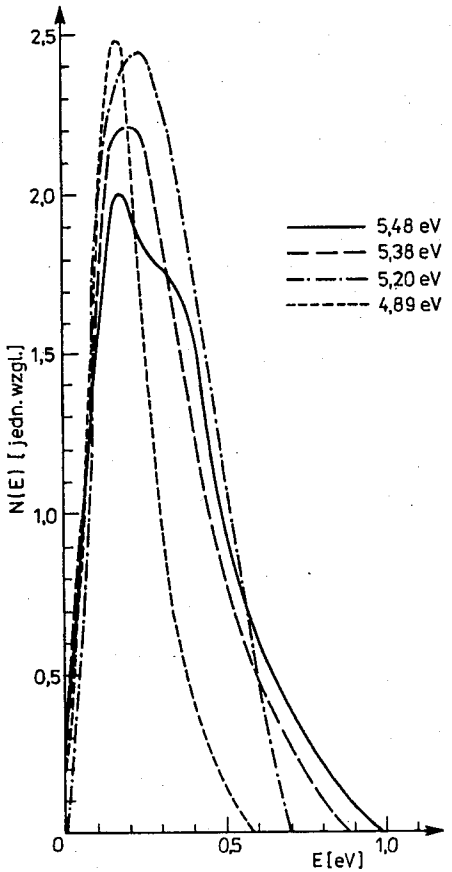
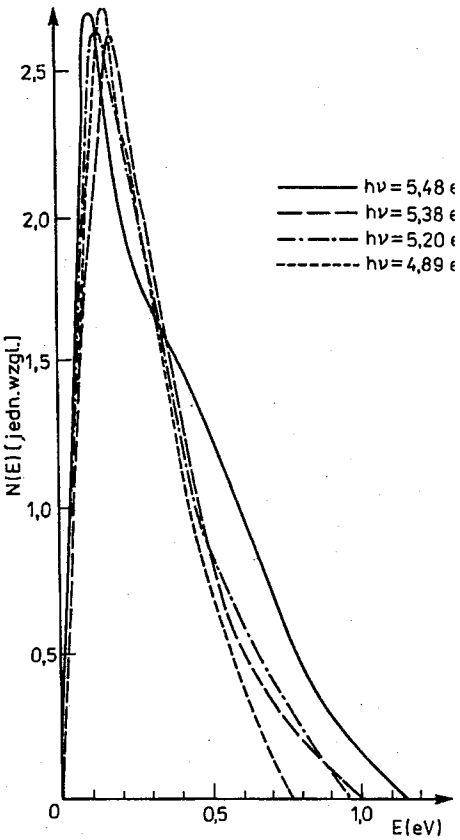
GaAs (100) i (110) typu n (rys. 7 i 8) ze wzrostem energii *hv* fotonów, świadczy (za Fischerem i innymi [29] [30]) o dominacji przejść skośnych w fotoemisji z tych próbek, co jest potwierdzeniem faktu dopasowania widm wydajności w pobliżu progu do wykładnika 3. Gdyby przeważały przejścia proste wykładnik *m* byłby zbliżony do jedności.



Rys.7. Rozkład energetyczny fotoelektronów dla próbki GaAs(100) typu n



Rys.8. Rozkład energetyczny fotoelektronów dla próbki GaAs(110) typu n

Na rys. 2a i b podane są reprezentatywne krzywe rozkładu energetycznego dla Si + Cs. Te krzywe mają kilka interesujących cech. Dla fotonów o energiach mniejszych niż około 3,0 eV krzywe są z grubsza gaussowskie w kształcie i przesuwają się do wyższej energii w miarę wzrostu *hv*. Jednakże, dla energii fotonów większych niż 3,0 eV, przy której zaczyna się pojawiać pierwsza silna absorpcja optyczna, rozkład po stronie niskoenergetycznej maksimum zmienia się bardzo mało w miarę wzrostu *hv*. Sugeruje to, że najniższe stany energetyczne dla wchodzących

w grę przejść optycznych są ustalone dla $h\nu > 3{,}0$ eV i nie zmieniają się znacznie w miarę wzrostu $h\nu$. Dla energii między 3,0 a 3,4 eV rozkład jest ciągle z grubsza gaussowski, ale ilości elektronów o wyższej energii zwiększają się ze wzrostem $h\nu$. Dla $h\nu > 3{,}4$ eV rozkład wydaje się być złożonym z dwu składników; jeden jest gaussowskim uzyskanym dla $h\nu = 3{,}4$ eV (scentrowany przy 1,0 eV), a drugi jest gaussowski przy wyższych energiach. Drugi rozkład osiąga swój ostateczny kształt (scentrowany przy 2,3 eV) dla energii fotonu 5,4 eV. Takich charakterystyk należy oczekiwać, jeżeli dominujące przejścia optyczne przyczyniające się do fotoemisji są przejściami do dwu grup pasma przewodnictwa [31]. Fakt, że każdy rozkład charakterystyczny osiąga swój kształt końcowy przy energii fotonu odpowiadającej mocnemu pikowi absorpcyjnemu (rys. 1) umożliwia jednoznaczne przypisania każdego rozkładu przejściu optycznemu: rozkładu scentrowanego przy 1,0 eV przejściu 3,4 eV, a rozkładu scentrowanego przy 2,3 eV przejściu 5,4 eV.

Z takich danych można uzyskać wartości absolutne energii stanów wchodzących w grę, jeśli będą małe zagięcia pasm i straty energii, gdyż czynniki te mają tendencję do „*rozmazania*" rozkładów. Ponieważ przejścia optyczne zachodzą między pasmami a nie stanami dyskretnymi i ponieważ poszerzenie związane z czasem życia nośników może zapobiec całkowitej ostrości krawędzi absorpcji, istnieje inne możliwe źródło błędu w łączeniu maksimum absorpcyjnego z odległością między ekstremami dwu stanów [32]. Jednakże, im bardziej płaskie i mniej skomplikowane są pasma w punkcie przejścia, tym bardziej jest prawdopodobne, że pik w absorpcji optycznej będzie się pokrywał z ekstremami pasm, a także że pik na krzywej rozkładu energetycznego będzie się pokrywał z ekstermami pasmowymi. Dla przejścia L_3, do L_3, pasma dla stanu początkowego, jak też i końcowego są względnie płaskie i są dobrze oddzielone od innych pasm [33], tak więc najlepsze dane odnośnie wartości absolutnych stanu końcowego powinny być uzyskane dla przejścia (L_3, do L_3) 5,4 eV.

W tablicy 2 przedstawione są wartości energii absolutnej (odniesionej do wierzchołka pasma walencyjnego) poziomów energetycznych uzyskanych doświadczalnie. Dla porównania są także podane wartości teoretyczne Phillipsa [34]. Tylko wartości dla przejść L_3, do L_3 i Γ_{25}, do Γ_{15} są uzyskane wyłącznie z danych rozkładu energetycznego. Możliwy błąd oparty na niepewności zera energii na krzywych rozkładu prędkości i powinowactwa elektronowego oceniono dla tych wartości na $\pm$ 0,2 eV.

Jak należy oczekiwać, zgodność pomiędzy doświadczeniem a teorią jest bardzo dobra dla stanów L_3 do $L_{3'}$. Widoczna niezgodność dla przejścia $\Gamma_{25'}$ do Γ_{15} jest znacznie większa od tej powodowanej nieokreślonością powinowactwa elektronowego i zera energii dla rozkładu energii lub innymi trudnościami doświadczalnymi. Główna przyczyna widocznej niezgodności staje się jasna, jeżeli się przebada diagramy E w zależności od k dla krzemu [35] [36]. W pobliżu punktu Γ zarówno najniższe pasmo przewodnictwa jak też najwyższe pasmo walencyjne są praktycznie równoległe i pochyłe w dół. Tak więc, dla $h\nu = 3{,}4$ eV będą powodowane przejścia do niżej leżących stanów pasma przewodnictwa jak też do stanów w punkcie Γ_{15}.

Przesuwa to maksimum krzywej rozkładu energetycznego do energii niższej jak to pokazuje rys. 2a.

Wykorzystując pomiary Spicera i Simona [19] można także nałożyć granice na energie stanów biorących udział w przejściach 3,7 i 4,4 eV. Ponieważ przejście 4,4 eV daje bardzo silne minimum na krzywej rozkładu widmowego (rys. 1), więc stan końcowy musi leżeć poniżej poziomu próżniowego, nakładając górną granicę 0,4 eV na energię stanu końcowego. Jest to w dobrej zgodności z przypisaniem tego piku przejściu $X_4 \rightarrow X_1$. Wykazano, że dane doświadczalne umiejscawiają stan początkowy dla przejścia optycznego 3,7 eV blisko (energetycznie) stanu L_3.

Pojawienie się dobrze określonego piku na rys. 1c odpowiadającego przejściu oznaczonemu zarówno od L_3, do L_1, jak i przejściu Λ_3 do Λ_1, a więc w zgodności z każdym oznaczeniem.

2.3. WIDMA BEZWZGLĘDNYCH WYDAJNOŚCI KWANTOWYCH GaAs

W celu wyznaczenia fotoelektrycznych progów rozkłady wydajności próbek GaAs ekstrapolowano funkcjami $\sqrt[3]{Y} = f(h\nu)$. Poszukiwano dopasowania doświadczalnych danych do funkcji matematycznej, przewidując ogólną zależność wydajności od $h\nu$ w postaci

$$Y = k(h\nu - h\nu_t)^m, \tag{6}$$

gdzie $h\nu_t$ – próg fotoelektryczny,

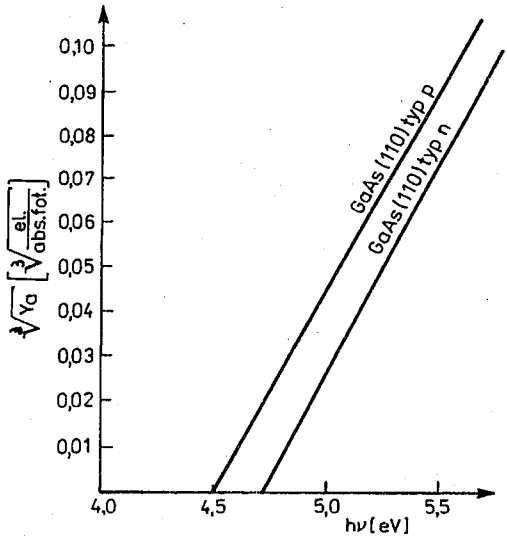
m – wykładnik potęgowy, który powinien być zbliżony do jednej z wartości podanych przez Kane [7].

Podobnie jak w pracy Gobeli i Allena [24] dane doświadczalne autora wykazały lepsze dopasowanie do potęgi 3 niż do wynikającej z teorii Kane dla przejść skośnej wartości 5/2. Dla sprawdzenia dopasowania trzeciej potęgi do swoich wyników autor zastosował następującą metodę:

Iloraz wartości Y przez jej pochodną Y' jest funkcją zmiennej $h\nu$ i po uproszczeniu wyraża się wzorem

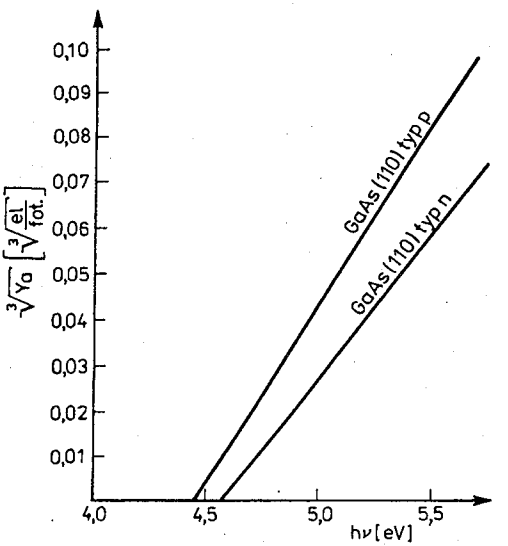
$$\frac{Y}{Y'} = \frac{h\nu - h\nu_t}{m}. \tag{7}$$

Jeśli m jest stałe, wykres $Y/Y' = f(h\nu)$ jest prostą odcinającą wartość $h\nu_t$ na osi $h\nu$. Z równania (7) łatwo jest określić wykładnik m, co uczyniono dla kilkunastu dowolnie wybranych charakterystyk wydajności próbek GaAs oraz otrzymano wartości m równe 2,8 ÷ 3,1. Jako średnią przyjęto więc wartość 3. Wykładnik ten daje (jak widać z rysunków 9 i 10 bardzo dobre dopasowanie do danych doświadczalnych w przedziale energii fotonów (4,7 ÷ 5,5) eV. Odchylenia występujące dla obu orientacji (100) i (110) przy mniejszych $h\nu$, autor tłumaczy emisją ponadwalencyjnych elektronów, której udział w całkowitej emisji rośnie przy zmniejszaniu się $h\nu$. Punkty przecięcia się prostych $\sqrt[3]{Y}$ z osiami $h\nu$ przyjęto za fotoelektryczne progi



Rys.9. Pierwiastek trzeciego stopnia z rozkładów widmowych bezwzględnych wydajności kwantowych w funkcji energii fotonów dla próbek typu p i n GaAs(100)

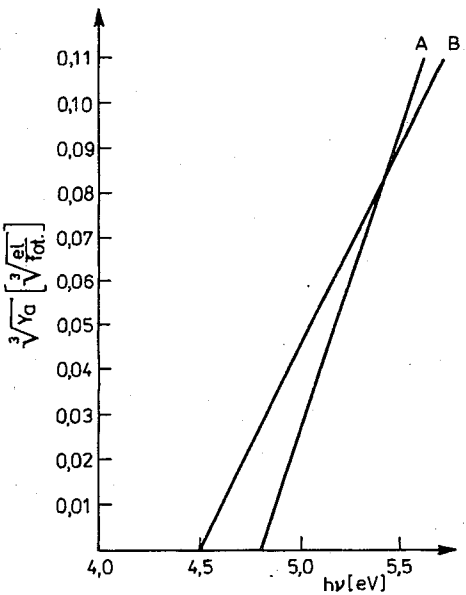
emisji. Z ekstrapolacji czterech rozkładów wydajności wyznaczono następujące wartości progów: $h\nu_{t(100)} = 4{,}70$ eV, i $h\nu_{t(110)} = 4{,}55$ eV dla próbek typu n oraz $h\nu_{t(100)} = h\nu_{t(110)} = 4{,}45$ eV dla próbek typu p (przedziały błędów wszystkich wartości $\pm$ 0,05 eV). W stosowanym zakresie energii fotonów wydajności badanych



Rys.10. Pierwiastek trzeciego stopnia z rozkładów widmowych bezwzględnych wydajności kwantowych dla próbek GaAs(110) typu p i n

czterech typów warstw przypowierzchniowych GaAs są rzędu ($10^{-4} \div 10^{-3}$) elektronów /abs. foton.

Z porównania wydajności wygrzewanych w wysokiej próżni próbek typu n i p wynika, że w przedziale fotonów (4,7 – 5,5) eV wydajności próbek typu p są co najmniej 1,5 krotnie wyższe. Zdaniem autora wpływają na to nieco niższe fotoelektryczne progi próbek typu p oraz mniejsze rozproszenia pobudzonych elektronów w pasmie przewodnictwa w próbkach typu p.



Rys.11. Zależność wydajności kwantowej Y od energii fotonów $h\nu$: (krzywa A dla GaAs(111A); Krzywa B dla GaAs(111B)

W odróżnieniu od rozkładów $Y = f(h\nu)$ powierzchni o orientacji (100) i (110), widma wydajności kwantowych powierzchni (111A) i (111B) GaAs bardziej się różnią między sobą. Jak widać z rys. 11, otrzymane z ekstrapolacji uśrednionej prostej $\sqrt[3]{Y} = k(h\nu)$ wartość fotoelektrycznego progu dla powierzchni (111A) jest większa o 0,3 eV od $(h\nu_t)_{111B}$. Wyższy jest także „*ogon*" emisji ponadwalencyjnej z powierzchni (111A), co znajduje uzasadnienie w energetycznej szerokości i stopniu zapełnienia elektronami pasma stanów związanych z tą powierzchnią.

Z rys. 11 widać, że proste $\sqrt[3]{Y}_{(111A)}$ ma większe nachylenie do osi $h\nu$, czyli wydajność fotoelektronów z tej powierzchni szybciej rośnie ze wzrostem $h\nu$, pomimo większej wartości fotoelektrycznego progu. Zdaniem autora ma to związek z rozpraszaniem pobudzonych elektronów przed ich emisją. Do oszacowania wpływu rozpraszania fotoelektronów na charakterystyki fotoemisji winny być znane: liczbowe parametry rozproszenia na fononach, rozproszenie elektron–elektron oraz

prawdopodobieństwa przejścia przez powierzchnię. Z analizy różnych parametrów powierzchni wynika, że efekt rozproszeń elektron–elektron w obszarze ładunku przestrzennego jest niewielki (przeważają rozproszenia elastyczne). Co do rozpraszania na fononach, to według [37] w związkach III–V w zakresie wartości hv: $(4 \div 6)$ eV średnia droga na to rozproszenie jest około 60 Å – ta sama co głębokość ucieczki fotoelektronów w GaAs. Ponadto wpływ rozproszeń na fononach na krzywe $Y = f(hv)$ byłby stały wzdłuż jednej orientacji w krysztale (w tym przypadku w kierunku [111]). Zatem oddziaływanie elektronów z fononami nie może tłumaczyć ostrzejszego wzrostu $Y = f(hv)$ z powierzchni (111A) niż z (111B). Zdaniem autora decydującymi są tu rozproszenia na samych powierzchniach. Wg Kane [37] element macierzowy P, opisujący prawdopodobieństwo ucieczki elektronów zależy od „*początkowych*" energii tych pobudzonych elektronów, które zjawiają się w obszarze ładunku przestrzennego oraz od ich stanów końcowych. Wykres funkcji $P = f(hv)$ obliczonej w przedziale $(4 \div 6)$ eV dla krzemu wykazuje ostry wzrost między 4 a 5 eV [35]. Można się spodziewać, że dla GaAs, którego właściwości fotoemisyjne są zbliżone do Si wzrost ten także istnieje. Ponieważ średnie drogi rozproszeń w tym niezbyt szerokim przedziale hv nie ulegają dużym zmianom, za wzrost jest odpowiedzialne prawdopodobieństwo T przejścia fotoelektronu przez powierzchnię. Silniejszy wzrost wydajności fotoelektrycznej z powierzchni (111A) niż z (111B) jest spowodowany również tym, że powierzchnie GaAs (111A) są o wiele bardziej chropawe niż powierzchnie GaAs (111B). W mikroskopie dają one obrazy „*księżycowe*" ze wzgórkami i „*kraterami*". Stąd rozproszenia nawet na wewnętrznych stronach tych powierzchni są zapewne znaczne. Z kolei przy porównaniu otrzymanych przez autora rozkładów $N(E) = f(E)$ z powierzchni (111A) i z (111B) widać, że pierwsze z nich mają bardziej rozbudowane wysokoenergetyczne części niż te, które otrzymano z powierzchni arsenowych. To też wpływa na silniejszy wzrost wydajności $Y_{(111A)}$ przy powiększaniu hv właśnie w zakresie $(4{,}5 \div 6)$ eV.

3. FOTOEMISYJNE BADANIA GĘSTOŚCI STANÓW KWANTOWYCH PASM WALENCYJNYCH I POWIERZCHNI

Gęstość stanów elektronowych N, czyli liczba Z stanów kwantowych przypadających na jednostkę energii w jednostce objętości jest funkcją energii E. Gęstość stanów definiuje się jako pochodną:

$$N(E) = \frac{dZ(E)}{dE}. \tag{8}$$

Badania fotoemisji zewnętrznej dostarczają sporo danych o gęstości stanów kwantowych pasm energetycznych z których emitowane są elektrony. Krzywe gęstości stanów dają więc informacje o strukturze górnych części pasm walencyjnych. Kinding i Spicer [6] podali sposób oszacowania energetycznego rozkładu gęstości stanów w tej części pasma, z której występuje fotoemisja posługując się przy tym rozkładami fotoelektronów i wartościami wydajności kwantowych.

Metoda ta opiera się na wstępnym założeniu, do którego należy wprowadzić wielkość fizyczną $\alpha'(\mathscr{E}, h\nu)$, którą możemy nazwać gęstością absorpcji optycznej $\frac{\partial\alpha}{\delta\mathscr{E}}$ wyrażoną w jednostkach $[\mathrm{eVcm}^3]^{-1}$.

To założenie jest następujące:

$$\frac{N(E, h\nu)}{Y(h\nu)} = \frac{\alpha'(\mathscr{E}, h\nu)}{(h\nu)} T(E), \tag{9}$$

gdzie $\mathscr{E}$ jest energią w emitującym pasmie (walencyjnym); E – poza pasmem; Y – wydajnością fotoemisji; $T(E)$ – jest funkcją ucieczki elektronów z kryształu i oznacza prawdopodobieństwo przejścia quasi swobodnych elektronów znajdujących się blisko powierzchni przez barierę powierzchniową. Dla energii fotonów $h\nu$ dość odległej od długofalowego progu $T(E) - 1$.

Opiera się ono na spostrzeżeniu, że fotony powodujące emisję elektronów z części pasma walencyjnego o energetycznej szerokości $\delta\mathscr{E}$ są pochłaniane selektywnie w funkcji energii $\mathscr{E}$ ich stanów kwantowych oraz, że maksimum α' w przedziale $\delta\mathscr{E}$ pokrywa się z maksimum krzywej $N(E)$. (Jest to słuszne w założeniu małych strat energii przy rozpraszaniu pobudzonych elektronów – w czasie ich wędrówki do powierzchni kryształu).

Metoda ta wraz z modyfikacją autora została przedyskutowana w pracy [38]. Autor badając fotoemisję z GaAs, wykorzystał rozkłady względnej ilości fotoelektronów w funkcji ich energii kinetycznej do obliczenia rozkładów gęstości stanów (wypełnionych elektronami) w emitujących pasmach tego półprzewodnika. Stosując zmodyfikowaną metodę autor otrzymał dla GaAs typu n (rys. 12a) zależności względnej gęstości stanów kwantowych d_v w funkcji energii $\mathscr{E}$ w pasmach walencyjnych; gęstość ta została określona względem wierzchołków tych pasm. Na rysunku tym, krzywa A jest uśrednioną dla kilku próbek GaAs (100) typu n, a krzywa B jest uśrednioną dla kilku próbek GaAs (110) typu n. Próbki tej samej orientacji i tego samego typu prawie nie różniły się między sobą koncentracją elektronów.

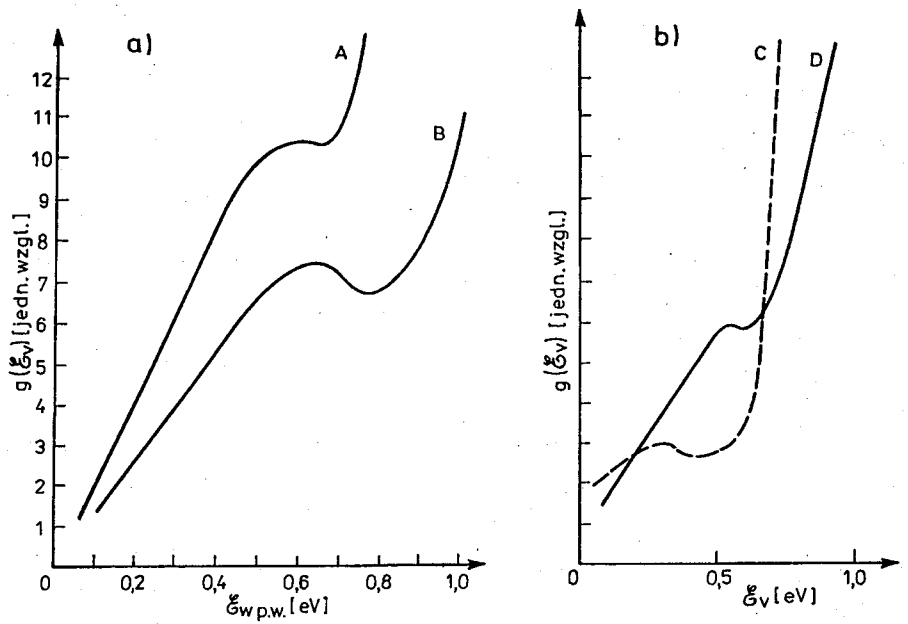
Dla dużo większej ilości próbek o orientacjach powierzchni (111A) i (111B) wyznaczono po jednej uśrednionej krzywej (rys. 12b) krzywe C i D). Przebieg krzywej gęstości dla próbek (111B) jest podobny do analogicznych krzywych z próbek o orientacjach (100) i (110) z tą próbką, że występujący w pobliżu szczytów pasm walencyjnych „*garb*" gęstości stanów jest tutaj nieco słabiej widoczny. Jego położenie energetyczne [$\mathscr{E} = (0{,}5 - 0{,}6)$ eV] jest jednak stałe dla wspomnianych trzech orientacji. Omawiany garb, zdaniem autora jest spowodowany nałożeniem się dodatkowych stanów energetycznych na pasma walencyjne GaAs w pobliżu ich szczytów. Podobny efekt obserwowali Eastman i Grobman [39] dla powierzchni GaAs (110) i przypisali go pasmu stanów powierzchniowych o szerokości około 1 eV nałożonemu na pasmo walencyjne GaAs.

Z przeglądu literatury światowej dotyczącej tych zagadnień wynika, że stosowanie energii fotonów w przedziale $6 \text{ eV} < h\nu < 12 \text{ eV}$ daje wzbudzanie elektronów do

obszaru przykrawędziowego pasma przewodnictwa co pozwala na uzyskanie większych prądów fotoemisji ze względu właśnie na dużą gęstość stanów w tym obszarze pasma przewodnictwa [38].

Krzywa rozkładu energii fotoelektronów EDC dla dowolnych wartości w wymienionym ostatnio obszarze energii dostarczają bardziej szczegółowych informacji o rozkładzie gęstości stanów w pasmie przewodnictwa N_c i pasmie walencyjnym N_v [21], [41–43]. Krzywe gęstości stanów dla (111A) mają nieco inny przebieg w pobliżu szczytów pasm walencyjnych. Są one prawie płaskie w przedziale energii $\mathscr{E}_v - \mathscr{E}_{v_0} = 0{,}6$ eV. ($\mathscr{E}_{v_0}$ – energia szczytu pasma walencyjnego). Jeśli zatem energie $h\nu$ nie są wystarczające do wywołania emisji z obszaru ostrego wzrostu gęstości stanów, to maksima na krzywych $N(E)$ będą przesuwały się w prawo ze wzrostem $h\nu$ właśnie ze względu na wspomniane *plateau*. Z kolei cofnięcie się w kierunku mniejszym energii maksimów rozkładów otrzymanych przy $h\nu = 5{,}8$ eV można tłumaczyć tak „*długą*" sondą świetlną, że emisja niskoenergetycznych elektronów odbywa się wtedy z obszaru silniejszego wzrostu gęstości stanów, to znaczy poczynając od 0,7 eV poniżej $\mathscr{E}_{v_0}$.

Z krzywych C i D na rys. 12 widać, że położenia „małych" maksimów na krzywych $d(\mathscr{E}_v)$ są pozornie przesunięte: $(\mathscr{E}_{d_{max}} - \mathscr{E}_{v_0})_{(111A)} = 0{,}30$ eV, oraz



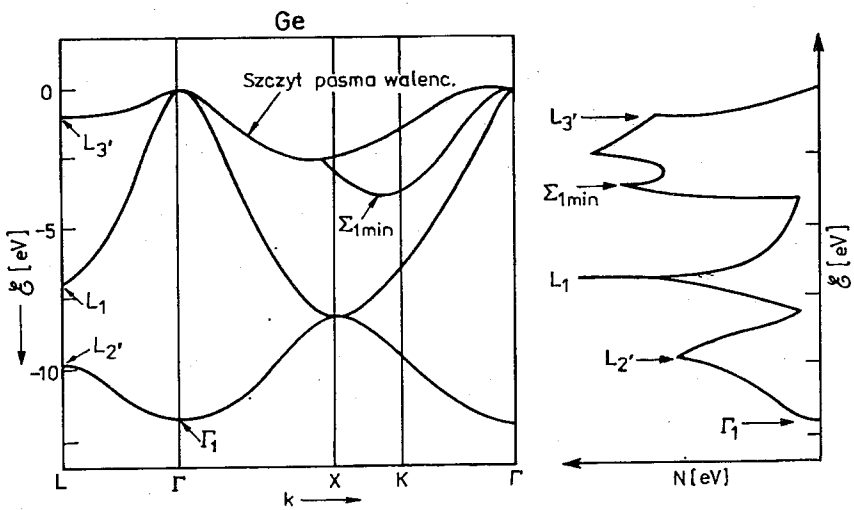
Rys.12. Badanie zależności gęstości stanów (g_s) energetycznych od ich energii w pasmach walencyjnych GaAs typu n

a) Krzywa A – uśredniona dla kilku próbek GaAs(100);
krzywa B – uśredniona dla kilku próbek GaAs(110);
b) Krzywa C – uśredniona dla kilku próbek GaAs(111A);
krzywa D – uśredniona dla kilku próbek GaAs(111B);

$(\mathscr{E}_{d_{max}} - \mathscr{E}_{v_0})_{(111B)} = 0{,}55$ eV. ($\mathscr{E}_{d_{max}}$ – energie „małych" maksimów). Nie trudno stąd wydedukować, że te maksima na krzywych $d(\mathscr{E}_v)$ – nazwijmy je *„garbami"* – są równo odległe od poziomu próżni dla obu typów powierzchni. Głębokość energetyczna sondy świetlnej 4,9 eV odpowiada położeniu garbów i powoduje emisję z nich pewnej grupy fotoelektronów. Dla większych *hv* różnice w przesuwaniu się maksimów $N(E)$ są spowodowane odmiennym przebiegiem krzywych $d(\mathscr{E}_v)$ odpowiednio w warstwach przy (111A) i przy (111B). Jeśli niewielki garb nakłada się na stromo wznoszącą się krzywą, a jest tak w przypadku $d(\mathscr{E}_v)$ dla (111B), to *„wydłużenie"* pobudzającej sondy świetlnej powoduje znaczną emisję z coraz głębiej energetycznie położonych poziomów walencyjnych. Innymi słowy, jeśli wzrost gęstości stanów jest taki, że kompensuje mniejsze prawdopodobieństwo ucieczki niskoenergetycznych fotoelektronów, to maksima na rozkładach $N(E)$ nie powinny się przesuwać.

Podsumowując, wielokrotnie powtarzane, badania gęstości stanów kwantowych w próbkach GaAs, stwierdza się, że krzywe gęstości stanów w górnych częściach pasm walencyjnych – otrzymane zmodyfikowaną (uproszczoną) metodą Kindinga i Spicera dla powierzchni (111A) i (111B) – wykazują niewielkie maksima, po których następuje ostry wzrost funkcji $d(\mathscr{E})$. Maksima te należy przypisać drugiej grupie stanów powierzchniowych nakładającej się na pasma walencyjne [44].

W pierwotnym widmie fotoelektronów to znaczy na krzywych $N(E) = f(E)$ są widoczne szczegóły w gęstości stanów kwantowych $d(\mathscr{E}_v)$ (w tym samym położeniu jak w gęstości stanów) odpowiadające krawędziom pasm: L_3, $\Sigma_{1_{mjn}}$ i Γ_1 (rys. 13) [45], [46], [16]. Można uzyskać rozsądną ogólną ocenę gęstości stanów kwantowych



Rys.13. Energia pasma walencyjnego w funkcji pędu krystalicznego $\vec{k}$ w zredukowanej strefie wykresu obliczonego metodą EMP. Pokazana jest jedno-elektronowa gęstość stanów korespondująca z tym pasmem obliczonym [45]. Punkty krytyczne pasma walencyjnego będą wykrywane (wykazywane) zarówno w pasmie energetycznym jak i w widmie gęstości stanów. Te pasma korespondują ze współczynnikiem pseudopotencjału dla Ge podanego w [16]

$d(\mathscr{E}_v)$ z krzywych rozkładu wysokoenergetycznego fotoemisji przez odjęcie bezstrukturowego widma elektronów wtórnych od krzywej emisji całkowitej, czyli gęstość pochodzącą z emisji pierwotnej; przy wyższych energiach fotonów można bowiem dokładniej rozdzielić emisję pierwotną od wtórnej.

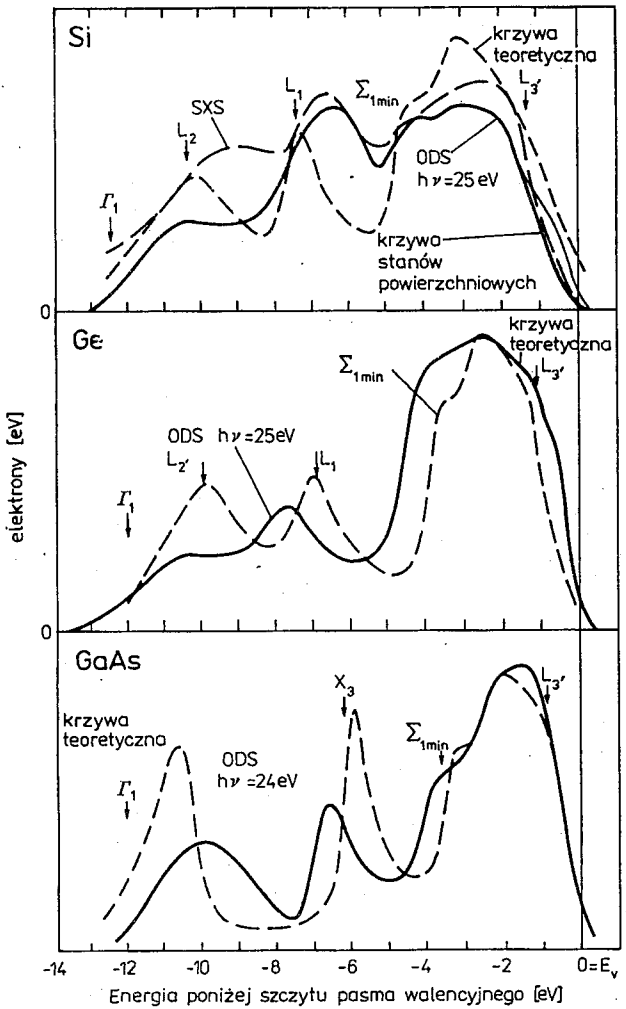
Grobman i Eastman [22] jako pierwsi uzyskali z krzywych energetycznych rozkładów fotoemisji w ultrafiolecie — otrzymanych przy energii fotonów około 25 eV — doświadczalne, optyczne gęstości stanów pasm walencyjnych dla GaAs. Autorzy ci dokładnie określili położenia kilku minimów pasm energetycznych dla drugiego i trzeciego pasma walencyjnego Ge i GaAs. Porównując wyżej wspomniane krawędzie pasm z obliczeniami teoretycznymi opartymi na danych optycznych, zaobserwowali znaczne różnice dla Ge i GaAs.

Pokażemy widma fotoemisyjnej spektroskopii ultrafioletowej (UPS — Ultraviolet Photoemission Spectroskopy) przy zastosowaniu energii fotonów $\leqslant 25$ eV, które dają ogólne spojrzenie na całe pasma walencyjne Si, Ge i GaAs szerokości ~ 12 eV. Z tych widm (UPS) uzyskuje się *„optyczną gęstość stanów"* (ODS — Optical Densities of States), które pokazują ogólną zgodność z teoretycznymi gęstościami stanów.

Grobman i Eastman dokładnie określili położenia kilku minimów pasm energetycznych względem krawędzi pasma walencyjnego E_v [47]. Dla GaAs uzyskali $E_v - \Sigma_{1\min} = (4{,}1 \pm 0{,}2)$ eV i $E_v - X_3 = (6{,}8 \pm 0{,}2)$ eV; obie te wartości są około 0,4 eV większe od wartości obliczonych teoretycznie. Wyniki Grobmana i Eastmana [47] przedstawiają nowe dane dające bezwzględne energetyczne położenia pasma walencyjnego w szerokim przedziale energetycznym poniżej E_v. ten typ danych w połączeniu z danymi optycznymi powinien przyczynić się do dokładniejszego określenia potencjałów krystalicznych dla półprzewodników.

Krawędź pasma walencyjnego dla GaAs można określić przez ekstrapolację krawędzi krzywej rozkładu energii fotoelektronów do zerowego natężenia emisji, które występuje około 0,1 eV powyżej E_v z powodu skończonej rozdzielczości spektrometrów. Uzyskane doświadczalnie optyczne gęstości stanów dla Si, Ge i GaAs odpowiednio dla $h\nu = 25$, 25 i 24 eV zostały przedstawione na rys. 14 jako linie ciągłe. Krzywe przerywane $N(E)$ są gęstościami stanów uzyskanymi ze współczynników pseudopotencjalnych Cohena—Bergstresser'a [16] Ge i GaAs i metody pseudopotencjału empirycznego [47].

Gęstość stanów można uzyskać również metodą spektroskopii miękkich promieni X (SXS — Soft X-ray Spectroscopy). Pokazano to dla Si na rys. 14 u góry. Ponieważ położenie E_v nie jest dokładnie znane w SXS, więc umiejscowiono tę krzywą (przerywaną, kropkowaną) tak, aby nakładała się na krzywą ODS w pobliżu E_v. Posługując się krzywą ODS z rys. 14 i innymi danymi fotoemisyjnymi dla $h\nu < 24$ eV, można określić absolutne wartości położeń punktów krytycznych względem E_v podane w tablicy 3. Wykonuje się to drogą obserwacji zabezpieczeń położeń punktów krytycznych $N(E)$ na rys. 14 względem elementów struktury krzywej $N(E)$ i uczynienie takiej samej odpowiedniości dla krzywych ODS. Na przykład $\Sigma_{1\min}$ odpowiada środkowi ostrego spadku emisji przy dnie dwu górnych



Rys.14. Porównanie krzywych gęstości optycznej stanów (ODS) [30] dla Si, Ge i GaAs z doświadczalnymi gęstościami stanów (z metody pseudopotencjału). Jest również przedstawiona gęstość stanów metodą SXS dla Si. Położenia punktów krytycznych zaznaczono na odpowiednich krzywych teoretycznych

pasm walencyjnych. Najlepiej określoną doświadczalną krawędzią dla GaAs jest krawędź X_3.

W tablicy 3 przedstawiono także wartości dla wybranych punktów krytycznych w GaAs określone przez Cohen'a i Bergstresser'a [16], jak również wyniki pracy Zucca i współpracowników [49] dla GaAs. Rozbieżności między doświadczeniem a teorią w tablicy 1 są spowodowane faktem, że te obliczone wartości były dopasowane do danych optycznych i nie były dopasowane do położeń pasm walencyjnych niższych energetycznie.

Tablica 3.

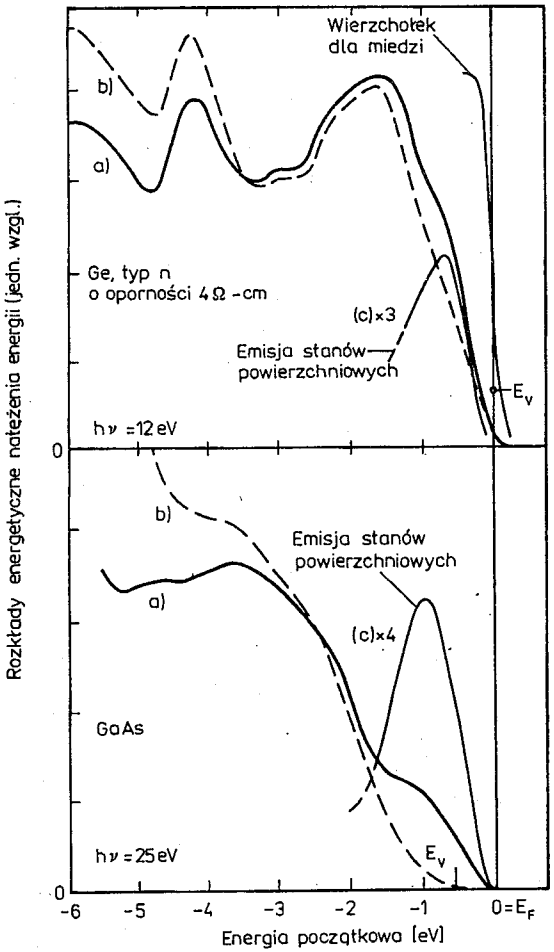
Punkty krytyczne energii poniżej E_v

		L_3 [eV]	$\Sigma_{1\,min}$ [eV]	X_3 [eV]	Γ_1 [eV]
GaAs	Eksperyment	0,8 ± 0,2	4,1 ± 0,2	6,8 ± 0,2	11,9 ± 0,6
	Teoria	0,9	3,6	6,15	12,3
	Teoria	0,9	3,2	6,3	12,0

Reasumując, wyniki własne, a przede wszystkim przytoczonych tu autorów, ilustrują zastosowanie fotoemisji przy wysokich energiach do uzyskania ogólnego spojrzenia na gęstość stanów walencyjnych, jak również dokładnych położeń minimów pasmowych dla półprzewodników [51]. Ostatnio przeprowadza się również pomiary mające na celu rozciągnięcie omawianej tu techniki fotoemisyjnej do wyższych energii fotonów. Pozwolą one zapewne na dokładniejsze określenie punktów krytycznych w najniższych pasmach walencyjnych i przerwy energetycznej w tych pasmach $X - X_3$ dla półprzewodników heteropolarnych, ponieważ emisja elektronowa pierwotna i wtórna dla tych pasm może być dokładniej rozdzielona przy wyższych energiach.

Dyskutując pracę Grobmana i Eastmana należy stwierdzić, że pomiary UPS przy wysokich energiach (około 25 eV) pozwalają na pełniejsze poznanie struktury pasma walencyjnego, czyli uzyskanie informacji odnośnie gęstości stanów w sposób podobny do zastosowanego w spektroskopii fotoemisji promieni X. Eastman i Grobman [39] wykonali również pomiary rozkładów energetycznych, fotoemisyjnych samoistnych stanów powierzchniowych dla Si, Ge i GaAs stosując promieniowanie synchrotronowe w przedziale 7 — 25 eV. Obserwowali oni pasma obsadzonych samoistnych stanów powierzchniowych o optycznych gęstościach stanów w kształcie krzywej Gaussa o szerokości około 0,8 do 1,0 eV, skoncentrowanych odpowiednio przy 0,75; 0,7 i 1,05 eV poniżej poziomu Fermiego, to jest biorąc pod uwagę zagięcie pasmowe, skoncentrowanych odpowiednio przy około 0,45; 0,75 i 0,5 eV poniżej krawędzi pasm walencyjnych.

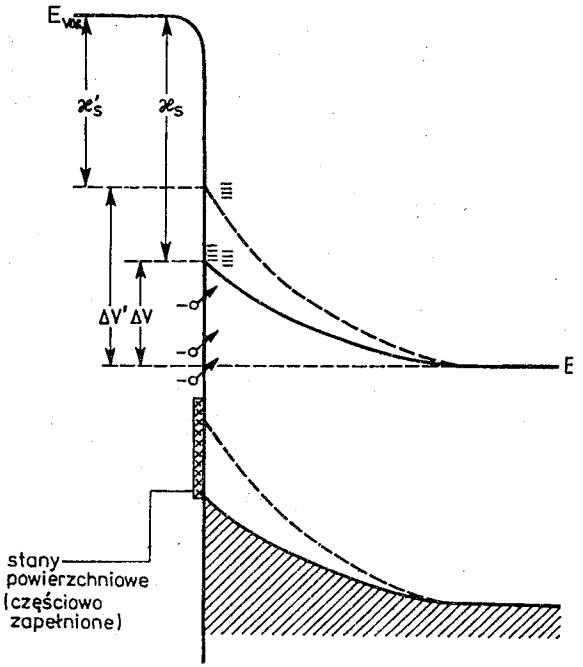
Rys. 15 przedstawia wyniki pomiarów [39] fotoemisyjnych rozkładów energetycznych samoistnych stanów powierzchniowych Ge i GaAs. Krzywe a), b) i c) odnoszą się odpowiednio do powierzchni: świeżo łupanych, powierzchni eksponowanych na gaz i wynikającej z nich optycznej gęstości stanów powierzchniowych. Wiadomym jest, że eksponowanie próbki na gaz — w dużym stopniu usuwa samoistne stany powierzchniowe [51—53] i tworzy niesamoistne stany powierzchniowe; obserwuje się wtedy zmniejszenie zagięcia pasm. Za Eastmanem i Grobmanem, autor również wykazał w GaAs pasma samoistnych stanów powierzchniowych szerokości około 1 ev skupione odpowiednio przy około 0,7 i 1 eV poniżej poziomu Fermiego, to jest przy około 0,7 i 0,5 eV poniżej maksimum pasma walencyjnego.



Rys.15. Rozkłady energetyczne natężenia fotoemisji dla Ge i GaAs. Dla Ge krzywą a) mierzono około 1 godziny po łupaniu ($p \sim 1 \cdot 10^{-9}$ Tr), a krzywą b) mierzono po wystawieniu na $\sim 2 \cdot 10^{-6}$ Torominut powietrza. Dla słabo domieszkowanego GaAs typu n, krzywą a) mierzono około 15 minut po łupaniu, a krzywą b) mierzono około 1 dzień później ($p \sim 3 \cdot 10^{-10}$ Tr). Krzywe różnicowe c) odpowidają gęstości optycznej samoistnych stanów powierzchniowych. Dla GaAs ta gęstość stanów jest poszerzona o ~ 0,35 eV przez pasmo przechodzenia fotonów [39

4. ZAGIĘCIE PASM PRZY POWIERZCHNI

Autor zaobserwował przypowierzchniowe zagięcie pasm do góry dla próbek GaAs typu n. Dla próbek o koncentracjach od $8 \cdot 10^{16}$ do $5 \cdot 10^{17} cm^{-3}$, dla orientacji (111A) bezwzględne zagięcie pasm wynosiło 0,80 eV, a dla orientacji (111B) – 1,30 eV. Zagięcie najmniejsze dla trzech badanych orientacji próbek GaAs wystąpiło przy powierzchniach „*galowych*", podczas gdy przy „*aresenowych*" pasma



Rys.16. Wpływ elektrododatniego aktywatora na zmianę zagięcia pasm przy powierzchni półprzewodnika

zaginają się prawie o całą energetyczną szerokość przerwy pasmowej. Wyniki te autor tłumaczy w następujący sposób: Mniejsze zagięcie pasm przy powierzchniach (111A) może być spowodowane obecnością monowarstwy galu na tych powierzchniach, którego działanie jest takie jak przy pokryciu powierzchni cezem. Gal bowiem jest pierwiastkiem słabo elektroujemnym i łatwo oddaje elektrony, których część wnika do obszaru przypowierzchniowego ładunku przestrzennego, przez co inwersja pasm przy powierzchni zmniejsza się. Rys. 16 ilustruje oddziaływanie galu na przypowierzchniowe zagięcie pasm energetycznych. Uzyskane przez autora zagięcia ΔV oraz to, że dla obu typów powierzchni prace wyjścia są zbliżone — są jakościowo zgodne z danymi pracy Jamesa i innych [54], którzy badali epitaksjalne warstwy GaAs po ich oczyszczeniu przez wygrzewanie. Cytowani autorzy otrzymali prawie takie same jak w tej pracy wartości przypowierzchniowych zagięć pasm ($\Delta V_{(111A)} = 0{,}6$ eV; $\Delta V_{(111B)} = 1{,}3$ eV) oraz pokrywające się dla obu warstw wartości prac wyjścia elektronów. Oczywiście liczbowe wartości $h\nu_t$ i φ tam otrzymane są znacznie mniejsze z powodu aktywacji powierzchni cezem; jednakże wyraźne zróżnicowanie wartości $h\nu_t$ w przypadku (111A) i (111B) jest widoczne na podstawie wyników Jamesa, podobnie jak w badaniach autora.

Wiemy, że duży wpływ na wartość φ ma termiczna obróbka próbek w próżni; pokrycie zaś cezem, równomiernie obniża powierzchniowe parametry. W cytowanej

pracy [54] wartości φ dla obu typów powierzchni są równe i wynoszą 1,32 eV, podczas gdy w modelach skonstruowanych przez autora różnią się conajmniej o 0,1 eV (co na dysponowane warunki jest zupełnie dobre) [55].

BIBLIOGRAFIA

1. H.R. Philipp, E.A. Taft: *Optical Constans of Silicon in the Region* 1 to 10 eV. Phys. Rev. 1960, v. 120, pp. 37–38
2. H. Ehrenreich, H.R. Philipp, J.C. Phillips: *Interband Transitions in Groups* 4, 3–5, and 2–6 *Semiconductors*. Phys. Rev. Let. 1962, v. 8, 59–61
3. L. Apker, E. Taft, J. Dickey: *Photoelectric Emission and Contact Potentials of Semiconductors*. Phys. Rev. 1948, v. 74 pp. 1462–1474
4. J. Woja s: *Analiza metody wyzanczania wielkości fotoemisyjnych*. Archiwum Elektrotechniki, 1983, Tom XXXII, s. 375–386
5. J. Woja s: *Rozkłady spektralne wydajności kwantowej a progi fotoemisji*. Elektronika, 1983, t. 9 (XXIII), s. 17–22
6. N. Kiding, W. Spicer: *Band structure of cadmium sulfide – Photoemission studies*. Phys. Rev. 1965, v. 138, pp. A561–A576
7. E.O. Kane: *Theory of photoelectric emission from semiconductors*. Phys. Rev. 1962, t. 127, pp. 131–141
8. E.O. Kane: *band structure of indium antimonide*. J. Phys. Chem. Solids, 1957, Nr 1, p. 249–261
9. J. Woja s: *O wydajności fotoemisji i jej długofalowych progach*. Archiwum Elektrotechniki, 1982, Tom XXXI, 375–386
10. M. Cardona: *Optical Absorption above the Fundamental Edge, Optical properties of* III–V *Compounds*. Semiconductor and Semimetals. 1967, v. 3, pp. 127–151
11. F. Bassani: *Physics of* III–V *Compounds*. Academic Press, Inc., New York 1966, vol. 1
12. J.C. Phillips: *The Fundamental Optical Spectra of Solids*. Solid State Phys., 1966, v. 18, pp. 55–164. (Advances in Research and Applications)
13. A.G. Thompson, J.C. Woolley, M. Rubenstein: *Reflectance of GaAs, GaP, and the* $GaAs_{1-x}P_x$ *Alloys*. Canad. J. Phys., 1966, v. 44, pp. 2927–2940
14. F. Bassani, M. Yoshimne: *Electronic Band Structure of Group* IV *Elements and* III–V *Compounds*. Phys. Rev., 1963, v. 130, pp. 20–33
15. F. Herman, R.L. Kortum, C.D. Kuglin, J.P. Van Dyke, S. Skillman: *Methods in Computational Physics*. Ed. B. Alden, S. Fernbach, M. Rotenberg, Academic Press, Inc., New York 1968, Vol. 8
16. M.L. Cohen, T.K. Berstresser: *Band Structures and Pseudopotential Form Factors for Fourteen Semiconductors of the Diamond and Zinc-blende Structures*. Phys. Rev., 1966, v.141, pp. 789–796
17. A.M. Gray: *Evaluation of Electronic Energy Band Structures of GaAs and GaP*. Phys. Stat. Sol., 1970, v. 37, pp. 11–28
18. W. Saslow, T.K. Bergstresser, C.K. Fong, M.L. Cohen, D. Brust: *Pseudopotential Calculation of* ε_2 *for the Zinc-blende Structure GaAs*. Solid State Commun. 1967, v. 5, pp. 667–669
19. W.E. Spicer, R.E. Simon: *Photoemissive studies of the band structure of silicon*. Phys. Rev. Letters 1962, v. 9, pp. 385–389
20. T.E. Fischer: *Determination of Semiconductor Surface Properties by Means of Photoelectric Emission*. Surface Science, 1969, v. 13, pp. 30–51
21. G.W. Gobeli, F.G. Allen: *Direct and Indirect Excitation Processes in Photoelectric Emission from Silicon*. Phys. Rev. 1962, v. 127, pp. 141–149

22. D.E. Eastman, W.D, Grobman: *Photoemission Valence-Band Densities of States for Si, Ge, and GaAs Using Synchrotron Radiation.* Phys. Rev. Letters 1972, v. 29, pp. 508–1512
23. T.E. Fischer, F.G. Allen, G.W. Gobeli: *Photoelectric Emission from InAs: Surface Properties and Interband Transitions.* 1967, v. 163, pp. 703–711
24. G.W. Gobeli, F.G. Allen: *Photoelectric Properies of Cleaved GaAs, GaSb, InAs, and InSb Surfaces; Comparison with Si and Ge.* Phys. Rev. 1965, v. 137, pp. A245–A254
25. M.L. Cohen, J.C. Phillips: *Spectral Analysis of Photoemissive Yields in Si, Ge, GaAs, GaSb, InAs, and InSb.* Phys. Rev. 1965, v. 139, pp. A912–A920
26. D. Brust, M.L. Cohen, J.C. Phillips: *Reflectance and Photoemission from Si.* Phys. Rev. Letters 1962, v. 9, pp. 389–392
27. H.R. Phillip, H. Ehrenreich: *Optical Properties of Semiconductors.* Phys. Rev. 1963, v. 129, pp. 1550–1560
28. D.L. Greenaway, M. Cardona: *Reflectivity measurements on InSb–In_2Te_3 and InAs – In_2Te_3 alloys and on pure InSb, InAs and In_2Te_3.* Proceedings of the International Conference on the Physics of Semiconductors, London 1962, pp. 666–672
29. T.E. Fischer, F.G. Allen, G.W. Gobeli: *Photoelectric Emission from InAs: Surface Properties and Interband Transitions.* Phys. Rev. 1967, v. 163, pp. 703–711
30. D.E. Eastman, W.D. Grobman: *Photoemission Energy Distributions for Au from 10 to 40 eV Using Synchrotron Radiation.* Phys. Rev. Let. 1972, v. 28, pp. 1327–1330
31. W.E. Spicer, R.E. Simon: *The use of photoemission to investigate the band structure of silicon and other semiconductors.* J. Phys. Chem. Solids, 1962, v. 23, pp. 1817(L)–1819(L)
32. J.C. Phillips, D. Brust, F. Bassani: *Critical points and ultraviolet reflectivity of semiconductors.* Phys. Rev. Let. 1962, v. 9, p. 94–97
33. J.C. Phillips: *Band Structure of Silicon, Germanium, and Related Semiconductors.* Phys. Rev. 1962, v. 125, pp 1931–1936
34. J.C. Phillips: *Ultraviolet Absorption of Insulators.* II. *Partially Ionic Crystals.* Phys. Rev. 1964, v. 133, pp. A452–A459
35. D.E. Eastman, D.W. Grobman, J.L. Freeouf, M. Erbudak: *Photoemission Spectroscopy using synchrotron radiation.* I. *Overviews of valence-band structure for Ge, GaAs, GaP, InSb, InSe, CdTe, and AgI.* Phys. Rev. B, 1974, v. 9, pp. 3473–3488
36. I.P. Ipatova, Yu.E. Kitayev, A.B. Subshiyev: *Symetry changes due to second-order phase transitions at the surface.* Pisma v Zh. Eksp. Teor. Fiz., 1980, v. 32, pp. 587–590
37. E.O. Kane: *Structure due to Transport Effects in Photoelectric Energy Distributions. Proceedings of the International Conference on the Physics of Semiconductors, Kyoto.* J. Phys. Society of Japan, 1966, v. 21, Supplement, pp. 37–45
38. J. Wojas: *Determination of the density of states of photoemitting bands in gallium arsenide by the method of Kinding and Spicer.* 1981, A60, pp. 767–781
39. D.E. Eastman, W.D. Grobman: *Photoemission Densities of Intrinsic Surface States for Si, Ge, and GaAs.* Phys. Rev. Letters, 1972, v. 28, pp. 1378–1381
40. B.A. Orłowski: *Fotoemisyjne badania struktury pamowej półprzewodników* $A_{II}B_{VI}$. Postępy Fizyki, 1976, t. 27, s. 463–476
41. F.G. Allen, G.W. Gobelli: *Comparison of the Photoelectric Properties of Cleaved, Heated, and Suttered Silicon Surfaces.* J. of Appl. Physics, 1964, v. 35, pp. 597–605
42. T. Fischer, F.G. Allen, G.W. Gobelli: *Photoelectric Emission from InAs: Surface Properties and Interband Transitions.* Phys. Rev. 1967, v. 163, pp. 703–711
43. P.E. Gregory, W.E. Spicer: *Photoemission study of surface states of the GaAs(110) surface.* Phys. Rev. 1976, v. 13, pp, 725–738
44. N.B. Kinding: *Effects of Band-Bending on Energy Distribution Curves in Photoemission.* J. of Appl. Phys. 1967. v. 38, pp. 3285–3290
45. R.A. Pollak i inni: *X-Ray Photoemission Valence-Band Spectra and Theoretical Valence-Band Densities of States for Ge, GaAs, and ZnSe.* Phys. Rev. Let. 1972, 29, pp. 1103–1105

46. L. L e y i inni: X-*Ray Photoemission Spectra of Crystalline and Amorphorus Si and Ge Valence Bands.* Phys. Rev. Let. 1972, v. 29, pp. 1088 – 1092
47. W.D. G r o b m a n, D.E. E a s t m a n: *Photoemission Valence-Band Densities of States for Si, Ge, and GaAs Using Synchrotron Radiation.* Phys. Rev. Let. 1972, v. 29, pp. 1508 – 1512
48. L.R. S a r a v i a, L. Casamayou: *Calculation of thew photoelectric effect in germianium.* J.Phys. Chem. Solids. 1971, v. 32, pp 1541 – 1552
49. R.L. Z u c c a i inni: *Wavelength modulation spectra od GaAs and Silicon.* Solid Commun. 1970, v. 8, pp. 627 – 632
50. J. C h e l i k o w s k y, D.J. C h a d i, M.L. C o h e n: *Calculated Valence – Band Densities of States and Photoemission Spectra of Diamond and Zinc-Blende Semiconductors.* Phys. Rev. B. 1973, v. 8, pp. 2786 – 2794
51. S.G. D a v i s o n, J.D. L e v i n e: *Surface States, in Solid State Physics* (Advances in Research and Applications), wyd. H. Ehrenreich, F. Sietz, F. Turnbull, New York – London: Academic Press, 1970, vol. 25, pp. 1 – 149
52. D.J. M i l l e r, D.L. H e r o n, D. H a n e m a n: *Proceedings of International Conference*, Surface Science, Boston, Massachusetts 1971
53. G. C h i a r o t t i i inni: *Optical Absorption of Surface States in Ultrahigh Vacuum Cleaved* (111) *Surfaces of Ge and Si.* Phys. Rev. 1971, vol. 4, pp. 3398 – 3402
54. L. J a m e s i inni: *Dependence on Crystalline Face of the Band Banding in Cs_2O – Activared GaAs.* J. Appl. Phys. 1971, vol. 42, pp. 4976 – 4980
55. R. W i l l i a m s: *Modern GaAs Processing Methods.* Boston 1990

J. WOJAS

THE EXAMINATION OF ENERGY STRUCTURE OF SEMICONDUCTORS BY APPLYING OF EXTERNAL PHOTOEMISION

S u m m a r y

In the paper, the possibility of applying of external photoemission for determining of the electronic and photoelectric parameters of the crystal surfaces and of the surface layers is described. In the second part of this work, the practical way of determining of the mentioned above parameters is shown by using of the experimental characteristics: EDC, quantum yield, reflection and absorption curves.

These experimental results follow on the defining of the energy structure of surface and the subsurface band bending.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1994, 40, z. 3, ss. 419—428

621.3.052.

Identyfikacja ogólnej struktury wielowymiarowych systemów stacjonarnych

GRZEGORZ CIESIELSKI

Instytut Podstaw Elektrotechniki, Politechnika Łódzka

Otrzymano 1994.06.09

Autoryzowano do druku 1994.11.10

W artykule omówione zostały zagadnienia związane z identyfikacją strukturalną wielowymiarowych systemów stacjonarnych opisanych za pomocą operatorów, które w ogólnym przypadku mogą być nieliniowe. Założono, że sygnały wejściowe i wyjściowe w rozważanych systemach, mogą przyjmować wartości rzeczywiste lub zespolone. Zdefiniowane zostały w sposób ścisły pojęcia stacjonarności i przyczynowości. Pokazano, że bezpośrednio z tych definicji wynikają pewne ściśle określone ogólne struktury operatorów opisujących takie systemy. Otrzymane struktury stanowią podstawę do sformułowania ogólnej teorii modelowania systemów, obejmującej między innymi, teorię Volterry oraz Wienera.

1. WSTĘP

Przedmiotem artykułu jest identyfikacja strumienia wielowymiarowych systemów stacjonarnych, które dają się opisać za pomocą określonych dalej odwzorowań nazywanych operatorami. Operatory te w ogólnym przypadku mogą być nieliniowe, a jedną z możliwych form ich reprezentacji może być układ równań różniczkowych, być może nieliniowych. Systemy rozważane w artykule są wielowymiarowe, czyli mają dowolną skończoną liczbę wejść i wyjść. Sygnały na poszczególnych wejściach i wyjściach mogą przyjmować wartości rzeczywiste lub zespolone.

Przypomnijmy, że identyfikacja strukturalna dowolnego systemu, jest jednym z elementów zadania modelowania, które ma na celu wyznaczenie matematycznego modelu tego sygnału. Zadanie modelowania obejmuje również identyfikację parametryczną modelu, którego struktura został wyznaczona w procesie identyfikacji strukturalnej (J. Jaworski, R. Morawski, J. Olędzki [10], s. 65).

Dowolny system można traktować jako odwzorowanie Φ zbioru sygnałów wejściowych w zbiór sygnałów wyjściowych tego systemu. Załóżmy, że Φ jest

elementem pewnej przestrzeni liniowej L_Φ. Wówczas zadanie identyfikacji strukturalnej poszukiwanego modelu systemu Φ, sprowadza się do określenia zbioru lub przestrzeni (klasy M_φ zawartej w przestrzeni L_Φ, do której należy poszukiwany model.

By to bliżej wyjaśnić załóżmy, że chcemy dokonać identyfikacji strukturalnej systemu statycznego Φ z jednym wejściem oraz jednym wyjściem, o którym wiemy, iż jest opisany za pomocą funkcji ciągłej na odcinku domkniętym $[x_{\min}, x_{\max}]$, a zatem przestrzeń $L_\Phi = C([x_{\min}, x_{\max}])$. Rozwiązaniem tego problemu może być znane twierdzenie Weierstrassa (G. Meinardus [11], s. 7 i A. Ralston [14], s. 35), z którego wynika, że każdy taki system możemy dowolnie dokładnie opisać w sensie normy w L_Φ, za pomocą odwzorowań należących do przestrzeni M_Φ wielomianów k-tego stopnia, zwiększając odpowiednio stopień k. Z kolei identyfikacja parametryczna jest czynnością polegającą na wyznaczeniu parametrów przyjętej struktury modelu systemu Φ, czyli jej celem jest wybór najlepszego modelu w sensie normy L_Φ, który należy do klasy modeli $M_\Phi \subset L_\Phi$ wyznaczonej w procesie identyfikacji strukturalnej Warto zauważyć, że zadanie to może być czasami niejednoznaczne, to znaczy, że jego rozwiązaniem będzie zbiór modeli danego systemu Φ. W przypadku przykładowego statycznego systemu z jednym wejściem oraz jednym wyjściem, rozwiązanie zadania identyfikacji parametrycznej modelu odnajdujemy w teorii dotyczącej aproksymacji funkcji (G. Meinardus [11], i A. Ralston [14], s. 218).

W światowej literaturze naukowej jest wiele publikacji, które w mniejszym lub większym stopniu wiążą się z przedmiotem niniejszego artykułu. Znakomita ich większość dotyczy rozwinięć operatorowych modeli systemów w szeregi Volterry, które będziemy dalej nazywali modelami Volterry. Z pewnością na najwyższą uwagę zasługują tu publikacje I.W. Sandberga i in. ([12] i [15]–[19]) oraz R.J.P. de Figueiredo i in. ([7] i [8]), które są matematycznie ścisłe. Z kolei w monografii M. Schetzena ([2]), ucznia N. Wienera, oraz monografii K.A. Pupkowa, W.I. Kapalina i A.S. Juszczenki ([13]) omówiono zarówno modele Volterry, jak również modele Wienera systemów nieliniowych. Monografie te są jednak obecnie opracowaniami o charakterze podstawowym i nie do końca formalnym. Brak jest w nich na przykład opisu klasy operatorów, dla których rozważane modele istnieją, a ponadto przedstawiona w nich teoria systemów nieliniowych składa się z dwóch rozłącznych teorii Volterry oraz Wienera.

Ponieważ w światowej literaturze nie podjęto próby w pełni formalnego, a zarazem jednolitego podejścia do zagadnienia modelowania systemów opisanych za pomocą operatorów, głównym celem tego artykułu jest stworzenie niezbędnych podstaw do sformułowania ogólnej teorii modelowania systemów obejmującej między innymi teorię Volterry oraz Wienera (G. Ciesielski [1]–[6]).

2. OPERATOROWY OPIS SYSTEMÓW

Dowolny system można traktować jako odwzorowanie zbioru U sygnałów wejściowych zwanych *wymuszeniami* w zbiór V sygnałów wyjściowych zwanych *odpowiedziami*, co zostało pokazane na rys. 1. Będziemy dalej zakładać, że oba te

Rys.1. System φ jako odwzorowanie zbioru wymuszeń U w zbiór odpowidzi V

zbiory są wyposażone w strukturę przestrzeni liniowej, a zatem określona jest w nich operacja dodawania sygnałów i mnożenia sygnału przez liczbę z ciała liczbowego F[1]. Zauważmy, że operację mnożenia sygnału przez liczbę z ciała F można interpretować jako wzmocnienie lub tłumienie tego sygnału. Ponadto, jeżeli F będzie ciałem liczb rzeczywistych, to wartości sygnałów wejściowych i wyjściowych będą rzeczywiste, a gdy F będzie ciałem liczb zespolonych, to wartości tych sygnałów będą zespolone.

Niech R[2] będzie zbiorem chwil czasowych, dla których określone są sygnały z U oraz V. Załóżmy ponadto, że $m \in \mathrm{N}$[3] jest liczbą wejść, a $n \in \mathrm{N}$ liczbą wyjść systemu. Wartości sygnałów wejściowych należą zatem do F^{m}[4], a wartości sygnałów wyjściowych do F^{n}. Wówczas przestrzeń sygnałów wejściowych $\mathrm{U_F}$[5] jest podprzestrzenią liniową przestrzeni $\mathrm{L_F^m(R)}$[6]. Podobnie przestrzeń sygnałów wyściowych $\mathrm{V_F}$ jest podprzestrzenią liniową przestrzeni $\mathrm{L_F^n(R)}$. Ponieważ U oraz V są przestrzeniami funkcyjnymi, których elementy określone są na tej samej dziedzinie R, a dowolne odwzorowanie Φ przestrzeni U w V nazywane jest *operatorem*, to każdy system można opisać za pomocą operatora, którego jedną z możliwych form reprezentacji jest układ równań różniczkowych. Zauważmy przy tym, że w ogólnym przypadku operator Φ nie musi być liniowy. Dla uproszczenia dalszej dyskusji pojęcia system opisany za pomocą operatora Φ i operator Φ będą więc traktowane jako równoważne.

3. OPERATORY PRZESUNIĘCIA I OBROTU FUNKCYJNEGO

Zobaczymy, że operatory przesunięcia i obrotu funkcyjnego pozwolą nam wniknąć w ogólną strukturę systemów stacjonarnych. W tym celu załóżmy, że $\sigma \in \mathrm{R}^k$.

---

(1) W całym artykule przez F oznaczamy ciało liczbowe. W szczególności ciało F oznacza ciało liczb rzeczywistych R lub ciało liczb zespolonych C.

(2) Zbiór X wyposażony w dowolną strukturę algebraiczną i/lub topologiczną będziemy oznaczać literą X, a zatem przez R oznaczamy zbiór liczb rzeczywistych, natomiast przez R ciało liczb rzeczywistych.

(3) Przez N oznaczamy zbiór liczb naturalnych.

(4) Dla dowolnego zbioru X przez X^n oznaczamy n-krotny iloczyn kartezjański (produkt) zbioru X.

(5) Przez $\mathrm{X_F}$ oznaczamy przestrzeń liniową nad ciałem skalarów F. Jeżeli F wynika z kontekstu, to będzie ono pomijane w oznaczeniu tej przestrzeni

(6) Przez $\mathrm{L_F^n}(X)$ oznaczamy przestrzeń odwzorowań określonych na X o wartościach z F^n. Oznaczenia $\mathrm{L_F^1}(X)$ oraz $\mathrm{L_F}(X)$ będziemy traktować jako równoważne, a gdy zbiór X wynika z kontekstu, to piszemy w skrócie $\mathrm{L_F^n}$ lub $\mathrm{L_F}$. Przestrzeń $\mathrm{L_F^n}$ będziemy również traktować jako n-krotny produkt przestrzeni $\mathrm{L_F}$.

Przesunięciem funkcyjnym o σ nazywamy operator oznaczany dalej przez $\nabla_\sigma: L_F^n(R^k) \to L_F^n(R^k)$, który $\forall u \in L_F^n(R^k)$ i $\forall t \in R^k$ spełnia warunek:

$$[\nabla_\sigma(u)](t) := u(t + \sigma).$$

Funkcję $\nabla_\sigma \circ u$ nazywamy *przesunięciem funkcji* u o σ.

Obrotem funkcyjnym nazywamy operator, który będziemy oznaczać przez $\nabla^*: L_F^n(R^k) \to L_F^n(R^k)$ taki, że $\forall u \in L_F^n(R^k)$ i $\forall t \in (R^k)$

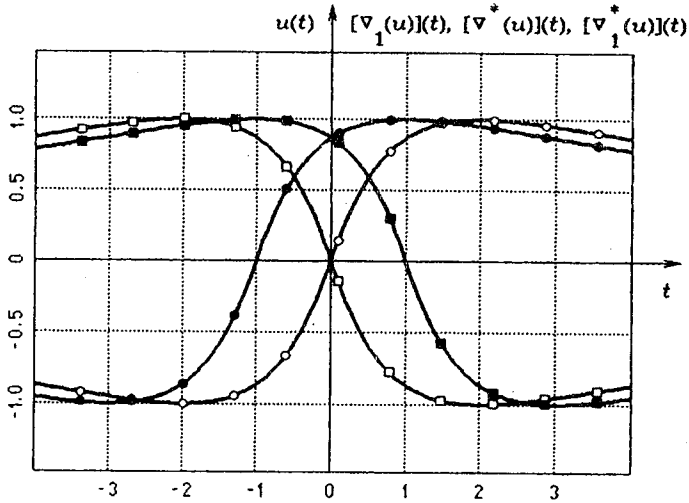
$$[\nabla^*(u)](t) := u(-t).$$

Funkcję $\nabla^* \circ u$ nazywać będziemy *obrotem funkcji* u.

Z kolei *splotowym przesunięciem funkcyjnym* o σ nazywać będziemy operator, który jest oznaczony przez $\nabla_\sigma^*: L_F^n(R^k) \to L_F^n(R^k)$ oraz $\forall u \in L_F^n(R^k)$ i $\forall t \in (R^k)$ spełnia warunek

$$[\nabla_\sigma^*(u)](t) := u(\sigma - t).$$

Złożenie $\nabla_\sigma^* \circ u$ nazywamy *przesunięciem spolotowym funkcji* u o σ.



Rys.2. Efekt działania przesunięcia funkcyjnego — ●, obrotu funkcyjnego — □ oraz splotowego przesunięcia funkcyjnego — ■ na funkcję — ○

Zauważmy tu, że przesunięcia ∇_σ i ∇_σ^* są operatorami liniowymi na przestrzeni $L_F^n(R^k)$ i odwzorowują przestrzeń $L_F^n(R^k)$ w nią samą. Ponadto

$$\nabla_\sigma \circ \nabla_{-\sigma} = \mathrm{id}^{(7)}, \quad \nabla_\sigma^* \circ \nabla_\sigma^* = \mathrm{id},$$
$$\nabla_\sigma^* = \nabla^* \circ \nabla_\sigma = \nabla_{\sigma-} \circ \nabla^* \quad \text{oraz} \quad \nabla_0^* = \nabla^*.$$

Efekt działania tych operatorów na funkcję $u \in L_F(R)$ pokazano na rys. 2. Zauważmy, że przesunięcie funkcji $\nabla_\sigma \circ (u)$ jest funkcją u przesuniętą w lewo o σ, obrót

(7) Przez id oznaczamy odwzorowanie identycznościowe (identyczność).

funkcji $\nabla^* \circ (u)$ jest funkcją u obróconą o π wokół osi rzędnych, a splotowe przesunięcie funkcyjne ∇^*_σ jest złożeniem operacji ∇^* z ∇_σ, których nie można wykonać w dowolnej kolejności.

4. STACJONARNOŚĆ I PRZYCZYNOWOŚĆ SYSTEMÓW

Operatory, a zatem i rozważane systemy, charakteryzują się pewnymi właściwościami, jedną z nich jest stacjonarność.

Operator Φ na $U_F \subset L_F^m$ nazywamy *stacjonarnym*, jeżeli

$$\forall u \in U \; \forall \sigma \in R \; \big((\nabla_\sigma(u) \in U) \wedge (\nabla_\sigma(\Phi(u)) = \Phi(\nabla_\sigma(u)))\big).$$

Innymi słowy, operator jest stacjonarny, gdy odpowiedź operatora na wymuszenie u przesunięte w czasie o σ jest równa przesuniętej w czasie o σ odpowiedzi operatora na wymuszenie u. Zauważmy, że przestrzeń liniowa U musi być zbiorem zamkniętym ze względu na przesunięcia funkcyjne, a zatem zawierać musi wszystkie przesunięcia funkcyjne swoich elementów. Kolejną ważną właściwością operatora jest przyczynowość.

Operator Φ na $U_F \subset L_F^m(R)$ nazywamy *przyczynowym*, jeżeli

$$\forall u, v \in U \quad \forall t \in R$$

$$\big(\forall \tau \in R \; ((\tau \leqslant t) \Rightarrow (u(\tau) = v(\tau)))\big) \Rightarrow \big([\Phi(u)](t) = [\Phi(v)](t)\big).$$

Wynika stąd, że operator jest przyczynowy, gdy wartość sygnału wyjściowego w chwili t zależy od wartości sygnału wejściowego do chwili t.

5. OGÓLNA STRUKTURA SYSTEMÓW STACJONARNYCH

Okazuje się, że bezpośrednio z przedstawionych w poprzedniej części definicji można wnioskować o ogólnej wewnętrznej strukturze operatorów stacjonarnych, a zatem i o ogólnej strukturze odpowiadających tym operatorom systemów.

Twierdzenie. *Załóżmy, że przestrzeń* $U \subset L_F^m(R)$ *spełnia warunek*:

$$\forall u \in U \quad \forall t \in R \; (\nabla_t(u) \in U).$$

(1) *Dla każdego przekształcenia*[8] $\varphi : U \to F^n$ *złożenie* $\varphi \circ \nabla$ *jest operatorem stacjonarnym na* U o *wartościach* z $L_F^n(R)$.

(2) *Operator* $\Phi : U \to L_F^n(R)$ *jest stacjonarny wtedy i tylko wtedy, gdy*

$$\forall u \in U \quad \forall t \in R \; \big([\Phi(u)](t) = [\Phi(\nabla_t(u))](0)\big).$$

[8] Jeżeli X oraz Y są zbiorami z pewną strukturą algebraiczną, to dowolne odwzorowanie $f : X \to Y$ nazywamy *przekształceniem*.

Załóżmy dalej, że U *spełnia dodatkowo warunek*:

$$\forall u \in U \left(\nabla^*(u_{-1})u \in U\right)^{(9)}.$$

(3) *Operator* Φ: $U \to L_F^n(R)$ *jest przyczynowy wtedy i tylko wtedy, gdy*

$$\forall u \in U \quad \forall t \in R \left([\Phi(u)](t) = [\Phi(\nabla_t^*(u_{-1})u)](t)\right).$$

(4) *Dla każdego przekształcenia* $\Phi: U \to F^n$ *spełniającego warunek*:

$$\forall u \in U \left(\varphi(u) = \varphi(\nabla^*(u_{-1})u)\right),$$

złożenie $\varphi \circ \nabla_\cdot$ *jest operatorem stacjonarnym przyczynowym na* U *o wartościach z* $L_F^n(R)$.

(5) *Operator* $\Phi: U \to L_F^n(R)$ *jest stacjonarny i przyczynowy wtedy i tylko wtedy, gdy*

$$\forall u \in U \quad \forall t \in R \left([\Phi(u)](t) = [\Phi(\nabla^*(u_{-1})\nabla_t(u))](0)\right).$$

D o w ó d

(1) Zauważmy, że $\forall u \in U$ oraz $\forall \tau \in R$ mamy:

$$\nabla_\tau([\varphi \circ \nabla_\cdot](u)) = \nabla_\tau(\varphi(\nabla_\cdot(u))) = \varphi(\nabla_{\cdot+\tau}(u)) = \varphi(\nabla_\cdot(\nabla_\tau(u))) =$$

$$= [\varphi \circ \nabla_\cdot](\nabla_\tau(u)),$$

a zatem operator $\Phi \circ \nabla_\cdot$ jest stancjonarny.

(2) W a r u n e k k o n i e c z n y. Zauważmy, że

$$\varphi := [\Phi(\cdot)](0)$$

jest przekształceniem z U w F^n, a zatem z części (1) wynika teza.

W a r u n e k d o s t a t e c z n y. Ponieważ z definicji operatora stacjonarnego wynika, że $\forall u \in U$ oraz $\forall t, \tau \in R$

$$[\Phi(u)](t) = [\Phi(u)](t - \tau + \tau) = [\nabla_{t-\tau}(\Phi(u))](\tau) = [\Phi(\nabla_{t-\tau}(u))](\tau),$$

to przyjmując $\tau = 0$ otrzymujemy, że

$$[\Phi(u)](t) = [\Phi(\nabla_t(u))](0).$$

(3) W a r u n e k k o n i e c z n y wykażemy nie wprost. Załóżmy zatem, że Φ nie jest operatorem przyczynowym. Ponieważ $\forall u \in U$ oraz $\forall t, \tau \in R$ wartość sygnału

$$[\nabla_t^*(u_{-1})](\tau)u(\tau) = \begin{cases} u(\tau), & \tau \leqslant t \\ 0, & \tau > t \end{cases},$$

to z definicji operatora przyczynowego wynika, że $\exists u, v \in U$ oraz $\exists t \in R$, iż z równości

$$[\nabla_t^*(u_{-1})]u = [\nabla_t^*(u_{-1})]v,$$

---

(9) W całym artykule przez u_{-1} oznaczamy skok jednostkowy.
Jeżeli $f: X \to F$ oraz $g: X \to F$, to fg określone jest przez wzór:

$$[fg](x) := f(x)g(x).$$

mamy

$$[\Phi(u)](t) \neq [\Phi(v)](t),$$

podczas gdy

$$[\Phi(\nabla_t^*(u_{-1})u)](t) = [\Phi(\nabla_t^*(u_{-1})v)](t).$$

Wynika stąd, że $\exists u \in U$ oraz $\exists t \in R$, dla których

$$[\Phi(u)](t) \neq [\Phi(\nabla_t^*(u_{-1})](u),$$

co przeczy założeniu warunku koniecznego.

W a r u n e k d o s t a t e c z n y. Z definicji wynika, że operator Φ jest przyczynowy, jeżeli jego wartość w chwili t zależy jedynie od wartości sygnału u do chwili t. Ponieważ $\forall u \in U$ oraz $\forall t, \tau \in R$ wartość sygnału

$$[\nabla_t^*(u_{-1})](\tau)u(\tau) = u(\tau),$$

dla $\tau \leqslant t$, to stąd otrzymujemy tezę.

(4) Z części (1) wynika, że złożenie $\varphi \circ \nabla_\cdot$ jest operatorem stacjonarnym. Ponieważ $\forall u \in U$ oraz $\forall t \in R$

$$\varphi(\nabla_t(u)) = \varphi(\nabla^*(u_{-1})\nabla_t(u)),$$

to oznaczając

$$[\Phi(u)](t) := \varphi(\nabla_t(u)),$$

otrzymujemy

$$\varphi(\nabla^*(u_{-1})\nabla_t(u)) = [\Phi(\nabla^*(u_{-1})\nabla_t(u))](0) =$$

$$= [\varphi(\nabla_{-t}(\nabla^*(u_{-1})\nabla_t(u)))](t) = [\varphi(\nabla_t^*(u_{-1})u)](t),$$

a stąd oraz części (3) widać, że złożenie $\varphi \circ \nabla_\cdot$ jest również operatorem przyczynowym.

(5) W a r u n e k k o n i e c z n y. Zauważmy, że

$$\varphi := [\Phi(\cdot)](0),$$

jest przekształceniem z U w F^n takim, że $\forall u \in U$

$$\varphi(u) = \varphi(\nabla^*(u_{-1})u),$$

a zatem z części (4) wynika teza.

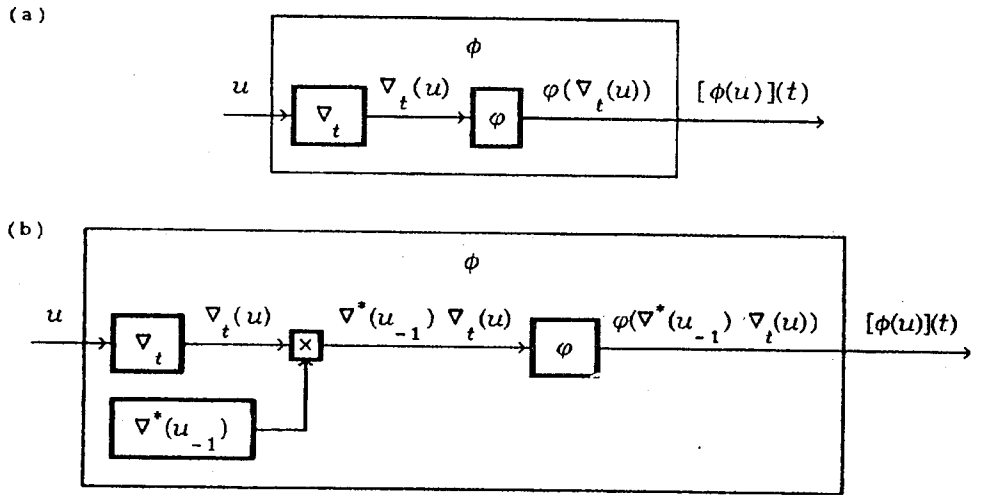
W a r u n e k d o s t a t e c z n y. Zgodnie z częścią (2), jeżeli operator φ jest stacjonarny, to $\forall u \in U$ oraz $\forall t \in R$

$$[\Phi(u)](t) = [\Phi(\nabla_t(u))](0).$$

Z części (3) wynika, że jeżeli operator Φ jest przyczynowy, to $\forall u \in U$ oraz $\forall t \in R$ spełniona jest równość:

$$[\Phi(u)](t) = [\Phi(\nabla_t^*(u_{-1})u)](t),$$

a zatem, przyjmując $t = 0$ otrzymujemy tezę. ∎



Rys.3. Ogólna struktura operatorów stacjonarnych (a) oraz operatorów stacjonarnych przyczynowych (b)

Na rys. 3 przedstawiono ogólną strukturę operatorów stacjonarnych oraz operatorów stacjonarnych przyczynowych. Struktury te wynikają bezpośrednio z udowodnionego wcześniej twierdzenia. Kolejnym ważnym wnioskiem wypływającym z tego twierdzenia jest fakt, że każdy operator stacjonarny Φ na przestrzeni U sygnałów wejściowych o wartościach z F^m można przedstawić jako złożenie przekształcenia φ przestrzeni U w przestrzeń F^n z operatorem ∇_t. Okazuje się, że z przedstawionych w części 4 definicji wynika również bardziej szczegółowa struktura operatorów stacjonarnych oraz operatorów stacjonarnych przyczynowych. Struktura ta wymaga zastosowania teorii dotyczącej aproksymacji średniokwadratowej oraz splotu sygnałów (G. Ciesielski [1]—[5]).

PODSUMOWANIE

W niniejszym artykule omówione zostały zagadnienia związane z identyfikacją strukturalną wielowymiarowych systemów stacjonarnych opisanych za pomocą operatorów, które w ogólnym przypadku mogą być nieliniowymi. Zdefiniowano w sposób ścisły pojęcia stacjonarności i przyczynowości. Ponadto założono, że wartości sygnałów wejściowych i wyjściowych rozważanych systemów mogą przyjmować wartości rzeczywiste lub zespolone. Pokazane zostało również, że bezpośrednio z definicji stacjonarności i przyczynowości systemu wynikają pewne ściśle określone ogólne struktury operatorów opisujących systemy stacjonarne oraz systemy stacjonarne przyczynowe. Struktury te pokazane zostały na rys. 3. Okazuje się, że przedstawione wyniki pozwalają nam wniknąć w bardziej szczegółowe struktury operatorów stacjonarnych przyczynowych (G. Ciesielski [1]—[5]).

BIBLIOGRAFIA

1. G. Ciesielski, S. Derlecki, Z. Marks-Wojciechowska: *Identyfikacja wielowymiarowych systemów pomiarowych opisanych operatorem nieliniowym.* Praca rozpoznawcza nr 112/5/90BP, Politechnika Łódzka, Łódź, 1990
2. G. Ciesielski, S. Derlecki, Z. Marks-Wojciechowska: *Identyfikacja i korekcja nieliniowa wielowymiarowych systemów pomiarowych opisanych operatorem nieliniowym.* Praca badawcza nr MEN 112/6/90 BP, Politechnika Łódzka, Łódź, 1990
3. G. Ciesielski: *Aproksymacja średniokwadratowa operatorów nieliniowych opisujących wielowymiarowe systemy pomiarowe.* XXIII Międzyuczelniana Konferencja Metrologów, Politechnika Warszawska, Warszawa, 1991 (referat nie opublikowany)
4. G. Ciesielski: *Modele Wienera wielowymiarowych systemów pomiarowych opisanych za pomocą operatorów nieliniowych.* Modelowanie i symulacja systemów pomiarowych, Materiały Sympozjum, AGH, Kraków, 1992
5. G. Ciesielski: *Operatorowe modele wielowymiarowych systemów pomiarowych.* Seminarium Naukowe Komisji Kształcenia Komitetu Metrologii i Aparatury Naukowej Polskiej Akademii Nauk, 1994 (referat nie opublikowany)
6. G. Ciesielski: *Identyfikacja ogólnej struktury wielowymiarowych systemów pomiarowych.* Modelowanie i symulacja systemów pomiarowych, Sympozjum, AGH, Kraków, 1994 (referat zgłoszony)
7. R.J.P. de Figueiredo, T.A.W. Dwyer, *A best approximation framework and implementation for simulation of large-scale nonlinear systems.* IEEE Trans. Curcuits Syst., 1980, vol. CAS–27, nr 11, pp. 1005–1014
8. R.J.P. de Figueiredo: Nonlinear circuits and systems applications of functional splines. Proc. of IEEE, 1992, pp. 1245–1247
9. J. Jakubiec: *Bieżące programowe odtwarzanie wartości chwilowych dynamicznych przebiegów wejściowych nieliniowych przetworników pomiarowych.* Rozprawa habilitacyjna, Zeszyty naukowe Politechniki Śląskiej, Elektryka, z. 111, Politechnika Śląska, Gliwice, 1988
10. J. Jaworski, R. Morawski, J. Olędzki: *Wstęp do metrologii i techniki eksperymentu.* WNT, Warszawa, 1992
11. G. Meinardus (tłum. z niem. L.L. Schumaker): *Approximation of Functions: Theory and Numerical Methods.* Springer Tracts in Natural Philosophy, vol. 13, Springer–Verlag, Berlin, 1967
12. J. Park, I.W. Sandberg: *Criteria for the approximation of nonlinear systems.* IEEE Trans. Circuits Syst.: Fundamental Theory and Applications, 1992, vol. 39, nr 8, pp. 673–676
13. K.A. Pupkow, W.I. Kapalin, S. Juszczenko: *Funkcjonalnyje rjady w tjeorii nieliniejnych sisjem.* Nauka, Moskwa, 1976
14. A. Ralston (tłum. z ang. R. Zuber): *Wstęp do analizy numerycznej.* PWN, Warszawa, 1983
15. I.W. Sandberg: *Criteria for the response of nonlinear systems to be L-asymptotic periodic.* Bell Syst. Tech. J., 1981, vol. 60, nr 10, pp. 2359–2371
16. I.W. Sandberg: *Series expansion for nonlinear systems.* Proc. IEEE, 1982, pp. 110–113
17. I.W. Sandberg: *Expansion for nonlinear systems.* Bell Syst. Tech. J., 1982, vol. 61, nr 2, pp. 159–199
18. I.W. Sandberg: *Volterra expansion for time-varying nonlinear systems.* Bell Syst. Tech. J., 1982, vol. 61, nr 2, pp. 201–225
19. I.W. Sandberg: *Multidimensional nonlinear systems and structure theoems.* J. Circuits, Syst. and Computers, 1992, vol. 2, nr 4, pp. 383–388
20. M. Schetzen: *The Volterra and Wiener Theories of Nonlinear Systems.* John Wiley & Sons, New York, 1980
21. M. Szyper: *Modelowanie systemów pomiarowych i ich elementów.* Modelowanie i symulacja systemów pomiarowych, Materiały Sympozjum, AGH, Kraków, 1992
22. R. Zielonko, A. Królikowski: *Metody pomiarowo-diagnostyczne analogowych układów elektronicznych.* WNT, Warszawa, 1988

23. A. Żuchowski: *Symulacyjna metoda wyznaczania momentów charakterystyki impulsowej liniowych przetworników pomiarowych.* Modelowanie i symulacja systemów pomiarowych. Materiały Sympozjum, AGH, Kraków, 1992

G. CIESIELSKI

GENERAL STRUCTURE IDENTIFICATION OF MULTIDIMENSIONAL TIME-INVARIANT SYSTEMS

Summary

In this paper, some problems of general structure identification of multidimensional time-invariant systems governed by operators that can be nonlinear are presented. It was assumed that the inputs and responses of considered systems are real or complex valued. The time-invariance and causality of these systems are formally defined. It is shown that these definitions directly involve some exactly determined general structures of operators which govern of such systems. Obtained structures are the base of the universal theory of system modelling that includes the Volterra and Wiener theories.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI, 1994, 40, z. 3, ss. 429–448

621.3.052

Modelowanie wielowymiarowych liniowych systemów stacjonarnych za pomocą splotów

GRZEGORZ CIESIELSKI

Instytut Podstaw Elektrotechniki, Politechnika Łódzka

Otrzymano 1994.06.09

Autoryzowano 1994.11.10

W artykule omówiono zagadnienia związane z modelowaniem wielowymiarowych systemów stacjonarnych za pomocą splotów. Założono, że sygnały wejściowe i wyjściowe w tych systemach mogą przyjmować wartości rzeczywiste lub zespolone, a zbiór sygnałów wejściowych ma strukturę przestrzeni Hilberta. Ponieważ rozważane systemy są wielowymiarowe, to uogólniono pojęcie splotu funkcji, definiując go dla funkcji w wartościach wektorowych. Udowodniono dwa twierdzenia dotyczące najistotniejszych właściwości tak zdefiniowanych splotów, oraz twierdzenie, z którego wynika związek splotu z iloczynem skalarnym. By związek ten pokazać, wprowadzono pojęcia przestrzeni stacjonarnej, splotowej i przyczynowej. Udowodniono następnie dwa twierdzenia, które rozwiązują zadanie identyfikacji strukturalnej oraz identyfikacji parametrycznej rozważanych systemów.

1. WSTĘP

Przedmiotem artykułu jest modelowanie wielowymiarowych liniowych systemów stacjonarnych za pomocą splotów. Przypomnijmy, że zadanie modelowania dowolnego systemu ma na celu wyznaczenie matematycznego modelu tego systemu. Zadanie to składa się z identyfikacji strukturalnej oraz identyfikacji parametrycznej modelu ([12]). W artykule rozważane są systemy, które dają się opisać za pomocą określonych dalej odwzorowań, nazywanych operatorami. Operatory te są odwzorowaniami liniowymi ciągłymi, dla których jedną z form reprezentacji jest układ liniowych równań różniczkowych. Systemy rozważane w artykule są wielowymiarowe, czyli mają dowolną skończoną liczbę wejść i wyjść. Sygnały na poszczególnych wejściach i wyjściach mogą przyjmować wartości rzeczywiste lub zespolone. Ponadto przyjęto, że zbiór sygnałów wejściowych ma strukturę przestrzeni Hilberta.

W literaturze naukowej jest wiele publikacji omawiających problem modelowania liniowych systemów stacjonarnych ciągłych ([2], [11], [13]–[15], [17]–[19], [21], [23] i [27]). Większość z nich ogranicza się do zadania modelowania systemów z jednym wejściem i wyjściem, które jak wiadomo, mogą być opisane za pomocą splotu funkcji (J. Kudrewicz [14], s. 76). Pewnym problemem jest jednak modelowanie wielowymiarowych liniowych systemów stacjonarnych ciągłych. Wymagają one bowiem wprowadzenia uogólnionej definicji splotu, w której występują funkcje o wartościach wektorowych.

Kolejny problem nie w pełni rozwiązany w literaturze związanej z tematyką tego artykułu, wynika bezpośrednio z właściwości splotów. Okazuje się, że niemal wszystkie najważniejsze twierdzenia dotyczą splotów funkcji całkowalnych z p-tą potęgą (J. Musielak [6], s. 324). Zauważmy jednak, że przestrzenie funkcji całkowalnych z p-tą potęgą nie zawierają chociażby tak podstawowego sygnału, jakim jest skok jednostkowy.

Tą pozorną sprzeczność rozwiązuje się przyjmując, że sygnały występujące w rozważanym liniowym systemie stacjonarnym należą do przestrzeni funkcji całkowalnych z p-tą potęgą w przedziale czasu $[[0, T]$, gdzie chwila T jest dowolnie dużą lecz skończoną liczbą rzeczywistą ([14], [15], [17] i [27]). Podejście to jest z wielu powodów niewygodne w modelowaniu systemów. Zauważmy bowiem, że zmieniając wartość chwili T, zmieniamy jednocześnie całą strukturę takiej przestrzeni, a w szczególności definicję normy, iloczynu skalarnego, czy też bazy ortonormalnej.

Oba przedstawione tu w zarysie problemy związane z modelowaniem wielowymiarowych liniowych systemów stacjonarnych ciągłych, zostaną rozwiązane w niniejszym artykule.

2. OPERATOROWY OPIS SYSTEMÓW

Dowolny system można traktować jako odwzorowanie zbioru U sygnałów wejściowych zwanych *wymuszeniami* w zbiór V sygnałów wyjściowych zwanych *odpowiedziami*. Będziemy dalej zakładać, że oba te zbiory wyposażone są w strukturę przestrzeni liniowej, a zatem określona jest w nich operacja dodawania sygnałów i mnożenia sygnału przez liczbę z ciała liczbowego F[1].

Niech R[2] Będzie zbiorem chwil czasowych, dla których określone są sygnały z U oraz V. Załóżmy ponadto, że $m \in N$[3] jest liczbą wejść, a $n \in N$ liczbą wyjść systemu. Wartości sygnałów wejściowych należą zatem do F^m[4], a wartości sygnałów wyjściowych do F^n. Wówczas przestrzeń sygnałów wejściowych U_F[5] jest podprzest-

[1] W całym artykule przez F oznaczamy ciało liczbowe. W szczególności ciało F oznacza ciało liczb rzeczywistych R lub ciało licz zaspolonych C.

[2] Zbiór X wyposażony w dowolną strukturę algebraiczną i/lub topologiczną będziemy oznaczać literą X, a zatem przez R oznaczamy zbiór liczb rzeczywistych, natomiast przez R ciało liczb rzeczywistych.

[3] Przez N oznaczamy zbiór liczb naturalnych.

[4] Dla dowolnego zbioru X przez X^n oznaczamy n-krotny iloczyn kartezjański (produkt) zbioru X.

[5] Przez X_F oznaczamy przestrzeń liniową nad ciałem skalarów F. Jeżeli F wynika z kontekstu, to będzie ono pomijane w jej oznaczeniu.

rzenią liniowwą przestrzeni $L_F^m(R)$[6]. Podobnie przestrzeń sygnałów wyjściowych V_F jest podprzestrzenią liniową przestrzeni $L_F^n(R)$. Ponieważ U oraz V są przestrzeniami funkcyjnymi, których elementy określone są na tej samej dziedzinie R, a dowolne odwzorowanie Φ takich przestrzeni nazywane jest operatorem, to każdy system można opisać za pomocą operatora, dla którego jedną z możliwych form reprezentacji jest układ równań różniczkowych. By uprościć dalszą dyskusję, pojęcia: system opisany za pomocą operatora Φ i operator Φ, będą więc traktowane jako równoważne.

W modelowaniu systemów wykorzystuje się trzy elementarne operatory: przesunięcie, obrót i splotowe przesunięcie funkcyjne (G. Ciesielski [4]–[10]).

Przesunięciem funkcyjnym o $\sigma \in R^k$ nazywamy operator oznaczany przez $\nabla_\sigma : L_F^m(R^k) \to L_F^m(R^k)$, który $\forall u \in L_F^m(R^k)$ i $\forall t \in R^k$ spełnia warunek:

$$[\nabla_\sigma(u)](t) := u(t + \sigma).$$

Funkcję $\nabla_s \circ u$ nazywamy *przesunięciem funkcji u* o σ.

Obrotem funkcyjnym nazywamy operator, który będziemy oznaczać przez $\nabla^* : L_F^m(R^k) \to L_F^m(R^k)$, taki, że $\forall u \in L_F^m(R^k)$ i $\forall t \in R^k$

$$[\nabla^*(u)](t) := u(-t).$$

Funkcję $\nabla^* \circ u$ nazywamy *obrotem funkcji u*.

Splotowym przesunięciem funkcyjnym o $\sigma \in R^k$ nazywać będziemy operator, który jest oznaczony dalej przez $\nabla_\sigma^* : L_F^m(R^k) \to L_F^m(R^k)$ oraz $\forall u \in L_F^m(R^k)$ i $\forall t \in R^k$

$$[\nabla_\sigma^*(u)](t) := u(\sigma - t).$$

Funkcję $\nabla_\sigma^* \circ u$ nazywać będziemy *przesunięciem splotowym funkcji u* o σ.

Zauważmy tu, że przesunięcia ∇_σ i ∇_σ^* są operatorami liniowymi na przestrzeni $L_F^n(R^k)$ i odwzorowują przestrzeń $L_F^n(R^k)$ w nią samą. Ponadto

$$\nabla_\sigma \circ \nabla_{-\sigma} = \mathrm{id}^{[7]}, \qquad \nabla_\sigma^* \circ \nabla_\sigma^* = \mathrm{id},$$

$$\nabla_\sigma^* = \nabla^* \circ \nabla_\sigma = \nabla_{\sigma-} \circ \nabla^* \quad \text{oraz} \quad \nabla_0^* = \nabla^*.$$

Zauważmy również, że przesunięcie funkcji $\nabla_\sigma \circ (u)$ jest funkcją u przesuniętą w lewo o σ, obrót funkcji $\nabla^* \circ (u)$ jest funkcją u obróconą o π wokół osi rzędnych, a splotowe przesunięcie funkcyjne ∇_σ^* jest złożeniem operacji ∇^* z ∇_σ, przy czym złożenia tego nie można wykonać w dowolnej kolejności (G. Ciesielski [10]).

Przejdźmy teraz do definicji podstawowych właściwości operatorów opisujących systemy.

Operator Φ na $U_F \subset L_F^m(R)$ nazywamy *stacjonarnym*, jeżeli

$$\forall u \in U \quad \forall \sigma \in R \quad \big((\nabla_\sigma(u \in U) \wedge (\nabla_\sigma(\Phi(u)) = \Phi(\nabla_\sigma(u)))\big).$$

---

[6] Przez $L_F^n(X)$ oznaczamy przestrzeń odwzorowań określonych na X o wartościach z F^n. Oznaczenia $L_F^1(X)$ oraz $L_F(X)$ będziemy traktować jako równoważne, a gdy zbiór X wynika z kontekstu, to piszemy w skrócie L_F^n lub L_F. Przestrzeń L_F^n będziemy również traktować jako n-krotny produkt przestrzeni L_F.

[7] Przez id oznaczamy odwzorowanie identycznościowe (identyczność).

Innymi słowy, operator jest stacjonarny, gdy odpowiedź operatora na wymuszenie u przesunięte w czasie o σ jest równa przesuniętej w czasie o σ odpowiedzi operatora na wymuszenie u. Zauważmy, że przestrzeń liniowa U musi być zbiorem zamkniętym ze względu na przesunięcia funkcyjne, a zatem zawierać musi wszystkie przesunięcia funkcyjne swoich elementów. Kolejną ważną właściwością operatora jest przyczynowość.

Operator Φ na $U_\Gamma \subset L^m_\Gamma(R)$ nazywamy *przyczynowym*, jeżeli

$$\forall u, v \in U \quad \forall t \in R$$
$$(\forall \tau \in R\ ((\tau \leqslant t) =: (u(\tau) = v(\tau))) =: ([\Phi(u)](t) = [\Phi(v)](t))).$$

Wynika stąd, że operator jest przyczynowy, gdy wartość sygnału wyjściowego w chwili t zależy jedynie od wartości sygnału wejściowego do chwili t.

3. SPLOTY

Sploty funkcji są odwzorowaniami, które umożliwiają modelowanie liniowych systemów stacjonarnych oraz są elementami modeli nieliniowych systemów stacjonarnych. Zanim jednak zdefiniujemy splot funkcji, wprowadzimy kilka pojęć bezpośrednio z nim związanych.

Załóżmy, że μ jest miarą na σ-ciele w X.

Jeżeli $1 \leqslant p < \infty$, to *przestrzenią* $L^m_{\Gamma p\mu}(X)$ nazywamy zbiór klas równoważności ze względu na relację $\mu - Æ$[8], funkcji mierzalnych (μ-mierzalnych) $u: X \to F^m$ takich, że

$$\int_X |u|^p \, d\mu < \infty.$$

Normą w $L^m_{\Gamma p\mu}(X)$ nazywamy funkcję, którą $\forall u \in L^m_{\Gamma p\mu}(X)$ oznaczamy przez

$$\|u\|_p := \left(\int_X |u|^p \, du \right)^{1/p}.$$

Funkcje należące do $L^m_{\Gamma 1\mu}(X)$ nazywamy *całkowalnymi* (*w sensie Lebesgue'a*).

Przestrzenią $L^m_{\Gamma \infty\mu}(X)$ nazywamy zbiór klas równoważności ze względu na relację $\mu - Æ$, funkcji mierzalnych (μ-mierzalnych) $u: X \to F^m$, dla których

$$\operatorname*{supess}_{x \in X} |u(x)| < \infty,$$

gdzie

$$\operatorname*{supess}_{x \in X} |u(x)| := \inf \{M \in \overline{R}_+ : \mu - Æ\ x \in X(|u(x)| \leqslant M)\}$$[9],

[8] Przez $\mu - Æ$ oznaczamy zwrot „μ – prawie wszędzie" lub „dla μ – prawie każdego".

[9] Przez $\overline{R}_+$ oznaczamy zbiór liczb rzeczywistych nieujemnych uzupełniony o liczbę ∞.

nazywamy *istotnym kresem górnym* funkcji $|u|$. *Normą* w $L^m_{F\infty\mu}(X)$ nazywamy funkcję, którą $\forall u \in L^m_{F\infty\mu}(X)$ oznaczamy przez

$$\|u\|_\infty := \operatorname*{supess}_{x \in X} |u(x)|.$$

Funkcje należące do przestrzeni $L^m_{F\infty\mu}(X)$ nazywamy *istotnie* (μ-*istotnie*) *ograniczonymi* (J. Musielak [16], s. 53 i 59).

Załóżmy dalej, że $1 \leqslant p \leqslant \infty$.

Przez $C^m_{Fp\mu}(X)$ będziemy oznaczać *podprzestrzeń funkcji ciągłych z przestrzeni* $L^m_{Fp\mu}(X)$.

Jeżeli zbiór X wynika z kontekstu, to będziemy pisać w skrócie $L^m_{Fp\mu}$ lub $C^m_{Fp\mu}$. Oznaczenia $L^1_{Fp\mu}$ i $C^1_{Fp\mu}$ oraz odpowiednio $L_{Fp\mu}$ i $C_{Fp\mu}$ będziemy traktować jako równoważne. W przypadku gdy miara μ jest miarą Lebesgue'a na R^k., którą będziemy oznaczać przez λ_k, w skrócie λ, to oznaczenia $L^m_{Fp\lambda}$ i $C^m_{Fp\lambda}$ oraz odpowiednio L^m_{Fp} i C^m_{Fp}, jak też $\lambda - Æ$ oraz Æ, są równoważne.

Przejdźmy teraz do definicji splotu funkcji.

Niech λ_k będzie miarą Lebesgue'a na R^k. *Splotem funkcji* mierzalnych u, $v: R^k \to F^m$ nazywać będziemy funkcję oznaczoną przez $u * v: R^k \to F$, która o ile istnieje, spełnia warunek:

$$\lambda^k - Æ\, t \in R^k \left([u * v](t) := \int_{R^k} u^T(t - \tau) v(\tau)\, d\tau \right).$$

Zwykle splot definiuje się dla funkcji o wartościach z F, a nie jak tutaj z F^m. Takie rozszerzenie pojęcia splotu wynika z jego związku z iloczynem skalarnym, o czym będzie mowa dalej. Zauważmy również, że całka Lebesgue'a może przyjmować wartości nieskończone, które nie należą do ciała F[10], podczas gdy wartości splotu funkcji należą do F. Z warunku definicyjnego wynika, że splot funkcji istnieje, gdy zbiór tych $t \in R^m$, dla których całka przyjmuje wartości spoza F jest miary zero.

Kolejne dwa twierdzenia dotyczą najistotniejszych właściwości zdefiniowanego wcześniej splotu funkcji.

---

[10] Załóżmy dla uproszczenia, że F = R. Mówimy, że *istnieje całka Lebesgue'a* funkcji μ-mierzalnej $f: X \to R$, jeżeli

$$\int_X f\, d\mu \in \overline{R},$$

gdzie przez $\overline{R}$ oznaczamy zbiór liczb rzeczywistych R uzupełniony o liczby $-\infty$ i $+\infty$. Z istnienia całki Lebesgue'a nie wynika zatem całkowalność funkcji f.

Twierdzenie 1. *Jeżeli* $1 \leqslant p,\ q \leqslant \infty,\ \frac{1}{p}+\frac{1}{q} \geqslant 1,\ \frac{1}{r} := \frac{1}{p}+\frac{1}{q}-1,\ u \in L^m_{Fp}(R^k)$ *oraz* $v \in L^m_{Fq}(R^k)$, to *prawdziwe są twierdzenia.*

(1) $\lambda - Æ t \in R^k \left(\int\limits_{R^k} |u^T(t-\tau)v(\tau)|\,d\tau < \infty \right)$.

(2) Jeżeli $\|\cdot\|_p$, $\|\cdot\|_q$ *oraz* $\|\cdot\|_r$ *są normami odpowiednio w* $L^m_{Fp}(R^k)$, $L^m_{Fq}(R^k)$ *oraz* $L_{Fr}(R^k)$, *to*

$$\exists u * v \in L_{Fr}(R^k)\ (\|u*v\|_r \leqslant \|u\|_p \|v\|_q).$$

D o w ó d w książce J. Musielaka ([16)], s. 325) przedstawiono dla $m = 1$, a stąd wynika dowód dla dowolnego $m \in N$. ∎

Twierdzenie 2. *Jeżeli* $1 \leqslant p,\ q \leqslant \infty, \frac{1}{p}+\frac{1}{q} = 1,\ u \in L^m_{Fp}(R^k)$ *oraz* $v \in L^m_{Fq}(R^k)$, *to prawdziwe są twierdzenia.*

(1) *Jeżeli* $\|\cdot\|_p$ i $\|\cdot\|_q$ *są normami odpowiednio* w $L^m_{Fp}(R^k)$ i $L^m_{Fq}(R^k)$, *to*

$$\forall t \in R^k \left(\int\limits_{R^k} |u^T(t-\tau)v(\tau)|\,d\tau \leqslant \|u\|_p \|v\|_q \right).$$

(2) $\exists u * v \in C_{F\infty}(R^k)\quad \forall t \in R^k \quad \left([u*v](t) = \int\limits_{R^k} u^T(t-\tau)v(\tau)\,d\tau \right)$.

D o w ó d.

(1) Ponieważ miara Lebsesgue'a jest symetryczna i niezmiennicza względem przesunięć[11] oraz prawdziwa jest nierówność

$$\|u^T\|_1 \leqslant \|u\|_p \|v\|_q$$

(W. Rudin [22], s. 74), to $\forall t \in R^k$

---

[11] Niech μ będzie miarą na R^n. *Miarę* μ nazywamy *symetryczną*, jeżeli dla każdego mierzalnego zbioru $A \subset R^n$ zachodzi równość:

$$\mu(A) = \mu(-A),$$

gdzie

$$-A := \{-a : a \in A\}.$$

Z kolei *miara* μ jest *niezmiennicza względem przesunięć*, jeżeli dla każdego mierzalnego zbioru $A \subset R^n$ i dla każdego $x \in R^n$ spełniona jest równość:

$$\mu(A) = \mu(A + x),$$

gdzie

$$A + x := \{a + x : a \in A\}.$$

$$\int_{R^k} |u^T(t-\tau)\, v(\tau)\, d\tau \leqslant \|\nabla_\tau^*(u)\|_p \|v\|_q = \|u\|_p \|v\|_q .$$

a stąd otrzymujemy tezę.

(2) Z części (1) wynika, że funkcja

$$w :\!= u * v$$

istnieje dla $\forall t \in R^k$, a z twierdzenia 1.(2) wynika, że należy ona do $L_{F\infty}$. Wystarczy zatem wykazać, że funkcja w jest ciągła.

Załóżmy zatem, że $t \in R^k$, $(t_i \in R^k)_N$ oraz $\lim_{i\to\infty} t_i = t$. Wówczas

$$|w(t) - w(t_i)| = \left|\int_{R^k} (u(t-\tau) - u(t_i-\tau))^T v(\tau)\, d\tau\right| \leqslant$$

$$\leqslant \int_{R^k} |(u(t-\tau) - u(t_i-\tau))^T v(\tau)|\, d\tau = \int_{R^k} |(\nabla_t^*(u) - \nabla_{t_i}^*(u))^T v|,$$

Z nierówności:

$$\|u^T v\|_1 \leqslant \|u\|_p \|v\|_q$$

(W. Rudin [22], s. 74) mamy, że

$$|w(t) - w(t_i)| \leqslant \|\nabla_{t_i}^*(w)\|_p \|v\|_q < \infty .$$

Ponieważ splot jest przemienny, to możemy przyjąć, że $p < \infty$. Dowodzi się wówczas (W. Rudin [22], s. 158), że istnieje takie $\alpha :\!= (\alpha_j \in F)_N$, zbiory mierzalne $(A_j \in R^k)_N$ oraz funkcje charakterystyczne $(\chi(A_j, \cdot))_N$, dla których

$$\|u\|_p :\!= \left(\int_{R^k} |u|^p\right)^{1/p} = \left(\int_{R^k} (\alpha^T(\chi(A_i, \tau))_N)^p\, d\tau\right)^{1/p}.$$

Miara Lebesgue'a jest symetryczna i niezmiennicza względem przesumięć, a zatem

$$(\|\nabla_t^*(u) - \nabla_{t_i}^*(u)\|_p :\!= \int_{R^k} |u(t-\tau) - (t_i - \tau)|^p\, d\tau =$$

$$= \int_{R^k} (\alpha^T |\chi(A_j, t-\tau) - \chi(A_j, t_i - \tau)|_{j\in N})^p\, d\tau =$$

$$= \int_{R^k} (\alpha^T |\chi(t - A_j, \tau) - \chi(t_i - A_j, \tau)|_{j\in N})^p\, d\tau =$$

$$\int_{R^k} (\alpha^T(\chi((t-A_j) \wedge (t_i - A_j), \tau))_{j\in N})^p\, d\tau^{(12)},$$

(12) Dla dowolnych zbiorów A i B przez $A \wedge B$ oznaczamy ich różnicę symetryczną.

Z ciągłości miary Lebesgue'a λ_k (P. Billingsley [3]. s. 162) wynika, że $\forall j \in \mathrm{N}$ granica

$$\lim_{i \to \infty} \lambda_k((t - A_j) \wedge (t_i - A_j)) = 0,$$

czyli

$$\lim_{i \to \infty} |w(t) - w(t_i)| \leqslant \|v\|_q \lim_{i \to \infty} \|\nabla_t^*(u) - \nabla_{t_i}^*(u)\|_p = 0,$$

co kończy dowód. ∎

Z ostatnich dwóch twierdzeń wynika, że $\forall u, v \in \mathrm{L}_{\mathrm{F}_2}^m$ splot $u * v$ możemy zdefiniować $\forall t \in \mathrm{R}^k$ w następujący sposób:

$$[u * v)(t)] := \int_{R^k} u^T(t - \tau) v(\tau)\, \mathrm{d}\tau,$$

Jest on wówczas funkcją ciągłą i ograniczoną na R^k, a zatem $u * v \in \mathrm{C}_{\mathrm{F}\infty}$.

4. PRZESTRZENIE STACJONARNE, SPLOTOWE I PRZYCZYNOWE

Przekonamy się dalej, że przestrzenie stacjonarne, splotowe i przyczynowe odgrywają bardzo ważną rolę w modelowaniu systemów stacjonarnych.

Załóżmy, że μ jest miarą na R^k oraz $1 \leqslant p \leqslant \infty$.

Przestrzenią stacjonarną oznaczaną przez $\mathrm{L}_{\mathrm{F}p\mu\nabla}^m(\mathrm{R}^k)$, nazywamy podzbiór klas równoważności z przestrzeni $\mathrm{L}_{\mathrm{F}p\mu}^m(\mathrm{R}^k)$ ze względu na relację $\mu - Æ$ o postaci:

$$\mathrm{L}_{\mathrm{F}p\mu\nabla}^m(\mathrm{R}^k) := \{u \in \mathrm{L}_{\mathrm{F}p\mu}^m(\mathrm{R}^k) : \forall t \in \mathrm{R}^k (\nabla_t(u) \in \mathrm{L}_{\mathrm{F}p\mu}^m(\mathrm{R}^k))\}.$$

Przestrzenią splotową oznaczaną przez $\mathrm{L}_{\mathrm{F}p\mu*}^m(\mathrm{R}^k)$, nazywamy podzbiór klas równoważności przestrzeni $\mathrm{L}_{\mathrm{F}p\mu}^m(\mathrm{R}^k)$, ze względu na relację $\mu - Æ$ o postaci:

$$\mathrm{L}_{\mathrm{F}p\mu*}^m(\mathrm{R}^k) := \{u \in \mathrm{L}_{\mathrm{F}p\mu}^m(\mathrm{R}^k) : \forall t \in \mathrm{R}^k (\nabla_t^*(u) \in \mathrm{L}_{\mathrm{F}p\mu}^m(\mathrm{R}^k))\}.$$

Przestrzenią przyczynową oznaczaną przez $\mathrm{L}_{\mathrm{F}p\mu+}^m(\mathrm{R}^k)$, nazywamy podzbiór klas równoważności z przestrzeni $\mathrm{L}_{\mathrm{F}p\mu}^m(\mathrm{R}^k)$ ze względu na relacje $\mu - Æ$ o postaci:

$$\mathrm{L}_{\mathrm{F}p\mu+}^m(\mathrm{R}^k) := \{u_{-1} u : u \in \mathrm{L}_{\mathrm{F}p\mu}^m(\mathrm{R}^k)\}^{(13)}.$$

---

(13) Przez u_{-1} oznaczamy *skok jednostkowy*, który $\forall t := (t_i \in \mathrm{R})_1^k$ zdefiniowany jest przez wzór:

$$u_{-1}(t) := \begin{cases} 1, & \forall i \in \{1, \ldots, k\} \quad (t_i \geqslant 0) \\ 0 & \exists i \in \{1, \ldots, k\} \quad (t < 0) \end{cases}.$$

Jeżeli $u : X \to \mathrm{F}$ oraz $v : \mathrm{X} \to \mathrm{F}^n$, to $\forall x \in \mathrm{X}$ *iloczyn funkcji* u i v oznaczany przez uv, zdefiniowany jest za pomocą wzoru:

$$[uv](x) := u(x) v(x).$$

Jeżeli zbiór R^k wynika z kontekstu, to będziemy pisać w skrócie $L^m_{Fp\mu\nabla}$, $L^m_{Fp\mu^*}$ oraz $L^m_{Fp\mu+}$. Oznaczenia $L^1_{Fp\mu\nabla}$, $L^1_{Fp\mu^*}$ i $L^1_{Fp\mu+}$ oraz odpowiednio $L_{Fp\mu\nabla}$, $L_{Fp\mu^*}$ i $L_{Fp\mu+}$ będziemy traktować jako równoważne. W przypadku gdy miarą na R^k jest miara Lebezgue'a, to jej symbol będzie pomijany w oznaczeniach zdefiniowanych przestrzeni.

Zauważmy, że zgodnie z definicją stacjonarności operatora, sformułowaną w części 2 tego artykułu, przestrzeń wymuszeń operatora stacjonarnego musi zawierać wszystkie przesunięcia funkcyjne swoich elementów. Przestrzeń stacjonarna $L^m_{Fp\mu\nabla}(R)$ ma tą właściwość, co pozwala nam definiować na niej operatory stacjonarne. Zauważmy tu również, że podobnie przestrzeń splotowa $L^m_{Fp\mu^*}(R)$ zawiera wszystkie splotowe przesunięcia funkcyjne swoich elementów. Z kolei elementy składające się na przestrzeń przyczynową $L^m_{Fp\mu+}(R)$ są funkcjami, które przyjmują wartość 0 dla ujemnych wartości zmiennej niezależnej. *Sygnały* należące do tej przestrzeni są *przyczynowymi* (A. Wojnar [26], s. 48).

Wniosek 1. *Załóżmy, że* μ *jest miarą na* R^k *oraz* $1 \leqslant p \leqslant \infty$.
(1) $L^m_{Fp\mu\nabla}$, $L^m_{Fp\mu^*}$ i $L^m_{Fp\mu+}$ *są podprzestrzeniami liniowymi przestrzeni* $L^m_{Fp\mu}$.
(2) $L^m_{Fp\mu\nabla} \subset L^m_{Fp\mu^*}$.
(3) *Jeżeli* μ *jest miarą symetryczną, to* $L^m_{Fp\mu\nabla} = L^m_{Fp\mu^*}$.
(4) $L^m_{Fp+} \subset L^m_{Fp\nabla} = L^m_{Fp^*} = L^m_{Fp}$.

D o w ó d.
(1) Oczywisty.
(2) Ponieważ $\forall t \in R^k$ i $\forall u \in L^m_{Fp\mu^*}$ $\nabla^*_t(u)$, $(\nabla^*_t(u)) \in L^m_{Fp\mu^*}$, to $\nabla_t(u) \in L^m_{Fp\mu^*}$, gdyż $\nabla^*(\nabla^*_t(u)) = \nabla_t(u)$.
(3) Oczywisty.
(4) Zawieranie jest oczywiste, a równości przestrzeni wynikają z faktu, że miara Lebesgue'a jest symetryczna i niezmiennicza względem przesunięć. ∎

Można wykazać (G. Ciesielski [4]—[8]), że istnieją również takie miary μ, różne od miary Lebesgue'a, że

$$L^m_{Fp\mu\nabla}(R) = L^m_{Fp\mu^*}(R) = L^m_{Fp\mu}(R),$$

a przestrzeń $L^m_{Fp\mu}$ zawiera wszystkie funkcje typu wykładniczego które jak wiadomo (J. Osiowski [17], s. 88), są bezwzględnie transformowalne w sensie jednostronnego przekształcenia Laplace'a. Przestrzeń ta jest więc wystarczająco uniwersalna by ją wybrać jako przestrzeń wymuszeń i odpowiedzi modelowanych systemów. Dlatego też zostanie ona szczegółowo omówiona w jednym z kolejnych artykułów. By jednak zaspokoić już teraz ciekawość Czytelnika należy tu wymienić książkę W.M. Amierbajewa i N.A. Utiembajewa [1], z której można wnioskować o strukturze wspomnianej przestrzeni.

5. ZWIĄZEK SPLOTU Z ILOCZYNEM SKALARNYM

Okazuje się, że splot funkcji jest ściśle powiązany z iloczynem skalarnym.

Twierdzenie 3. *Jeżeli* ω *jest gęstością miary* μ *względem miary Lebesgue'a na* R^k[14], $u \in L^m_{F2\mu^*}(R^k)$, $v \in L^m_{F2\mu}(R^k)$ *oraz* $\langle . , . \rangle_\mu$ *jest iloczynem skalarnym w* $L^m_{F2\mu}(R^k)$, *to* $\forall t \in R^k$

$$[w\bar{v} * u](t) = \langle \nabla_t^*(u), v \rangle_\mu \quad (1)$$[15].

D o w ó d. Zauważmy, że $\forall t \in R^k$ funkcja $\nabla_t^*(u)$ należy do przestrzeni $L^m_{F2\mu}$, a ztaem $\forall t \in R^k$ istnieje iloczyn skalarny we wzorze (1). Z kolei splot

$$[w\bar{v} * u](t) = \int_{R^k} w(t-\tau)v^*(t-\tau)u(\tau)\,d\tau = \int_{R^k} w(\tau)v^*(\tau)u(t-\tau)\,d\tau =$$

$$= \int_{R^k} v^*(\tau)u(t-\tau)w(\tau)\,d\tau = \int_{R^k} v^*(\tau)u(t-\tau)\,d\mu(\tau) = \langle \nabla_t^*(u), v \rangle_\mu ,$$

a zatem twierdzenie zostało udowodnione. ∎

Zobaczymy dalej, że twierdzenie to odegra istotną rolę w identyfikacji strukturalnej rozważanych systemów.

6. IDENTYFIKACJA STRUKTURALNA

Zauważmy, że systemy, które dają się opisać za pomocą splotu funkcji, odwzorują przestrzeń liniową sygnałów wejściowych $L^m_{Fp}(R)$ lub $L^m_{Fq}(R)$ w przestrzeń liniową sygnałów wyjściowych $L_{Fr}(R)$, o czym mówi twierdzenie 1.(2). Można łatwo sprawdzić, że systemy takie są liniowe i stacjonarne. Z kolei twierdzenia 1 oraz 2 mówią niemal o wszystkich najważniejszych właściwościach splotów, a więc i o właściwościach systemów opisanych za ich pomocą. Okazuje się jednak, że twierdzeń tych nie można bezpośrednio wykorzystać do modelowania liniowych

---

[14] Będziemy zakładać, że gęsstość w miary μ na X względem miary ν również na X jest funkcją ν-mierzalną o wartościach z $\overline{R}_+$. Piszemy wówczas

$$d\mu := w\,d\nu.$$

[15] Oznaczmy dalej przez A^* transpozycję ze sprzężeniem macierzy A, a zatem

$$A^* := \overline{A}^T.$$

Iloczyn skalarny w przestrzeni $L^n_{F2\mu}(X)$ definiujemy w następujacy sposób:

$$\forall u, v \in L^n_{F2\mu}\left(\langle u, v \rangle := \int_{R^x} v^* u\,d\mu\right).$$

systemów stacjonarnych. Głównym tego powodem jest fakt, że przestrzeń sygnałów wejściowych L^m_{Fp} lub L^m_{Fq} dla skończonej wartości p lub q, nie zawiera tak podstawowego sygnału, jakim jest na przykład skok jednostkowy. Spostrzeżenie to mogłoby sugerować, że za pomocą splotu funkcji nie można modelować nawet liniowych systemów stacjonarnych. Z kolejnego twierdzenia wynika, że nie jest to jednak prawdą.

Twierdzenie 4. *Jeżeli w jest taką gęstością miary μ względem miary Lebegue'a na* R, *że miara μ jest symetryczna oraz*

$$L^m_{F2\mu\nabla}(R) = L^m_{F2\mu}(R),$$

to dla dowolnego liniowego operatora stacjonarnego ciągłego $\Phi: L^m_{F2\mu}(R) \to C_{F\infty\mu}(R)$ prawdziwe są następujące twierdzenia.

(1) *Istnieje dokładnie jedna funkcja $g \in L^m_{F2\mu}(R)$ taka, że $\forall u \in L^m_{F2\mu}(R)$*

$$\Phi(u) = wg * u.$$

Załóżmy dalej, że $g \in L^m_{F2\mu}(R)$ jest funkcją o której mowa w części (1).

(2) *Jeżeli $\|\cdot\|_2$ jest normą w $L^m_{F2\mu}(R)$,* to

$$\|\Phi\| = \|g\|_2 \text{ [16]},$$

Załóżmy dalej, że operator φ jest przyczynowy.

(3) $g \in L^m_{F2\mu+}(R)$.

(4) *Jeżeli miara μ jest skończona, to $wg \in L^m_{F1+}(R)$.*

(5) *Jeżeli $wg \in L^m_{F1}(R)$ oraz $wg \in L^m_{F\infty}(R)$,* to $wg \in L^m_{F2+}(R)$.

(6) *Jeżeli $\langle . , . \rangle$ jest iloczynem skalarnym* w $L^m_{F2}(R)$, $v \Leftarrow (v_i)_N$ *jest bazą ortonormalną w $L^m_{F2+}(R)$ oraz $wg \in L^m_{F2}(R)$, to istnieje taki wektor*

$$c := \langle wg, v\rangle := (\langle wg, v_i\rangle_N \text{ [17]},$$

dla którego $\forall u \in L^m_{F2\mu}(R)$ *mamy*

$$\Phi(u) = (c^T v) * u.$$

---

[16] Niech X i Y będą przestrzeniami unormowanymi z normami odpowiednio $\|\cdot\|_X$ i $\|\cdot\|_Y$. Odwzorowanie $\Phi: X \to Y$ nazywane dalej *przekształceniem*, jest *ograniczone*, jeżeli $\exists M > 0$, że $\forall x \in X$

$$\|\Phi(x)\|_Y \leqslant M\|x\|_X,$$

a gdy φ jest przekształceniem liniowym ciągłym, to *normą przekształcenia φ* nazywamy liczbę oznaczaną przez

$$\|\Phi\| \Leftarrow \inf\{M > 0 : \forall x \in X(\|\Phi(x)\|_Y \leqslant M\|x\|_X)\}$$

[17] Jeżeli $f: X \to Y$ oraz

$$A := (a_{ij} \in X)^{m,n}_{i,j=1},$$

to dla uproszczenia zapisu przyjmujemy, że

$$f(A) := (f(a_{ij}))^{m,n}_{i,j=1},$$

o ile nie będzie to prowadzić do nieporozumień.

D o w ó d.

(1) Niech

$$\varphi := [\Phi(\cdot)](0).$$

Ponieważ operator Φ jest stacjonarny oraz z założenia przestrzeń

$$L^m_{F2\mu\nabla} = L^m_{F2\mu},$$

to można wykazać (G. Ciesielski [10]), że

$$\Phi = \varphi \circ \nabla.$$

Z założenia ciągłości operatora φ i z twierdzenia Banacha (J. Musielak [16], s. 139) wynika jego ograniczoność, a zatem $\forall u \in L^m_{F2\mu}$

$$\|\Phi(u)\|_\infty \leqslant \|\Phi\| \|u\|_2,$$

gdzie $\|\cdot\|_\infty$ jest normą w $L^m_{F\infty\mu} \supset C_{F\infty\mu}$, a $\|\cdot\|_2$ jest normą w $L^m_{F2\mu}$. Ponieważ elementy przestrzeni $C_{F\infty\mu} \subset L_{F\infty\mu}$ są funkcjami ciągłymi, to

$$\|\Phi(u)\|_\infty := \operatorname*{supess}_{t\in R} |[\Phi(u)](t)| = \sup_{t\in R} |\Phi(u)](t)|,$$

a stąd

$$|\varphi(u)| \leqslant \|\Phi\| \|u\|_2,$$

czyli funkcjonał φ jest rownież ograniczony oraz

$$\|\varphi\| = \|\Phi\|.$$

Z kolei z założenia liniowości operatora φ wynika, że funkcjonał φ też jest liniowy.

Przestrzeń $L^m_{F2\mu}$ z iloczynem skalarnym, który $\forall u, v \in L^m_{F2\mu}$ określony jest przez wzór:

$$\langle u, v\rangle := \int_R v^* u \, d\mu,$$

jest przestrzenią Hilberta (J. Musielak [16], s. 53). Dowodzi się wówczas, że dla każdego ciągłego lub zgodnie z twierdzeniem Banacha (J. Musielak [16], s. 139) ograniczonego funkcjonału liniowego $\varphi : L^m_{F2\mu} \to F$ istnieje dokładnie funkcja $f \in L^m_{F2\mu}$ taka, że

$$\varphi =: \langle \cdot, f\rangle$$

(W. Rudin [22], s. 90). Niech

$$g := \nabla^*(f).$$

Z założenia symetrii miary μ wynika, że funkcja $\bar{g} \in L^m_{F2\mu}$, a zatem również i funkcja $g \in L^m_{F2\mu}$. Ponadto $\forall u \in L^m_{F2\mu}$

$$\varphi(u) = \langle u, f\rangle = \langle \nabla^*(u), \bar{g}\rangle.$$

Z kolei z wniosku 1.(3) oraz z założenia:

$$L^m_{F2\mu\nabla} = L^m_{F2\mu}$$

wynika równość:

$$L^m_{F2\mu^*} = L^m_{F2\mu},$$

a stąd i z twierdzenia 3 mamy, że $\forall t \in R$

$$[\Phi(u)](t) = \varphi(\nabla_t(u)) = \langle \nabla^*_t(u), \bar{g} \rangle = [wg * u](t).$$

(2) Z nierówności Schwarza (J. Musielak [16], s. 67) $\forall u \in L^m_{F2\mu}$ mamy

$$|\varphi(u)| = |\langle \nabla^*(u), \bar{g} \rangle| \leqslant \|g\|_2 \|u\|_2.$$

Z kolei

$$|\varphi(\nabla^*(\bar{g}))| = |\langle \nabla^*(\nabla^*(\bar{g})), \bar{g} \rangle| = |\langle \bar{g}, \bar{g} \rangle| = \|g\|_2^2.$$

Można zatem wywnioskować, że

$$\|\varphi\| = \|g\|_2,$$

a stąd oraz z równości

$$\|\Phi\| = \|\varphi\|$$

mamy

$$\|\Phi\| = \|g\|_2.$$

(3) Dowód oczywisty.

(4) Ponieważ miara μ jest z założenia skończona, to można dowieść (J. Musielak [16], s. 54), że

$$L^m_{F2\mu} \subset L^m_{F1\mu},$$

a zatem $g \in L^m_{F1\mu}$, czyli $wg \in L^m_{F1}$. Stąd i z części (3) wynika ostatecznie, że $wg \in L^m_{F1+}$.

(5) Niech $\|\cdot\|_1$, $\|\cdot\|_2$ oraz $\|\cdot\|_\infty$ będą normami odpowiednio w $L^m_{F_1}, L^m_{F_2}$ oraz $L^m_{F_\infty}$. Ponieważ z założenia $wg \in L^m_{F_1} \cap L^m_{F_\infty}$, to można wykazać (W. Rudin [22], s. 74), że

$$\|wg\|_2 = \|wg^* wg\|_1 \leqslant \|wg\|_1 \|wg\|_\infty < \infty,$$

a zatem $wg \in L^m_{F_2}$. Stąd i z części (3) wynika ostatecznie, że $wg \in L^m_{F_2+}$.

(6) Istnienie wektora c jest oczywiste. Niech zatem $\|\cdot\|_2$ będzie normą w $L^m_{F_2}$. Ponieważ z części (3) wynika, że $wg \in L^m_{F2+}$, to można wykazać (J. Musielak [16], s. 77), że

$$\|wg - c^T v\|_2 = 0,$$

czyli

$$wg - c^T v = 0$$

prawie wszędzie na R, a stąd

$$(wg * u)\big((c^T v)\big) * u) = (wg - c^T v) * u = 0 * u = 0,$$

co kończy dowód. ∎

Z udowodnionego twierdzenia wynika, że każdy liniowy system stacjonarny φ ciągły na $L^m_{F2\mu}$ można opisać za pomocą splotu funkcji. Jeżeli założymy przy tym,

że miara μ na R jest skończona, to odpowiedź impulsowa *wg* systemu Φ jest całkowalna, a gdy ponadto odpowiedź *wg* jest istotnie ograniczona, to jest ona również całkowalna w kwadracie. Widać też, że struktura liniowego systemu stacjonarnego ciągłego Φ jest powiązana ze strukturą przestrzeni wymuszeń $L^m_{F2\mu}$ przez gęstość w miary μ.

Warto tutaj zauważyć, że założenie skończoności miary μ nie ogranicza obszaru zastosowań udowodnionego twierdzenia. Założenie to sprawia bowiem, że przestrzeń $L^m_{F2\mu}$ zawiera chociażby tak ważny sygnał jakim jest skok jednostkowy. Można też wówczas wykazać (J. Musielak [16], s. 54), że zachodzi bardzo wygodne w modelowaniu systemów zawieranie przestrzeni:

$$L^m_{F\infty\mu} \subset \ldots \subset L^m_{Fq\mu} \subset L^m_{Fp\mu} \subset \ldots \subset L^m_{F1\mu},$$

gdzie $1 \leqslant p < q \leqslant \infty$.

Zauważmy jeszcze na koniec, że twierdzenie 4 dotyczy systemów m-wejściowych, czyli systemów mających m wejść i jedno wyjście. Nie jest to jednak żadnym ograniczeniem, gdyż dowolny system φ, który ma m wejść i n wyjść, można traktować jak n systemów m-wejściowych $(\varphi_i)_1^n$ o odpowiedziach impulsowych $(wg_i)_1^n$. W szczególności z części (6) tego twierdzenia wynika wówczas, że dla takiego systemu φ istnieje macierz

$$\mathbf{C} = (c_{ij})^n_{i=1,j\in N} := (\langle wg_i, v\rangle^T)_1^n \Leftarrow (wg_i, v_j\rangle)^n_{i=1,j\in N},$$

dla której $\forall u \in L^m_{F2\mu}$ mamy

$$\varphi(u) \Leftarrow (\varphi_i(u))_1^n = (Cv) * u \Leftarrow (([C]_i v) * u)_1^n \text{(18)}.$$

7. IDENTYFIKACJA PARAMETRYCZNA

Zajmiemy się teraz identyfikacją parametryczną liniowych systemów stacjonarnych i przyczynowych ciągłych na $L^m_{F2\mu}$. Z twierdzenia 4.(6) wynika, że każdy taki system Φ można przedstawić jako kombinację liniową liniowych systemów stacjonarnych i przyczynowych ciągłych na $L^m_{F2\mu}$, których odpowiedzi impulsowe $(v_i)_N =: v$ tworzą bazę ortonormalną w L^m_{F2+}. Dla systemu φ, który ma m wejść i n wyjść mamy wówczas, że $\forall u \in L^m_{F2\mu}$

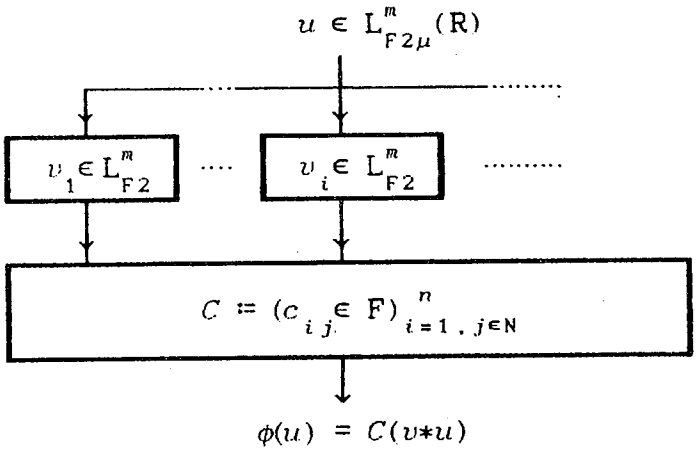
$$\varphi(u) = (Cv) * = C(v * u) := C(v_i * u)_N,$$

gdzie

$$C := (c_{ij})^n_{i=1,j\in N},$$

a zatem system Φ można schematycznie przedstawić tak, jak to zostało pokazane na rys. 1. Identyfikacja prametryczna sprowadza się więc do postępowania mającego na celu wyznaczanie współczynników kombinacji liniowej, które składają się na macierz

(18) Dla dowolnej macierzy A przez $[A]_i$ oznaczamy i-ty wiersz tej macierzy.



Rys.1. Struktura liniowych systemów stacjonarnych i przyczynowych ciągłych na $L^m_{F2\mu}(R)$

C. Z twierdzenia 4.(6) wynika też, że postępowanie to jest równoważne wyznaczeniu odpowiednich iloczynów skalarnych w $L^m_{F_2}$, gdyż

$$C \Leftarrow (c_{ij})^n_{i=1,\, j\in N} = (\langle wg_i, v_j\rangle)^n_{i=1,\, j\in N}.$$

Jeżeli system Φ jest dany w postaci analitycznej, to zadanie identyfikacji parametrycznej jest dość prostym problemem obliczeniowym. Zajmiemy się zatem dalej systemami danymi w postaci nieanalitycznej, a ściślej, jedną z najważniejszych reprezentacji tej postaci, dla której mamy dostępne wejścia i wyjścia systemu Φ. Będziemy przy tym wykorzystywać pojęcie wartości średniej sygnału.

Wartość średnia sygnału mierzalnego $u: R \to F^m$, oznaczana dalej przez

$$A(u) := \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u(t)\, dt.$$

Kolejne twierdzenie pozwala rozwiązać zadanie identyfikacji parametrycznej.

Twierdzenie 5. *Jeżeli* $\langle .\, ,\, .\rangle$ *jest iloczynem skalarnym w* $L^m_{F_2}(R)$,

$$x := (x_i : R \to F)^m_1$$

jest sygnałem złożonym z nieskorelowanych ergodycznych szumów białych oraz S_x *jest gęstością widmową mocy sygnału x,* to $\forall u,\, v \in L^m_{F_2}(R)$

$$A((u * x)(\overline{v * x})) = \langle S_x\, u,\, v\rangle.$$

D o w ó d. Przypomnijmy, że proces stochastyczny biały ma zgodnie z definicją, stałą gęstość widmową mocy, a zatem ergodyczny szum biały, który jest jego

ergodyczną realizacją ma również stałą gęstość widmową mocy. Wynika stąd, że gęstość widmowa mocy sygnału x złożonego zgodnie z założeniem, z nieskorelowanych ergodycznych szumów białych, ma postać:

$$S_x = \operatorname{diag}(S_{x_i})_1^{m(19)},$$

gdzie $\forall i \in \{1, \ldots, m\}$ elementy diagonalne

$$S_{x_i} = \text{const}$$

są gęstościami widmowymi mocy szumów x_i składających się na sygnał x. Wynika stąd, że $\forall t \in \mathrm{R}$ autokorelacja sygnału x ma postać:

$$R_x(t) :=: A(\nabla_t(x)x^*) :=: \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t+\tau)x^*(\tau)\,\mathrm{d}\tau =^{(20)}$$

$$= \mathscr{F}^{-1}(S_x) = u_0(t)S_x^{(21)}.$$

---

(19) Przez $\operatorname{diag}(a_i)_1^n$ oznaczamy *macierz diagonalną* o *elementach diagonalnych* $(a_i)_1^n$, która ma posatć:

$$A = \left(\alpha_{ij} :=: \begin{cases} a_i, & i = j \\ 0, & i \neq j \end{cases}\right)_1^{n,n}.$$

(20) Jeżeli $a \Leftarrow (a_i \in \mathrm{F})_1^m$ oraz $b \Leftarrow (b_i \in \mathrm{F})_1^n$, to dla uproszczenia zapisu

$$ab^T :=: (a_i\, b_j)_{i,j=1}^{m,n}.$$

(21) Przez $\mathscr{F}$ oznaczamy przekształcenie Fouriera.

Załóżmy, że λ jest miarą Lebesgue'a na R, $\forall \varepsilon \in \mathrm{R}$ $(\varepsilon > 0)$ i $\forall t \in R$

$$u_\varepsilon(t) :=: \begin{cases} \dfrac{1}{2\varepsilon}, & t \in [-\varepsilon, \varepsilon] \\ 0, & t \notin [-\varepsilon, \varepsilon] \end{cases}$$

oraz δ jest funkcjonałem, który dla każdej lokalnie całkowalnej funkcji u na R, określony jest przez wzór:

$$\delta(u) :=: \int_{\mathrm{R}} u_0(t)u(t)\,\mathrm{d}t \Leftarrow \lim_{\varepsilon\to 0} \int_{\mathrm{R}} u_\varepsilon(t)u(t)\,\mathrm{d}t.$$

Jeżeli funkcja u jest ciągła, to można wykazać (P. Billingsley [3], s. 221), że

$$\delta(u) = \lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} u(t)\,\mathrm{d}t = u(0).$$

Tak zdefiniowany funkcjonał δ na przestrzeni Schwartza $\mathscr{D}$ (J. Musielak [16], s. 289) jest *dystrybucją Diraca* (J. Musielak [16], s. 300), a symbol u_0 nazywamy dalej *impulsem Diraca*. Warto podkreślić, że impuls Diraca nie jest funkcją lokalnie całkowalną (J. Musielak [16], s. 300), chociaż w literaturze fizycznej i technicznej nazywany jest funkcją Diraca.

Wartość średnia

$$A((u*v)(\overline{v*x})) = \mathbf{A}((u*x)(\overline{v}*\overline{x})) \Leftarrow$$

$$\Leftarrow \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [u*x](t)[\overline{v}*\overline{x}](t)\,dt =$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [x*u](t)[\overline{x}*\overline{v}](t)\,dt =$$

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left(\int_{\mathrm{R}} x^T(t-\tau)\,u(\tau)\,d\tau \int_{\mathrm{R}} x^*(t-\tau)\,\overline{v}(\tau)\,d\tau \right) d\tau =$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left(\int_{\mathrm{R}}\int_{\mathrm{R}} (x^T(t-\tau_1)u(\tau_1))(x^*(t-\tau_2)\overline{v}(\tau_2))\,d\tau_1\,d\tau_2 \right) dt =$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left(\int_{\mathrm{R}}\int_{\mathrm{R}} (u^T(\tau_1)x(t-\tau_1))(x^*(t-\tau_2)\overline{v}(\tau_2))\,d\tau_1\,d\tau_2 \right) dt =$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \left(\int_{\mathrm{R}}\int_{\mathrm{R}} u^T(\tau_1)(x(t-\tau_1)x^*(t-\tau_2))\overline{v}(\tau_2)\,d\tau_1\,d\tau_2 \right) dt =$$

$$= \int_{\mathrm{R}}\int_{\mathrm{R}} u^T(\tau_1) \left(\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t-\tau_1)x^*(t-\tau_2)\,dt \right) \overline{v}(\tau_2)\,d\tau_1\,d\tau_2 =$$

$$= \int_{\mathrm{R}}\int_{\mathrm{R}} u^T(\tau_1)(u_0(\tau_2-\tau_1)S_x)\overline{v}(\tau_2)\,d\tau_1\,d\tau_2 =$$

$$\int_{\mathrm{R}}\int_{\mathrm{R}} v^*(\tau_2)(u_0(\tau_2-\tau_1)S_x)u(\tau_1)\,d\tau_1\,d\tau_2 =$$

$$\int_{\mathrm{R}} v^*(\tau_2) \left(\int_{\mathrm{R}} u_0(\tau_2-\tau_1)\,S_x u(\tau_1)\,d\tau_1 \right) d\tau_2 =$$

$$\int_{\mathrm{R}} v^*(\tau_2) S_x\, u(\tau_2)\,d\tau_2 = \langle S_x\, u,\, v\rangle,$$

a stąd otrzymujemy tezę. ∎

Z udowodnionego twierdzenia można wywnioskować, że jeżeli elementy diagonalne $(S_{x_i})_1^m$ macierzy S_x mają jednakową wartość $a > 0$, czyli $\forall i \in \{1, \ldots, m\}$ mamy

$$S_{x_i} = a,$$

to macierz współczynników kombinacji liniowej

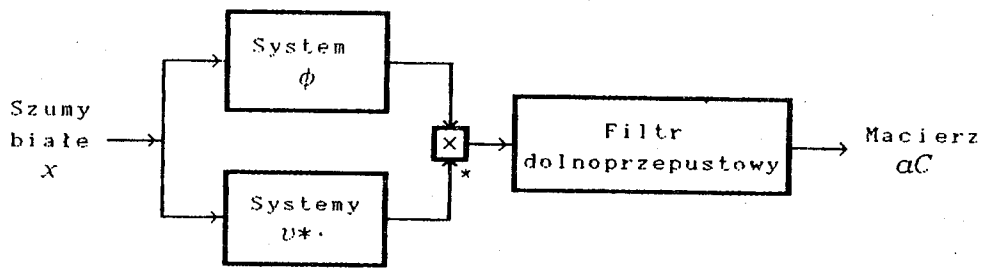
$$C \Leftarrow (c_{ij})_{i=1, j \in N}^n = (\langle wg_i, v_j \rangle)_{i=1, j \in N}^n =$$

$$= \frac{1}{a} (A((wg_i * x)(\overline{v_j * x})))_{i=1, j \in N}^n := \frac{1}{a} A((wg * x)(v * x)^*),$$

gdzie

$$g := (g_i)_{i=1}^n.$$

Wynika stąd, że macierz ta może być wyznaczona w bardzo prostym układzie, którego schemat blokowy przedstawiono na rys. 2. Schemat ten można również odczytać jako schemat blokowy algorytmu wyznaczenia macierzy C.



Rys.2. Schemat blokowy do eksperymentalnego wyznaczania macierzy C

PODSUMOWANIE

W artykule omówione zostały zagadnienia związane z modelowaniem wielowymiarowych liniowych systemów stacjonarnych ciągłych na przestrzeni Hilberta $L_{F2\mu}^m$. Założono przy tym, że wartości sygnałów wejściowych i wyjściowych rozważanych systemów przyjmują wartości z ciała liczbowego F, które może być ciałem liczb rzeczywistych lub zespolonych. Uogólniono pojęcie splotu funkcji, definiując go dla funkcji o wartościach wektorowych, oraz sformułowano twierdzenia 1 i 2, dotyczące jego podstawowych właściwości. Takie rozszerzenie definicji splotu było wynikiem wykazanego dalej, jego związku z iloczynem skalarnym. By związek ten pokazać wprowadzono pojęcia przestrzeni stacjonarnej $L_{Fp\mu\nabla}^m$, przestrzeni splotowej $L_{Fp\mu*}^m$ i przestrzeni przyczynowej $L_{Fp\mu+}^m$, oraz udowodniono twierdzenie 3. Rozważania te doprowadziły do twierdzenia 4, z którego wynika między innymi, że każdy liniowy system stacjonarny ciągły na $L_{F2\mu}^m$ można opisać za pomocą splotu

oraz można go przedstawić jako kombinację liniową liniowych systemów stacjonarnych ciągłych na $L^m_{F2\mu}$, których odpowiedzi impulsowe tworzą bazę ortonormalną w przestrzeni przyczynowej L^m_{F2+}. Z kolei w twierdzeniu 5 pokazano w jaki sposób wyznaczyć współczynniki wspomnianej kombinacji liniowej.

BIBLIOGRAFIA

1. W.M. Amierbajew, N.A. Utiembajew: *Czislennyj analiz łagierowskiego spiektra.* Nauka, Ałma-Ata, 1982
2. J.S. Bendat, A.G. Piersol: *Metody analizy i pomiaru sygnałów losowych.* Biblioteka Naukowa Inżyniera, PWN, Warszawa, 1976
3. P. Bilinglsey: *Prawdopodobieństwo i miara.* PWN, Warszawa, 1987
4. G. Ciesielski, S. Derlecki, Z. Marks-Wojciechowska: *Identyfikacja wielowymiarowych systemów pomiarowych opisanych operatorem nieliniowym.* Praca badawcza nr 112/5/90 BP, Politechnika Łódzka, Łódź, 1990
5. G. Ciesielski, S. Derlecki, Z. Marks-Wojciechowska: *Identyfikacja i korekcja nieliniowa wielowymiarowych systemów pomiarowych opisanych operatorem nieliniowym.* Praca badawcza nr MEN 112/6/90 BP, Politechnika Łódzka, Łódź, 1990
6. G. Ciesielski: Aproksymacja średniokwadratowa operatorów nieliniowych opisujących wielowymiarowe systemy pomiarowe. XXIII Międzyuczelniana Konferencja Metrologów, Politechnika Warszawska, Warszawa, 1991 (referat nie opublikowany)
7. G. Ciesielski: Modele Wienera wielowymiarowych systemów pomiarowych opisanych za pomocą operatorów nieliniowych. Modelowanie i symulacja systemów pomiarowych, *Materiały Sympozjum*, AGH, Kraków, 1992
8. G. Ciesielski: Operatorowe modele wielowymiarowych systemów pomiarowych. Seminarium Naukowe Komisji Kształcenia Komitetu Metrologii i Aparatury Polskiej Akademii Nauk, 1994 (referat nie opublikowany)
9. G. Ciesielski: *Klasyfikacja ogólnej struktury wielowymiarowych systemów pomiarowych.* Modelowanie i symulacja systemów pomiarowych, AGH, Kraków, 1994 (referat zgłoszony)
10. G. Ciesielski: *Identyfikacja ogólnej struktury wielowymiarowych systemów stacjonarnych.* Kwart. Elektroniki i Telekom. 1994, vol. 40, z. 3, ss. 103–112
11. P. Eykhoff: *Identyfikacja w układach dynamicznych*, Biblioteka Naukowa Inżyniera, PWN, Warszawa, 1980
12. J. Jaworski, R. Morawski, J. Olędzki: *Wstęp do metrologii i techniki eksperymentu.* Warszawa, 1992
13. T. Kaczorek: *Teoria sterowania.* t. 1 i 2, PWN, Warszawa, 1977
14. J. Kudrewicz: *Częstotliwość metody w teorii nieliniowych układów dynamicznych.* WNT, Warszawa, 1970
15. R. Kulikowski: *Sterowanie w wielkich systemach.* WNT, Warszawa, 1974
16. J. Musielak: *Wstęp do analizy funkcjonalnej.* PWN, Warszawa, 1989
17. J. Osiowski: *Zarys rachunku operatorowego. Teoria i zastosowania w elektrotechnice.* WNT, Warszawa, 1981
18. A. Papoulis: *Prawdopodobieństwo, zmienne losowe i procesy stochastyczne.* WNT, Warszawa, 1972
19. J. Piotrowski: *Teoria pomiarów. Pomiary w fizyce i technice.* PWN, Warszawa, 1986
20. A. Plucińska, E. Pluciński: *Elementy probabilistyki.* Matematyka dla politechnik, PWN, Warszawa, 1979
21. K.A. Pupkow, W.I. Kapalin, A.S. Juszczenko: *Funkcjonalnyje rjady w tjeorii nieliniejnych sistjem.* Nauka, Moskwa, 1976

22. W. R u d i n: *Analiza rzeczywista i zespolona.* PWN, Warszawa, 1986
23. M. S c h e t z e n: *The Volterra nad Wiener Theories of Nonlinear Systems.* John Wiley & Sons, New York, 1980
24. J. S z a b a t i n: *Podstawy teorii sygnałów.* WNT, Warszawa, 1982
25. M. S z y p e r: *Modelowanie systemów pomiarowych i ich elementów.* Modelowanie i symulacja systemów pomiarowych, Materiały Sympozjum, AGH, Kraków, 1992
26. A. W o j n a r: *Teoria sygnałów.* Podręczniki akademickie EIT, WNT, Warszawa, 1980
27. J. Z a b c z y k: *Zarys matematycznej teorii sterowania.* Biblioteka Matematyczna, t. 73, PWN, Warszawa, 1991
28. R. Z i e l o n k o, A. K r ó l i k o w s k i: *Metody pomiarowo-diagnostyczne analogowych układów elektronicznych.* WNT, Warszawa, 1988
29. A. Ż u c h o w s k i: *Symulacyjna metoda wyznaczania momentów charakterystyki impulsowej przetworników pomiarowych.* Modelowanie i symulacja systemów pomiarowych, Materiały Sympozjum, AGH, Kraków, 1992

G. CIESIELSKI

MODELLING OF MULTIDIMENSIONAL LINEAR AND TIME-INVARIANT SYSTEMS VIA CONVOLUTION

S u m m a r y

In this paper, some problems of multidimensional linear and time-invariant systems modelling via convolutions are presented. It was assumed that the inputs and responses of these systems can be real or complex valued, and the set od inputs have the Hilbert space structure. Because the considered systems are multidimesional, the definition of functions convolution is generalized by defining this convolution for the vector valued functions. There are proven two theorems on the main features of such defined convolutions, and theorem that involves relation between the convolution and the inner product. To show this relation, the stationary, convolution and causal spaces are defined. The next two theorems that are then proven sove the tasks of structure identification and parametric identification of considered systems.

SPIS TREŚCI

CONTENTS

Cena zł 4,—
40.000,—

Kwartalnik Elektroniki i Telekomunikacji

WARUNKI PRENUMERATY

Wpłaty na prenumeratę przyjmowane są na okresy roczne

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

Termin przyjmowania prenumeraty rocznej krajowej i na zagranicę przez RUCH S.A. i Zakład Kolportażu Wydawnictwa SIGMA—NOT
do 25.XI. roku poprzedzającego

Termin przyjmowania prenumeraty rocznej krajowej przez Pocztę Polską
do 25.XI. roku poprzedzającego

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

Adresy przedsiębiorstw kolportażowych i ich konta bankowe:

Zakład Kolportażu Wydawnictwa SIGMA—NOT, 00-716 Warszawa, ul. Bartycka 20, skr. poczt. 1004
Konto PBK III Oddział Warszawa nr 370015-1573-139-11
Prenumerata ze zleceniem wysyłki za granicę jest dwukrotnie wyższa od ceny krajowej. Zlecający powinien podać dokładny adres odbiorcy. Dodatkowe informacje można otrzymać pod nr telefonu 40-30-86 lub 40-00-21 wew. 249, 295, 299.

Ruch S.A. Oddział Warszawa, 00-958 Warszawa, ul. Towarowa 28
Konto PBK XIII Oddział Warszawa nr 370044-1195-139-11
Dostawa odbywa się przesyłką zwykłą w ramach opłaconej prenumeraty, z wyjątkiem zlecenia dostawy pocztą lotniczą, której koszt w pełni pokrywa zleceniodawca.
Prenumerata ze zleceniem dostawy za granicę jest o 100% wyższa od krajowej. Od osób lub instytucji, zamieszkałych lub mieszczących się w miejscowościach, w których nie ma jednostek kolportażowych RUCH prenumeratę Kwartalnika można zamówić w Oddziale Warszawskim RUCH

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

W roku 1995	
przewidywana cena 1 egz. wyniesie	**50.000 zł**
prenumerata roczna w kraju	**200.000 zł**
prenumerata roczna za granicę	**90 $ USA**

POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

Indeks 363189
PL ISSN 0867-6747

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

TOM XL — ZESZYT 4

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1995

Subscription for external subscribers:

The promotional subscription price in 1995 is $ 90 including postage for institutions. Subscriptions should be sent to the publisher, Polish Scientific Publishers PWN Ltd, Journal Division, Miodowa 10, 00-250 Warsaw, POLAND, fax (48) (22) 26 09 50, (48) (22) 26 71 63, with a copy by fast to Electronics and Telecommunications Quarterly 00-665 Warsaw, ul. Nowowiejska 15/19, p. 470. Subscription is occupated after showing a checque or transfer documents. Our bank account is as follows: Bank account: PBK VIII O/Warszawa nr 370028-1052. At subscripber's request this journal will be air mailed at additional postage to 50% dross price to European countries and 65% overseas.

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

POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

TOM XL — ZESZYT 4

WYDAWNICTWO NAUKOWE PWN
WARSZAWA 1995

Wydanie kwartalnika sponsoruje „Spółka Telekomunikacja Polska S.A."

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

Dyżury Redakcji: środy i piątki, godz. 14—16
tel. 628 89 81 21007-737

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

WYDAWNICTWO NAUKOWE PWN
Warszawa, ul. Miodowa 10

Ark. wyd. 9,00 Ark. druk. 7,75	Podpisano do druku w styczniu 1995 r.
Papier offsetowy kl. III 70 g. B-1	Druk ukończono w lutym 1995 r.

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

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1994. 40, z. 4. ss. 453—487

621.39
519.68

Porównanie właściwości procesorów sygnałowych

PIOTR REMLEIN

Instytut Elektroniki i Telekomunikacji, Politechnika Poznańska

Otrzymano 1994.10.19

Autoryzowano do druku 1994.12.16

Artykuł prezentuje właściwości procesorów sygnałowych stało- i zmiennoprzecinkowych dwóch znanych firm Texas Instruments i Analog Devices. Często procesory oceniane są na podstawie liczby operacji wykonywanych w ciągu jednej sekundy tzw. MIPS. Takie podejście nie oddaje w pełni rzeczywistych możliwości układów.

W tej pracy porównano procesory sygnałowe pod względem ich cech architektonicznych, zdolności adresowych i sterowania programem.

Zaprezentowano przykładowe realizacje DSP wykorzystujące omawiane procesory sygnałowe. Oceniono procesory na podstawie złożoności i czasu wykonywania poszczególnych algorytmów.

1. WSTĘP

W ostatnich kilku latach zauważalny jest znaczny wzrost zainteresowania systemami cyfrowego przetwarzania sygnałów. Dzieje się tak na skutek coraz większej dostępności do specjalizowanych układów — cyfrowych procesorów sygnałowych (DSP). Umożliwiają one spełnienie zarówno wymagań dotyczących szybkości przetwarzania jak i elastyczności modelowanych systemów. Dzięki swym możliwościom procesory sygnałowe znajdują dzisiaj zastosowanie w wielu różnorodnych dziedzinach np.: w biomedycynie, akustyce, radiolokacji, sejsmologii, telekomunikacji, w kodowaniu mowy, transmisji danych, przetwarzaniu obrazów. Używane w tych zatosowaniach procesory zawierają w jednym układzie scalonym procesor arytmetyczny, pamięci ROM i RAM, generatory adresów, często również układy konwersji c/a i a/c oraz układy bezpośrdniego adresowania pamięci.

Obecnie na rynku istnieje wiele firm produkujących procesory sygnałowe umożliwiające wykonywanie w czasie rzeczywistym skomplikowanych algorytmów matematycznych. Niniejszy artykuł prezentuje i porównuje cechy dwóch

popularnych rodzin procesorów sygnałowych TMS firmy Texas Instruments [2], [4] i ADSP firmy Analog Devices [1], [3]. Przedstawiono zarówno procesory zmienno- jak i stałoprzecinkowe produkowane przez obie wymienione firmy. Ukazano różnice w budowie procesorów i sposobie projektowania systemów cyfrowych przy ich pomocy. Porównania procesorów dokonano na podstawie rozważań następujących zagadnień:

- zakresu i sposobu reprezentacji liczb;
- szybkości i elastyczności układów arytmetycznych;
- możliwości pobierania dwóch operandów w czasie pojedynczego cyku maszynowego;
- realizacji cyklicznego pobierania danych;
- sposobu realizacji instrukcji pętli i skoku;
- cech posiadanych pamięci.

Następne rozdziały omawiają dokładnie wymienione problemy. W rozdziale drugim przedstawione zostały sposoby reprezentacji liczb w analizowanych procesorach, ich architektura i możliwości układów arytmetycznych. Rozdział trzeci prezentuje sposoby adresowania danych w procesorach sygnałowych. Część czwarta omawia układy sterowania programem. W piątym rozdziale przedstawione zostały różnice w możliwościach adresowania pamięci wewnętrznych i zewnętrznych występujące między poszczególnymi procesorami. Część szósta zawiera opis przykładowych zastosowań procesorów sygnałowych oraz porównanie ich działania na podstawie realizacji wybranych algorytmów cyfrowego przetwarzania sygnałów stosowanych w telekomunikacji. Ostatni, siódmy rozdział przedstawia wnioski wynikające z przeprowadzonych analiz i porównań oraz praktycznych zastosowań.

2. WŁAŚCIWOŚCI ARYTMETYCZNE

Właściwości arytmetyczne są jednym z ważniejszych czynników wpływających na sposób działania procesorów sygnałowych. Cyfrowe procesory sygnałowe powinny umożliwiać wykonywanie operacji mnożenia i dodawania w czasie jednego cyklu maszynowego. Powinny szybko wykonywać standardowe operacje arytmetyczne i logiczne.

Wszystkie omawiane procesory posiadają zmodyfikowaną architekturę harwardzką. Program i dane przechowywane są w dwóch oddzielnych blokach pamięci, do których istnieje dostęp poprzez odrębne magistrale. Dzięki temu istnieje możliwość jednoczesnego korzystania z argumentów znajdujących się w pamięci programu i danych. Procesory zmiennoprzecinkowe TMS 320C30 i ADSP 21020 nie mają stałego podziału pamięci. Najczęściej stosowana jest jednak architektura harwardzka. Dowolne skonfigurowanie pamięci tych procesorów pozwala na budowę bardziej optymalnych kodów programu w zależności od realizowanych algorytmów. Zwiększa to znacznie moc obliczeniową procesora.

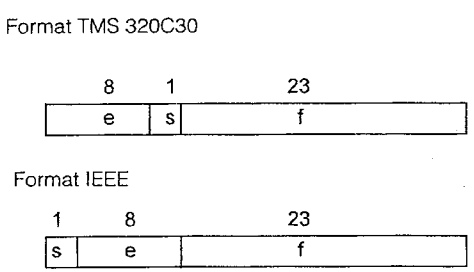
2.1. STOSOWANE FORMATY DANYCH

W procesorach sygnałowych stosuje się różne sposoby przedstawiania liczb i różne ich zakresy. W procesorach stałopozycyjnych liczby ujemne reprezentowane są na ogół w postaci uzupełnienia do dwóch. W procesorze stałoprzecinkowym ADSP 2101 możliwe są dwa tryby reprezentacji liczb w uzupełnieniu do dwóch. W jednym iczby traktowane są jako ułamek reprezentowany przez jeden bit znaku i 15 bitów znaczących. W drugim jako liczby całkowite przedstawiane przy pomocy 16 bitów.

W procesorze TMS 320C50 liczby reprezentowane są w postaci 16 bitowego uzupełnienia do dwóch. W obu układach nie są wykorzystywane inne formaty reprezentacji liczb np. znak moduł, uzupełnienie do 1, czy kod BCD.

Procesory zmiennopozycyjne ADSP 21020 i TMS 320C30 umożliwiają wykonywanie operacji arytmetycznych zarówno w formacie zmiennopozycyjnym jak i w stałopozycyjnym. W procesorze ADSP 21020 dane reprezentowane są przez liczby 32 lub 40 bitowe, zmiennoprzecinkowe. Liczba 32 bitowa odpowiada standardowi IEEE 754/854. W formacie tym 24 bity oznaczają mantysę, a 8 bitów reprezentuje cechę. W 40-sto bitowym formacie 32 bity przeznaczone są dla mantysy i 8 dla cechy [3].

W procesorze TMS 320C30 32 bitowa liczba zmiennopozycyjna jest określona przez 24 bitową mantysę i 8 bitową cechę. Jednak ten format nie odpowiada standardowi IEEE. Rysunek 1 przedstawia format zmiennopozycyjny według



Rys. 1. Sposób reprezentacji liczb zmiennoprzecinkowych według standardu IEEE i TMS 320C30.

standardu IEEE oraz sposób reprezentacji liczby zmiennoprzecinkowych stosowany w procesorze TMS. W formacie IEEE pierwszy bit jest bitem znaku, następnych 8 bitów oznacza cechę, określoną wyrażeniem e-127, a kolejne 23 bity reprezentują bezwzględną wartość mantysy. W formacie zmiennoprzecinkowym stosowanym w procesorze TMS 320C30 pierwszych osiem bitów oznacza cechę w uzupełnieniu do dwóch. Następne 24 bity określają mantysę. Szczególny przypadek w tej reprezentacji liczby występuje dla $e = -128$ wówczas przedstawiana liczba wynosi 0 niezależnie od wartości bitu znaku s i bitów mantysy f. Różnice przedstawiania liczb zmiennoprzecinkowych w obu reprezentacjach zawiera tablica 1.

Tablica 1

Różnice reprezentacji liczb w formacie TMS i IEEE

format zmiennoprzecinkowy w procesorze TMS 320C30		format zmiennoprzecinkowy według standardu IEEE	
$2^e*(01.f)$	jeśli s=0	$(-1)^s*2^{e-127}*(01.f)$	dla $0<e<255$
$2^e*(10.f)$	jeśli s=1	$(-1)^s*0.0$	dla e=0 i f=0
0	jeśli e=−128	$(-1)^s*2^{-126}*(0.f)$	dla e=0 i f<>0
		$(-1)^s$*nieskończoność	dla e=255 i f=0
		stan nieokreślony	dla e=255 i f<>0

Uwzględniając różnice w definicji reprezentacji zmiennoprzecinkowej dla procesora TMS i standardu IEEE można przeprowadzić programową konwersję z jednej postaci na drugą. Rysunek 2 przedstawia wydruk programu napisanego w asemblerze procesora TMS 320C30, który służy do zmiany postaci liczby reprezentowanej w formacie IEEE na postać używaną przez procesor TMS. Konwersja ta trwa 12 cykli maszynowych i może być przeprowadzona dla przypadku, gdy wartość cechy mieści się w przedziale od 0 do

```
*
* R0    liczba przekształcana
* AR1 wskaźnik do tablicy stałych
*
* (0) 0xFF800000 <-AR1
* (1) 0xFF000000
* (2) 0x7F000000
* (3) 0x80000000
* (4) 0x81000000
*
* Rejestry z danymi wejściowymi R0, AR1
* Rejestry modyfikowne R0, R1
* Rejestr zawierający wynik R0

        .global FMIEEE

FMIEEE AND3  R0,*AR1,R1          ; zastąpienie ułamka zerem
             BND    NEG          ; sprawdzenie bitu znaku s
             ADDI   R0,R1        ; przesunięcie bitów znaku i cechy
             LDIZ   *+AR1(1),R1  ; jeśli same zera to generowane jest 0 w formacie 'C30
             SUBI   *+AR1(2),R1  ; wyznaczenie cechy
             PUSH   R1
             POPF   R0;          ; wyznaczenie liczby zmiennoprzecinkowej
             RETS
*
NEG          PUSH   R1
             POPF   R0           ; wyznaczenie liczby zmiennoprzecinkowej
             NEGF   R0,R0        ; negacja jeśli bit znaku wynosi 1
             RETS
```

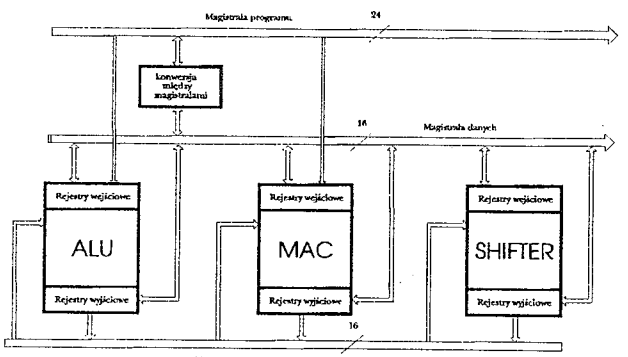
Rys. 2. Wydruk programu dokonującego zmiany formatu IEEE na reprezentację stosowaną w procesorach TMS

255. W pozostałych przypadkach przedstawionych w tablicy 1 konwersja trwa dłużej około 23 cykli. Liczba przekształcana znajduje się w niższych 32 bitach rejestru R0, a wynik konwersji umieszczany jest w wyższych 32 bitach tego rejestru.

Należy zaznaczyć, że w procesorze TMS 320C30 operacje zmiennoprzecinkowe dla liczb 40 bitowych mogą być wykonywane tylko przez układ ALU, układ mnożący korzysta jedynie z danych 32 bitowych. Oba procesory ADSP 21020 i TMS 320C30 mogą również wykonywać działania na liczbach stałopozycyjnych. 21020 korzysta z danych 32 bitowych zarówno dla układu mnożącego jak i ALU. Procesor 'C30 korzysta w układzie mnożącym z liczb 24 bitowych stałopozycyjnych i 32 bitowych w układzie ALU. Stosowanie procesorów zmiennoprzecinkowych umożliwia używanie liczb z szerszego zakresu oraz uwalnia projektanta od przeprowadzania operacji skalowania, które należy wykonywać przy stosowaniu procesorów stałopozycyjnych.

2.2. ARCHITEKTURA UKŁADÓW ARYTMETYCZNYCH PROCESORA ADSP 2101

Procesor ADSP 2101 posiada trzy niezależne układy liczące: jednostkę arytmetyczno-logiczną ALU, układ mnożący i akumulator (MAC) i układ przesuwny (SHIFTER), rys. 3 [1]. Wszystkie są połączone wspólną magistralą wyników. Dzięki temu rejestry wyjściowe poszczególnych układów mogą zawierać argumenty wykorzystywane w następnym cyklu przez inne układy. Dodatkowo układy ALU i MAC są bezpośrednio połączone do magistrali danych i programu. Operandy dla ALU i MAC mogą pochodzić zarówno z pamięci danych jak i programu oraz z dowolnych rejestrów danych procesora. Oczywiście wszystkie rejestry układów arytmetycznych mogą być rejestrami ogólnego przeznaczenia.



Rys. 3. Schemat blokowy układów arytmetycznych procesora ADSP 2101

Układ ALU procesora ADSP 2101 posiada dwa podwójne 16 bitowe rejestry wejściowe X i Y, tzn. AX0, AX1 oraz AY0, AY1 [1]. Wszystkie operacje przeprowadzane przez ten układ odbywają się na operandach znajdujących się

w tych rejestrach. Mogą one być ładowane z pamięci programu, pamięci danych lub z dowolnego rejestru procesora. Wynik operacji składowany jest w rejestrze AR lub rejestrze powrotnym AF. Rejestry te oraz rejestry przechowujące wyniki w układzie mnożącym MAC i układzie przesuwnym mogą być użyte jako operandy wejściowe X i Y do operacji wykonywanych w układzie ALU. Poniżej przedstawiono przykładową instrukcję wykonywaną przez ALU.

$$AR=AX0+AY1,\ AX0=DM(I0,M0),\ AY1=PM(I4,M4)$$

Jest to instrukcja wielofunkcyjna. Pierwsza jej część do przecinka oznacza operację dodawania zawartości rejestrów AX0 i AY1, druga część określa załadowanie rejestru AX0 przez operand znajdujący się w pamięci danych, trzecia instrukcja oznacza przesłanie do rejestru AY1 argumentu z pamięci programu. Ta złożona instrukcja wykonywana jest przez procesor w jednym cyklu maszynowym trwającym 60 ns przy częstotliwości zegara 16.67 MHz.

Podobnie jak jednostka arytmetyczno-logiczna tak i układ mnożący wyposażony jest w dwa podwójne rejestry wejściowe MX0, MX1 i MY0, MY1 [1]. Oba układy: MAC i ALU w procesorze ADSP 2101 działają niezależnie. Różni to je od budowy układów arytmetycznych procesora TMS 320C50.

Układ MAC wykonuje działania na operandach zawartych w rejestrach X i Y. Mogą one być ładowane zarówno z pamięci danych jak i programu oraz z dowolnych rejestrów procesora. Wynik operacji przechowywany jest w rejestrze MR lub rejestrze powrotnym MF. Rejestry te podobnie jak w ALU mogą przesyłać argumenty na wejście układu mnożącego. Na wejście to mogą być również podawane operandy bezpośrednio z rejestrów przechowujących wyniki w układach przesuwnym i ALU. Poniżej przedstawione są dwie przykładowe instrukcje wykonywane przez układ mnożący.

$$MR=MX0*MY0\ (SS)$$
$$MR=MR+MX1*MY1\ (SU),\ MX1=DM(I0,M0),\ MY1=PM(I4,M4)$$

Pierwsza z nich oznacza mnożenie dwóch liczb bez znaku zawartych w rejestrach MX0 i MY0. Druga instrukcja oznacza mnożenie i dodawanie wyniku mnożenia liczby ze znakiem i bez znaku. Instrukcja ta składa się z trzech części. Druga składowa opisuje przesłanie operandu do rejestru X z pamięci danych, a trzecia załadowanie rejestru Y przez argument znajdujący się w pamięci programu.

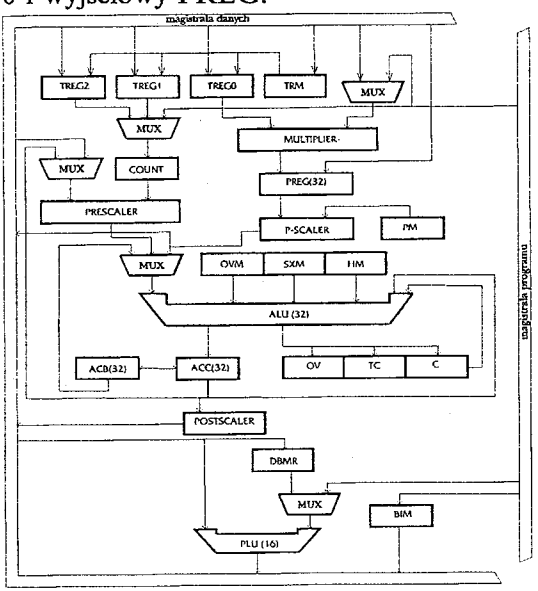
Rejestr MR, który służy do przechowywania wyników jest 40 bitowy i składa się z dwóch rejestrów 16 bitowych MR1 i MR0 oraz z rejestru 8 bitowego MR2. Rejestr MR2 zabezpiecza układ mnożący przed przepełnieniem jakie może wystąpić przy operacjach mnożenia i dodawania. Wszystkie operacje realizowane przez układ MAC również wielofunkcyjne są wykonywane w jednym cyklu maszynowym.

Układ przesuwny w procesorzez ADSP 2101 posiada rejestr wejściowy SI. Może on być ładowany przez operandy zawarte w rejestrach wynikow układów MAC i ALU oraz z rejestru SR zawierającego wynik opracji wykonywanych

w układzie przesuwnym. Rejestr wyników jest złożony z dwóch 16 bitowych rejestrów SR0 i SR1. Omawiany układ posiada także rejestr SE przeznaczony do przechowywania w nim cechy przy operacjach zmiennoprzecinkowych. Jest on automatycznie inicjalizowany przez instrukcje służące do przeprowadzania normalizacji. Rejestr przesuwny może umieścić w czasie jednego cyklu 16 bitową liczbę w dowolne miejsce 32 bitowego pola. Operand wejściowy może być przesunięty o dowolną liczbę bitów w prawo lub w lewo. Budowa układu przesuwającego umożliwia wykonywanie w jednym cyklu maszynowym między innymi operacje normalizacji i denormalizacji.

2.3. ARCHITEKTURA UKŁADÓW ARYTMETYCZNYCH PROCESORA TMS 320C50

Rysunek 4 przedstawia schemat blokowy układów arytmetycznych procesora TMS 320C50. Składa się on z układu mnożącego, ALU, jednostki równoległej logiki (PLU), 16 bitowego rejestru przesuwnego i dodatkowych rejestrów przesuwnych na wyjściu akumulatora i układu mnożącego [2]. Układ mnożący ma bezpośredni dostęp do magistrali danych i programu. Posiada rejestr wejściowy TREG0 i wyjściowy PREG.



Rys. 4. Schemat blokowy układów arytmetycznych procesora TMS 320C50

Jednostka arytmetyczno-logiczna ALU ma bezpośredni dostęp jedynie do magistrali danych. Wyniki są zawsze przesyłane do szyny danych lub do rejestrów akumulatora. W niektórych przypadkach wynik operacji musi być najpierw umieszczony w pamięci, by mógł być używany jako operand wejściowy do następnych obliczeń. Takie operacje jak dodawanie dwóch argumentów z pamięci lub mnożenie i dodawanie mogą wymagać dodatkowych cykli maszynowych.

Podobnie jak w TMS 320C25 tak i w procesorze 'C50 układ mnożący nie ma możliwości jednoczesnego akumulowania wyników mnożenia. Operacja taka często występuje w algorytmach przetwarzania sygnałów. Musi wówczas być używany zarówno ALU jak i układ mnożący, a czasami konieczna jest zmiana realizowanego algorytmu. Problemy te nie występują w układzie ADSP 2101.

Jednostka ALU procesora TMS 320C50 do wykonania operacji potrzebuje jednego operandu pochodzącego z akumulatora a drugiego z wyjścia układu mnożącego, bufora akumulatora (ACCB), z magistrali danych lub układu przesuwnego akumulatora (Prescaler) rys. 4. Aby wykonać dodawanie dwóch liczb przy pomocy ALU należy najpierw załadować do akumulatora pierwszą liczbę, po czym dopiero można dodać do niej drugą liczbę.

Instrukcje asemblera dla operacji wykonywanych przy pomocy układu ALU odznaczają się specyficzną mnemoniką. Poniżej przedstawione są dwie instrukcje potrzebne do wykonania dodawania dwóch liczb:

ZALR	⟨adres pamięci danych⟩;	wpisanie do 16 starszych bitów akumulatora wskazanego operandu i wyzerowanie 15 młodszych bitów, na pozycję 15 jest wpisywana wartość 1
ADD	⟨adres pamięci danych⟩;	dodanie do akumulatora wskazanego operandu

Wynik powyższych operacji jest składowany w akumulatorze i może być użyty jako operand wejściowy dla innych niż ALU układów arytmetycznych po uprzednim zapamiętaniu w pamięci danych. Czas trwania większości operacji wykonywanych przez układ ALU wynosi 1 cykl maszynowy 50 ns, przy częstotliwości zegara 20 MHz. Istnieją jednak instrukcje jak np. przedstawione powyżej, które potrzebują dwóch cykli maszynowych. Wśród instrukcji ALU są również takie, które nie mogą być wykonywane cyklicznie jedna po drugiej. Należą do nich m.in. ADD #k dodająca do zawartości akumulatora liczbę k, SUB #k odejmująca od zawartości akumulatora wartość k oraz ADRK #k dodająca do wybranego rejestru pomocniczego liczbę k.

Budowa układu mnożącego procesora TMS 320C50 różni się od posiadanego przez ADSP 2101. Dla wykonania operacji mnożenia i akumulacji procesor 'C50 musi użyć obu układów ALU i MAC. Przy wykonaniu prostego mnożenia należy najpierw załadować do rejestru TREGO pierwszy operand, który następnie jest mnożony przez liczbę pochodzącą z magistrali danych. Poniżej przedstawiono dwie przykładowe instrukcje wykonujące mnożenie:

LT ⟨adres pamięci danych⟩
MPY ⟨adres pamięci danych⟩

Wynik operacji jest otrzymywany po dwóch cyklach maszynowych.
Do realizacji mnożenia i akumulacji służy instrukcja MAC:
MAC ⟨adres pamięci programu⟩⟨adres pamięci danych⟩

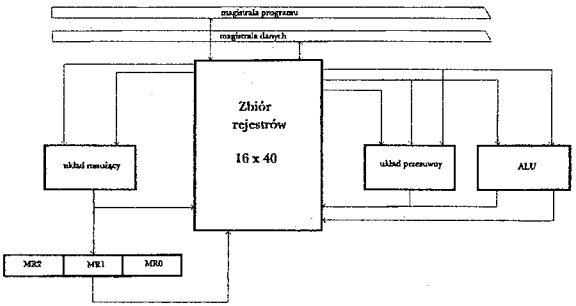
Wymaga ona dwóch słów adresowych pamięci. Czas wykonania tej instrukcji z dwoma operandami z wewnętrznej pamięci procesora wynosi w trybie bez powtórzeń trzy cykle maszynowe (3×50 ns), a w trybie z powtórzeniami $2+n$ cykli maszynowych, gdzie n oznacza liczbę powtórzeń. Istnieją cztery różne zapisy mnemoniczne dla funkcji wykonujących mnożenie i akumulację: MAC, MACD, MADD, MADS. Akumulator procesora TMS 320C50 posiada 31 bitów i jeden dodatkowy bit przeznaczony do oznaczenia przepełnienia. W procesorze ADSP 2101 akumulator ma 40 bitów. W procesorze 'C50 po wystąpieniu więcej niż jednego przepełnienia dalsze obliczenia są błędne i nie jest możliwe automatyczne zaokrąglanie wyników.
W układzie ADSP 2101 jest możliwe wystąpienie 256 przepełnień bez stracenia danych, a automatyczne zaokrąglanie jest wykonywane w tym samym cyklu co operacja mnożenia.

Procesor TMS 320C50 posiada trzy rodzaje układów przesuwnych. Rejestr P-scaler umożliwia przesunięcie liczby o 0, 1 lub 4 bity w lewo i 6 bitów w prawo. Rejestr prescaler może przesuwać dane podawane na wejście jednostki ALU od 0 do 16 bitów w prawo lub w lewo. Rejestr Post-scaler znajduje się na wyjściu ALU. Może przesuwać dane pochodzące z akumulatora (ACC) od 0 do 7 bitów w lewo. Dzięki tym rejestrom możliwe jest skalowanie liczb podczas ich przesyłania między układami. Nie potrzeba wykonywać dodatkowych operacji przesuwania.

2.4. ARCHITEKTURA UKŁADÓW ARYTMETYCZNYCH PROCESORA ADSP 21020

Procesor zmiennoprzecinkowy ADSP 21020 posiada zbiór rejestrów wejściowych i wyjściowych dla układów arytmetycznych, rysunek 5 [3]. Składa się on z szesnastu 40 bitowych rejestrów i 8 portów danych. W jednym cyklu maszynowym możliwe są dwa dostępy do pamięci, dostarczenie dwóch argumentów do ALU, do układu mnożącego i zapamiętanie pochodzących z nich wyników. Rejestry układu arytmetycznego podzielone są na dwie grupy po 8 rejestrów. Każda grupa posiada odpowiadające jej rejestry zapasowe, które są przełączane w razie korzystania z przerwań lub podprogramów. Zatem łączna liczba wszystkich rejestrów wynosi 32. Wszystkie układy arytmetyczne tego procesora realizują operacje w jednym cyklu maszynowym. Są one połączone ze sobą równolegle. Operand wyjściowy z dowolnego układu arytmetycznego może być podawany bezpośrednio na wejście innego w następnym cyklu. Przy obliczeniach wielofunkcyjnych ALU i układ mnożący mogą działać niezależnie. Istnieje specjalny zbiór rejestrów przeznaczony do zapamiętywania wyników i przenoszenia danych między układami arytmetycznymi a magistralami danych.

Układ ALU procesora ADSP 21020 może realizować wszystkie standardowe operacje logiczne i arytmetyczne oraz umożliwia wykonanie w jednym cyklu maszynowym funkcji, które nie są realizowane przez wiele tradycyjnych procesorów sygnałowych. Umożliwia stałoprzecinkowe



Rys. 5. Schemat blokowy układów arytmetycznych procesora ADSP 21020

dodawanie i odejmowanie o zwiększonej precyzji z zachowaniem przeniesienia i pożyczki. Ponadto może prowadzić szybkie obliczenia funkcji odwrotnej do danego operandu (1/x) i funkcji odwrotnej z pierwiastka kwadratowego. Procesor zmiennoprzecinkowy firmy Texas Instruments TMS 320C30 wymaga do tych samych obliczeń więcej cykli maszynowych. Tablica 2 porównuje czasy wykonania przykładowych instrukcji obu procesorów zmiennoprzecinkowych.

Tablica 2

Zestawienie czasów realizaci wybranych operacji arytmetycznych przez procesory ADSP 21020 i TMS 320C30

operacja	ADSP 21020	TMS 320C30
1/X	6 cykli	35 cykli
Y/X	6 cykli	36 cykli
1/SQRT (X)	9 cykli	74 cykle
SQRT (X)	10 cykli	39 cykli

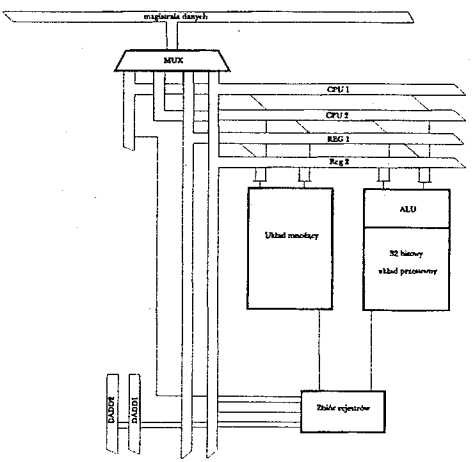
Układ mnożący ADSP 21020 może wykonywać operacje na 32 bitowych liczbach stałopozycyjnych i 32 lub 40 bitowych liczbach zmiennopozycyjnych. Przy mnożeniu 32 bitowym, stałopozycyjnym wyższe 32 bity z 64 bitowego wyniku mogą być przesłane do któregoś z rejestrów lub 64 bitowy wynik może zostać dodany (lub odjęty) do zawartości jednego z dwóch 80 bitowych akumulatorów. Stałoprzecinkowe operandy mogą być indywidualnie wybrane jako liczby bez znaku lub ze znakiem i jako liczby ułamkowe lub całkowite. Natomiast przy mnożeniu zmiennoprzecinkowym 32 lub 40 bitowy wynik musi zostać przesłany do jednego z szesnastu 40 bitowych rejestrów.

W zbiorze instrukcji procesora ADSP 21020 istnieje wiele rozkazów pozwalających na wykonywanie działań na poszczególnych bitach operandów. Możliwe jest przesuwanie danych w prawo lub w lewo zarówno arytmetyczne jak i logiczne oraz przesuwanie kodowe. Przesuwany argument może pochodzić z dowolnego rejestru lub być określony przez słowo instrukcji. Układ przesuwny

procesora ADSP 21020 posiada zdolność do zerowania, ustawiania i sprawdzania poszczególnych bitów w rejestrach. Umożliwia pobranie dowolnego bitu z jednego słowa i wstawienie go do innego. Operacje przesuwania danych realizowane są na 32 bitowych słowach.

2.5. ARCHITEKTURA UKŁADÓW ARYTMETYCZNYCH PROCESORA TMS 320C30

Rysunek 6 przedstawia schemat blokowy układów arytmetycznych procesora zmiennoprzecinkowego TMS 320C30 [4]. Zbiór rejestrów składa się z ośmiu 40 bitowych rejestrów rozszerzonej precyzji. Z rejestrów tych mogą być pobierane do układu mnożącego dwa operandy gdy układ ALU pobiera dwa argumenty bezpośrednio z pamięci. Wyniki operacji dodawania i mnożenia są zawsze przesyłane do rejestrów arytmetycznych. Rejestry te nie posiadają dublujących ich rejestrów zapasowych jak to ma miejsce w procesorze ADSP 21020. Dla zwiększenia szybkości działania procesor ten wyposażony jest w podwójne magistrale adresowe (DADDR), rejestrów (REGØ i jednostki centralnej (CPU), rys. 6.



Rys. 6. Schemat blokowy układów arytmetycznych procesora TMS 320C30

Układ ALU procesora TMS 320C30 umożliwia realizację podstawowych funkcji arytmetycznych i logicznych (m.in. dodawanie, odejmowanie, obliczanie wartości bezwzględnej), przy jego pomocy nie można jednak prosto zrealizować wielu funkcji dostępnych w układzie ALU procesora ADSP 21020. Na przykład do wyznaczenia liczby większej z dwóch danych w ADSP 21020 jest realizowane przy pomocy pojedynczej instrukcji R7=MAX (R5, R6), a w TMS 320C30 potrzeba do tego celu aż trzech rozkazów:

```
LDF    R6,R7
CMPF   R5,R7
LDFLT  R5,R7
```

Układ mnożący procesora TMS 320C30 wykonuje działania na operandach 24 bitowych stałopozycyjnych lub 32 bitowych zmiennopozycyjnych [4]. W przypadku mnożenia liczb stałoprzecinkowych (24 bitowych) wynik jest 48 bitowy. Wszystkie operandy wejściowe są przyjmowane jako liczby całkowite ze znakiem. Nie istnieje ułamkowa reprezentacja liczb stałopozycyjnych. Do wykonywania operacji akumulowania stałoprzecinkowych iloczynów wykorzystywany jest 32 bitowy układ ALU. Wynik mnożenia stałoprzecinkowego jest ograniczony do 32 bitów, bo tylko 32 niższe bity są zapisywane do rejestrów roboczych. Wynik mnożenia zmiennoprzecinkowego jest 40 bitowy. Układ przesuwny procesora 'C30 udostępnia użytkownikowi tylko niewielki podzbiór instrukcji oferowanych przez procesor ADSP 21020. Należą do nich przesunięcia logiczne, arytmetyczne i kołowe.

2.6. PORÓWNANIE CECH UKŁADÓW ARYTMETYCZNYCH PROCESORÓW SYGNAŁOWYCH

Porównując cechy architektoniczne układów arytmetycznych omawianych procesorów sygnałowych można stwierdzić, że procesor stałoprzecinkowy ADSP 2101 pozwala na prostszą realizację wielu algorytmów cyfrowego przetwarzania sygnałów niż jego odpowiednik firmy Texas Instruments TMS 320C50. W procesorze 'C50 występuje niewygodna zależność układów ALU i MAC. Przy realizacji operacji mnożenia i sumowania oba te układy muszą być użyte. W procesorze ADSP 2101 nie występuje konieczność zapamiętywania wyników operacji w pamięci, mogą one być przechowywane w rejestrach lokalnych i w następnym cyklu użyte jako operandy wejściowe do innych układów. Eliminowane są zatem zbędne cykle pobrania operandów z pamięci.

Tablica 3

Zestawienie cech układów arytmetycznych procesorów ADSP 2101 i TMS 320C50

cechy	ADSP 2101	TMS 320C50
wszystkie operacje ALU wykonywane są w jednym cyklu		nie
mnożenie w jednym cyklu		nie
operacje mnożenia i akumulacji w jednym cyklu		
operacja przesuwania w jednym cyklu	w prawo lub w lewo o 0–32 bity	w prawo lub w lewo o 0–16 bitów 0–7 bitów w lewo 1 lub 4 bity w lewo 6 bitów w prawo
zabezpieczenie akumulatora przed przepełnieniem	8 bitowe	1 bitowe
mnożenie liczb ze znakiem, bez znaku, mieszanych		nie posiada trybu mieszanego
normalizacja w jednym cyklu		nie

Zalety tej nie ma TMS 320C50. Zestawienie właściwości układów arytmetycznych obu procesorów zawiera tablica 3.

Natomiast porównując cechy układów arytmetycznych omawianych procesorów zmiennoprzecinkowych należy zauważyć, że procesor ADSP 21020 posiada mniejszą liczbę rejestrów roboczych od procesora TMS 320C30, ale wszystkie są 40 bitowe. W 'C30 jest tylko 8 rejestrów o rozszerzonej precyzji. W ADSP możliwe jest przedstawienie liczb w formacie IEEE, a w TMS należy do tego celu użyć specjalnej procedury. Ponadto procesor 21020 posiada w swym zbiorze instrukcji wiele rozkazów przydatnych do realizacji złożonych algorytmów DSP, których brakuje w liście rozkazów procesora 'C30. Tablica 4 zawiera zestawienie właściwości układów arytmetycznych obu procesorów zmiennoprzecinkowych.

Tablica 4

Zestawienie wybranych cech układów arytmetycznych procesorów ADSP 21020 i TMS 320C30

cechy	ADSP 21020	TMS 320C30
operacje zmiennoprzecinkowe 40 bitowe		
operacje stałoprzecinkowe 32 bitowe		
dodawanie i odejmowanie wielokrotnej precyzji		
instrukcje obliczania funkcji: min, max, skalowania		nie
instrukcja wyznaczania wartości bezwzględnej		nie
instrukcja wyznaczania liczby odwrotnej		nie
operacje przesuwania		
mnożenie i dodawanie w jednym cyklu		
mnożenie, dodawanie i odejmowanie w jednym cyklu		nie

3. SPOSOBY ADRESOWANIA DANYCH

Możliwości adresowania danych w procesorach sygnałowych są bardzo ważnym elementem ich architektury. Bezużyteczne są bowiem duże zdolności obliczeniowe procesora, jeśli pobieranie danych trwa długo. Budowa procesora musi zatem umożliwiać odpowiednio szybki dostęp do operandów, adekwatny do szybkości przetwarzania procesora.

Maksymalną efektywność pracy mikroprocesora można uzyskać, jeśli:

— w jednym cyklu maszynowym pobierane są dwa operandy, potrzebne do określonej operacji,

— jednocześnie dostępne są dwa różne obszary pamięci,

— można używać zarówno pamięci danych jak i programu do zapamiętywania danych,

— istnieje możliwość użycia bufora kołowego, często stosowanego w algorytmach DSP,

— można zastosować wskaźniki adresowe.

Nie wszystkie procesory spełnieją te wymagania. Dla większości podstwowymi sposobami adresowania są:

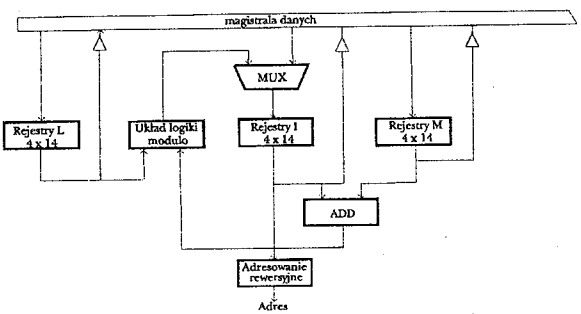
— adresowanie bezpośrednie — adres operandu zawarty jest w kodzie instrukcji,

— adresowanie pośrednie rejestrowe — adres operandu umieszczony jest w jednym z rejestrów roboczych.

3.1. MOŻLIWOŚCI ADRESOWE PROCESORA ADSP 2101

Procesor ADSP 2101 posiada dwa niezależne generatory adresów (rys. 7) [1]. Jeden umożliwia adresowanie pamięci programu, drugi pamięci danych. Dzięki temu zmodyfikowana architektura harwardzka procesora jest efektywnie wykorzystywana.

Każdy generator adresów ma cztery rejestry indeksowe I, które służą do zapamiętywania wskaźników adresowych, cztery rejestry modyfikujące M — przeznaczone do zmiany adresów wskazywanych przez rejestry I oraz cztery rejestry L — długości, które określają zakres zwiększania adresu, rys. 7. Wymienione rejestry wykorzystywane są przy adresowaniu typu modulo lub do tworzenia bufora kołowego. Mogą one również być używane jako rejestry ogólnego przeznaczenia do zapamiętywania danych. Generatory adresów posiadają także zdolność negacji wszystkich bitów adresowych, co jest przydatne przy realizacjach algorytmów FFT.



Rys. 7. Schemat blokowy generatora adresów procesora ADSP 2101

Inną cechą generatorów adresów procesora ADSP 2101 jest możliwość wykorzystywania ich do automatycznego buforowania danych przesyłanych do odbieranych przy pomocy portu szeregowego procesora. Procesor stałoprzecinkowy ADSP 2101 umożliwia adresowanie bezpośrednie, pośrednie rejestrowe i buforowanie kołowe. Przy adresowaniu bezpośrednim adres operandu zawarty jest bezpośrednio w kodzie instrukcji. Jest to możliwe, bo adres jest 14 bitowy, a kod rozkazu jest 24 bitowy. Poniżej przedstawiona jest instrukcja używająca adresowania bezpośredniego do czytania danych z pamięci:

MX0=DM (etykieta);

Adresowanie pośrednie jest realizowane przy użyciu wspomnianych już rejestrów I, M, L. W tym samym cyklu gdy dana jest pobierana z pamięci o adresie wskazywanym przez rejestr I rejestr ten jest modyfikowany przez zawartość rejestru M i przygotowywany tym samym do wskazania następnego operandu. Poniższa instrukcja ilustruje opisany mechanizm.

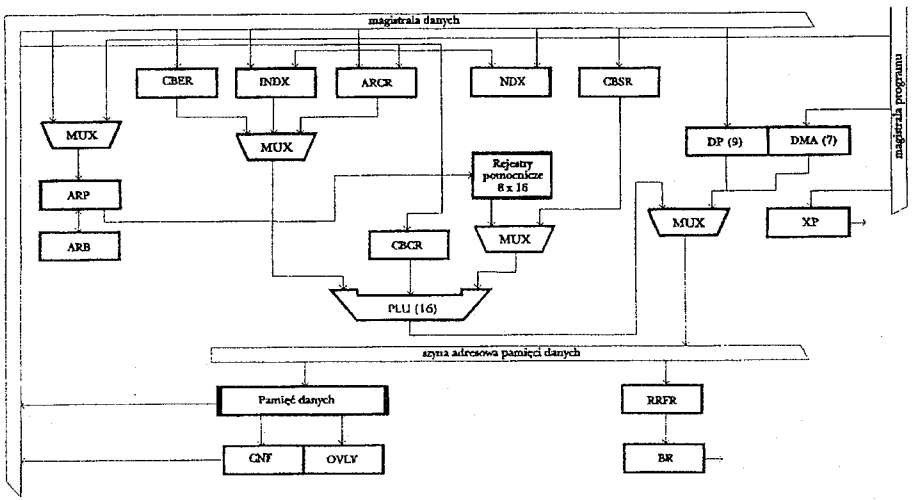
AX0=DM (I0,M3)

Do rejestru AX0 ładowany jest operand z pamięci danych i rejestr I0 jest modyfikowany przez rejestr M3. Dzięki istnieniu trzeciego rejestru L można realizować adresowanie cykliczne, modulo L, tworzyć bufor o określonej przez rejestr L długości. Możliwości te są szczególnie przydatne przy realizacji algorytmów filtrów cyfrowych, układów opóźniających itp.

3.2. MOŻLIWOŚCI ADRESOWE PROCESORA TMS 320C50

W procesorze TMS 320C50 zbiór rejestrów pomocniczych wykorzystywany jest do zapamiętywania i modyfikowania adresów. W danej chwili tylko jeden adres może być tworzony w zbiorze rejestrów pomocniczych. Zatem, niemożliwe jest adresowanie w jednym cyklu maszynowym dwóch operandów tak, jak ma to miejsce w ADSP 2101. Zbiór rejestrów pomocniczych jest połączony z jednostką arytmetyczną, co umożliwia automatyczne zwiększanie zawartości rejestrów pomocniczych lub ich zerowanie. W procesorze TMS 320C50 możliwa jest modyfikacja adresu polegająca na jego zwiększaniu lub zmniejszaniu o 1. Przy pomocy adresowania bezpośredniego w procesorze 'C50 można zaadresować 128 słów pamięci — w ADSP 2101 16 K słów. Do zwiększenia dostępnej przestrzeni adresowej wykorzystywany jest w połączeniu z adresem bezpośrednim 9 bitowy rejestr stronicowy — DP — rys. 8 [2].

Do tworzenia bufora kołowego wykorzystywane są rejestry CBSR i CBER. Pierwszy z nich wskazuje adres początkowy bufora, drugi adres końcowy. Jako wskaźniki stosowane są rejestry pomocnicze. Z tej przyczyny nie można realizować bufora kołowego w pamięci programu. Przy jego realizacji sprawdza się tylko, czy wskaźnik następnego adresu jest równy adresowi końca bufora. Nie istnieje sprzętowa kontrola o ile wskaźnik przekroczył adres końcowy. Możliwe jest tylko zwiększanie wskaźnika adresu o 1. W ralizacjach algorytmów, które wymagają większego niż 1 cyklicznego przyrostu adresu, np. przy interpolacji filtrów, należy programowo kontrolować wartość wskaźnika, na co potrzeba dodatkowego cyklu maszynowego dla każdego adresowanego słowa danych. Maksymalny rozmiar bufora kołowego w procesorze TMS 320C50 jest ograniczony do 256 słów. Ma to wpływ na rozmiar aplikcji (np. stopień filtru), które można przy jego pomocy realizować.



Rys. 8. Schemat blokowy generatora adresów procesora TMS 320C50

Mówiąc o adresowaniu w procesorze 'C50 nie można pominąć specyficznego sposobu zapisu adresów w niektórych operacjach. Przedstawiają to dwie instrukcje:

```
ADD*    ;
MPY*0+  ;
```

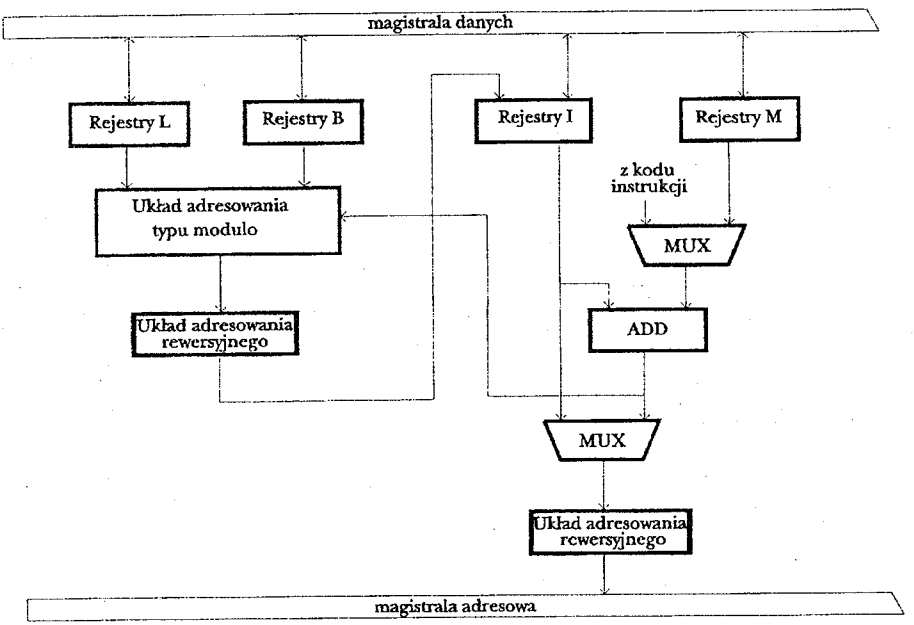
Pierwsza dodaje do akumulatora zawartość komórki pamięci wskazywanej przez rejestr pomocniczy, druga mnoży zawartość akumulatora przez operand wskazywany przez rejestr pomocniczy po czym modyfikuje zawartość rejestru dodając do niego cyfrę 0. Instrukcje te mogą być mało zrozumiałe, ponieważ nie używają bezpośrednio nazw rejestrów. Informacja ta zawarta jest w rejestrze pomocniczym ARP określającym, który z rejestrów wskazuje adres danej.

Podobnie jak w procesorze ADSP 2101 tak i w TMS 320C50 istnieje możliwość negacji bitów adresu, tzw. adresowanie rewersyjne stosowane do realizacji FFT.

3.3. MOŻLIWOŚCI ADRESOWE PROCESORA ADSP 21020

Procesor ADSP 21020 posiada możliwości bezpośredniego adresowania całej pamięci programu i pamięci danych w jednym cyklu i jedną instrukcją. Zawartość żadnych rejestrów ogólnego przeznaczenia lub rejestrów danych nie musi być zapamiętywana w pamięci. Procesor może wykonywać w pojedynczym cyklu zarówno operacje związane z układami arytmetycznymi jak i bezpośrednio czytać lub zapisywać dane do pamięci. Sposób zapisu adresowania bezpośredniego danych jest bardzo czytelny. Na przykład załadowanie do rejestru R2 danej z pamięci danych o adresie 1000000 hex wygląda następująco:

R2=DM(0x1000000). Do adresowania pośredniego procesor ADSP 21020 wykorzystuje rejestry z układów generatora adresów. Przykładowa instrukcja tego typu adresowania ma postać PM(I9,M13)=R1. Oznacza ona zapis do komórki pamięci programu wskazywanej przez rejestr I9, zawartości rejestru R1 i zmodyfikowanie zawartości rejestru I9 o wartość rejestru M13. Wszystkie te operacje wykorzystywane są w jednym cyklu. Procesor ADSP 21020 wyposażony jest w dwa generatory adresów (DAGs) (rys. 9) [3]. Jeden jest stosowany do wskazywania danych pamięci danych, durgi adresuje pamięć programu.



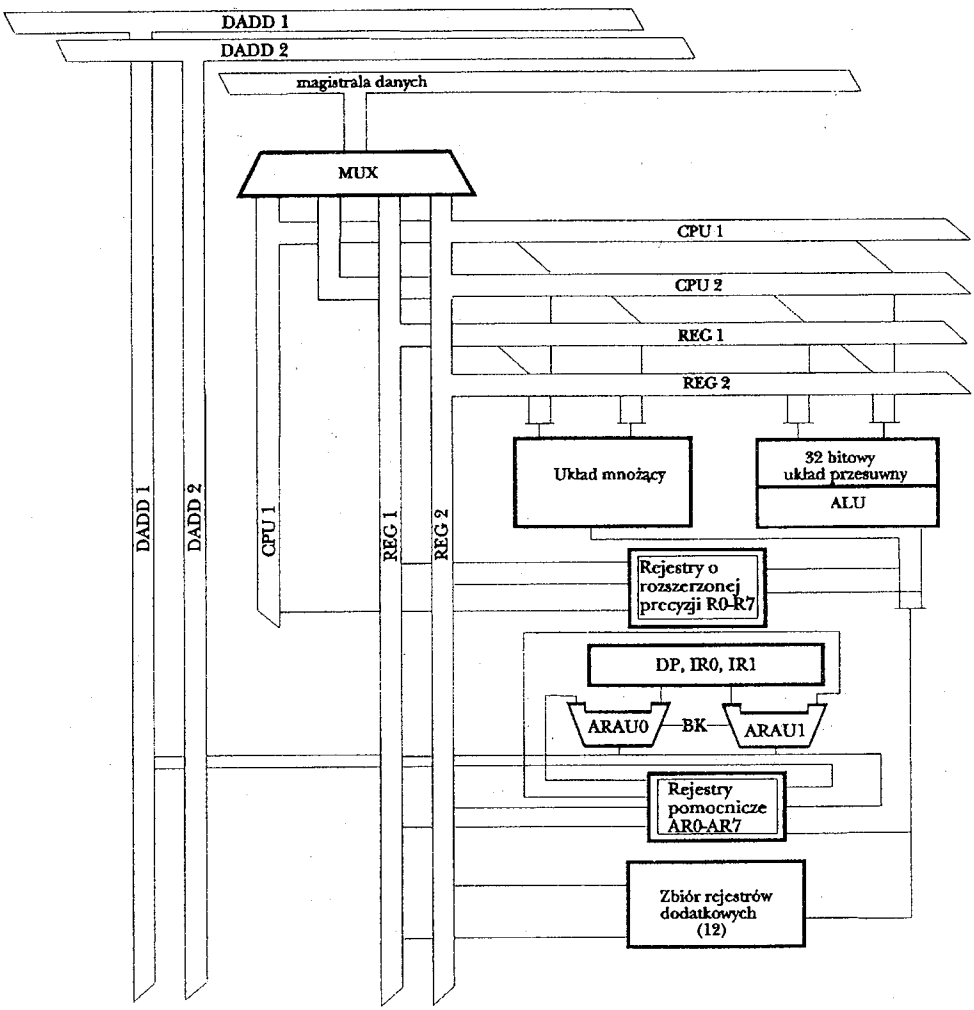
Rys. 9. Schemat blokowy generatora adresów procesora ADSP 21020

Każdy z generatorów adresów rys. 9 zawiera po 8 rejestrów indeksowych I i modyfikujących M. Każdy z rejestrów indeksowych może być użyty do adresowania innej przestrzeni pamięci. Rejestry te mogą być modyfikowane przez jeden z 8 rejestrów modyfikujących M przed lub po wykonaniu operacji. Każdy rejestr indeksowy posiada odpowiadający mu rejestr bazowy B i rejestr długości L. Służą one do konstruowania bufora kołowgo. Rejestr bazowy zawiera początkowy adres bufora, a odpowiadający mu rejestr L określa długość tego bufora. Po każdej modyfikacji rejestru I sprawdza się czy nowy adres jest zawarty w określonym przez B i L przedziale, jeśli nie wartość rejestru I jest poprawiana przez operację modulo. Ta właściwość określana jest jako adresowanie typu modulo lub kołowe. W związku z tym, że istnieje 16 rejestrów I, L, M, B, L można dowolnie tworzyć 16 buforów kołowych. Procesor 21020 posiada także możliwość adresowania rewersyjnego.

3.4. MOŻLIWOŚCI ADRESOWE PROCESORA TMS 320C30

Procesor TMS 320C30 może zaadresować bezpośrednio jedną 64 K stronę pamięci. Wynika to z długości słowa instrukcji — 32 bitowego (ADSP 21020 — ma rozkazy 48 bitowe). Strona wskazywana jest przez rejestr wskaźnika strony DP (data page pointer). Układ 'C30 może wykonywać operacje arytmetyczne w układzie ALU i mnożącym (ale nie obie jednocześnie) i równocześnie, bezpośrednio czytać dane z obszaru jednej strony. Bezpośrednie adresowanie pamięci dla zapisu danych i jednoczesne wykonywanie operacji obliczeniowych jest niemożliwe.

Procesor TMS 320C30 umożliwia wykonywanie adresowania indeksowego i rewersyjnego, posiada osiem 24 bitowych rejestrów adresowych AR0 — AR7,



Rys. 10. Jednostka centralna procesora TMS 320C30

rys. 10 [4]. Rejestry te mogą być używane do adresowania różnych obszarów pamięci. 'C30 może pobierać dwa operandy z pamięci w jednym cyklu używając dwóch z ośmiu rejestrów adresowych. Rejestry te mogą być modyfikowane przed lub po wykonaniu operacji przez liczbę 0, 1 lub przez zawartość jednego z dwóch rejestrów IR1, IR0. TMS320C30 posiada bufor kołowy, ale ma tylko jeden rejestr określający jego długość BK. Zatem, jeśli w podprogramie jest używanych kilka buforów kołowych, muszą one być jednakowej długości lub po każdej operacji zawartość rejestru BK musi być zmieniana. Adres początkowy bufora jest określany przez jego długość.

3.5. PORÓWNANIE MOŻLIWOŚCI ADRESOWYCH PROCESORÓW SYGNAŁOWYCH

Dla uzyskania dużej szybkości przetwarzania procesora ważnym jest, by czas czytania i zapisu danych do i z pamięci był jak najkrótszy. Możliwości adresowe danego procesora decydują więc o szybkości jego działania. Tablica 5 zawiera porównanie cech adresowych omawianych procesorów stałopozycyjnych ADSP 2101 i TMS 320C50.

Tablica 5

Zestawienie możliwości adresowych procesorów ADSP 2101 i TMS 320C50

cechy	ADSP 2101	TMS 320C50
pobranie dwu operandów z pamięci wewnętrznej w jednym cyklu		nie
operacje mnożenia i akumulacji w jednym cyklu		
modyfikacja dwóch adresów przez dwie różne wartości w każdym cyklu		nie
możliwość adresowania rewersyjnego		
automatyczne zwiększanie wskaźnika instrukcji przy adresowaniu kołowym		
adresowanie kołowe i typu modulo		nie

Prezentowane procesory zmiennoprzecinkowe ADSP 21020 i TMS 320C30 mają podobne możliwości adresowe. Procesor ADSP posiada większą liczbę rejestrów służących do adresowania pośredniego i może zaadresować bezpośrednio większy obszar pamięci. Tablica 6 zawiera porównanie obu procesorów ze względu na ich zdolności adresowe.

Tablica 6

Zestawienie wybranych możliwości adresowych procesorów ADSP 21020 i TMS 320C30

cechy	ADSP 21020	TMS 320C30
adresowanie bezpośrednie	32 bitowe	16 bitowe
liczba rejestrów indeksowych	16	8
liczba rejestrów modyfikujących	16	2
liczba rejestrów długości	16	1
liczba rejestrów bazowych	16	0
jednoczesne czytanie i zapisywanie operandów adresowanych pośrednio oraz wykonanie dwóch obliczeń		nie
jednoczesne czytanie dwóch operandów adresowanych pośrednio i wykonanie dwóch obliczeń		
jednoczesne zapisanie dwoch operandów adresowanych pośrednio i wykonanie dwóch obliczeń		nie
jednoczesne zapisanie dwóch operandów adresowanych pośrednio i wykonanie jednego obliczenia		nie
adresowani kołowe i typu modulo		nie

4. UKŁAD STEROWANIA PROGRAMEM

Aby procesor sygnałowy mógł przeprowadzać operacje w czasie rzeczywistym musi oprócz dużej szybkości obliczeniowej i krótkiego dostępu do pamięci w bardzo efektywny sposób kontrolować i sterować wykonywany program. Efektywność ta zależy od wielu czynników, ale najważniejsze to, w jaki sposób realizowane są instrukcje pętli, skoków, skoków warunkowych, wywołania podprogramów obsługi przerwań. Operacje te bardzo często występują w algorytmach cyfrowego przetwarzania sygnałów. Od czasu ich trwania uzależniona jest przydatność procesorów do realizacji określonych algorytmów.

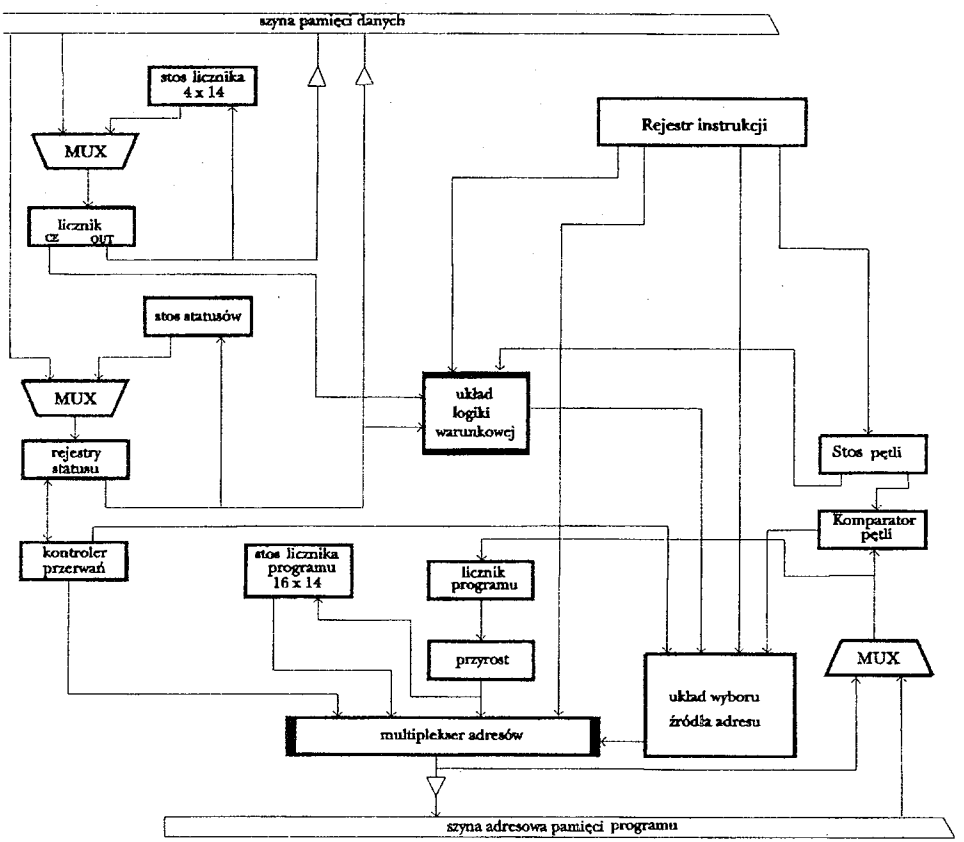
4.1. UKŁAD STEROWANIA PROGRAMEM PROCESORA ADSP 2101

Układ sterowania programem procesora ADSP 2101 umożliwia wybór adresu następnego rozkazu jako jednego z:

— zawartości licznika programu,
— 14 bitów adresu zawartych w kodzie instrukcji, dla bezpośrednich skoków lub wywołań podprogramów,
— stosu licznika programu, przy powrotach z podprogramów i przerwań,
— adresu automatycznej obsługi przerwań zewnętrznych.

Wszystkie instrukcje wykonywane są w jednym cyklu, niepotrzebne są instrukcje potokowe, dzięki czemu przebieg programu jest prosty do zrozumienia. Gdy występuje przerwanie, wszystkie rejestry statusu procesora zostają automatycznie zapamiętane na stosie statusu, a przy powrocie z procedury obsługi przerwań są ponownie z niego pobierane, rys. 11. Procesor ADSP 2101 pozwala na bardzo sprawną realizację pętli „DO UNTIL". Używa do tego celu stosu licznika (count stack), stosu pętli (loop stack) i komparatora pętli (loop comparator), rys. 11, [1]. Przy ich pomocy określa, czy pętla powinna się zakończyć i wyznacza adres następnej instrukcji. Wielkość pętli „DO UNTIL" może być praktycznie dowolna od jednej instrukcji do rozmiaru określonego wielkością pamięci programu. Wyjście z pętli jest możliwe, gdy zostanie wyzerowany 14 bitowy licznik pętli, lub gdy zostanie spełniony określony warunek arytmetyczny. Poniżej przedstawiony został przykład pętli wykonującej 100 razy trzy instrukcje:

CNTR = 100;
Do etykieta UNTIL CE;



Rys. 11. Układ sterowania programem procesora ADSP 2101

```
                    pierwsza instrukcja pętli;
                    druga instrukcja pętli;
etykieta:  ostatnia instrukcja pętli;
           pierwsza instrukcja poza pętlą;
```

Pierwsza instrukcja powoduje załadowanie licznika powtórzeń wartością 100. Instrukcja DO UNTIL zawiera adres ostaniej instrukcji w pętli i określa warunek zakończenia wykonywania pętli (w przykładzie adres jest reprezentowany przez etykietę a wyjście z pętli może nastąpić po wyzerowaniu przerzutnika CE). Na początku wykonywania pętli adres pierwszej instrukcji występującej w pętli zostaje załadowany na stos PC, a adres ostatniej na stosie pętli. W czasie wykonywania operacji pętli komparator sprawdza czy instrukcja, która ma zostać wykonana jest ostatnią w pętli czy nie. Jeśli jest to ostania instrukcja i jednocześnie jest spełniony warunek, że ma nastąpić wyjście z pętli, to ze stosu PCV pobierana jest pierwsza instrukcja pętli a licznik programu — PC jest zwiększony o 1, aby przejść do wykonywania następnej instrukcji poza pętlą. Mechanizm wykonywania pętli DO jest realizowany automatycznie bez uczestnictwa użytkownika. Procesor ADSP 2101 posiada możliwość zagnieżdżania pętli do czterech poziomów. Jest to szczególnie przydatne przy realizacji operacji z macierzami i wykonywaniu złożonych obliczeń.

W zbiorze instrukcji układu ADSP 2101 istnieje wiele instrukcji warunkowych. Większość rozkazów arytmetycznych tak samo jak instrukcje skoków, wywołań podprogramów, powrotów z procedur obsługi przerwań mogą być poprzedzone warunkami. Układ sterowania programem podejmuje decyzję, czy warunek jest spełniony i jakie operacje mają być wykonane. Oto przykłady zapisu kilku instrukcji warunkowych:

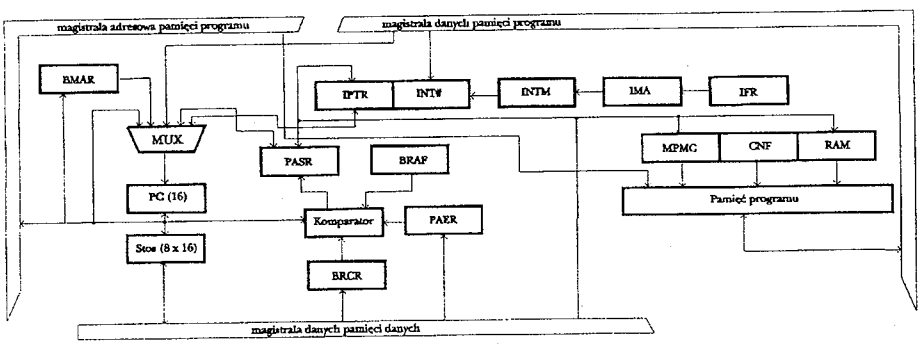
```
IF warunek JUMP etykieta;
IF warunek JUMP I4;
IF warunek CALL etykieta;
IF warunek CALL I4;
IF warunek RTS;
```

W przykładzie tym rejestr I4 zawiera adres następnej instrukcji, „warunek" odpowiada jednemu z 16 możliwych warunków arytmetycznych, etykieta oznacza adres instrukcji w przestrzeni adresowej programu, która będzie wykonywana po spełnieniu żądanego warunku.

4.2. UKŁAD STEROWANIA PROGRAMEM PROCESORA TMS 320C50

Układ sterowania programem procesora TMS 320C50 składa się z licznika programu, stosu i układów pomocniczych. Procesor posiada blok funkcji powtórzeń kontrolowany przez trzy rejestry PASR, PAER, BRCR, które pamiętają pierwszy i ostatni adres pętli oraz liczbę powtórzeń, rys. 12, [2]. Ze względu na potokowość wykonywanych instrukcji pętla musi składać się z co

najmniej trzech rozkazów. Jest wykonywana automatycznie, ale ponieważ nie ma lokalnego stosu i pamięci do pamiętania zawartości licznika pętli, trudno jest realizować zagnieżdżanie pętli. Istniejąca logika umożliwia wykonanie nie więcej niż 256 powtórzeń jednej instrukcji. W procesorze 'C50 przy wykonywaniu instrukcji stosowane jest czteropoziomowe przetwarzanie potokowe, składające się z pobrania rozkazu, dekodowania, pobrania operandu i wykonania operacji. Wspomniana potokowość powoduje, że przy realizacji niektórych instrukcji np. ładowaniu danych do rejestrów, wykonywaniu pętli, obsłudze przerwania potrzeba dodatkowych cykli maszynowych. Cecha ta różni procesor TMS od ADSP 2101, który nie ma żadnych ograniczeń szybkości przetwarzania związanych z potokowością wykonywanych instrukcji. Należy dodać, że potok pobieranych instrukcji może zostać zakłócony przy wykonywaniu rozkazów skoku, wywołania podprogramów lub przerwań. Potrzeba wówczas przynajmniej trzech cykli maszynowych do ponownego odzyskania utraconej potokowości.



Rys. 12. Układ sterowania programem procesora TMS 320C50

Procesor 'C50 posiada rejestr — licznik pobieranych instrukcji PFC, w którym przechowywany jest adres następnego rozkazu. Rozkaz ten ładowany jest do rejestru instrukcji IR, chyba że zawiera on jeszcze aktualnie wykonywaną instrukcję. W tym przypadku jest on zapamiętywany tymczasowo w rejestrze kolejkowym instrukcji QIR (rys. 12). Potokowość instrukcji w połączeniu z wykonywaniem rozkazów w kilku cyklach maszynowych może powodować, że program napisany w asemblerze 'C50 będzie mało czytelny i łatwo mogą powstawać błędy przy jego realizacji. Przykład na rys. 13 ukazuje jak niewiele różniące się dwa programy, jedynie instrukcją NOP, dają zupełnie inne wyniki. Podobnych problemów związanych z zachowaniem potokowości pozbawiony jest procesor ADSP 2101 i wszystkie instrukcje wykonuje w czasie jednego cyklu maszynowego.

W procesorze TMS 320C50 licznik programu może być wykorzystywany do adresowania sekwencyjnego, a pojedynczy stos licznika programu o pojemności 8 słów może służyć zarówno do zapamiętywania adresu powrotnego z podprogramu jak również danych z akumulatora.

```
PROB1   LAR     AR2,#067h;  do rejestru AR2 wpisywana jest wartość 67
        LACC    #064h;      do akumulatora ACC wpisywana jest wartość 00000064
        SAMM    AR2;        przyjęcie rejestru AR2 za domyślny przy
                            następnych dwóch instrukcjach, ARP = 2
        LACC    *-;         do akumulatora wpisywana jest wartość adresowana
                            przez rejestr AR2, a jego zawartość zmniejszana jest o 1,
                            czyli AR2 = 66
        ADD     *-;         do akumulatora dodawany jest argument wskazywany
                            przez rejestr AR2 i jego zawartośc zmniejszana jest o 1,
                            czyli AR2 = 65

PROB2   LAR     AR2,#067h;  do rejestru AR2 wpisywana jest wartość 67
        LACC    #064h;      do akumulatora ACC wpisywana jest wartość 00000064
        SAMM    AR2;        przyjęcie rejestru AR2 za domyślny przy
                            następnych instrukcjach, ARP = 2
        LACC    *-;         do akumulatora wpisywana jest wartość adresowana
                            przez rejestr AR2, a jego zawartość zmniejszana jest o 1,
                            czyli AR2 = 66
        NOP     ;           mimo rozkazu NOP zawartość rejestru AR2 ulega
                            zmianie i AR2 = 64
        ADD     *-;         do akumulatora dodawany jest argument wskazywany
                            przez rejestr AR2 i jego zawartośc zmniejszana jest o 1,
                            czyli AR2 = 63
```

Rys. 13. Przykładowe fragmenty programów napisane w asemblerze procesora TMS 320C50

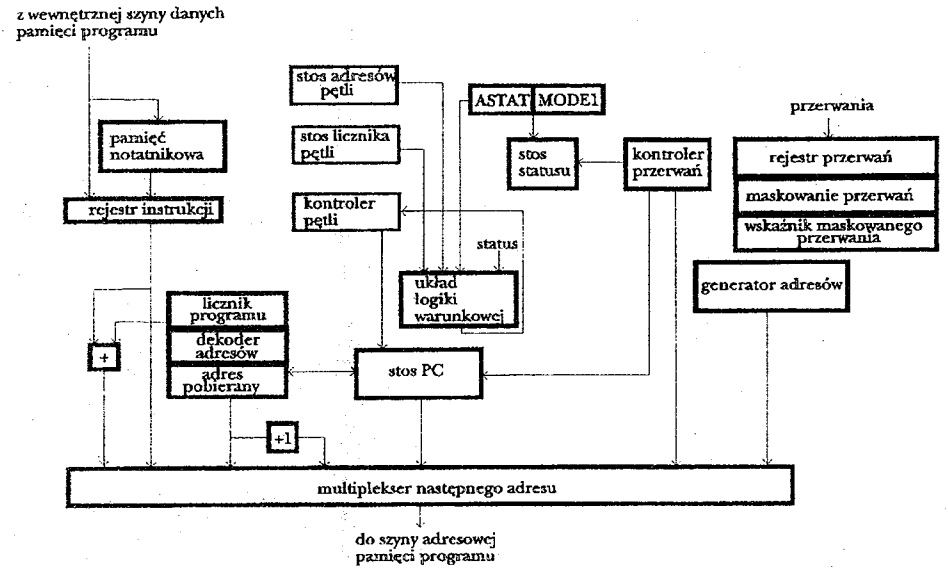
Rejestr flagowy — statusu — przerwań IFR jest wykorzystywany do wywołania procedur obsługi przerwań. Inaczej niż w ADSP 2101 bity statusu przerwań nie są automatycznie zapisywane i programista musi pamiętać o użyciu funkcji zapisu i odczytu stanu wektora przerwań.

W 'C50, którego pamięć programu jest 16 bitowa, instrukcje rozgałęzienia zawierające w swym kodzie bezpośredni adres operandu zajmują więcej niż jedno słowo pamięci programu. W związku z tym potrzebują co najmniej dwóch cykli pobrania, co zwiększa czas wykonywanych operacji. Przykładem tego są rozkazy: B 191 — skok bezwarunkowy do adresu 191 lub BANZ 191 — skok do adresu 191, jeśli zawartość aktualnego rejestru pomocniczego AR jest różna od zera. Ponadto układ ten wymaga często przy wykonywaniu instrukcji warunkowych i skoku dodatkowych cykli maszynowych. Ich liczba nie jest stała i zależy od rodzaju warunku.

4.3. UKŁAD STEROWANIA PROGRAMU PROCESORA ADSP 21020

Układ sterowania programem procesora ADSP 21020 pozwala w łatwy i efektywny sposób realizować program posiadający odwołania do podprogramów, wywołania procedur przerwań, operacje warunkowe, pętle typu DO. Umożliwia wykonywanie złożonych operacji warunkowych. Od tego samego warunku mogą być jednocześnie uzależnione operacje arytmetyczne i przesyła-

nia danych. Adres rozgałęzienia programu może być zawarty bezpośrednio w kodzie instrukcji lub określony pośrednio przez rejestr indeksowy, a 24 bitowy adres pamięci programu może być określony przez dowolny skok warunkowy lub bezwarunkowy, rys. 14.



Rys. 14. Układ sterowania programem procesora ADSP 21020

Układ 21020 posiada trzypoziomowy potok instrukcji składający się z fazy pobrania, dekodowania i wykonania. Oznacza to, że rozkaz pobierany z pamięci zostanie wykonany dopiero po dwóch cyklach maszynowych. Podczas realizacji skoku przetwarzanie potokowe zostaje zakłócone. Po rozkazie skoku zamiast instrukcji umieszczonych w sekwencji przetwarzanej wykonywane są dwie instrukcje NOP. Istnieje możliwość wyeliminowania zbędnych operacji NOP przez użycie instrukcji tzw. skoku opóźnionego, wówczas dwie instrukcje znajdujące się w przetwarzanym potoku są zrealizowane przed wykonaniem skoku. Podobnie istnieje możliwość opóźnienia, o dwa cykle, wywołania podprogramu. Podczas wywoływania procedury adres powrotny jest zapamiętywany na stosie licznika programu (program counter stack), rys. 14, [3].

Procesor ADSP 21020 umożliwia przyjmowanie przerwań generowanych wewnętrznie, zewnętrznie i programowo. Odpowiada na przerwanie z opóźnieniem nie większym niż cztery cykle. Rejestr maskujący przerwań MODE1 pozwala na indywidualne włączanie i wyłączanie odpowiednich przerwań oraz posiada bit IRPTEN, który maskuje wszystkie przerwania. Pod pierwszym z nich może być umieszczony rozkaz skoku opóźnionego do adresu procedury obsługi przerwania. Każdorazowo, gdy występuje przerwanie na stos statusu, zapisywane są automatycznie zawartości rejestrów statusowych: arytmetycznych, trybu i maskowania przerwań. Przy wywoływaniu podprogramów

programista sam musi wykonywać ładowanie na stos odpowiednich rejestrów statusowych. ADSP 21020 pozwala na zagnieżdżenie przerwań. Posiada również specjalną instrukcję IDLE, która wprowadza procesor w stan oczekiwania na przerwanie. Gdy przerwanie wystąpi, procesor przystępuje do jego obsługi po czym powraca do następnego rozkazu w programie po IDLE.

Po wykonaniu podprogramu lub procedury obsługi przerwania układ sterowania programem powoduje, że do licznika programu — PC wpisywany jest adres instrukcji występującej w programie po wywołanej procedurze. ADSP 21020 umożliwia dodatkowo stosowanie powrotów warunkowych z podprogramów oraz tzw. skoków opóźnionych. Dzięki nim nie występuje potrzeba stosowania dodatkowych cykli maszynowych przy powrotach do programu głównego — przetwarzanie potokowe nie zostaje zakłócone.

Procesor zmiennoprzecinkowy firmy Analog Devices posiada sprzętowo rozwiązaną realizację pętli DO. Przed wykonywaniem instrukcji występujących w pętli zapamiętywany jest adres ostatniego rozkazu oraz warunek pętli. Dzięk istnieniu sześciopoziomowego stosu pętli możliwe jest zagnieżdżanie sześciu instrukcji pętli. Każda z nich posiada swój własny 32 bitowy licznik. Wyjście z pętli może nastąpić na skutek spełnienia warunku arytmetycznego, odpowiedniej zawartości rejestru statusowego pętli, testowania określonych bitów, statusu wejść lub bezwarunkowo. Wyjście z pętli połączone jest z automatycznym zdjęciem ze stosu pętli odpowiadajacych jej adresów.

4.4. UKŁAD STEROWANIA PROGRAMEM PROCESORA TMS 320C30

Układ sterowania programem zmiennoprzecinkowego procesora sygnałowego TMS 320C30 podobnie jak układ jego odpowiednika firmy Analog Devices posiada wszystkie podstawowe cechy jakie muszą zawierać procesory sygnałowe dla sprawnego wykonywania algorytmów cyfrowego przetwarzania sygnałów. Istnieją jednak między nimi subtelne różnice. Między innymi w 'C30 jedynie skok bezwarunkowy może dotyczyć przestrzeni pamięci określonej 24 bitami, skoki warunkowe mogą odbywać się w obszarze 32K (adres 16 bitowy).

Przetwarzanie potokowe przebiega w czterech cyklach: pobrania instrukcji i uaktualnienia licznika programu (PC), zdekodowania instrukcji i modyfikacji rejestrów pomocniczych oraz wskaźnika stosu, pobrania operandu z pamięci — jeśli istnieje taka potrzeba, wykonania rozkazu. Istnieje możliwość wykonania tzw. skoków opóźnionych, wówczas potok wykonywanych opracji nie zostaje zakłócony. W przeciwnym razie potrzeba czterech dodatkowych cykli maszynowych. W procesorze 'C30 nie istnieje jednak możliwość wykonywania opóźnionych wywołań podprogramów i obsługi procedur przerwań. W tych przypadkach zakłócenie potokowości jest nieuniknione.

TMS 320C30 umożliwia przejście do obsługi procedury przerwania z opóźnieniem nie większym niż pięć cykli maszynowych. Wszystkie przerwania mogą być maskowane indywidualnie lub przez jeden głowny bit (GIE). Każde przerwanie posiada tylko jeden adres, pod którym może zostać zapisany skok do

procedury jego obsługi. Powoduje to ponowne zakłócenie przetwarzania potokowego i tym samym całkowite opóźnienie wprowadzane przez wystąpienie przerwania wynosi aż 9 cykli [4]. Procesor ten nie posiada stosu statusu, przez co programista musi pamiętać o zapisywaniu na stos zawartości rejestrów statusu przy wykonywaniu procedur obsługi przerwań. Podobnie jak w liście rozkazów ADSP 21020 tak i wśród instrukcji procesora 'C30 istnieje rozkaz IDLE wprowadzający procesor w stan oczekiwania na przerwanie. Oprócz przerwań zewnętrznych istnieje możliwość generowania przerwań wewnętrznych, przez dwa układy zegarowe, oraz przerwań programowych. Nie można jednak generować przerwań pod wpływem warunków arytmetycznych i przy pomocy rejestrów adresowych. Przy powrotach z podprogramów lub procedur obsługi przerwań zawsze potrzeba trzech dodatkowych cykli maszynowych. Nie ma możliwości zastosowania powrotów opóźnionych.

W przeciwieństwie do ADSP 21020 w 'C30 możliwe jest wykonywanie jedynie jednej instrukcji pętli DO. Nie istnieje możliwość zagnieżdżania tych pętli. Pętla DO jest realizowana przez rejestry RS (rejestr adresu początkowego), RE (rejestr adresu końcowego pętli), licznik powtórzeń — RC, rys. 10. Jeśli w trakcie wykonywania pętli w programie głównym wystąpi przerwanie, to programista musi w procedurze obsługi zapamiętać stan rejestrów RE, RS, RC; ich zawartość nie jest zapamiętywana automatycznie. Wykonywanie pętli może zostać przerwane jedynie przez załadowanie 0 do rejestru powtórzeń RC.

4.5. PORÓWNANIE WŁASNOŚCI UKŁADÓW STEROWANIA PROGRAMEM PROCESORÓW SYGNAŁOWYCH

Większość algorytmów cyfrowego przetwarzania sygnałów składa się z operacji wektorowych i cyklicznych, które wykonywane są w pętli. Dlatego ważnym jest, by procesory sygnałowe sprawnie realizowały instrukcję pętli DO. Istotnym jest również, by procesory mogły wykonywać bez dodatkowych cykli maszyno-

Tablica 7

Zestawienie cech układów sterowania programem procesorów ADSP 2101 i TMS 320C50

cechy	ADSP 2101	TMS 320C50
rozmiar stosu licznika programu PC	16	8
możliwość zagnieżdżania pętli	4 poziomy	nie
instrukcje warunkowe arytmetyczne		nie
dowolne opuszczenie pętli		nie
wymagane przetwarzanie potokowe dla uzyskania pełnej mocy obliczeniowej	nie	4 poziomowy sposób przetwarzania
automatyczne zapisanie na stosie rejestrów statusu przy wystąpieniu przerwania obliczeń		nie

wych rozkazy skoków. Sprawność układu sterującego programem wpływa w sposób decydujący na szybkość przetwarzania procesora. Tablice 7 i 8 przedstawiają porównanie własności układów sterujących programem omawianych procesorów stało i zmiennoprzecinkowych.

Tablica 8

Zestawienie cech układów sterowania programem procesorów ADSP 21020 i TMS 320C30

cechy	ADSP 21020	TMS 320C30
skok warunkowy		nie
przetwarzanie potokowe	3 poziomy	4 poziomy
skok opóźniony		nie
opóźnione wywoływanie podprogramów		nie
wielkość wektora obsługi przerwań	8	1
opóźnione powroty z podprogramów		nie
możliwość zagnieżdżania pętli	6 poziomów	1 poziom
warunkowe wyjście z pętli		nie
stos statusu		nie

5. PAMIĘCI PROCESORÓW SYGNAŁOWYCH

Wszystkie omawiane procesory sygnałowe wyposażone są w wewnętrzne pamięci. Są one wykorzystywane do przechowywania danych i programu. Dzięki temu, że znajdują się w wewnętrznej strukturze procesorów zmniejszają czas dostępu do argumentów i instrukcji. Tym samym czas wykonywanych operacji znacznie maleje.

Procesor ADSP 2101 posiada pamięć wewnętrzną RAM danych o pojemności 1K × 16 bitów i programu 2K × 24 bity. Procesor jest w stanie w jednym cyklu pobrać dwa operandy z pamięci wewnętrznej danych i programu oraz następną instrukcję z wewnętrznej pamięci programu. W związku z tym nie jest wyposażony w pamięć notatnikową. Wewnętrzna pamięć programu może być ładowana z zewnętrznej pamięci EPROM tzw. BOOT automatycznie po wyzerowaniu procesora w zależności od sygnału sterującego na wejściu MMAP. Pamięć BOOT jest podzielona na 8 stron o pojemności 8K × 8 każda. Wszystkie wewnętrzne magistrale adresowe i danych są rozszerzone na zewnątrz procesora. Całkowity obszar pamięci programu i danych wynosi, dla każdej z pamięci, 16 K słów.

Układ TMS 320C50 posiada wewnętrzną programowaną pamięć ROM 2K × 16, która jest ładowana po wyzerowaniu procesora z zewnętrznych znacznie wolniejszych pamięci ROM lub EPROM. Jest ona sterowana sygnałem

na wejściu MP/$\overline{MC}$. Jeśli na tym wejściu jest stan niski, przestrzeń tej pamięci zajmuje pamięć zewnętrzna. Oprócz wymienionej pamięci istnieje wewnętrzna pamięć danych RAM IK × 16 o podwójnym dostępie. Część tej pamięci może być wykorzystana jako pamięć programu. W wewnętrznej strukturze procesora 'C50 znajduje się także pamięć programu RAM 9K × 16. Szesnastobitowa magistrala adresowa pozwala na zaadresowanie 64K przestrzeni pamięci.

Omawiane procesory zmiennoprzecinkowe ADSP 21020 i TMS 320C30 wyposażone są w wewnętrzne pamięci notatnikowe (Cache memory), które służą do przechowywania często powtarzanych instrukcji np. wykonywanych w pętli. Procesor ADSP 21020 posiada pamięć notatnikową o wielkości 32 słów. Zapamiętywane są w niej ostatnie 32 instrukcje wymagające dostępu do pamięci danych, pamięć notatnikowa jest zamrażana. Układ 21020 jest pozbawiony innych wewnętrznych pamięci RAM, posiada natomiast bardzo efektywny aparat dostępu do dużego obszaru pamięci zewnętrznych zarówno danych jak i programu. Czas dostępu do pamięci przy zapisie i czytaniu wynosi jeden cykl, chyba że programista wyraźnie zadeklaruje większą wartość oczekiwania. Stosowane pamięci mogą mieć czas dostępu 35 ns. Całkowita pamięć programu może mieć wielkość 16M słów (24 bitowa szyna adresowa), pamięć danych może osiągnąć aż 4 G słów (32 bitowa szyna adresowa).

Procesor TMS 320C30 nie ma tak efektywnego połączenia z pamięciami zewnętrznymi. Zapis do nich zawsze trwa dwa cykle i w większości przypadków tyle samo trwa odczyt. Posiada jednak pamięci wewnętrzne: dwa bloki pamięci RAM o pojemności 1K × 32 każdy, jeden blok pamięci ROM 4K × 32 oraz pamięć notatnikową. Ta ostatnia składa się z 64 słów. Zapamiętywane są w niej wszystkie 64 ostatnio wykonywane instrukcje korzystające z zewnętrznej pamięci bez względu na to, czy powodują konflikt magistrali czy nie. W przypadku realizacji pętli większej niż 64 instrukcje pamięć notatnikowa może być zamrażana. Istnieje również możliwość zerowania i wyłączania tej pamięci. Należy zwrócić uwagę na to, że w przypadku procesora ADSP 21020 może być wykonywana pętla złożona z więcej niż 64 instrukcji, byleby tylko liczba rozkazów powodujących konflikt magistral była mniejsza niż 32. Całkowity obszar pamięci jaki może być zaadresowany przez procesor 'C30 wynosi 16M słów 32 bitowych (24 bitowa magistrala adresowa).

6. PRZYKŁADOWE REALIZACJE DSP

Dla zilustrowania niektórych właściwości procesorów sygnałowych omawianych w poprzednich rozdziałach, przedstawione zostaną fragmenty programów, napisanych w asemblerze, realizujące przykładowe algorytmy cyfrowego przetwarzania sygnałów, bardzo często stosowane w praktycznych układach telekomunikacyjnych. Przy pomocy tych algorytmów wykazane są różnice między odpowiednimi procesorami stało i zmiennopozycyjnymi. Układy ADSP 2101 i TMS 320C50 porównane są na przykładzie realizacji prostego filtru o skoń-

czonej odpowiedzi impulsowej FIR, a ADSP 21020 i TMS 320C30 na podstawie algorytmu szybkiej transformaty fouriera FFT.

Rys. 15 przedstawia fragment kodu programu napisanego w asemblerze procesora ADSP 2101 realizującego filtr o skończonej odpowiedzi impulsowej według wzoru:

$$Ck+1 = Ck + Alfa*Beta*A(n) \tag{6.1}$$

```
AR=DM(Beta);
        MY1=Alfa;                                          {pobranie wartości Alfa      }
        MF=AR*MY1(RND), AY0=PM(I4,M4), MX0=DM(I0,M0); {MF=Alfa*Beta, pobież Ck, A }
        MR=MX0*MF(RND);                                    {MR=Alfa*Beta*A(n)           }
        CNTR=A;                                            {ustalenie parametrów pętli  }
        DO etykieta UNTIL CE;
              AR=MR1+AY0, AY0=PM(I4,M6), MX0=DM(I0,M1); {AR=cK+Alfa*Beta*A(n)  }
etykieta:     PM(I4,M7)=AR, MR=MX0*MF(RND);                {zapamiętanie Ck+1           }
        RTS;                                               {wyjście z podprogramu       }
```

Rys. 15. Fragment programu napisanego w asemblerze procesora ADSP 2101 realizujący filtr o skończonej odpowiedzi impulsowej — FIR

We fragmencie tego programu zastosowana została instrukcja pętli DO UNTIL, której liczba powtórzeń zależna od stopnia filtru może być łatwo zmieniana przez załadowanie do licznika powtórzeń CNTR. Ponadto zostało użyte adresowanie pośrednie do adresowania bufora zawierającego współczynnik i filtru Ck i dane wejściowe A(n). Realizują je rejestry indeksowe I0 i I4. Procedura rozpoczyna się od załadowania rejestru wyników ALU AR współczynnikiem Beta z pamięci danych. Następny rozkaz ładuje do rejestru wejściowego układu mnożącego MY1 zawartość Alfa. Obie wymienione zmienne są mnożone, a ich wynik jest zaokrąglany do 16 najbardziej znaczących bitów. W tym samym czasie pobierany jest współczynnik Ck z pamięci programu oraz dana wejściowa A(n) z pamięci danych. Należy pamiętać, że rownocześnie z tymi operacjami modyfikowane są rejestry I przez zawartość rejestrów M. Ta cecha adresowania i zdolności wykonywania aż 3 instrukcji w jednym cyklu maszynowym nie jest możliwa do wykonania na procesorze TMS 320C50. Wynik mnożenia przechowywany jest w rejestrze MF, który w następnym cyklu może być ponownie wykorzystany do realizacji operacji mnożenia przez daną wyjściową A(n) przechowywaną w rejestrze MX0. Następnie wykonywana jest pętla DO UNTIL, której główna instrukcja oblicza wartość $Ck+1$ i aktualizuje zawartości rejestrów dla następnych operacji. Wynik zapamiętywany jest w pamięci programu na końcu bloku pętli. Ostatnią instrukcją jest rozkaz powrotu z podprogramu RTS. Wszystkie instrukcje wykonywane są w czasie jednego cyklu. Zatem łączny czas realizacji tej procedury dla procesora ADSP 2101 wynosi $7+n*2$ cykli, gdzie n oznacza liczbę współczynników, od której zależy z kolei liczba powtórzeń w pętli.

Na rys. 16 przedstawiony jest kod programu napisanego w asemblerze procesora TMS 320C50 realizujący filtr FIR również według równania (6.1). Przykład ten oparty jest o wykorzystanie rozkazu RTPB powtarzającego wykonanie bloku instrukcji określoną liczbę razy. Do adresowania bufora współczynników filtru Ck i bufora danych wejściowych zastosowano adresowanie pośrednie. Jest ono realizowane przez rejestry pomocnicze AR2 i AR3. Wykonywanie procedury rozpoczyna się od przepisania wartości Beta z pamięci o adresie BETA do rejestru TREG0, aby następnie pomnożyć ją przez liczbę Alfa znajdującą się w pamięci o adresie ALFA. Wynik mnożenia znajduje się w rejestrze P układu mnożącego, ale aby mógł być wykorzystany w następnych obliczeniach, musi być przepisany do akumulatora — czyni to instrukcja PAC. Wynik mnożenia może zostać zaokrąglony do 16 najbardziej znaczących bitów przy pomocy rozkazu ADD ONE, I4 i zapamiętany w pamieci rozkaz SACH. Jak widać operacja zaokrąglenia potrzebuje dodatkowego cyklu maszynowego i nie jest jak w ADSP 2101 częścią operacji mnożenia.

```
        LT      BETA;          do T załaduj Beta
        MPY     ALFA;          P=Alfa*Beta
        PAC;                   ACC=P
        ADD     ONE, 14;       zaokrąglanie
        SACH    BETAM,1;       zapamiętanie ACC w pamięci
        LACC    #126;
        SAMM    BRCR;          ustalenie liczby powtórzeń
        LAR     AR2, #COEFFD;  wskaźnik współczynników filtru
        LAR     AR3, #LASTAP;  wskaźnik danych wejściowych
        LT      BETAM;
        MPY     *-, AR2;       P=Alfa*Beta*x(i-255)
        RPTB    ETY - 1;       od I=0 do I<=126, I++
ADR     ZALR    *, AR3;        załaduj ACCH przez rej. Ak
        MPYA    *-, AR2;       wykonaj końcowe operacje arytmetyczne w pętli
        SACH    *+;            zapamiętanie wyniku pośredniego
ETY     ZALR    *,AR3;
        RETD;                  powrót do programu głównego
        APAC;
        SACH    *+;            zapamiętanie wyniku końcowego
```

Rys. 16. Fragment programu napisanego w asemblerze procesora TMS 320C50 realizujący filtr o skończonej odpowiedzi impulsowej — FIR

Następnie wykonywana jest pętla obliczająca współczynniki filtru. Inicjują ją dwie instrukcje LACC i SAMM ustawiające licznik powtórzeń pętli BRCR w tym przypadku na wartość 127. Adresowanie pętli odbywa się przy użyciu dwóch rejestrów AR2 i AR3 inicjalizowanych instrukcją LAR. Zapamiętany w pamięci wynik mnożenia Alfa*Beta jest pobierany ponownie do rejestru T dla wykonania mnożenia z daną wejściową A(n). Pobranie wykonuje instrukcja LT BETAM, a mnożenie zawartości rejestru TREG0 z komórką pamięci o adresie wskazywanym przez rejestr AR rozkaz MPY*,

AR2. Jak widać z tego zapisu, w procesorze 'C50 należy pośredni wynik mnożenia współczynników Alfa i Beta przechowywać w pamięci. Operacji tej nie potrzeba było wykonywać w procesorze ADSP 2101. Główną część pętli realizuje rozkaz RPTB powtarzający wszystkie operacje do adresu opatrzonego etykietą umieszczoną w argumencie instrukcji. W przykładzie są to trzy instrukcje ZALR, MPYA i SACH wykonywane 127 razy. Należy zauważyć, że procesor ADSP 2101 ma dużo prostszy i bardziej czytelny sposób realizacji pętli DO UNTIL. Może ona składać się z dowolnej liczby rozkazów, a nie jak instrukcja RPTB w układzie 'C50, która musi zawierać co najmniej trzy rozkazy. Ostatnimi instrukcjami procedury filtru FIR jest zapamiętanie otrzymanego wyniku w pamięci danych oraz wykonanie instrukcji powrotu opóźnionego z procedury. Instrukcja ta jest konieczna dla zachowania przetwarzania potokowego występującego w 'C50. Rozkaz taki jest zbyteczny w ADSP 2101, ponieważ procesor ten wykonuje wszystkie operacje w jednym cyklu. Całkowita liczba instrukcji omawianej procedury wynosi 19. Niektóre z rozkazów są wykonywane w więcej niż w jednym cyklu. Łączny czas realizacji podprogramu można w przybliżeniu określić wyrażeniem 17+n*3, gdzie n oznacza rząd filtru. Jest on więc o (10+n) razy dłuższy niż czas realizacji takiego samego algorytmu na procesorze ADSP 2101.

Bardzo często stosowanym algorytmem cyfrowego przetwarzania sygnałów jest algorytm szybkiej transformaty Fouriera FFT. Jest on używany we wszystkich układach wykorzystujących widmo sygnałów. Od czasu jego wykonania i złożoności realizacyjnej zależy jaka moc obliczeniowa procesora może być przeznaczona na inne operacje. Na rysunku 17 przedstawiony został fragment realizacji najważniejszej części algorytmu FFT — obliczeń motylkowych — napisany w asemblerze procesora zmiennoprzecinkowego ADSP 21020. Jak widać, składa się z czterech operacji wielofunkcyjnych. Każda z instrukcji wykonywana jest w jednym cyklu. Zatem łączny czas tej procedury wynosi cztery cykle maszynowe. Taki krótki czas może być osiągnięty dzięki możliwości wykonywania przez układ 21020 równocześnie operacji mnożenia, dodawania lub odejmowania, dodawania i odejmowania (instrukcja druga), zapisania do pamięci zawartości rejestru oraz załadowanie rejestru określoną komórką pamięci. W czasie jednego cyklu może więc być wykonanych aż pięć różnych operacji.

```
R8=R1*R6,    R14=R11-R14,  DM(I2,M0)=R10,  R9=PM(I11,M8);
R11=R1*R7,   R3=R9-R14,    DM(I2,M0)=R13,  R7=PM(I8,M8);
R14=R0*R6,   R12=R2-R12,   R8=DM(I0,M0),   PM(I10,M10=R9);
R12=R0*R7,   R13=R8-R12,   R6=DM(I0,M0),   PM(I10,M10)=R3;
```

Rys. 17. Fragment programu napisanego w asemblerze procesora ADSP 21020 realizujący obliczenia motylkowe algorytmu FFT

Kod programu realizujący ten sam algorytm dla procesora zmiennoprzecinkowego TMS 320C30 jest dużo bardziej złożony i zarazem jest mniej czytelny (rys. 18). Procedura wykonująca obliczenia motylkowe składa się z 14 instrukcji. Niektóre z nich wykonywane są równolegle np. zmiennoprzecinkowe dodawa-

```
        subf   *ar2, *ar0, r2
        subf   *+ar2, *+ar0, r1
        mpy    r2, r6, r5
||      addf   *+ar2, *+ar0, r3
        mpyf   r1, *+ar4 (IR1), r3
||      stf    r3, *+ar0
        subf   r0, r3, r4
        mpyf   r1, r6, r0
||      subf   *+ar2, *+ar0, r3
        mpyf   r2, *+ar4(IR1), r3
||      stf    r3, ar0++ (IR0)
        addf   r0, r3, r5
        stf    r5, *ar2++ (IR0)
||      stf    r4, *+ar2
```

Rys. 18. Fragment programu napisanego w asemblerze procesora TMS 320C30 realizujący obliczenia motylkowe algorytmu FFT

nia dwóch operandów i mnożenie lub ładowanie wskazanej wartości do rejestru i mnożenie. Podprogram ten jest znacznie dłuższy niż analogiczny dla układu ADSP 21020. Główną przyczyną jest brak możliwości realizacji przez TMS 320C30 w jednym cyklu składowania zawartości rejestru do pamięci z jednoczesnym wykonywaniem dodawania i mnożenia. Nie może także być wykonywane w jednej instrukcji dodawanie i odejmowanie operujące na tych samych operandach wejściowych. Całkowity czas realizacji algorytmu szybkiej transformaty Fouriera drugiego rzędu dla 1024 punktów zmiennych zespolonych wynosi dla tego procesora 50660 cykli, co przy czasie cyklu 50 ns wynosi 2,533 ms. Ten sam algorytm wykonywany przez procesor ADSP 21020 trwa 21314 cykli, przy czasie realizacji instrukcji 40 ns daje prawie trzykrotny zysk 0,852 ms.

7. WNIOSKI

Oferowane współcześnie procesory sygnałowe chociaż umożliwiają realizację złożonych algorytmów cyfrowego przetwarzania sygnałów, różnią się między sobą niektórymi cechami architektonicznymi. Różnice te powodują, że niektóre procesory mogą w lepszy sposób — bardziej efektywny realizować określone algorytmy DSP niż inne. Procesory stałopozycyjne, mniej kosztowne, wymagają bardziej złożonych kodów programów i dłużej wykonują żądane operacje. Często jednak ich możliwości zupełnie wystarczają i względy ekonomiczne decydują o ich stosowaniu. Procesory zmiennoprzecinkowe pozwalają na realizację działań na liczbach z bardzo dużego zakresu, a jednocześnie uwalniają programistę od wykonywania kłopotliwej operacji skalowania liczb koniecznej przy operacjach stałoprzecinkowych. Z analizy cech omawianych w poprzednich rozdziałach procesorów sygnałowych ADSP i TMS wynika, że dużo łatwiej

i sprawniej realizowane są algorytmy DSP przy pomocy procesorów firmy Analog Devices. Potwierdzają to również własne doświadczenia autora. Procesory TMS stało- i zmiennoprzecinkowe posiadają bardzo specyficzny zapis mnemoniczny realizowanych operacji. Jest on mało czytelny nawet dla zaawansowanych programistów. Układy te w porównaniu z ich odpowiednikami ADSP pozbawione są kilku ważnych możliwości, jak np.: jednoczesnego wykonywania operacji arytmetycznych mnożenia i dodawania z dostępem do pamięci, opóźnionej instrukcji skoku, zagnieżdżania pętli, automatycznego zapisywania przy przerwaniu rejestrów statusu na stos, cyklicznego adresowania modulo. Wymagają również zwracania dużej uwagi na zachowanie przetwarzania potokowego dla uzyskania pełnej mocy obliczeniowej. Programy pisane w asemblerze procesorów ADSP są bardzo czytelne nie zawierają skomplikowanych odwołań do rejestrów i pamięci. Lista rozkazów tych układów jest bogatsza niż ich odpowiedników firmy Texas Instruments. Wśród rozkazów ADSP 21020 istnieją realizujące złożone operacje matematyczne np. obliczanie funkcji odwrotnej lub odwrotnej wartości pierwiastka kwadratowego. Wszystkie instrukcje nawet wielofunkcyjne są wykonywane w jednym cyklu, a potok przetwarzanych operacji procesora 21020 jest krótszy niż TMS'C30. Układy ADSP posiadają możliwość wykonywania pętli dowolną liczbę razy złożoną z instrukcji, których liczba ograniczona jest jedynie wielkością pamięci. Oprócz stosu programowego posiadają dodatkowe stosy licznika programu, pętli, statusu i układu licznika. Sposób realizacji operacji pętli, adresowania i wykonywania wielu funkcji w jednym cyklu maszynowym sprawiają, że tworzone przy ich pomocy aplikacje są bardzo efektywne i spełniają bardzo wygórowane żądania co do czasu i dokładności przetwarzania.

BIBLIOGRAFIA

1. Analog Devices, Inc., ADSP-2100 Family User's Manual, 1993
2. Texas Instruments, Inc. TMS 320C50 User's Guide, 1993
3. Analog Devices, Inc., ADSP-21020 Family User's Manual, 1993
4. Texas Instruments, Inc., TMS 320C30 User's Guide, 1990.

P. REMLEIN

COMPARISON THE PROPERTIES OF THE DIGITAL SIGNAL PROCESSOR'S

Summary

This paper presents the properties of fixed- and floating point digital signal processors of two known firms Texas Instruments and Analog Devices. The processors often are characterized by them MIPS rate. This approach doesn't give the reality performance of the processors. In this paper the digital signal processors are compared on the ground of them architecture, addressing modes and

program sequencing capabilities. The implementation of some algorithms DSP on these signal processors are presented. The complexities and time realisation of the algorithms for processors have been computed and compared.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1994. 40, z. 4. ss. 489–497

621.39
519.68

Nieliniowa korekcja interferencji z przewidywaniem przyszłej decyzji

JERZY KISILEWICZ

Instytut Sterowania i Techniki Systemów, Politechnika Wrocławska

Otrzymano 1994.09.25

Autoryzowano do druku 1994.12.20

W pracy opisano metodę korekcji szczególnego rodzaju interferencji międzysymbolowej przy użyciu korektora z decyzyjnym sprzężeniem zwrotnym bez korektora wstępnego. Korektor wstępny zastąpiono przewidywaniem decyzji. Na podstawie odebranego sygnału określa się prawdopodobieństwa aktualnie odbieranych danych binarnych i danych, które będą odbierane w następnym takcie. Przeprowadzono symulację transmisji przy użyciu opisanego korektora oraz tradycyjnego korektora z decyzyjnym sprzężeniem zwrotnym i korektorem wstępnym. Porównano częstości występowania błędów w obecności białego szumu gaussowskiego.

1. 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 $t=hT$ $(0\leqslant h\leqslant g)$, 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}.

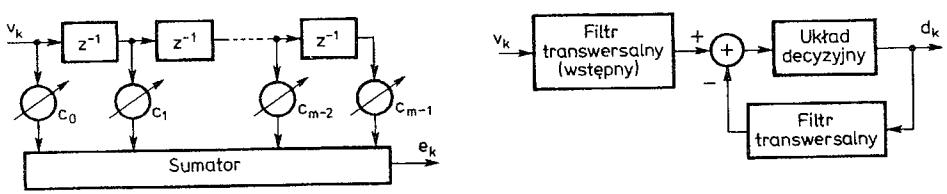
Występujące we wzorze (1) składniki $a_{k-i}y_i$ dla $0 \leqslant i \leqslant g$ oraz $i \neq h$ są komponentami interferencji międzysymbolowej.

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, 8, 10, 13].

Dołączenie korektora do kanału podstawowego ma na celu zmniejszenie prawdopodobieństwa błędnego odbioru poprzez redukcję interferencji międzysymbolowej. Do najczęściej stosowanych korektorów należą: korektor z decyzyjnym sprzężeniem zwrotnym oraz prosty filtr transwersalny.

Filtr transwersalny zbudowany jest z linii opóźniającej z odczepami oraz z wielowejściowego sumatora o regulowanych wzmocnieniach na każdym wejściu. Sygnały z odczepów linii opóźniającej zsumowane wagowo dają sygnał wyjściowy z korektora jak to pokazano na rys. 1a. Korektory te są zawsze stabilne, ale wymagają stosowania dużej liczby odczepów. Zaletą korektorów transwersalnych prostych jest zdolność do korekcji wszystkich komponentów interferencji. Pogarszają one jednak stosunek mocy sygnału do mocy szumu.

Korektory transwersalne z decyzyjnym sprzężeniem zwrotnym mają w pętli sprzężenia zwrotnego nieliniowy układ decyzyjny, który najczęściej jest demodulatorem transmitowanych danych. Schemat blokowy takiego korektora pokazano na rys. 1b. Użyty w sprzężeniu filtr transwersalny posiada cyfrową linię opóźniającą, w której pamiętane są kolejne decyzje detektora. Sygnał na wyjściu tego filtru nie zawiera szumu, tak więc korektory z decyzyjnym sprzężeniem zwrotnym nie pogarszają stosunku mocy sygnału do mocy szumu. Przy założeniu, że w linii opóźniającej nie ma błędnych decyzji, korekcja tych prążków interferencji, które występują po prążku głównym jest dokładna.



Rys. 1. Schemat: a) prostego filtru transwersalnego, b) korektora z decyzyjnym sprzężeniem zwrotnym

Ponieważ próbka v_k reprezentuje dane a_{k-h}, to w decyzyjnym sprzężeniu pamiętane są decyzje d_i o danych a_i dla $i<k-h$. Tak więc z próbki v_k zgodnie ze wzorem (1) można usunąć tylko te składniki interferencji $a_{k-i}y_i$, dla których $k-i<k-h$, czyli dla $i>h$. Pozostałe składniki (dla $i<h$) interferencji koryguje wstępny filtr transwersalny. Algorytmy syntezy korektorów opisano w [4 i 7].

Gdy w charakterystykach amplitudowych kanału występują zera (np. kanały radiowe opisane w [6]), liniowy filtr transwersalny jest mało efektywny jako korektor [6]. Ponadto korektor wstępny pogarsza stosunek mocy sygnału do

mocy szumu. Celowe jest tu więc poszukiwanie innych (niż korektor wstępny) metod korekcji tych składowych interferencji, które zależą od tych danych, co do których decyzja jeszcze nie została podjęta. W niniejszej pracy zaproponowano metodę korekcji jednego takiego składnika interferencji.

REGUŁA DECYZYJNA

Przyjmimy, 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.

Przyjmijmy taką interferencję kanału, że zawiera ona jeden istotny składnik nie korygowany przez decyzyjne sprzężenie zwrotne. Uwzględniając te założenia można z (1) napisać wzór na wartości próbek na wejściu układu decyzyjnego

$$V_k = a_{k+1} y_0 + a_k y_1. \tag{2}$$

Przy braku zakłóceń wzór (2) wyznacza możliwe wartości sygnału dla wszystkich możliwych kombinacji danych a_{k-1}, a_k. Zakładając transmisję danych binarnych otrzymuje się (jak to pokazano na rys. 2) 4 wartości:

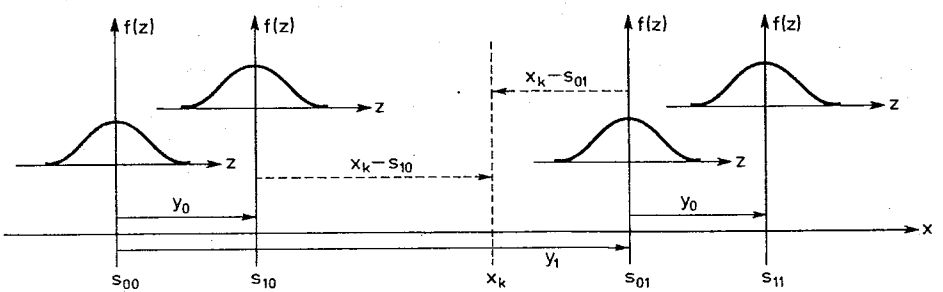
$$s_{00}=0,\ s_{10}=y_0,\ s_{01}=y_1,\ s_{11}=y_0+y_1.$$

Układ decyzyjny na podstawie znajomości tych wartości, oraz gęstości rozkładu prawdopodobieństwa zakłóceń z_k wyznaczy prawdopodobieństwa p_{00}, p_{10}, p_{01} i p_{11} powstania danego sygnału wejściowego

$$x_k = V_k + Z_k$$

dla wszystkich 4 kombinacji danych a_{k+1}, a_k. Układ decyzyjny wyznaczy taką wartość d_k danej a_k, dla której sygnał wejściowy y_k jest najbardziej prawdopodobny. Oznaczając przez $f(z)$ rozkład gęstości prawdopodobieństwa zakłóceń oraz przez Δ dokładność pomiaru x_k można wyznaczyć prawdopodobieństwo tego, że $z_k = y_k - s_{ij}$ $(ij=0,1)$ jako

$$p_{ij} - f(x_k - s_{ij})\,\Delta,\ (i,j=0,1).$$



Rys. 2. Nominalne poziomy sygnału x_k

Prawdopodobieństwa tego, że $a_{k+1}=i$ oraz $a_k=j$ przy odebranym sygnale x_k wynoszą

$$\overline{p_{ij}}=\frac{f(x_k-s_{ij})}{f(x_k-s_{00})+f(x_k-s_{10})+f(x_k-s_{01})+f(x_k-s_{11})}. \tag{3}$$

Jeśli przez p_i^{k+1} i p_j^k oznaczymy kolejno prawdopodobieństwa tego, że $a_{k+1}=i$ oraz tego, że $a_k=j$ to

$$\overline{P}_i^{k+1}=p_{i0}+p_{i1}, \tag{4}$$

$$P_j^k=p_{0j}+p_{1j}. \tag{5}$$

P_j^k są prawdopodobieństwami, że nadano $a_k=j$ przy odebranym sygnale x_k, natomiast $\overline{P}_i^{k+1}$ są to prawdopodobieństwa, z którymi przewiduje się że $a_{k+1}=i$.

Decyzję o tym jaką wartość ma a_{k+1} podejmuje się w następnym kroku na podstawie odbioru x_{k+1}. W aktualnym momencie k podejmowana jest decyzja o wartości a_k. Pierwsza proponowana reguła decyzyjna ma postać:

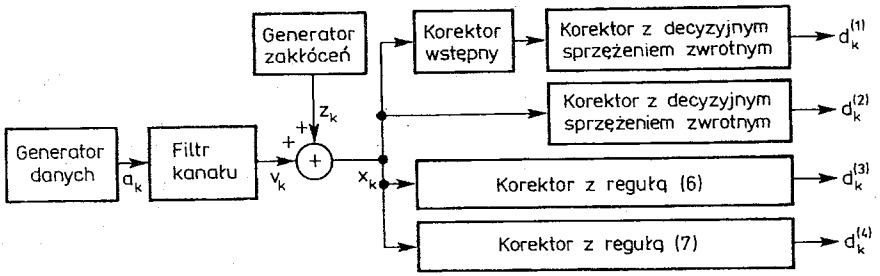
$$\bar{d}_k=\begin{cases}0 & \text{gdy } P_0^k\geqslant P_1^k\\ 1 & \text{gdy } P_0^k<P_1^k\end{cases}. \tag{6}$$

Zauważmy, że powyższa reguła decyzyjna nie wykorzystuje wszystkich dostępnych informacji. Podobnie jak tu na podstawie próbki x_k przewiduje się, że $a_{k+1}=i$ z prawdopodobieństwem $\overline{P}_i^{k+1}$, tak na podstawie poprzedniej próbki x_{k-1} przewidywano $a_k=i$ z prawdopodobieństwem $\overline{P}_i^k$. Reguła decyzyjna może wykorzystać również i te prawdopodobieństwa. Na podstawie próbki x_{k-1} w chwili poprzedniej wyznaczono w oparciu o wzory (4) i (5) prawdopodobieństwa $\overline{P}_i^k$ oraz P_j^{k-1}. Druga reguła decyzyjna porównuje sumy prawdopodobieństw wyznaczonych w dwu kolejnych krokach decyzyjnych. Decyzja d_k zostaje podjęta na posdtawie prawdopodobieństw tego, że $a_k=0$ oraz tego, że $a_k=1$ wyznaczonych w chwili $k-1$ oraz w chwili k. Tak więc

$$d_k=\begin{cases}0 & \text{gdy } P_0^k+\overline{P}_0^k\geqslant P_1^k+\overline{P}_1^k\\ 1 & \text{gdy } P_0^k+\overline{P}_0^k<P_1^k+\overline{P}_1^k\end{cases}. \tag{7}$$

EKSPERYMENT

W celu porównania zaprezentowanej tu metody korekcji przeprowadzono eksperyment obliczeniowy. Symulowano pracę systemu złożonego ze źródła danych, kanału z zakłóceniami i 4 korektorów w układzie pokazanym na rys. 3.



Rys. 3. 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, ponieważ spośród wielu rodzajów zakłóceń najczęściej jest rozważany biały szum gaussowski, choć najważniejszymi czynnikami powodującymi przekłamania są inne rodzaje zakłóceń, jak np. szum impulsowy i krótkie przerwy transmisji [6, 9]. Wyniki badań wskazują na to, że przy porównaniu dwóch systemów ten, który jest odporniejszy na szum gaussowski, jest na ogół również lepszy ze względu na inne rodzaje zakłóceń [6, 8]. 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 wejścia 4 korektorów z decyzyjnym sprzężeniem zwrotnym: z korektorem wstępnym, bez korektora wstępnego, z przewidywaniem przeszłej decyzji, z regułą decyzyjną (6) oraz z regułą decyzyjną (7). Dla różnych poziomów zakłóceń porównywano dane odbierane $d_k^{(i)}$ z danymi nadanymi a_k ($i=1,2,...,4$). Dla każdego korektora wyznaczono liczbę błędnie odebranych bitów i odniesiono tę liczbę do wszystkich odebranych bitów.

Eksperyment powtarzano zmieniając odstęp sygnału od szumu od 25 dB do 7 dB co 0.5 dB. W każdym eksperymencie generowano ciąg 65536 bitów.

Do eksperymentu wybrano kanał radiowy użyty przez A. Dąbrowskiego w [6]. Powodem wyboru tego kanału było stwierdzenie w [6], że w jego przypadku liniowa korekcja transwersalna jest mało skuteczna oraz, że kanały tego typu (posiadające zera w charakterystyce częstotliwościowej) powinny być korygowane metodami nieliniowymi. Ponieważ sygnał wyjściowy kanału zakłócano białym szumem gaussowskim, to prawdopodobieństwa p_{ij} obliczano ze wzoru (3) przyjmując

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

gdzie σ^2 reprezentuje moc szumu.

WYNIKI I WNIOSKI

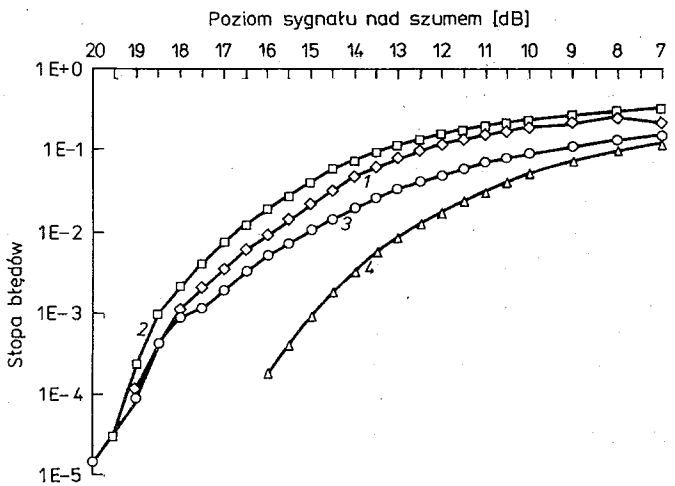
Próbki odpowiedzi kanałów rozważanych w [6] przedstawiono w tabeli 1.

Tabela 1

	g	Wartości próbek y_k	Wzór (2)
Kanał 1	2	0.407 0.817 0.407	$V_k=0.407a_{k+1}+0.817a_k$
Kanał 2	4	0.227 0.460 0.688 0.460 0.227	$V_k=0.277a_{k+1}+0.460a_k$

Na rysunkach 4 i 5 przedstawiono dla tych kanałów wykresy obliczonej elementowej stopy błędów (stosunek liczby błędnie odebranych bitów do liczby wszystkich odebranych bitów) w funkcji odstępu sygnału od szumu (stosunek mocy sygnału do mocy szumu). Dodatkowo rozważono kanał 2 pozostawiając nieskorygowany prążek $y_0=0.227$ w korektorze bez korekcji wstępnej i w korektorach realizujących algorytmy (6) i (7), opóźniając w nich moment podejmowania decyzji do chwili odbioru prążka o maksymalnej wartości $y_2=0.688$. Dla tego przypadku we wzorze (2) pojawi się dodatkowy pomijany w rozważaniach składnik związany z danymi a_{k+2}

$$V_k=0.227a_{k+2}+460a_{k+1}+0.668a_k \tag{9}$$



Rys. 4. Zależność stopy błędów od poziomu sygnału nad szumem dla kanału 1

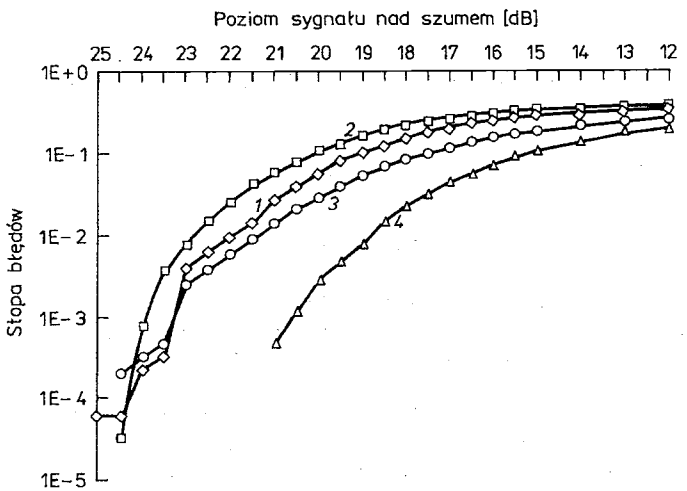
Na wszystkich wykresach poszczególne krzywe oznaczono numerami od 1 do 4 przyporządkowując je wynikom dla następujących korektorów z decyzyjnymi sprzężeniami zwrotnymi:

1. z korektorem wstępnym,
2. bez korektora wstępnego,

3. korektor z przewidywaniem decyzji algorytmem (6),
4. korektor z przewidywaniem decyzji algorytmem (7).

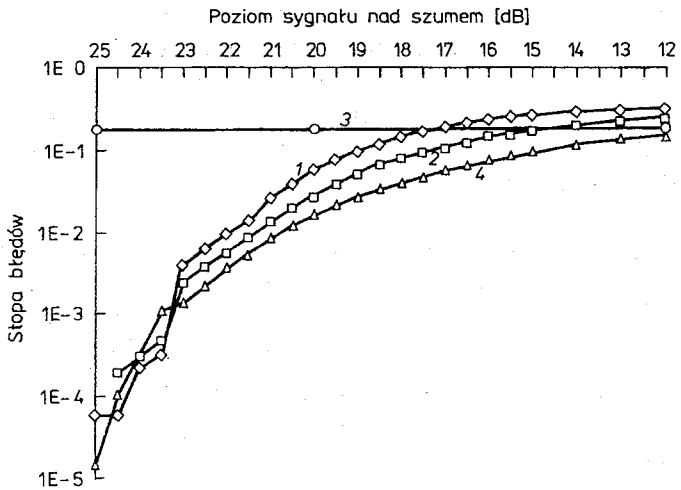
We wszystkich przypadkach zdecydowanie najlepszym okazał się korektor 4 z przewidywaniem decyzji algorytmem (7). Pozostałe korektory wymagają odstępu sygnału od szumu większego co najmniej o 2.5 dB aby osiągnąć stopę błędów dorównującą stopie uzyskiwanej przy użyciu korektora 4. Przy większych zakłóceniach ta różnica wzrasta powyżej 4.5 dB.

W pierwszych dwu przypadkach kolejnymi co do wierności decyzji były: korektor 2 z przewidywaniem decyzji algorytmem (6); korektor 1 — typowy korektor z decyzyjnym sprzężeniem zwrotnym i korektorem wstępnym, natomiast najgorszym był korektor 1 z decyzyjnym sprzężeniem zwrotnym bez korekcji wstępnej. Jak widać, korektory z przewidywaniem decyzji są tu wyraźniej lepsze.



Rys. 5. Zależność stopy błędów od poziomu sygnału nad szumem dla kanału 2

W trzecim przypadku (rys. 6) korektor bez filtru wstępnego i korektory z przewidywaniem decyzji nie korygowały pierwszego prążka interferencji o wartości $y_0 = 0.227$. Korektor 1 pracował jak poprzednio korygując całą interferencję. O ile poprzednio (rys. 5) korektor 2 podejmował decyzję w oparciu o $y_0 = 0.227$, to obecnie podejmuje ją na podstawie $y_1 = 0.460$, co jest bardziej bezpieczne (większy odstęp od zakłóceń, bo $y_1 > y_0$). Tak więc w tym przypadku korektor 2 okazał się trochę lepszy od korektora 1. Był jednak gorszy od korektora 4. Korektor 3 praktycznie nie spełniał swojej roli. Zauważmy z (9), że poziomy $s_{01} = y_2 = 0.680$ oraz $s_{10} = y_1 = 0.460$ zniekształcone inteferencją y_0 mogą nie być rozróżnialne, ponieważ $y_1 + y_0 \approx y_2$. Uwzględnienie w korektorze 4 prognozy a_n podczas odbioru poprzedniego bitu a_{n11} radykalnie poprawia działanie tego korektora. Stwarza to nadzieję na dobre efekty w studiach nad algorytmami korekcji, które będą uwzględniać i prognozować więcej przyszłych decyzji.



Rys. 6. Zależność stopy błędów od poziomu sygnału nad szumem dla kanału 2 bez korekcji prążka y_0

BIBLIOGRAFIA

1. A b r a m s o n, F.F. K u o (red.): *Sieci telekomunikacyjne komputerów.* WNT, Warszawa 1978
2. W.R. B e n n e t t, J.R. D a v e y: *Data Transmission.* McGraw-Hill, New York 1965
3. A.B. C a r l s o n: *Communication Systems.* McGraw-Hill, Tokyo 1975
4. A.P. C l a r k: *Advanced Data-Transmission Systems.* Pentech Press, London 1977
5. A.P. C l a r k: *Principles od Digital Data Transmissoin.* Pentech Press, London 1976
6. A. D ą b r o w s k i: *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. K i s i l e w i c z: *Prblemy poprawy wierności transmisji danych w sieciach komputerowych.* Prace Naukowe Instytutu Sterowania i Techniki Systemów Politechniki Wrocławskiej, Seria: Monografie nr 6, WPWr, Wrocław 1991
8. R.W. L u c k y, J. S a l z, E.J. W e l d o n: *Principles of Data Communiation.* McGraw-Hill, New York 1968
9. T. N o w i c k i, T. S t a c h o w i a k, W. S t a ń c z a k: *Wybrane zagadnienia sieci teleinformatycznych.* PWN, Warszawa 1985
10. W. S o b c z a k: *Wprowdzenie do teleinformatyki.* WKŁ, Warszawa 1979
11. W. S o b c z a k (red.): *Problemy teleinformatyki.* WKŁ, Warszawa 1984
12. M. S t e r n a d, A. A h l e n: *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
13. A.S. T a n e n b a u m: *Sieci komputerowe.* WNT, Warszawa 1988

J. KISILEWICZ

NONLINEAR EQUALIZATION OF INTERSYMBOL INTEREFERENCE WITH FURTHER DECISION PREDICTION

Summary

The method of particular intersymbol intereference equalization is described. The nonlinear equalizer with decision feedback but without linear feedforward transversal filter at the input is used. The intersymbol intereference component that is not eliminated by the nonlinear equalizer, is elininated by further decision prediction. Basing on received signal, the probabilities of further possible data bits incoming in the next sample moment are calculated. The data transmission systems with described equalizer, common equalizer with decision feedback and with linear feedforward transversal filter at the input were simulated. The error bit rate of these systems with presence of white Gaussian noise was calculated and compared.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1994. 40, z. 4. ss. 499–510

621.39
519.68

Algorytmy ważonych filtrów medianowych o strukturze wielostopniowej

EUGENIUSZ KORNATOWSKI, WIESŁAW CZERWIŃSKI

Instytut Elektroniki i Informatyki, Politechnika Szczecińska

Otrzymano 1994.10.15

Autoryzowano do druku 1994.

W pracy zaproponowano algorytmy wielostopniowych ważonych filtrów medianowych (WMMF). Dwa filtry tej klasy: centralnie ważone i warstwowo ważone dwukierunkowe, zostały przetestowane i porównane z jednokierunkowymi i dwukierunkowymi nieważonymi filtrami medianowymi. Porównania dokonano w sensie dwu wskaźników jakości: MSE (Mean Squared Error) i MAE (Mean Absolute Error).

1. WSTĘP

Zastosowanie filtrów nieliniowych dla potrzeb przetwarzania obrazów spowodowane zostało istotnymi ograniczeniami filtracji liniowej, szczególnie uciążliwymi w przypadku występowania impulsowych zakłóceń sygnału. Powstanie w wyniku tego licznych możliwości praktycznego wykorzystania algorytmów filtracji nieliniowej wywołało intensywne prace badawcze nad ich właściwościami, co z kolei pozwoliło na osiągnięcie na przestrzeni ostatnich lat znacznego postępu w tej dziedzinie, np. [1, 2, 11].

Wprawdzie już za pomocą prostych, jednostopniowych filtrów medianowych [6] można osiągnąć bardzo dobre wyniki w zakresie tłumienia szumu impulsowego, jednakże proces filtracji powoduje jednocześnie istotne zniekształcenia przetwarzanego sygnału. W przeciwieństwie do algorytmów liniowych, mamy tu do czynienia nie tyle z rozmazywaniem krawędzi, ile z usuwaniem drobnych elementów obrazu. W celu ograniczenia skutków tego niepożądanego efektu możliwe jest wprawdzie odpowiednie zmodyfikowanie kształtu użytej maski filtru. Wprowadzenie algorytmów medianowych filtrów ważonych pozwala na dalszą poprawę jakości przetwarzania, ale nie prowadzi to jednak do zasadniczej poprawy w tym względzie.

Istotny postęp na tym polu pozwala osiągnąć dopiero użycie algorytmów wielostopniowych [11] (multilevel filtering, multistage filtering). Oparte na tej zasadzie filtry wykazują znacznie korzystniejsze właściwości niż techniki jednostopniowe, głównie z uwagi na mniejsze zniekształcenia struktury obrazu. Na bazie algorytmów wielostopniowych powstało wiele różnorodnych filtrów nieliniowych, pozwalających na uzyskiwanie szczególnie dobrych rezultatów na danych obszarach zastosowań.

W niniejszym artykule zaproponowano algorytmy wielostopniowych, ważonych filtrów medianowych, a następnie przeprowadzone zostało ich porównanie z przedstawionym w [13] wielostopniowym filtrem medianowym (multistage median filters) oraz z zaprezentowanym w [11] wielostopniowym filtrem medianowo-liniowym (FIR-Median Hybrid). Wybór tych filtrów podyktowany był ich bardzo dobrymi właściwościami ze względu na tłumienie zakłóceń impulsowych oraz zachowanie struktury obrazu. Dodatkowo zamieszczone zostały również wyniki uzyskane dla klasycznych filtrów jednostopniowych z maską kwadratową (nieważonych).

W rozdziale 2 pokazany został sposób opisu apertur dla filtrów medianowych wzorowany na [13], po czym na tej bazie zdefiniowane zostały apertury dla podokien filtrów wielostopniowych i wielostopniowych ważonych.

Rozdział 3 zawiera algorytmy rozpatrywanych wielostopniowych, ważonych filtrów medianowych oraz użytego w celach porównawczych wielostopniowego filtru medianowego i wielostopniowego filtru medianowo-liniowego. Analiza właściwości proponowanych algorytmów z uwagi na zachowanie struktury obrazu zamieszczona została w rozdziale 4, natomiast zdolność tłumienia szumów przez prezentowane filtry zbadana została w rozdziale 5. Rozdział 6 zawiera podsumowanie uzyskanych wyników i wypływające z nich wnioski.

2. APERTURY PODOKIEN

Realizacja filtrów medianowych opiera się, podobnie jak w przypadku filtrów liniowych, na zasadzie tzw. ruchomej maski [6]. W metodzie tej maska o zadanej wielkosci jest przesuwana nad poszczególne elementy macierzy obrazu, poprzez co wybrane zostają argumenty dla obliczenia wartości mediany w kolejnych punktach. Poprzez podanie maski filtru zostają więc jednoznacznie określone punkty sąsiednie punktu (m, n), które zostaną uwzględnione przy obliczaniu wartości wyjściowej. Jeżeli przyjmiemy, że włączeniu danego punktu do obliczeń odpowiada wartość „$\neq 0$" a jego pominięciu wartość „0", to możliwe jest przedstawienie masek o różnych kształtach za pomocą macierzy dwuwymiarowych.

Niech $A_{(2N+1)\times(2N+1)}$ będzie macierzą maski filtru o elementach $a_{i,j}$, $i,j \in \{-N,...,0,...,N\}$. Wówczas wartość mediany $g'_{m,n}$ w punkcie $g_{m,n}$ dana jest równaniem [6]

$$g'_{m,n}=\operatorname{med}(g_{m+i,n+j};\ (i,j)\in W)=\underset{W}{\operatorname{med}}(g_{m,n}), \tag{1}$$

przy czym W oznacza zbiór wszystkich tych indeksów, które odpowiadają niezerowym elementom maski filtru

$$(i,j)\in W \Leftrightarrow a_{i,j}\neq 0 \qquad i,j\in\{-N,...,0,...,N\} \tag{2}$$

i nazywany jest aperturą.

W przypadku median ważonych konieczne jest pewne rozszerzenie powyższego modelu tak, aby uwzględnione zostały wagi poszczególnych pikseli. Cel ten można osiągnąć poprzez przyjęcie wartości elementów macierzy filtru $a_{i,j}$ równych wagom odpowiadających im pikseli. Wówczas w aperturze W poszczególne indeksy (i,j) zawarte są k-krotnie, przy czym k odpowiada wadze piksela $g_{m+i,n+j}$ i zachodzi $k=a_{i,j}$.

Poniżej zdefiniowane zostały na bazie przedstawionego modelu apertury podokien dla rozważanych filtrów w rozdziale 3. Przyjęto przy tym rozmiar okna równy $(2\cdot N+1)=5$, tzn. $i,j\in\langle -2,2\rangle$. Możliwe jest jednak rozszerzenie opisanych struktur dla dowolnie wybranego $N\in IN$.

Okna jednokierunkowe (unidirectoinal windows) dane są poprzez następujące rownania [13]:

$$W_1=\{(i,0):-2\leqslant i\leqslant 2\}, \tag{3}$$

$$W_2=\{(i,i):-2\leqslant i\leqslant 2\}, \tag{4}$$

$$W_3=\{(0,i):-2\leqslant i\leqslant 2\}, \tag{5}$$

$$W_4=\{(i,-i):-2\leqslant i\leqslant 2\}. \tag{6}$$

Z kolei na bazie struktur jednokierunkowych można przedstawić okna dwukierunkowe (bidirectonal windows) jako [13]:

$$W_{1,3}=W_1\cup W_3 \tag{7}$$

$$W_{2,4}=W_2\cup W_4. \tag{8}$$

Analogicznie określone są również apertury dla filrów ważonych. Dla okien dwukierunkowych i wagi punktu centralnego trzykrotnie większej od wag pozostałych punktów struktury otrzymujemy

$$W_{c1,3}=\left\{\begin{matrix}(0,0),(0,0),(0,0),\\(-2,0),(-1,0),(1,0),(2,0)(0,-2),(0,-1),(0,1),(0,2)\end{matrix}\right\} \tag{9}$$

$$W_{c2,4}=\left\{\begin{matrix}(0,0),(0,0),(0,0),\\(-2,-2),(-1,-1),(1,1),(2,2),(-2,2),(-1,1),(1-1),(2,-2)\end{matrix}\right\}. \tag{10}$$

Filtry oparte o apertury dane równaniami (9), (10) będą w poniżej nazywane filtrami centralnie ważonymi.

Przyjęcie zróżnicowanych wag poszczególnych punktów w zależności od ich odległości od punktu centralnego w stosunku 1:2:5 prowadzi do struktur wyrażonych równaniami

$$W_{s1,3}=\begin{Bmatrix}(0,0),(0,0),(0,0),(0,0),(0,0),\\(-1,0),(-1,0),(0,1),(0,1),(1,0),,(0,-1),(0,-1)\\(-2,0),(0,2),(2,0),(0,-2)\end{Bmatrix} \tag{11}$$

$$W_{s1,3}=\begin{Bmatrix}(0,0),(0,0),(0,0),(0,0),(0,0),\\(-1,-1),(-1,-1),(-1,1),(-1,1),(1,1),(1,1),(1,-1),(1,-1),\\(-2,-2),(-2,2),(2,2),(2,-2).\end{Bmatrix} \tag{12}$$

W dalszej części artykułu filtry wykorzystujące apertury wyrażone równaniami (11), (12) będą nazywane filtrami warstwowo ważonymi.

3. ALGORYTMY FILTRACJI WIELOSTOPNIOWEJ

Zasadnicza rożnica pomiędzy technikami jedno- i wielostopniowymi wynika już z odmiennego charakteru apertury filtru. Podczas gdy w przypadku algorytmów jednostopniowych maskę filtru zawsze tworzy jednolita struktura, to maski używane w filtracji wielostopniowej mogą być traktowane jako złożenie kilku oddzielnych podokien, stanowiących apertury poszczególnych filtrów bazowych.

Każdy z filtrów bazowych zapewnia tłumienie zakłóceń przy stosunkowo niewielkich zniekształceniach sygnału w jednym, zadanym kierunku. Mimo, iż teoretycznie możliwe jest zastosowanie dowolnie złożonych struktur w celu uzyskania dobrych właściwości algorytmu w zakresie zachowania struktury obrazu, to jednak użycie takich kompleksowych masek znacznie pogarsza charakterystyki filtru ze względu na przetwarzanie drobnych szczegółów obrazu. Z tego powodu w praktyczne znaczenie mają jedynie struktury jedno- i dwukierunkowe. W dalszej części przedstawione zostaną, oparte o powyższe struktury, wybrane algorytmy filtracji wielostopniowej.

Niech $\mathbf{G}_{Z\times S}$ będzie macierzą obrazu o elementach $g_{m,n}\in\mathbf{G}$. Ponadto niech dane będą następujące filtry bazowe [13]:

$$z_1(m,n)=\underset{W_1}{\operatorname{med}}(g_{m,n}); \qquad 1\leqslant 1\leqslant 4, \tag{12}$$

$$z_{1,3}(m,n)=\underset{W_{1,3}}{\operatorname{med}}(g_{m,n}), \tag{13}$$

$$z_{2,4}(m,n)=\underset{W_{2,4}}{\operatorname{med}}(g_{m,n}), \tag{14}$$

z aperturami W_1, W_2, W_3, W_4, $W_{1,3}$, $W_{2,4}$ określonymi w rozdziale 2.

Definicja 1: Wartość $g'_{m,n}$ na wyjściu jendokierunkowego, wielostopniowego filtru medianowego w punkcie $g_{m,n}$ dana jest równaniem [13]

$$g'_{m,n}=\text{med}(y_{1,3}(m,n),y_{2,4}(m,n),g_{m,n}), \tag{15}$$

przy czym

$$y_{1,3}(m,n)=\text{med}(z_1(m,n),z_3(m,n),g_{m,n}) \tag{16}$$

oraz

$$y_{2,4}(m,n)=\text{med}(z_2(m,n),z_4(m,n),g_{m,n}) \tag{17}$$

Definicja 2: Wartość $g'_{m,n}$ na wyjściu dwukierunkowego, wielostopniowego filtru medianowego w punkcie g_{mn} wyrażona jest wzorem [13]:

$$g'_{m,n}=\text{med}(z_{1,3}(m,n),z_{2,4}(m,n),g_{m,n}) \tag{18}$$

Na podstawie powyższych definicji można zauważyć, że obliczanie wartości wyjściowej dla filtrów jednokierunkowych odbywa się w trzech stopniach, natomist dla filtrów dwukierunkowych w dwóch kolejnych krokach. Z tego też względu określane są one odpowiednio jako trójstopniowy filtr medianowy (3-Level Median filter) oraz dwustopniowy filtr medianowy (2-Level FIR-Median hybrid filter) i oznaczane symbolami 3LM – i 2LM + [1].

W celu uzyskania lepszych właściwości algorytmów wielostopniowych w zakresie zachowania struktury obrazu przy jednoczesnej poprawie współczynnika tłumienia zakłóceń impulsowych sygnału, w dalszej cześci artykułu rozpatrzone zostało zastosowanie jako filtrów bazowych ważonych filtrów medianowych. Z wielu możliwych realizacji, do analizy opartych na tej zasadzie algorytmów wielostopniowych wybrano filtry centralnie i warstwowo ważone o aperturach przedstawionych w rozdziale 2.

Niech $\mathbf{G}_{Z\times S}$ będzie macierzą obrazu o elementach $g_{m,n}\in\mathbf{G}$. Wówczas filtry bazowe centralnie ważone dane są równaniami

$$z_{c1,3}(m,n)=\text{med}\begin{pmatrix} g_{m,n},g_{m,n},g_{m,n}, \\ g_{m-2,n},g_{m-1,n},g_{m+1,n},g_{m+2,n},g_{m,n-2},g_{m,n-1},g_{m,n+1},g_{m,n+2}\end{pmatrix}$$

$$=\underset{W_{c1,3}}{\text{med}}(g_{m,n}) \tag{19}$$

i

$$z_{c2,4}(m,n)=\text{med}\begin{pmatrix} g_{m,n},g_{m,n},g_{m,n}, \\ g_{m-2,n-2},g_{m-1,n-1},g_{m+1,n+1},g_{m+2,n+2},g_{m+1,n-1},g_{m-1,n+1},g_{m-1,n+2}\end{pmatrix}$$

$$=\underset{W_{c2,4}}{\text{med}}(g_{m,n}) \tag{20}$$

z aperturami $W_{c1,3}$, $W_{c2,4}$ określonymi są przez równania (9), (10).

[1] Znak '–' wskazuje tu na użycie struktur jednokierunkowych, natomiast znak '+' na zastosowanie struktur dwukierunkowych.

Podobnie filtry bazowe warstwowo ważone wyrażone są zależnościami

$$z_{s1,3}(m,n) = \operatorname{med}\begin{pmatrix} g_{m,n},g_{m,n},g_{m,n},g_{m,n},g_{m,n} \\ g_{m-1,n},g_{m-1,n},g_{m+1,n},g_{m+1,n},g_{m,n-1},g_{m,n-1},g_{m,n+1},g_{m,n+1} \\ g_{m-2,n},g_{m+2,n},g_{m,n-2},g_{m,n+2} \end{pmatrix} =$$

$$= \underset{W_{s1,3}}{med}(g_{m,n}) \tag{21}$$

oraz

$$z_{s24}(m,n) = \operatorname{med}\begin{pmatrix} g_{m,n},g_{m,n},g_{m,n},g_{m,n},g_{m,n}, \\ g_{m-1,n-1},g_{m-1,n-1},g_{m+1,n+1},g_{m+1,n+1},g_{m+1,n-1},g_{m+1,n-1},g_{m-1,n+1},g_{m-1,n+1} \\ g_{m-2,n-2},g_{m+2,n+2},g_{m+2,n-2},g_{m-2,n+2} \end{pmatrix} =$$

$$= \underset{W_{s2,4}}{med}(g_{m,n}), \tag{22}$$

przy czym apertury $W_{s1,3}$, $W_{s2,4}$ zdefiniowane są przez równania (11), (12).

Definicja 3: Wartość $g'_{m,n}$ na wyjściu dwukierunkowego, centralnie ważonego, wielostopniowego filtru medianowego w punkcie $g_{m,n}$ dana jest równaniem

$$g'_{m,n} = \operatorname{med}(z_{c1,3}(m,n),\ z_{c2,4}(m,n), g_{m,n}). \tag{23}$$

Definicja 4: Wartość $g'_{m,n}$ na wyjściu dwukierunkowego, warstwowo ważonego, wielostopniowego filtru medianowego w punkcie $g_{m,n}$ wyrażona jest wzorem

$$g'_{m,n} = \operatorname{med}(z_{s1,3}(m,n),\ z_{s2,4}(m,n), g_{m,n}). \tag{24}$$

Zarówno dla filtru centralnie jak i warstwowo ważonego obliczanie wartości wyjściowej następuje w dwóch krokach. Zgodnie więc z przyjętymi powyżej oznaczeniami, filtry te określane będą jako dwustopniowy, centralnie ważony filtr medianowy (2-Level Central-weighted Median filter) i dwustopniowy, warstwowo ważony filtr medianowy (2-Level Shell-weighted Median filter) i oznaczane odpowiednio symbolami 2LCM+ i 2LSM+[2].

Właściwości powyższych ważonych algorytmów wielostopniowych zostały zbadane w rozdziale 4 oraz 5. Jednocześnie porównano je z przedstawionymi w [13] nieważonymi filtrami medianowymi. Przeanalizowane zostało zachowanie się rozważanych filtrow z uwagi na zachowanie struktury obrazu oraz ze względu na tłumienie zakłóceń sygnału.

[2] Znak '+' oznacza wykoryzstanie apertur dwukierunkowych.

4. ZACHOWANIE STRUKTURY OBRAZU

Dla porównania właściwości poszczególnych filtrów konieczne jest zestawienie stopnia zakłóceń oraz zniekształceń obrazu przed i po filtracji. Obok subiektywnej, wizualnej oceny otrzymanego obrazu istotne jest również zestawienie odpowiednich obiektywnych wskaźników jakości danego algorytmu filtracji.

Znalezienie uniwersalnego kryterium, które możliwie dokładnie odzwierciedlałoby jakość obrazu, jest jednak niezmiernie trudne. Zdecydowanie najczęściej stosowane są w tym celu statystyczne metody oszacowań błędów na podstawie wartości średniego błędu bezwzględnego oraz średniego błędu kwadratowego, np. [13].

Niech $\mathbf{G}$, $\mathbf{G}'$ będą macierzami o rozmiarze $\mathrm{M}\times\mathrm{N}$, a $g_{i,j}\in\mathbf{G}$, $g'_{i,j}\in\mathbf{G}$, $i\in\{0,1,...,\mathrm{M}-1\}$, $j\in\{0,1,...,\mathrm{N}-1\}$ elementami tych macierzy. Ponadto niech $\mathbf{G}$ odpowiada obrazowi oryginalnemu, natomiast $\mathbf{G}'$ obrazowi po filtracji. Wówczas błąd bezwzględny wyrażony jest zależnością

$$\mathrm{mae}=\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}|\,g_{i,j}-g'_{i,j}\,|, \tag{25}$$

a błąd kwadratowy dany jest równaniem

$$\mathrm{mse}=\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(g_{i,j}-g'_{i,j})^2, \tag{26}$$

przy czym mae (*mean absolute error*) i mse (*mean square error*) oznaczają odpowiednio wartości błędu bezwzględnego i kwadratowego.

W celu maksymalnego uniezależnienia powyższych wskaźników od charakteru filtrowanego obrazu, ich wartości zostają zazwyczaj dodatkowo normalizowane. W wyniku tego wyrażenia (25), (26) zmieniają się do postaci

$$\mathrm{MAE}=\frac{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}|\,g_{i,j}-g'_{i,j}\,|}{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}g_{i,j}}, \tag{27}$$

$$\mathrm{MSE}=\frac{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(g_{i,j}-g'_{i,j})^2}{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}(g_{i,j})^2}. \tag{28}$$

MAE i MSE oznaczają wartości odpowiednio średniego błędu bezwzględnego i średniego błędu kwadratowego. Równania (27) i (28) podane są w tej formie również w [13] i będą stosowane dla matematycznej oceny jakości obrazu w dalszej części artykułu.

Na podstawie powyższych definicji nietrudno zauważyć, że równomiernie rozłożone zniekształcenia o niewielkich amplitudach silniej wpływają na średni błąd bezwzględny, podczas gdy zakłócenia o charakterze impulsowym decydują głównie o wartości średniego błędu kwadratowego. Poprzez porównanie wartości obydwu wskaźnikow można więc wnioskować nie tylko o poziomie, ale też o rodzaju zniekształceń.

Dla potrzeb analizy algorytmów filtracji zaproponowanych zostało wiele różnorodnych sygnałów testowych, które istotnie ułatwiają zbadanie poszczególnych właściwości. W niniejszym artykule jako obrazu testowego użyto odchylanego liniowo sygnału cyfrowego o modulacji częstotliwościowej. Jedną z pierwszych prac, w której użyto tego testu jest [10] i obecnie jest on szeroko stosowany szczególnie dla oceny właściwości filtracji ze względu na zachowanie struktury obrazu. W układzie współrzędnych biegunowych sygnał ten opisany jest zależnością

$$g(r)=\begin{cases} A\cdot\cos\left(2\cdot\pi\cdot\dfrac{f_0\cdot r^2}{R}\right) & \text{dla} \quad r\leqslant\dfrac{R}{2} \\ A\cdot\cos\left(2\cdot\pi\cdot\dfrac{f_0\cdot r^2-f_0\cdot(r-R)^2}{R}\right) & \text{dla} \quad \dfrac{R}{2}<r\leqslant\dfrac{3\cdot R}{2} \end{cases} \tag{29}$$

gdzie r jest promieniem mierzonym od środka obrazu, A określa amplitudę sygnału, natomiast R i f_0 oznaczają odpowiednio okres i maksymalną częstotliwość chwilową sygnału modulującego.

Wskaźniki jakości procesu filtracji dla czterech rozważanych filtrów wielostopnowych przy masce o rozmiarze 5×5, jak również dla jednostopniowego filtru medianowego z maską kwadratową wielkości 3×3, przedstawione są w tabeli 1.

Tabela 1

Wartości średniego błędu kwadratowego (MSE) oraz średniego błędu bezwzględnego (MAE) dla poddanego danym typom filtracji obrazu testowego.
Najwyższe i najniższe wartości zaznaczone są odpowiednio przez (*) i (▽). Skrót 1LM■ odnosi się do jednostopniowego filtru medianowego (1-level median filter) o masce kwadratowej

Filtr	N	Rozmiar okna	Wskaźniki błędów	
			MSE	MAE
1LM■	1	3	2,3038▽	0,1236▽
3LM–	2	5	0,0467*	0,0056*
2LM+	2	5	0,8321	0,0612
2LCM+	2	5	0,2204	0,0235
2LSM+	2	5	0,1546	0,0159

Zarówno wizualna ocena otrzymanych w wyniku filtracji obrazów, jak również porównanie matematycznych wskaźników ich jakości, wykazuje znakomite zachowanie algorytmów trójstopniowych, które w zakresie zachowania struktury obrazu posiadają znacznie lepsze właściwości niż techniki dwustop-

niowe. Jednocześnie wyraźnie zauważalna jest poprawa jakości przetwarzanych obrazów w wyniku zastosowania algorytmów ważonych. Znajduje ona odzwierciedlenie również w znacznie mniejszych wskaźnikach błędów, które w przypadku filtrów ważonych są blisko czterokrotnie niższe niż dla ich nieważonych odpowiedników i osiągają wartości porównywalne z otrzymanymi przy użyciu technik trójstopniowych.

Zgodnie z oczekiwaniami, zdecydowanie najgorsze rezultaty uzyskane zostały dla jednostopniowego filtru medianowego, dla którego zniekształcenia wywołane w wyniku procesu filtracji powodują znaczne obniżenie jakości obrazu. Również wskaźniki błędów są w tym przypadku o rząd wielkości większe niż dla algorytmów wielostopniowych.

5. TŁUMIENIE SZUMÓW

Obok znajomości zachowania filtru dla obrazów testowych, istotna jest również analiza właściwości poszczególnych algorytmów na bazie obrazów rzeczywistych. Szczególnie badanie współczynnika tłumienia szumów przy wykorzystaniu występujących w praktyce obrazów jest dużo bardziej miarodajne niż z zastosowaniem sztucznie stworzonych obrazów testowych. Poniżej przedstawione zostały wyniki filtracji typowego zdjęcia pejzażowego zakłóconego szumem o rozkładzie normalnym oraz impulsowym, uzyskane za pomocą rozpatrywanych algorytmów.

Zdolność eliminacji zakłóceń o charakterze impulsowym stanowi jedną z głównych zalet filtracji medianowej w stosunku do technik liniowych i decyduje o możliwościach zastosowań danego algorytmu. Z tego też względu dobre właściwości filtru na tym polu są szczególnie pożądane. Istotną rolę odgrywa w tym miejscu przede wszystkim stopień zaszumienia obrazu, w znaczący sposób wpływający na wynik filtracji.

W niniejszym artykule rozpatrzono kolejno przypadki małego, średniego i dużego stopnia zaszumienia addytywnym szumem impulsowym. Przyjęto

Tabela 2

Wartości średniego błędu kwadratowego oraz średniego błędu bezwzględnego dla zakłóconego szumem impulsowym obrazu rzeczywistego poddanego filtracji za pomocą różnych algorytmów. Stopień zaszumienia sygnału określony jest poprzez prawdopodobieństwo p_0 wystąpienia próbki zakłóconej (patrz też tabela 1 dla objaśnienia użytych oznaczeń)

Filtr	Wskaźniki błędów					
	$p_0=0{,}05$		$p_0=0{,}1$		$p_0=0{,}3$	
	MSE	MAE	MSE	MAE	MSE	MAE
1LM■	0,0402▽	0,0417▽	0,0450	0,0437▽	0,1221*	0,0632
3LM –	0,0255	0,0153*	0,0598▽	0,0244	0,6001▽	0,1321▽
2LM +	0,0307	0,0309	0,0350	0,0328	0,1271	0,0552*
2LCM +	0,0202	0,0185	0,0337*	0,0230	0,3581	0,0905
2LSM +	0,0197*	0,0169	0,0405	0,0228*	0,4636	0,1084

amplitudę zakłóceń równą 90% amplitudy sygnału i prawdopodobieństwo wystąpienia próbki zakłóconej równe odpowiednio $p_0=0.05$, $p_0=0.1$ oraz $p_0=0.3$.

Wartości wskaźników błędów, będące wynikiem eksperymentu zebrane są w tabeli 2.

Porównanie otrzymanych statystycznych wartości błędów, jak również wizualna ocena jakości przefiltrowanych obrazów, wykazuje bardzo dobre zachowanie filtrów ważonych w szerokim zakresie stopnia zaszumienia sygnału. Jedynie w przypadku występowania szczególnie silnych zakłóceń, lepsze rezultaty osiągnięto przy użyciu dwustopniowego filtru nieważonego czy też klasycznego filtru jednostopniowego. Przede wszystkim jednak, przy stosunkowo często występujących średnich stopniach zaszumienia obrazu, zastosowanie algorytmów ważonych prowadzi do zauważalnej poprawy wyników w porównaniu z rezultatami uzyskanymi na bazie metod nieważonych. Dopiero dla bardzo słabych zakłóceń podobne efekty dają także filtry trójstopniowe.

Również w przypadku szumów o rozkładzie normalnym filtracja na bazie ważonych algorytmów medianowych pozwala na poprawę jakości przetworzonych obrazów, aczkolwiek w nieco mniejszym stopniu niż dla zakłóceń o charakterze impulsowym. Ze względu jednak na węższy zakres zastosowań filtrów medianowych dla tłumienia tego typu szumów, ich właściwości na tym polu nie są już tak istotne.

Wskaźniki jakości filtrów uzyskane przy pomocy poszczególnych algorytmów dla zakłóconego szumem o rozkładzie normalnym obrazu rzeczywistego zestawione zostały w tabeli 3. Addytywny szum zakłócający posiadał wariancję var=140 i wartość oczekiwaną $\mu=0$ przy 256 poziomach szarości testowanego obrazu.

Tabela 3

Wartości średniego błędu kwadratowego oraz średniego błędu bezwzględnego dla zakłóconego szumem Gaussa obrazu rzeczywistego poddanego filtracji przy pomocy różnych algorytmów (patrz też tabela 1 dla objaśnienia użytych oznaczeń)

Filtr	Wskaźniki błędów	
	MSE	MAE
1LM■	0,0466▽	0,0607
3LM−	0,0361	0,0645▽
2LM+	0,0384	0,0560*
2LCM+	0,0342*	0,0600
2LSM+	0,0343	0,0613

6. PODSUMOWANIE

Mimo szybkiego postępu, jaki dokonał się na polu cyfrowego przetwarzania obrazów w ciągu ostatnich lat, kwestia doboru algorytmów zapewniających dobre właściwości filtru w zakresie tłumienia szumów przy możliwie niewielkich zniekształceniach sygnału pozostaje wciąż otwarta. Wprawdzie wykorzystanie algorytmów wielostopniowych pozwoliło osiągnąć znacznie lepsze rezultaty niż miało to miejsce w przypadku filtrów jednostopniowych, jednak dla wielu zastosowań poprawa ta jest nadal niewystarczająca.

W niniejszym artykule zaproponowano klasę wielostopniowych algorytmów ważonych i zbadano charakterystyki dwóch opartych na nich filtrów. Przeprowadzone porównanie z filtrami nieważonymi wykazało bardzo dobre właściwości technik ważonych (w sensie wyrażonych liczbowo wskaźników jakości, jak i wrażeń subiektywnych) ze względu na zachowanie struktury obrazu, a takze z uwagi na tłumienie szumów. W szczególności przy średnich poziomach zakłóceń impulsowych wykorzystanie filtrów ważonych prowadzi do istotnej poprawy jakości obrazu i jedynie w ekstremalnych przypadkach korzystniejsze jest stosowanie filtrów nieważonych.

Oprócz przedstawionych powyżej filtrów centralnie i warstwowo ważonych, możliwe jest zaprojektowanie na bazie wielostopniowych algorytmów ważonych całej gamy zmodyfikowanych filtrów tej klasy. Odpowiedni dobór wag pikseli maski wpływa przy tym bezpośrednio na właściwości filtru, co pozwala na dopasowanie danego algorytmu do wymagań charakterystycznych dla poszczególnych zastosowań.

BIBLIOGRAFIA

1. B. Jähne: *Digitale Bildverarbeitung.* 2. Auflage, Springer 1991
2. P. Haberäcker: *Digitale Bildverarbeitung: Grundlagen und Anwendungen.* 4 Auflage Hanser 1991
3. J.S. Lim: *Two-dimensional signal and image processing.* Prentice-Hall 1990
4. R.C. Gonzales, P. Wintz: *Digital image processing.* Addison-Wesley 1977
5. M.P. Ekstrom: *Digital image processing techniques.* Academic Press 1984
6. T.S. Huang: *Two-dimensional digital signal processing II: transforms and median filters.* Springer 1981
7. E.R. Dougherty, C.R. Giardina: *Image processing — continuous to discrete.* Vol. 1, Prentice-Hall 1987
8. N.C. Gallagher, G.L. Wise: *Passband and stepband properties of median filters.* IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-29, Dec. 1981
9. T.A. Nodes, N.C. Gallagher: *Median filters: some modifications and their properties.* IEEE Trans. Acoust. Speech, signal Processing, vol. ASSP-30, Oct. 1982
10. T. Thong: *Digital image processing test patterns.* IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-31, Jun. 1983
11. A. Nieminen, P. Heinonen, Y. Nuevo: *A new class of detail preserving filters for image processing.* IEEE Trans. Pattern Anal. Machine Intell, vol. PAMI-9, Jan 1987

12. G.E. A r c e, M.P. McL o u g h l i n: *Theoretical analysis of the max/median filter.* IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-35, Jan. 1987
13. G.E. A r c e, R.E. F o s t e r: *Detail-preserving ranked-order based filters for imge processing.* IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-37, Jan. 1989
14. Y-H. L e e, S. T a n t a r a t a n a: *Decision-based order statistic filters.* IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-38, Mar. 1990
15. R. W i c h m a n, J.T. A s t o l a, P.J. H e i n o n e n, Y.A. N e u v o: *FIR-median hybrid filters with excellent transient response in noisy conditions.* IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-38, Dec. 1990

E. KORNATOWSKI

THE ALGORITHMS OF WEIGHTED MULTILEVEL MEDIAN FILTERS

S u m m a r y

In this paper we introduce a new class of multilevel median filters which we call weighted multilevel median filters (WMMF). Two filters of this class, the central weigthed and shell weighted bidirectional median filter were examined and compared to unidirectional and bidirectoinal non-weighted median filters. The comparisons were made using the mean squared error (MSE) and the measn absolute error (MAE) criteria.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1994. 40, z. 4. ss. 511 – 528

621.372.54
621.375

Warianty włączenia wzmacniacza operacyjnego do układu pasywnego

ZYGMUNT WRÓBEL

Zakład Metrologii, Instytut Problemów Techniki, Uniwersytet Śląski

Otrzymano 1994.09.28

Autoryzowano do druku 1994.11.10

W pracy przedstawiono możliwe sposoby włączenia idealnego wzmacniacza operacyjnego do układów czwórników i trójników pasywnych, oraz związane z nimi ograniczenia w realizacji poszczególnych parametrów układu.

Poszczególne parametry z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij} czwórników i trójników zawierających idealny wzmacniacz operacyjny, można wyznaczyć poprzez sumę „cząstkowych" dopełnień algebraicznych z pełnej macierzy admitancji węzłowych $\mathbf{Y}_w^p$ charakteryzującej tylko pasywną część układu zawierającego idealny wzmacniacz operacyjny.

1. WPROWADZENIE

Wzmacniacz operacyjny jest obecnie najczęściej i najchętniej stosowanym elementem czynnym w analogowych układach elektronicznych [6]. Parametry scalonego rzeczywistego wzmacniacza operacyjnego, można zamodelować w postać różnorodnych schematów zastępczych uwzględniających poszczególne jego nieidealne własności [1].

W wielu zastosowaniach a w szczególności w procesie wstępnej analizy układów zawierających wzmacniacz operacyjny wprowadza się pojęcie: idealnego wzmacniacza operacyjnego. Z punktu widzenia teorii obwodów idealny wzmacniacz operacyjny to układ, który w zasadniczy sposób decyduje o immitancjach i transmitancjach układu elektronicznego, mimo iż jego parametry małosygnałowe nie występują w tych wyrażeniach [7, 8]. Wydaje się, że dla potrzeb syntezy struktur kanonicznych układów aktywnych, idealny wzmacniacz operacyjny o dwóch wejściach i jednym wyjściu można zamodelować w postaci nulatora i noratora [3, 4].

W pracach [7, 8] wykazano, że układ aktywny zawierający idealny wzmacniacz operacyjny można traktować jako układ pasywny z ograniczeniami

związanymi z włączeniem nolatora i noratora modelujących idealny wzmacniacz operacyjny.

W niniejszej pracy dla rozróżnienia układy pasywnych i aktywnych ich parametry będziemy oznaczać odpowiednio indeksami p i a. Jak wynika z pracy [11, 12] dany parametr F_{ij} ($F_{ij} \in \{z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij}\}$) charakteryzujący czwórniki i trójniki pasywne można w ogólnym przypadku zapisać w postaci funkcji wymiernej typu:

$$F(t_1,t_2)=\frac{H^v_{ka}(t_1,t_2)}{H^v_{kb}(t_1,t_2)}, \tag{1}$$

gdzie:

$H^v_{ka}(t_1,t_2)$, $H^v_{kb}(t_1,t_2)$ — odpowiednio sumy iloczynów admitancji elementów typu $\{t_1\}$ i $\{t_2\}$ tworzących dendryty ka-drzewowe $T^v_{ka}(t_1,t_2)$ w liczniku i dendryty kb-drzewowe $T^v_{kb}(t_1,t_2)$ w mianowniku funkcji wymiernej $F(t_1,t_2)$ opisującej parametry układu.

Parametry czwórników i trójników można zapisać w postaci sześciu układów równań, które zestawiono w tabeli 1.

Tabela 1

Układy równań określające parametry $F_{ij} \in \{z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij}\}$ czwórników i trójników

F	Układy równań	F_{11}	F_{12}	F_{21}	F_{22}
z	$U_1=z_{11}\cdot I_1+z_{12}\cdot I_2$ $U_2=z_{21}\cdot I_1+z_{22}\cdot I_2$	$\left.\frac{U_1}{I_1}\right\|_{I_2=0}$	$\left.\frac{U_1}{I_2}\right\|_{I_1=0}$	$\left.\frac{U_2}{I_1}\right\|_{I_2=0}$	$\left.\frac{U_2}{I_2}\right\|_{I_1=0}$
y	$I_1=y_{11}\cdot U_1+y_{12}\cdot U_2$ $I_2=y_{21}\cdot U_1+y_{22}\cdot U_2$	$\left.\frac{I_1}{U_1}\right\|_{U_2=0}$	$\left.\frac{I_1}{U_2}\right\|_{U_2=0}$	$\left.\frac{I_2}{U_1}\right\|_{U_2=0}$	$\left.\frac{I_2}{U_2}\right\|_{U_1=0}$
a	$U_1=a_{11}\cdot U_2-a_{12}\cdot I_2$ $I_1=a_{21}\cdot U_2-a_{22}\cdot I_2$	$\left.\frac{U_1}{U_2}\right\|_{I_2=0}$	$\left.\frac{U_1}{I_2}\right\|_{U_2=0}$	$\left.\frac{I_1}{U_2}\right\|_{I_2=0}$	$-\left.\frac{I_1}{I_2}\right\|_{U_2=0}$
b	$U_2=b_{11}\cdot U_1-b_{12}\cdot I_1$ $I_2=b_{21}\cdot U_1-b_{22}\cdot I_1$	$\left.\frac{U_2}{U_1}\right\|_{I_1=0}$	$-\left.\frac{U_2}{I_1}\right\|_{U_1=0}$	$\left.\frac{I_2}{U_1}\right\|_{I_1=0}$	$-\left.\frac{I_2}{I_1}\right\|_{U_1=0}$
h	$U_1=h_{11}\cdot I_1+h_{12}\cdot U_2$ $I_2=h_{21}\cdot I_1+h_{22}\cdot U_2$	$\left.\frac{U_1}{I_1}\right\|_{U_2=0}$	$\left.\frac{U_1}{U_2}\right\|_{I_1=0}$	$\left.\frac{I_2}{I_1}\right\|_{U_2=0}$	$\left.\frac{I_2}{U_2}\right\|_{I_1=0}$
g	$I_1=g_{11}\cdot U_1+g_{12}\cdot I_2$ $U_2=g_{21}\cdot U_1+g_{22}\cdot I_2$	$\left.\frac{I_1}{U_1}\right\|_{I_2=0}$	$\left.\frac{I_1}{I_2}\right\|_{U_2=0}$	$\left.\frac{U_2}{U_1}\right\|_{I_2=0}$	$\left.\frac{U_2}{I_2}\right\|_{U_1=0}$

Jak wynika z rozważań pracy [12] poszczególne parametry F_{ij} czwórników i trójników pasywnych można wyrazić poprzez odpowiednie dopełnienia algebraiczne Δ^p macierzy admitancji węzłowych $\mathbf{Y}^p_w$ układow, lub poprzez odpowiednie sumy H^v_k iloczynów admitancji elementów tworzących dendryty

k-drzewowe T_k^v, co ilstruje tabela 2. Celem niniejszej pracy jest przedstawienie możliwych wariantów włączenia wzmacniacz operacyjngo do struktur czwórników i trójników pasywnych, oraz związane z nimi ograniczenia w realizacji poszczególnych parametrów układu. W niniejszej pracy wykorzystano pewne pojęcia zdefiniowane w pracach [11, 12].

Tabela 2

Parametry $F_{ij} \in \{z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij}\}$ charakteryzujce własnosci czwórników i trójników pasywnych

F	F_{11}	F_{12}	F_{21}	F_{22}
z	$\frac{P(2)}{P(1)}$	$\frac{P(3)}{P(1)}$	$\frac{P(4)}{P(1)}$	$\frac{P(5)}{P(1)}$
y	$\frac{P(5)}{P(6)}$	$\frac{P(3)}{P(6)}$	$\frac{P(4)}{P(6)}$	$\frac{P(2)}{P(6)}$
a	$\frac{P(2)}{P(4)}$	$\frac{P(6)}{P(4)}$	$\frac{P(1)}{P(4)}$	$\frac{P(5)}{P(4)}$
b	$\frac{P(5)}{P(3)}$	$\frac{P(6)}{P(3)}$	$\frac{P(1)}{P(3)}$	$\frac{P(2)}{P(3)}$
h	$\frac{P(6)}{P(5)}$	$\frac{P(3)}{P(5)}$	$\frac{P(4)}{P(5)}$	$\frac{P(1)}{P(5)}$
g	$\frac{P(1)}{P(2)}$	$\frac{P(3)}{P(2)}$	$\frac{P(4)}{P(2)}$	$\frac{P(6)}{P(2)}$

Wyrażone poprzez wyznaczniki Δ^p macierzy admitancji węzłowych $\mathbf{Y}_w^p$ i sumy H_k^v dendrytów k-drzewowych, T_k^v — gdzie:

Dla czwórników:

$P(1) = \Delta^p = H^v$

$P(2) = \Delta^p_{11} = H^v_{1,0}$

$P(3) = \Delta^p_{(3+2)1} = H^v_{12,30} - H^v_{13,20}$

$P(4) = \Delta^p_{1(3+2)} = H^v_{13,20} - H^v_{12,30}$

$P(5) = \Delta^p_{(3+2)(3+2)} = H^v_{2,3}$

$P(6) = \Delta^p_{11,(3+2)(3+2)} = H^v_{12,30} + H^v_{1,2,30} + H^v_{13,2,0} + H^v_{1,3,20}$

Dla trójników:

$P(1) = \Delta^p = H^v$

$P(2) = \Delta^p_{11} = H^v_{1,0}$

$P(3) = \Delta^p_{21} = H^v_{21,0}$

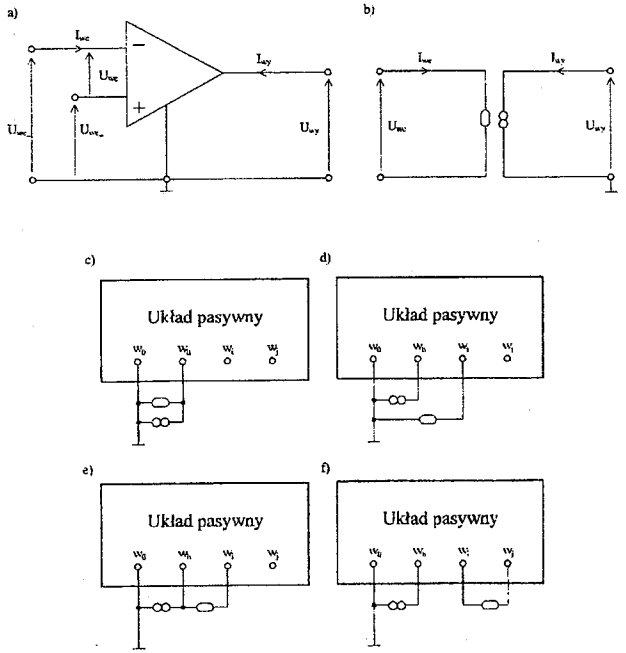
$P(4) = \Delta^p_{12} = H^v_{12,0}$

$P(5) = \Delta^p_{22} = H^v_{2,0}$

$P(6) = {}^p_{11,22} = H^v_{1,2,0}$

2. SPOSOBY WŁĄCZENIA IDEALNEGO WZMACNIACZA OPERACYJNEGO DO UKŁADU PASYWNEGO

Reprezentujący idealny wzmacniacz operacyjny nulator i norator można teoretycznie włączyć do układu pasywnego na cztery (Rys. 1) różne sposoby:



Rys. 1. Idealny wzmacniacz operacyjny, a) schemat ideowy, b) model nulatorowo-noratorowy, c, d, e, f) warianty włączenia idealnego wzmacniacza operacyjnego do układu pasywnego

1. Nulator i norator są włączone równolegle między wierzchołkiem $\{w_h\}$ a wierzchołkiem odniesienia którym zawsze będziemy przyjmować $\{w_0\}$ (Rys. 1c). Takie włącznie idealnego wzmacniacza operacyjnego do układu pasywnego prowadzi do wykreślenia h-tej kolumny i h-tego wiersza z macierzy admitancji węzłowych $\mathbf{Y}_w^p$ układu pasywnego, co odpowiada zwarciu $\{w_h\}$ wierzchołka układu z wierzchołkiem odniesienia $\{w_0\}$. Niżej ten rodzaj włączenia idealnego wzmacniacza operacyjnego nie będzie rozpatrywany.

2. Nulator włączony między wierzchołkiem $\{w_i\}$ a wierzchołkiem odniesienia $\{w_0\}$. Norator włączony między wierzchołkiem $\{w_h\}$ a wierzchołkiem odniesienia $\{w_0\}$ (Rys. 1d).

Takie włączenie idealnego wzmacniacza operacyjnego do układu pasywnego prowadzi do wykreślenia i-tej kolumny i h-tego wiersza z macierzy admitancji węzłowych $\mathbf{Y}_w^p$ układu pasywnego.

Dopełnienie algebraiczne $\triangle^a$ macierzy admitancji węzłowych $\mathbf{Y}_w^a$ układu aktywnego (z włączonym idealnym wzmacniaczem operacyjnym) ma postać:

$$\triangle^a = \triangle^p_{hi} = H^v_{hi,0}, \tag{2}$$

gdzie:

$\triangle^p_{hi}$ — dopełnienie algebraiczne admitancji węzłowych $\mathbf{Y}^p_w$ części pasywnej układu.

$H^v_{hi,0}$ — suma iloczynów admitancji poszczególnych elementów układu pasywnego tworzących dendryty 2-drzewowe $T^v_{hi,0}$.

Zbiór dendrytów 2-drzewowych typu $T^v_{hi,0}$ tworzy się jako przecięcie zbioru dendrytów 2-drzewowych typu $T^v_{h,0}$ i zbioru dendrytów 2-drzewowych typu $T^v_{i,0}$. Dendryty 2-drzewowe $T^v_{h,0}$ nie zawierają elementów włączonych między wierzchołkami $\{w_h, w_0\}$, natomiast dendryty 2-drzewowe $T^v_{i,0}$ nie zawierają elementów włączonych między wierzchołkami $\{w_i, w_0\}$. Innymi słowy elementy włączone między wierzchołkami $\{w_h, w_0\}$ (równoległe do noratora) i wierzchołkami $\{w_i, w_0\}$ (równolegle do nulatora), nie będą uczestniczyć w funkcji wymiernej F_{ij} opisującej poszczególne parametry układu.

3. Nulator włączony między zbiorem wierzchołków $\{w_i, w_h\}$. Norator włączony między wierzchołkiem $\{w_h\}$ a wierzchołkiem odniesienia $\{w_0\}$ (Rys. 1e).

Takie włączenie idealnego wzmacniacza operacyjnego do układu pasywnego prowadzi do sumowania elementów i-tej i h-tej kolumny i wykreślenia i-tego wiersza z macierzy $\mathbf{Y}^p_w$. Dopełnienie algebraiczne macierzy admitancji węzłowych $\mathbf{Y}^a_w$ układu aktywnego ma postać:

$$\triangle^1 = \triangle^p_{h(h+i)}. \tag{3}$$

Indeks $(h+i)$ oznacza, że elementy kolumny „h" są sumowane z elementami kolumny „i", a rezultat sumowania zapisywany jest w miejsce kolumny „h". Wiadomo [10], że:

$$\triangle_{a(b+c)} = -\triangle_{a(c+b)} \tag{4}$$

co oznacza, że przy sumowaniu kolumn znak dopełnienie algebraiczne zmienia się na przeciwny, jeśli wynik sumowania nie umieścimy w kolumnie „b" a w kolumnie „c". Ponieważ funkcja F_{ij} charakteryzująca parametry układu jest ilorazem sum dendrytów k-drzewowych H^v_{ka} i H^v_{kb} łatwo można wykazać, że zmiana położenia rezultatów sumowania kolumn nie wpływa na funkcję F_{ij}. (Znak licznika i mianownika funkcji F_{ij} zmieniają się jednocześnie).

Dopełnienie algebraiczne macierzy admitancji węzłowych układu pasywnego z włączonym wzmacniaczem operacyjnym ma postać:

$$\triangle^a = \triangle^p_{h(h+i)} = \triangle^p_{hi} - \triangle^p_{hh} \tag{5}$$

Jak wynika z rozważań pracy [9]:

$$\triangle^p_{hi} = H^v_{hi,0} \tag{6}$$

$$\triangle^p_{hh} = H^v_{h,0} \tag{7}$$

stąd:

$$\triangle^{p}_{h(h+i)} = H^{v}_{hi,0} - H^{v}_{h,0}. \tag{8}$$

Ponieważ jak pokazano w pracy [9]:

$$H^{v}_{h,0} = H^{v}_{hi,0} + H^{v}_{h,i0} \tag{9}$$

stąd:

$$\triangle^{p}_{h(h+i)} = -H^{v}_{h,i0}. \tag{10}$$

Zbiór dendrytów 2-drzewowych typu $T^{v}_{h,0}$ tworzy się jako przecięcie zbioru dendrytów 2-drzewowych typu $T^{v}_{h,0}$ i zbioru dendrytów 2-drzewowych typu $T^{v}_{h,i}$. Dendryty 2-drzewowe $T^{v}_{h,i}$ nie zawierają elementów włączonych między zbiorami wierzchołkami $\{w_i, w_h\}$, natomiast dendryty 2-drzewowe $T^{v}_{h,0}$ nie zawierają elementów włączonych między zbiorami wierzchołkami $\{w_h, w_0\}$. Innymi słowy elementy włączone między wierzchołkami $\{w_h, w_0\}$ (równolegle do noratora) i wierzchołkami $\{w_i, w_h\}$ (równolegle do nulatora), nie będą uczestniczyć w funkcji wymiernej F_{ij} opisującej poszczególne parametry układu.

4. Nulator włączony między zbiorem wierzchołków $\{w_i, w_j\}$. Norator włączony między wierzchołkiem $\{w_h\}$ a wierzchołkiem odniesienia $\{w_0\}$ (Rys. 1f).

Takie włączenie idealnego wzmacniacza operacyjnego do układu pasywnego prowadzi do sumowania elementów i-tej i j-tej kolumny i wykreślenia h-tego wiersza z macierzy $\mathbf{Y}^{p}_{w}$. Dopełnienie algebraiczne macierzy $\mathbf{Y}^{a}_{w}$ układu pasywnego z włączonym idealnym wzmacniaczem operacyjnym ma postać:

$$\triangle^{a} = \triangle^{p}_{h(h+j)} = \triangle^{p}_{hj} - \triangle^{p}_{hi} = H^{v}_{hj,0} - H^{v}_{hi,0}. \tag{11}$$

Ponieważ:

$$H^{v}_{hj,0} = H^{v}_{hij,0} + H^{v}_{hj,i0} \tag{12}$$

$$H^{v}_{hi,0} = H^{v}_{hij,0} + H^{v}_{hi,j0}, \tag{13}$$

stąd:

$$\triangle^{p}_{h(i+j)} = H^{v}_{hj,i0} + H^{v}_{hi,j0}. \tag{14}$$

Z powyższych zależności wynika, że w formułowaniu macierzy $\mathbf{Y}^{a}_{w}$ uczestniczą tylko zbiory dendrytów 2-drzewowych typu $T^{v}_{hj,i0}$ i $T^{v}_{hi,j0}$. W obu przypadkach wirzchołki $\{w_h, w_0\}$ i $\{w_i, w_j\}$ znajdują się w różnych częściach układu, a tym samym elementy włączone między tymi wierzchołkami równolegle do noratora i nulatora, nie będą uczestniczyć w tworzeniu funkcji wymiernej F_{ij} opisującej poszczególne parametry układu.

W praktyce są wykorzystywane dwa sposoby włączenia wzmacniacza operacyjnego do układu pasywnego.

1. W układzie z uziemionym nulatorem (Rys. 1d).
2. W układzie z nieuziemionym nulatorem (Rys. 1f).

Włączenie wzmacniacza operacyjnego według rysunku 1e jest przypadkiem szczególnym włączenia wzmacniacza operacyjnego z rysunku 1f.

Z powyższych rozważań dotyczących włączenia idealnego wzmacniacza operacyjnego do układu pasywnego wynika, że w ogólnym przypadku poszczególne parametry z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij} charakteryzujące własnosci układów aktywnych zawierających jeden wzmacniacz operacyjny można wyrazić poprzez odpowiednie dopełnienia algebraiczne macierzy admitancji węzłowych $\mathbf{Y}_w^p$ części

Tabela 3

Parametry $F_{ij} \in \{z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij}\}$ charakteryzujce własności czwórników i trójników zawierających wzmacniacz operacyjny włączony według pierwszego sposobu

F	F_{11}	F_{12}	F_{21}	F_{22}
z	$\frac{P(2)}{P(1)}$	$\frac{P(3)}{P(1)}$	$\frac{P(4)}{P(1)}$	$\frac{P(5)}{P(1)}$
y	$\frac{P(5)}{P(6)}$	$\frac{P(3)}{P(6)}$	$\frac{P(4)}{P(6)}$	$\frac{P(2)}{P(6)}$
a	$\frac{P(2)}{P(4)}$	$\frac{P(6)}{P(4)}$	$\frac{P(1)}{P(4)}$	$\frac{P(5)}{P(4)}$
b	$\frac{P(5)}{P(3)}$	$\frac{P(6)}{P(3)}$	$\frac{P(1)}{P(3)}$	$\frac{P(2)}{P(3)}$
h	$\frac{P(6)}{P(5)}$	$\frac{P(3)}{P(5)}$	$\frac{P(4)}{P(5)}$	$\frac{P(1)}{P(5)}$
g	$\frac{P(1)}{P(2)}$	$\frac{P(3)}{P(2)}$	$\frac{P(4)}{P(2)}$	$\frac{P(6)}{P(2)}$

Wyrażone poprzez wyznaczniki Δ^p macierzy admitancji węzłowych $\mathbf{Y}_w^p$ i sumy H_k^v dendrytów k-drzewowych, T_k^v części pasywnej układu — gdzie:

Dla czwórników:

$$P(1) = \Delta^a = \Delta^p_{ht} = H^v_{ht,0}$$

$$P(2) = \Delta^a_{11} = \Delta^p_{ht,11} = H^v_{ht,1,0}$$

$$P(3) = \Delta^a_{(3+2)1} = \Delta^p_{ht,(3+2)1} = H^v_{ht,21,0} - H^v_{ht,13,0}$$

$$P(4) = \Delta^a_{1(3+2)} = \Delta^p_{ht,1(3+2)} = H^v_{ht,12,0} - H^v_{ht,13,0}$$

$$P(5) = \Delta^a_{(3+2)(3+2)} = \Delta^p_{ht,(3+2)(3+2)} = H^v_{ht,2,0} + H^v_{ht,3,0} - 2H^v_{ht,1,23,0}$$

$$P(6) = \Delta^a_{11,(3+2)(3+2)} = \Delta^p_{ht,11,(3+2)(3+2)} = H^v_{ht,11,3,0} + H^v_{ht1,2,0} - 2H^v_{ht,123,0}$$

Dla trójników:

$$P(1) = \Delta^a = \Delta^p_{ht} = H^v_{ht,0}$$

$$P(2) = \Delta^a_{11} = \Delta^p_{ht,11} = H^v_{ht,1,0}$$

$$P(3) = \Delta^a_{21} = \Delta^p_{ht,21} = H^v_{ht,21,0}$$

$$P(4) = \Delta^a_{12} = \Delta^p_{ht,12} = H^v_{ht,12,0}$$

$$P(5) = \Delta^a_{22} = \Delta^p_{ht,22} = H^v_{ht,2,0}$$

$$P(6) = \Delta^a_{11,22} = \Delta^p_{ht,11,22} = H^v_{ht,1,2,0}$$

pasywnej układu lub sumy H_k^v iloczynów admitancji elementów tworzących dendryty k-drzewowe T_k^v. W tabelach 3 i 4 zestawiono poszczególne dopełnienia algebraiczne $\triangle_{ij}^p$ macierzy $\mathbf{Y}_w^p$ i sumy H_k^v, charakteryzujące poszczególne parametry $z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij}$ układów pasywnych, dla pierwszego (Rys. 1d) i drugiego (Rys. 1e i Rys. 1f) sposobu włączenia idealnego wzmacniacza operacyjnego.

Tabela 4

Parametry $F_{ij} \in \{z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij}\}$ charakteryzujce własności czwórników i trójników zawierających wzmacniacz operacyjny włączony według drugiego sposobu

F	F_{11}	F_{12}	F_{21}	F_{22}
z	$\frac{P(2)}{P(1)}$	$\frac{P(3)}{P(1)}$	$\frac{P(4)}{P(1)}$	$\frac{P(5)}{P(1)}$
y	$\frac{P(5)}{P(6)}$	$\frac{P(3)}{P(6)}$	$\frac{P(4)}{P(6)}$	$\frac{P(2)}{P(6)}$
a	$\frac{P(2)}{P(4)}$	$\frac{P(6)}{P(4)}$	$\frac{P(1)}{P(4)}$	$\frac{P(5)}{P(4)}$
b	$\frac{P(5)}{P(3)}$	$\frac{P(6)}{P(3)}$	$\frac{P(1)}{P(3)}$	$\frac{P(2)}{P(3)}$
h	$\frac{P(6)}{P(5)}$	$\frac{P(3)}{P(5)}$	$\frac{P(4)}{P(5)}$	$\frac{P(1)}{P(5)}$
g	$\frac{P(1)}{P(2)}$	$\frac{P(3)}{P(2)}$	$\frac{P(4)}{P(2)}$	$\frac{P(6)}{P(2)}$

Wyrażone poprzez wyznaczniki $\triangle^p$ macierzy admitancji węzłowych $\mathbf{Y}_w^p$ i sumy H_k^v dendrytów k-drzewowych, T_k^v części pasywnej układu — gdzie:

Dla czwórników:

$$P(1) = \triangle^a = \triangle^p_{h(i+j)} = H^v_{hj,0} - H^v_{hi,0}$$

$$P(2) = \triangle^a_{11} = \triangle^p_{h(i+j),11} = H^v_{hj,1.0} - H^v_{hi,1.0}$$

$$P(3) = \triangle^a_{(3+2)1} = \triangle^p_{h(i+j),(3+2)1} = H^v_{hj,21.0} + H^v_{hi,31.0} - H^v_{hj,31.0} - H^v_{hi,21.0}$$

$$P(4) = \triangle^a_{1(3+2)} = \triangle^p_{h(i+j),1(3+2)} = H^v_{hj,12.0} + H^v_{hi,13.0} - H^v_{hj,13.0} - H^v_{hi,12.0}$$

$$P(5) = \triangle^a_{(3+2)(3+2)} = \triangle^p_{h(i+j),(3+2)(3+2)} = H^v_{hj,3.0} + H^v_{hj,2.0} - 2H^v_{hj,23.0} - H^v_{hi,3.0} - H^v_{hi,2.0} + H^v_{hi,23.0}$$

$$P(6) = \triangle^a_{11,(3+2)(3+2)} = \triangle^p_{h(i+j),11,(3+2)(3+2)} = H^v_{hj,1.3.0} + H^v_{hj1.2.0} - 2H^v_{hj,1.23.0} - H^v_{hi,1.3.0} - H^v_{hi,1.2.0} + 2H^v_{hi,1.23.0}$$

Dla trójników:

$$P(1) = \triangle^a = \triangle^p_{h(i+j)} = H^v_{hj,i0} - H^v_{hi,j0}$$

$$P(2) = \triangle^a_{11} = \triangle^p_{h(i+j),11} = H^v_{hj,1.0} - H^v_{hi,1.0}$$

$$P(3) = \triangle^a_{21} = \triangle^p_{h(i+j),21} = H^v_{hj,21.0} - H^v_{hi,21.0}$$

$$P(4) = \triangle^a_{12} = \triangle^p_{h(i+j),12} = H^v_{hj,12.0} - H^v_{hi,12.0}$$

$$P(5) = \triangle^a_{22} = \triangle^p_{h(i+j),22} = H^v_{hj,2.0} - H^v_{hi,2.0}$$

$$P(6) = {}^a_{11,22} = \triangle^p_{h(i+j),11,22} = H^v_{hj,1.2.0} - H^v_{hi,1.2.0}$$

W tworzeniu poszczególnych sumy H_k^v iloczynów admitancji elementów tworzących dendryty k-drzewowe T_k^v wykorzystano następujące zależności:

$$\triangle^p_{hi,(m+l)p} = \triangle^p_{hi,lp} - \triangle^p_{hi,mp} = H^v_{hi,lp,0} - H^v_{hi,mp,0}$$

$$\triangle^p_{hi,(m+l)(m+l)} = \triangle^p_{hi,mm} + \triangle^p_{hi,ll} - \triangle^p_{hi,lm} = H^v_{hi,m,0} + H^v_{hi,l,0} - 2H^v_{hi,lm,0}$$

$$\triangle^p_{hi,pp,(m+l)(m+l)} = \triangle^p_{hi,pp,mm} + \triangle^p_{hi,pp,ll} - \triangle^p_{hi,pp,ml} - \triangle^p_{hi,pp,lm} = H^v_{hi,p,m,0} + H^v_{hi,p,l,0} - 2H^v_{hi,p,lm,0}$$

$$\triangle^p_{h(i+j),(m+l)p} = \triangle^p_{jj,lp} + \triangle^p_{hi,mp} - \triangle^p_{hi,jp} - \triangle^p_{hj,mp} = H^v_{hj,lp,0} + H^v_{hi,mp,0} - H^v_{hj,mp,0} - H^v_{hi,lp,0}$$

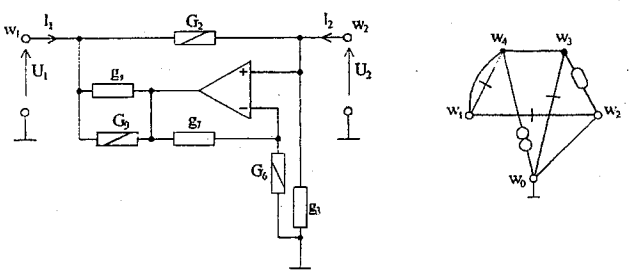
$$\triangle^p_{h(i+j),(m+l)(m+l)} = (\triangle^p_{hj,mm} + \triangle^p_{hj,ll} - \triangle^p_{hj,lm} - \triangle^p_{hj,ml}) - (\triangle^p_{hi,mm} + \triangle^p_{hi,ll} - \triangle^p_{hi,lm} - \triangle^p_{hi,ml})$$

$$= H^v_{hj,m,0} + H^v_{hj,l,0} - 2H^v_{hj,lm,0} - H^v_{hi,m,0} - H^v_{hi,l,0} + H^v_{hi,lm,0}$$

$$\triangle^p_{h(i+j)pp,(m+l)(m+l)} = H^v_{hj,p,m,0} + H^v_{hj,p,l,0} - 2H^v_{hj,p,lm,0} - H^v_{hi9,p,m,0} - H^v_{hi,p,l,0} + 2H^v_{hi,p,lm,0}$$

3. ELEMENTY ANALIZY UKŁADÓW ZAWIERAJĄCYCH IDEALNY WZMACNIACZ OPERACYJNY

Analizie układów aktywnych poświęcono wiele prac [1, 2, 7, 8]. „*Tradycyjny*" sposob obliczania poszczególnych wielomianów P(1)–P(6) licznika i mianownika funkcji wymiernej charakteryzującej parametry ukladów aktywnych zilustrujemy na przykładzie trójnika z rysunku 2.



Rys. 2. Trójnik pasywny z włączonym wzmacniaczem operacyjnym

I tak, aby obliczyć poszczególne wielomiany P(1)–P(6), należy utworzyć macirzy admitancji węzłowych $\mathbf{Y}_w^a$ pozbawioną parametrów wzmacniacza operacyjnego, która dla układu z rysunku 2 ma postać:

$$\mathbf{Y}_w^a = \begin{bmatrix} g_9 + G_2 x + G_9 x & -G_2 x & 0 & -g_9 - G_9 x \\ -G_2 x & g_3 + G_2 x & 0 & 0 \\ 0 & 0 & g_7 + G_6 x & -g_7 \\ -g_9 - G_9 x & 0 & -g_7 & g_7 + g_9 + G_9 x \end{bmatrix}$$

Jak wiadomo,[7, 10] poszczególne wielomiany P(*i*) to dopełnienia algebraiczne typu $\triangle^{a}_{(p_1+r_1)(m_1+l_1),(p_2+r_2)(m_2+l_2)}$. Wskaźniki przy $\triangle$ pokazują operacje jakie należy wykonać na macierzy $\mathbf{Y}^{p}_{w}$. W ogólnym przypadku elementy wierszy p_i dodajemy do elementów odpowiednich wierszy r_i po czym skreślamy wiersze p_i. Analogicznie elementy kolumn m_i dodajemy do elementów odpowiednich kolumn l_i, po czym skreślamy kolumny m_i. Uzyskany w ten sposob wyznacznik, należy pomnożyć przez $(-1)^{\sigma+\varkappa}$, gdzie: σ — to suma numerów skreślonych wierszy i kolumn, $\varkappa$ — całkowita liczba przedstawień w ciągach skreślonych wierszy i kolumn, potrzebnych do uszeregowania ichw porządku rosnącym.

Powyższe rozważania zilustrujemy na przykładzie układu z rysunku 2. Wielomian P(1) to dopełnienie algebraiczne typu $\triangle_{4(3+2)}$. W rozpatrywanym przypadku z macierzy $\mathbf{Y}^{a}_{w}$, należy wykreślić 4-ty wiersz oraz 3-cią kolumnę, przy czym skreślenia kolumny 3-ej można dokonać dopiero po dodaniu jej elementów do kolumny 2-ej. Wykonując powyższe operacje dopełnienie algebraiczne $\triangle_{4(3+2)}$ ma postać:

$$\triangle_{4(3+2)}=(-1)^{4+3}\begin{vmatrix} g_9+G_2x+G_9x & -G_2x & -g_9-G_9x \\ -G_2x & g_3+G_2x & 0 \\ 0 & g_7+G_6x & -g_7 \end{vmatrix}=-[-g_3g_7g_9-$$

$$-g_7g_9G_2x-g_3g_7G_2x-g_7G_2G_2x^2-g_7G_2G_9x^2+g_3G_7G_9x+g_7g_9G_2x+G_7G_2G_9x^2+$$
$$+g_9G_2G_6x^2+G_2G_6G_9x^3+g_7G_2G_2x2].$$

Po uproszczeniu otrzymujemy:

$$\mathrm{P}(1)=\triangle_{4(3+2)}=-[G_2G_6G_9x^3+g_9G_2G_6x^2-g_3g_7G_9x-g_3g_7G_2x-g_3g_7g_9].$$

Wielomian P(2) jest dopełnieniem algebraicznym typu $\triangle_{4(3+2),11}$. W tym przypadku z macierzy $\mathbf{Y}^{p}_{w}$, należy wykreślić 4-ty i 1-wszy wiersze oraz 3-cią i 1-szą kolumny, przy czym skreślenia kolumny 3-ej można dokonać dopiero po dodaniu jej elementów do kolumny 2-ej.

Wykonując powyższe operacje dopełnienie algebraiczne $\triangle_{4(3+2),11}$ ma postać:

$$\triangle_{4(3+2),11}=(-1)^{4+3+1+1+1+1}\begin{vmatrix} G_2x+g_3 & 0 \\ g_7+G_6x & -g_7 \end{vmatrix}=-[-g_7G_2x-g_3g_7].$$

O znaku dopełnienia algebraicznego decyduje suma wslaźników skreślonych wierszy (4+1) i kolumn (3+1) oraz liczba inwersji, jaką należy wykonać na każdej z tych grup wskaźników z osobna, aby tworzyły one ciągi rosnące. W rozpatrywanym przypadku na wierszach należy wykonać jedną inwersję i jedną inwersję na kolumnach. Pozostałe wielomiany P(3)—P(6) można wyznaczyć analogicznie wykreślając odpowiednie wiersze i kolumny. Sposób obliczania wielokrotnych sumarycznych dopełnień algebraicznych przedstawiono w pracy [10].

Jak łatwo zauważyć (tabele 2, 3 i 4) dopełnienie algebraiczne $\triangle^{p}_{hj,p,m}$ w ogólnym przypadku to sumą *cząstkowych* dopełnień algebraicznych $H^{v}_{hj,p,m,0}$ z pełnej macierzy admitancji węzłowej $\mathbf{Y}^{p}_{w}$ układu.

Jak wynika z pracy [13] „cząstkowe" dopełnienia algebraiczne $H^{v}_{hj,p,m,0}$ z pełnej macierz admitancji węzłowej $\mathbf{Y}^{p}_{w}$ układu można zapisać w postaci:

$$H^{v}_{hn,ol,pr}=(-1)^{m+n+o+l+p+r}\det \mathbf{Y}_{-mn,ol,pr}, \tag{15}$$

gdzie:

$\mathbf{Y}_{-mn,ol,pr}$ — pełna macierz admitancji węzłowej $\mathbf{Y}^{p}_{w}$ z której wykreślono m-ty, o-ty i p-ty wiersze, oraz n-tą, l-tą i r-tą kolumnę. „*Cząstkowe*" dopełnienie algebraiczne $H^{v}_{hj,p,m,0}$ zawierają tylko sumy iloczynów elementów tworzących dendryty k-drzewowych typu $T^{v}_{mn,ol,pr}(x)$.

Zgodnie z powyższymi rozważaniami, oraz pracami [9, 12, 13] przed wykreśleniem poszczególnych wierszy i kolumn z pełna macierz admitancji węzłowej $\mathbf{Y}^{p}_{w}$, należy w miejsce *niedozwolonych* gałęzi łączących zbiory wierzchołków tworzących dany typ dendrytu k-drzewowego, wstawiać wartość zero. Na przykład *niedozwolonymi* gałęziami dla dendrytu $T^{v}_{1,3,20}$ jest zbiór elementów $\{e_1, e_2, e_4, e_5, e_6\}$ $(e_1\in\{w_1, w_0\}, e_2\in\{w_2, w_1\}, e_4\in\{w_2, w_3\}, e_5\in\{w_3, w_1\}, e_6\in\{w_3, w_0\})$.

Korzystając z zapisu poszczególnych parametrów z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij} układów aktywnych w postaci *cząstkowych* dopełnień algebraicznych $H^{v}_{hj,p,m,0}$ (tabele 3 i 4), można zaproponować nieco inny *prostszy* algorytm analizy układów zawierających idealny wzmacniacz operacyjny. Poszczególne parametry układów zawierających idealny wzmacniacz operacyjny, można otrzymać *wykreślając* z pełnej macierzy admitancji węzłowych $\mathbf{Y}^{p}_{w}$ części pasywnej układu odpowiednie wiersze i kolumny.

Sposob *wykreślania* poszczególnych wierszy i kolumn w macierzy macierzy admitancji węzłowych $\mathbf{Y}^{p}_{w}$ zilustrujemy na przykładzie układu z rysunku 2 (w którym wzmacniacz operacyjny jest włączony według drugiego sposobu ($h=4$, $i=3$, $j=2$)).

Pełna macierz admitancji węzłowej $\mathbf{Y}^{p}_{w}$ części pasywnej układu ma postać:

$$\mathbf{Y}^{p}_{w}=\begin{bmatrix} g_3+G_6x & 0 & -g_3 & -G_6x & 0 \\ 0 & g_9+G_2x+G_9x & -G_2x & 0 & -g_9-G_9x \\ -g_3 & -G_2x & g_3+G_2x & 0 & 0 \\ -G_6x & 0 & 0 & g_7+G_6x & -g_7 \\ 0 & -g_9-G_9x & 0 & -g_7 & g_9+g_7+G_9x \end{bmatrix}$$

Poszczególne wielomiany P(1)—P(6) konieczne do wyznaczenia parametrów rozpatrywanego trójnika korzystając z tabeli 4 można wyznaczyć, obliczając odpowiednie cząstkowe dopełnienia algebraiczne $H^{v}_{hi,kl,np}$ z pełnej macierzy admitancji węzłowych $\mathbf{Y}^{p}_{w}$. I tak, aby obliczyć P(1) należy wyznaczyć $H^{5}_{42,30}(x)$ i $H^{5}_{43,20}(x)$. Cząstkowe dopełnienie algebraiczne $H^{5}_{42,30}(x)$. Cząstkowe dopełnienie algebraiczne $H^{5}_{42,30}(x)$ powstaje z macierzy $\mathbf{Y}^{p}_{w}$, po wykreśleniu z niej 4-ego i 3-ego wierszy, oraz 2-ej i 0-ej kolumn. Otrzymujemy wtedy:

$$H^5_{42,30}(x)=(-1)^{(4+2+3+0)}\begin{vmatrix} 0 & -G_6x & 0 \\ g_9+G_2x+G_9x & 0 & -g_9-G_9x \\ -G_2x & 0 & 0 \end{vmatrix}=G_2G_6G_9x^2+g_9G_2G_6x^2$$

Analogicznie obliczamy $H^5_{43,20}(x)$. Wtedy:

$$H^5_{43,20}(x)=(-1)^{(4+3+2+0)}\begin{vmatrix} 0 & -g_3 & 0 \\ g_9+G_2x+G_9x & -G_2x & -g_9-G_9x \\ 0 & 0 & g_7 \end{vmatrix}=+g_3g_7G_2x+g_3g_7G_9x+g_3g_7g_9.$$

Wielomian $P(1)=H^5_{42,30}(x)-H^5_{43,20}(x)$ ma postać:

$$P(1)=G_2G_6G_9x^3+g_9G_2G_6x^2-g_3g_7G_2x-g_3g_7G_9x-g_3g_7g_9$$

Wielomian P(2) ma postać: $P(2)=H^5_{42,1,0}(x)-H^5_{43,1,0}(x)$. Cząstkowe dopełnienie algebraiczne $H^5_{42,1,0}(x)$ powstaje z macierzy $\mathbf{Y}^p_w$, po wykreśleniu z niej 4-go, 1-go i 0-go wierszy, oraz 2-ej, 1-ej i 0-ej kolumn. Otrzymujemy wtedy:

$$H^5_{42,1,0}(x)=(-1)^{(4+2+1+1+0+0)}\begin{vmatrix} 0 & 0 \\ g_7+G_6x & -g_7 \end{vmatrix}=0.$$

Analogicznie obliczamy $H^5_{43,1,0}(x)$. Otrzymujemy wtedy:

$$H^5_{43,1,0}(x)=(-1)^{(4+3+1+1+0+0)}\begin{vmatrix} g_3+G_2x & 0 \\ 0 & -g_7 \end{vmatrix}=+g_7G_2x+g_3g_7,$$

stąd wielomian $P(2)=H^5_{42,1,0}(x)-H^5_{43,1,0}(x)=-g_7G_2x-g_3g_7$.

Pozostałe wielomiany P(3)–P(6) można wyznaczyć analogicznie wykreślając odpowiednie wiersze i kolumny.

4. OGRANICZENIA REALIZACJI WIELOMIANÓW P(1)–P(6) CHARAKTERYZUJĄCYCH PARAMETRY z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij}

Ponieważ parametry czwórników i trójników są określone poprzez prądy i napięcia $\{I_1, U_1\}$ i $\{I_2, U_2\}$ związane odpowiednio z wierzchołkami wejściowymi (w niniejszej pracy przyjęto $\{w_1, w_0\}$) i wyjściowymi (w niniejszej pracy przyjęto $\{w_2, w_3\}$ dla czwórników i $\{w_2, w_0\}$ dla trójników), dołączenie nulatora lub noratora do wierzchołków brzegowych wprowadza pewne dodatkowe ograniczenia na realizację niektórych parametrów F_{ij} układów.

Mogą tutaj występować następujące ograniczenia w realizacji parametrów F_{ij}.

1. Nulator włączony równolegle do wierzchołków wejściowych $\{w_1, w_0\}$, układu pasywnego. Jak wiadomo, nulator to element, przez który nie płynie prąd ($I=0$) i na zaciskach którego nie występuje napięcie ($U=0$). Takie

włączenie nulatora powoduje, że $U_1 = 0$. Tym samym nie mogą być określone te parametry F_{ij} układów (Tab. 1) w których jako sygnał informacyjny wystepuje U_1 i (lub) I_1. Przykład ilustrujący ten sposób włączenia wzmacniacza operacyjnego do układu pasywnego pokazano na rysunku 1 w tabeli 5.

2. Nulator włączony równolegle do wierzchołków wyjściowych $\{w_2, w_3 \vee w_2, w_0\}$ układu pasywnego. Takie włączenie nulatora powoduje, że $U_2 = 0$. Tym samym nie mogą być określone te parametry F_{ij} układów (Tab. 1) w których jako sygnał informacyjny występuje U_2 i (lub) I_2. Przykład ilustrujący ten sposób włączenia wzmacniacza operacyjnego do układu pasywnego pokazano na rysunku 2 w tabeli 5.

3. Norator włączony równolegle do wierzchołków wejściowych $\{w_1, w_0\}$ układu pasywnego. Jak wiadomo, norator jest elementem, przez który płynie prąd o dowolnej wartości ($I = 0 + \infty$) i na zaciskach którego występuje napięcie o dowolnej wartości ($U = 0 + \infty$). Należy zwrócić uwagę, że napięcie i prąd noratora nie są między sobą związane. Takie włącenie noratoa powoduje, że napięcie wejściowe U_1 i prąd wejściowy I_1 mogą przyjmować dowolne wartości ($U_1 = 0 \div \infty$, $I_1 = 0 \div \infty$). Przykład ilustrujący ten sposob włączenia wzmacniacza operacyjnego do układu pasywnego pokazano na rysunku 3 w tabeli 5.

4. Norator włączony równolegle do wierzchołków wyjściowymi $\{w_2, w_3 \vee w_2, w_0\}$ układu pasywnego. Takie włączenie noratora powoduje, że napięcie wyjściowe U_2 i prąd wyjściowy I_2 mogą przyjmować dowolne wartości ($U_2 = 0 \div \infty$, $I_2 = 0 \div \infty$). Przykład ilustrujący ten sposób włączenia wzmacniacza operacyjnego do układu pasywnego pokazano na rysunku 4 w tabeli 5.

5. Norator włączony równolegle do wierzchołków $\{w_1, w_2\}$ trójnika pasywnego. W tym przypadku $U_1 = U_2$. Przykład ilustrujący ten sposób włączenia wzmacniacza operacyjnego do układu pasywnego pokazano na rysunku 5 w tabeli 5.

Pełny zestaw ograniczeń realizacji poszczególnych parametrów z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij} układów, można otrzymać wychodząc z następujących rozważań. Poszczególne wielomiany P(1) – P(6) występujących w liczniku lub mianowniku funkcji wymiernej opisującej dany parametr F_{ij} czwórnika lub trójnika (tabele 3 i 4) są określone *cząstkowymi* dopełnieniami algebraicznymi z macierzy admitancji węzłowych $\mathbf{Y}_w^p$ zapisanymi w postaci $H^v_{hi,kl,np}(x)$. Dane *cząstkowe* dopełnienie algebraiczne $H^v_{hi,kl,np}(x)$ można zrealizować wykreślając z pełnej macierzy admitancji węzłowych $\mathbf{Y}_w^p$ odpowiednie wiersze i kolumny.

Cząstkowe dopełnienie algebraiczne $H^v_{hi,kl,np}(x)$ nie można obliczyć wtedy, jeśli z pełnej macierzy admitancji węzłowych $\mathbf{Y}_w^p$ należałoby wielokrotnie wykreślić dany wiersz lub kolumnę. W tym przypadku układ nie realizuje $H^v_{hi,kl,np}(x)$ — części składowej wielomianu P(i). Na przykład, trójnik z rysunku 1 z tabeli 5 nie realizuje wielomianu P(2), ponieważ w dopełnieniu algebraicznym $H^4_{21,11}(x)$, należałoby dwukrotnie wykreślić 1-szą kolumnę.

Pełny zestaw ograniczeń realizacji wielomianów P(1) – P(6) występujących w liczniku lub mianowniku funkcji wymiernej opisującej dany parametr z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij} czwórnika lub trójnika zawierających idealny wzmacniacz operacyjny włączony według pierwszego i drugiego sposobu zestawiono odpowiednio w tabelach 6 i 7.

Tabela 5

Przykłady ograniczeń w realizacji funkcji wymiernej opisującej dany parametr F_{ij} trójnika związane z włączeniem nulatora i noratora do jego wierzchołków brzegowych

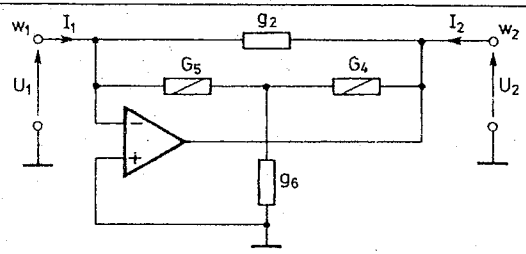
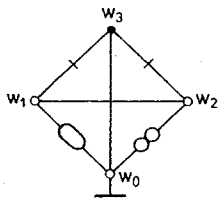
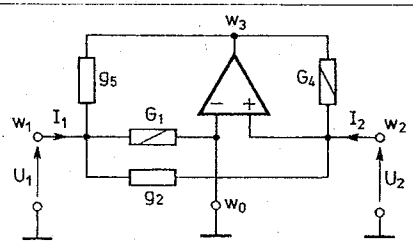
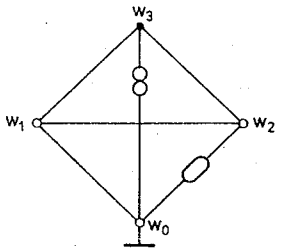
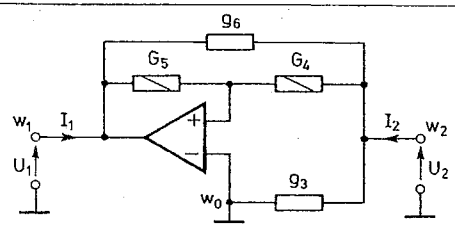
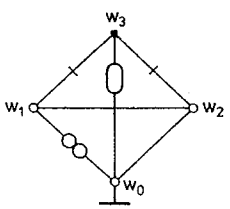
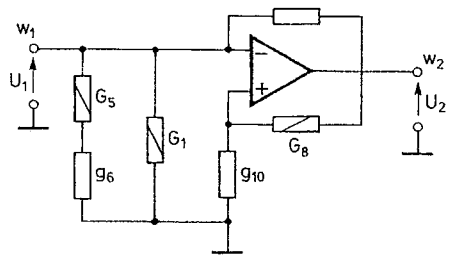
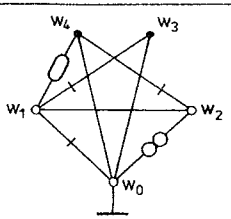
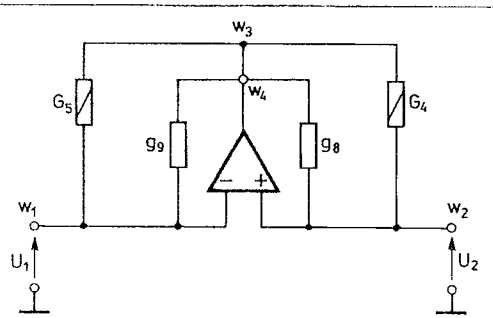
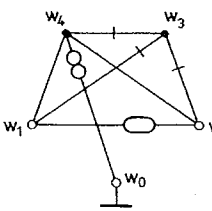
Lp.	Przykład struktury układu	
1		
2		
3		
4		
5		

Tabela 6

Ograniczenia realizacji wielomianów P(1)—P(6) występujących w liczniku lub mianowniku funkcji wymiernej F_{ij} opisującej dany parametr trójnika (T) lub czwórnika (Cz) zawierających idealny wzmacniacz operacyjny włączony według pierwszego sposobu

Lp	Norator	Nulator	P(1)		P(2)		P(3)		P (4)		P (5)		P (6)	
	$\{w_0,w_h\}$	$\{w_0,w_i\}$	T	Cz	T	Cz	T	Cz	T	Cz	T	Cz	T	Cz
1	$\{w_0,w_1\}$	$\{w_0,w_2\}$	1	1	0	0	1	1	0	0	0	1	0	0
2	$\{w_0,w_1\}$	$\{w_0,w_3\}$	1	1	0	0	1	1	0	0	1	1	0	0
3	$\{w_0,w_1\}$	$\{w_0,w_m\}$	1	1	0	0	1	1	0	0	1	1	0	0
4	$\{w_0,w_2\}$	$\{w_0,w_1\}$	*1*	*1*	*0*	*0*	*0*	*0*	*1*	*1*	*0*	*1*	*0*	*0*
5	$\{w_0,w_2\}$	$\{w_0,w_3\}$	1	1	1	1	0	1	1	1	0	1	0	0
6	$\{w_0,w_2\}$	$\{w_0,w_m\}$	1	1	1	1	0	1	1	1	0	1	0	1
7	$\{w_0,w_3\}$	$\{w_0,w_1\}$	1	1	0	0	0	0	1	1	1	1	0	0
8	$\{w_0,w_3\}$	$\{w_0,w_2\}$	1	1	1	1	1	1	0	1	0	0	0	1
9	$\{w_0,w_n\}$	$\{w_0,w_1\}$	1	1	0	0	0	0	1	1	1	1	0	0
10	$\{w_0,w_n\}$	$\{w_0,w_2\}$	1	1	1	1	1	1	0	1	0	1	0	1

1 — układ realizuje dany wielomian P(i)
0 — układ nie realizuje wielomianu P(i)

Tabela 7

Ograniczenia realizacji wielomianów P(1)–P(6) występujących w liczniku lub mianowniku funkcji wymiernej F_{ij} opisującej dany parametr trójnika (T) lub czwórnika (Cz) zawierających idealny wzmacniacz operacyjny włączony według drugiego sposobu

Lp	Norator	Nulator	P(1)		P(2)		P(3)		P (4)		P (5)		P (6)	
	$\{w_0,w_h\}$	$\{w_0,w_j\}$	T	Cz	T	Cz	T	Cz	T	Cz	T	Cz	T	Cz
1	$\{w_0,w_1\}$	$\{w_1,w_2\}$	1	1	0	0	1	1	0	0	0	1	0	0
2	$\{w_0,w_1\}$	$\{w_1,w_3\}$	1	1	0	0	1	1	0	0	1	1	0	0
3	$\{w_0,w_1\}$	$\{w_1,w_m\}$	1	1	0	0	1	1	0	0	1	1	0	0
4	$\{w_0,w_1\}$	$\{w_2,w_3\}$	*1*	*1*	*0*	*0*	*0*	*0*	*1*	*1*	*0*	*1*	*0*	*0*
5	$\{w_0,w_1\}$	$\{w_2,w_m\}$	1	1	1	1	0	1	1	1	0	1	0	0
6	$\{w_0,w_1\}$	$\{w_3,w_m\}$	1	1	1	1	0	1	1	1	0	1	0	1
7	$\{w_0,w_2\}$	$\{w_2,w_1\}$	1	1	0	0	0	0	1	1	1	1	0	0
8	$\{w_0,w_2\}$	$\{w_2,w_3\}$	1	1	1	1	1	1	0	1	0	0	0	1
9	$\{w_0,w_2\}$	$\{w_2,w_m\}$	1	1	0	0	0	0	1	1	1	1	0	0
10	$\{w_0,w_2\}$	$\{w_1,w_3\}$	1	1	1	1	1	1	0	1	0	1	0	1
11	$\{w_0,w_2\}$	$\{w_1,w_m\}$	1	1	1	1	0	1	1	1	0	1	0	1
12	$\{w_0,w_2\}$	$\{w_3,w_m\}$	1	1	1	1	0	1	1	1	0	1	0	1
13	$\{w_0,w_3\}$	$\{w_3,w_m\}$	1	1	1	1	1	1	1	1	1	1	1	1
14	$\{w_0,w_n\}$	$\{w_1,w_2\}$	1	1	1	1	1	1	1	1	1	1	0	1

1 — układ realizuje dany wielomian P(i)
0 — układ nie realizuje wielomianu P(i)

PODSUMOWANIE

Jak wynika z powyższych rozważań, oraz prac [7, 8] *dołączenie* idealnego wzmacniacza operacyjnego do układu pasywnego, nie tylko nie komplikuje obliczeń poszczególnych parametrów powstałego układu aktywnego, ale nawet je nieco upraszcza w stosunku do bazowego układu pasywnego. Włączenie do pasywnej struktury układu idealnego wzmacniacza operacyjnego powoduje, że wszystkie parametry F_{ij} charakteryzujące układ aktywny są określane tylko za pomocą wartości elementów jego części pasywnej (elementy włączone równolegle do nulatora i noratora nie uczestniczą w tworzeniu parametrów F_{ij}). Przedstawienie dopełnienia algebraicznego $\triangle^a$ macierzy admitancji węzłowych $\mathbf{Y}_w^a$ układu aktywnego w postaci cząstkowych dopełnień algebraicznych $H_{hj,p,m,0}^v$ pełnej macierzy admitancji węzłowych $\mathbf{Y}_w^p$ części pasywnej układu pozwala w prosty sposób wyznaczyć (korzystając z tabel 3 i 4) poszczególne parametry układu aktywnego zawierającego idealny wzmacniacz operacyjny.

Główną zaletą takiego przedstawienia parametrów układów aktywnych jest fakt, że analiza i synteza układów zawierających idealny wzmacniacz operacyjny jest w rzeczywistości analizą i syntezą części pasywnych tych układów.

Zdefiniowane ograniczenia dotyczące możliwości włączenia wzmacniacza operacyjnego do układu pasywnego (tabele 6 i 7), pozwolą w procesie syntezy struktur kanonicznych zminimalizować moce rozpatrywanych zbiorów układów pretendentów.

BIBLIOGRAFIA

1. A.C. Aleksenko, B.T. Zujew, V.F. Lamiekin, I.A. Romanow: *Makromodilirowanie analogowych intiegralnych mikrschem.* Radio i Swiaz, Moskwa, 1983
2. W.I. Aliew: *K analizu cepiej sodierżaszczich idiealnyje opiercjinnyje usilitieli.* Radiotechnika i Elektronika 1974, tom XIX, Nr 10, ss. 2089—2095
3. M. Białko: *Zastosowanie nulatorów i noratorów do tworzenia równoważnych układów aktywnych.* Rozprawy Elektrotechniczne, 1971, t. XVII, z. 3, ss. 399—414
4. S. Bolkowski: *Elektrotechnika teoretyczna — teoria obwodów elektrycznych.* WNT, Warszawa 1981
5. W.A. Kisiel: *Raściot ciepej s mnogopolusnymi opieracjonnymi usilitielami.* Radiotechnika i Elektronika 1982, tom XXVII, Nr 11, ss. 2197—2203
6. Z. Kulka, M. Nadachowski: *Liniowe układy scalone i ich zastosowania.* WKiŁ, Warszawa 1974
7. L. Lasek, J.J. Witkowski: *General approach to the analysis of networks having ideal operational amplifiers.* IEEE Journal Electronic and Systems, 1977, vol. 1, nr 4, pp. 133—136
8. W.P. Sigorskij, K. Thylasiraman: *Algoritmy analiza elektronnych schem.* Tiechnika, Kijew 1970
9. M.N.S. Swamy, K. Thulasiraman: *Graphs, networks and algorithms.* Wiley-Interscience, New York 1981
10. J.J. Witkowski: *Skuteczna metoda obliczania wielokrotnych sumarycznych dopełnień algebraicznych.* Zeszyty Naukowe Pol. Śląskiej, Automatyka, z. 66, ss. 49—61, Gliwice 1983

11. Z. W r ó b e l: *Funkcje strukturalne grafów układów dwójnikow i czwórników składających się z dwóch rodzajów elementów.* Kwart. Elektr. i Telekom. 1991, t. 37, z. 1—2, ss. 35—52
12. Z. W r ó b e l: *D-rownoważność struktur układów realizujących funkcje wymierne* Kwart. Elektr. i Telekom. 1993, t. 39, z. 3, ss. 523—539

Z. WRÓBEL

ALTERNATIVE DESIGNS OF INCORPORATING AN OPERATIONAL AMPLIFIER INTO THE PASSIVE NETWORK

Summary

The paper presents possible ways of incorporating an operational amplifier into the structures of passive four- and three- terminal networks, and relevant restrictions on the realisation of the individual parameters of the networks. The individual parameters f_{ij} ($F_{ij} \in \{z_{ij}, y_{ij}, a_{ij}, b_{ij}, h_{ij}, g_{ij}\}$) characterising the properties of the four- and three- terminal networks with one operational amplifier can be expressed by an appropriate algebraic determinant of the nodal admittance matrices $\mathbf{Y}_w^p$ of the passive part of the networks.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1994. 40, z. 4. ss. 529–539

681.325

Minimisation of capacitances recharging errors in charge-distribution analog-to-digital converters

ANDRZEJ MURSAJEW

Instytut Informatyki i Automatyki Morskiej, Politechnika Szczecińska

ROŚCISŁAW GRUSZWICKI, ANDRZEJ SIDOROW

Electrotechnical University, Sankt Petersburg

Received 1994.09.25

Authorized 1994.12.19

Charge injection significantly influences the accuracy of devices based on switched-capacitances technique. The comparison of different methods of charge injection compensation in analog-to-digital charge distribution converters is presented in the paper. Results of simulation have shown, that the 16–18 bit resolution may be achieved using optimally related parameters of switches and capacitors shunting those switches.

1. INTRODUCTION

Switching-capacitor technique is a perspective approach to develop monolithic digital-analog devices. Significant place in the market have successive-approximation analog to digital converters realising charge-distribution method. The main advantages of this technique are the execution in a single unit of sample-and-hold and conversion operations and uniform technology of both digital and analog circuits production. Nowadays charge-distribution analog-to-digital converters (CDADC) are the leaders among sampling converters within the range of 16–18 bit accuracy and 5–25 μs conversion time [1].

Since the first attempts to design CDADC in earlier 70-th [2] much efforts of explores and manufacturers were aimed to overcome problems typical for this technique — the compensation of technological variances of capacitors area and the decrease of charge injection from control circuits into capacitors through switches.

Technological variances compensation nowadays is achieved either by means of compulsory calibration of by self calibration.

Compulsory calibration presumes changing of parameters of manufactured elements according to the results of tests using external equipment. Evaporating parts of elements by means of laser heating is a commonly used technology [3, 4], but this method violates the crystal structure and makes the time stability worse. Another and nowadays becoming more popular method is based on the implementation additional resistor or capacitor arrays, connected with the main DAC array into a chip. Due to the test results these additional arrays are reconfigurated using evaporation of connectors by laser beam or any other PROM-like technology [5].

The regulation made once may not guarantee the stability of parameters for a long time and within wide temperature range, that is why self-calibration based on periodical testing has gone into practice. In the most advanced form self-calibration is realised in the family of CDADC's produced by Crystal Semiconductors Inc. [6, 7]. In CS's approach sets of regulating capacitors and switches are connected in parallel to each capacitor of the DAC array. In the self-calibration mode such set of signals applied to the switches is chosen, so that the necessary proportion of weighting circuits capacitances is achieved. This set, named a correcting code, is written into the on-chip RAM and stored for the time of normal work (conversion mode). The alternative self-calibrating ADC structure uses correcting code not for circuit parameters' control, but for calculation of the summary error corresponding to the current digital value according to special algorithm [8]. That error estimation is subtracted from the value, obtained in the main converting device.

Nevertheless calibration, that is actually the parameter regulation, is not enough for charge-injection compensation. First, the equivalent influence of charge injection is much stronger than the influence of errors, caused by technological variances, and, as minimum, requires significant increase of the array, occupied by the correcting circuitry on a chip. Secondly, parasite capacitors nonliniarity causes error dependence on the converted signal value. Such errors may not be fully compensated by means of changing the capacitances (areas) of components; it is necessary to use special techniques to *take aside* the charges flowing through the parasite components or to introduce additional charges into capacitors of charge-distribution array. This article deals with the methods to design commutative cells of CDADC minimising errors of recharging with possibly minimal time requirements.

2. CHARGE-INJECTIONS EFFECTS IN CDADC

The typical structure of charge-distribution ADC is presented in fig. 1 [9, 10]. In this network DAC is an array of binary-weigthed capacitors. All legs of the array share common node at the comparator input and summary capacitances is $C=2^{n-1}C_0$.

Conversion process takes two stages. When conversion command is issued the sampling stage begins. Switch S is opened and all capacitors are connected to the line U_{in}. This traps charge $Q_{in} = U_{in}C$ on the comparator side of the capacitor array and creates a floating node at the comparator input. This capacitor array accumulates the charge that is proportional to the input signal. Conversion algorithm operates at fixed charge and state of the analog input pin is ignored. In effect the entire DAC capacitor array serves as analog memory and no external sampling network is not necessary.

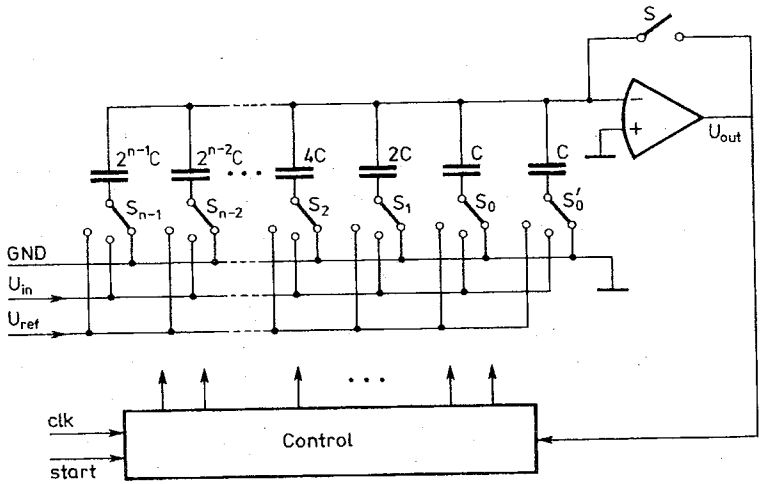


Fig. 1. Structure of a charge-distribution ADC

At the beginning of the second stage all capacitors are connected to zero-voltage line (GND). Voltage on the input of the comparator raises to the level $-U_{in}$. The conversion consists of manipulating the free plates of the capacitor array to U_{ref} and GND to form a capacitor divider and drive the voltage at the floating point to zero. Since the charge at the floating node remains fixed the voltage at that node depends on the proportion of capacitances tied to U_{in} versus GND. The successive approximatin algorithm is used to find the necessary proportion of capacitance.

For example to determine the most significant bit (MSB) the capacitor $2^{n-1}C_0$ is tied to U_{ref} by setting corresponding input to logical ONE, while the others are tied to GND. Voltage at the floating node raises to

$$U = -U_i + 0.5 \times U_{ref}.$$

According to the comparator response the capacitor $2^{n-1}C_0$ is retied to GND or remains in the previous state. Then the same operations are done with all capacitors sequentially in the turn of there capacitances decrease. When floating node is zeroed the charge in the array Q is

$$Q = U_{ref}C^*,$$

where C — capacitance, tied to U_{ref}.

If no additional charge is injected into the capacitor array ($Q=Q_{in}$) the procedure results such proportion of summary capacitances:

$$\frac{C^*}{C}=\frac{U_{in}}{U_{ref}}=\frac{\sum_{i=0}^{n-1}\alpha_i C_i}{C_0+\sum_{i=0}^{n-1}C_i}=\frac{\sum_{i=0}^{n-1}\alpha_i 2^i C_0}{2^n C_0}=\sum_{i=0}^{n-1}\alpha_i\, 2^{i-n}, \tag{1}$$

where $\alpha_i=1$ if i-th capacitor is tied to U_{ref}, otherwise $\alpha_i=0$.

Code $\{\alpha_{n-1},\alpha_{n-2},\ldots,\alpha_0\}$ presents the digital output of the converter.

Actually Q differs from Q_{in} because of the injection of additional charges through the switches pins while commutating, and it is a very valuable part of summary error. Influence of charge injection may be decreased by the increase of capacitances in the DAC array or by the decrease of switches arrays. But it effects the increase of conversion time. To achieve the convenient speed-to-accuracy relation it is necessary to guarantee quick recharging with minimal errors.

For example, to achieve 10 μS conversion time with 16-bit accuracy and 5 V of full scale, the capacitor of the MSB (usually of 100 pF) in the worst case must be recharged from U_{ref} to zero in less than 50 nS, and the error may be not more than 50 μV. That requires the decrease of switches resistances, actually the increase of width-to-length (W/L) relation of MOS-transistors in the switching circuits. But the increasing of W causes the increase of gate-to-channel parasite capacitances of switches and, as the result, increases errors caused by charge injection from control circuits into DAC array while commutating. So speed and accuracy are the contradictory parameters. To improve both it is necessary to use techniques to minimise the influence of injection.

In recent years many methods to diminish the influence of charge injection have been proposed, among them those are to be mentioned.

1. **Usage of compensating capacitors.** One plate of any compensating capacitor is tied with the output of the commutating switch and inverted control signal is applied to another plate. A charge flying through this capacitor partially compensates the charge injected through the switch's parasite capacitance. But because of the nonlinearity of MOS transistor capacitances full compensation is unachievable. Better compensation is achieved by using dummy transistors with their gate-drain capacitance as an equivalent of a compensating capacitor [11].
2. **Usage of CMOS switches.** Partial compensation is achieved because parasite charges flying through CMOS transistor pair are complement. But even an ideally balanced transistor pair does not guarantee full compensation within the whole range of commutated signals because the signal on the commutating node differently affects the values of internode capacitances in P and N branches.

3. **Variation of switches' parameters** (usage of different quantity of transistors or different array of transistors to realise several switches of the DAC). In the ideal case if the summary capacitances of switches in different bits are in binary relation, the injected charges will be also binary-weighted and the conversion function will be linearised. Really the difference of the circuit configuration in the sampling and recharging stages (while sampling the common node of all capacitors is virtually zeroed and while recharging it is not) effects the loss of direct capacitor-to-injection correspondence. Optimal ratio of transistors areas (or quantities of transistors) in various bits becomes not binary and needs additional investigations.
4. **Shunting of the switches by the capacitors**. Pulses of current between the gate and the channel of MOS-switch which appear during recommutation are relatively short and it is natural to try to get these pulses aside from DAC array through some reactive circuit.

The first two methods take more space on the chip to implement compensating circuitry and are not enough for 16–18 bit accuracy. Results of simulation of CDADC based on CMOS switches are presented in this paper. Our investigations have shown, that better compensation with smaller space demands can be achieved with the usage of 3-rd and 4-th methods together and corresponding optimisation of switches and shunting capacitors parameters.

The problem of charge injection compensation for the case of a single MOS switch connected to discrete or shared $R-C$ — circuits was discussed in [12, 13, 14]. But no papers dealing with the analysis of multi-switch charge-distribution circuits have been published.

3. SIMULATION AND OPTIMISATION

Simplified equivalent circuit of a CDADC at the beginning of the K-th bit testing is presented in Fig. 2.

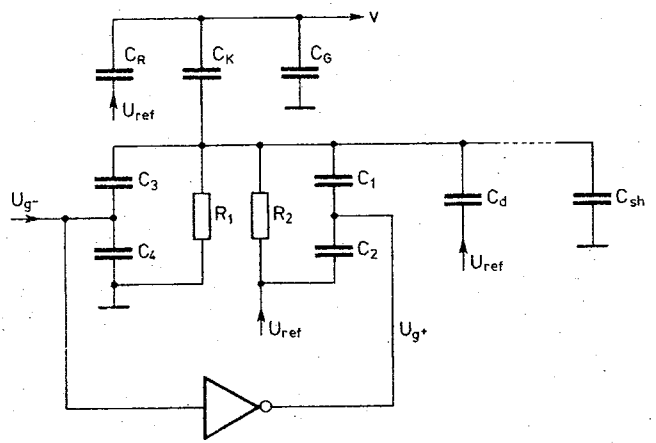


Fig. 2. Equivalent circuit of an ADC bit

In this figure C_R and C_G present sums of weighting capacitances tied to U_{ref} and GND respectively at the current step of successive approximation. Their ratio is defined during previous steps of conversion. Node V is the input of the comparator. The parasite capacitance at this node is negligible in comparison with C_R and C_G. Capacitors C_1 and C_2 present the parasite capacitances of the switch, that connects the lower plate of C_k with U_{ref}, C_3 and C_4 present the switch, connecting that plate with GND. Resistances of switches R_1 and R_2 are the functions of voltages applied to the gates of MOS-transistors. C_d presents summary capacitance of the switch, connected with U_{in} and C_{sh} — the optional shunting capacitor, that may be introduced to decrease charge-injection influence.

Analytic investigation of such circuit seems to be impossible because of the MOS internode capacitances and resistances nonlinearity. For precision CDADC calculations are to be done with the accuracy of some tens of mikrovolts. That is why faint effects must be taken into consideration: offset voltage of the opened switch, leakage current of the closed switches, base-to-substrate parasite capacitances, subthreshold effect and so on. The problem may be solved using computer simulation of the circuit. Computer analysis makes it possible to calculate states of the circuit in different modes considering various parasite effects and to optimise parameters according to accuracy and technological constraints.

For our simulation we defined the capacitance of the highest significant bit to be 120 pF and the minimal channel length to be 5 μm. Switches are controlled with signals of 5 V amplitude and 1 nS raise/fall time. Determining of parameters sufficient to achieve 16-bit accuracy with 10 μS conversion time was the goal.

Practical results presented further were achieved using the models of MOS-transistors, presented in [15]. Some parameters of this model were changed due to our technology (used values of the most essential parameters are listed in table 1). But the simulation with significantly changed parameters also confirmed the effectiveness of proposed techniques for charge-injection compensation.

The simulation of the sampling stage has shown that recharging with necessary accuracy is guaranteed using the NMOS transistor of 400 μm width as a switch in the most significant bit. The widths of other switching transistors may be decreased proportionally to the bit capacitances.

The recharging stage was simulated as follows. The "ideal" response code corresponding to current U_{in} was previously calculated using (1). The sequences of signals that would occur while processing successive-approximation algorithm under negligible error conditions were applied to the nodes of switches U_{g-} and U_{g+}. In that case the voltage at V-node at the end of conversion must be zero of no errors exist. Actually because of charge injection the charges in the array are not fully balanced. That remaining disbalance was chosen as an estimation of charge-injection influence.

Table 1

Symbol	Parameter	NMOS	PMOS
L	Model level (complexity)	2	2
U_{to}	Threshold voltage [V]	0.9	1.1
T_{ox}	Oxide thickness [m]	4.5×10^{-8}	4.5×10^{-8}
V_{sub}	Substrate impurity concentration [1/cm³]	3.16×10^{16}	2.0×10^{16}
γ	Factor of substrate influence [$V^{1/2}$]	0.467	0.452
P_{hi}	Metal-semiconductor junction voltage [V]	0.747	0.602
P_b	Voltage of subsurface layer inversion [V]	0.9	0.792
X_j	Junction depth [m]	9.5×10^{-7}	8.5×10^{-7}
L_d	Length of side diffusion space [m]	5.5×10^{-7}	5.5×10^{-7}
C_j	P-N-junction specific capacitance [F/m²]	5.72×10^{-4}	3.3×10^{-4}
C_{jsw}	Side surface specific capacitance [F/m]	6.5×10^{-10}	7.3×10^{-10}
J_s	Saturation current density [A/m²]	1.0×10^{-8}	1.0×10^{-8}
C_{gdo}	Gate-to-drain specific capacitance [F/m]	7.67×10^{-11}	7.67×10^{-11}
C_{gso}	Gate-to-source specific capacitance [F/m]	7.67×10^{-11}	7.67×10^{-11}
V_o	Surface carrier mobility [cm²/V/cm]	600	200
V_{exp}	Empirical const. specifying carrier mobil.	0.094	0.166
V_{max}	Maximal speed of carrier drift [m/S]	6.0×10^6	0.8×10^5

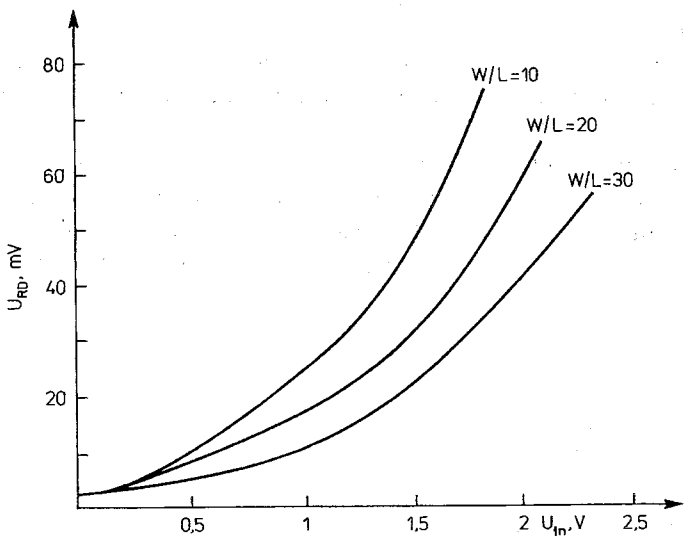


Fig. 3. Remaining disbalance in the circuit without compensation

Fig. 3 presents the results of simulation of the DAC, realised using NMOS switches with equal widths and no shunting capacitors. There the error estimator U_{ref} for various values of W/L ratio for MSB is plotted against the value of converted voltage. U_{rd} is growing unlinealy with the growth of U_{in}, actually with the summary capacitance of bits that were switched from ONE-state to ZERO-state and not reswitched during conversion. For this structure even the ideal binary relations of weighting capacitors can not guarantee the accuracy of

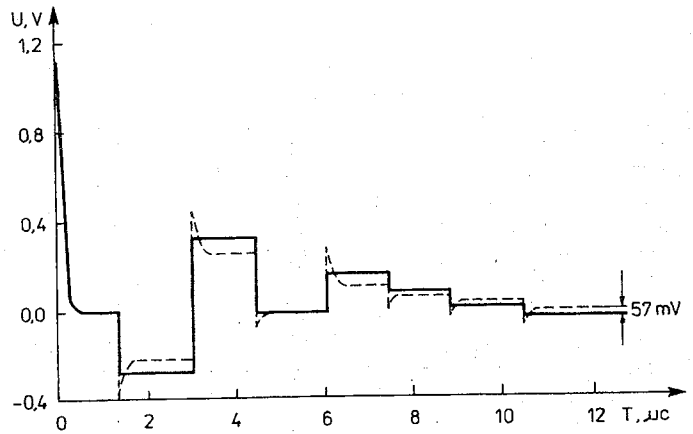


Fig. 4. Wave-forms at V-node while conversion:
(– – –) — without compensation, $C_{ahi}=0$, $M_i=40$ pF, $i=1$; n;
(———) — usage of optimaly related shunt capacitances and transistors

more than 8 bits. Glitches at the node V caused by raises and falls of control signals achieve double value of U_{in} (see fig. 4) and may destroy the normal work of the comparator.

Simulation of a DAC created with the usage of CMOS switches has shown that the glitches and errors are decreased approximately thrice and maximum value of U_{rd} becomes 15 mV.

Introducing of shunting capacitors (C_{sh} in fig. 2) is a better way to reduce glitches and remaining disbalance. These capacitors "get away" short parasite current pulses and do not influence the "valuable" part of capacitor charges, which is determined only by the voltages at central nodes of switches after transients.

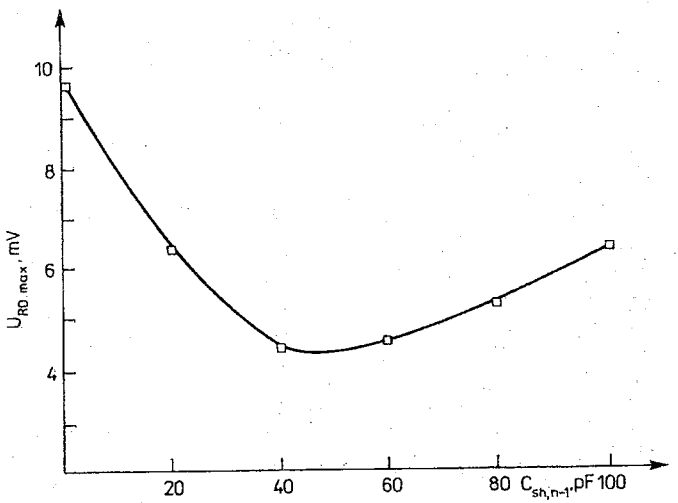


Fig. 5. Effect of shunt capacitance

While investigating the DAC with introduced shunting capacitors it was found that the U_{rd} is a function not only of C_{sh} value for a single bit, but also of the C_{sh} relations in various bits. To optimise parameters we varied C_{sh} for MSB ($C_{sh,n-1}$) and for each $C_{sh,n-1}$ stochastic searching procedure in the parameter space $\{C_{sh,n-1}, C_{sh,n-2},\ldots,C_{sh,0}\}$ was done. The plot of U_{rd} against $C_{sh,n-1}$ is presented in fig. 5.

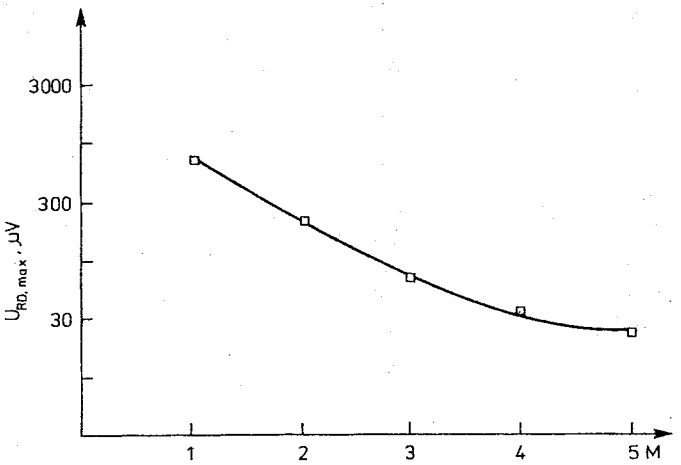


Fig. 6. Usage of paralelly tide switching capacitors

Plotted values correspond to optimally related shunting capacitances. The minimum error (about 4 mV) corresponds to $C_{sh,n-1}=40$ pF with presented rato of shunting capacitances:

$$C_{sh,n-1} : C_{sh,n-2} : C_{sh,n-3} : \ldots : C_{sh,0} = 1 : 1 : 2^{-1} : 2^{-2} : \ldots : 2^{-n+2}.$$

Such parameters are enough to achieve 11 bit accuracy of the converter. Actually the influence of charge-injection in lower bits is relatively small and it is enough to use shunting capacitors only in 4–5 higher bits. Substitution of NMOS switches by CMOS decreases the error twice.

Additional decrease of charge-injection influence may be achieved using unindentical switches in different bits (see p. 4 in section 2). Such difference may be obtained either by using MOS transistors of different width or by tying several identical transistors in parallel to realise different switches. The second approach seems to be more convenient because in this case the parasite time factor of such complicated switch remains constant with the changes of summary width of transistors. Variation of width of a single transistor affects changes of parasite time factor and makes optimisation more difficult. Coupling of several transistors per switch results in additional useful effects: averaging of technological variances, improvement of noise

parameters. The space, occupied by several transistor with summary width W does not differs essentially from the space occupied by one transistor with the same width W.

The simulation has shown that optimal compensation of charge-injection is achieved if the quantities of transistors used to realise different switches correspond to the ratio

$$M_{n-1}:M_{n-2}:M_{n-3}:M_{n-4}: \ldots\ldots = M_{n-1}:(M_{n-1}-1):(M_{n-1}-2):(M_{n-1}-2):\ldots, \quad (2)$$

where M_i — the quantity of tied MOS-transistors tied in parallel in the switch of the i-th bit.

Fig. 6 present U_{rd} as a function M_{n-1} for the case of usage both shunting capacitors and different quantity of MOS transistors per switch (both parameters optimised). The accuracy convenient to realise 16-bit ADC is achieved at $M_{n-1} \geqslant 3$ with M_i defined as 1 if (2) gives smaller value. Glitches at V-node in such circuit configuration are practically eliminated (see fig. 3).

CONCLUSION

The problem of charge-injection compensation in charge-distribution ADC-s must be solved using the complex of circuit-design methods. The use of optimally chosen shunting capacitors and quantities of parallel tied switching MOS-transistors allows the significant improvement of conversion accuracy. Combining these techniques with self-calibration makes possible the realisation of 16–18 bit ADC with 5–10 μs conversion time, having excellent stability in a wide range of external conditions.

REFERENCES

1. F. Goodenaugh: *High resolution ADCs up dynamic range in more aplications.* Electronic Design, Apr. 1991, v. 37, No 7, pp. 65–79
2. B.M. Gordon: *Linear Electronic Analog/Digital Conversion Architecture. Their Origins, Parameters, Limitations and Applications.* IEEE Trans. on Circuits and Systems, Vol. CAS-25, No. 7, July 1978, pp. 391–418
3. Z. Kulka, A. Libura, M. Nadachowski: *Przetworniki analogowo-cyfrowe i cyfrowo-analogowe.* WKL, Warszawa 1987
4. H. Dajkiti, A. Tosio, A. Hidze: *Calibrating capacitors in IC.* Pat. 55-43620 Japan, MKI HO1I 27/01, HO1G 4/40, 1990
5. M. Risemann: *High-resolution ADC design. Problems and perspectives.* Electronics, 1988, vol. 61, No 4, pp. 62–72
6. H.S. Lee, D.A. Hodges, P.R. Gray: *A self-calibrating 15-bit CMOS a/d converter.* IEEE J. of Solid State Circuits, v. 19, No 6, 1984, pp. 813–819
7. Data acquisition seminar application book. Crystal Semiconductor Corp., 1991

8. A. Mursajew, R. Gruszwicki, A. Czajkowski: *Strukturalne metody minimalizacji błędów przetworników analogowo-cyfrowych i cyfrowo-analogowych*. Kwartaln. Elektroniki i Telekomunikacji, v. 39, 1993, No 4, pp. 93 – 107
9. J. Mulawka: *Układy mikroelektroniczne z przełączanymi pojemnościami*. WKL, Warszawa 1987
10. C.C. Lee: *A new switched-capacitor realisation for cyclic analog-to-digital converter*. Proc. ISCAS'83, Newport, CA, 1983, pp. 1261 – 1265
11. C. Eichenbegger, W. Guggenbuhl: *Dummy transistor compensation of analog MOS switches*. IEEE J. Solid-State Circuits, 24, No 4, 1989, pp. 1143 – 1146
12. G. Wegmann, E. Vittoz, F. Rahali: *Charge injection in analog MOS switches*. IEEE j. Solid-State Circuits, v. 22, No 6, 1987, pp. 1091 – 1097
13. J. Dobrynin, B. Manchew, A. Mursajew: *Compensation of glitches in capacitor analog memory*. Kontrolno-izmeritelnaja technika, v. 32, Lwów, Wyższa Szkoła, No 9, 1982, pp. 137 – 141 (in Russian)
14. J. Kuo, R. Duttom, B. Wooley: *MOS pass transistor turn-off transient analysis*. IEEE Trans. Electron Devices, Vol. ED-33, No 10, 1986, pp. 1545 – 1555
15. P. Allen, E. Sanches-Sinencio: *Switched capacitor circuits*. VNR Company, New York 1984

A. MURSAJEW, R. GRUSZWICKI, A. SIDOROW

MINIMALIZACJA BŁĘDÓW PRZEŁADOWANIA POJEMNOŚCI W PRZETWORNIKACH ANALOGO-CYFROWYCH Z DYSTRYBUCJĄ ŁADUNKU

Streszczenie

Injekcja ładunków przy przełączeniach istotnie wpływa na dokładność układów z pojemnościami przełączanymi. Artykuł przedstawia porównanie metod kompensacji injekcji ładunków w przetwornikach analogo-cyfrowych z pojemnościami przełączanymi. Wyniki symulacji udowodniły, że rozdzielczość o 16-18 bitów może być osiągnięta z wykorzystaniem optymalnie wybranych parametrów przełączników oraz pojemności, które bocznikują te przełączniki.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1994. 40, z. 4. ss. 541–565

621.396.6

Obciążenia elementów pierścieni SDH ruchem telekomunikacyjnym

ANDRZEJ DROBNIK

Instytut Telekomunikacji, Politechnika Warszawska

Otrzymano 1994.09.03

Autoryzowano do druku 1994.11.08

W opracowaniu przedstawiono zależności analityczne, na podstawie których można wyznaczyć obciążenia elementów pojedynczych pierścieni SDH ruchem telekomunikacyjnym. Spośród struktur pierścieniowych SDH rozpatrzono pierścień jednokierunkowy (USHR) oraz dwukierunkowy o 2 i 4 włóknach (BSHR-2 i BSHR-4). Dla każdego z tych pierścieni wyznaczone zostały obciążenia jego gałęzi oraz wielkości ruchu transferowego i tranzytowego w każdym z węzłów dla ogólnej postaci macierzy zainteresowań oraz dla trzech szczególnych jej przypadków, obejmujących ruch jednorodny, scentralizowany oraz szeregowy. Omówiono preferencje aplikacyjne poszczególnych struktur pierścieniowych w sieciach SDH.

1. WSTĘP

W nowoczesnych sieciach telekomunikacyjnych coraz powszechniej, by nie powiedzieć jedynie, wykorzystuje się systemy synchronicznej hierarchii cyfrowej SDH (Synchronous Digital Hierarchy). Stosowane są one na ogół wraz z łączami światłowodowymi, stanowiącymi w tych systemach najbardziej typowe medium transmisyjne.

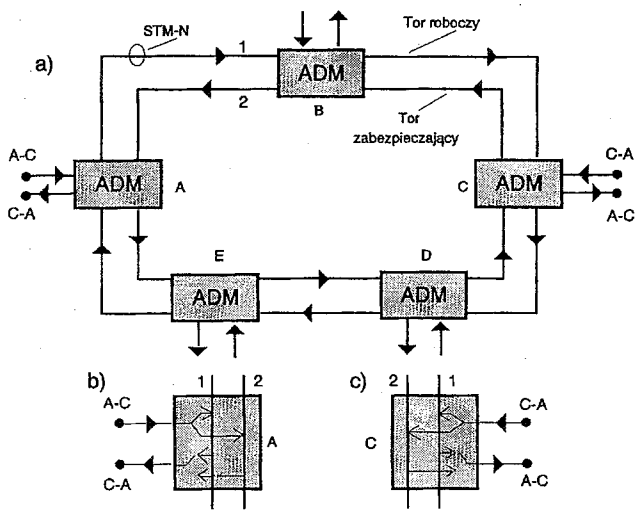
Wprowadzenie systemów SDH zaowocowało nie tylko powstaniem nowej generacji urządzeń telekomunikacyjnych, lecz także opracowaniem i wykorzystaniem nowych topologii sieci. Do tych, które nabrały szczególnego znaczenia i rozpowszechnienia w sieciach SDH w konsekwencji narzucanych na te sieci wysokich wymagań w zakresie ich niezawodności przy jednoczesnym dążeniu do jak najdalej idącego obniżenia ich kosztów, należą struktury pierścieniowe. Ich właściwości, a w tym i zalety, zostały szczegółowo omówione w licznych publikacjach [1], [2], [3], [4], [8], [9], [10], [11].

W niniejszym opracowaniu nasze zainteresowanie dotyczyć będzie wszystkich najpowszechniej stosowanych konfiguracji struktur pierścieniowych. Należą do nich:

— pierścień jednokierunkowy 2-włóknowy;
— pierścień dwukierunkowy 2-włóknowy;
— pierścień dwukierunkowy 4-włóknowy.
Poniżej przedstawione zostanie ich krótkie omówienie.

Pierścień jednokierunkowy 2-włóknowy

Struktura pierścienia jednokierunkowego (USHR — Unidirectional Self-Healing Ring — jednokierunkowy pierścień samonaprawialny) pokazana jest na rys. 1a. Charakterystyczną jej cechą jest jednokierunkowość transmisji sygnałów zarówno nadawczych, jak i odbiorczych. W pierścieniu występują dwa niezależne tory transmisyjne: tor roboczy (1) i tor zabezpieczający (2). Przy transmisji między węzłami A i C — jak pokazano to na rysunku — ruch nadawczy kierowany jest z węzła A do C torem roboczym, zgodnie z kierunkiem ruchu wskazówek zegara. Ruch odbiorczy połączenia A – C, tzn. ruch z węzła C do A, kierowany jest w pierścieniu torem roboczym również w tym samym kierunku, drogą przechodzącą przez węzły CDEA, stanowiącą dopełnienie drogi z A do C.



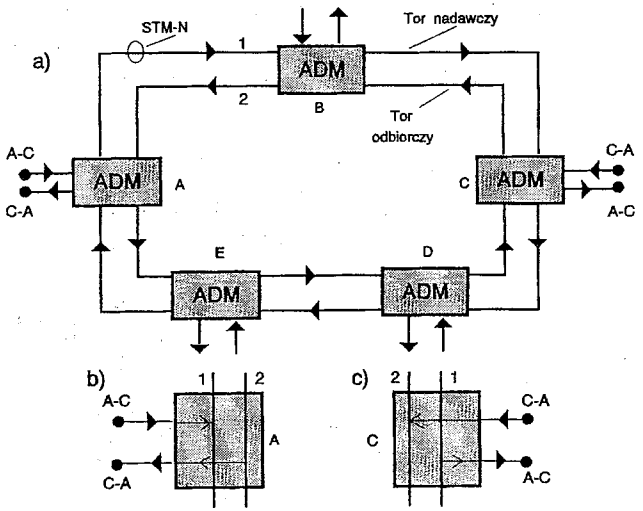
Rys. 1. Pierścien jednokierunkowy 2-włóknowy
a) struktura
b) nadawanie i odbiór sygnałów w węźle A
c) nadawanie i odbiór sygnałów w węźle C

Dla podwyższenia niezawodności ruch w kierunku nadawczym jest kierowany nie tylko do toru roboczego (1), lecz także i do zabezpieczającego (2). W tym drugim przypadku jest on jednak kierowany przeciwnie do ruchu wskazówek zegara. W końcowym węźle odbiorczym sygnały docierające dwoma niezależnymi drogami są ze sobą wzajemnie porównywane w oparciu o określone

kryteria jakości transmisji. W sposób w pełni zautomatyzowany wybierany jest lepszy z sygnałów. Sposób wprowadzania i odbioru sygnału w węźle początkowym A oraz końcowym C pokazany jest na rysunkach 1b i 1c.

Pierścień dwukierunkowy 2-włóknowy

Pierścień dwukierunkowy (BSHR — Bidirectional Self-Healing Ring — dwukierunkowy pierścień samonaprawialny bądź SPR lub SP SHR — Shared Protection Ring — pierścień z dzielonym zabezpieczeniem lub rezerwą) 2-włóknowy różni się od pierścienia jednokierunkowego sposobem wykorzystania obydwu torów. Jak pokazano na rys. 2, jeden z nich jest torem nadawczym, a drugi odbiorczym.



Rys. 2. Pierścien dwukierunkowy 2-włóknowy
a) struktura
b) nadawanie i odbiór sygnałów w węźle A
c) nadawanie i odbiór sygnałów w węźle C

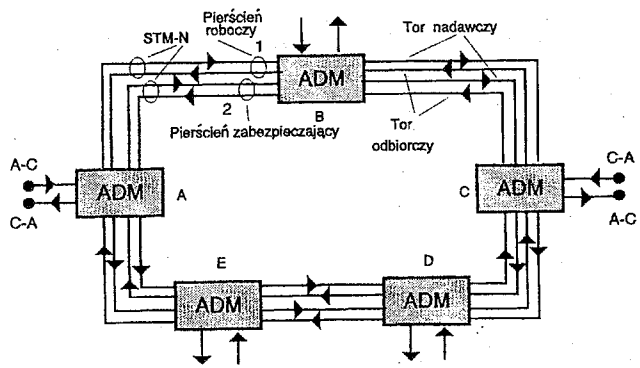
Sygnały nadawcze i odbiorcze krążą w pierścieniu dwukierunkowym, w odróżnieniu od pierścienia USHR, w przeciwnych kierunkach. Inny jest również w tym pierścieniu mechanizm zabezpieczenia przed awarią toru transmisyjnego lub węzła. Pasmo przepustowe dzielone jest na dwie, zazwyczaj równe części: roboczą i zabezpieczającą. W stanie awaryjnym po obu stronach uszkodzonego przęsła lub węzła dokonuje się odpowiednio przełączeń. Mają one na celu utworzenie pętli zwrotnej omijającej uszkodzony element oraz wprowadzenie dopływających strumieni do zabezpieczających części pasma w torze o przeciwnym kierunku.

Inna organizacja transmisji sygnałów w pierścieniu BSHR powoduje na ogół efektywniejsze wykorzystanie jego elementów niż w pierścieniu USHR. Obsługa

ruchu bowiem między poszczególnymi parami węzłów (w szczególności sąsiednimi) angażuje w pierścieniu dwukierunkowym jedynie część tego pierścienia, podczas gdy w przypadku pierścienia USHR ruch ten obsługiwany jest przez wszystkie węzły pierścienia, niezależnie od ich wzajemnego położenia.

Pierścień dwukierunkowy 4-włóknowy

Pierścień dwukierunkowy 4-włóknowy można traktować jako strukturę sieciową złożoną z dwu pierścieni dwukierunkowych 2-włóknowych, z których jeden stanowi pierścień roboczy, a drugi — zabezpieczający. W każdym węźle pierścienia występują dwie pary torów nadawczo-odbiorczych: robocza i zabezpieczająca. W odróżnieniu jednak od pierścienia 2-włóknowego w 4-włóknowym nie zachodzi potrzeba przeznaczenia połowy pasma na zabezpieczenie. Jest ono bowiem realizowane za pośrednictwem pierścienia zabezpieczającego. Architektura pierścienia dwukierunkowego 4-włóknowgo pokazana jest na rys. 3.



Rys. 3. Pierścień dwukierunkowy 4-włóknowy

Pierścień dwukierunkowy 4-włóknowy bywa w literaturze określany skrótem BSHR-4 w odróżnieniu od pierścienia 2-włóknowego, w stosunku do którego używa się skrótu BSHR-2.

2. ZAŁOŻENIA WSTĘPNE

Niniejsze opracowanie poświęcone jest problematyce obciążeń ruchem telekomunikacyjnym elementów pojedynczego pierścienia SDH w warunkach realizacji określonych, zadanych *a priori* zapotrzebowań na usługi telekomunikacyjne między poszczególnymi węzłami pierścienia. Pod używanym tutaj i niżej pojęciem ruchu rozumiany będzie strumień o określonej przepływności przenoszony przez element sieci lub zaalokowany połączeniu w danej relacji międzywęzłowej.

W dalszym ciągu opracowania zostanie przyjęte założenie, że zapotrzebowania na usługi telekomunikacyjne określone są przez macierz zainteresowań:

$$\mathbf{T}=[t_{ij}]. \tag{1}$$

Elementy t_{ij} macierzy zainteresowań określają wielkość strumieni między węzłami i i j wyrażoną liczbą przyjętych jednostek, np. liczbą strumieni odpowiadających kontenerom wirtualnym VC12. Macierz określona przez (1) jest macierzą kwadratową o wymiarach $\mathbf{N}\times\mathbf{N}$, jeśli przez $\mathbf{N}$ oznaczona zostanie liczba węzłów pierścienia.

Zakładać się będzie symetrię zainteresowań między poszczególnymi węzłami pierścienia, jak również ich brak wewnątrz każdego z węzłów. Założenia te znajdują pełne uzasadnienie w praktyce. Wyrażają się one następującymi zależnościami:

$$t_{ij}=t_{ji} \begin{cases} \forall\ i \in \{1,2,...,N\} \\ \forall\ j \in \{1,2,...,N\} \end{cases} \tag{2}$$

oraz

$$t_{ij}=0 \qquad \forall \quad i=j. \tag{3}$$

Całkowite zapotrzebowanie T_{tot} na ruch telekomunikacyjny między wszystkimi węzłami pierścienia dla jednego kierunku transmisji określone jest, przy spełnieniu ograniczeń (2) i (3), równaniem:

$$T_{tot}=\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} t_{ij}. \tag{4}$$

Natomiast wielkość całkowitego zapotrzebowania T_{tot_2} na ruch dla obu kierunków transmisji ze względu na symetrię macierzy zainteresowań wyrażoną równaniem (2) jest równa:

$$T_{tot\_2}=2\ T_{tot}.$$

Z uwagi na rozłączność elementów pierścienia obsługujących oba kierunki transmisji w poszczególnych relacjach połączeń międzywęzłowych wygodniejszą spośród tych dwu wielkości całkowitego zapotrzebowania do stosowania w praktyce jest wielkość T_{tot} odniesiona do jednego kierunku transmisji.

Wielkością determinującą możliwości transmisyjne pierścienia jest jego pojemność (przepustowość) C. Definiowana jest ona jako maksymalna wielkość ruchu telekomunikacyjnego, który może być przesłany przez pojedyncze przęsło pierścienia, tzn. odcinek między sąsiednimi jego węzłami. W zgodzie z realiami praktycznymi zakłada się równość pojemności wszystkich przęseł.

Dla oceny efektywności wykorzystania pojemności C pierścienia dla obsługi całkowitego ruchu stosowana będzie wielkość określona mianem współczynnika efektywności pierścienia **WEP**, który to współczynnik zdefiniowany zostanie jako:

$$\mathbf{WEP}=\frac{T_{tot}}{C}. \tag{5}$$

Współczynnik ten określa zatem relatywną wielkość całkowitego ruchu telekomunikacyjnego w jednym kierunku transmisji, który może być przeniesiony przez pierścień o danej przepustowości.

Obciążenia elementów pierścienia uzależnione są przede wszystkim od jego struktury oraz wielkości i rozkładu wartości elementów macierzy zainteresowań. W analizie będą brane pod uwagę wspomniane już wyżej struktury USHR oraz BSHR-2 i BSHR-4. Dla tych struktur pierścieni SDH obciążenia ich elementów będą rozpatrywane dla 4 przypadków postaci ruchu telekomunikacyjnego. Jeden z nich jest przypadkiem ogólnym, pozostałe zaś szczególnymi. Oto one:

a) przypadek ogólny, w którym na wielkości strumieni między poszczególnymi węzłami pierścienia — a zatem i na elementy t_{ij} macierzy zainteresowań — nie będą narzucane żadne, poza określonymi przez zależności (2) i (3), ograniczenia; nie narzuca się więc ograniczeń ani na strukturę topologiczną, ani na wielkość zainteresowań w obrębie pierścienia;

b) przypadek ruchu jednorodnego, tzn. takiego, w którym ruch telekomunikacyjny występuje między wszystkimi parami węzłów struktury, a ograniczenia dotyczą jedynie jego wielkości — mianowicie ruch między poszczególnymi parami węzłów jest jednakowy, a w szczególnym przypadku — jednostkowy; dla tak wyidealizowanej postaci ruchu elementy macierzy zainteresowań są określone następująco:

$$t_{ij}=\begin{cases}1 & \begin{cases}\forall\ i \in \{1,2,...,N\},\\ \forall\ j \in \{1,2,...,N\},\end{cases} \quad i\neq j\\ 0 & \forall\ i=j\end{cases} \tag{6}$$

c) przypadek ruchu scentralizowanego — takiego, w którym dochodzi do transmisji sygnałów jedynie między pewnym węzłem centralnym pierścienia c, a wszystkimi pozostałymi;

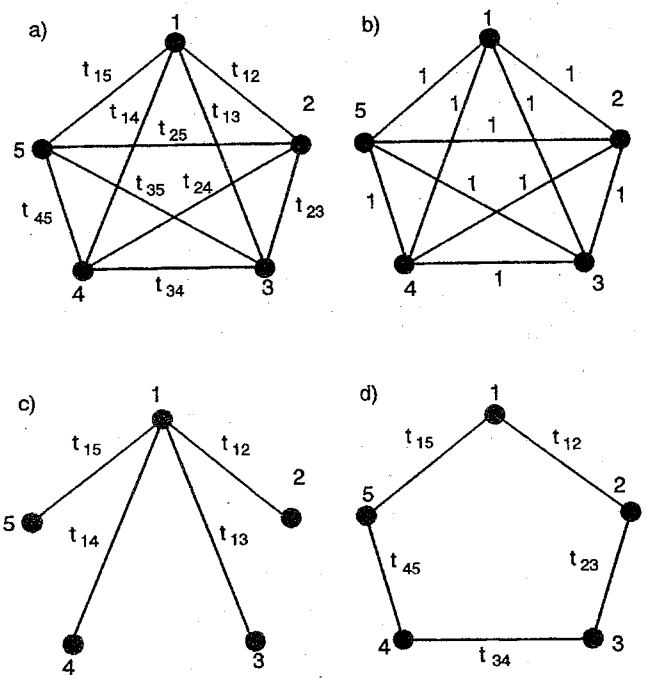
$$t_{ij}=\begin{cases}t_{ij} & \forall \begin{cases}i=c \bigwedge j\neq c\\ i\neq c \bigwedge j=c\end{cases}\\ 0 & \forall\ \ i\neq c \bigwedge j\neq c\end{cases} \tag{7}$$

d) przypadek ruchu szeregowego, kiedy zapotrzebowanie na usługi telekomunikacyjne w obrębie pierścienia dotyczy jedynie węzłów sąsiednich w stosunku do siebie; można to zapisać w postaci:

$$\forall\ i \in \{1,2,...,N\}$$

$$t_{ij}=\begin{cases} t_{ij} & \forall\ j=\begin{cases}1+i \bmod N \bigwedge\\ 1+(N+i-2) \bmod N\end{cases}\\ \\ 0 & \forall\ j\neq\begin{cases}1=i \bmod N \bigwedge\\ 1+(N+i-2) \bmod N\end{cases}\end{cases} \tag{8}$$

Sposób kształtowania się ruchu między węzłami pierścienia dla powyższych przypadków zilustrowany jest na rys. 4



Rys. 4. Warianty kształtowania ruchu telekomunikacyjnego a) przypadek ogólny, b) ruch jednorodny, c) ruch scentralizowany, d) ruch szeregowy

Dotychczas mówiąc o obciążeniach elementów pierścienia nie precyzowano, o jakie elementy chodzi. Z punktu widzenia projektanta systemu SDH istotne są obciążenia gałęzi sieci, określające minimalne przepustowości — w przypadku pierścieni — jego przęseł, oraz obciążenia węzłów, a ściślej rzecz ujmując — minimalne niezbędne przepustowości krotnic transferowych typu ADM (Add-Drop Multiplexer), jako urządzeń instalowanych na ogół w poszczególnych węzłach pierścienia.

W tym drugim przypadku istotne jest nie tylko ogólne obciążenie węzła, tzn. wielkość całkowitego ruchu przezeń przechodzącego, ale także wyróżnienie w nim składnikow pochodzących od ruchu tranzytowego i transferowego. W przpadku ruchu transferowego może być istotną informacja o wielkości tego ruchu w obydwu kierunkach pierścienia, z uwagi na to, że pozwala to w niektórych konfiguracjach pierścienia efektywniej wykorzystywać przepustowość zainstalowanego systemu. Obciążenia krotnicy węzła m ruchem transferowym dla obu kierunków będą oznaczane symbolami $L_{ADM-L,m}$ i $L_{ADM-P,M}$, natomiast wynikające z ruchu tranzytowego — symbolem $L_{ADM-T,M}$.

W niniejszym opracowaniu zakładać się będzie rozłączność szczelin czasowych wykorzystywanych dla obsługi ruchu w różnych relacjach. Dyktowane to

jest rzadkością występowania w praktyce przypadków tak szczególnych postaci ruchu telekomunikacyjnego, iż powyższe założenie byłoby dla nich niesłuszne. Założenie to implikuje rozłączność ruchu transferowego i tranzytowego w każdym z węzłów, co oznacza, iż ten sam strumień danych nie podlega w węźle lub węzłach pierścienia jednocześnie transferowaniu i przesyłaniu go do co najmniej jednego następnego węzła.

Zestawienie stosowanych oznaczeń:

N — liczba węzłów pierścienia o minimalnej wartości równej 3;

C — minimalna pojemność pierścienia;

$L_{ADM-L,M}$ — wielkość strumienia transferowanego w/z kierunku przeciwnego w stosunku do kierunku ruchu wskazówek zegara w m-tym węźle pierścienia;

$L_{ADM-P,m}$ — wielkość strumienia transferowanego w/z kierunku zgodnego z kierunkiem ruchu wskazówek zegara w m-tym węźle pierścienia;

$L_{ADM-T,m}$ — wielkość strumienia tranzytowego przenoszonego przez m-ty węzeł pierścienia;

L_m — obciążenie ruchem telekomunikacyjnym m-tej gałęzi pierścienia, tzn. wielkość strumienia przez nią przesyłanego;

m-ta gałąź — gałąź między węzłem $w=m$ i węzłem v określonym następująco:

$$v=\begin{cases} w+1 & \text{dla } w<N \\ 1 & \text{dla } w=N \end{cases}$$

gdzie:

$$m \in \{1,2,...,N\},$$
$$w \in \{1,2,...,N\},$$
$$v \in \{1,2,...,N\}.$$

T — macierz zainteresowań;

t_{ij} — wielkość strumienia zaalokowanego połączeniu między węzłami i i j;

T_{tot} — całkowite zapotrzebowanie na ruch telekomunikacyjny między wszystkimi parami węzłów pierścienia dla jednego kierunku transmisji;

WEP — współczynnik efektywności pierścienia.

3. OBCIĄŻENIA RUCHEM W PIERŚCIENIU JEDNOKIERUNKOWYM

W pierścieniu jednokierunkowym ruch dopływający do każdego z węzłów musi być przeniesiony przez cały pierścień, niezależnie od wzajemnego usytuowania węzłów w określonej relacji połączenia. Dlatego też obciążenia poszczególnych elementów pierścienia będą niezależne od ich położenia w pierścieniu. W konsekwencji, obciążenia wszystkich gałęzi czy też węzłów pierścienia USHR będą sobie równe. Poniżej przedstawione zostaną zależności określające obciążenia gałęzi oraz węzłów pierścienia USHR.

1) Przypadek ogólny ruchu telekomunikacyjnego

Obciążenia gałęzi, pojemność pierścienia oraz całkowitą wielkość ruchu w pierścieniu dla jednego kierunku transmisji będzie można wyznaczyć z poniższej zależności:

$$L_m = C = T_{tot} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} t_{ij}. \tag{9}$$

Wielkość ruchu transferowego w węźle m jest wyznaczona przez:

$$L_{\mathrm{ADM-L},m} = L_{\mathrm{ADM-P},m} = \sum_{j=1}^{N-1} t_{ml} \tag{10}$$

gdzie:

$$l = \begin{cases} m+j & \forall\ j \leqslant N-m \\ m+j-N & \forall\ j > N-m \end{cases}. \tag{11}$$

Wielkość strumienia ruchu tranzytowego w węźle:

$$L_{\mathrm{ADM-T},m} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} t_{ij} \qquad \forall\ \{i \neq m \wedge j \neq m\}. \tag{12}$$

2) Przypadek ruchu jednorodnego

Obciążenia gałęzi, pojemność pierścienia i wielkość T_{tot}:

$$L_m = C = T_{tot} = \frac{N(N-1)}{2}. \tag{13}$$

Ruch transferowy:

$$L_{\mathrm{ADM-L},m} = L_{\mathrm{ADM-P},m} = N-1. \tag{14}$$

Ruch tranzytowy:

$$L_{\mathrm{ADM-T},m} = \binom{N-1}{2} = \frac{(N-1)(N-2)}{2}. \tag{15}$$

3) Przypadek ruchu scentralizowanego

Obciążenia gałęzi pierścienia, pojemność pierścienia i całkowite zapotrzebowanie są równe:

$$L_m = C = T_{tot} = \sum_{j=1}^{N-1} t_{cl}. \tag{16}$$

gdzie c jest węzłem centralnym, a l określone jest przez:

$$l = 1 + (c+j-1) \bmod N \tag{17}$$

W warunkach „ujednorodnienia" ruchu scentralizowanego uzyskuje się:

$$L_m = C = T_{tot} = N - 1. \tag{18}$$

Wielkość ruchu transferowego w węźle m zależy od tego, czy jest on węzłem centralnym c, czy też nie. W pierwszym przypadku, tzn. gdy $m = c$ będziemy mieli:

$$L_{\mathrm{ADM-L},c} = L_{\mathrm{ADM-P},c} = \sum_{j=1}^{N-1} t_{cl}, \tag{19}$$

gdzie l określone jest zależnością (17).

W przypadku, gdy węzeł m nie jest węzłem centralnym c ($m \neq c$), ruch transferowy jest równy:

$$L_{\mathrm{ADM-L},m} = L_{\mathrm{ADM-P},m} = t_{mc} \quad \forall\ m \neq c. \tag{20}$$

Wielkość ruchu tranzytowego również jest uzależniona od położenia węzła m względem węzła centralnego c. W przypadku, gdy węzły te są różne, tzn. $m \neq c$:

$$L_{\mathrm{ADM-T},m} = \sum_{i=1}^{N-1} t_{ck} \qquad \forall\{m \neq c \wedge k \neq m\}, \tag{21}$$

gdzie:

$$k = \begin{cases} m+i & \forall\ i \leqslant N-m \\ m+i-N & \forall\ i > N-m \end{cases} \tag{22}$$

Brak jest natomiast ruchu tranzytowego w węźle centralnym pierścienia USHR, tzn. $L_{\mathrm{ADM-T},c} = 0$.

4) Przypadek ruchu szeregowego

Obciążenia gałęzi, pojemność pierścienia i całkowite zapotrzebowanie na ruch są równe:
gdzie:

$$L_m = C = T_{tot} = \sum_{i=1}^{N} t_{ik}. \tag{23}$$

$$k = 1 + i \bmod N$$

Dla jednorodnej postaci ruchu będzie:

$$L_m = C = T_{tot} = N \tag{24}$$

Ruch transferowy:

$$L_{\mathrm{ADM-L},m} = L_{\mathrm{ADM-P},m} = t_{mi} + t_{mj}, \tag{25}$$

gdzie:

$$i = \begin{cases} m-1 & \forall\ m > 1 \\ N & \forall\ m = 1 \end{cases}$$

$$j = \begin{cases} m+1 & \forall\ m < N \\ 1 & \forall\ m = N \end{cases} \tag{26}$$

Ruch tranzytowy:

$$L_{\mathrm{ADM-T},m}=\begin{cases}\sum_{i=1}^{N-2} t_{jk} & \text{dla} \quad N>3\\ t_{jk} & \text{dla} \quad N=3\end{cases} \tag{27}$$

gdzie:

$$j=\begin{cases}m+i & \forall\ i\leqslant N-m\\ m+i-N & \forall\ i>N-m\end{cases}$$

$$k=\begin{cases}m+i+1 & \forall\ i\leqslant N-m-1\\ m+i+1-N & \forall\ i>N-m-1\end{cases} \tag{28}$$

Oczywisty jest fakt, że we wszystkich wyszczególnionych wyżej przypadkach współczynnik efektywności pierścienia USHR WEP = 1. Dla każdego z przypadków postaci ruchu telekomunikacyjnego obciążenia poszczególnych gałęzi pierścienia USHR są niezależne od położenia tych gałęzi w pierścieniu, tzn. zachodzi relacja:

$$L_1=L_2=\ \ldots\ =L_m=\ \ldots\ =L_N \tag{29}$$

4. OBCIĄŻENIA RUCHEM W PIERŚCIENIU DWUKIERUNKOWYM

W pierścieniu dwukierunkowym — w odróżnieniu od jednokierunkowego — ruch między dowolną parą węzłów nie przechodzi tranzytem przez wszystkie pozostałe węzły pierścienia. Ponadto, w stosunku do podstawowej drogi połączenia dwu węzłów, istnieje zawsze droga alternatywna, stanowiąca dopełnienie tej pierwszej. Wynika to z faktu, że rozróżnienie na tor nadawczy i odbiorczy, pokazane na rysunkach 2 i 3, ma charakter wyłącznie logiczny, a nie fizyczny. Bez trudu można wyobrazić sobie i zrealizować sytuację, w której role tych torów na pewnych odcinkach są przeciwne. Dlatego też zawsze możliwe jest połączenie węzłów A-C z rysunków 2 i 3 drogą A-B-C lub A-E-D-C, bądź tymi obydwoma drogami. W związku z tym na obciążenia elementów pierścieni BSHR wpływ będą miały nie tylko zainteresowania między poszczególnymi parami węzłów, ale także i sposób wyboru drogi podstawowej oraz — ewentualnie — rozdziału ruchu na obie drogi.

Znanych jest kilka algorytmów kierowania ruchem w pierścieniu BSHR [5], [6], [7]. Do najbardziej rozpowszechnionych ze względu na swoją prostotę i efektywność należy algorytm MHR (Minimum Hop Routing)[1]. Zakłada on połączenie między dwoma węzłami drogą najkrótszą. Miarą długości drogi jest liczba węzłów pośrednich, przez które droga przechodzi. Określenie w ten sposob zdefiniowanej drogi najkrótszej jest jednoznaczne dla nieparzystej

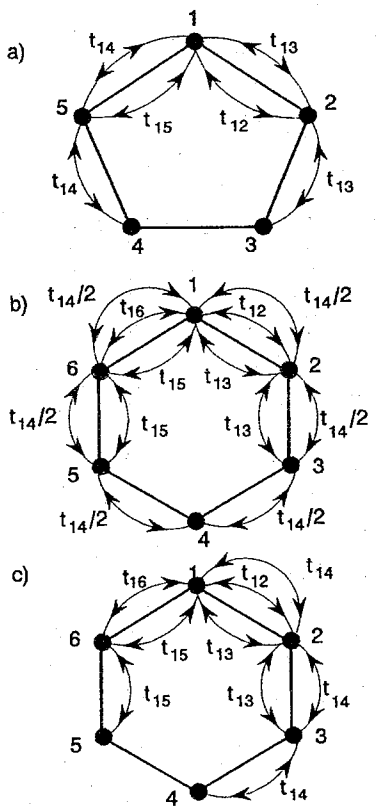
[1] W literaturze anglojęzycznej algorytm MHR bywa również określany mianem algorytmu SPR (Shortes Path Routing).

wartości N. Dla parzystej wartości N istnieją jednak zawsze dwie drogi o równej sobie długości do węzła przeciwległego w stosunku do węzła początkowego drogi. Algorytm MHR ma z tego powodu 2 warianty:

a) — z rozdziałem ruchu — ruch między danym węzłem a węzłem przeciwległym jest dzielony na dwa równe lub bliskie sobie strumienie i przesyłany obydwoma równorzędnymi drogami;

b) — bez rozdziału ruchu — ruch do węzła przeciwległego kierowany jest drogą podstawową.

Poniżej podane zostaną zależności określające wielkość obciążenia ruchem telekomunikacyjnym gałęzi L_m pierścienia oraz jego węzłów, przy sterowaniu tym ruchem w oparciu o algorytm MHR. Zależności te zostaną podane dla wariantu z rodziałem ruchu. Ich modyfikacja dla przypadku bez rozdziału ruchu może być dokonana bez większych trudności.



Rys. 5. Obsługa ruchu telekomunikacyjnego w/g algorytmu MHR (pokazano połączenia tylko węzła 1)

a) N — nieparzyste (N = 5),
b) N — parzyste (N = 6) z podziałem ruchu 1 – 4
c) N — parzyste (N = 6) bez podziału ruchu 1 – 4

Wielkość obciążeń elementów pierścienia wyznaczane będą dla pierścienia BSHR-4. Różnice bowiem między pierścieniami BSHR-4 i BSHR-2 sprowadzają się przede wszystkim do sposobu zabezpieczenia i konieczności przeznaczenia na ten cel w pierścieniu BSHR-2 połowy przepustowości systemu.

1) Przypadek ruchu bez ograniczeń jego struktury i wartości

Dla nieparzystej wartości N:
Obciążenie gałęzi:

$$L_m = \sum_{i=1}^{\frac{N-1}{2}} \sum_{j=1}^{i} t_{kl}, \tag{30}$$

gdzie:

$$k = \begin{cases} m+i-\dfrac{N-1}{2} & \forall\ i > \dfrac{N-1}{2} - m \\ m+i+\dfrac{N+1}{2} & \forall\ i \leqslant \dfrac{N-1}{2} - m \end{cases} \tag{31}$$

$$l = \begin{cases} m+j & \forall\ j \leqslant N-m \\ m+j-N & \forall\ j > N-m. \end{cases}$$

Ruch transferowy wyznaczony będzie przez:

$$L_{\mathrm{ADM-L},m} = \sum_{i=1}^{\frac{N-1}{2}} t_{km} \tag{32a}$$

oraz

$$L_{\mathrm{ADM-P},m} = \sum_{i=1}^{\frac{N-1}{2}} t_{ml}, \tag{32b}$$

gdzie:

$$k = \begin{cases} m-i & \forall\ i \leqslant m-1 \\ m-i+N & \forall\ i > m-1 \end{cases}$$

$$l = \begin{cases} m+i & \forall\ i \leqslant N-m \\ m+i-N & \forall\ i > N-m. \end{cases} \tag{33}$$

Ruch tranzytowy w węźle m dla pierścienia BSHR o liczbie węzłów $N>3$ będzie równy:

$$L_{\mathrm{ADM-T},m} = \sum_{i=1}^{\frac{N-1}{2}} \sum_{j=1}^{i} t_{kl} \tag{34}$$

gdzie k i l jest dane przez (33).

W trójwęzłowym pierścieniu BSHR ruch tranzytowy nie występuje w żadnym z węzłów.

Dla pierścienia o parzystej liczbie gałęzi N odpowiednie zależności będą miały postać następującą:

$$L_m = \sum_{i=1}^{\frac{N}{2}-1} \sum_{j=1}^{i} t_{k'l} + \frac{1}{2} \sum_{\substack{i=1 \\ j=1}}^{\frac{N}{2}} t_{k''l} \tag{35}$$

gdzie:

$$k' = \begin{cases} m+i+1-\dfrac{N}{2} & \forall\ i > \dfrac{N}{2}+1-m \\ m+i+1+\dfrac{N}{2} & \forall\ i \leqslant \dfrac{N}{2}+1-m \end{cases} \tag{36}$$

$$k'' = \begin{cases} m+i+\dfrac{N}{2} & \forall\ i \leqslant \dfrac{N}{2}-m \\ m+i-\dfrac{N}{2} & \forall\ i > \dfrac{N}{2}-m \end{cases} \tag{37}$$

$$l = \begin{cases} m+j & \forall\ j \leqslant N-m \\ m+j-N & \forall\ j > N-m\,. \end{cases} \tag{38}$$

Ruch transferowy będzie równy:
gdzie k'' określone jest zależnością (37), zaś:

$$L_{\mathrm{ADM-L},m} = \sum_{i=1}^{\frac{N}{2}-1} t_{k''m} + \frac{1}{2} t_{k'''m}, \tag{39}$$

gdzie k'' określone jest zależnością (37), zaś:

$$k''' = \begin{cases} m+\dfrac{N}{2} & \forall\ m \leqslant \dfrac{N}{2} \\ m-\dfrac{N}{2} & \forall\ m > \dfrac{N}{2} \end{cases} \tag{40}$$

oraz

$$L_{\mathrm{ADM-P},m} = \sum_{i=1}^{\frac{N}{2}-1} t_{ml} + \frac{1}{2} t_{k'''m}, \tag{41}$$

gdzie l określone jest przez (33), a k''' — przez (40).

Wielkość obciążenia ruchem tranzytowym wynosi:

$$L_{\mathrm{ADM-T},m} = \sum_{i=1}^{\frac{N}{2}-2} \sum_{j=1}^{i} t_{k't} + \frac{1}{2} \sum_{i=1}^{\frac{N}{2}-1} t_{k''l}, \tag{42}$$

gdzie k', k'' i l są wyznaczane z zależności (36), (37) i (38).

Dla określenia minimalnej przepustowości pierścienia C niezbędne jest wyznaczenie maksymalnego obciążenia jego gałęzi $L_{\max}$.

$$L_{\max}=\max_{m}(L_m). \tag{43}$$

Przepustowości pierścienia dwukierunkowego 4-oraz 2-włóknowego, odpowiednio C_{BSHR-4} oraz C_{BSHR-2} będą określone z zależnościami:

$$\begin{gathered} C_{BSHR-4}\geqslant L_{\max} \\ C_{BSHR-2}=2\ C_{BSHR-4}\geqslant 2\ L_{\max} \end{gathered} \tag{44}$$

Przepustowość pierścienia BSHR-2 musi być zatem 2-krotnie większa od przepustowości pierścienia BSHR-4. Wynika to z konieczności przeznaczenia, dla potrzeb zabezpieczenia systemu przed awarią, pasma o szerokości równej szerokości pasma roboczego.

2) Przypadek ruchu jednorodnego

Całkowite zapotrzebowanie na ruch T_{tot} w pierścieniu BSHR jest takie samo jak w pierścieniu USHR i wyraża się w związku z tym zależnością (13). Dla ruchu jednorodnego obciążenia elementów pierścienia dwukierunkowego będą wynosić:

Dla N nieparzystego:

Obciążenia gałęzi pierścienia:

$$L_1=L_2=\ \ldots\ =L_m=\ \ldots\ =L_N=\frac{N^2-1}{8}. \tag{45}$$

Ruch transferowy:

$$L_{ADM-L,m}=L_{ADM-P,m}=\frac{N-1}{2}. \tag{46}$$

Ruch tranzytowy:

$$L_{ADM-T,m}=\frac{(N-1)(N-3)}{8}. \tag{47}$$

Współczynnik efektywności pierścienia:

$$\mathbf{WEP}=\frac{4N}{N+1} \tag{48}$$

Dla N parzystego:

Obciążenia gałęzi:

$$L_1=L_2=\ \ldots\ =L_m=\ \ldots\ =L_N=\frac{N^2}{8}. \tag{49}$$

Ruch transferowy:

$$L_{ADM-L,m}=L_{ADM-P,m}=\frac{N-1}{2}. \tag{50}$$

Ruch tranzytowy:

$$L_{\mathrm{ADM-T},m}=\frac{N(N-2)}{8}. \tag{51}$$

Współczynnik efektywnosci pierścienia BSHR-4:

$$\mathbf{WEP}=\frac{4(N-1)}{N}. \tag{52}$$

W ruchu jednorodnym obciążenia wszystkich gałęzi pierścienia BSHR są sobie równe. Podobne stwierdzenie dotyczy również obciążeń węzłów.

Nietrudno stwierdzić, że współczynniki efektywności pierścienia BSHR o nieparzystej liczbie węzłów $N=2k-1$ oraz o parzystej $N=2k$ dla $k=2,3,4,\ldots$ są sobie równe dla tej samej wartości parametru k. Wartość współczynnika **WEP** dla $N>=3$ zawiera się w przedziale [3,4), a dla stosowanych w praktyce wielkości liczby węzłów pierścienia $N \in [3,16]$ wartości **WEP** mieszczą się w zakresie:

$$3<=\mathbf{WEP}<=3{,}75.$$

3) Przypadek ruchu scentralizowanego

Obciążenia gałęzi i węzłów pierścienia przy założeniu, że węzłem centralnym będzie węzeł $c=1$, można wyznaczać na podstawie zależności:

Dla N nieparzystego:

Obciążenie gałęzi jest równe

$$L_m=\begin{cases}\displaystyle\sum_{i=1}^{\frac{N+1}{2}-m} t_{1k} & \forall\ m<\dfrac{N+1}{2} \wedge k=m+i\\ 0 & \forall\ m=\dfrac{N+1}{2}\\ \displaystyle\sum_{i=1}^{m-\frac{N+1}{2}} t_{1k} & \forall\ m>\dfrac{N+1}{2} \wedge k=1+(N+m-i) \bmod N\end{cases} \tag{53a}$$

oraz dla N parzystego:

$$L_m=\begin{cases}\displaystyle\sum_{i=1}^{\frac{N}{2}-m} t_{1k}+\frac{1}{2}\,t_{1k^{\mathrm{IV}}} & \forall\ m\leqslant\dfrac{N}{2}-1 \wedge k=m+i\\ \dfrac{1}{2}\,t_{1k^{\mathrm{IV}}} & \forall\ m=\dfrac{N}{2} \wedge m=\dfrac{N}{2}+1\\ \displaystyle\sum_{i=1}^{m-\frac{N-2}{2}} t_{1k}+\frac{1}{2}\,t_{1k^{\mathrm{IV}}} & \forall\begin{cases} m>\dfrac{N}{2}+1 \wedge\\ k=1+(N+m-i) \bmod N,\end{cases}\end{cases} \tag{53b}$$

gdzie:

$$k^{IV} = N/2 + 1 .$$

Przy jednorodnej postaci ruchu scentralizowanego obciążenie gałęzi jest równe: dla N nieparzystego:

$$L_m = \left| \frac{N+1}{2} - m \right| \tag{54a}$$

oraz dla N parzystego:

$$L_m = \begin{cases} \left| \frac{N+1}{2} - m \right| & \forall \; m \neq \frac{N}{2} + 1 \\ \frac{1}{2} & \forall = \frac{N}{2} + 1 . \end{cases} \tag{54b}$$

Maksymalna wartość obciążenia gałęzi i minimalna pojemność pierścienia wynoszą:

$$C = L_{max} = (N-1)/2 .$$

Gałęziami maksymalnie obciążonymi są: rozpoczynająca się i kończąca się w węźle centralnym. Obciążenia gałęzi maleją liniowo w miarę oddalania się od węzła centralnego.

Całkowite obciążenie pierścienia, podobnie jak pierścienia USHR (vide równanie (18)), wynosi:

$$T_{tot} = N - 1 .$$

W związku z tym współczynnik efektywności pierścienia w tych warunkach jest równy:

$$\mathbf{WEP} = 2 .$$

Ruch transferowy dla $m = c$ przy nieparzystej wartości N określony jest przez zależności (32) i (33), a dla N parzystego — przez (39), (40) i (41).

Dla węzła nie będącego węzłem centralnym ($m \neq c$) będzie:

$$L_{ADM-L,m} = 0, \qquad L_{ADM-P,m} = t_{cm} \qquad \forall \begin{cases} m < c \;\wedge\; c - m < \frac{N}{2} \\ m > c \;\wedge\; m - c > \frac{N}{2} \end{cases}$$

$$L_{ADM-L,m} = L_{ADM-P,m} = \frac{1}{2} t_{cm} \qquad \forall \; | m - c | = \frac{N}{2} \tag{55}$$

$$L_{ADM-L,m} = 0, \qquad L_{ADM-P,m} = 0 \qquad \forall \begin{cases} m < c \;\wedge\; c - m < \frac{N}{2} \\ m > c \;\wedge\; m - c > \frac{N}{2} \end{cases}$$

Powyższa zależność jest słuszna dla parzystych wartości N. Podobną do niej można uzyskać dla N nieparzystego.

Ruch tranzytwy w pierścieniu BSHR o scentralizowanym ruchu nie będzie występował w żadnym z węzłów dla $N=3$ oraz w niektórych węzłach przy $N>3$. Brak tego ruchu, tzn.

$$L_{\mathrm{ADM-T},m}=0$$

ma miejsce dla $N>3$ w węzłach, dla których m spełnia warunki:

$$m=\begin{cases} c \\ 1+\left(c+\dfrac{N-\alpha}{2}\right) \bmod N \end{cases} \tag{56}$$

gdzie:

$\alpha=1{,}3$ dla N nieparzystego

oraz

$\alpha=2$ dla N parzystego.

4) Przypadek ruchu szeregowego

W sytuacji, gdy w pierścieniu BSHR obsługiwany jest ruch szeregowy, obciążenia gałęzi i węzłów pierścienia będą wynosić:

$$L_m=t_{mk} \tag{57}$$

gdzie:

$$k=1+m \bmod \mathrm{N}. \tag{57a}$$

Wielkość całkowitego ruchu w pierścieniu wynosi:

$$T_{tot}=\sum_{m=1}^{N} t_{mk}. \tag{58}$$

gdzie k określone jest jak w zależności (57a).

Dla jednorodnej postaci ruchu:

$$C=L_{\max}=L_m=1$$

oraz

$$T_{tot}=N.$$

Współczynnik efektywności pierścienia w tych warunkach będzie równy:

$$\mathbf{WEP}=\mathrm{N}.$$

Ruch transferowy w węźle m jest równy:

$$\begin{aligned} L_{\mathrm{ADM-L},m}&=t_{ml} \\ L_{\mathrm{ADM-P},m}&=t_{mk} \end{aligned} \tag{59}$$

gdzie:

$$\begin{aligned} l&=1+(m+\mathrm{N}-2) \bmod \mathrm{N} \\ k&=1+m \bmod \mathrm{N}. \end{aligned}$$

Ruch tranzytowy nie występuje w żadnym z węzłów, a więc:

$$L_{\mathrm{ADM-T},m}=0.$$

4. PORÓWNANIE WŁASNOŚCI PIERŚCIENI SDH Z PUNKTU WIDZENIA ICH MOŻLIWOŚCI OBSŁUGI RUCHU TELEKOMUNIKACYJNEGO

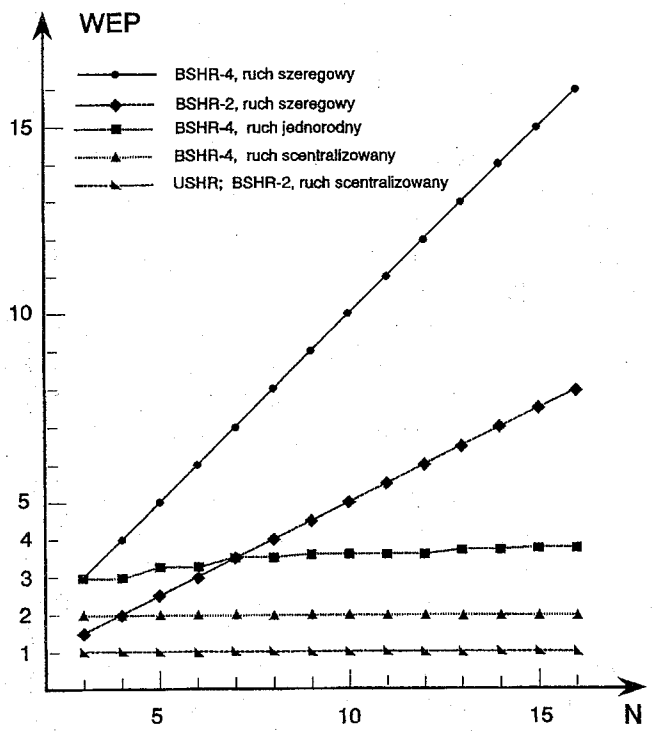
Odmienności w architekturze poszczególnych struktur pierścieniowych SDH powodują zróżnicowanie ich własności, a tym samym i ich przydatności do określonych zastosowań.

Pierścień dwukierunkowy zarówno z dzieloną rezerwą (BSHR-2) jak i z dedykowaną rezerwą (BSHR-4) pozwala na obsługę na ogół znacznie większej liczby strumieni transferowanych w poszczególnych jego węzłach niż pierścień jednokierunkowy. W przypadku zainstalowania w pierścieniu systemu STM-n, gdzie $n = 1,4$ lub 16, przy pomocy pierścienia BSHR-4 możliwe jest przesłanie maksymalnie N modułów STM-n, jeśli transmisja dotyczyć będzie tylko odcinków między sąsiednimi węzłami pierścienia, czyli w sytuacji, gdy obsługiwany ruch telekomunikacyjny będzie ruchem szeregowym. W tych warunkach bowiem w każdym z węzłów pierścienia wydzielany będzie strumień STM-n i do zajmowanej przezeń szczeliny czasowej może być wprowadzony nowy, o przepustowości równej przepustowości systemu — w tym przypadku STM-n. Tak więc maksymalna wielkość ruchu telekomunikacyjnego obsługiwana w tych warunkach przez pierścień będzie równa pojemności N systemów STM-n. Warto zwrócić uwagę, że tak określona maksymalna pojemność systemu jest rzadko osiągana w warunkach praktycznych. Stanowi ona jednak dobre odniesienie i miarę porównawczą.

W pierścieniu BSHR-2 pojemność w tych samych warunkach będzie równa przepustowości N/2 systemów STM-n, a zatem jest 2-krotnie mniejsza od pojemności pierścienia BSHR-4. Wynika to z konieczności zarezerwowania połowy przepustowości pierścienia na cele jego zabezpieczenia.

W przypadku zastosowania pierścienia USHR przesłanie między sąsiednimi węzłami modulu STM-n spowoduje pełne wykorzystanie możliwości transmisyjnych pierścienia, przez zajęcie pozostałej części pierścienia dla potrzeb transmisji w drugim kierunku. Tak więc maksymalna wielkość ruchu telekomunikacyjnego obsłużona przez pierścień USHR będzie równa pojemności tylko jednego systemu STM-n. Jest ona zatem N-krotnie i N/2-krotnie mniejsza od analogicznych wielkości uzyskiwanych w pierścieniach — odpowiednio — BSHR-4 i BSHR-2.

Dobrą miarą wielkosci całkowitego ruchu telekomunikacyjnego obsługiwanego przez pierścień SDH w stosunku do przepustowości zainstalowanego w tym pierścieniu systemu transmisyjnego jest współczynnik efektywności pierścienia **WEP** określony przez równanie (5). Określa on efektywność wykorzystania przepustowości pierścienia SDH. Zależność wartości współczynnika **WEP** od liczby węzłów pierścienia dla różnych struktur pierścieniowych oraz różnych postaci ruchu telekomunikacyjnego zilustrowana jest na rys. 6. Należy podkreślić, że zależności pokazane na wykresach na rys. 6 uzyskane zostały dla szczególnych przypadków, gdy zainteresowania między odpowiednimi parami węzłów były sobie równe.



Rys. 6. Współczynnik efektywności pierścienia SDH w zależnosci od liczby jego węzłów

Najmniejszą efektywnością wykorzystania przepustowości zainstalowanego systemu charakteryzuje się pierścień jednokierunkowy USHR. Jego efektywność jest niezależna zarówno od postaci obsługiwanego ruchu telekomunikacyjnego jak i od liczby węzłów pierścienia. Zatem główne preferencje dla zastosowania pierścieni jednokierunkowych obejmują przede wszystkim przypadki, wymagające mniejszej przepustowości użytkowej systemu lub takie, w których większość obciążeń ruchem telekomunikacyjnym dotyczy jednego węzła. Z tą ostatnią sytuacją mamy do czynienia np. w przypadku małego ruchu lokalnego w obrębie pierścienia i znacznego zapotrzebowania na ruch zewnętrzny. Z takim rozkładem ruchu telekomunikacyjnego spotykamy się często na najniższym poziomie organizacji sieci dostępu.

Taką samą efektywnością, jaką ma pierścień USHR, charakteryzuje się pierścień BSHR-2 przy obsłudze ruchu scentralizowanego, 2-krotnie zaś większą — pierścień BSHR-4, obsługujący ruch scentralizowany. Jednakże warto pamiętać, iż wyższa — w porównaniu z efektywnością pierścienia BSHR-2 — efektywność pierścienia BSHR-4 jest okupiona większym kosztem jego instalacji. Z tego też powodu struktura BSHR-4 nie znajduje na ogół zastosowania dla obsługi ruchu scentralizowanego. Efektywność pierścieni SDH przy obsłudze takiego ruchu jest niezależna od wielkości (liczby węzłów) pierścienia.

Najodpowiedniejszą konfiguracją pierścienia SDH do obsługi ruchu jednorodnego jest pierścień dwukierunkowy. Jego efektywność zawiera się w przedziałach [3,4) oraz [1,5,2) dla pierścieni — odpowiednio — BSHR-4 i BSHR-2.

Jest ona w niewielkim stopniu zależna od liczby węzłów pierścienia, rosnąc z jej wzrostem. Dla N = 16 efektywność pierścienia BSHR przy obsłudze ruchu jednorodnego osiąga wartość:

$$\mathbf{WEP}_{\mathbf{BSHR-4,\,jednorodny}} = 3{,}75$$

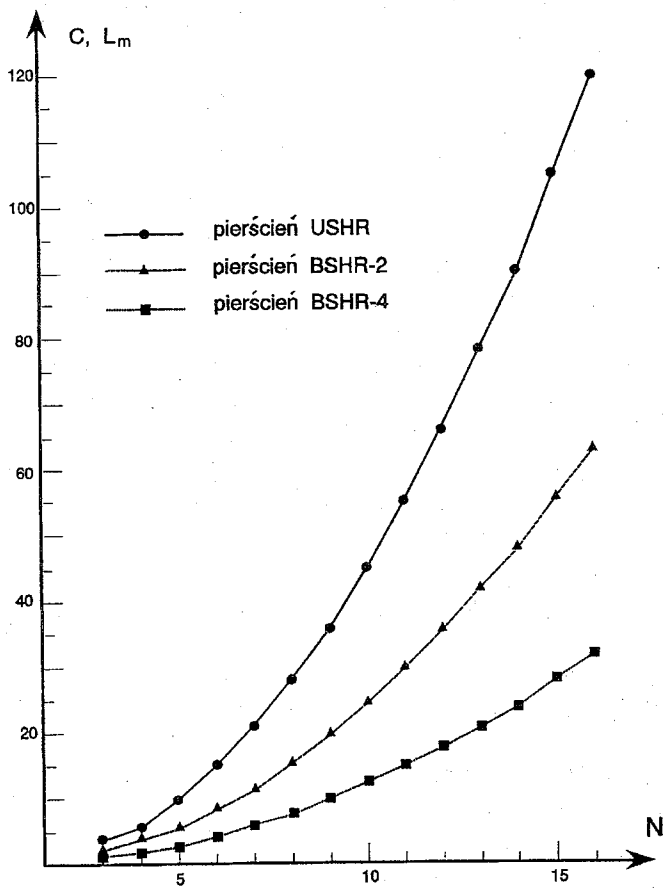
$$\mathbf{WEP}_{\mathbf{BSHR-2,\,jednorodny}} = 1{,}875$$

Postacią ruchu, przy której pierścień dwukierunkowy uzyskuje największą efektywność w wykorzystaniu swoich możliwości transmisyjnych, jest ruch szeregowy. Współczynnik efektywności pierścienia BSHR-4 — przy zuniformizowaniu tej postaci ruchu — określony jest wartością N i rośnie zatem liniowo ze wzrostem liczby węzłów pierścienia. Jak powiedziano już wyżej, pierścień BSHR-4 z zainstalowanym systemem STM-n przy ruchu szeregowym jest w stanie obsłużyć w najkorzystniejszych warunkach ruch o wielkości równej przepustowości N systemów STM-n. Te najkorzystniejsze warunki powstają przy równości ruchu między sąsiednimi węzłami i maksymalnym wykorzystaniu przepustowości przęseł pierścienia. Tak wysoka efektywność pierścienia BSHR przy obsłudze takiego ruchu jest konsekwencją braku ruchu tranzytowego we wszystkich węzłach.

Wspomniana wyżej cecha pierścieni BSHR implikuje główną sferę ich zastosowań. Pierścienie dwukierunkowe, zarówno z dzieloną, jak i dedykowaną rezerwą, są wykorzystywane na ogół w przypadkach, gdy występuje znaczne zapotrzebowanie przede wszystkim na ruch między sąsiednimi lub bliskimi sobie węzłami. Dotyczy to np. takich sieci miejskich, w których ruch międzycentralowy kształtuje się tak, że może być obsługiwany przede wszystkim przez krótkie odcinki pierścienia. Strukturę ruchu zbliżoną do tej, jaką ma ruch szeregowy, przewiduje się również w przypadkach wykorzystywania sieci SDH dla transmisji sygnałów związanych z zintegrowanymi usługami szerokopasmowymi, opartych na asynchronicznej technice transferu ATM (Asynchronous Transfer Mode). Zastosowanie pierścieni BSHR w sieciach SDH przy takich konfiguracjach ruchu znajduje najwyraźniejsze uzasadnienie ekonomiczne.

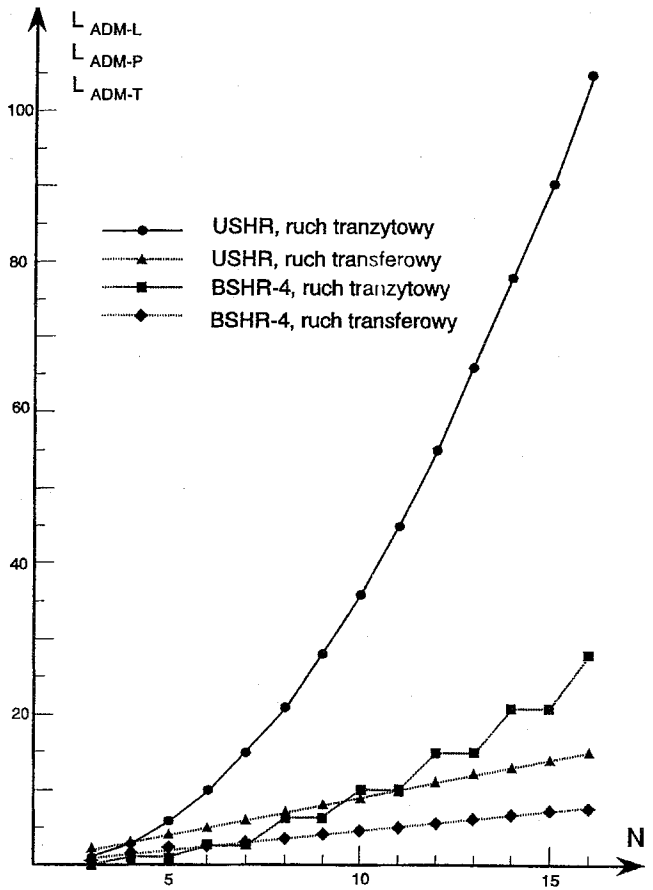
Z punktu widzenia efektywności wykorzystania przepustowości pierścienia BSHR, liczba jego węzłów winna być odpowiednio duża, np. około 20 [5]. Jednakże w praktyce — ze względu na trudności związane z zarządzaniem pierścieniami o dużej liczbie węzłów, a także zmniejszenie efektywności ekonomicznej takich pierścieni — występuje ograniczenie liczby ich węzłów do wielkości nie przekraczającej 16 [6], [7], [10]. O wadze tego ograniczenia świadczy fakt przeznaczenia przez CCITT w polu adresowym kanału automatycznego przełączania zabezpieczenia APS (Automatic Protection Swtching) jedynie 4 bitów dla oznaczenia numeru węzła pierścienia.

Zależności minimalnej pojemności pierścienia SDH, a także wielkości obciążenia jego gałęzi, od liczby węzłów pierścienia dla różnych jego struktur oraz jednorodnego ruchu pokazane są na rys. 7. Wykresy wyraźnie ilustrują większą efektywność pierścienia dwukierunkowego od jednokierunkowego w zastosowaniach do obsługi ruchu jednorodnego lub o bliskim jednorodnemu rozkładzie.



Rys. 7. Minimalna pojemność pierścienia SDH oraz obciążenia jego gałęzi dla ruchu jednorodnego

Na wykresach przedstawionych na rys. 8 przedstawione zostały zależności obciążeń ruchem transferowym i tranzytowym węzłów pierścieni USHR i BSHR-4 przy obsłudze ruchu jednorodnego od liczby węzłów pierścienia. Dla pierścienia BSHR-2 odpowiednie wykresy różnią się jedynie skalą od podanych dla BSHR-4. Nietrudno stwierdzić, że większość obciążeń węzłów w pierścieniach SDH przy odpowiednio dużej liczbie węzłów pochodzi z ruchu tranzytowego. Stąd też dążenie do jego zminimalizowania prowadzić będzie do większej efektywności pierścienia. Możliwe jest to przez uwzględnienie w procesie projektowania pierścienia czynników, wynikających ze wzmożonego zapotrzebowania na ruch w określonych relacjach.



Rys. 8. Obciążenie węzłow pierścienia SDH jednorodnym ruchem transferowym i tranzytowym

5. ZAKOŃCZENIE

W niniejszym opracowaniu przedstawiono zależności analityczne, na podstawie których można wyznaczyć obciążenia elementów pojedynczych pierścieni SDH ruchem telekomunikacyjnym. Spośród struktur pierścieniowych SDH rozpatrzono pierścień jednokierunkowy (USHR) oraz dwukierunkowy o 2 i 4 włóknach (BSHR-2 i BSHR-4). Dla każdego z tych pierścieni wyznaczone zostały obciążenia jego gałęzi oraz wielkości ruchu transferowego i tranzytowego w każdym z węzłów dla ogólnej postaci macierzy zainteresowań oraz dla trzech szczególnych jej przypadków, obejmujących ruch jednorodny, scentralizowany oraz szeregowy.

W przypadku pierścieni dwukierunkowych podane w opracowaniu zależności zostały określone przy założeniu kierowania ruchem w obrębie pierścienia

zgodnie z algorytmem MHR. Algorytm ten charakteryzuje się dużą prostotą i efektywnoscią. Jego wadą jest jednak to, że bardzo często prowadzi do znacznego niezrównoważenia obciążeń w poszczególnych elementach pierścienia. Powoduje to w efekcie określenie większej wymaganej przepustowości pierścienia niż jest to uzyskiwane w warunkach lepszego zrównoważenia obciążeń w obrębie pierścienia [6].

Z kolei efektywniejsze pod tym względem algorytmy kierowania ruchem w pierścieniu [5], [6], [7], [11] są na tyle skomplikowane i uzależnione od rozkładu ruchu, iż nie pozwalają na analityczne wyznaczenie wielkosci obciążeń pierścienia. Nawiasem mówiąc warto stwierdzić, że algorytmy te nie prowadzą w ogólnym przypadku do określenia minimalnej pojemności pierścienia [6] i ich efektywność jest w znacznym stopniu uzależniona od rozkładu zapotrzebowania na usługi telekomunikacyjne między węzłami pierścienia. W szczególności brak jest uzasadnienia dla ich stosowania w przypadku ruchu jednorodnego.

W przedstawionej wyżej analizie obciążeń pierścieni SDH zgodnie z tym, co powiedziano w założeniach wstępnych, nie brano pod uwagę ruchu o strukturze „transfer + tranzyt"., tzn. takiego, w którym ten sam strumień podlega w danym węźle transferowi i jest jednocześnie przesyłany do co najmniej jednego, dalej położonego węzła. Z ruchem takim ma się do czynienia w przypadku transmisji np. sygnałów telewizyjnych czy telekonferencyjnych. W przypadku wystąpienia takiego ruchu należy się liczyć z mniejszymi obciążeniami niektórych elementów pierścienia, niż by to wynikało z obciążeń określonych przez macierz zainteresowań i wyznaczonych w opracowaniu zależności. W tej szczególnej postaci ruchu istneje bowiem możliwość wykorzystywania tych samych szczelin czasowych dla obsługi ruchu w różnych relacjach.

Ograniczenie wprowadzone w niniejszym opracowaniu, polegające na rozpatrywaniu obciążeń ruchem telekomunikacyjnym jedynie w pojedynczych pierścieniach SDH, nie stanowi istotnego zawężenia ogólności rozważań i uzyskanych wynikow. Wszelki bowiem ruch zewnętrzny w stosunku do pierścienia może być w obliczeniach uwzględniony poprzez wprowadzenie dodatkowego węzła lub dodatkowych węzłów logicznych transferujących wyłącznie ten ruch. Odpowiedniej modyfikacji wymaga naturalnie w takim przypadku macierz zainteresowań.

Autor pragnie złożyć gorące podziękowania dr hab. inż. Andrzejowi Dąbrowskiemu za inspirację, owocną dyskusję i wsparcie w trakcie przygotowania niniejszej pracy oraz dr inż. Januszowi Maliszewskiemu za cenne uwagi.

BIBLIOGRAFIA

1. James K. Conlisk: Topology and Survivability of Future Transport Networks. ICC'89, 1989, pp. 23.5.1 – 23.5.8
2. W.D. Grover: The Self-Healing Network: A Fast Distributed Restoration Technique for Networks using Digital Cross-connect Machines. GLOBECOM'87, Nov. 1987, pp. 28.2.1 – 28.2.6
3. Lzax Haque, Wilehlm Kremer, Kamal Raychaudhuri: Self-healing Rings in a Synchronous Environment, IEEE LTS, Nov. 1991, pp. 30 – 37

4. Steven H. Hersey, Mark J. Soulliere: Architectures and applications of SONET in a Self-healing Network. ICC'91
5. Stephen Liese: Surviving Today' Competetive Climate: Telephony, June 1992, pp. 113 – 122
6. Jianxu Shi, Johng Fonseka: Dimensioning of Self Healing Rings and Their Interconnections, ICC '93, May 1993 Geneva, pp. 1579 – 1583
7. Ching-Chir Shur, Ying-Ming Wu, Chun-Hsien Chen: A Capacity Comparison of SONET Self Healig Ring Networks, ICC'93, 1993, pp. 1574 – 1578
8. M. Wehr: Protection des resaux de transmission synchrone. Commutation & Transmission, Nr 4, 1993, pp. 5 – 12
9. Tsong-Ho Wu: Fiber Network Survivability, ARtech House Inc. 1992, Boston
10. Tsong-Ho Wu, M. Burrowes: Feasibility of a High-Speed SONET Self-Healing Ring architectures for Future Interoffice Fiber Networks, IEEE Comm. Mag., 18/11, 1990, pp. 33 – 43
11. Tson-Ho Wu, D.J. Kolar, R.H. Cardwell: High speed Self-Healing Ring Architectures for Future Interoffice Networks, IEEE GLOBECOM'1989 Nov. 1989, pp. 23.1.1 – 23.1.7

A. DROBNIK

LOAD OF SDH SELF-HEALING RING ELEMENTS UNDER TRAFFIC DEMAND

Summary

Load-carrying capacity of Self-Healing Ring (SHR) elements have been presented as a result of given communications traffic demand matrix. SHRs span, add-drop and flow through stream loads for single USHR, BSHR-2 and BSHR-4 structures have been considered in the case of different traffic demand patterns. These patterns include general, uniform, central and cyclic demands between ring nodes. Efficiency of the SHRs capacity utilization has been determined and discussed.

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI. 1994. **40**, z. 4. ss. 00–00

SPIS TREŚCI

CONTENTS

POLSKA AKADEMIA NAUK
KOMITET ELEKTRONIKI I TELEKOMUNIKACJI

KWARTALNIK ELEKTRONIKI I TELEKOMUNIKACJI

ELECTRONICS AND TELECOMMUNICATIONS QUARTERLY

SPIS TREŚCI DO TOMU XL

WYDAWNICTWO NAUKOWE PWN

WARSZAWA 1994

SPIS TREŚCI DO TOMU XL — 1994

ZESZYT NR 1

ZESZYT NR 2

ZESZYT NR 3

ZESZYT NR 4

CONTENTS

NR 1

NR 2

NR 3

NR 4

Cena 4 zł / 40000 zł

Kwartalnik Elektroniki i Telekomunikacji

WARUNKI PRENUMERATY

Wpłaty na prenumeratę przyjmowane są na okresy roczne

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

Termin przyjmowania prenumeraty rocznej krajowej i na zagranicę przez RUCH S.A. i Zakład Kolportażu Wydawnictwa SIGMA-NOT
do 20.XI. roku poprzedzającego

Termin przyjmowania prenumeraty rocznej krajowej przez Pocztę Polską
do 25.XI. roku poprzedzającego

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

Adresy przedsiębiorstw kolportażowych i ich konta bankowe:

Zakład Kolportażu Wydawnictwa SIGMA-NOT, 00-716 Warszawa, ul. Bartycka 20, skr. poczt. 1004
Konto PBK III Oddział Warszawa nr 370015-1573-139-11
Prenumerata ze zleceniem wysyłki za granicę jest dwukrotnie wyższa od ceny krajowej. Zlecający powinien podać dokładny adres odbiorcy. Dodatkowe informacje można otrzymać pod nr telefonu 49-30-86 lub 40-00-21 wew. 249, 295, 299.

Ruch S.A. Oddział Warszawa, 00-958 Warszawa, ul. Towarowa 28
Konto PBK XIII Oddział Warszawa nr 370044-1195-139-11
Dostawa odbywa się przesyłką zwykłą w ramach opłaconej prenumeraty, z wyjątkiem zlecenia dostawy pocztą lotniczą, której koszt w pełni pokrywa zleceniodawca. Prenumerata ze zleceniem dostawy za granicę jest o 100% wyższa od krajowej. Od osób lub instytucji, zamieszkałych lub mieszczących się w miejscowościach, w których nie ma jednostek kolportażowych RUCH prenumeratę Kwartalnika można zamówić w Oddziale Warszawskim RUCH na konto: PBK XIII Oddział Warszawa, nr 370044-1195-139-11.

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

W roku 1995	
przewidywana cena 1 egz. wyniesie	50.000 zł
prenumerata roczna w kraju	200.000 zł
prenumerata roczna za granicę	90 $ USA