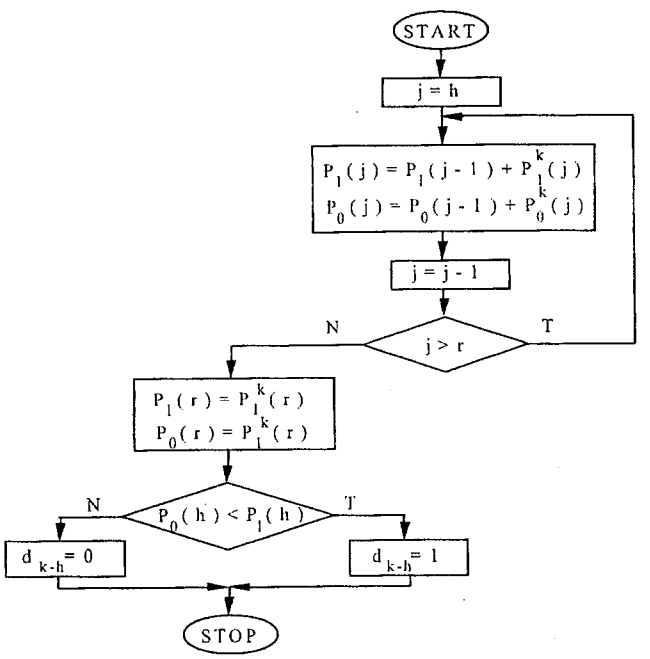
$$a_{k-h} = a_{(k-j)-(h-j)} \quad \text{dla} \quad (j = r, r+1, \ldots, h).$$

Ostatecznie prawdopodobieństwo P_1^k tego, że $a_{k-h} = 1$ oraz prawdopodobieństwo P_0^k tego, że $a_{k-h} = 0$ obliczamy jako wartości średnie

$$\begin{aligned} P_1^k &= \frac{1}{h-r+1} \sum_{j=r}^{h} P_1^{k-j}(h-j), \\ P_0^k &= \frac{1}{h-r+1} \sum_{j=r}^{h} P_0^{k-j}(h-j) = 1 - P_1^k. \end{aligned} \tag{10}$$



Rys. 4. Algorytm obliczania $P_1^k(i)$ oraz $P_0^k(i)$



Rys. 5. Algorytm obliczania P_1^k, P_0^k oraz d_{k-h}

Układ decyzyjny podejmuje decyzję bardziej prawdopodobną

$$d_{k-h} = \begin{cases} 0 & \text{gdy} \quad P_0^k \geqslant P_1^k \\ 1 & \text{gdy} \quad P_0^k < P_1^k \end{cases} \tag{11}$$

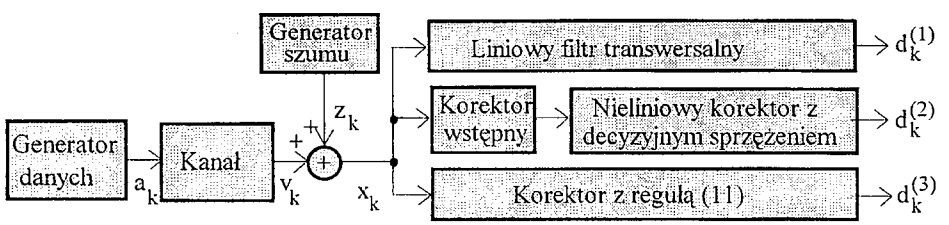
Algorytm realizujący wzory (10) i (11) przedstawiono na rys. 5.

EKSPERYMENT

W celu zbadania właściwości opracowanego algorytmu (11) przeprowadzono szereg eksperymentów obliczeniowych. Symulowano pracę systemu złożonego ze źródła danych, kanału z zakłóceniami i 3 korektorów:

— prostego filtru transwersalnego minimalizującego błąd średniokwadratowy,
— korektora z decyzyjnym sprzężeniem zwrotnym i z liniowym korektorem wstępnym,
— korektora z decyzyjnym sprzężeniem zwrotnym i z regułą (11).

System ten pokazano na rys. 6.



Rys. 6. Schemat symulowanego systemu

Danych binarnych dostarczył generator pseudoprzypadkowych liczb całkowitych. Ten sam generator posłużył do wygenerowania próbek białego szumu gaussowskiego o zadanej mocy. W eksperymencie symulowano biały szum gaussowski, choć ważniejszymi czynnikami powodującymi przekłamania są inne rodzaje zakłóceń, jak np. szum impulsowy i krótkie przerwy transmisji [6, 10]. Błędy powstałe w wyniku zakłóceń impulsowych lub krótkich przerw eliminuje się przez stosowanie kodowania nadmiarowego, ponieważ efektywność eliminacji tych błędów w kanale ziarnistym jest niewielka [6].

Modelem kanału był prosty filtr transwersalny, który wprowadzał zadaną interferencję międzysymbolową. Próbki x_k zaszumionego sygnału na wyjściu kanału podawano na wejście liniowego filtru transwersalnego oraz na wejścia 2 korektorów z decyzyjnym sprzężeniem zwrotnym: z korektorem wstępnym, bez korektora wstępnego z przewidywaniem przyszłych decyzji z regułą decyzyjną (11). Dla różnych poziomów zakłóceń porównywano dane odbierane $d_k^{(i)}$ z danymi nadanymi $a_i (i = 1, 2, 3)$. Dla każdego korektora wyznaczono liczbę błędnie odebranych bitów i odniesiono tę liczbę do liczby wszystkich odebranych bitów.

Eksperyment powtarzano zmieniając odstęp sygnału od szumu oraz liczbę przewidywanych deyczji, czyli liczbę korygowanych próbek interferencyjnych występujących przed próbką główną. Pozostałe nie korygowane (początkowe) próbki interferencyjne zostały uwzględnione w formie dodania ich mocy (sumy kwadratów $\sigma_y^2 = y_0^2 + y_1^2 + ... + y_{r-1}^2$) do mocy szumu we wzorze (5). W każdym eksperymencie generowano ciąg 65536 bitów.

Do eksperymentu wybrano kanały użyte przez A.P. Clarka w [5] oraz kanały użyte przez A. Dąbrowskiego w [6]. Ponieważ sygnał wyjściowy kanału zakłócano białym szumem gaussowskim, to prawdopodobieństwa $p(D_k)$ obliczano przyjmując we wzorze (7)

$$f(x_k - s_{ij}) = \exp\left(-\frac{(x_k - s_{ij})^2}{2\sigma^2}\right), \tag{12}$$

gdzie $\sigma^2 = \lambda^2(\sigma_N^2 + \sigma_y^2)$, σ_N^2 jest mocą szumu w kanale a λ jest dobranym eksperymentalnie współczynnikiem. W początkowych eksperymentach przyjmowano $\lambda = 1$, a następnie $\lambda = \frac{1}{y_h}$ odnosząc moc zakłóceń do mocy głównego prążka.

WYNIKI I WNIOSKI

Próbki odpowiedzi użytych kanałów przedstawiono w Tabeli 1.

Tabela 1

	g	h	Wartości próbek y_k								
Kanał 1	4	2	0,227	0,460	**0,688**	0,460	0,227				
Kanał 2	12	6	0,002	−0,005	−0,015	−0,040	0,108	−0,294	**0,800**	−0,439	0,161
			−0,059	−0,022	−0,008	0,003					
Kanał 3	10	5	0,04	−0,06	0,07	−0,21	−0,5	**0,72**	0,36	0,00	0,21
			0,03	0,07							
Kanał 4	4	2	0,167	0,471	**0,707**	0,471	0,167				

Na rysunkach od 7 do 10 przedstawiono dla tych kanałów wykresy obliczonej elementowej stopy błędów (stosunek liczby błędnie odebranych bitów do liczby wszystkich odebranych bitów) w funkcji wyrażonego w decybelach stosunku mocy sygnału do mocy szumu.

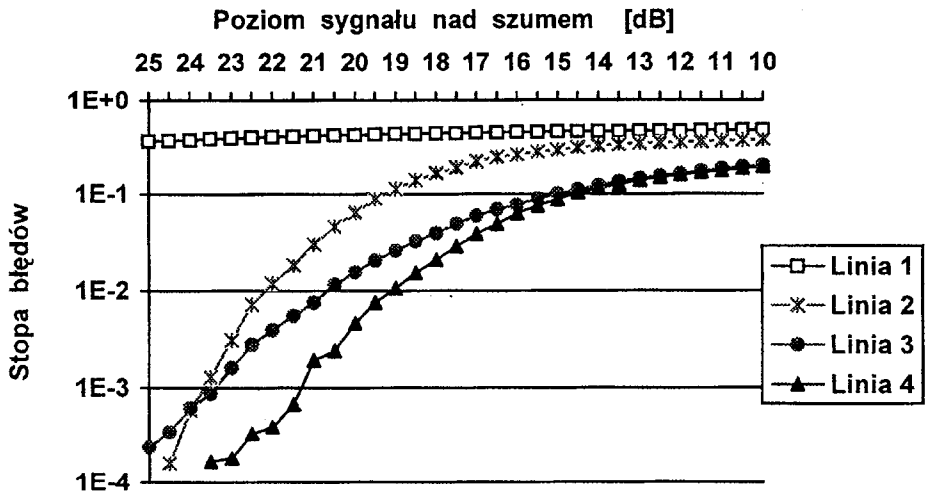
Na wszystkich wykresach poszczególne krzywe oznaczono numerami przyporządkowując je wynikom dla następujących korektorów:

Linia 1. prosty filtr transwersalny minimalizujący błąd średniokwadratowy,

Linia 2. korektor z decyzyjnym sprzężeniem zwrotnym i z liniowym korektorem wstępnym,

Linia 3. (i dalsze numery) korektor z decyzyjnym sprzężeniem zwrotnym i z regułą (11) dla różnych liczb (rosnąco) przewidywanych decyzji.

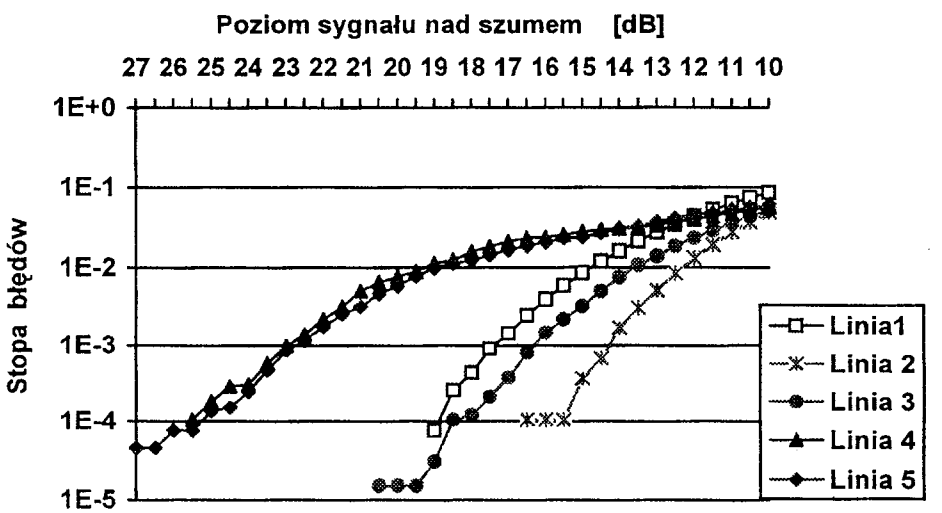
Na rys. 7 linia 3 reprezentuje wyniki dla pojedynczej decyzji, natomiast linia 4 — dla dwu przewidywanych decyzji. Przewidywanie dwu decyzji daje lepszą stopę błędów od pozostałych algorytmów. W porównaniu z typowym korektorem z decy-



Rys. 7. Zależnosć stopy błędow od poziomu sygnału nad szumem dla kanału 1

zyjnym sprzężeniem zwrotnym poprawa jest równoważna spadkowi mocy zakłóceń o ok. 2,5 do 5 dB. Przy większych zakłóceniach (poziom sygnału poniżej 16 dB) algorytm z przewidywaniem pojedynczej deyczji jest równie dobry jak z przewidywaniem dwu decyzji.

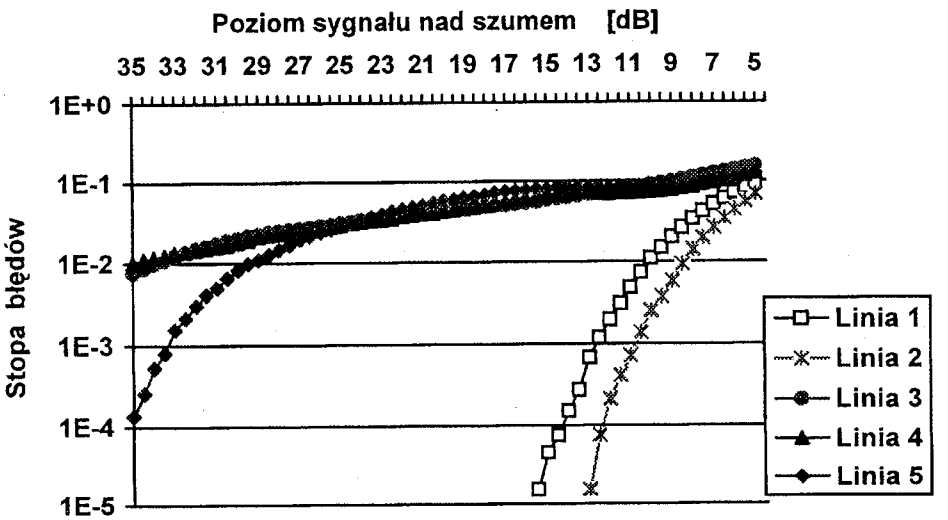
Na rys. 8 linie 3, 4, 5 prezentują wyniki dla algorytmu przewidującego kolejno 1, 2 i 3 decyzje. W przypadku kanału 2 najlepsze efekty daje korektor tradycyjny oraz



Rys. 8. Zależnosć stopy błędów od pziomu sygnału nad szumem dla kanału 2

algorytm przewidujący tylko jedną decyzję. Uwzględnienie większej liczby decyzji pogarsza stopę błędów tak jak wzrost mocy szumów o ok. 7 dB. Algorytmy przewidujące 2 i 3 decyzje są prawie równoważne.

Na rys. 9 linie 3, 4, 5 prezentują wyniki dla algorytmu przewidującego kolejno 1, 2 i 3 decyzje. W przypadku kanału 3 najlepsze efekty daje korektor tradycyjny oraz prosty filtr transwersalny. Prezentowany tu algorytm (11) jest wyraźnie gorszy niezależnie od liczby przewidywanych decyzji, choć uwzględnienie trzeciej decyzji poprawia jego działanie. Algorytmy przewidujące 1 i 2 decyzje są prawie równoważne.

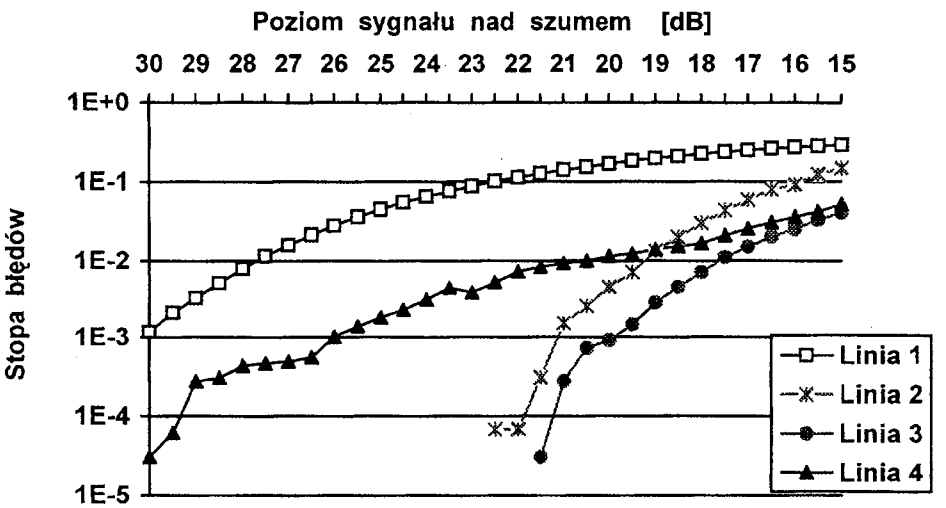


Rys. 9. Zależnosć stopy błędów od poziomu sygnału nad szumem dla kanału 3

Na rys. 10 linia 3 reprezentuje wyniki dla pojedynczej decyzji, natomiast linia 4 — dla dwu przewidywanych decyzji. W przypadku kanału 4 zdecydowanie najlepsze efekty daje przewidywanie jednej decyzji. W porównaniu z typowym korektorem z decyzyjnym sprzężeniem zwrotnym, algorytm (11) poprawia odbiór tak jak zmniejszenie mocy szumu o ok. 1 do 2 dB.

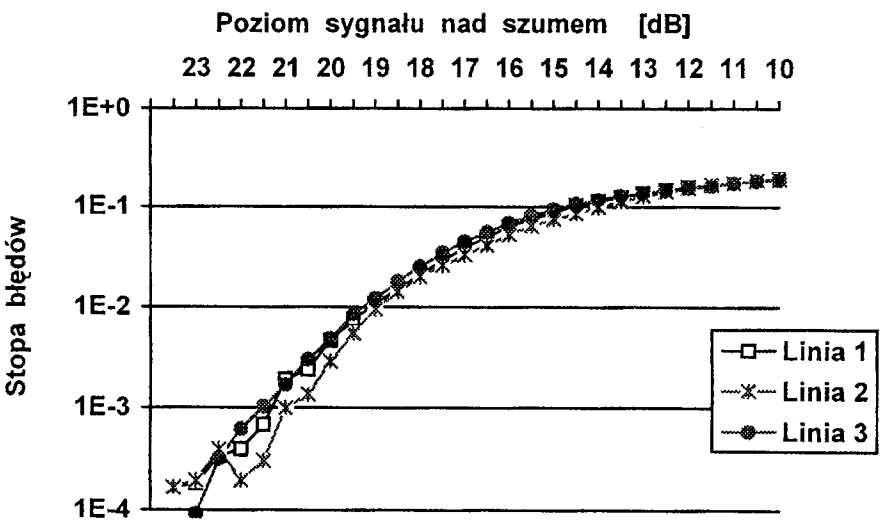
W dotychczasowych eksperymentach zakładano we wzorze (12) $\lambda = 1$. Kolejne eksperymenty wykonano, aby zbadać wpływ sposobu uwzględniania we wzorze (12) zakłóceń na skuteczność opisywanego tu korektora z przewidywaniem przyszłych decyzji. W kolejnych eksperymentach założono w (12) $\lambda = \frac{1}{y_h}$ oraz założono $\sigma^2 = \sigma_N^2$ pomijajac zakłócające działanie próbek interferencyjnych $y_0, y_1, \ldots, y_{h-r-1}$.

Wykresy na rysunkach od 11 do 14 ilustrują uzyskane zależności elementowej stopy błędów od poziomu sygnału nad szumem kolejno dla kanałów od 1 do 4 oraz dla optymalnych liczb przewidywanych decyzji. Na każdym rysunku przedstawiono 3 wykresy:

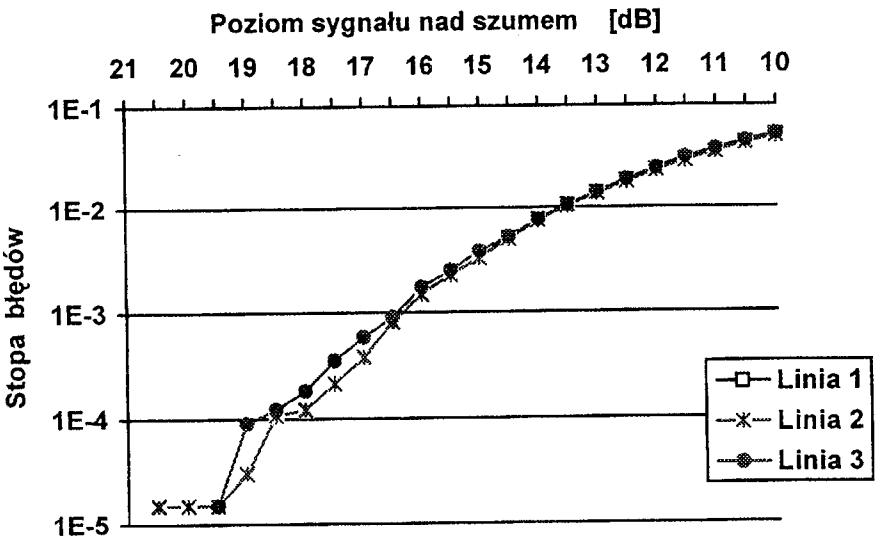


Rys. 10. Zależność stopy błędów od poziomu sygnału nad szumem dla kanału 4

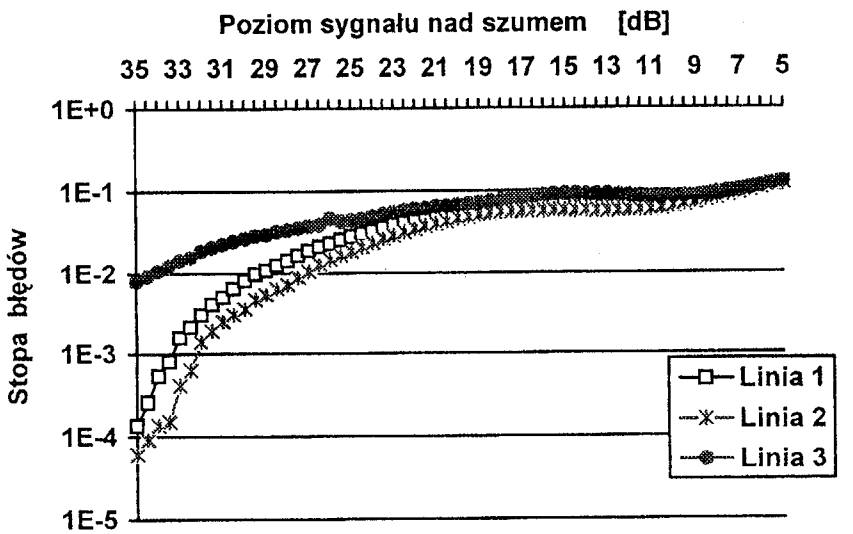
Linia 1. algorytm oryginalny użyty w poprzednim eksperymencie z $\lambda = 1$,
Linia 2. algorytm, który w (7 i 12) odnosi moc szumu i moc pominiętych próbek do wartości głównej z $\lambda = \frac{1}{y_h}$,
Linia 3. algorytm, który w (7 i 12) nie uwzględnia początkowych nie korygowanych próbek interferencyjnych przyjmując $\sigma^2 = \sigma_N^2$.



Rys. 11. Zależność stopy błędów od poziomu sygału nad szumem dla kanału 1

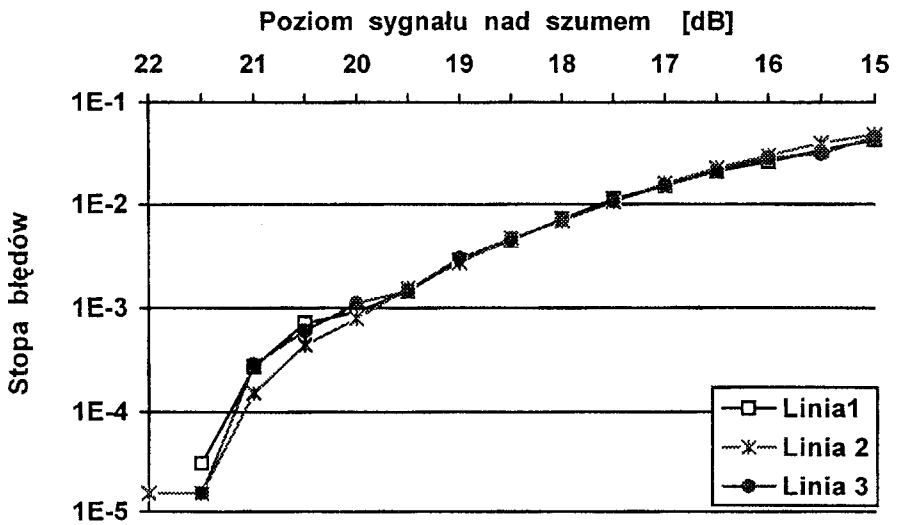


Rys. 12. Zależność stopy błędów od poziomu sygnału nad szumem dla kanału 2



Rys. 13. Zależność stopy błędów od poziomu sygnału nad szumem dla kanału 3

Jak widać z wykresów we wszystkich przypadkach użycie mocy szumu powiększonej o moc początkowych próbek interferencyjnych i odniesionej do wartości próbki daje nieco lepsze wyniki w postaci nieco mniejszej stopy błędów. Interesujące jest w tym to, że poprawa (choć znikoma) wystąpiła dla wszystkich badanych kanałów.



Rys. 14. Zależność stopy błędów od poziomu sygnału nad szumem dla kanału 4

BIBLIOGRAFIA

1. N. Abramson, F.F. Kuo (red.): *Sieci telekomunikacyjne komputerów*. WNT, Warszawa 1978
2. W.R. Bennett, J.R. Davey: *Data Transmission*. Mc Graw-Hill, New York 1965
3. A.B. Carlson: *Communication Systems*. McGraw-Hill, Tokyo 1975
4. A.P. Clark: *Advanced Data-Transmission Systems*. Pentech Press, London 1977
5. A.P. Clark: *Principles of Digital Data Transmission*. Pentech Press, London 1976
6. A. Dąbrowski: *Odbiór sygnałów transmisji danych w obecności zniekształceń interferencyjnych i zakłóceń*. Rozdz. 4, Pr. zb. pod kier. Z. Barana: Podstawy transmisji danych, WKŁ, Warszawa 1982
7. J. Kisilewicz: *Problemy poprawy wierności transmisji danych w sieciach komputerowych*. Prace Naukowe Instytutu Sterowania i Techniki Systemów Politechniki Wrocławskiej, Seria: Monografie nr 6, WPWr, Wrocław 1991
8. J. Kisilewicz: *Nieliniowa korekcja interferencji z przewidywaniem przyszłej decyzji*. Kwartalnik Elektroniki i Telekomunikacji, 1994, z. 4, ss. 489—497
9. R.W. Lucky, J. Salz, E.J. Weldon: *Principles of Data Communication*. McGraw-Hill, New York 1968
10. T. Nowicki, T. Stachowiak, W. Stańczak: *Wybrane zagadnienia sieci teleinformatycznych*. PWN, Warszawa 1985
11. W. Sobczak: *Wprowadzenie do teleinformatyki*. WKŁ, Warszawa 1979
12. W. Sobczak (red.): *Problemy teleinformatyki*. WKŁ, Warszawa 1984
13. M. Sternad, A. Ahlen: *The Structure and Design of Realizable Decision Feedback Equalizers for IIR Channels with Colored Noise*. IEEE, July 1990, vol. IT-36, no. 4, p. 848—858
14. A.S. Tanenbaum: *Sieci komputerowe*. WNT, Warszawa 1988

J. KISILEWICZ

INTERSYMBOL INTERFERENCE EQUALIZATION USING FURTHER DECISIONS PREDICTION

Summary

The method of particular intersymbol interference equalization is described. The nonlinear equalizer with decision feedback but without linear feedforward transversal filter at the input is used. The intersymbol interference components that are not eliminated by the nonlinear equalizer, are eliminated by further decisions prediction method. Basing on received signal, the probabilities of further possible data bits incoming in the next sample moments are calculated. The decision is worked out to maximize the medium probability of received data. The data transmission systems with described equalizer, common equalizer with decision feedback and with linear feedforward transversal filter at the input where simulated. The error bit rate these systems with presence of white Gaussian noise was calculated and compared for four channels and different numbers of predicated decisions.

Key words: data transmission, transversal equalizer, decision feedback.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1997, **43**, z. 3, ss. 421–437

Evaluation of bandwidths in spectral analysis and micrometry

BOHDAN WOŁCZAK

Instytut Elektroniki i Informatyki, Politechnika Szczecińska

Received 1997.02.13

Authorized 1997.07.28

The aim of this work is to present new limits of application of some often applied criteria of evaluating and measurements bandwidths (linewidths) and to compare the results. There were given results of comparison bandwidths calculated for twelve typical signals according to six definitions of bandwidths applied and 12 coefficients for "equal areas criterion" of bandwidths. It was demonstrated that the problem of evaluating or measuring bandwidths cannot be solved by applying any definition of bandwidth separately, because there exist some cases for which the antinomies are very strong.

Key words: antinomies, spectral bandwidths, micrometry, linewidth measurement.

1. INTRODUCTION

One of the principal notion of the signal theory, spectral analysis, radio — and optoelectronics and laser micrometry is the notion and definition of the bandwidth or the interference fringes. Bandwidth can be defined variously and the definition used are not directly interrelated because they imply from different conventions. This situation does not cause too much difficulties when statistic features of the investigated structures (interference patterns) do not change. Otherwise one has to face extreme difficulties not only in the determination and measurement of structure elements but even in qualitative estimation

whether the element of the interference pattern under discussion has enlarged or narrowed.

The paper reports some new results of numerical calculation of band of spectral widths for different signals (images).

2. PRINCIPAL THEORETICAL DEPENDENCES

The frequency bandwidth is calculated from the radiant power spectrum $S(f)$ possessing the following property:

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

what corresponds to 100% of the general signal power [1–3]. The $S(f)$ spectrum is related to the unnormalized spectrum of power density $G(f)$ by the dependence:

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

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

where: $y(t)$ is the analyzed signal, t is the time and f is the frequency.

The analytical form of $y(t)$ is not known for random (stochastic) signals and therefore formula (3) is not useful for specific calculations.

Notice: In laser micrometry [4] or in detection of structure edges — images [5] the $S(f)$ function is a function of grayness and variable f is a geometrical variable.

Knowing, for a random signal, the corresponding normalized autocorrelation function $a(\tau)$ having the properties:

$$a(0) = 1 \qquad a(\tau) = a(-\tau) \qquad |a(\tau)| <= 1$$

and

$$\lim_{\tau \to \infty} a(\tau) = 0$$

one can calculate the radiant power spectrum $S(f)$ using the Wiener transformations:

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

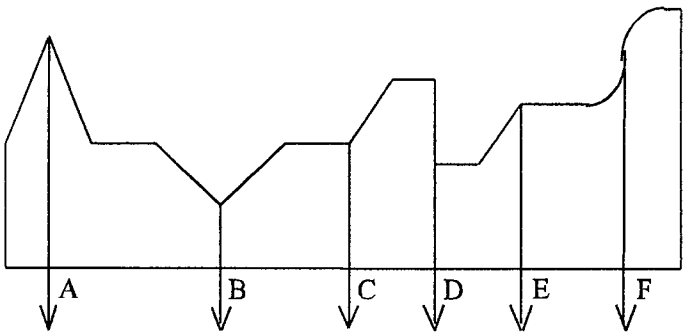


Fig. 1. Principal types of edges: (A) convex roof, (B) concave roof, (C) linear concave jump, (D) jump, (E) linear convex jump, (F) unlinear jump

The functions of the grayness level $S(f)$ called previously [1 – 3] as spectra can be interpreted as a microscopic pattern of different type of element edges of given (investigated) image structures. The principal types of edges, taken from [5] are illustrated in Fig. 1. These functions can be given in a triple way:

a) analytical, b) numerical, c) graphical.

In the present paper within 12 form functions analyzed 7 were given according to a) and b) and 4 were given according to c).

The plots of the different form functions of $S(f)$ adopted for numerical studies are given in figures $S_1 \div S_{12}$ in the appendix I and II.

The detection of edges in image recognition consists in the extraction of:

- discontinuity point in the level of grayness (points type D)
- discontinuity point in the first derivative (points type A, B, C, E)
- zeroths of the second derivative (points type F)

However, in the spectral analysis one applies other, internationally recognized criteria for the estimation of the bandwidth, i.e. the criterion of the determination (detection) of the coordinate of the defined band limit (or defined width of a given function).

3. THE APPLIED CRITERIA FOR THE ESTIMATION OF CHANGES IN BANDWIDTHS

The following, previously published [1 – 3], criteria were applied in numerical studies to the estimation of the changes in bandwidths (the changes in the location of a detector of a structure element edge image).

1. Criterion for a normalized spectrum moment — formula (5)
2. Criterion for a unit area in the band — formula (6)

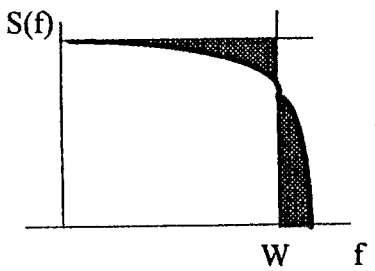


Fig. 2. Criterion of equalization of blackened areas [4]

3. Criterion for a drop of the spectrum to a given value, i.e. to 3 dB — formula (7)
4. Energetical criterion for 90% of signal power in the band — formula (8) with $p = 0{,}9$
5. Energetical criterion for 99% of signal power in the band — formula (8) with $p = 0{,}99$
6. Informational criterion for W_0 — formula (9)
7. Criterion for "equal areas" [4]

3.1. Criterion for a normalized spectrum moment:

$$W_1 = (\mathrm{d}f)_1 = 0.5\sqrt{\int_0^\infty f^2 S(f)\,\mathrm{d}f}. \tag{5}$$

3.2. Criterion for a unit area in the band:

$$W_2 = (\mathrm{d}f)_2 = \frac{1}{\max S(f)}. \tag{6}$$

3.3. Criterion for a drop of the spectrum to a given value, i.e. to 3 dB:

$$0.5\,S(W_3) = 2\int_0^\infty a(\tau)\,\mathrm{d}\tau. \tag{7}$$

3.4. Energetical criterions for p. 100% of signal power in the band:

$$0 < \int_0^{(\mathrm{d}f)_4} S(f)\,\mathrm{d}f = p < 1 \tag{8}$$

standard $p = 0.9$ or $p = 0.99$.

3.5. Informational criterion for W_0:

$N(W_0) = 0$, where

$$N(W) = 0.5\ln W + \frac{1}{2W}\int_0^W \ln S(f)\,\mathrm{d}f - \ln\int_0^W S(f)\,\mathrm{d}f \tag{9}$$

$-\infty < N(W) < \infty$.

4. THE DISCUSSION OF THE RESULTS

For. "1 – 7" numbers group of $S(f)$ function below we list the number of the investigated spectra in sequence of the narrowing of particular bands.

It is seen, from the Table 1, that there exists no identical distribution of spectra numbers for any two definitions. Some contradictions can be observed, for example, when changing the characteristics of signal No 2 to characteristics of that No 3 in

Table 1

The list of spectra numbers 1 – 7 in sequence of band narrowing

criterion 1	1,6	6,1	4	3	7	2
criterion 2	4,6	6,4	1,5	**3**	**2**	7
criterion 3	4	6	5,1	**2**	**3**	7
criterion 4	1	6	**4**	**3**	7,5	2
criterion 5	6	1	**3**	7	**4**	2,5
criterion 6	4	1	3	5,2	7	6

criterion 2 what is in contrast with the values resulting from criterion 3, and also for signals 1 and 6 as well as 3, 4 and 7 in criterion 4 and 5.

The list of spectra nuemration "8 – 12" for particular signals in sequence of band narrowing are presented below.

Table 2

The list of spectra numbers 8 – 12 in sequence of band narrowing

criterion 1	8	11	12	**9**	**10**
criterion 2	8	11	12	**10**	**9**
criterion 3	8	11	12	9,10	10,9
criterion 4	8	11	12	9	10
criterion 5	8	11	12	9	10
criterion 6	8	11	12	9	10

In the above list, a consistent spectra numeration is obtained for criteria 3 – 6, whereas from the calculation resulting from the application of criterion 1 and 2 for signals 9 and 10 one can observe a contradiction. the obtained consistency in the sequence of narrowing of bandfunctions S_8, S_{11}, S_{12} can be explained by a considerable similarity of such shapes for spectra of power density of signals or by corresponding changes in distributions of the grayness level of structure elements (images).

Table 3
The results of bandwidths calculation W of function $S(f)$ according criterion 7

Nr $S(f)$	W	Symbol
1	0,2438	j
2	0,1410	l
3	0,1591	k
4	0,4999	g
5	0,2947	i
6	0,3868	h
7	0,1250	d
8	1,7569	a
9	1,0434	f
10	1,1053	e
11	1,4090	b
12	1,1478	c

For criterion 7 ("*equal areas*") the results of calculation are listed in Table 3 and Table 4.

These results are similar with the results obtained using criterion 2 — "*unit* (elementary) *area in the defined band*"

Table 4
The list of spectra numbers 1 – 12 in sequence of band narrowing with criterion 7 applied

Symbol	a	b	c	d	e	f	g	h	i	j	k	l
Nr $S(f)$	8	11	12	7	10	9	4	6	5	1	3	2

CONCLUSIONS

A numerical verification has collaborated the correctness of previously published conclusions and also yielded some new results. The programmes applied can be used to the calculation of the width changes for different curves (grayness distributions of image contours or edges of structure elements) for the shapes different from those considered so far. This can be valuable not only in the laser micrometry.

However, the problem connected with an issue of the evaluation of bandwidths in images remain still current and cannot be unequivocally solved in their complexity in the sense of removing the antinomies in the results of measurements (or calculations) for different physically realisable spectra or grayness functions $S(f)$ taking into a account a variety of presently existing criteria.

The problem under discussion is of great importance during the recognition of images and the detection of contour edges of images, in microelectronics (the detection of conduction paths) and in laser microscopy.

$$S_3(f) = 2\Pi b e^{-2\Pi b f}$$

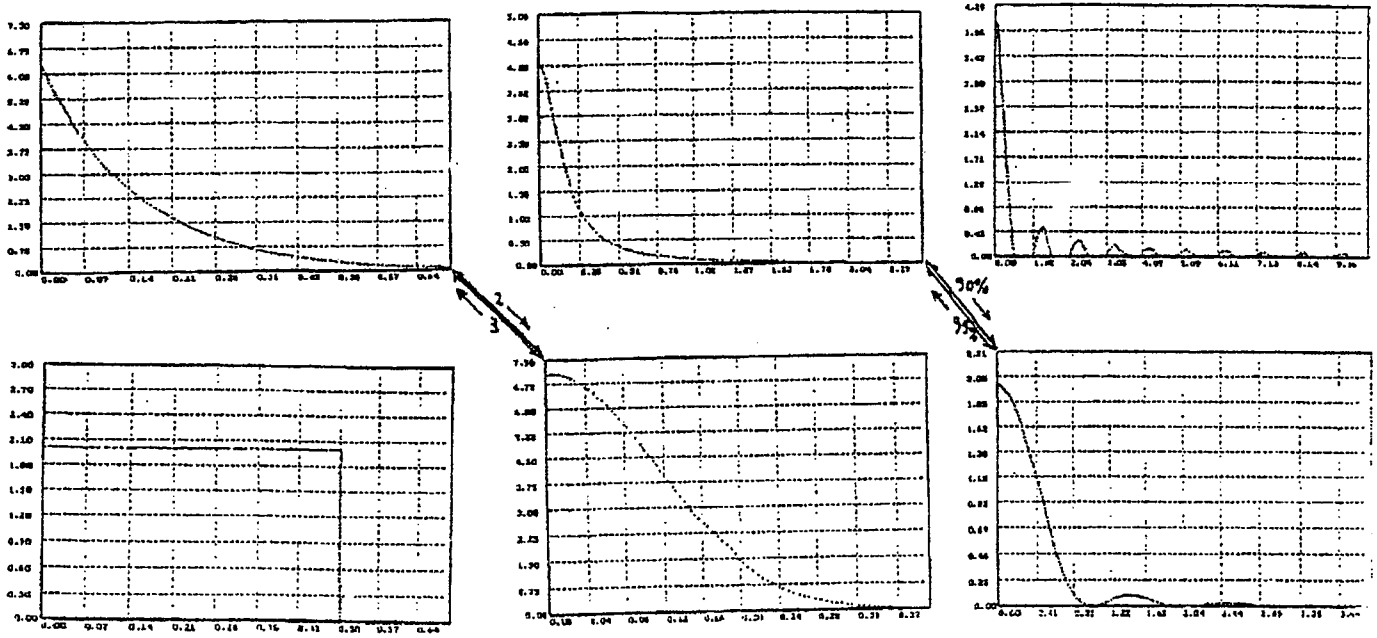
$$S_1(f) = \frac{4b}{1+(2\Pi b f)^2}$$

$$S_5(f) = \frac{2\sin 2\Pi b f}{\Pi f}$$

$$\frac{n}{b} \le f \le \frac{n}{b} + \frac{1}{2b}$$

dla $b>0$ oraz

$n = 0, 1, 2, 3, \ldots$

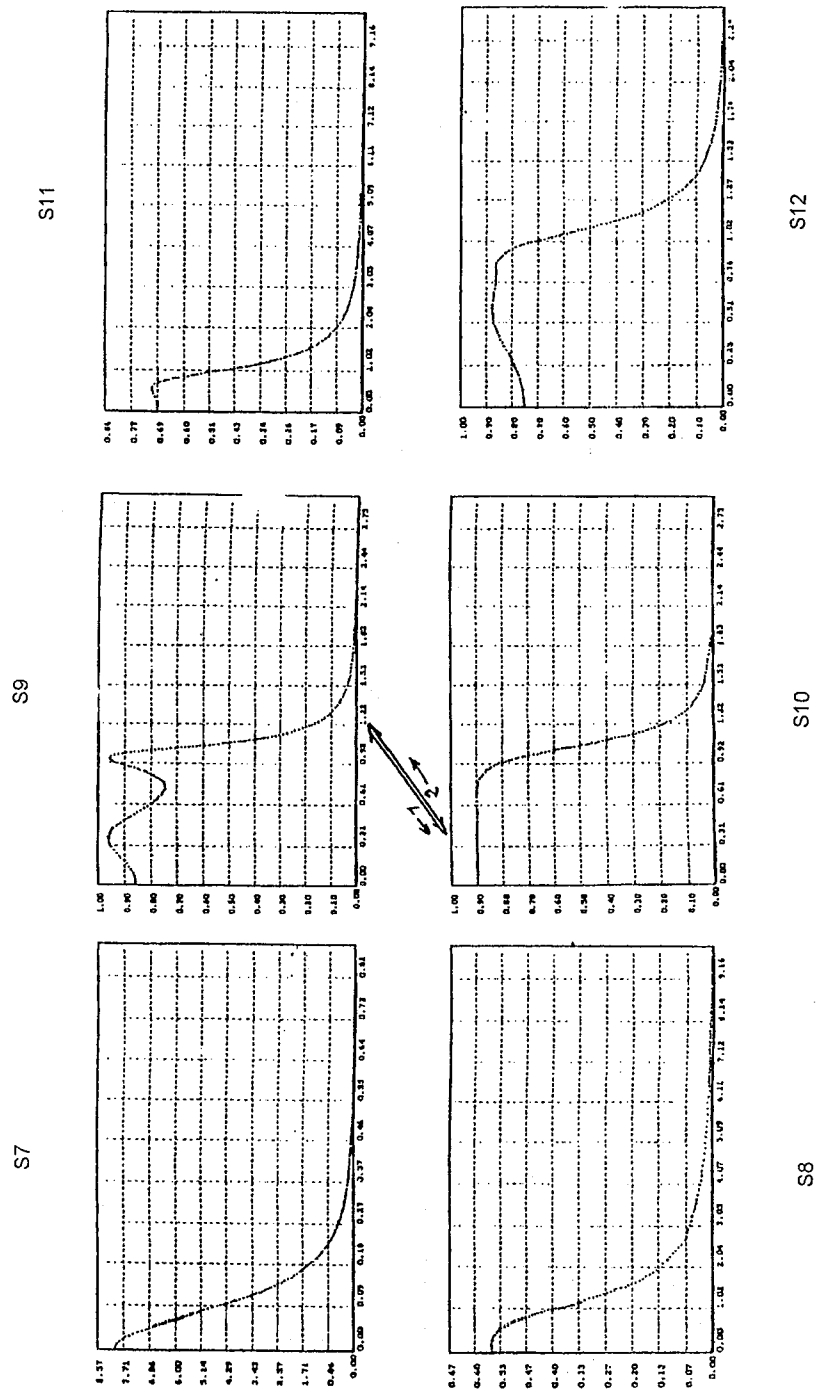


$$S_4(f) = \begin{cases} 2b \Rightarrow f < 2b^{-1} \\ b \Rightarrow f = 2b^{-1} \\ 0 \Rightarrow f < 2b^{-1} \end{cases}$$

$$S_2(f) = 4b\sqrt{\Pi} e^{-(2\Pi b f)^2}$$

$$S_6(f) = \frac{2\sin^2(\Pi b f)}{b\Pi^2 f^2}$$

APPENDIX II



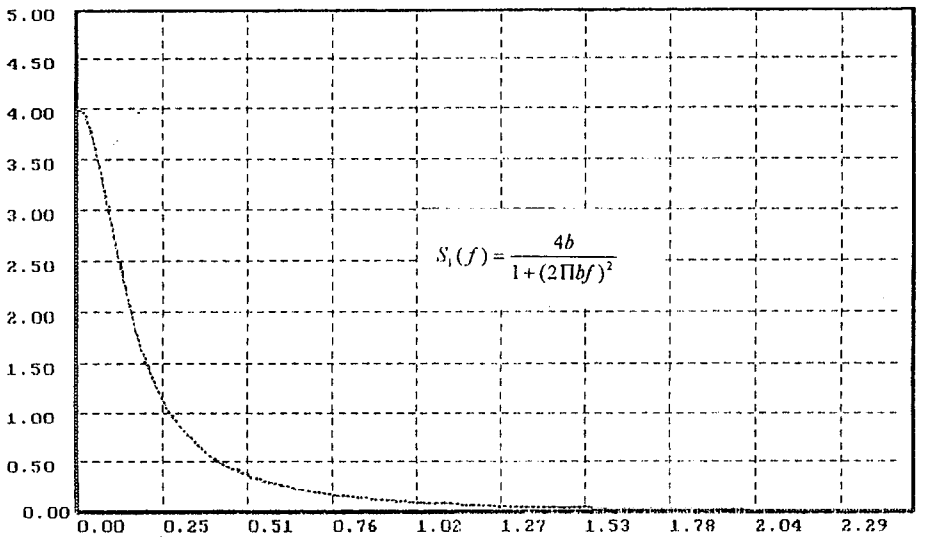


Fig. 3 — S1(f)

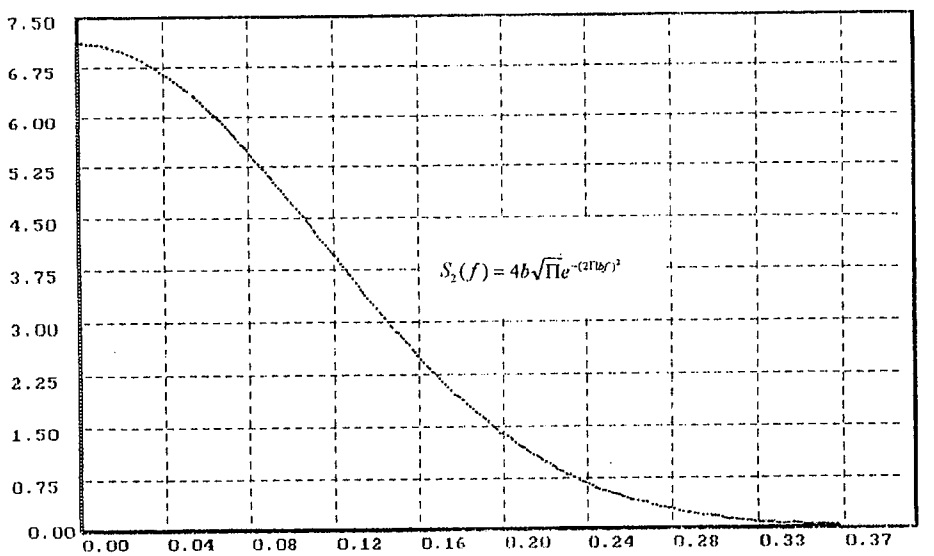


Fig. 4 — S2(f)

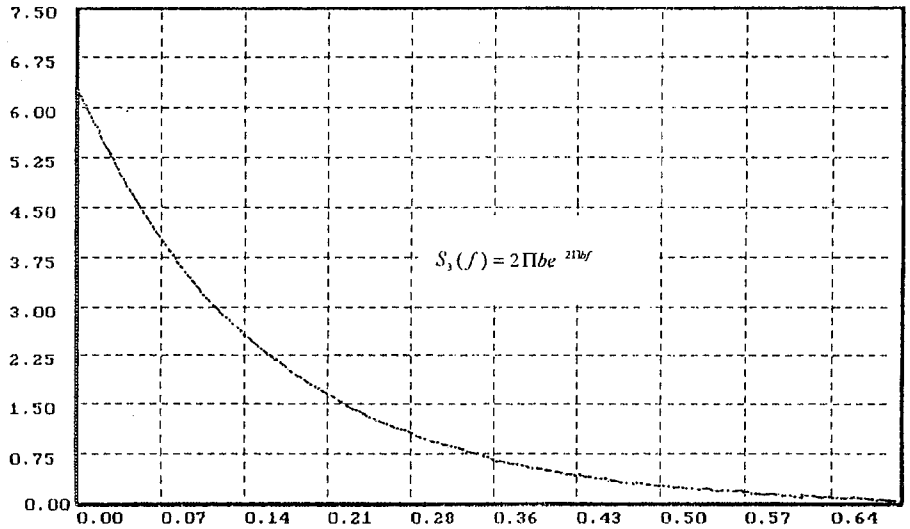


Fig. 5 — S3(f)

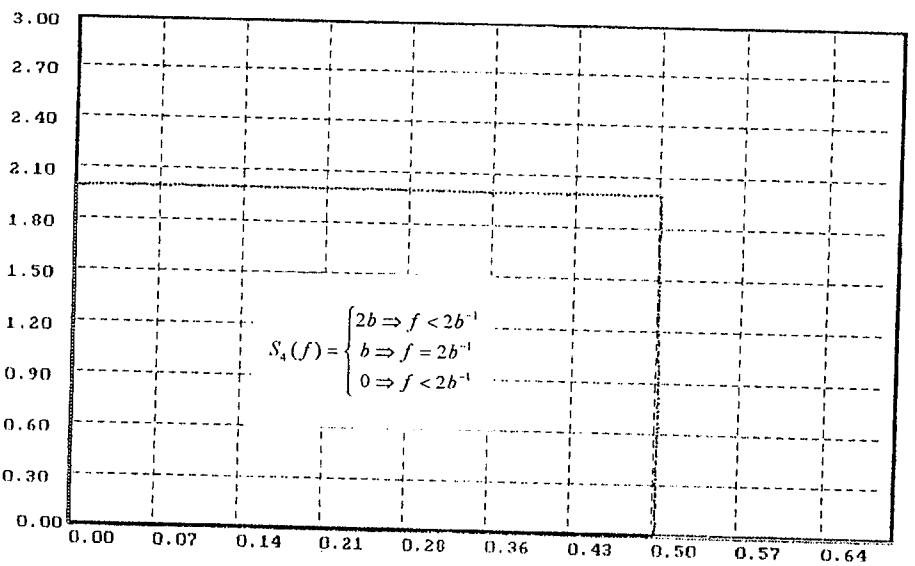


Fig. 6 — S4(f)

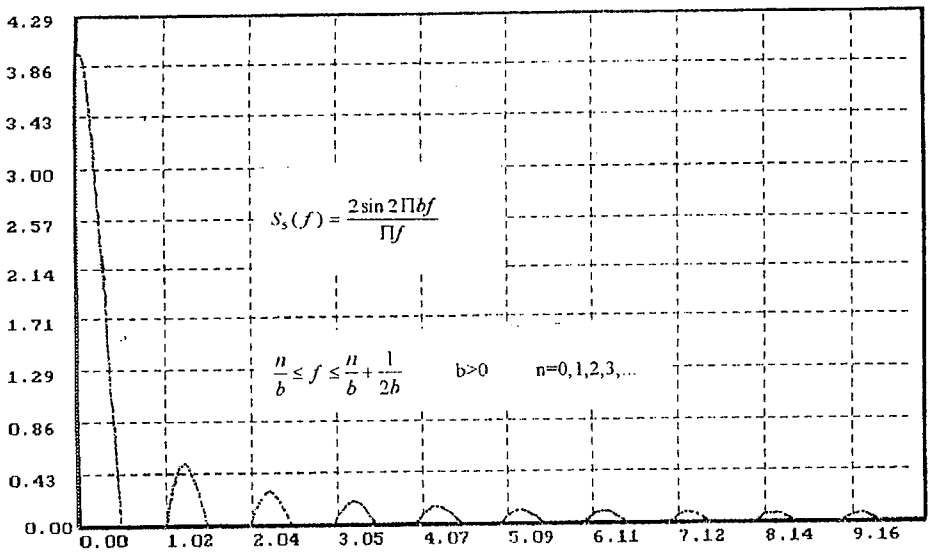


Fig. 7 — S5(f)

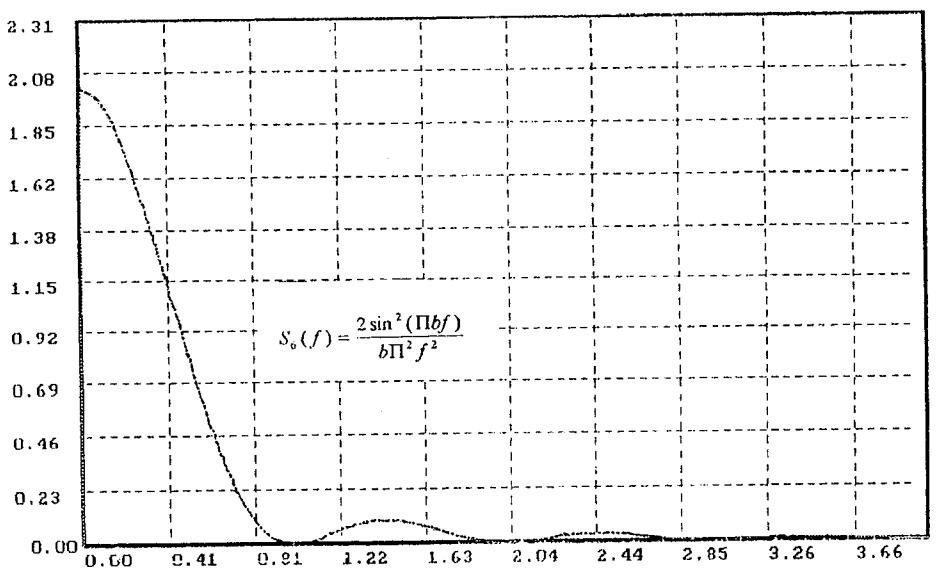


Fig. 8 — S6(f)

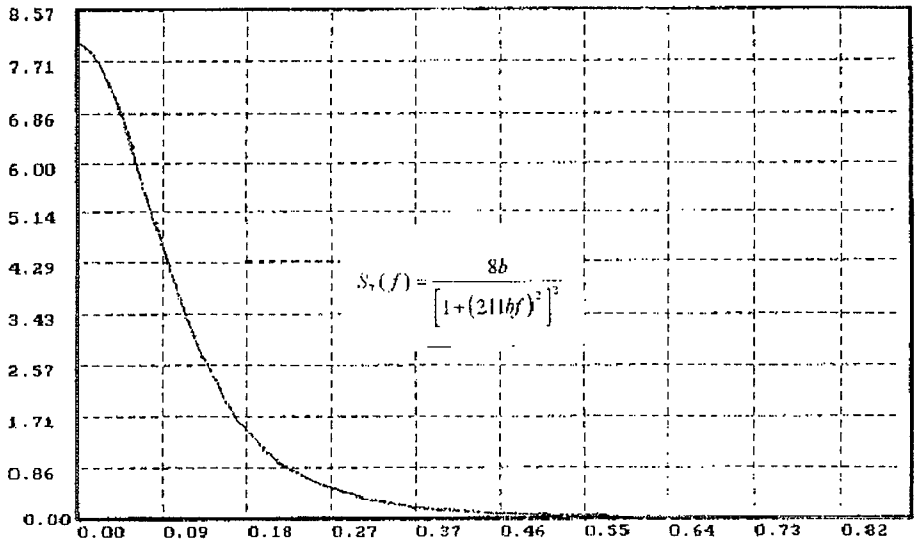


Fig. 9 — S7(f)

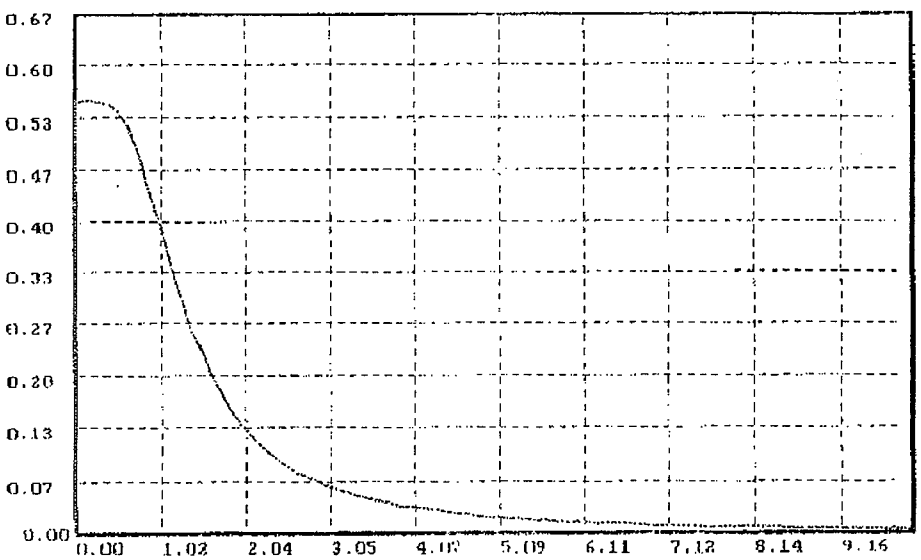


Fig. 10 — S8(f)

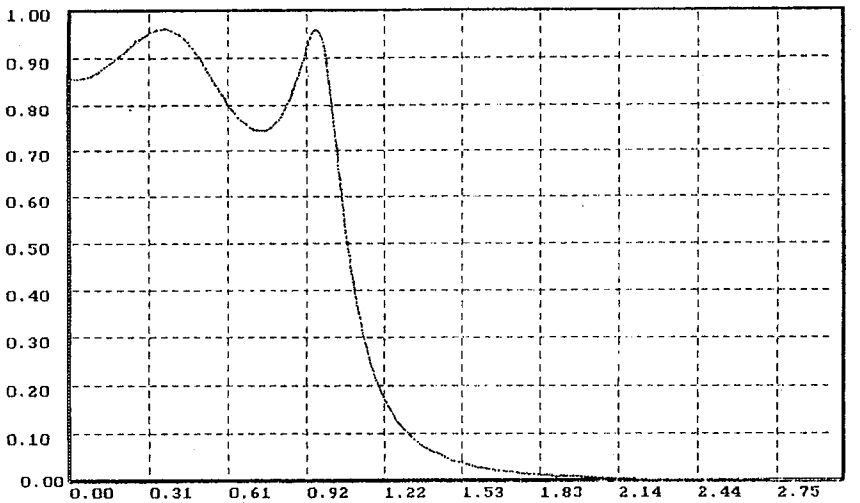


Fig. 11 — S9(f)

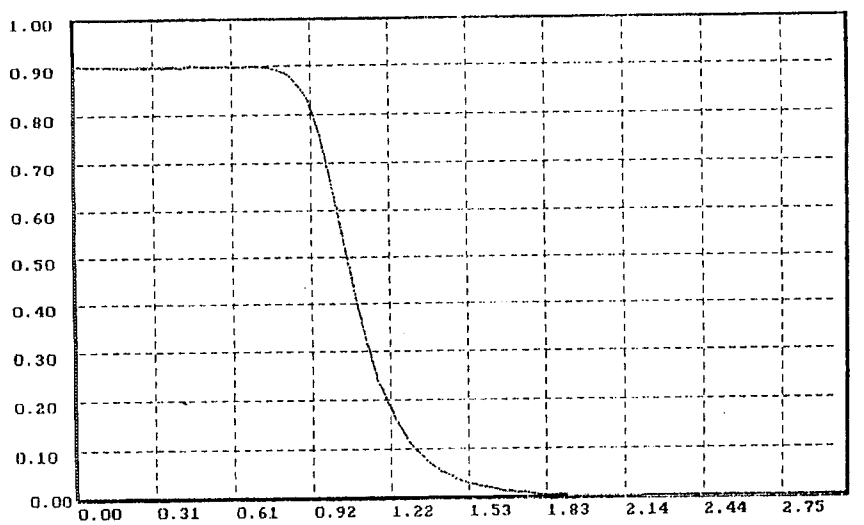


Fig. 12 — S10(f)

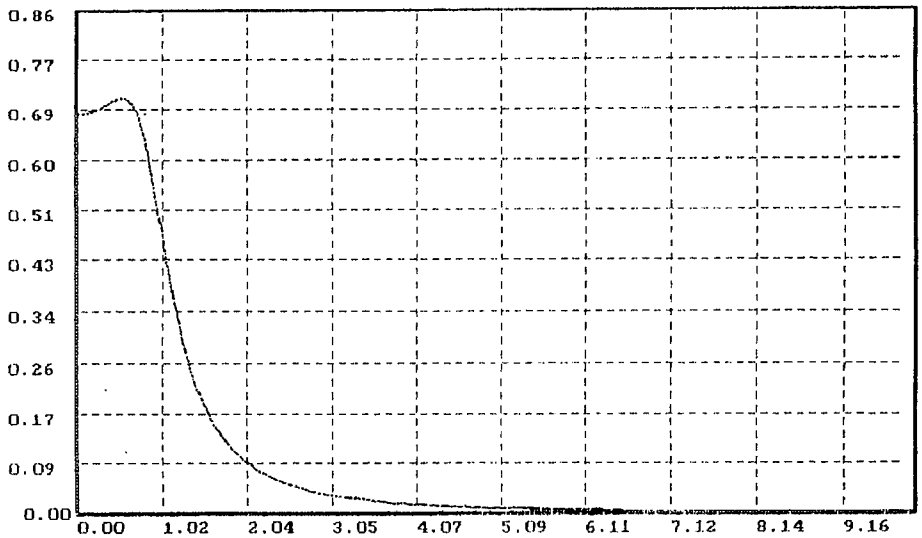


Fig. 13 — S11(f)

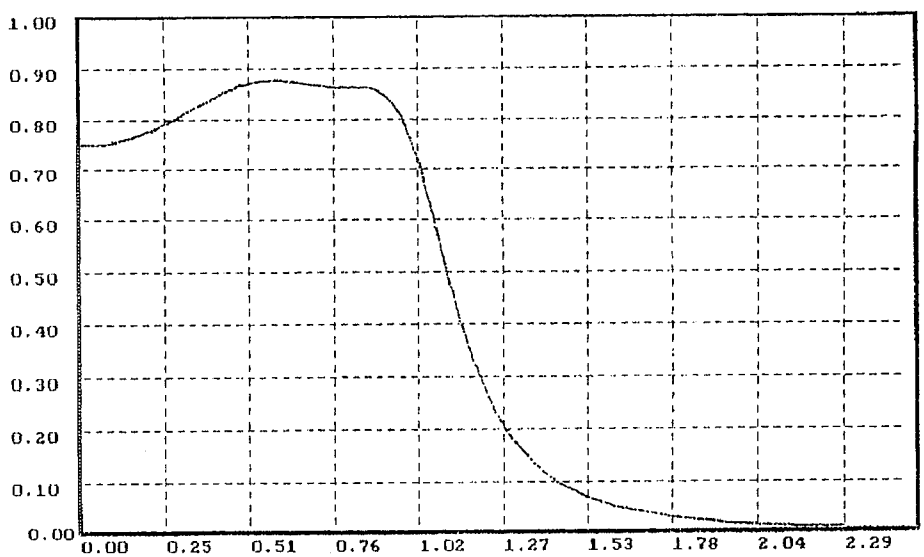


Fig. 14 — S12(f)

Table 5

The results of $S_1 \div S_7$ bandwidths calculation according criterions 1 ÷ 6
Letters a, b, c, d, ... in sequence of band narrowing

b = 1 in formulas S (f)	Criterion 1: „normalized spectrum moment"	Criterion 2: „unit area in the band"	Criterion 3: „drop to 3 dB"	Criterion 4: „90% in the band"	Criterion 5: „99% in the band"	Criterion 6: „informa-tional W_0"
S_1	∞ a	0.2500 b	0.1600 d	0.9580 a	9.5540 b	0.4600 b
S_2	0.0563 e	0.1410 d	0.1400 e	0.1830 g	0.2890 g	0.2000 e
S_3	0.1125 c	0.1592 c	0.1200 f	0.3590 d	0.7280	0.3200 c
S_4	0.1443 b	0.5000 a	0.5000 a	0.4480 c	0.4950 e	0.5000 a
S_5	–	0.2500 b	0.3100 c	0.2590 e	0.3020 f	0.3000 d
S_6	∞ a	0.5000 a	0.4500 b	0.7920 b	9.5890 a	0.0100 g
S_7	0.0788 d	0.1250 e	0.1100 g	02120 f	0.5310 d	0.1900 f

[435]

Table 6

The results of $S_8 \div S_{12}$ bandwidths calculation according criterions 1 ÷ 6
Letters a, b, c, d, ... in sequence of band narrowing

b = 1 in formulas S (f)	Criterion 1: „normalized spectrum moment"	Criterion 2: „unit area in the band"	Criterion 3: „drop to 3 dB"	Criterion 4: „90% in the band"	Criterion 5: „99% in the band"	Criterion 6: „informa-tional W_0"
S_8	∞ a	1.8000 a	1.3200 a	4.1090 a	13.4030 a	2.6500 a
S_9	0.3498 d	1.0422 e	1.0800 d	1.0350 d	1.5600 d	1.2000 d
S_{10}	0.3445 e	1.1100 d	1.0800 d	1.0240 e	1.5420 e	1.1900 e
S_{11}	2.7271 b	1.4203 b	1.1700 b	2.1240 b	8.9790 b	1.8000 b
S_{12}	0.3908 c	1.1406 c	1.1300 c	1.1080 c	1.6900 c	1.2900 c

REFERENCES

1. B. Wołczak: *Porównanie rozmaitych kryteriów oceny zmian szerokości pasma losowych sygnałów.* Archiwum Elektrotechniki, 1968, vol. 17, z. 3, s. 499–514
2. B. Wołczak: *Antinomies of different criteria for evaluation bandwidths in images and laser measurements.* Laser Technology IV: Research Trends Instrumentation, and Applications in Metrology and Materials Processing. W. Woliński, Z. Jankiewicz, Editors: Proc. SPIE 1995, vol. 2203, s. 582–587
3. B. Wołczak: *Nowe aspekty — antynomie — detekcji krawędzi obrazów i mikrometrii laserowej.* Kwartalnik Elektroniki i Telekomunikacji, 1996 vol. 42, z. 3, s. 343–357
4. M. Pluta: „Mikrometria laserowa. V Krajowa Szkola Optoelektroniki „Metrologia laserowa". Waplewo 14–18 października 1991 część 2
5. B. Smolka: *Zastosowanie rozkładu kanonicznego Gibbsa w detekcji krawędzi.* Zeszyty Naukowe Politechniki Śląskiej Nr 120, Gliwice 1997

B. WOŁCZAK

OCENA SZEROKOŚCI PASM W ANALIZIE WIDMOWEJ I MIKROMETRII

Streszczenie

Celem pracy jest przedstawienie nowych ograniczeń dla niektórych często stosowanych kryteriów oceny i (lub) sposobów pomiarów szerokości pasm (szerokości mikrolinii) oraz porównania uzyskanych rezultatów. Podano wyniki porównań szerokości pasm (linii) obliczonych dla dwunastu typowych sygnałów według sześciu definicji szerokości pasm i 12 współczynników dla „kryterium równych pól". Pokazano, że problem oceny lub (i) pomiarów szerokości pasm nie może być jednoznacznie rozwiązany przy zastosowaniu jakiejkolwiek pojedynczej („uniwersalnej") definicji szerokości pasma, ponieważ istnieją przypadki, dla których antynomie pomiędzy poszczególnymi kryteriami oceny zmian szerokości pasm są bardzo wyraziste.

Słowa kluczowe: antynomie, szerokość spektralna, mikrometria, pomiary mikrolinii

INFORMACJE DLA AUTORÓW

Redakcja przyjmuje do publikowania prace oryginalne, przeglądowe i monograficzne wchodzące w zakres szeroko pojętej elektroniki. Ponieważ KWARTALNIK ELEKTRONIKI i TELEKOMUNIKACJI jest czasopismem Komitetu Elektroniki i Telekomunikacji Polskiej Akademii Nauk, w związku z tym na jego łamach znajdują się prace naukowe dotyczące podstaw teoretycznych i zastosowań z zakresu elektroniki, telekomunikacji, mikroelektroniki, optoelektroniki, radiotechniki i elektroniki medycznej.

Artykuły powinno charakteryzować oryginalne ujęcie zagadnienia, własna klasyfikacja, krytyczna ocena (teorii lub metod), omówienie aktualnego stanu, lub postępu danej gałęzi techniki oraz omówienie perspektyw rozwojowych.

Artykuły publikowane w innych czasopismach nie mogą być kierowane do druku w Kwartalniku Elektroniki i Telekomunikacji w drugiej kolejności zgłoszenia.

Objętość artykułu nie powinna przekraczać 30 stron po około 1800 znaków na stronie.

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

Sposób pisania tekstu: Tekst powinien być pisany bez używania wyróżnień, a w szczególności nie dopuszcza się spacjowania, podkreślania i pisania tekstu dużymi literami z wyjątkiem wyrazów, które umownie pisze się dużymi literami (np. FORTRAN). Proponowane wyróżnienia Autor może zaznaczyć w maszynopisie zwykłym ołówkiem za pomocą przyjętych znaków adiustacyjnych) np. podkreślenie linią przerywaną oznacza spacjowanie (rozstrzelenie), podkreślenie linią ciągłą — pogrubienie, podkreślenie wężykiem — kursywa. Tekst powinien być napisany z podwójnym odstępem między wierszami, tytuły i podtytuły małymi literami. Marginesy z każdej strony powinny mieć około 35 mm. Przy podziale pracy na rozdziały i podrozdziały cyfrowe ich oznaczenia nie powinny być większe niż III stopnia (np. 4.1.1).

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

Sposób pisania wzorów matematycznych: Rozmieszczenie znaków, cyfr, liter i odstępów powinno być zbliżone do rozmieszczenia elementów druku. Wskaźniki i wykładniki potęg powinny być napisane wyraźnie i być prawidłowo obniżone lub podwyższone w stosunku do linii wiersza podstawowego. Znaki nad literami i cyframi, całkami i in. symbolami (strzałki, linie, kropki, daszki) powinny być umieszczone dokładnie nad tymi elementami, do których się odnoszą. Numery wzorów cyframi arabskimi powinny być kolejne i umieszczone w nawiasach okrągłych z prawej strony. Nazwy jednostek, symbole literowe i graficzne powinny być zgodne z wytycznymi IEE (International Electronical Commision) oraz ISO (International Organization of Standarization).

Powołania: Powołania na publikacje powinny być umieszczone na ostatnich stronach tekstu pod tytułem „Bibliografia", opatrzone numeracją kolejną bez nawiasów. Numeracja ta powinna być zgodna z odnośnikami w treści artykułu. Przykłady opisu publikacji:

— periodycznej: F. Valdoni: A new millimetre wave satelite. E.T.T. I990, vol. 2, nr 5, p. 553

— nieperiodycznej: K. Andersen: A resource allocation framework. XVI International Symposium. Stockholm (Sweden), May 1991, paper A 2.4

— książki: Y.P. Tvidis: Operation and modelling of the MOS transistors. Mc Grâw-Hill, New York 1987, p. 141—148

Materiał ilustracyjny: Rysunki powinny być wykonane wyraźnie, na papierze gładkim, lub milimetrowym w formacie nie mniejszym niż 9×12 cm. Mogą być także w postaci wydruku komputerowego. Fotografie lub diapozytywy przyjmowane są raczej czarno-białe w formacie nie przekraczającym 10×15 cm. Na marginesie każdego rysunku i na odwrocie fotografii powinno być napisane ołówkiem imię i nazwisko Autora oraz skrót tytułu artykułu, do którego są przeznaczone. Spis podpisów pod rysunki i fotografie powinien być umieszczony na oddzielnej stronie.

Streszczenia: Do każdego artykułu musi być dołączone obszerne — do 60 wierszy — streszczenie z tytułem artykułu. Streszczenia (Summary) muszą być w języku polskim i angielskim. Streszczenie powinno wyjaśniać główny cel pracy, wskazywać korzyści i ograniczenia, możliwe zastosowania i zalecenia dla dalszego rozwoju danej gałęzi techniki. Pod streszczeniami powinny być podane słowa kluczowe.

Autorowi przysługuje bezpłatnie 20 odbitek artykułu. Dodatkowe egzemplarze odbitek, lub cały zeszyt Autor może zamówić u wydawcy na własny koszt.

Autora obowiązuje korekta autorska, którą powinien wykonać w ciągu 3 dni od daty otrzymania tekstu z Redakcji i zwrócić osobiście, lub listownie pod adresem Redakcji. Korekta powinna być naniesiona na przekazanych Autorowi szpaltach na marginesach ew. na osobnym arkuszu w przypadku uzupełnień tekstu większych niż dwa wiersze. W przypadku nie zwrócenia korekty w terminie, korektę przeprowadza Redakcja Techniczna Wydawcy.

Redakcja prosi Autorów o powiadomienie ją o zmianie miejsca pracy i adresu prywatnego.

INFORMATION FOR THE AUTHORS OF K.E.T.

The KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI (K.E.T.) — ELECTRONICS AND TELECOMMUNICATION QUARTERLY is a journal of the Committee for Electronics and Telecommunications of the Polish Academy of Sciences and is a direct continuation for 43 years of the quarterly ROZPRAWY ELEKTROTECHNICZNE, that has been published also under the auspices of the Polish Academy of Sciences.

Its pages contain scientific articles concerned with theory and applications of electronics, telecommunication, microelectronics, optoelectronics, radioengineering and medical electronics.

Accordingly, its pages contain scientific articles, that are original works both contributory and concerned with comparative analyses of methods or systems synthetic reviews of a given field, state-of-art discussions of a given field, and presentations of developments in a given technology.

Regular papers are accepted for publication in Polish or English, but title, the extensive summary and figure caption must be bilingual. Bilingual summary must shown the significance of the results shown in the paper, emphasizing limitations and advantages, possible applications and recommendations for further works.

K.E.T. is meant for specialists and students with above subjects.

FORMAL PREPARATION OF MANUSCRIPT

Texts

Texts to be submitted to K.E.T. may be typed on one side of the paper only with double spacing between the lines and a 4 cm margin or uswing of wordprocessing with text formatters based on MS DOS, APPLE/MAC such as Microsoft word. In any case the text formatter must be specified on the diskette. Each text should be preceded by: a sheet with paper title, name (in full) and surname of the Author/s, their Affilitions and relevant addresses.

All text pages should be numbered. Regular papers should have a lenght of no more than 30 pages.

Printing types and symbols

For letteral symbols, units and graphical symbols, the recommendations of IEE (International Electrotechnical Commission) and ISO (International Organisation of Standarisation) must be followed.

- In case of manuscripts conventionally typed (i.e. without wordprocessing), letters to be printed in italic should be underlined; letters to be printed in bold should be double interlined.
- Formulae and footnotes should be numbered in accordance with the quotation order followed in the text.
- The serial number of each formula should be written at right side, between round brackets.
- The serial numberr of footnotesw should be written between round brackets and upper.
- Mathematical details, anciliary to the main discussion of the paper may be included in one or more appendixes, that are printed after conclusions and before references.

In manuscript containing lot of symbols, a glossary may be included after introduction.

References

References are usually gathered at the end of the text and numbered according to the order of quotation in the text. The serial numbers should be written without square brackets (i.e. 1., 2. ...)

Illustrations

The originals of illustrations may be either drawings or photographs (black or colour). Each copy of paper must include two complete set of illustrations.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1997, 43, z. 3

SPIS TREŚCI

CONTENTS