PEŁNOMOCNIK RZĄDU
DO SPRAW
ELEKTRONICZNEJ TECHNIKI OBLICZENIOWEJ

ELEKTRONICZNA TECHNIKA OBLICZENIOWA

POLITECHNIKA ŚLĄSKA BIBLIOTEKA GŁÓWNA Gliwice
P. 3057/68

NOWOŚCI

2/1968

INSTYTUT MASZYN MATEMATYCZNYCH
DZIAŁOWY OŚRODEK INTE

Wydaje

INSTYTUT MASZYN MATEMATYCZNYCH

D z i a ł o w y O ś r o d e k I n f o r m a c j i N a u k o w e j

T e c h n i c z n e j i E k o n o m i c z n e j

KOMITET REDAKCYJNY

Eugeniusz Nowak, Dorota Prawdzic /red. nacz./,
Ryszard Rawski, Zbigniew Świątkowski
Witold Wudel /sekr. red./

Adres Redakcji: Warszawa, ul. Koszykowa 79,
tel. 21-84-41, wewn. 131 lub 28-37-29

Wykonano w Instytucie Maszyn Matematycznych

Nakł. 500 egz. pap. offset kl. III g. 80 zam. nr 72/68, GP-IV/1843/67

P. 3057/68

POLITECHNIKA ŚLĄSKA BIBLIOTEKA GŁÓWNA GLIWICE

ELEKTRONICZNA TECHNIKA OBLICZENIOWA

N O W O Ś C I

Rok VII Nr 2/18/ maj 1968

S p i s t r e ś c i

str.

Mgr inż. JERZY DAŃDA, mgr inż. JERZY FIETT: Postępy automatyzacji wytwarzania maszyn cyfrowych. Część II 3

Mgr inż. JERZY DAŃDA. mgr inż. ZDZISŁAW WRZESZCZ: H a r d w a r e i s o f t w a r e techniczno-funkcjonalnych systemów pamięciowych 47

Mgr inż. MIROSŁAW MIKA: Zastosowanie półprzewodnikowych układów scalonych do konstrukcji EMC. Część I 63

Dr inż. ZYGMUNT SAWICKI: Kryteria oceny i dynamiczne metody pomiarowe parametrów drutów magnetycznych 89

Mgr inż. TADEUSZ WERNER: Pamięć wirtualna maszyny cyfrowej i algorytmy zamiany 107

Mgr inż. ZOFIA KLAUZNICER: Analiza kosztów ferrytowych pamięci operacyjnych w systemach 2,5D i 3D 121

Tabela: Przewidywana wartość produkcji maszyn matematycznych i urządzeń w USA 133

mgr inż. Jerzy Dańda
Instytut Maszyn Matematycznych
mgr inż. Jerzy Fiett
Przemysłowy Instytut Telekomunikacji

681.322:681.3.06:
681.322.001.24

POSTĘPY AUTOMATYZACJI WYTWARZANIA MASZYN CYFROWYCH

C Z Ę Ś Ć I I

Wstęp - streszczenie

W części I niniejszego artykułu omówiono ogólne zagadnienia automatyzacji projektowania maszyn cyfrowych wraz ze szkicem historycznego rozwoju. Następnie rozwinięto szerzej zagadnienia stosowania języków symulacyjnych typu GPSS, oraz SOL do analizy prekonstrukcyjnej systemów automatycznego przetwarzania informacji. Kolejnym tematem była automatyzacja projektowania hardware'u w zakresie automatycznego projektowania struktury logicznej oraz symulacji działania projektowanych sieci logicznych. Część I zakończono krótką charakterystyką sytuacji krajowej w zakresie zagadnień omówionych w tej części.

Część II rozpoczynamy od dość szczegółowego opisu systemu programów symulujących, zastosowanych przy projektowaniu maszyny cyfrowej rakiety Saturn V. Artykuł [13] na którym oparty jest ten opis jest bardzo szczegółowy i pozwala, naszym zdaniem, wyrobić sobie pogląd jak winien wyglądać system programów symulacyjnych o praktycznej użyteczności.

Ponadto w artykule tym zawarto szczegółowe informacje o chwytach programowych zwiększających efektywność działania programów symulujących, która to efektywność jest, ogólnie biorąc, słabą stroną większości programów symulujących.

Ponieważ wydaje się, że szczegóły te zostały dobrze pomyślane i mogą być cenne dla osób konstruujących takie systemy symulacji w kraju, postanowiliśmy je zamieścić w niniejszym artykule, godząc się z pewnym zachwianiem proporcji materiału.

Kolejno omawiamy zagadnienia automatycznej generacji software'u szczególnego typu, a mianowicie testów kontrolnych oraz diagnostycznych. Ze względu na objętość artykułu, zagadnienie to aczkolwiek niesłychanie ważne, zostało potraktowane zwięźle.

Wydaje się, że rozważania nad testami kontrolnymi i diagnostycznymi, generowanymi automatycznie, winny być poprzedzone opisem występujących tutaj węzłowych problemów, które istnieją też przy normalnym /nieautomatycznym/ pisaniu takich programów testujących. Omówienie tych zagadnień wykracza jednak poza ramy niniejszego artykułu. W związku z tym ograniczmy się tylko do naszkicowania różnych podejść, które zarysowują się obecnie, odsyłając Czytelnika do odpowiednich pozycji literatury fachowej.

Jeżeli chodzi o zagadnienia automatycznej generacji software'u w szerszym znaczeniu tego słowa, to ograniczymy się do pobieżnego zreferowania tylko jednego kierunku, rozwijanego w Massachusetts Institute of Technology. Wybór tego kierunku został podyktowany jego związkiem z zagadnieniami automatyzacji i projektowania w ogólnym sensie tego słowa.

Pozostałą treść części II wypełniają zagadnienia realizacji fizycznej układów i sieci logicznych. Do zreferowania wybrane zostały programy ACCEL i CADIC. Pierwszy z nich służy do automatycznego rozmieszczania elementów na płytkach obwodu drukowanego, drugi używany jest przy projektowaniu obwodów scalonych. Pierwszy jest przykładem podejścia wykorzystującego batch processing, drugi jest typowym programem konwersacyjnym, stosującym urządzenia komunikacji graficznej do współpracy na bieżąco człowieka z maszyną. Wreszcie omawiamy zagadnienia rozmieszczania pakietów i wykonywania połączeń międzypakietowych.

Część II kończą uwagi na temat stanu prac podejmowanych w tym zakresie w naszym kraju i pewne postulaty odnośnie dalszego przebiegu tych prac.

8. Programowy symulator maszyny cyfrowej rakiety Saturn

W celu zbadania poprawności logicznej maszyny cyfrowej stosowanej w rakiecie Saturn opracowano w IBM Federal Systems Division symulator programowy, pracujący na maszynie IBM 7090. W programie tym zastosowano szereg nowych rozwiązań, uprzednio nie publikowanych i z tych względów wydaje się celowe krótkie omówienie tych elementów nowości.

Prócz typowych dla programów symulacyjnych zastosowań, do jakich można zaliczyć sprawdzenie poprawności schematów logicznych maszyny przed zbudowaniem jej modelu, za cel postawiono:

- analizę skutków uszkodzeń trwałych oraz chwilowych /intermittent failures/,
- ocenę skuteczności działania programów diagnostycznych maszyny Saturn w odniesieniu do błędów trwałych i chwilowych,
- ocenę proponowanych zmian przed wprowadzaniem ich do projektu.

Program pozwala na symulację całej jednostki centralnej maszyny Saturn /wraz z pamięcią tej maszyny/. Do cech charakterystycznych programu można zaliczyć:

- Zastosowanie kompilowanego programu symulacyjnego, mającego przewagę nad programami typu interpretacyjnego głównie dzięki większej szybkości symulacji. Program symuluje działanie obejmujące okres 2000 taktów zegara maszyny Saturn w ciągu jednej minuty pracy maszyny IBM 7090.
- Możliwość symulacji uszkodzeń pojedyńczych i wielokrotnych.
- Możliwość elastycznego określania symulacji wybranych fragmentów symulowanej maszyny, wybierania wyjść, których stany są drukowane i drukowania stanów wyjściowych w normalnej skali czasu, względnie w tych tylko chwilach /taktach zegarowych/ w których zmienia się stan wybranego wyjścia.

. Możliwość stosowania różnych "modeli logicznych" do układów występujących w maszynie, przez odpowiednie zmiany podprogramów w programie kompilującym.

. Zastosowanie interface'u programowego, umożliwiającego stosowanie jako danych wejściowych, opisujących strukturę logiczną symulowanej maszyny, zapisu na taśmie magnetycznej tzw. SLDA Logic Master Tape. Taśma ta jest wynikiem działania innego systemu /Solid Logic Design Automation/ programowego, stosowanego przy projektowaniu logicznym całej maszyny. Taśma ta uaktualniana w trakcie nanoszenia zmian do projektu stanowi źródłowy dokument projektowy systemu i na jej podstawie są sporządzane jako wtórne wszelkie dokumenty, które przy tradycyjnym sposobie projektowania są dokumentami pierwotnymi /np. schematy logiczne wykreślane na kalce/. Z taśmy tej spo - rządza się też "wyciągi" służące np. do sterowania maszynami automatycznie wykonującymi połączenia.

Posługiwanie się w programie symulacji tą właśnie taśmą na wielkie znaczenie, usuwa bowiem możliwość pomyłek przy ręcznej transkrypcji danych, a poza tym umożliwia natychmiastowe sprawdzenie za pomocą symulatora, konsekwencji jakie pociągają zmiany logicznę, względnie techniczne, wprowadzane do projektu.

Stosunkowo dużą sprawność działania programu symulacji uzyskano dzięki zastosowaniu kilku chwytów programowych. Pierwszym z nich jest symulowanie za pomocą jednego obiegu testu równocześnie wielu imitowanych uszkodzeń. W tym konkretnym programie istnieje możliwość symulowania 33 pojedyńczych uszkodzeń lub 33 grup uszkodzeń wielokrotnych /do 25-ciu uszkodzeń w grupie/. Uzyskuje się to przez odpowiednie wykorzystywanie poszczególnych bitów słowa maszynowego. Wprawdzie sposób ten zmusza do dokonywania operacji na poszczególnych bitach słowa, co musi być robione przy pomocy programu i co obniża szybkość działania programu, jednak w sumie uzyskuje się znaczne usprawnienie programu.

Drugim chwytem jest "omijanie grup rozkazów" /instruction bypass/. W programie symulującym nie operuje się /w miejscach, gdzie jest to możliwe/ pojedyńczymi układami typu "lub", "i" oraz "nie", lecz pewnymi ich zespołami, np. tworzącymi przerzutnik. Odcinek podprogramu symulujący ta-

ki przerzutnik złożony jest z co najmniej kilkunastu rozkazów. Przy klasycznych programach symulujących, w których operuje się na poziomie podstawowych elementów logicznych, wykonuje się symulację działania takiego przerzutnika, za każdym obiegiem programu symulującego. Dokonuje się tego nawet w tych przypadkach, gdy aktualnie stan przerzutnika nie ulega zmianie.

W podejściu zastosowanym przez twórców programu symulacji maszyny Saturn, przed wejściem do odcinka programu symulującego przerzutnik, sprawdza się /trzema rozkazami/ czy aktualnie przerzutnik może zmienić stan. Jeśli nie, omija się odcinek symulujący pracę przerzutnika - stąd też nazwa instruction bypass. Podejście to ma dodatkową zaletę, istotną ze względów ergonomicznych. Projektant łatwiej posługuje się programem symulacji, jeśli może operować większymi blokami układowymi, a więc przerzutnikami zamiast bramkami a także inwerterami składającymi się na przerzutnik.

Trzecim chwytem, zwiększającym globalną sprawność systemu symulacji jest rozdzielenie procesu kompilacji programu symulującego maszynę "dobrą" od procesu wprowadzania imitowanych uszkodzeń. Te ostatnie mogą być wprowadzane dowolnie przez operatora systemu i ich wprowadzenie nie wymaga powtórnej kompilacji programu symulującego. Przeprowadza się ją tylko wtedy, jeśli do struktury logicznej maszyny "dobrej" wprowadzane są zmiany.

W dalszej części artykułu poza ogólnym opisem posługiwania się programem symulacji, zatrzymamy się nieco dłużej nad realizacją takich fragmentów systemu symulacji, które zapewniają mu dużą sprawność. Nie będziemy natomiast referowali wyników badań, przeprowadzonych za pomocą symulatora i ich wpływu na zmiany w strukturze maszyny Saturn, skądinąd bardzo ciekawych, odsyłając Czytelników do źródłowego artykułu [13].

8.1. Równoczesna imitacja wielu uszkodzeń

W programie symulacji w wypadku równoczesnej imitacji wielu uszkodzeń wykorzystuje się poszczególne bity słowa maszynowego. Każdy bit

takiego słowa określa rodzaj uszkodzenia prostego, względnie grupy uszkodzeń prostych występujących równocześnie i tworzących tzw. uszkodzenia wielokrotne. Ze względu na długość słowa maszyny IBM 7090 wynoszącą 36 bitów i z pewnych technicznych względów nakazujących zarezerwować dla maszyny "dobrej" trzy bity - pozostałe 33 bity pozwalają na imitowanie równoczesne 33 pojedyńczych uszkodzeń, względnie 33 grup uszkodzeń wielokrotnych. Każde uszkodzenie wielokrotne może zawierać maksymalnie 25 uszkodzeń "pojedyńczych".

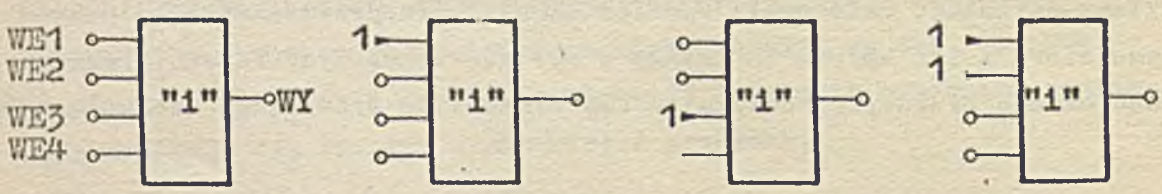
Wykorzystanie słowa maszynowego i uproszczony obraz działania programu symulacyjnego może być zilustrowany następującym przykładem. Załóżmy symulację działania układu typu iloczynu logicznego 4-ro wejściowego, przedstawionego na rys. 1. Chcemy zbadać działanie tego układu dla różnych informacji przykładanych na jego wejścia, a mianowicie:

WE 1	WE 2	WE 3	WE 4	
0	1	1	1	I-szy wariant danych wejściowych
0	0	1	1	II-gi wariant danych wejściowych

przy założeniu następujących trzech różnych uszkodzeń:

a. WE 1 ma stale wartość logiczną 1 /uszkodzenie pojedyńcze/
b. WE 3 ma stale wartość logiczną 1 /uszkodzenie pojedyńcze/
c. WE 1 i WE 2 mają stale /i równocześnie/ wartość logiczną 1 /uszkodzenie podwójne/.

Uszkodzenia tego typu powstają np. przy zwarciach lub rozwarciach obwodu wyjściowego /zależnie od stosowanej konwencji/ lub wejściowego podstawowego układu logicznego. Rozpatrywane warianty uszkodzeń przedstawiono na rys. 1.



Rys. 1. Czterowejściowy układ logiczny typu "i" i warianty uszkodzeń tego układu

Tabela 1

Program	zawartość rejestru A	zawartość miejsca pamięciowego WY
UMA E1	00000000....0	XXXXXXXX....X
DOL E1+4	00010100....0	XXXXXXXX....X
PZA E1+7	00000000....0	00010100....0
UMA E1+1	11111111....1	00010100....0
DOL E1+5	11111111....1	00010100....0
MNL E1+7	00010100....0	00010100....0
PZA E1+7	00000000....0	00010100....0
UMA E1+2	11111111....1	00010100....0
DOL E1+6	11111111....1	00010100....0
MNL E1+7	00010100....0	00010100....0
MNL E1+3	00010100....0	00010100....0
PZA E1+7	00000000....0	00010100....0

.....
.....

```
E1                  :DANE WEJSCIOWE - WE1=0,WE2=1,WE3=1,WE4=1
00000000....0       :WE1
11111111....1       :WE2
11111111....1       :WE3
11111111....1       :WE4
00010100....0       :MASKA 1
00000100....0       :MASKA 2
00001000....0       :MASKA 3
XXXXXXXX....X       :WY - PRZED WYKONANIEM PROGRAMU SYMULACJI
```

```
E1                  :INNE DANE WEJSCIOWE - WE1=0,WE2=0,WE3=1,WE4=1
00000000....0       :WE1
00000000....0       :WE2
11111111....1       :WE3
11111111....1       :WE4
00010100....0       :MASKA 1
00000100....0       :MASKA 2
00001000....0       :MASKA 3
00000100....0       :WY - PO WYKONANIU PROGRAMU SYMULACJI
```

Program symulujący dla bloku iloczynu o czterech wejściach, zapewniający możliwość imitowania uszkodzeń lub ich kombinacji przedstawiony został w tabeli 1. Dla zilustrowania działania programu obok odpowiednich rozkazów programu * zamieszczono zawartość akumulatora oraz roboczego miejsca pamięci "WY", jaka w tych rejestrach istnieje po wykonaniu rozkazu zawartego w danym wierszu.

W bloku E1 przekonuje się w kolejnych słowach zadane stany wejściowe układu, którego działanie symulujemy oraz maski, określające uszkodzenia. I tak, w czwartym bicie słowa "Maska 1" występuje jedynka oznaczająca, że chodzi nam o symulację działania układu z uszkodzonym pierwszym wejściem, mającym stałe wartość logiczną "jeden". Podobnie "jedynka" w piątym bicie słowa "Maska 2" oznacza, że kolejne imitowane uszkodzenie polega na stałej wartości logicznej "jeden" na wejściu WE 2. Natomiast jedynki na szóstej pozycji bitowej słów "Maska 1" i "Maska 2" oznaczają, że tym razem chodzi o symulację uszkodzenia podwójnego, polegającego na równoczesnym przyjmowaniu /na stałe/ wartości logicznej "jeden" przez wejścia WE 1 i WE 2.

---

* Fragment programu umieszczony w tabeli 1 napisany jest w języku PJES - języku symbolicznym maszyny ZAM-41Z. Wybrano go ze względu z pewnością na większą powszechność tego języka w Polsce niż języka FAP dla maszyny IBM 7090, w którym są zapisane przykłady podane w referowanej pracy [13].

Szczegółowy opis języka PJES można znaleźć w pracy [14]. Tu podamy tylko, że symbole mnemotechniczne rozkazów, użytych w tym przykładzie mają następujące znaczenie:

UMA - Umieść w Akumulatorze,
DOL - Dodaj Logicznie,
MNL - Mnóż Logicznie,
PZA - Pamiętaj i Zeruj Akumulator.

E1 jest etykietą bloku danych. Rozkaz "DOL E1+4" oznacza, że do zawartości akumulatora należy dodać logicznie zawartość czwartego /licząc od zera/ miejsca pamięciowego z bloku danych oznaczonych etykietą E1.

Wynik działania programu można odczytać z zawartości miejsca pamięciowego WY w ostatnim wierszu programu /PZA E1+7/, którego zawartość wynosi: 00010100.....0. Jak widać pierwsze uszkodzenie prowadzi do tego, że dla zadanej kombinacji wejść stan wyjścia układu jest inny /"jedynka" na czwartej pozycji słowa "WY"/ niż dla układu poprawnie działającego /pierwsze trzy "zera" w słowie "WY"/. Równocześnie jednak okazuje się, że drugie uszkodzenie ciągła "jedynka" na WE 3 nie daje innego stanu "WY" niż przy poprawnej pracy układu /"zero" na piątej pozycji bitowej słowa "WY"/.

W tabeli 1 pokazano też w uproszczony sposób wynik działania tego samego programu tzn. dla takich samych uszkodzeń imitowanych, lecz dla innych danych wejściowych /0,0,1,1/. W rezultacie otrzymamy wtedy wynik "WY" = 00000100....0, co oczywiście oznacza, że przy takich danych wejściowych stan wyjścia układu przy uszkodzeniu polegającym na przyjmowaniu przez WE 1 wartości logicznej "jeden" również nie będzie się różnił od stanu wyjścia przy normalnym działaniu układu.

8.2. O m i j a n i e g r u p r o z k a z ó w

Omijanie grup rozkazów w zestawionym /skompilowanym/ programie symulującym może być pokazane na przykładzie programu symulującego działanie przerzutnika, którego struktura została przedstawiona na rys. 2. Program symulujący pracę tego zespołu układów podstawowych, napisany jest w języku PJES ma poniższą postać *:

---

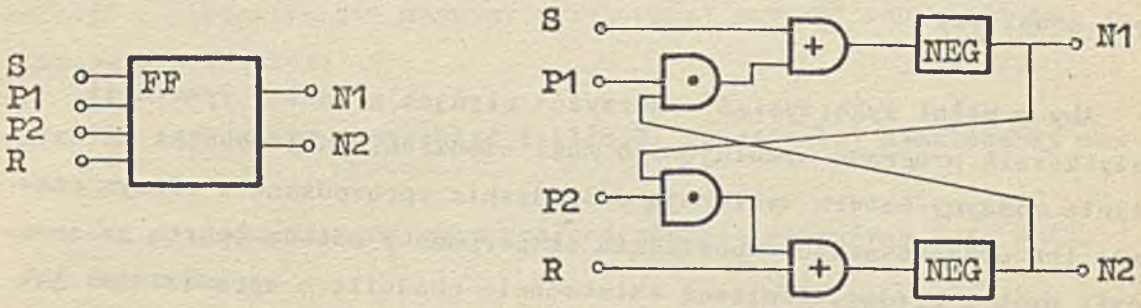
* OSL jest symbolem mnemotechnicznym alternatywy wyłączającej, której jednym argumentem jest zawartość akumulatora a drugim argumentem zawartość miejsca pamięciowego wskazanego przez adres rozkazu. W programie operacja ta wykorzystywana do tworzenia dopełnienia /negacji/przez wykonanie operacji OSL E1+6. To miejsce pamięciowe zawiera "same jedynki", co oznaczone jest przez szablon oktalny X77777777.

```
UMA E1          : A:=S
DOL E1+1        : A:=S+R
SZA Q+14        : ominięcie trzynastu rozkazów
UMA E1+2        : A:=P1
MNL E1+5        : A:=P1.N2
DOL E1          : A:=(P1.N2)+S
OSL E1+6        : A:=(P1.N2)+S
MNL E1+3        : A:=((P1.N2)+S).P2
DOL E1+1        : A:=(((P1.N2)+S).P2)+R
OSL E1+6        : A:=(((P1.N2)+S).P2)+R
PZA E1+5        : N2:=(((P1.N2)+S).P2)+R
UMA E1+5        : A:=N2
MNL E1+2        : A:=N2.P1
DOL E1          : A:=(N2.P1)+S
OSL E1+6        : A:=(N2.P1)+S
PZA E1+4        : N1:=(N2.P1)+S
.................
.................

E1
S
R
P1
P2
N1
N2
X77777777
```

Jako komentarz do poszczególnych rozkazów programu służy opis zmian zawartości akumulatora względnie miejsc roboczych programu, umieszczonych w bloku E1, po wykonaniu operacji mieszczącej się w danym wierszu programu. W komentarzu zastosowano znak " + " jako symbol operacji sumy logicznej i znak " . " jako symbol mnożenia logicznego. Miejsca "S", "R", "P1", "P2" bloku danych E1 określają stany wejść sterujących przerzutnika, w miejsca N1 i N2 natomiast w wyniku pracy programu symulującego zawarte są stany wyjść przerzutnika.

Pierwsze trzy rozkazy programu służą do ominięcia następnych trzynastu. Jeśli suma logiczna wejść sterujących S+R, tworzona pierwszymi dwoma rozkazami, okaże się równa zeru, trzeci rozkaz sterujący SZA Q+14 /Skocz przy Zerze Akumulatora/ spowoduje zwiększenie stanu licznika rozkazowego o 14 i następnym wykonywanym rozkazem będzie rozkaz znajdujący się po rozkazie PZA E1+4.



Rys. 2. Schemat blokowy i struktura logiczna przerzutnika R-S

Należy podkreślić, że w tym przypadku oszczędność na czasie wykonywania programu będzie znaczna lecz nie maksymalna, gdyż oszczędność czasu będzie spowodowana wyeliminowaniem tylko trzynastu rozkazów. Gdyby jednak bloki składały się z większej ilości układów podstawowych oszczędności byłyby znacznie większe.

8.3. Imitowanie uszkodzeń trwałych i chwilowych

Jak już powiedzieliśmy wcześniej, proces kompilowania programu symulującego maszynę nieuszkodzoną jest oddzielony od procesu wprowadzania do programu uszkodzeń imitowanych. Zalety jakie uzyskuje się wskutek takiego podejścia zostały już wymienione. W celu wprowadzenie do programu symulacji imitowanego uszkodzenia operator systemu podaje /za pomocą karty perforowanej/ tylko następujące dane o uszkodzeniu: numer arkusza schematu logicznego, symbol bloku logicznego, numer wejścia względnie wyjścia bloku logicznego, do którego wprowadzane jest imitowane uszkodzenie i typ tego uszkodzenia tzn. czy mamy do czynienia ze stałym wprowadzaniem wartości logicznej "1" czy też "0".

W przypadku wprowadzenia błędu chwilowego, operacje stają się nieco bardziej złożone. Istnieją ponadto dwie możliwości. Przy pierwszej z nich, prócz danych wprowadzanych w sposób identyczny jak przy imitowaniu uszkodzeń trwałych, operator systemu wprowadzał informacje o czasie /liczonym w cyklach zegarowych maszyny symulowanej/, w którym uszkodzenie ma się pojawić i o czasie, w którym uszkodzenie to ma być usunięte.

Aby w pełni wykorzystać korzyści płynące z takiej symulacji, użytkownik programu symulującego musi wiedzieć, jakie skutki na działanie maszyny Saturn wywierają uszkodzenia wprowadzane w różnym czasie lub przeprowadzać odpowiednie eksperymenty metodą opartą na zasadzie prób i błędów. Ponieważ ostatecznie chodziło o sprawdzenie jak skuteczne są projektowane w maszynie Saturn środki wykrywania błędów /zarówno kontrola układowa jak i programowa/ w następnej wersji usprawniono /i uproszczono z punktu widzenia operatora programu symulującego/ sposób definiowania uszkodzeń chwilowych. W tej drugiej wersji użytkownik określa tylko miejsce, w którym ma być wprowadzony błąd chwilowy, nie podaje natomiast czasów, kiedy ma być on wprowadzony i kiedy usunięty. Czynności te wykonuje za niego program symulujący.

Program symulujący przyjmuje, że zadeklarowane przez operatora uszkodzenie chwilowe jest uszkodzeniem trwałym i to założenie jest utrzymywane w mocy do chwili wykrycia przez program diagnostyczny /którego działanie w maszynie symulowanej jest również symulowane niejako "do kwadratu" przez ten program/ odchylenia od normalnej pracy maszyny, wywołanego wprowadzonym uszkodzeniem. Wtedy ulega zmianie stan pewnego zespołu liczników /programowych/ zawartych w programie symulującym. Na podstawie zmian stanu tych liczników operatorowi zostają dostarczone po przebiegu programu symulującego następujące informacje:

- ile razy imitowany błąd naruszył prawidłową pracę układów logicznych,
- jaki procent cyklów zegarowych maszyny symulowanej stanowiły te cykle, w czasie których układy logiczne działały źle,
- ile razy imitowany błąd został wykryty,

. jaki procent błędów został wykryty przez środki detekcji błędów maszyny symulowanej,
. czas, po jakim nastąpiła detekcja uszkodzenia,
. okres między kolejnymi detekcjami uszkodzeń.

Dane statystyczne tego typu pozwalają użytkownikowi programu symulacji i projektantowi maszyny symulowanej uzyskać odpowiedź na następujące kwestie:

. Ile razy wprowadzony błąd zakłócił pracę całej symulowanej maszyny?
. Czy czułość maszyny na określone rodzaje uszkodzeń jest duża, czy mała?
. Czy skutki jakie powoduje uszkodzenie odbijają się w sposób "trwały" na dalszym działaniu systemu, czy też mają znaczenie "przemijające"?
. Jak dużo błędów wywierających wpływ na poprawną pracę maszyny jest w istocie wykrywana przez programy diagnostyczne?
. Jak długo /w mikrosekundach/ musi trwać uszkodzenie chwilowe, aby być "zauważone" przez program diagnostyczny?
. Jakie jest prawdopodobieństwo wykrycia określonego imitowanego uszkodzenia chwilowego, jeśli założymy odpowiednie zmiany okresu czasu trwałości uszkodzenia?

Krótko mówiąc: co jest w tej maszynie błędem chwilowym, jak wpływa on na pracę maszyny i czy może być wykryty i zlokalizowany?

8.4. Praca programu symulującego

Prace przygotowawcze, które powinny być wykonane przed puszczeniem w ruch programu symulującego można skrótowo ująć w następujących punktach:

. Przetworzyć szkicowe schematy logiczne symulowanej maszyny na zapis odpowiedni do przetwarzania maszynowego. Ten etap jest przygotowany w ramach projektu logicznego, wykonywanego z użyciem systemu projektowania Solid Logic Design Automation. W wyniku tego powstaje taśma 1 /rys. 3./.

- Wybrać z taśmy magnetycznej, na której zapisane są wszystkie informacje o strukturze projektowanego systemu, te fragmenty i w takiej kolejności, aby uzyskać sprawną pracę programu symulującego, czyli bez zbytecznego "pętlenia". Praca ta na obecnym etapie rozwoju systemu jest wykonywana ze znacznym udziałem człowieka, który określa kolejność symulowania. Rozpoczyna się od symulacji działania układów czasujących /zegarów/, kolejno symuluje się sieci kombinacyjne i wreszcie układy sekwencyjne. W wyniku tego powstaje taśma magnetyczna 2 /rys. 3/. Przy opracowywaniu tej taśmy projektant systemu posługuje się programem wybierającym /SELECT PROGRAM/ oznaczonym symbolem A na rys. 3.
- Na podstawie informacji zawartych w taśmie 2 oraz ewentualnych informacji o uszkodzeniach imitowanych, zawartych w taśmie magnetycznej 3, wprowadzonych tam z kolei w wyniku działania Programu Wprowadzania Imitowanych Uszkodzeń /FAILURE INJECTION PROGRAM/ - trzeba sporządzić /skompilować/ program symulujący przy pomocy Programu Kompilującego /COMPILER PROGRAM, oznaczony symbolem C, na rys. 3/. Program ten w wyniku swego działania daje prócz zasadniczej taśmy 5, zawierającej program symulujący dwie inne taśmy magnetyczne: Taśmę Miejsc /LOCATION TAPE/ - czyli taśmę 4 oraz Taśmę Uszkodzeń /FAILURE TAPE/ - czyli taśmę 6, których rola zostanie omówiona później.

Aby móc przeprowadzić symulację, należy jeszcze sprecyzować szereg innych danych, potrzebnych do pracy Programu Wykonawczego /EXECUTIVE PROGRAM/, dla którego informacjami wejściowymi są taśmy 4 i 5 oraz dane wprowadzone przez Program Ładowania Pamięci Maszyny Saturn /MEMORY LOADER PROGRAM/.

Działanie tego ostatniego programu zostanie omówione później, obecnie podamy w skrócie dodatkowe dane wprowadzane przez użytkownika systemu dla sprecyzowania trybu symulacji.
Są to następujące dane:

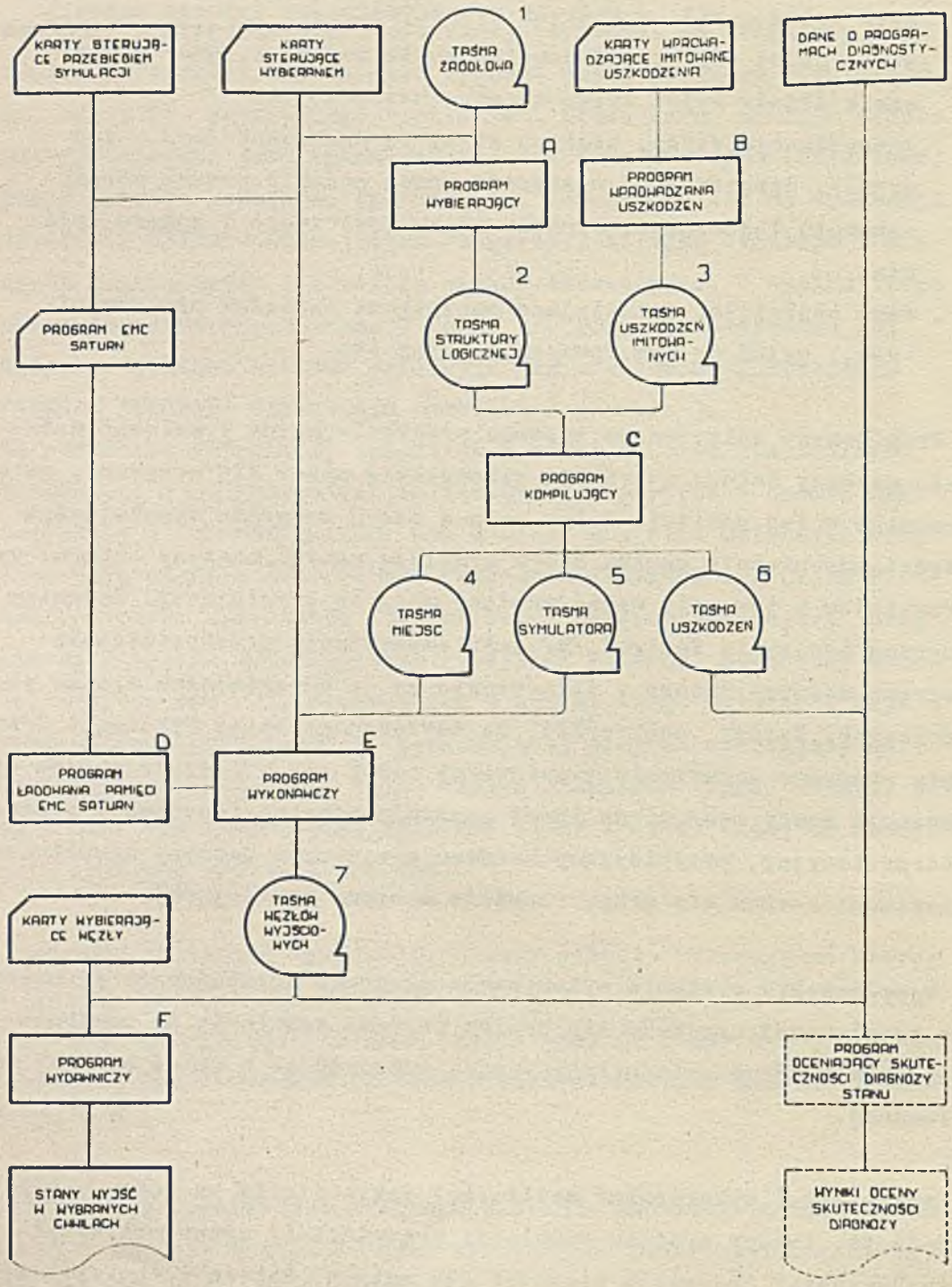
- symulacja maszyny nieuszkodzonej względnie z imitowanymi uszkodzeniami,
- ilość rozkazów, których działanie ma być symulowane /czas symulacji/,

- sposób wydawania wyników: dla każdego cyklu zegara lub dla tych tylko cykli, w których wyspecyfikowane wyjście zmienia stan. Można też wybrać dowolne cykle zegara i żądać drukowania stanów wyjść tylko w tych chwilach,
- specyfikacja wyjść, których stany są drukowane /max. 100 wyjść/. Specyfikacja następuje przez podanie numeru strony schematu logicznego, symbolu bloku logicznego i numeru wyjścia.
- dane wejściowe, ustawiające początkowe wartości przerzutników, wejść do symulowanej maszyny itp.

Przy pomocy opisywanego systemu programów można symulować działanie maszyny Saturn w trakcie wykonywania przez nią programu, umieszczonego w jej pamięci. W tym celu w skład programu symulującego wchodzi odpowiednia część, która symuluje pamięć maszyny Saturn, wykorzystując w tym celu część pamięci maszyny symulującej. Zadaniem Programu Ładowania Pamięci /D/ jest odpowiednie przekształcenie programu maszyny Saturn i załadowanie go do odpowiednich miejsc pamięciowych. Należy podkreślić, że zastosowana tutaj symulacja działania programu w maszynie symulowanej różni się w istotny sposób od symulacji pracy maszyny na innej maszynie poprzez odpowiedni program interpretacyjny, przypisujący każdemu z rozkazów maszyny symulowanej odpowiedni rozkaz czy grupę rozkazów maszyny symulującej.

W opisywanym systemie wykonywanie programu działającego w maszynie symulowanej uzyskuje się bowiem poprzez symulację na poziomie działania każdego układu logicznego, wchodzącego w skład maszyny symulowanej.

Aby w pełni wykorzystać możliwości jakie niesie za sobą takie podejście, twórcy systemu symulacji przewidzieli nawet możliwość "uruchamiania" programów pisanych dla maszyny Saturn w tym systemie symulacyjnym. W tym celu wyposażono system w możliwość drukowania tzw. post mortemów /memory dump/ tzn. wykazów stanów wybranych miejsc pamięciowych po wykonaniu jakiegoś odcinka programu maszyny



Rys. 3. Schemat czynności i przepływu informacji w systemie programowej symulacji EMC SATURN

Saturn. Istnieje nawet możliwość drukowania tzw. "śladu pamięciowego" /memory trace/. Składa się on z wykazu kolejno pobieranych rozkazów i operandów w trakcie wykonywania programu na maszynie Saturn. Ślad pamięciowy wykonuje się w zasadzie dla maszyny nieuszkodzonej. W ten sposób informacja określająca adres następnego rozkazu nie jest zniekształcona przez wadliwe działanie układów logicznych. Jeśli jednak wykonujemy program maszyny Saturn z wprowadzonymi imitowanymi uszkodzeniami, to drukuje się również te rozkazy, względnie operandy, które zostałyby pobrane, gdyby błędy wynikłe w układach logicznych zadziałały na rejestr adresowy symulowanej pamięci. W ten sposób powstają rozgałęzienia programu, jednakże dla symulacji działania następnych rozkazów programu poruszamy się tylko wzdłuż śladu dla maszyny nieuszkodzonej.

Taka informacja dodatkowa jest cenna dla analizy skutków uszkodzeń na działanie programu maszyny Saturn. Możliwość przeprowadzania takiej symulacji istnieje dzięki temu, że równocześnie z symulacją maszyny z uszkodzeniami przeprowadza się symulację maszyny nieuszkodzonej. Jak wiadomo w słowie maszynowym trzy bity przeznacza się na symulację maszyny nieuszkodzonej, a pozostałe 33 bity na symulację różnych błędów pojedyńczych lub wielokrotnych

Dla zakończenia zwięzłego opisu działania systemu programów symulacyjnych pozostało nam wyjaśnić rolę taśmy 4 /Taśma Miejsc/ oraz taśmy 7. Taśma Miejsc zawiera skorowidz symboli bloków logicznych i przyporządkowanie ich zmiennym występującym w programie symulacyjnym , mieszczącym się na taśmie 5. Ponadto na taśmie tej wyróżnione są wyjścia układów logicznych sekwencyjnych, co pozwala na wykrywanie występujących w trakcie symulacji "wyścigów". Użytkownik jest informowany o każdym takim fakcie i decyduje czy w danym przypadku wyścig jest szkodliwy i ma być wyeliminowany przez odpowiednią zmianę struktury logicznej maszyny, czy można go zignorować.

W przypadku wystąpienia "wyścigu" stan przerzutnika nie może być jednoznacznie określony. Jednak w pewnych przypadkach sytuacja taka nie wywołuje błędnej pracy maszyny i dlatego wystarczy informować operatora systemu o wszystkich takich przypadkach. Na przykładzie tej właściwości widać wyraźnie zalety współpracy Człowiek-Maszyna.

W wyniku pracy Programu Wykonawczego, który jest właściwym programem symulacyjnym powstaje taśma 7, w której zastosowany jest gęsty zapis binarny stanów wyjść wyspecyfikowanych przez użytkownika, w chwilach również przez niego określonych. Takie podejście umożliwia maksymalne wykorzystanie pamięci operacyjnej maszyny IBM 7090, która ma pojemność 32 tys. słów, co pozwala z kolei na symulację systemów złożonych z około 4000 bloków logicznych. Taśma 7 jest opracowana przez program wydawniczy F, który pozwala na zastosowanie różnych formularzy danych wyjściowych, określanych przez użytkownika. Z tej samej taśmy 7 mogą być sporządzane tabele wyników o różnych postaciach, przez odpowiednie sterowanie pracą programu wydawniczego.

Na rys. 3 przedstawione są liniami przerywanymi fragmenty systemu programów symulacyjnych jeszcze w tej chwili niezrealizowanych. Zadaniem ich będzie określanie skuteczności metod diagnozy uszkodzeń.

Czasy działania poszczególnych programów systemu, przy symulacji systemu przetwarzania informacji złożonego z 4000 bloków logicznych, są następujące:

- Program Wybierający A , 5 minut,
- Program Wprowadzania Uszkodzeń B ,ok. 1 minuta,
- Program Kompilujący C , 6 minut,
- Program Wykonawczy E działa z szybkością 12 rozkazów maszyny Saturn na minutę pracy maszyny IBM 7090. Ponieważ symulacja każdego rozkazu maszyny Saturn wymaga symulowania działania maszyny w czasie 168 cykli zegarowych tej maszyny /maszyna Saturn jest maszyną szeregową/, to w ciągu jednej minuty pracy maszyny 7090 można symulować działanie 2016 cykli maszyny Saturn,

. Program Wydawniczy F , wydanie wyników dla krótkiego programu maszyny Saturn /rzędu 20 do 30 rozkazów/ z imitacją wszystkich 33 uszkodzeń wymaga około trzech minut.

System programów symulacyjnych opracowany był przez dziewięć osób. Okres opracowania nie jest znany.

9. Automatyczna generacja testów kontrolnych i diagnostycznych

Charakterystyczną cechą nowoczesnego rozwiązania sprawy tworzenia testów wykrywających i lokalizujących uszkodzenia w układach logicznych maszyn cyfrowych jest zaniechanie sprawdzania funkcji wykonywanych przez zbiór rozkazów na rzecz sprawdzania poprawności działania poszczególnych układów logicznych. O ile przy pierwszym podejściu twórcą programu sprawdzającego jest z reguły matematyk, operujący na poziomie rozkazów, o tyle w drugim twórcą programów jest z reguły inżynier-konstruktor struktury logicznej maszyny.

Pierwsze rozwiązanie pozostawiało na ogół poza sprawdzeniem znaczną część układów, niemożliwe bowiem było sprawdzenie działania danej operacji arytmetycznej przy wszystkich kombinacjach, jakie mogą wystąpić w operandach tej operacji.

Drugie rozwiązanie, przynajmniej teoretycznie pozwala pod warunkiem podziału układów logicznych na odpowiednio małe podukłady dokonać sprawdzenia wyczerpującego możliwe stany wejść tego podukładu. Możliwe jest nawet podanie na wejścia tych podukładów takich stanów, które w trakcie normalnej pracy, przy posługiwaniu się rozkazami maszyny wystąpić nie mogą. Tak więc o ile pierwsze podejście zwano sprawdzaniem poprawności wykonywania funkcji,drugie zwie się sprawdzaniem poprawności działania układów /hardware'u/. Obecnie nie ma wątpliwości co do tego, że drugie podejście jest bezporównania rozwiązaniem poprawniejszym.

Jeśli przejdziemy do rozwiązań "technicznych" tego zagadnienia, to możemy wyróżnić dwa różne kierunki. Pierwsze z nich zakłada istnienie

tzw. niezawodnego rdzenia /hard core/. Jest to grupa układów logicznych albo wprowadzona do maszyny dodatkowo, albo stanowiąca ich integralną część, o których zakładamy, że są bardzo niezawodne, względnie /lub równocześnie/, że zostają sprawdzone przed rozpoczęciem sprawdzania programowego, przy pomocy innych metod /przeważnie "ręcznie" z użyciem oscyloskopu/.

W drugiej metodzie założeń takich nie czynimy - procedura diagnostyczna układana jest tak, żeby można było się posługiwać układami tworzącymi maszynę, przy czym wszystkie te układy mają jednakową niezawodność.

Pierwsze podejście ma już ugruntowaną pozycję, zastosowano je m. in. w maszynach IBM/360, drugie podejście posiada na razie raczej charakter możliwości teoretycznej, chociaż zaproponowano już strukturę maszyny odpowiednią do celów takiej samodiagnozy. Dla pierwszego kierunku reprezentacyjnymi artykułami są w kolejności ich publikacji następujące artykuły: K. Maling i E.L. Allen [18], W.C. Carter et al. [19] oraz J.P. Roth et al. [20]. Reprezentacyjnym ujęciem wspomianego drugiego podejścia jest artykuł E. Manning'a [17].

Na szczególną uwagę zasługuje artykuł [20], przedstawiono w nim bowiem ulepszone efektywne algorytmy DALG II i TEST-DETECT. Algorytm pierwszy gwarantuje znalezienie /dla zadanej sieci logicznej i zadanego uszkodzenia/ takiego testu, który ten błąd wykryje o ile ten test w ogóle istnieje. Algorytm drugi zaś dla danego testu i układu znajduje wszystkie uszkodzenia przez ten test wykrywane.

Te wyniki teoretyczne, oparte na "rachunku D-kostek" calculus of D-cubes J.P. Roth'a, wprowadzonym przez niego w artykule [21], rozwiązują praktycznie problem automatycznego generowania testów wykrywających i lokalizujących uszkodzenia elementów logicznych, co zdaje się wyczerpywać problemy, występujące przy podejściu pierwszym.

Wyżej wymienione algorytmy zostały przytoczone w cytowanej pracy [20]. Użyto w nich tzw. notacji Iverson'a, lub raczej języka A Programming

Language - APL, [22], którym posługują się powszechnie w firmie IBM. Formalny opis systemu maszyn IBM-360 został wykonany w notacji Iverson'a. Programy DALG-II i TEST-DETECT były wykonywane z użyciem interpretatora języka APL. Czas wykonywania obu programów na maszynie IBM-360 Model 50 dla bloku złożonego z 50 układów logicznych zawiera się w granicach kilkudziesięciu sekund /bez operacji wejścia-wyjścia/.

10. Automatyczna generacja software'u

W Massachusetts Institute of Technology przy rozwiązywaniu zagadnień automatyzacji projektowania zastosowano podejście na tyle ogólne, że może być ono wykorzystywane również przy projektowaniu i automatycznym generowaniu software'u. Proces projektowania traktowano tam bardzo ogólnie jako zagadnienie rozwiązywania problemów niekompletnie zdefiniowanych. W zagadnieniu tak postawionym wyróżnia się następujące etapy:

1. Opracowanie leksykograficzne /leksykalne/ ciągu znaków opisujących problem, dające w wyniku ciąg słów.
2. Opracowanie syntaktyczne ciągu słów, dające w wyniku zdania, opisujące problem.
3. Opracowanie semantyczne, dające w wyniku symboliczny model problemu.
4. Rozwiązanie modelu problemu, dającego w wyniku poszukiwaną odpowiedź czyli rozwiązanie problemu.

Na dotychczasowym etapie prac, opisanych w [23] uzyskano pełne rozwiązanie etapów 1 i 2 oraz częściowe rozwiązanie zagadnień 3 i 4. Podejście to zastosowano przy konstrukcji kompilatora AED-1 /Automated Enginering Design/. Kompilator AED-1 jest programem zarówno niezależnym od języka /language independent/ jak i niezależnym od maszyny /machine independent/. Przez wymianę opisów definiujących język źródłowy i maszynę, dla której kompilator AED-1 ma produkować program binarny, uzyskuje się pełną elastyczność tego kompilatora.

Inne podejście do tego zagadnienia można znaleźć w pracy przeglądowej [24], w której zreferowane zostały metody zwane boot-strapping polegające na wykorzystywaniu już zaprogramowanych części translatora do translowania jego fragmentów. Podobna technika była zastosowana przy tworzeniu translatora języka LOGOL [25].

Praca [26] opisuje kompilator kompilatorów.

11. Zagadnienia realizacji fizycznej układów i sieci logicznych

Wysoki stopień komplikacji współczesnych systemów analogowych i cyfrowych oraz ciągle wzrastająca szybkość pracy układów wchodzących w ich skład stawiają przed projektantami szczególnie trudne zadania. Nanosekundowe czasy propagacji powodują konieczność zagęszczania montażu poprzez powiększanie płytek montażowych z obwodami drukowanymi, jak i zagęszczania elementów elektronicznych na płytkach oraz w przypadku układów scalonych - rozszerzania funkcji poszczególnych układów poprzez ich komplikację.

Zabiegi te mają przede wszystkim na celu skrócenie połączeń i zmniejszenie na tej drodze opóźnień sygnałów, a także zmniejszenie przesłuchów oraz redukcję pojemności szkodliwych. Jednakże podwyższanie gęstości montażu wyspomnianymi metodami znacznie komplikuje proces projektowania. Dotyczy to zarówno trudności z prowadzeniem połączeń na płytkach drukowanych, a szczególnie z wyprowadzaniem połączeń zewnętrznych, jak i optymalizacją rozmieszczenia elementów na płytce.

Dodatkowe specyficzne trudności powstają przy projektowaniu mikroukładów o rozszerzonych funkcjach. Pozostaje to w związku z podwyższeniem wymagań co do dokładności rozmieszczenia elementów składowych, a co za tym idzie, zwiększeniem wymagań odnośnie opracowania masek. Do tego wszystkiego dołączają się coraz to ostrzejsze wymagania stawiane przez zamawiających co do terminów realizacji gotowych projektów z pełną dokumentacją konstrukcyjną i pomocami technologicznymi.

W tej sytuacji nie pozostało nic innego, jak odwołanie się do pomocy maszyn.

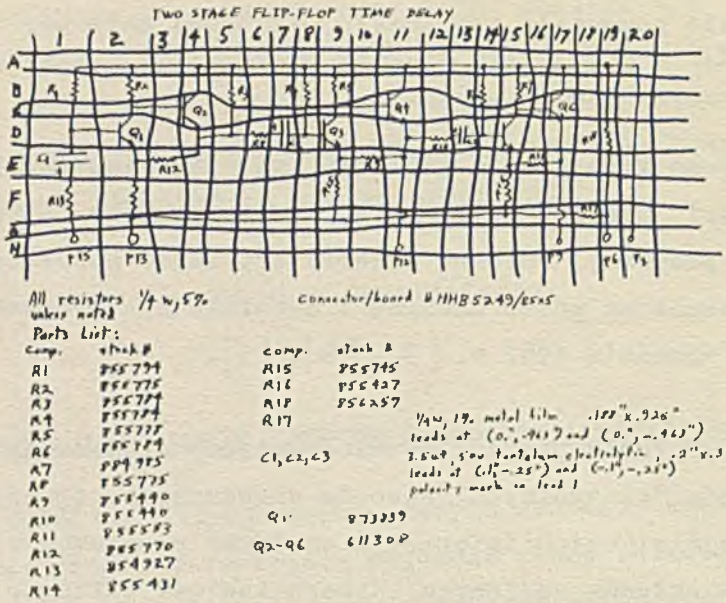
Wyżej wymienione zagadnienia znalazły już efektywne rozwiązanie na drodze programowej. Spośród najbardziej znanych należy tu wymienić trzy podstawowe programy, opisane w serii artykułów poświęconych projektowaniu wspomaganemu przez maszynę i opublikowane w czasopiśmie Electronics z 4 września 1967 r. [3], [4], [5].

W niniejszym artykule poświęcimy nieco więcej miejsca automatycznemu projektowaniu "klasycznych" obwodów drukowanych, ograniczając się jedynie do ogólnikowych informacji na temat programów służących do projektowania układów scalonych. Wyboru takiego dokonano zarówno ze względu na większą obecnie w kraju aktualność pierwszego tematu i jego praktyczne znaczenie jak i z uwagi na stosunkowo niewielki - przynajmniej na razie - krąg osób zainteresowanych szczegółami dotyczącymi projektowania układów scalonych.

11.1. Projektowanie obwodów drukowanych. Program ACCEL

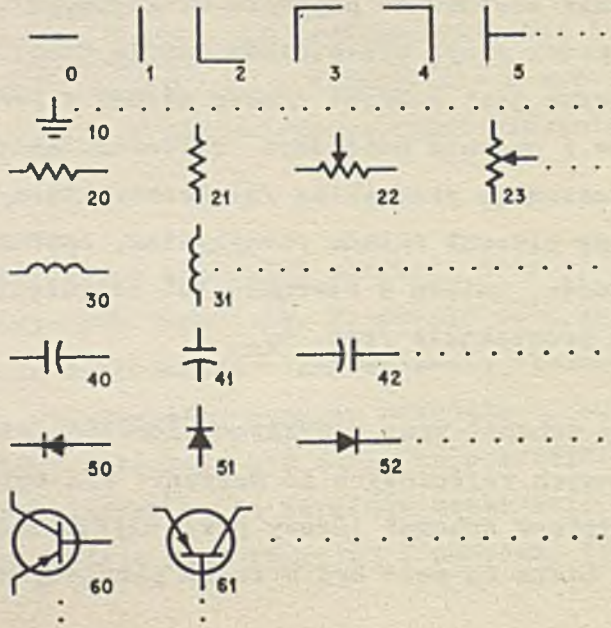
Skrót ACCEL oznacza: Automatic Circuit Card Etching Layout /automatyczne rozmieszczenie obwodów na płytkach drukowanych/. Program ten został opracowany przez grupę pracowników Sandia Corp. Punktem wyjścia dla działania programu jest schemat ideowy układu z jednoznacznym określeniem elementów i umownie nałożonym rasterem. Raster może być nanoszony przez przyuczonego pracownika /niekonstruktora/ w taki sposób, aby każdy wyróżniony element układu /tranzystor, opornik, kondensator, skrzyżowanie przewodów, zmiana w kierunku lub rozgałęzienie itp/ znalazł się w osobnym prostokącie /rys. 4/.

Tak spreparowany schemat wraz z wykazem elementów stanowi podstawę do przygotowania danych wejściowych do maszyny. Dla celów kontrolnych maszyna drukuje najpierw schemat ideowy przy użyciu znaków z listy pokazanej na rys. 5. Lista ta może być w razie potrzeby rozszerzana o dodatkowe symbole.



Rys. 4. Informacja wejściowa od konstruktora z siatką naniesioną przez pracownika pomocniczego

Na podstawie wprowadzonych informacji maszyna tworzy w pamięci odwzorowanie struktury układu. Projektant, podając specyfikację stosowanych w układzie elementów może ograniczyć się do wpisania numeru katalogowego dla elementów umieszczonych w katalogu lub też podać niezbędne dane /patrz lista elementów w dolnej części rys. 4/.



Rys. 5. Lista symboli graficznych programu ACCEL

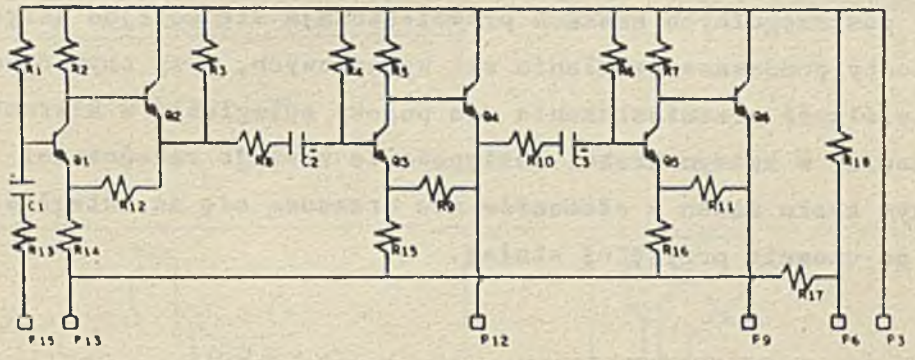
W wyniku działania programu ACCEL maszyna dokonywuje rozmieszczenia elementów na płytce montażowej, określa sposób prowadzenia połączeń oraz daje w rezultacie następującą dokumentację:

- schemat ideowy /rys. 6/,
- wykaz elementów /rys. 7/,

} powtórzenie wprowadzonych informacji

- rysunek zestawieniowy /rys. 9/,
- lista wiercenia otworów /rys. 8/,
- matryce połączeń /rys. 10a i 10b/,
- rysunek mostków /rys. 10c/.

} wynik działania programu



Rys. 6. Schemat ideowy układu z rys. 4 wydrukowany w wyniku działania programu ACCEL

R1		855704
R10		855444
R11		855355
R12		855770
R13		854927
R14		855431
R15		855745
R16		855427
R17	12.0 K, 1 0/0, 1/4W, METALFILM	
R18	.10 MEG, 1 0/0, 1/4W, METALFILM	856257
R2		855775
R3		855704
R4		855704
R5		855770
R6		855704
R7		854905
R8		855775
R9		855440
C1	3.5 000000, 50 V, TANTALUME	
C2	3.5 000000, 50 V, TANTALUME	
C3	3.5 000000, 50 V, TANTALUME	
Q1	JEDEC TA-206	873839
Q2	JEDEC TA-202	611308
Q3	JEDEC TA-202	611308
Q4	JEDEC TA-202	611308
Q5	JEDEC TA-202	611308
Q6	JEDEC TA-202	611308

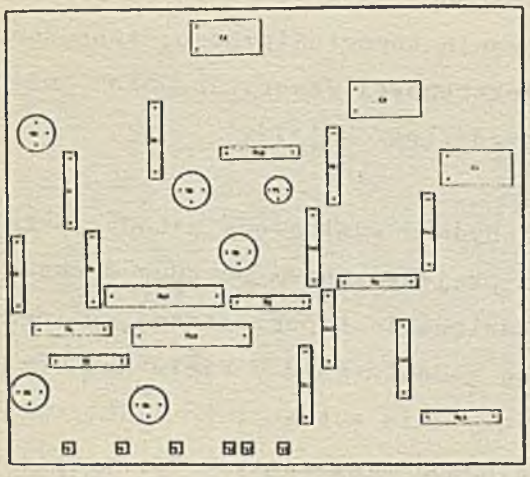
Rys. 7. Wykaz elementów układu sporządzony przez program ACCEL

Rozmieszczenie elementów na płytce dokonywane jest metodą zrównoważenia "sił i momentów". Wprowadza się cztery rodzaje umownych oddziaływań wzajemnych: "siły przyciągania" pomiędzy elementami połączonymi wspólną ścieżką przewodzącą, "siły odpychania" pomiędzy wszystkimi elementami a brzegami płytki oraz "momenty obrotowe" od każdego punktu połączenia elementu do wszystkich wspólnych punktów połączenia rozpatrywanego elementu z pozostałymi. Zakłada się przy tym, że "siły przyciągania" są wprost proporcjonalne do odległości.

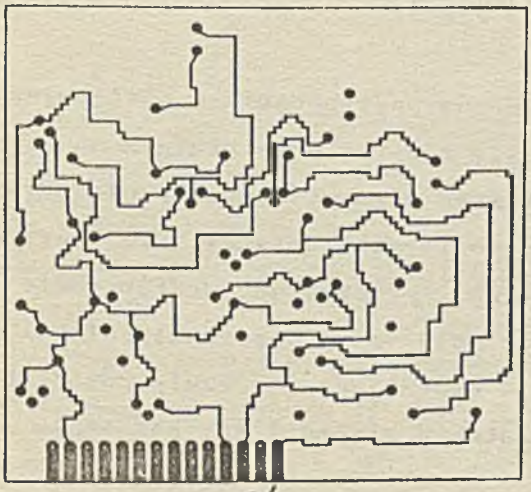
Na początku postępowania elementy są rozrzucone na płytce w sposób dowolny. Przemieszczanie elementów dokonywane jest metodą iteracyjną. W poszczególnych krokach przemieszczają się kolejno wszystkie elementy poddawane działaniu sił wypadkowych, przy czym ogranicza się wielkość przemieszczenia /do połowy odległości w kierunku ograniczenia/ w każdym kroku. Postępowanie zostaje zakończone, gdy w kolejnym kroku żaden z elementów nie przesuwa się na odległość większą od umownie przyjętej stałej.

		SCHEMATIC NO SANDIA			HOLE DRILLING LIST			DATE 08/09/87			PAGE 1		
X	Y	DIAM	X	Y	DIAM	X	Y	DIAM	X	Y	DIAM	X	Y
3.468	4.583	0.050	0.610	4.580	0.050	0.774	4.580	0.050	0.930	4.580	0.050	1.086	4.583
1.242	4.580	0.050	1.398	4.580	0.050	1.554	4.580	0.050	1.710	4.580	0.050	1.866	4.583
2.322	4.580	0.050	2.170	4.580	0.050	2.334	4.580	0.050	2.490	4.580	0.050	2.372	3.531
0.946	3.531	0.038	3.933	2.033	0.052	3.825	2.233	0.052	3.731	4.310	0.038	4.301	4.313
2.104	3.299	0.038	2.670	3.299	0.038	2.392	2.307	0.038	2.552	2.307	0.038	2.472	2.227
3.757	2.496	0.038	3.757	2.980	0.038	2.598	1.976	0.038	2.025	1.976	0.038	0.360	1.070
0.360	1.670	0.038	0.460	1.770	0.038	0.102	3.318	0.038	0.102	2.744	0.038	1.407	2.135
1.407	1.561	0.038	1.776	0.855	0.052	1.778	1.053	0.052	0.662	2.578	0.038	0.662	2.334
0.850	3.205	0.038	0.855	2.706	0.038	1.618	2.301	0.038	1.818	2.301	0.038	1.718	2.271
3.531	3.518	0.038	3.531	4.092	0.038	2.950	1.819	0.038	2.950	2.393	0.038	1.218	4.232
1.419	4.238	0.038	1.319	4.152	0.038	2.892	3.621	0.038	2.892	3.247	0.038	3.149	1.427
3.145	1.427	0.052	3.032	3.128	0.038	3.606	3.128	0.038	0.517	3.815	0.038	1.091	3.813
2.322	2.853	0.038	2.222	2.853	0.038	2.122	2.753	0.038	2.769	2.532	0.038	2.769	3.126
2.691	3.736	0.038	2.691	4.310	0.038	0.181	4.086	0.038	0.381	4.086	0.038	0.281	3.986
1.834	3.239	0.052	1.008	3.238	0.052	1.242	3.588	0.052	2.168	3.588	0.052		

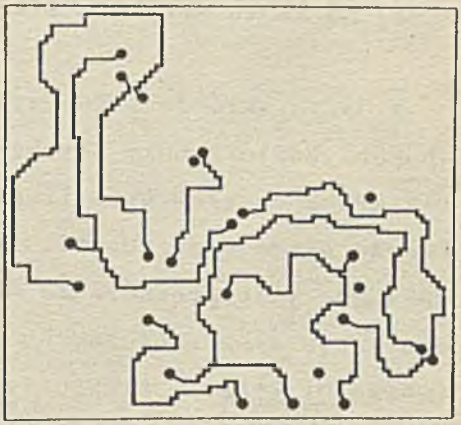
Rys. 8. Lista wiercenia otworów



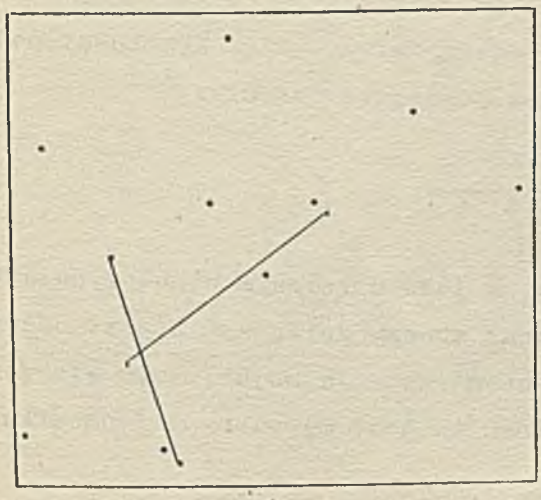
Rys. 9. Rysunek zestawieniowy z rozmieszczeniem elementów



a/



b/



c/

Rys. 10 a, b, c. Matryce połączeń na obu stronach płytki /a i b/ oraz rysunek mostków /c/

Program po dokonaniu rozmieszczenia przystępuje do rozwiązywania problemu połączeń. Przyjęto tutaj postępowanie kombinowane, wykorzystując dość dawno znaną metodę modelowania topograficznego, stosowaną do rozwiązywania zagadnień optyki elektronowej /Zworykin/ oraz metodę poszukiwania najkrótszych połączeń /Lee/ - [11].

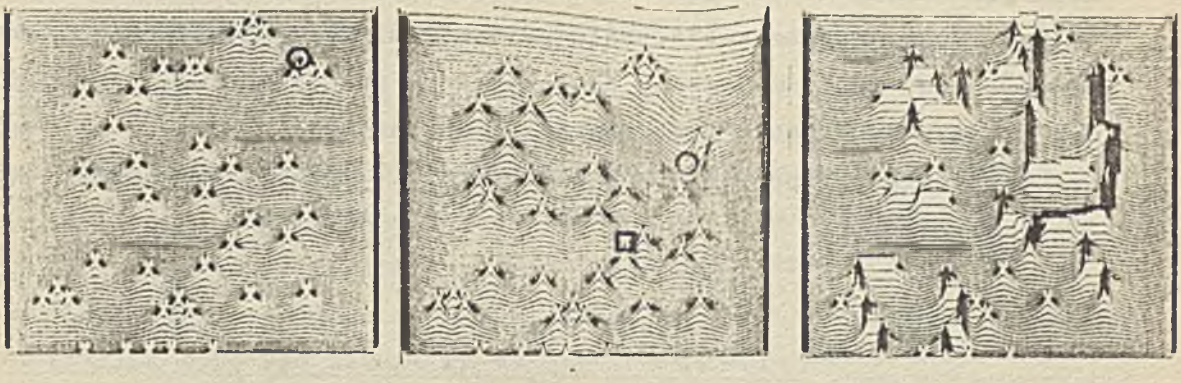
Autorzy programu ACCEL rozwinęli obydwie wymienione metody, uwzględniając konkretne potrzeby związane z projektowaniem obwodów drukowanych. Wprowadzono tzw. dynamiczne modelowanie topograficzne, polegające na budowaniu połączeń, startując jednocześnie z wielu punktów. Ponadto przewidziano możliwość rozwiązywania sytuacji konfliktowych, uwzględniono możliwość nakładania warunków wynikających ze specyficznych wymagań stawianych dla wybranych połączeń /np. żądanie minimalnej pojemności połączenia, odseparowanie ze względu na wyższe napięcie i tp/. Szczegółowy opis wspomnianej metody można znaleźć w cytowanej na końcu artykułu literaturze [3].

W dużym skrócie można wyjaśnić zasadę postępowania w następujący sposób. Powierzchnię płytki dzieli się rastrem prostokątnym na szereg drobnych elementów, Punkty montażowe elementów elektronicznych tworzą ostre wierzchołki. Fragmenty powierzchni płytki, dla których zakazane jest prowadzenie ścieżek - np. przewodzące powierzchnie elementów elektronicznych, stykające się z powierzchnią płytki - tworzą plateau, a krawędzie płytki - otaczające ją granie. Wyodrębnia się również takie punkty jak kontakty złącza i tp. Pozostałe elementy tworzą doliny.

Przykład tak uformowanej mapy podano na rys. 11a. Wysokości poszczególnych elementów określa się na podstawie formuły:

$$A = \frac{K}{CxD + 1}$$

gdzie A oznacza wysokość elementu, K jest ustalonym współczynnikiem, C określa stromość wierzchołków oraz zboczy dolin zaś D jest odległością między rozpatrywanym elementem rastru a najbliższym wierzchołkiem lub krawędzią płytki. Odległość ta jest mierzona w jednostkach rastru.



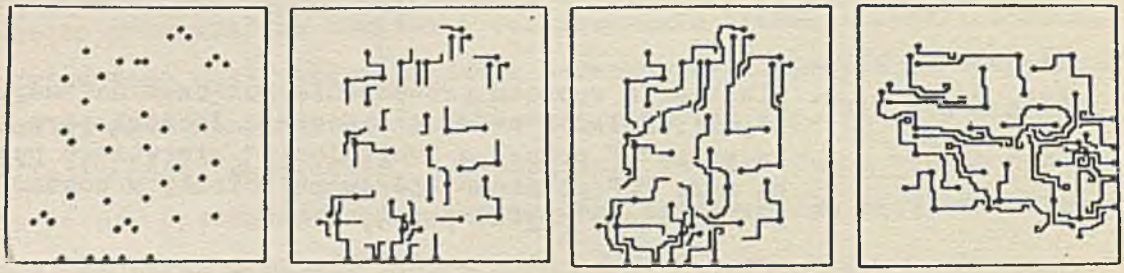
Rys. 11 a, b, c. Ilustracja sposobu prowadzenia połączeń na "mapie" płytki. Kółko i kwadrat oznaczają łączoną parę punktów. Na rys. "b" pokazano "wciśnięcie" pierwszego punktu. Na rys. "c" pokazano utorowaną ścieżkę w postaci "grani" łączącej obydwa punkty

Formuła, według której jest określana wysokość elementów może podlegać modyfikacjom w trakcie realizacji programu. Proces poszukiwania połączeń pomiędzy poszczególnymi punktami polega na poszukiwaniu, obrazowo mówiąc, linii największego spadku pomiędzy parami punktów, przy czym jeden z punktów pary jest przejściowo"wciskany" w głąb, pociągając za sobą otoczenie i tworząc coś w rodzaju leja, którego powierzchnia zachowuje względne wzniesienia i zagłębienia /rys. 11b/. Dokonywane jest to na drodze przeliczenia wysokości elementów rastru według odpowiednio zmodyfikowanej formuły, o czym już była mowa poprzednio.

Postępowanie takie zakłada rozrastanie się ścieżek po jednym elemencie rastru w każdym kroku, od wszystkich punktów jednocześnie. Przyrosty następują wyłącznie w dwu wzajemnie prostopadłych kierunkach wyznaczonych przez raster, przy czym za każdym razem analizowane jest bezpośrednie sąsiedztwo chwilowych wędrujących końców ścieżek. W przypadku spotkania się wędrujących końców tej samej ścieżki uważa się ją za zakończoną. Taka zakończona ścieżka tworzy nową grań łączącą kończące ją wierzchołki /rys. 11c/.

Jeśli końce ścieżek zostają przed połączeniem się uwięzione - ścieżki te przestają się chwilowo rozrastać i próba ich zakończenia jest

podejmowana dopiero w następnej fazie działania programu realizowanego już według nieco zmienionych zasad. Jeśli również w tych następnych fazach próba zakończenia połączenia nie zostanie uwieńczona powodzeniem, program wpisuje niezakończone połączenie na listę mostków wykonywanych ponad połączeniami drukowanymi. Kolejne kroki postępowania opisanego powyżej bardzo pobieżnie - ilustrują rysunki 12a, 12b, 12c, 12d.



Rys. 12. Kolejne fazy tworzenia połączeń

Jak już powiedziano na początku opisu działania programu ACCEL, daje on jako rezultat kompletną dokumentację, niezbędną dla produkcji obwodów drukowanych. Czas potrzebny na całkowite zaprojektowanie płytki drukowanej dla układu podanego na rys. 4 wynosi dla maszyny IBM 7090 ca 12 minut. Przewidziano przytem szereg udogodnień dla użytkownika. Tak na przykład projektant może określić na ilu płaszczyznach mają być rozmieszczone połączenia, to znaczy czy ma być użyty laminat jednostronnie foliowany, dwustronnie foliowany czy też ma być wykonane połączenie na obwodach wielowarstwowych. Można również stawiać specjalne wymagania co do sposobu rozmieszczania elementów /gęsty lub luźny montaż/. Normalny reżim przewiduje luźny montaż na laminacie dwustronnie foliowanym i przy ścieżkach rozrzuconych.

Aktualne ograniczenia programu ACCEL umożliwiają rozmieszczanie układów, zawierających do 500 elementów na płytkę, o nie więcej niż 200 ścieżkach dla warstwy drugiej i warstw dalszych /dla warstwy pierwszej praktycznie nie ma ograniczeń/ i o nie więcej niż 300 węzłach na płytkę. Przy rastrze 0,025'' maksymalna powierzchnia płytki może wynosić 90 cali kwadratowych. Elementy mogą być rozmieszczane wy-

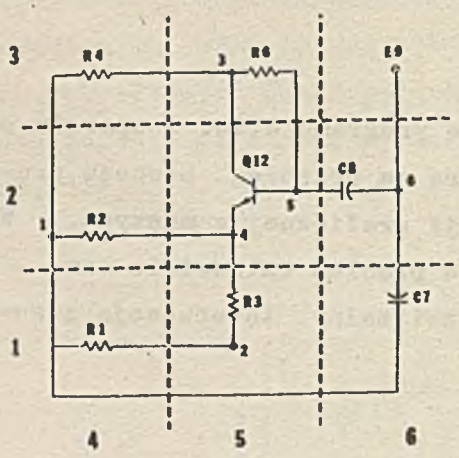
łącznie na jednej powierzchni płytki. Program został napisany dla maszyn IBM serii 7090-7094 w języku FORTRAN II /w 90%/, i w języku FAP /w 10%/.

Przewidywane jest dalsze ulepszanie programu m.in. w oparciu o umożliwienie oddziaływania konstruktora na przebieg procesu projektowania za pośrednictwem komunikacji graficznej z maszyną. W obecnej formie program albo rozwiązuje problem całkowicie samodzielnie, albo też uznaje go za nierozwiązalny. Interwencja z zewnątrz jest niemożliwa.

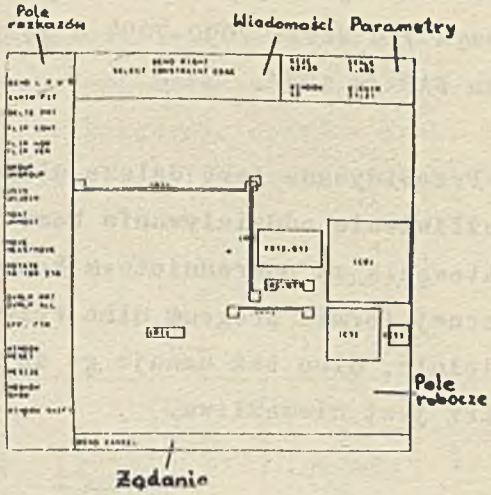
11.2. Projektowanie układów scalonych. Program CADIC

W odróżnieniu od programu ACCEL, który jak poprzednio wspomniano, działa w swej obecnej postaci całkowicie "samodzielnie" bez możliwości bezpośredniej interwencji projektanta, w sytuacji prowadzących do negatywnych rezultatów, CADIC jest typowym systemem konwersacyjnym, włączającym projektanta w cykl roboczy i dającym mu możliwość podejmowania na bieżąco /on line/ decyzji na pośrednich etapach projektowania. W tym celu program CADIC /Computer Aided Design of Integrated Circuits/ w szerokim zakresie korzysta z możliwości jakie daje komunikacja graficzna maszyna-człowiek-maszyna przy zastosowaniu wskaźników oscyloskopowych wyposażonych w pióro świetlne oraz specjalną klawiaturę.

System CADIC jest przeznaczony do pomocy przy projektowaniu układów scalonych. Punktem wyjścia - podobnie jak w przypadku programu ACCEL - jest gotowy schemat ideowy układu, który po odpowiednim zakodowaniu jest wprowadzony do maszyny wraz z danymi dotyczącymi typów i wartości elementów oraz ich wyprowadzeń, a także przybliżonym położeniem tych elementów, określonym na podstawie siatki naniesionej na schemat /rys. 13/.



Rys. 13. Przykładowy schemat



Rys. 14. Podział ekranu wskaźnika na pola

System CADIC na podstawie wprowadzonych do maszyny danych określa wymiary i kształt wzorów masek wszystkich elementów, przy czym wymiary oporników i kondensatorów są obliczane na podstawie danych wejściowych, zaś informacje dotyczące innych elementów są pobierane z kartoteki maszynowej, która może być w razie potrzeby rozszerzana o nowe typy elementów standardowych. Następnie na ekranie wskaźnika pojawia się w odpowiednim polu /rys. 14/ uproszczony obraz wszystkich elementów składowych w układzie pozwalającym na łatwą identyfikację elementów ze schematem ideowym oraz odpowiednio rozsuniętych celem uniknięcia nakładania się tych elementów na siebie.

Kiedy operator jest zainteresowany w szczegółach dotyczących poszczególnych masek, może spowodować wyświetlenie odpowiednich dodatkowych informacji na wskaźniku, za pośrednictwem klawiatury /rys. 15/ oraz pióra świetlnego. Tak na przykład naciśnięcie przycisku EMITTER wywołuje pojawienie się na wskaźniku obrazu maski dyfuzyjnej.

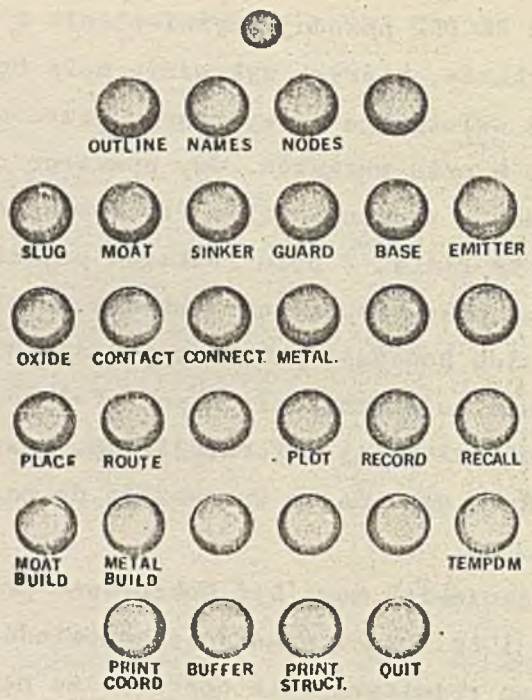
Do zmiany reżimu pracy służy cała grupa kluczy: np. kluczem PLOT można powodować wykreślenie przy pomocy pisaka obrazu pojawiającego się na wskaźniku oscyloskopowym, zaś kluczami RECORD i RECALL można wywoływać przepisywanie treści zobrazowania do pamięci dyskowej oraz powodować wyświetlanie zawartości tej pamięci w dogodnym czasie.

Naciśnięcie klucza RECORD powoduje wyświetlenie w lewej części ekranu repertuaru rozkazów, których wykonanie może być wywołane przez zbliżenie pióra świetlnego do wybranego przez operatora napisu mieszczącego się w polu rozkazów. Gdy operator chce dokonać przemieszczenia elementów wywołuje reżim "rozmieszczenie" przy pomocy naciśnięcia klucza PLACE. W polu rozkazów pojawia się odpowiedni dla tego reżimu pracy repertuar rozkazów i wskazanie piórem świetlnym jednego z nich powoduje pożądane przesunięcie. Wykonaniu operacji towarzyszy pojawienie się w polu "wiadomości" /lewa górna część ekranu/ odpowiedniej informacji, np. przypomnienie operatorowi jakie dodatkowe dane ma on wprowadzić do maszyny.

W reżimie "rozmieszczenie" mogą być dokonywane przesunięcia, obroty, zmiana skali i tp. przez wskazanie odpowiedniego rozkazu. Po zakończeniu procesu rozmieszczania operator za pomocą naciśnięcia klucza NODES /po uprzednim skasowaniu nazw elementów/ wywołuje wyświetlenie numerów pól kontaktowych poszczególnych elementów, które mają być połączone ścieżkami metalizowanymi. Następną fazą rozpoczynającą się naciśnięciem klucza ROUTE jest wykonywanie połączeń. W przypadku, gdy operator decyduje się na automatyczne prowadzenie połączeń, wskazuje piórem świetlnym odpowiedni rozkaz.

Algorytm wykonywania połączeń stosowany w systemie CADIC jest pewną modyfikacją algorytmu opracowanego przez C.Y. Lee [11]. Na tym samym algorytmie bazuje poprzednio opisywany program ACCEL.

W przypadku trudności z wykonywaniem połączeń program sygnalizuje ten fakt przez wpisanie w dowolnym polu ekranu wskaźnika numeru połączenia ze znakiem minus. Operator może dla usunięcia blokady powrócić do reżimu "rozmieszczenie" i dokonać odpowiednich przemieszczeń, na przykład obrotu elementów w celu zbliżenia łączonych pól kontaktowych tych elementów,po czym można ponownie wrócić do programu automatycznego poszukiwania połączeń. Możliwe jest również ręczne wprowadzanie poprawek za pomocą pióra świetlnego.



Rys. 15. Klawiatura systemu CADIC

Końcowy wynik działania programu CADIC w postaci geometrycznego obrazu masek nie może być bezpośrednio wykorzystywany do wytwarzania masek produkcyjnych na drodze fotograficznego zmniejszenia ze względu na zbyt małą dokładność otrzymywanych rysunków. Na razie otrzymywane wyniki dają projektantowi jedynie wskazówki, jak mają wyglądać maski. W przyszłości przewiduje się połączenie wyników działania programu CADIC z aparaturą do bezpośredniego wytwarzania masek.

Opisy programów związanych bezpośrednio z produkcją masek dla integracji wielkoskalowej /LSI - Large Scale Integration/ można znaleźć między innymi w artykułach J.W. Lathropa i innych [27], [5].

11.3. Zagadnienia rozpakietowania, rozmieszczenia i prowadzenia połączeń

Poprzednio omówiono problemy związane z nadaniem zamkniętej postaci konstrukcyjnej zespołom maszyn i urządzeń cyfrowych najniższego

szczebla, składającym się z elementów elektronicznych takich jak oporniki,kondensatory, tranzystory, diody i tp., zarówno w montażu konwencjonalnym jak i w przypadku układów scalonych.

Obecnie zostaną rozpatrzone pokrótce ważniejsze aspekty konstrukcji zespołów szczebla wyższego. W tym miejscu należy zwrócić uwagę na dość umowny podział na szczeble. Tak na przykład ostatnio lansowana integracja wielkoskalowa/LSI/ powoduje umieszczenie na jednej podstawie montażowej wspomnianego najniższego szczebla złożonych funkcjonalnie układów, które są tworzone bezpośrednio z elementów elektronicznych. W tym przypadku problem połączeń pomiędzy poszczególnymi układami podstawowymi takimi jak przerzutniki, inwertery i tp., rozwiązywany jest w procesie scalania, podczas gdy dla "zwykłych", pojedyńczych układów scalonych przenoszony był do fazy konstruowania pakietów /modułów/ wykonywanych w postaci obwodów drukowanych, jedno lub wielowarstwowych.

Wzrastająca stale szybkość pracy układów cyfrowych sprawia również, że problemy walki z opóźnieniami i zakłóceniami typowe od niedawna dla połączeń w większych zespołach /ramy, szafy/ przenoszą się stopniowo na połączenia wykonywane w ramach modułów.

Dla większej przejrzystości ograniczymy się poniżej do rozważań dotyczących automatyzacji projektowania rozpakietowania oraz rozmieszczenia i prowadzenia połączeń typowych dla realizacji fizycznej maszyn i urządzeń cyfrowych drugiej generacji.

Zakłada się zwykle, że w momencie przystąpienia do implementacji fizycznej dysponuje się kompletną dokumentacją logiczną, uwzględniającą ograniczenia co do maksymalnej dopuszczalnej ilości wejść, maksymalnego wzmocnienia logicznego, opóźnienia i innych cech zestawu układów podstawowych, z których ma być budowana maszyna czy urządzenie. W zależności od stopnia zaawansowania rozwiązań technicznych, stanowiących podstawę do dalszych etapów projektowania, można rozpatrywać dwa warianty:

1. Dokonanie wyboru optymalnego zestawu pakietów /modułów/ standardowych ze względu na minimalną ilość różnych typów. Wybór

taki jednocześnie zapewnia wysoki stopień wykorzystania zmontowanych na pakietach układów.

2. Projektowanie w oparciu o pozostający w dyspozycji, poprzednio dobrany i sprawdzony w praktyce zestaw pakietów standardowych.

Wariant "1" może uwzględniać specyficzne własności projektowanego urządzenia, wynikające ze schematu logicznego, np. występowanie dużej liczby określonych typów układów podstawowych w ustalonej konfiguracji.

Dalsze etapy projektowania obejmują przypisanie poszczególnych układów podstawowych, występujących na schematach logicznych, oddzielnym pakietom standardowym, rozmieszczenie pakietów na ramach montażowych, opracowanie tabel połączeń i wreszcie określenie sposobu wykonywania połączeń międzypakietowych. Wymienione czynności dotyczą przypadku typowego dla płaskiego montażu pakietów, kiedy na pakietach nie występują połączenia logiczne /tzn. wszystkie wejścia i wyjścia układów podstawowych umieszczonych na pakiecie wyprowadzane są na zewnątrz pakietu/.

Dla wyżej wymienionych czynności opracowano mniej lub bardziej doskonałe metody projektowania automatycznego. Zasadnicza trudność polega jednakże na tym, że wszystkie te czynności są wzajemnie powiązane. Ponadto ścisłe metody optymalizacyjne prowadzą zwykle do nader długotrwałych obliczeń, dyskwalifikujących ich praktyczną przydatność.

Jak można wnioskować z odnośnej literatury oraz własnych obserwacji i bardzo skromnych doświadczeń, najbardziej efektywne jest postępowanie heurystyczne bazujące na algorytmicznym, optymalnym rozwiązywaniu problemów cząstkowych. Należy w tym miejscu dodać, że bogato w publikacjach przeglądowych cytowana literatura omawianego przedmiotu jest w ogromnej części niedostępna /wewnętrzne raporty znanych firm zagranicznych/. Dotyczy to przede wszystkim najbardziej interesujących opracowań o wyraźnie konkretnym i praktycznym charakterze. Nie brak natomiast publikacji czysto teoretycznych, których wykorzystanie w realnych warunkach nie jest możliwe.

Zajmiemy się nieco bliżej zagadnieniami rozmieszczania i prowadzenia połączeń między pakietami. Jeśli mamy N danych obiektów, które można rozmieścić na M pozycjach, to ilość możliwych rozmieszczeń wynosi jak wiadomo M!/M-N/! /przyczem M ≥ N/. Zwykle jako kryterium optymalizacji rozmieszczenia przyjmuje się kryterium minimalizacji połączeń ze względu na związaną z tym redukcję pojemności międzyprzewodowych i przesłuchów. Z kolei przy rozwiązywaniu problemu połączeń poszukuje się takiego minimalnego /w sensie sumy długości gałęzi/ drzewa o n-1 gałęziach, które łączy n wierzchołków o ustalonych położeniach. Ilość wszystkich możliwych drzew łączących n wierzchołków wynosi $n^{/n-2/}$.

Przy rozwiązywaniu "czystego" problemu rozmieszczenia zakłada się, że określone jest drzewo minimalne, natomiast przy rozwiązywaniu "czystego" problemu połączeń zakłada się, że określone jest rozmieszczenie łączonych punktów. Oczywiście w rzeczywistości obydwa problemy są wzajemnie powiązane i w zasadzie powinny być rozwiązywane jednocześnie, powstają jednak przy tym poważne ograniczenia co do ilości rozmieszczanych obiektów [9].

Efektywne rozwiązanie problemu połączeń przy założeniu, że znane jest rozmieszczenie, a także w przypadku prowadzenia połączeń w jednej płaszczyźnie i po liniach prostych, podano w pracy [8]. Przyjęte tam postępowanie jest optymalne dla przypadku, gdy krotność węzłów konstruowanego drzewa połączeń nie przekracza trzech. W przeciwnym przypadku rozwiązanie może być nieoptymalne.

Jak można sądzić na podstawie pośrednich informacji podanych w [1] wspomniane powyżej postępowanie jest bardzo zbliżone do opisanego w [10]. Sprawa znacznie się komplikuje, gdy przy rozmieszczeniu pragnie się uwzględnić optymalizację połączeń.

Zadanie rozwiązuje się wówczas zwykle w dwu fazach: pierwsza obejmuje wstępne rozmieszczenie, druga zaś korektę rozmieszczenia. W skrajnym przypadku wstępne rozmieszczenie może być dokonane w sposób przypadkowy, po czym musi być przeprowadzona korekcja. Jeśli natomiast pierwsza faza nie opiera się na losowym rozmieszczeniu obiektów, wówczas dokonywane jest ono w dwu krokach. Pierwszy z nich obejmuje wybór

kolejnego obiektu, którego położenie ma być określone.

Drugi krok polega na wyborze miejsca, w którym ma być umieszczony wybrany obiekt. W rozpatrywanej klasie problemów, do której ograniczyliśmy nasze rozważania, obiektem może być np. układ podstawowy, który należy umieścić na odpowiednim pakiecie posiadającym taką możliwość. Należy jednak uwzględniać tutaj dodatkowe warunki, wynikające z przyjętych realnych ograniczeń /ilość układów na pakiecie, ilość kontaktów i tp./.

Wybór miejsca ulokowania obiektu jest dokonywany zwykle według kryterium minimalizacji pewnej funkcji odległości /długości połączeń/. Proces ten jest powtarzany aż do rozmieszczenia wszystkich obiektów. Jeśli wynik tego początkowo rozmieszczenia nie jest zadawalający, realizuje się fazę korekcyjną.

Druga faza postępowania polega zwykle na iteracyjnym procesie przemieszczania dwu lub więcej obiektów. Dla każdego przemieszczenia określa się przytem miarę ulepszania, opierającą się na wzajemnych połączeniach obiektu przemieszczanego z ustalonymi. W niektórych przypadkach dopuszcza się tylko takie przemieszczenia, które dają maksymalne ulepszenie. Jako miarę tego ulepszania zwykle wprowadza się pewną funkcję odległości - np. długości gałęzi minimalnego drzewa połączeń.

Zagadnienie optymalizacji połączeń zależy w dużej mierze od wielu czynników związanych ze stosowaną konstrukcją i technologią, trudno więc mówić o uniwersalnych metodach projektowania programowego.

12. Sytuacja krajowa i potrzeby w zakresie automatyzacji projektowania

Zagadnienia omawiane w niniejszym artykule stanowią przedmiot coraz liczniej pojawiających się publikacji. Trudno ostatnio znaleźć numer czasopisma fachowego z zakresu elektroniki, względnie automatycznego przetwarzania informacji, w którym nie można byłoby spotkać przynajmniej wzmianki na temat automatyzacji projektowania. Listopa-

dowy numer Proceedings of the IEEE z 1967 roku jest w całości poświęcony tym zagadnieniom: projektowaniu wspomaganemu przez maszynę.) ile kilka lat temu artykuły z tego zakresu miały z reguły czysto teoretyczny charakter [6], [11], to najnowsze publikacje mają coraz to bardziej użytkowy i praktyczny charakter [3], [4], [5]. Pozostaje to między innymi w związku z coraz powszechniejszym wprowadzaniem integracji wielkoskalowej /LSI - Large Scale Integration/.

Jak na tym tle kształtuje się sytuacja w kraju? Na pytanie to trudno dać odpowiedź pełną i ścisłą. Wiadomo natomiast z całą pewnością, że problematykę tę podjęto w conajmniej kilku ośrodkach. Jak dotychczas prace te są mało zaawansowane i prowadzone raczej wycinkowo, ale obserwuje się wyraźne tendencje do poszukiwania rozwiązań użytecznych dla celów praktyki. Już w 1963 roku pojawiła się praca [15] poświęcona modelowaniu analizatora równań różniczkowych na maszynie cyfrowej. Równolegle z niniejszym artykułem zgłoszono do druku publikacje poświęcone modelowaniu systemów elektronicznych [16] oraz automatycznemu projektowaniu rozmieszczenia połączeń [8]. Ponadto zespół pracowników Instytutu Maszyn Matematycznych podjął szeroko zakrojone prace zmierzające do formalizacji opisu struktury logicznej w celu przygotowania następnych etapów automatyzacji projektowania, takich np. jak rozpakietowanie, rozmieszczenie, prowadzenie połączeń, symulacja działania.

W centrum Obliczeniowym PAN prowadzi się również interesujące prace, na razie przede wszystkim o charakterze rozpoznawczym, w dziedzinie automatyzacji projektowania logicznego. Ośrodek ten organizuje w dniach 27 - 28 czerwca br. Sympozjum na temat automatycznego projektowania maszyn cyfrowych. Można mieć nadzieję, że sympozjum to przyczyni się do wypracowania najwłaściwszych dla specyficznej sytuacji krajowej form współpracy i koordynacji badań.

Specyfika ta polega przede wszystkim na znacznym rozproszeniu zespołów badawczych, zainteresowanych omawianą problematyką. Jak się wydaje, rozproszenie to jest jedną z poważnych przyczyn niedostatecznego

obecnie stanu badań i zastosowań praktycznych w dziedzinie automatycznego projektowania w naszym kraju. Rozproszenie to z kolei jest wywołane koniecznością realizowania bieżących zadań przez poszczególne placówki, które nie są w stanie wyodrębnić dostatecznie silnych grup, mogących rozwiązać kompleks zagadnień na tyle obszerny, aby wyniki mogły dać widoczne efekty praktyczne.

Niewątpliwie jednak zbliża się chwila, kiedy sytuację tę pomimo trudności organizacyjnych, trzeba będzie zmienić. O ile dotychczasowe projekty z dziedziny API, realizowane w kraju w ramach tzw. drugiej generacji maszyn, mogły się jeszcze obywać bez oparcia o automatyczne projektowanie, to nie ma się co łudzić, że bez doprowadzenia do absurdów ekonomicznych uda się utrzymać dotychczasowy stan rzeczy przy przejściu do projektowania maszyn i urządzeń trzeciej generacji. Dotyczy to wszystkich nieomal etapów projektowania, a w szczególności projektowania hardware'u.

Możliwe są w zasadzie dwa rozwiązania:

- Sprowadzenie niezbędnych urządzeń i programów umożliwiających bez żadnych dodatkowych prac projektowanie maszyn cyfrowych trzeciej generacji w sposób nowoczesny.
- Rozwiązanie zadań związanych z przygotowaniem bazy technicznej a przede wszystkim programowej własnymi, krajowymi środkami.

Rozwiązanie pierwsze wydaje się mało prawdopodobne z wielu względów. Poszukiwanie rozwiązania na drugiej drodze może być dokonane w dwóch wariantach. Pierwszy z nich, teoretycznie poprawniejszy, polegałby na próbie ścisłego sformułowania zadań kompleksowych oraz ich rozdziału pomiędzy poszczególne ośrodki, przy jednoczesnym pełnym ujednoliceniu form działania i środków takich jak stosowane języki, opisy, formalizmy.

Drugi - bardziej realistyczny - bazowałby na zapewnieniu niezbędnego minimum wymiany doświadczeń i usług, zapewniających wszystkim zainteresowanym instytucjom osiąganie efektów praktycznych, bez istotnego zwiększania środków, przeznaczonych na ten cel.

Mało realnym wydaje się pokładanie nadziei w tym, że któryś z krajowych potentatów z zakresu automatycznego przetwarzania informacji mógłby w rozsądnym terminie rozwiązać samodzielnie problem automatycznego projektowania w rozmiarach niezbędnych dla potrzeb projektowania maszyn trzeciej generacji. Przedstawiony w niniejszym artykule, z konieczności bardzo skrótowo i w sposób niepełny, zarys nowoczesnych metod projektowania maszyn cyfrowych nie pozwala na wyrobienie poglądu na wagę i stopień trudności omawianych zagadnień. Dopiero wnikliwe przestudiowanie dostępnej literatury może dać bardziej pełny obraz tych możliwości i korzyści jakie te metody dają oraz tych trudności, jakie należy pokonać, aby stworzyć system użyteczny i zaspakający przynajmniej podstawowe nasze potrzeby.

Literatura

[1] M.A. BREUER: General Survey of Design Automation of Digital Computers. Proc. IEEE, December 1966.

[2] J. MITTLEMAN: Computer-aided Design: Part 11., Short Cuts to IC's and p-c Boards. Electronics, September 4, 1967.

[3] C.J. FISK, D.L.CASKEY, L.E. WEST: Taking the Puzzle out of p-c Design. ibid, [2].

[4] A. SPITALNY, M. GOLBERG: Drawing Board for IC's, ibid. [2].

[5] H. FREITAG: Generating IC's Masks Automatically, ibid. [2].

[6] W.S. LINSKIJ: Ałgoritmičeskoje projektirovanie vyčišlitielnych cifrovych ustrojstv. AN CCCP, Vyčisl. Cientr. Soobsčenija po vyčislitielnoj technikie, vypusk 2.

[7] N.E. WISEMAN: Application of List Processing Methods to the Design of Interconnections for a Fast Logic System. University Math. Lab. Cambridge, Techn. Mem. No. 63/3, June 1963.

[8] J. FIETT, B. MAZUREK: Program opracowania sposobu łączenia punktów montażowych według tabel połączeń. Złożone w redakcji Prac PIT.

[9] M.A. BERUER: The Formulation of Some Allocation and Connection Problems as Integer Programs. Naval Research Logistics Quarterly, vol. 13, March 1966.

[10] M.N. WEINDLING: A Method for Best Placement of Units on a Plane. 1964 Share Design Automation Workshop Proc. and Douglas Paper 3108, Douglas Aircraft Co., Santa Monica, California.

[11] C.Y. LEE: An Algorithm for Path Connections and its Application. IRE Trans. EC, vol. EC-10, September 1961.

[12] E.L. LAWLER: Electrical Assemblies with a Minimum Number of Interconnections. IRE Trans. on Electronic Computers, vol. EC-11, February 1962.

[13] F.H. HARDIE, R.J. SUCHOCKI: Design and Use of Fault Simulation for Saturn Computer Design. IEEE Trans. on Electronic Computers, vol. EC-16, nr 4. August 1967.

[14] S. CHOROMAŃSKI et al. pod redakcją S. SAWICKIEGO: Opis Podstawowego Języka Symbolicznego - Maszyna ZAM-41 - System Operacyjny SO41B. Opracowanie T-OF-378-D, Instytut Maszyn Matematycznych, Warszawa 1967.

[15] R. MARCZYŃSKI, W. PULCZYN: Modelowanie cyfrowego analizatora równań różniczkowych na maszynie cyfrowej. Algorytmy, Zeszyt Specjalny Nr 1, Warszawa 1963.

[16] J. JANOWSKI, T. SINKIEWICZ, J. WINKOWSKI: Modelowanie systemów elektronicznych. Złożone w redakcji Archiwum Automatyki i Telemechaniki.

[17] E. MANNING: On Computer Self-diagnosis. Part I - Experimental Study of a Processor, Part II - Generalizations and Design Principles. IEEE Trans. on Electronic Computers, vol. EC-15, nr 6, December 1966.

[18] K. MALING, E.L. ALLEN: A Computer Organization and Programming System for Automated Maintenance. IEEE Trans. on Electr. Computers, vol. EC-12, nr 6, December 1963.

[19] W.C. CARTER et al. : Design of Serviceability Features for the IBM System/360. IBM Jour. of Res. and Dev., April 1964.

[20] J.P. ROTH et al. : Programmed Algorithms to Compute Tests to Detect and Distinguish between Failures in Logic Circuits. IEEE Trans. on Electr. Comp., vol. EC-16, nr 5 October 1967.

[21] J.P. ROTH: Diagnosis of Automata Failures: a Calculus and a Method. IBM Jour. of Res. and Dev., vol. 10, July 1966.

[22] K.E. IVERSON: A Programming Language. Wiley, New York 1963.

[23] D.T. ROSS: The AED Approach to Generalized Computer-aided Design. Report ESL-R-305, April 1967, Massachusetts Institute of Technology.

[24] W.H. WATTENGURG: Design Automation for Computer Software. IEEE Trans. on Electr. Comp., vol. EC-15, nr 3, June 1966.

[25] J. MYSIOR, S. WALIGÓRSKI: LOGOL - język algorytmiczny dla przetwarzania napisów. Biul. IMM Nowości Techniczno-Ekonomiczne Rok IV, nr 2/10/, czerwiec 1965.

[26] R.A. BROOKER et al.: The Compiler Compiler. Annual Review in Automatic Programming 3, Pergamon Press, 1963.

[27] J.W. LATHROP et al. : A Discrectionary Wiring System as the Interface between Design Automation and Semiconductor Array Manufacture. PIEEE, vol. 55, No 11, November 1967.

Mgr inż. JERZY DAŃDA
Mgr inż. ZDZISŁAW WRZESZCZ
Zakład Techniki Cyfrowej IMM

621.317:681.327.6

H A R D W A R E I S O F T W A R E TECHNICZNO-FUNKCJONALNYCH BADAŃ SYSTEMÓW PAMIĘCIOWYCH

1. Wstęp

Zagadnienia projektowania i konstrukcji maszyn cyfrowych posiadają bogatą literaturę fachową. Zupełnie inna sytuacja istnieje w dziedzinie urządzeń pomocniczych, niezbędnych do produkcji, uruchamiania i eksploatacji maszyn cyfrowych i poszczególnych zespołów wchodzących w skład tych maszyn. Literatura tego tematu jest nadzwyczaj uboga. Praktycznie dostępne są wyłącznie ulotki firmowe, opisujące konkretne realizacje takich urządzeń. Brak jest natomiast pozycji opisujących zasady ich projektowania.

Niniejszy artykuł stanowi próbę częściowego choćby wypełnienia luki w tym zakresie. Ograniczymy się w nim jednak do zagadnień występujących przy techniczno-funkcjonalnych badaniach pamięci operacyjnych i pamięci stałych maszyn cyfrowych.

Starając się sprecyzować w tytule artykułu zakres naszych zainteresowań użyliśmy pewnych sformułowań, które wymagają wyjaśnień. Przez h a r d w a r e rozumiemy tutaj aparaturę pomiarową, używaną przy tych badaniach, przez s o f t w a r e - w szerokim znaczeniu tego słowa - rozumiemy metodykę badań, oraz wszystkie szczegółowe przepisy służące do realizacji programu badań techniczno-funkcjonalnych, w szczególności w przypadku stosowania aparatury programowanej wliczyć możemy tu oprogramowanie w sensie klasycznym.

Wytwarzaniu i eksploatacji maszyn cyfrowych względnie ich modułów towarzyszą nieprzerwanie dwa typy badań: techniczne i funkcjonalne. Pierwsze mają na celu stwierdzenie zgodności aktualnych parametrów wyrobu z

parametrami zazwyczaj określanymi przed rozpoczęciem projektowania urządzenia i zawartymi w wymaganiach technicznych dotyczących tego urządzenia. Drugie natomiast mają na celu stwierdzenie poprawności funkcjonowania urządzenia w rzeczywistych lub symulowanych warunkach pracy.

Aby zorientować Czytelnika o jaki typ urządzeń realizujących te badania nam chodzi, wymienimy niektóre z nich:

- Urządzenia do badania pakietów mieszczących złożone układy logiczne,
- urządzenia do badania pakietów specjalnych, np. wchodzących w skład pamięci,
- urządzenia do badania pamięci.

Należy podkreślić, że największą złożonością charakteryzują się urządzenia ostatniej grupy, służące do badania pamięci, przy czym z reguły wymagane są oddzielne przyrządy dla każdego typu pamięci. W dalszym ciągu przedmiotem szczególnej naszej uwagi będą urządzenia do sprawdzania pamięci o natychmiastowym dostępie.

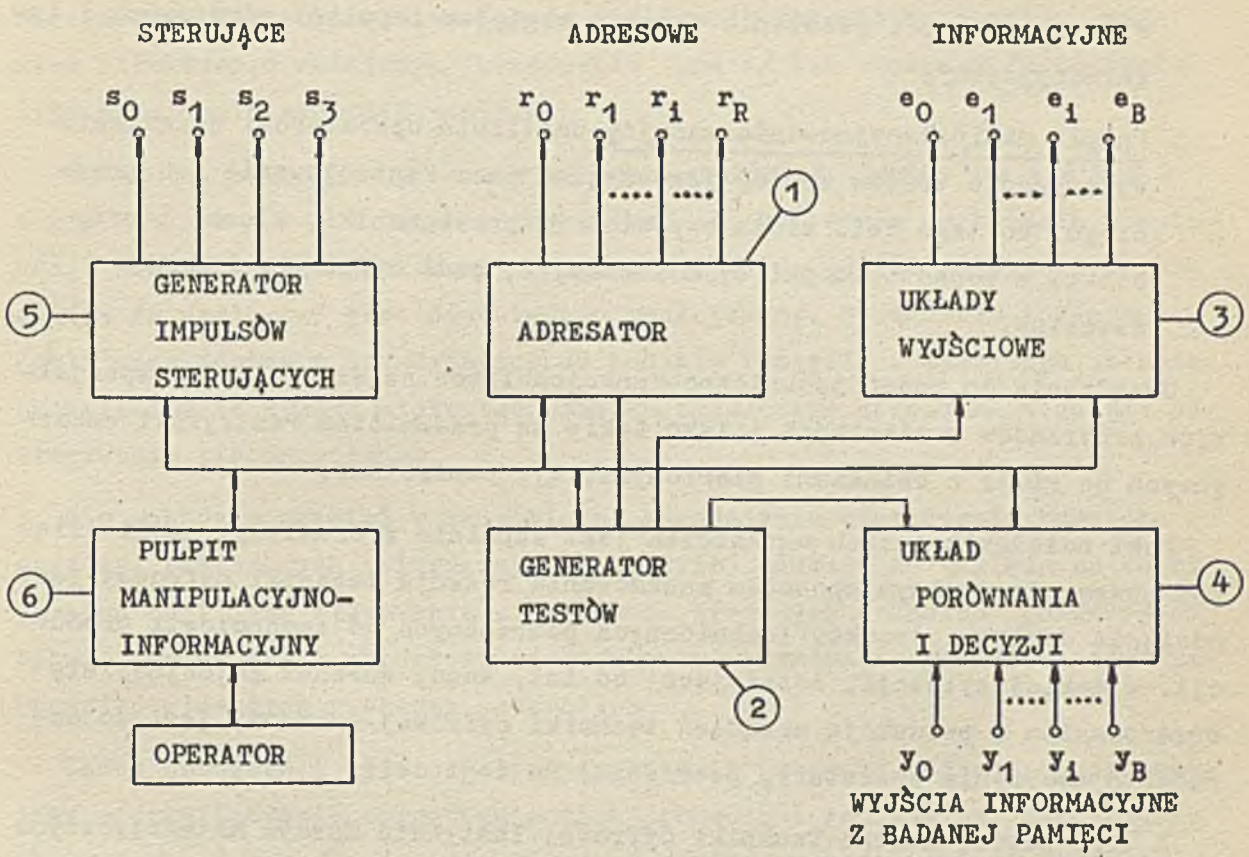
2. Urządzenia do badania pamięci operacyjnych /testery/

Urządzenia tego typu w literaturze anglosaskiej noszą miano m e m o r y e x e r c i s e r. Nazwa ta trudna jest do wiernego przekładu na język polski. Określa ona urządzenia używane przy uruchamianiu oraz badaniu pamięci operacyjnych i zawierające zwykle następujące bloki funkcjonalne /rys. 1/:

- Adresator /1/
- generator testów /2/
- układy wyjściowe /3/
- układy porównania i decyzji /4/
- generator impulsów sterujących /5/
- pulpit manipulacyjno-informacyjny /6/

Zadania wymienionych wyżej bloków są następujące:

- Adresator generuje sekwencję adresów, które podawane są na wejścia badanej pamięci.



Rys. 1. Schemat blokowy testera /typu m e m o r y e x e r c i s e r/ do sprawdzania pamięci

- Generator testów produkuje informację, która następnie zapisywana jest w miejscach pamięci wyznaczonych przez adresator. Generatory testów zazwyczaj umożliwiają zapis do pamięci ograniczonych postaci informacji, np. samych jedynek, samych zer, oraz tzw. testów najtrudniejszych, co dla celów badawczych jest zupełnie wystarczające.
- Układy wyjściowe dostarczają na wejścia informacyjne pamięci impulsów o odpowiednich parametrach napięciowo-czasowych.
- Układ porównania i decyzji przyjmuje informację odczytaną z pamięci, porównuje ją z informacją pochodzącą z generatora testów i wydaje decyzję odnośnie dalszego kontynuowania badania względnie zatrzymania biegu testu i wyświetlenia informacji o wykrytym błędzie.
- Generator impulsów sterujących wytwarza impulsy sterujące wykonaniem poszczególnych mikrooperacji w badanej pamięci. Zazwyczaj istnieje

możliwość regulacji pewnych czasowych parametrów impulsów sterujących /szerokość, położenie czasowe względem impulsów adresowych i informacyjnych/.

- Pulpit manipulacyjno-informacyjny umożliwia operatorowi urządzenia wykonywanie testów według jego życzeń oraz kontrolowanie ich przebiegu. Do tego celu służą odpowiednie przełączniki, klucze i przyciski, a ponadto lampki sygnalizacyjne, bądź cyfrowe wskaźniki świetlne.

Urządzenia do badań techniczno-funkcjonalnych należą do grupy specjalnych przyrządów pomiarowych i jako takie są przedmiotem restrykcji embargowych na równi z obiektami mierzonymi, tj. pamięciami.

Fakt nałożenia takich ograniczeń jest zupełnie zrozumiały, jeśli wziąć pod uwagę, że realnym sposobem zahamowania rozwoju techniki cyfrowej jest odcięcie dopływu środków technicznych potrzebnych do technologii produkcji. W takiej sytuacji, istniejącej od lat, każdy ośrodek zajmujący się opracowaniem i produkcją urządzeń techniki cyfrowej zmuszony jest do doraźnego tworzenia aparatury, potrzebnej do tego celu.

Z tego powodu Zakład Techniki Cyfrowej Instytutu Maszyn Matematycznych na przestrzeni ostatnich lat - równolegle do opracowania konstrukcji i technologii pamięci operacyjnych - prowadził prace nad wyspecjalizowanymi generatorami programowanymi typu m e m o r y e x e r c i s e r, oznaczonymi symbolem MOPS.

Poszczególne egzemplarze tych generatorów związane były z kolejno opracowywanymi typami pamięci operacyjnych oraz ich wersjami. Ogólnie można powiedzieć, że ich konstrukcja uwzględniała wymagania funkcjonalne stawiane generatorom tego typu z pewnymi odstępstwami wynikającymi z faktu, że konstruowane pamięci były pamięciami z wybieraniem liniowym. Przyrządy typu MOPS będą krótko omówione w końcowym rozdziale niniejszej pracy.

3. Metodyka techniczno-funkcjonalnych badań jako s o f t w a r e aparatury kontrolno-pomiarowej maszyn cyfrowych

Wróćmy powtórnie do tytułu, a mianowicie do słowa s o f t w a r e. Jego znaczenie wyjaśniliśmy już we Wstępie. Należy jeszcze uzasadnić potrzebę stosowania s o f t w a r e' u.

Urządzenie pomiarowe jest zarazem elementem złożonego systemu, a także narzędziem. Narzędzie to, mając do realizacji określone funkcje, może mieć strukturę o mniejszym /urządzenie typu A/ lub większym /urządzenie typu B/ stopniu skomplikowania.

Znane są rozwiązania urządzeń amerykańskich typu m e m o r y e x e r c i s e r o stopniu skomplikowania zbliżonym do stopnia komplikacji średniej wielkości maszyny cyfrowej. Wtedy udział człowieka w obsłudze /w systemie/ jest odpowiednio zmniejszony. Urządzenie pomiarowe /cały czas mówimy o urządzeniach do badania pamięci/ o mniejszym stopniu skomplikowania wymaga większego wysiłku umysłowego człowieka w czasie posługiwania się przyrządem.

Nie wchodząc głębiej w ekonomiczne zagadnienia użytkowania naszych urządzeń pomiarowych twierdzimy, że bardziej słuszne ze względu na koszt eksploatacji jest stosowanie w warunkach krajowych urządzeń pomiarowych typu A, to znaczy o uproszczonej strukturze. Wymaga to jednak stosowania bardziej złożonych procedur badawczych.

Chcąc uprościć i ujednolicić sporządzanie procedur badawczych proponujemy sformalizowanie opracowania tych procedur i ich zapisu. Stąd użycie słowa s o f t w a r e dla urządzeń pomiarowych do pamięci ferrytowych.

Doświadczenie uczy, że opracowanie software'u angażuje nie mniejsze środki niż opracowania czysto konstrukcyjne. Zamierzone prace nad softwarem dla urządzeń pomiarowych muszą być z tego względu realizowane etapami. Chcąc określić zawartość opracowania software'u dla urządzeń pamięciowych, podamy najpierw cechy jakie powinny charakteryzować metodykę badań pamięci operacyjnych i pamięci stałych, która jest dla nas interesująca. Metodyka ta winna posiadać następujące cechy:

a. Przy badaniu prototypu urządzenia /pamięci operacyjnej lub stałej/ według opracowanej metodyki winno być możliwe pełne sprawdzenie wszystkich cech technicznych urządzenia, włączając takie, które tradycyjnie są badane dopiero po dołączeniu pamięci do maszyny cyfrowej, z którą pamięć ma w finalnym systemie współpracować.

b. Metodyka winna umożliwiać sprawne wykonanie wszystkich technicznych badań prototypu, interesujących w sposób szczególny konstruktora urządzenia. Jako przykłady można wymienić badanie obszarów

sprawności, ich zmiany z upływem czasu i pod wpływem czynników mechanoklimatycznych;

c. Przy badaniu serii informacyjnej winno się używać odpowiednio uproszczonej wersji metodyki, uwzględniającej wykonanie badania w ograniczonym czasie i z ograniczonym wyposażeniem w urządzenia pomiarowe. Tym niemniej owa uproszczona wersja winna dawać wyniki rzetelne i dostarczać maksimum informacji. Sprawa ta jest szczególnie podkreślana, ponieważ przy badaniu prototypów można przyjąć za uzasadnione stanowisko, że pewną część pomiarów wykonuje się niejako na wyrost;

d. Metodyka stosowana przy badaniach serii informacyjnej winna być tak pomyślana, żeby po niewielkich modyfikacjach mogła być użyta do sprawdzania jakości serii produkcyjnej, przez wprowadzanie tej metodyki do warunków technicznych na urządzenia.

Wymagania zawarte w punktach a - d można streścić następująco:

Przy opracowaniu metodyki należy uwzględnić wszystkie etapy opracowania urządzenia od badań modelu do kontroli ostatecznej produkowanego seryjnie urządzenia. Na tych wszystkich etapach metodyka winna być w swej istocie jednolita. Na każdym stanowisku kontrolnym, gdzie stosuje się opisaną metodykę, w maksymalnym stopniu powinny być uwzględniane potrzeby użytkownika oraz wygoda operatora prowadzącego bezpośrednie badania i pomiary.

To ostatnie stwierdzenie pociąga za sobą konieczność sformułowania dodatkowych, pomocniczych wymagań odnośnie metodyki:

e. Przy tworzeniu metodyki należy zwrócić uwagę na szczególnie czytelne opracowanie badań, protokołów do dokumentacji wyników, a przy wersjach do badań serii informacyjnych i produkcyjnych - przygotowanie schematów decyzyjnych. Wynika to stąd, że przy badaniu serii bierze udział personel techniczny o niższych kwalifikacjach.

f. W miarę możliwości winno być zapewnione automatyczne sporządzanie dokumentacji przez stosowanie wszelkiego rodzaju automatycznych rejestratorów. Automatyczne sporządzanie dokumentacji wyników badań ma szczególne /chociaż przeważnie niedoceniane/ znaczenie na etapie badań prototypu, względnie modelu laboratoryjnego. Automatyczne dokumentowanie wyników badań ułatwia przeprowadzającemu badania inżynierowi koncentrację na właściwym problemie, tj. na kierowaniu pracami, wnioskowaniu i podejmowaniu decyzji pośrednich oraz końco-

wych. Ponadto zautomatyzowanie rejestracji danych pozwala w tym samym czasie zanotować znacznie więcej informacji, co wcale nie jest łatwe przy stosowaniu metody manualnej z użyciem kilku protokolantów. Wreszcie, automatyczna rejestracja danych pozwala /przy założeniu wyposażenia w odpowiednią aparaturę/ na prowadzenie rejestracji pewnych parametrów "na wyrost". Jak to jest ważne, można się przekonać, gdy po zakończeniu badań, w czasie analizy wyników okazuje się, że gdyby notowane były pewne parametry /które przez oszczędność czasu i środków zostały pominięte/, badania miałyby znacznie większą wartość.

g. Urządzenia wchodzące w skład poszczególnych stanowisk kontrolnych winny być ergonomicznie poprawne, tzn. dostosowane do psychofizycznych właściwości operatora.

Wymienione dotychczas cechy charakteryzują opisywaną metodykę w sposób ogólny. Postaramy się obecnie wymienić choćby zasadnicze jej elementy składowe. Dokonamy w tym celu przeglądu znanej nam literatury, starając się podkreślić sugestie dotyczące elementów procedury.

Metodyka badań urządzeń pamięciowych, podawana w ulotkach firmowych wytwórców zachodnich, opiera się o stosowanie generatora programowego i polega na przeprowadzaniu testów: zapis i odczyt "jedynek", "zer", "testu najtrudniejszego" oraz "dopełnienie testu najtrudniejszego" w zmiennych warunkach zewnętrznych. Najczęściej poddaje się zmianom amplitudę prądów przełączających oraz amplitudę prądu zakazu. Wielkość odchyleń jest rzędu 5%. Spotyka się wzmianki /P h i l i p s, pamięć typu 4B5, dokument 32-238-BE 7/67/ o przeprowadzaniu każdego z wyżej wymienionych testów w warunkach jednoczesnej zmiany wszystkich napięć zasilających o 5%, przy wyczerpaniu wszystkich kombinacji odchyleń. Ponadto, w tym samym dokumencie mówi się o przeprowadzaniu tych badań przy trzech temperaturach otoczenia: $0^{\circ}C$, $25^{\circ}C$ oraz $65^{\circ}C$. W innym dokumencie /RCA M e m o r y A c c e p t a n c e T e s t/ podaje się minimalny czas przeprowadzania badań, wynoszący dla każdego wariantu 8 godzin bezbłędnej pracy pamięci.

Należy podkreślić, że tego typu dane spotyka się w prospektach firm znanych z wysokiej niezawodności wyrobów. Znacznie częściej sprawa ta jest kwitowana ogólnikowym stwierdzeniem, że w trakcie projektowania i produkcji podejmowane są wszelkie środki, mające zapewnić wysoką jakość produktu. Ponadto przytoczone powyżej dane P h i l i p s a podane są

tylko przy jednym typie pamięci, przy którym, jak się wydaje, osiągnięto wskutek długiej praktyki produkcyjnej wysoką niezawodność. Pamięć ta jest stosunkowo wolna - czas cyklu wynosi 5 usek. Przy innych, nowszych produktach firmy dane o metodyce badań są znacznie mniej szczegółowe.

Bardziej złożone badania pamięci operacyjnej oparte są o wykonywanie dwuwymiarowych przekrojów obszaru sprawności w czasie wykonywania testu najgorszego przypadku /w o r s t c a s e p a t t e r n/. Jako współrzędne wybiera się z reguły amplitudę prądów X - Y oraz amplitudę prądu zakazu. Parametrem pomiaru jest temperatura. Dane tego typu otrzymaliśmy od firmy Plessey. W terminologii angielskiej badania tego typu określa się jako s h m o o p l o t t i n g. Wyniki tych badań są przeznaczone tylko dla odbiorcy urządzenia. Jedynym znanym dotychczas wyjątkiem jest wspomniany wyżej dokument firmy Philips, zawierający dla pamięci typu FX - 12 wykres obszaru sprawności dla dwóch temperatur. Należy jednak podkreślić, że w opisie wyników brak jest szeregu istotnych danych.

Badania typu s h m o o p l o t t i n g są bardzo pracochłonne i ich znaczenie w metodyce byłoby znacznie większe, pod warunkiem ich automatyzacji. Pokusiła się o to firma IBM, zestawiając do badań nowych systemów pamięciowych układ zapewniający automatyczne tworzenie takich wykresów. Szczegóły techniczne są zawarte w publikacji [1].

Z zamieszczonego skróconego opisu metodyki można wyłowić następujące elementy, zestawiając je w listę:

- Testy,
- parametry zewnętrzne /np. temperatura/, ich zmienność,
- parametry wewnętrzne /napięcia zasilające/, ich zmienność,
- obszary sprawności, warunki ich określania, dobór parametrów zewnętrznych i wewnętrznych dla obszarów płaskich,
- rejestracja obszarów płaskich,
- czas badań.

Po ułożeniu listy elementów procedury, pozostaje zagadnienie zawartości ich treści, np. Testy: należy określić rodzaje, stopień złożoności, kolejność stosowania, czas wykonywania, itp.

Oddzielnego omówienia wymaga sprawa adekwatności warunków stwarzanych przez generator programowy /m e m o r y e x e r c i s e r/ do warun-

ków występujących w czasie nominalnej eksploatacji pamięci w urządzeniu finalnym. W oczywisty sposób warunki występujące przy sprawdzaniu urządzeniem typu m e m o r y e x e r c i s e r różnią się od warunków występujących w maszynie pod następującymi względami:

1. W testerze pamięciowym z reguły miejsca pamięci adresowane są w sposób sekwencyjny. W maszynie sytuacja taka występuje tylko dla pewnych odcinków programu.

2. W testerze pamięciowym nie występuje nigdy wielokrotne wybieranie pojedynczego adresu jako reżim pracy, który może być wykonywany automatycznie. Wybieranie takie może być jedynie realizowane ręcznie, co wydłuża czas badań. Tymczasem wiadomo, że może ono występować w maszynie, powodując z reguły większe obciążenia termiczne elementów pamięci, a szczególnie rdzeni pamięciowych i rdzeni przełącznikowych. Może się więc okazać, że wyniki badań przy pomocy testera są znacznie korzystniejsze niż badania eksploatacyjne pamięci, przeprowadzane przy współpracy z maszyną cyfrową.

3. Przy pracy pamięci w maszynie cyfrowej z reguły nie występują informacje tworzące w o r s t c a s e p a t t e r n s, które stanowią podstawowe testy przy badaniu generatorem programowanym.

Tak więc pamięć badana testerem jest mierzona w trudniejszych warunkach /ze względu na p. 3/, ale równocześnie w łatwiejszych ze względu na p. 2. Odmienny sposób adresowania /p. 1/ wydaje się mieć mniejsze znaczenie, szczególnie w pamięciach typu r a n d o m . a c c e s s, którymi się zajmujemy. Nie można jednak bez ryzyka stwierdzić, że trudniejsze warunki stwarzane przez stosowanie w o r s t c a s e p a t t e r n s zapewniają odpowiedni margines bezpieczeństwa, pozwalający ignorować łagodniejsze warunki, stwarzane przez tester, w którym nie stosuje się wielokrotnego wybierania każdego adresu.

Sprawa ta poruszana jest prawie wyłącznie w literaturze radzieckiej, w której ten reżim pracy określa się mianem męczenia, nękania /"dołblenije"/.

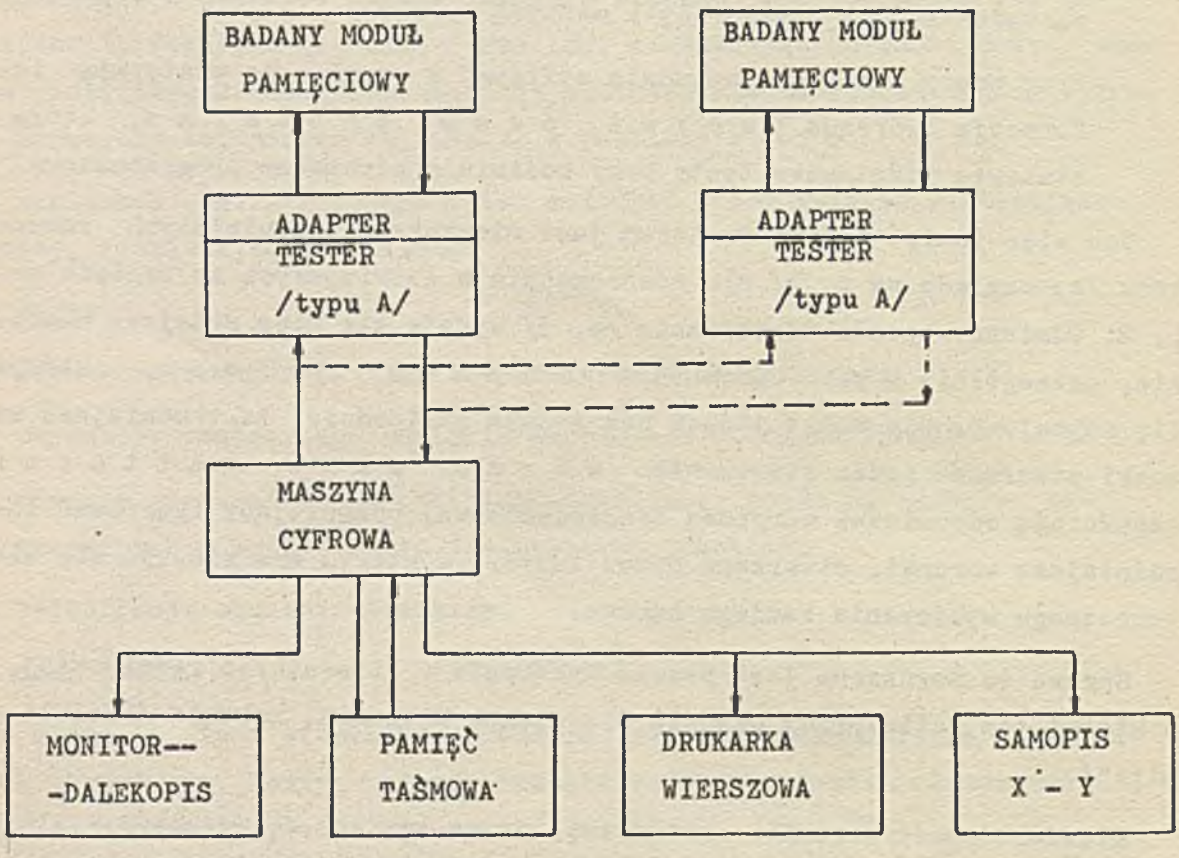
Nieadekwatność warunków sprawdzania przez testery do warunków występujących w maszynie cyfrowej znalazła swoje odbicie w literaturze dopiero w ostatnich latach. Przykładem może być publikacja [2], w której przeprowadzono rozważania o prawdopodobieństwie występowania podczas normalnej

pracy w o r s t c a s e p a t t e r n s oraz pewnych modyfikacji takich wzorów.

W konkluzji autor pracy [2] dochodzi do wniosku, że producenci zapewniają odpowiedni margines bezpieczeństwa, gdyż pamięć jest sprawdzana w znacznie trudniejszych warunkach. Jest to jednak naszym zdaniem stanowisko jednostronne, ze względów omówionych wyżej w p. 2.

Sądzimy, że z powyższych rozważań wynika dość jasno konieczność uzupełnienia wspomnianej listy jeszcze jednym elementem, który będzie uwzględniał problem współpracy badanego modułu pamięciowego z innymi członami systemu /np. z częścią centralną maszyny cyfrowej/. Nazwijmy ten element warunkami współpracy.

Z powyższego wynika już nasza linia postępowania przy rozwiązywaniu problemu. Prosty schemat blokowy /rys. 2/ obrazuje potrzebne do tego celu środki techniczne.



Rys. 2. Schemat blokowy zestawu pomiarowego do badań techniczno-funkcjonalnych pamięci

Do kompletu potrzebny jest jeszcze pakiet następujących programów:

. Plan nadrzędny badań.
. Programy badań szczegółowych.
. Programy sporządzania wyników.
. Strategia wnioskowania.
. Źródłowy język problemowy do pisania programów z p. 2 i 3.

Przejdźmy do opisu schematu blokowego. Istotnym elementem systemu służącego do przeprowadzania badania jest tester typu A. Jest to urządzenie o stosunkowo prostej strukturze, wyposażone w "adapter", dopasowujący poziomy używanych napięć w badanym module i w testerze. Ponadto tester wraz z adapterem służy jako "transformator szybkości". Z tego wynika, że istnieje możliwość stosowania maszyny cyfrowej o szybkości działania wielokrotnie mniejszej od szybkości badanego modułu pamięciowego.

Zastosowana maszyna musi być odpowiednio wyposażona w urządzenia zewnętrzne, potrzebne do realizacji programu badań. Należy tutaj wymienić: monitor-dalekopis, który stanowi środek techniczny komunikacji człowieka z systemem sprawdzającym, drukarkę wierszową, która służy do wydawania większych masywów informacji /pewne wyniki testów/, samopis /Y - Y p l o t t e r/, który służy do automatycznego wykreślania obszarów sprawności, a ściślej biorąc dwuwymiarowych przekrojów tego wielowymiarowego obszaru. Ostatnim wreszcie elementem systemu jest pamięć taśmowa, służąca do przechowywania archiwalnej dokumentacji wyników badań.

Nasz sposób stosowania metody opiera się o maszynę ZAM-41Z.

Pakiet programów złożony będzie z następujących głównych części:

. Test programowy, używany w czasie badania pierwotnego i badania diagnostycznego pamięci. Odpowiada on drugiemu punktowi pakietu wymienionego wyżej.
. Program analizujący, używany w obydwu fazach testowania. Program ten tworzy wykazy tzw. czułych komórek pamięciowych.
. Program wstępnego przetwarzania, który wykorzystuje informację zawartą w wykazach czułych komórek pamięciowych. Wykazy te otrzymywane są w czasie pierwotnego badania pamięci. Program tworzy na ich podstawie listy eksperymentów reprezentatywnych.

. Program rozpoznający, który porównuje listy reprezentatywnych eksperymentów z wynikami aktualnego badania diagnostycznego, określając w ten sposób "najbliższy" obraz przekłamania reprezentatywnego. Program ten podaje również dodatkowe dane o różnicach między przechowywanymi obrazami przekłamań i obrazem przekłamania obserwowanym w czasie badania diagnostycznego.

Wymienione powyżej programy łączą się także z drugim punktem pakietu programów. Punkt trzeci tego pakietu - to programy dla ustalenia formatu wydawniczego. Wreszcie czwartym składnikiem pakietu programów jest program sterujący, który umożliwia operatorowi wykonanie diagnostycznych i pierwotnych badań pamięci zgodnie z jego planem badania.

Komunikacja między operatorem i maszyną posiada pewne cechy ograniczonego konwersacyjnego sposobu współpracy. Wykonuje się ją przy pomocy monitora-dalekopisu, podłączonego do maszyny o n - l i n e.

Wymienione ostatnio funkcje posiadają cechy pozwalające je zaliczyć do punktu pierwszego pakietu programów, tzn. planu nadrzędnego badań oraz do punktu czwartego pakietu, tzn. strategii wnioskowania.

Warunkiem efektywnej realizacji tak pomyślanego programu badań techniczno-funkcjonalnych jest m.in. opracowanie właściwego źródłowego języka zewnętrznego, posiadającego pewne cechy języków konwersacyjnych.

Zagadnienie to jest w tej chwili najmniej zaawansowane. Przewiduje się w tej sprawie współdziałanie Zakładu Techniki Cyfrowej z Pracownią Systemów Operacyjnych Maszyny. Szersze omówienie zagadnień programowych, występujących przy określaniu obszarów sprawności według naszej metodyki, można znaleźć w artykule [4].

Przed przystąpieniem do opracowania zarysów nowej metodyki badania, przedstawionej w niniejszym artykule, wykonano pewne prace rozpoznawcze. Prace te polegały na próbie sformułowania wytycznych do opracowania metodyki badań urządzeń kompletnych. Aby prace te były konkretne, oparliśmy się przede wszystkim na opracowaniu metodyki badania pamięci PAO-5. Opracowana metodyka posłużyła do ustalenia warunków technicznych, jakie ma spełniać pamięć operacyjna PAO-5.

Szczegółowy opis metodyki znajduje się w publikacji [3].

4. Stan opracowań przyrządów serii MOPS

Urządzenia do sprawdzania pamięci operacyjnych serii MOPS opracowywane były w Zakładzie Techniki Cyfrowej i ze względu na potrzeby i posiadane środki wykonywane w postaci użytkowych modeli laboratoryjnych.

Dla zorientowania Czytelnika w zakresie rezultatów osiągniętych w omawianej dziedzinie wyliczymy i krótko scharakteryzujemy odpowiednie opracowania Zakładu Techniki Cyfrowej IMM:

MOPS-1 był pierwowzorem serii przyrządów i posiadał bardzo skromne właściwości funkcjonalne. Przeznaczony był do sprawdzania modelu laboratoryjnego pamięci PAO-3.

MOPS-2 był modelem użytkowym, przeznaczonym do uruchamiania i badania pamięci PAO-3. Ponieważ pamięć PAO-3 wyposażona była we wbudowane układy kontroli poprawności działania, MOPS-2 nie był wyposażony w układy porównania. Do sygnalizacji błędu wykorzystywany był impuls pochodzący z zespołów kontroli pamięci, który zatrzymywał pracę adresatora przyrządu MOPS-2.

MOPS-3 był modelem użytkowym, przeznaczonym do uruchamiania i badania pamięci PAO-5. Przyrząd ten był zrealizowany na układach techniki germanowej S-400. Posiadał on szeregowy układ porównania i decyzji. Był dość bogato wyposażony w różne reżimy pracy, nie wykorzystywane nawet w pełni w trakcie eksploatacji. Przyrząd ten w zasadzie był przeznaczony do badania pamięci z wybieraniem liniowym, jednak wyposażony został m.in. w generator testów o szerokich możliwościach, pozwalający na użytkowanie go również przy sprawdzaniu pamięci koincydencyjnych.

Poza modelem użytkowym zostały wyprodukowane dwa dodatkowe egzemplarze, z których jeden był używany przez WZE ELWRO przy produkcji pamięci PAO-5.

MOPS-3A był modelem użytkowym, służącym do uruchamiania i sprawdzania drugiej wersji pamięci PAO-5, przeznaczonej do maszyny ZAM-41 i ZAM-41Z. Funkcjonalnie przyrząd ten był prawie identyczny z przyrządem MOPS-3, usunięto w nim jedynie pewne możliwości badań, które - jak wykazała praktyka - nie były wykorzystywane. Dotyczyło to możliwości regulacji w szerokich granicach parametrów czasowych impulsów sterujących, adresowych i informacyjnych.

MOPS-4 jest modelem użytkowym, przeznaczonym do uruchamiania i sprawdzania koincydencyjnej pamięci PAO-6 o pojemności 16.000 słów i czasie cyklu 2 usek. Zrealizowany został na pakietach techniki krzemowej S-50. Zastosowano w nim, w odróżnieniu od poprzednich modeli, równoległe układy porównania i decyzji, zapewniające przyspieszenie procesu sprawdzania. Ponadto w przyrządzie zwrócono szczególną uwagę na wygodę operatora. Ilość manipulacji, jakie należy wykonać dla ustawienia poszczególnych reżimów, została zminimalizowana. Było to możliwe dzięki temu, że metodyka sprawdzania pamięci koincydencyjnych przy pomocy takich przyrządów jest znacznie lepiej zdefiniowana i opisana niż analogiczna metodyka dla pamięci z wybieraniem liniowym.

Jakkolwiek przyrząd MOPS-4 również posiada postać laboratoryjnego modelu użytkowego, to jednak jego dokumentacja po odpowiednim przetworzeniu może służyć za podstawę do zbudowania prototypu.

Co więcej, obserwując rynek tego typu przyrządów, można bez zbytniego ryzyka stwierdzić, że zarówno krajowe jak i zagraniczne zapotrzebowanie uzasadnia nawet podjęcie produkcji seryjnej.

5. Zakończenie

Autorzy mają nadzieję, że linia postępowania, nakreślona w niniejszej pracy, jest dość jasna. Komplikowanie się struktury logicznej pamięci, zwiększające się wymagania odnośnie stosowania testów, zaostrzające się wymagania w stosunku do jakości, wyrażające się w zwiększonej liczbie sprawdzanych parametrów konstruowanych, wytwarzanych i eksploatowanych urządzeń - z jednej strony, a istniejące od lat trudności w uzyskaniu właściwych elementów konstrukcyjnych - z drugiej strony, doprowadziły nas do wniosku, że realizacja wszystkich funkcji urządzenia sprawdzającego na drodze czysto układowej przestaje się opłacać. W wyniku tego poglądu zamierzamy przejść na realizację mieszaną - h a r d w a r e - s o f t - w a r e.

Najbliższe lata pokażą czy koncepcja ta, w porównaniu do koncepcji czysto układowej, jest bardziej ekonomicznym rozwiązaniem problemu w naszych warunkach.

Literatura

[1] J.E. GERSHBACH: The Great Shmoo Plot: Testing Memories Automatically. Electronics, July 25, 1966, str. 127-134.

[2] B.R. HEARN: Statistical Probability of Worst-case Patterns in Coincident-current Ferrite-core Stores. Proc. IEE. vol. 114, Nr 2, February 1967, str. 187-194.

[3] A. SIKORSKI, J. DAŃDA: Badanie i kontrola ferrytowej pamięci operacyjnej. Biuletyn informacyjny Nowości Techniczne, rok IV, Nr 3/15/, lipiec 1967, str. 31-50.

[4] J. DAŃDA: An Application of a Digital Computer for Extended Shmoo-Plotting of Ferrite-core Memories /Preliminary draft/. Instytut Maszyn Matematycznych, październik 1967 r., Warszawa /Maszynopis na prawach rękopisu, dostępny w Bibliotece IMM/.

Mgr inż. MIROSŁAW MIKA
Zakład Elektroniki Cyfrowej

621.38:049.7-181.4:681.322

ZASTOSOWANIE PÓŁPRZEWODNIKOWYCH UKŁADÓW SCALONYCH DO KONSTRUKCJI EMC

Część I

1. Wstęp

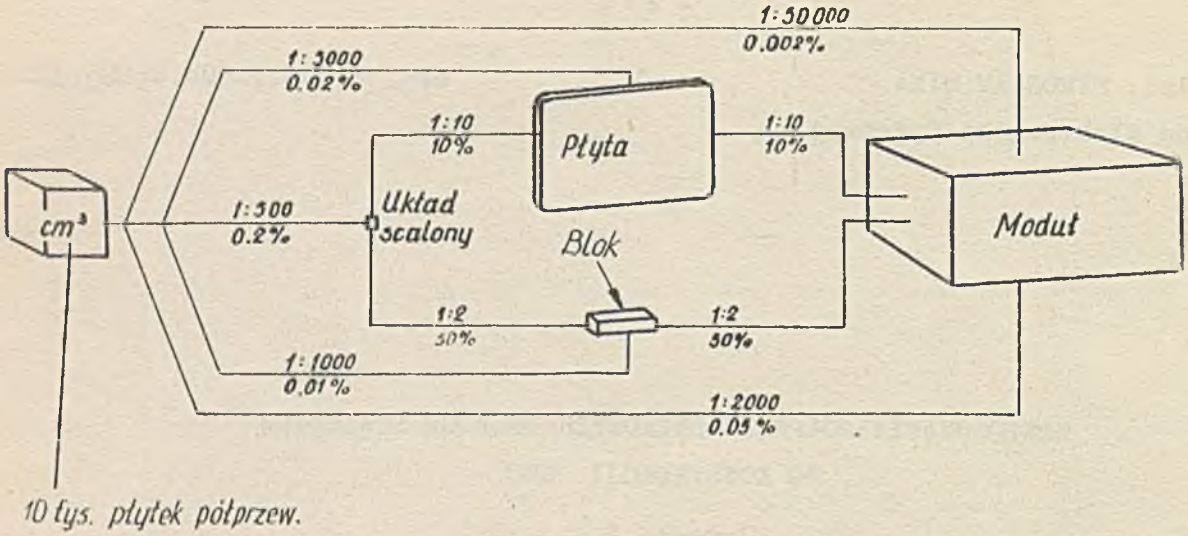
Dynamiczny rozwój mikroelektronicznych układów scalonych, wyrażający się w ciągłym wzroście szybkości działania i niezawodności oraz redukcji objętości, ciężaru i zużycia mocy spowodował, że projektowanie, produkcja i eksploatacja systemów mikroelektronicznych stały się bardzo złożone. Zadaniem specjalistów od upakowania układów mikroelektronicznych /p a c k a g i n g/ jest powiązanie układów scalonych w system, przy maksymalnym wykorzystaniu ich zalet.

Przez upakowanie układów należy rozumieć proces przestrzennego rozmieszczenia, łączenia, chłodzenia i zabezpieczenia układów mikroelektronicznych.

2. Upakowanie systemów mikroelektronicznych

Przy opracowywaniu konstrukcji mikroelektronicznych należy brać pod uwagę następujące aspekty:

O b j ę t o ś ć: Układ mikroelektroniczny, spełniający określoną funkcję, może być ukształtowany w płytce półprzewodnikowej o powierzchni mniejszej od 1 mm^2. Aby z takich układów zrealizować system cyfrowy wystarczy zużyć 10.000 płytek, które umieszczone jedna przy drugiej, zajmują objętość 1 cm^3. Przy obecnych możliwościach, objętość takiej kostki sześciennej po upakowaniu układów /po dodaniu obudowy, połączeń itp./ w najlepszym wypadku powiększa się 2.000 razy /rys. 1/.



Rys. 1. Skuteczność upakowania układów scalonych w systemach mikroelektronicznych

N i e z a w o d n o ś ć: Wzrost gęstości upakowania elementów i połączeń wpływa na obniżenie niezawodności systemu. Mikroskopijne połączenia i przewody są trudniejsze do wykonania, a nowe techniki połączeń nie są jeszcze zbadane. Czas życia układu maleje od 20 do 50% na każde $10^{o}C$ przyrostu temperatury począwszy od $25^{o}C$. Wymaga to szczególnie ostrożnego potraktowania problemu odprowadzenia ciepła.

S z y b k o ś ć d z i a ł a n i a: Układy scalone są coraz szybsze, opracowano układy z opóźnieniem mniejszym od 1 ns. Istnieje możliwość wykonania systemów cyfrowych z cyklem zegarowym o częstotliwości ok. 100 MHz. Dlatego bardzo istotne są ograniczenia spowodowane długością przewodów łączących. Przy upakowywaniu układów powstała konieczność szczegółowej analizy przeników w okablowaniu.

T e c h n i k a p o ł ą c z e ń: Zastosowanie dużych wiązek przewodów i standardowych złącz w systemach mikroelektronicznych jest znikome. Montaż elementów i układów wykonuje się pod mikroskopem. Dokonanie wyboru techniki montażu powinno nastąpić po porozumieniu się z wytwórcą. Przedmiot łączenia wyprowadzeń jest jednym z najbardziej dyskusyjnych. Duża ilość nowych, obiecujących, ale jeszcze ostatecznie nie przebadanych technik wykonywania połączeń utrudnia podjęcie decyzji. Niemniej jednak intensywność uszkodzeń wynosi około 0,00001% na 1,000 godzin dla połączeń w porównaniu z wartością 0,001% dla monolitycznych układów scalonych.

E k s p l o a t a c j a: W systemach mikroelektronicznych bardzo istotną jest sprawa lokalizacji uszkodzeń. Wyposażenie do zlokalizowania i usunięcia uszkodzenia może okazać się droższe od wyposażenia produkcyjnego. Mimo dobrze zaprojektowanego urządzenia, umożliwiającego wymianę elementów, raczej opłaca się wyrzucać całe moduły niż próbować je naprawiać.

C z ę ś c i z a p a s o w e: Jak wynika z przeprowadzonych rozważań, ze względu na niezawodność zaleca się konstruowanie dużych modułów podstawowych, ale wówczas ich powtarzalność jest bardzo mała. Z tego powodu konstruktor ma do rozstrzygnięcia problem, czy projektować duże moduły dla zwiększenia niezawodności, czy też pozostawić je małymi dla ułatwienia zaopatrzenia w części zapasowe.

K o s z t y: Ceny układów scalonych szybko maleją. Układ w postaci scalonej jest znacznie tańszy od jego odpowiednika, zbudowanego z elementów dyskretnych. Natomiast koszty projektu i produkcji systemu wzrastają. Decyduje o tym dążność do zwiększenia gęstości upakowania układów, zmniejszenia długości przewodów łączących oraz zwiększenia niezawodności przez ograniczenie ilości kontaktów ciernych oraz połączeń lutowanych lub zgrzewanych [1].

Wykorzystanie w konstrukcjach elektronicznych układów scalonych powoduje przesunięcie montażu od elementów układu do układów funkcjonalnych. Jednak nie zmieniło to podstawowych decyzji odnośnie podziału systemu na mniejsze zespoły.

Stosowane są trzy poziomy upakowania:

. Poziom I. Pakiet - kilka połączonych układów realizujących funkcję logiczną.

. Poziom II. Blok - zespół wzajemnie połączonych pakietów poziomu I.

. Poziom III. System - szereg wzajemnie połączonych bloków.

Gęstość upakowania na poszczególnych poziomach określa się przez liczbę układów scalonych na jednostkę powierzchni lub objętości.

Trzypoziomowy podział systemu pociąga za sobą konieczność podjęcia decyzji odnośnie typu obudowy układu scalonego i metody łączenia wyprowadzeń na poziomie I, optymalnej wielkości pakietów i ilości kontaktów złącza, sposobu odprowadzania ciepła oraz metody realizacji połączeń między pakietami.

Najtrudniej jest ustalić granice podziału systemu na poszczególne poziomy. Poziom I, na przykład, może zawierać od kilku do kilkuset układów. Wyniki analizy zależą od tego, które z parametrów zostaną uznane za najbardziej istotne. Pracę tę powinien wykonać mieszany zespół specjalistów od systemów, upakowania, logiki i układów, uwzględniając wszystkie ograniczenia i zasady upakowania układów. Należy uwzględnić opóźnienie wnoszone przez układy i przewody łączące, poziom zakłóceń i przeników, łatwość zminimalizowania sprzężeń źródeł zasilających, amplitudy przebiegów i zniekształcenia fazowe, średnią długość przewodów łączących, zakres szybkości układów przy założonych obciążeniach, koszty projektowania i badań, koszty masowej produkcji, elastyczność projektu, podatność eksploatacyjną, średni czas naprawy, koszty części zamiennych, objętość, ciężar, odprowadzenie ciepła i niezawodność systemu, ogółem ok. 15 czynników. Coraz częściej stosuje się maszyny cyfrowe do rozwiązywania tych problemów [3].

Ustalenie rozmiarów pakietu jest najbardziej krytyczną decyzją w projektowaniu systemu. Najogólniej biorąc jest to rodzaj poszukiwania kompromisu pomiędzy wymaganiami związanymi z niezawodnością i eksploatacją. Zestawienie w tabeli 1 wskazuje jak bardzo zmieniają się wymiary pakietów w systemach mikroelektronicznych.

Pakiety, w których wszystkie wyjścia układów są wyprowadzone do kontaktów złącza są szeroko stosowane w konwencjonalnych maszynach cyfrowych. Ponieważ płytka montażowa z uwagi na jej wymiary może pomieścić tylko kilka układów zbudowanych z elementów dyskretnych, jej powtarzalność jest bardzo duża. Produkcja i eksploatacja takiego systemu jest znacznie uproszczona. Płytka z wszystkimi wejściami i wyjściami układów, wyprowadzonymi do jednego z jej bloków nie znalazła zastosowania w maszynach cyfrowych, budowanych w technice scalonej. Na niewielkiej powierzchni można pomieścić bardzo dużo układów. Gdyby wszystkie wyjścia wyprowadzić do krawędzi płytki, to ich ilość byłaby bardzo duża, a mikroskopijne wymiary kontaktów skomplikowałyby konstrukcję złącza. Natomiast zastosowanie małej płytki nadmiernie rozbudowuje poziom II montażu.

Celem zwiększenia gęstości upakowania na pakietach wprowadza się funkcjonalny podział systemu. Na jednej płytce umieszcza się grupę układów, realizującą po połączeniu pewną funkcję i stanowiącą niezależny podzespół. Otrzymujemy w wyniku mniej wyjść, mniej złącz i mniej przewodów łączących na poziomie II.

Tabela 1. Mikroelektroniczne systemy cyfrowe

System	Zastosowanie	Poziom I Ilość ukł.scal. /szt./	Poziom I Wymiary modułów /mm/	Poziom I Rodzaj montażu	Poziom I Metoda łączenia wypr.ukł. scal.	Poziom I Połączenia między ukł. scal.	Poziom I Całkowita ilość poł. wej.wyj.	Połączenia między modułami poziomu I i II	Poziom II Ilość modułów poziomu I /szt./	Poziom II wymiary modułów /mm/	Poziom II Połączenia między modułami poziomu I	Metoda odprowadzenia ciepła
Honeywell H-200	Cywilne	50	127x152	-	lutowanie zanurzeniowe lub na fali	wielowarstwowe /4 warstwy/	80	-	-	600x600x150	Połączenia bezpośrednie drutowe plus płyty obwodów drukowanych	-
Computer Control Co	Cywilne	22	74x69x6	1 lub 2-str.	lutowanie oporowe /po wstępnym cynowaniu/	trawione połączenia dwu- lub wielowarstwowe	34	złącze	24 do 11	150x310x130	połączenia bezpośrednie drutowe /owijane/	wymuszony ruch powietrza
I B M System 360	Cywilne	6 do 24	38x41	płaski 1-str.	lutowanie	trawione połączenia dwuwarstwowe	24	złącze	ponad 66	210x320	wielowarstwowe /4 warstwy/ plus połączenia bezpośrednie drutowe	wymuszony ruch powietrza
R C A Spectra - 70	Cywilne	16	64x127	płaski	lutowanie na fali	trawione połączenia dwuwarstwowe	32 do 48 /zmienna/	złącze	ponad 130	430x430x130	wielowarstwowe /6 warstw/ plus połączenie bezpośredn. drutowe	-
Honeywell M-387	Specjalne /lotnicze/	160 /na każdej z dwu połączonych płytek/	152x304x5	płaski 2-str.	zgrzewanie elektrodami równoległ.	dwie płytki wielowarstwowe	420 /podwójna płytka/	elastyczne wyprowadzenie /38-przewodowe segmenty/	6 /2160 ukł. scalonych/	130x190x480	wielowarstwowe	przewodzenie lub wymuszony ruch powietrza
General Electric A-200	Specjalne /statki kosmiczne/	18	13x13x89	objętościowy /sągowy/	lutowanie	dwie płytki dwuwarstwowe	31	złącze	13 /234 ukł. scalonych/	130x180x25	wielowarstwowe	przewodzenie
Sylwania MS P-24	wojskowe naziemne	178	190x130	płaski 2-str.	lutowanie wyprowadzeń położonych płasko	zestaw dwu dwuwarstwowych płytek sygnałów i napięć połączonych z warstwą masy	200	złącze	40 /7.120 ukł. scalonych/	430x200x180	połączenie drutowe bezpośrednie	wymuszony ruch powietrza
Autonetics D 37 C	wojskowe powietrzne /pociski balistyczne/	196	104x140	płaski 2-str.	lutowanie wyprowadzeń położonych płasko	wielowarstwowe /8 warstw/	160	złącze	23 /cały system/	530x250x150	wielowarstwowe	przewodzenie

Jednocześnie odnalezienie uszkodzenia i oddzielenie go jest łatwiejsze i nie wymaga specjalnego, złożonego wyposażenia, ponieważ w każdej jednostce konkretne wejście jest związane z jednoznacznie określonym wyjściem.

Jak już stwierdzono ze wzrostem ilości układów na pakiecie maleje ilość połączeń wejścia-wyjścia przypadających na jeden układ scalony tego pakietu. Zmniejszenie ilości kontaktów stykowych, stosowanych do połączenia pakietów, wpływa na zwiększenie niezawodności systemu. Ale jednocześnie ze wzrostem ilości układów powtarzalność pakietów w systemie maleje - przy ok. 50 układach faktycznie wszystkie pakiety w danym systemie różnią się między sobą [23].

Wyniki analizy wielu systemów, określające zależność pomiędzy wielkością pakietów, ilością połączeń wejścia-wyjścia i powtarzalnością pakietów podano w tabeli 2.

Tabela 2. Wielkość pakietu, liczba połączeń wejście-wyjście i powtarzalność pakietów w systemie [23].

Liczba układów scalonych w pakiecie /szt./	1	5	10	20	40	50	100	250
Liczba połączeń wejście-wyjście w odniesieniu do jednego układu scalonego /szt./	10	5*	4*	3,5*	3*	2*	1*	0,85*
Liczba typów pakietów w systemie złożonym z 1600 układów scalonych /szt./	-	10	15	18	25	31	16	8
Powtarzalność pakietów /%/	-	96,8	90,6	77,5	37,5	3,1	-	-

* Bez wejść zasilających.

Układy scalone są rozmieszczone na płaskich płytkach montażowych lub na konstrukcjach nośnych trójwymiarowych.

Ponadto istnieje możliwość tworzenia kombinacji obu sposobów np. przez płaskie wiązanie pakietów wykonanych metodą przestrzenną lub odwrotnie, przez przestrzenne rozmieszczenie na poziomie II płaskich pakietów.

M o n t a ż p ł a s k i występuje w wypadku kiedy układy scalone można montować po jednej lub po obu stronach płaskiej struktury połączeń elektrycznych. Płaskie konstrukcje są doskonale przystosowane do chłodzenia przez przewodzenie, ponieważ układy można montować na metalowych paskach, połączonych do podłoża. Ze względu na łatwy dostęp do obu stron płaszczyzny, nie ma kłopotów z wymianą elementów. Niemniej jednak duża płaska konstrukcja nośna jest wrażliwa na wibracje.

Najdoskonalszą konstrukcją płaską jest wielowarstwowa płytka drukowana /WPD/, która może pomieścić dużą ilość układów likwidując trudności w realizacji połączeń między nimi. Wielowarstwowa płytka drukowana jest rozwiniętą formą obwodu drukowanego, w której poszczególne, odizolowane warstwy ścieżek drukowanych, zostały połączone między sobą przez wykonanie metalizowanych otworów w miejscach połączeń międzywarstwowych.

Montaż płaski znajduje najszersze zastosowanie w urządzeniach profesjonalnych.

M o n t a ż p r z e s t r z e n n y zapewnia największą gęstość upakowania układów /do 60% całej objętości zajmują układy i ich wyprowadzenia/.

U p a k o w a n i e o b j ę t o ś c i o w e umożliwia wykonanie zwartego, sztywnego systemu o minimalnym ciężarze i objętości.

Przykładem może być zestaw układów umieszczonych jeden na drugim. Połączenia wykonuje się przy pomocy taśmy niklowej spawanej do wyprowadzeń usytuowanych równolegle do podstawy płaskiego układu scalonego. Zbędne połączenia usuwa się przez wycięcie odpowiednich odcinków taśmy.

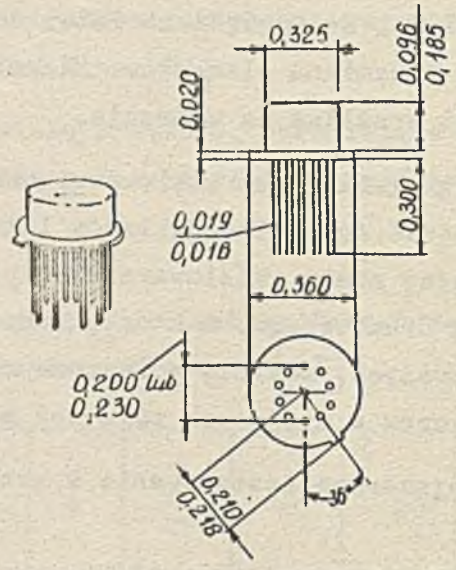
Możliwość operowania w przestrzeni trójwymiarowej wpłynęła na dużą różnorodność konstrukcji objętościowych.

3. Obudowy układów scalonych

Kształt obudowy układu scalonego jest tak bardzo powiązany z ogólną konstrukcją systemu, że już w trakcie dokonywania jej wyboru należy znać postać i właściwość całego systemu.

3.1. O b u d o w a t y p u T O

Obudowa typu TO jest adaptacją tradycyjnej obudowy tranzystorowej do układów mikroelektronicznych poprzez zmniejszenie w niektórych przypadkach wymiarów gabarytowych i powiększenie ilości wyprowadzeń /rys. 2/.



Rys. 2. Obudowa typu TO-5 /wymiary w calach/

Ze względu na opanowaną produkcję obudowa TO należy do najtańszych i najbardziej niezawodnych konstrukcji. Wyróżnia się ona szczególnie dobrą hermetycznością dzięki dużej szczelności połączenia pomiędzy szkłem i Kovarem[x], które nie pogarsza się przy znacznym podwyższeniu temperatury otoczenia.

Najbardziej popularny jest typ TO-5 o wymiarach:

	min.	max.
. średnica obudowy	-	0,325" /8,3 mm/
. wysokość obudowy	0,096" /2,4 mm/	0,185" /4,7 mm/
. długość wyprowadzeń	0,300" /7,6 mm/	1,5" /38 mm/
. średnica rozstawienia wyprowadzeń	0,200" /5,1 mm/	0,230" /5,3 mm/

[x] Stop specjalny o odpowiednio dużej oporności cieplnej, stosowany przy zgrzewaniu oporowym.

. kąt rozstawienia wyprowadz.	36°	-
. ilość wyprowadzeń	3	12

Wyprowadzenia wykonano w postaci giętkich drutów kovarowych, najczęściej złoconych i dostosowanych do znanych metod ręcznego lub masowego lutowania.

Układy scalone w tej obudowie przeznaczone są do montażu na płytkach drukowanych jedno - lub wielowarstwowych przez umieszczenie wyprowadzeń w metalizowanych otworach. Ciasne rozstawienie wyprowadzeń narzuca ścisłe tolerancje otworów i ścieżek drukowanych, co wydatnie zwiększa koszty produkcji przy nienajlepszej gęstości upakowania układów [18].

Niemniej trudne i niemożliwe do zautomatyzowania staje się umieszczenie układów w płytce.

Bardzo blisko położone otwory nie pozwalają na poprowadzenie ścieżek pod podstawą obudowy, a małe odległości między polami lutowniczymi sprzyjają formowaniu się mostków szczególnie przy lutowaniu kąpielowym.

Dość duża powierzchnia obudowy pozwala na stosowanie chłodzenia z wymuszonym obiegiem powietrza lub płynu, ale ze względu na trudności w realizacji pewnego kontaktu z radiatorem uniemożliwia zastosowanie bardziej skutecznych metod chłodzenia przez przewodzenie.

Prawie wszystkie formy oferują ten typ obudowy, choć wydaje się, że po opanowaniu trudniejszych technologii montażu znamiennych dla innych typów, dojdzie do stopniowego wycofania tej konstrukcji.

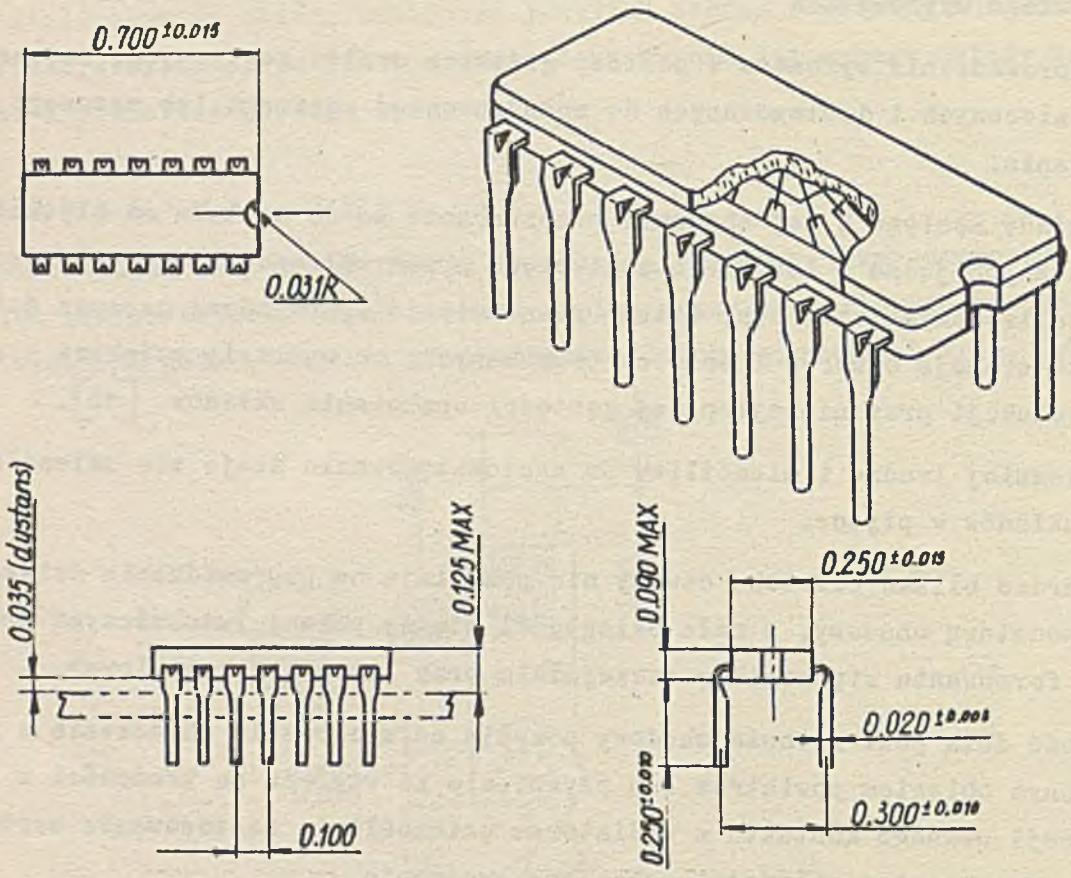
3.2. O b u d o w a t y p u p a j ą k*. Płaska obudowa z pionowymi wyprowadzeniami w dwu rzędach

Obudowa typu pająk jest stosunkowo dużym, płaskim pudełkiem z wyprowadzeniami prostopadłymi do podstawy, rozmieszczonymi wzdłuż dwu równoległych linii /rys. 3/. Najczęściej stosowanym materiałem konstrukcyjnym jest ceramika.

Jak we wszystkich przyrządach półprzewodnikowych, tak i tutaj materiałem stosowanym na wyprowadzenie jest Kovar lub Dumet pokryty spoiwem lutowniczym /cynowany/ lub złotem.

---

* Określenie proponowane przez autora. Odpowiednik nazwy angielskiej DIL /D u a l i n - l i n e/.



Rys. 3. Obudowa typu pająk /wymiary w calach/

Do zasadniczych wymiarów należą:

. wysokość obudowy	0,090" /2,3 mm/,
. wymiary podstawy	0,700" x 0,250" /18x6,4 mm/,
. rozstawienie wyprowadzeń /w linii/	0,1" /2,5 mm/,
. odległość między dwoma rzędami wyprowadzeń	0,3" /7,6 mm/,
. średnica lub długość większego boku prostokątnego przekroju wyprowadzeń	0,020" /0,51 mm/,
. długość wyprowadzeń	0,22" /5,6 mm/,
. ilość wyprowadzeń	14.

Wyprowadzenia są sztywnymi kołkami o przekroju prostokątnym lub okrągłym, które można umieszczać w otworach płytek drukowanych jedno- lub wielowarstwowych. Metalizowanie otworów w obwodach wielowarstwowych jest konieczne przede wszystkim z uwagi na zmniejszenie odległości między powierzchniami części lutowanych, która w prawidłowo uformowanym połączeniu powinna wynosić nie więcej niż 0,25 mm.

Luźne rozstawienie wyprowadzeń nie narzuca zbyt ostrych tolerancji położenia otworów i pól lutowniczych. Poszerzenie końcówek przy podstawie obudowy powoduje, że pozostaje ona w pewnej odległości od powierzchni płytki, dzięki czemu można wykorzystać całą powierzchnię płytki dla ścieżek drukowanych i otworów.

Podobnie jak dla obudowy typu TO najbardziej uzasadnione jest odprowadzanie ciepła przez wymuszony obieg gazów lub płynów. Tak więc ten typ obudowy dostosowano do metod masowego lutowania i automatycznego montażu, unikając wszelkich niedogodności występujących w obudowie typu TO przy porównywalnej gęstości upakowania układów. Należy podkreślić, że obudowa typu pająk występuje w najbardziej popularnych, takich konstrukcjacj modułów poziomu I, oferowanych w postaci gotowych dwustronnych płytek obwodów drukowanych, zawierających od kilku do kilkunastu wlutowanych układów scalonych.

3.3. O b u d o w a t y p u k r a b [*]. Płaska obudowa z wyprowadzeniami osiowymi

Zaletą tej konstrukcji jest zredukowana wielkość obudowy przystosowanej do skutecznego upakowania objętościowego układów /rys. 4/. Spłaszczenie podkreśla jej szczególną przydatność do chłodzenia przez przewodzenie wzdłuż metalowych pasków przewodzących, tym bardziej, że materiałem konstrukcyjnym jest ceramika z tlenków berylu /BeO lub Be_2O_3/ o dużym współczynniku przewodności cieplnej /wartość taka jak dla aluminium/.

Całkowite hermetyczne uszczelnienie obudowy jest na razie trudne do osiągnięcia.

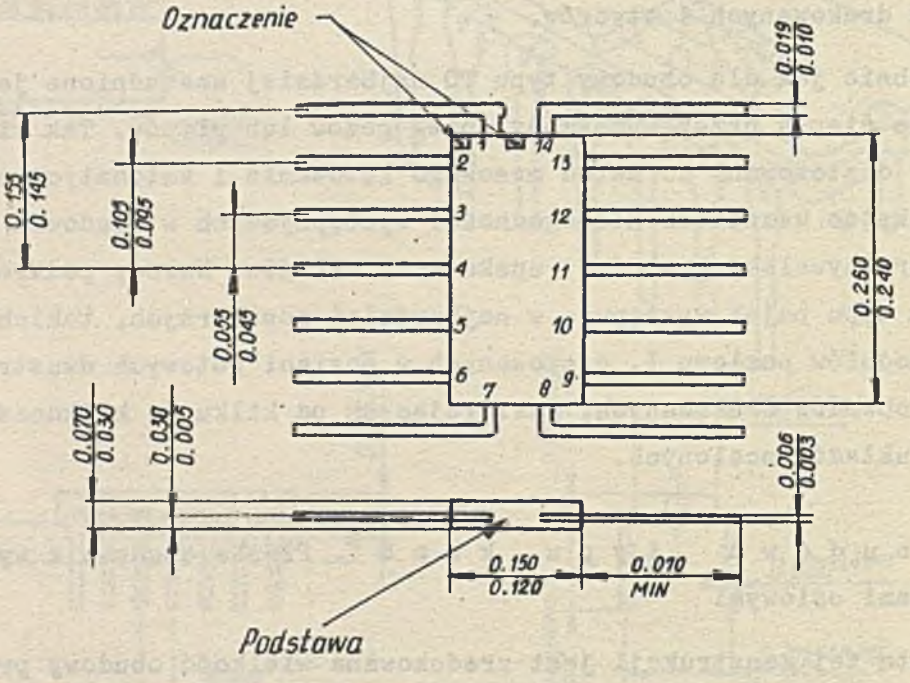
Najczęściej stosowany typ obudowy posiada wymiary:

	min.	max.
. wysokość obudowy	0,030" /0,76 mm/	0,070" /1,8 mm/,
. wymiary podstawy obudowy	0,240 x 0,120" /6,1 x 3 mm/	0,260 x 0,150", /6,6 x 3,8 mm/,
. rozstawienie wyprowadzeń	0,045" /1,15 mm/	0,055" /1,4 mm/,
. długość wyprowadzeń	0,070" /1,8 mm/,	

[*] Określenie proponowane przez autora. W nomenklaturze angielskiej: F l a t p a c k.

. wymiary prostokątnego przekroju wyprowadzeń 0,010 x 0,003" 0,019 x 0,006"
/0,25 x 0,076 mm/ /0,48 x 0,15 mm/,

. ilość wyprowadzeń 14.



Rys. 4. Obudowa typu krab /wymiary w calach/

O s i o w e usytuowanie wyprowadzeń /po dwu stronach, równolegle do podstawy/ wykonanych z płaskiej złoconej taśmy kovarowej umożliwia wykorzystanie bardzo wielu metod ich łączenia. Do najbardziej popularnych należy lutowanie oporowe lub gazowe, zgrzewanie oporowe, ultradźwiękowe, elektronowe i laserowe.

M a ł a o b j ę t o ś ć układu scalonego i bliskie położenie wyprowadzeń praktycznie eliminuje metody ręcznego montażu.

W konstrukcjach modułowych wykorzystuje się dwustronne płytki drukowane /z folią oporową/, wielowarstwowe płytki drukowane, a także przestrzenne struktury połączeń.

Zestawienie w tabeli 3 wskazuje, że układ scalony w płaskiej obudowie z wyprowadzeniami osiowymi daje większe możliwości zastosowań niż układy w obudowie typu TO, chociaż jego stopień hermetyczności jest niższy, a ceny wysokie [12].

Ciągła dążność do zwiększenia szybkości pracy systemów i związanego z nią wzrostu gęstości upakowania układów upoważnia do stwierdzenia, że płaska obudowa z wyprowadzeniami osiowymi jest do takich zamierzeń najbardziej przystosowana.

Tabela 3. Właściwości obudowy typu TO i płaskiej obudowy z wyprowadzeniami osiowymi /typu krab/

Właściwość	TO	Krab
Konstrukcje na obwodach drukowanych?	tak	tak
montaż dwustronny?	nie	tak
konstrukcja bloków modułowych?	nie	tak
łączenie przez lutowanie?	tak	tak
łączenie przez zgrzewanie?	nie	tak
łatwość usunięcia przy naprawie?	nie	tak
wysoka gęstość upakowania?	nie	tak
standardyzacja wymiarów?	tak	nie
całkowicie hermetyczne zamknięcie?	tak	wątpliwe
koszty obudowania?	niskie	wysokie

4. Technologia łączenia układów scalonych

Łączenie wyprowadzeń układów mikroelektronicznych nie jest problemem rozwiązanym, ponieważ ciągle jeszcze pojawiają się nowe metody i materiały, dla których jeszcze nie opracowano podstaw teoretycznych, umożliwiających przeprowadzenie rzetelnej analizy i oceny jakości połączeń. Dla wielu metod nie opracowano jeszcze danych niezawodnościowych, choć wydaje się, że one będą bardzo pozytywne. Wytwórcy, nie czekając na wyniki, śmiało wprowadzają do produkcji nowe metody łączenia.

Wyróżnia się trzy klasy połączeń:

- Połączenia stałe np. zgrzewane, w których jedna lub obie części łączone muszą ulec uszkodzeniu przy ich rozłączeniu.
- Połączenia półstałe np. lutowane, w których można rozłączyć części łączone bez ich uszkodzenia przy użyciu specjalnych narzędzi lub procesów.

. Połączenia cierne lub szybkorozłączalne np. w postaci kontaktów wtykowych.

Do realizacji w/w typów połączeń w mikroelektronice stosowane są różne technologie. Do najważniejszych należy zaliczyć:

. Lutowanie przy zastosowaniu lutownicy ręcznej, urządzenia zanurzeniowego lub z ciekłą falą lutu, oporowego lub indukcyjnego ogrzewania, ultradźwiękowej lub optycznej energii oraz strumienia gorącego gazu,

. zgrzewanie oporowe przy zastosowaniu elektrod usytuowanych przeciwlegle, równolegle lub pod kątem ostrym,

. zgrzewanie przy pomocy strumienia elektronowego lub laserowego,

. łączenie ultradźwiękowe,

. łączenie mechaniczne przy zastosowaniu połączeń owijanych, sprężynujących zacisków, klamerek itp.,

. połączenia cierne.

Prowadzone są prace nad łączeniem przez dyfuzję metali, przy pomocy warstw osadzanych przez metalizowanie, napylanie próżniowe, pyrolizę, rozpylanie katodowe i natryskiwanie metali oraz przy pomocy klejów przewodzących.

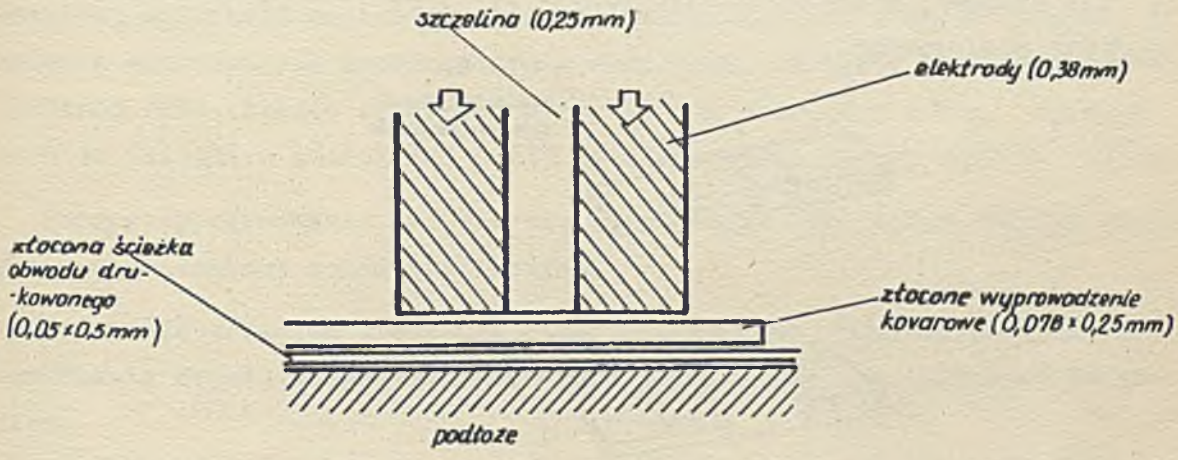
4.1. L u t o w a n i e

Lutowanie ze względu na swoje rozpowszechnienie i dużą tradycję w tym zakresie nie będzie omawiane szczegółowo. Stosowane jest do układów w tradycyjnej obudowie typu TO, jak również do układów w płaskiej obudowie typu pająk.

Stosowane są techniki lutowania zanurzeniowego lub na fali ciekłego lutu. Wynika to ze specjalnie dopasowanej konstrukcji tych układów do metod typowych dla techniki obwodów drukowanych jedno- lub wielowarstwowych [18].

4.2. Zgrzewanie oporowe

Do łączenia wyprowadzeń układów typu krab najczęściej stosuje się zgrzewanie oporowe. W tym wypadku wykonuje się połączenia przy pomocy urządzenia z równolegle usytuowanymi elektrodami ze szczeliną powietrzną lub wypełnioną izolacją, które jako połączony zestaw dociska się do płaskiej powierzchni zgrzewanego wyprowadzenia /rys. 5/.



Rys. 5. Usytuowanie elektrod na powierzchni wyprowadzenia układu scalonego zgrzewanego do ścieżki obwodu drukowanego

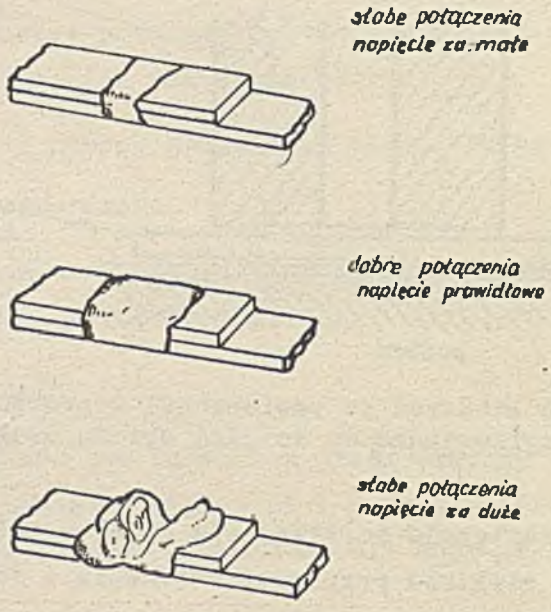
Proces zgrzewania następuje po przejściu impulsu prądu stałego lub zmiennego od jednej z elektrod przez wyprowadzenie i ścieżkę obwodu drukowanego do drugiej elektrody. Impuls prądu stałego jest regulowany zależnie od wymagań przez zmianę czasu i energii rozładowania kondensatora, a prąd zmienny przez zmianę czasu, amplitudy i kształtu fali.

Oporność styku i samych części łączonych na drodze przepływu prądu decyduje o wartości generowanej energii cieplnej, a więc i wzrostu temperatury, która powoduje powstanie połączenia.

Łączeniu podlegają złocone kowarowe wyprowadzenia układów scalonych zgrzewane do folii niklowej, również złoconej, z której wykonano ścieżki obwodu drukowanego. Złoto dzięki niskiemu punktowi topliwości przekształca zgrzewanie w lutowanie twarde, co pozwala wydatnie zmniejszyć temperaturę łączenia /zgrzewanie w temp. około 1500°C, lutowanie twarde w temp. około 950°C/ tak, by nie uszkodzić podłoża.

Oczywiście jest możliwe zgrzewanie metali bez złotego pokrycia.

Ponieważ skuteczność zgrzewania zależy od wzrostu temperatury zależnej od oporności elektrycznej połączenia i części zgrzewanych powstaje konieczność dokładnego kontrolowania wszystkich czynników, wpływających na oporność i gęstość prądu /rys. 6/. Należą do nich: warunki zgrzewania tzn. charakterystyka energii elektrycznej oraz nacisk i powierzchnia styku elektrod, jak również dane charakterystyczne części zgrzewanych tzn. wymiary przewodów i powierzchni kontaktowej, kompozycje metali /lub stopów/, grubość metalizowanego pokrycia oraz wykończenie i gładkość powierzchni.



Rys. 6. Wpływ napięcia impulsu zgrzewającego na jakość połączenia

Niewielkie zmiany czystości materiałów, stosunkowej zawartości dopuszczalnych zanieczyszczeń, grubości niklu, kovaru, złota, warstwy kleju pod folią niklową obwodu drukowanego lub parametrów urządzenia zgrzewającego może uniemożliwić wykonanie prawidłowego połączenia [19].

Jednak większość wytwórców szczególnie systemów wojskowych wybiera właśnie tę metodę łączenia, opłacając wysoką niezawodność urządzenia dobrym przygotowaniem materiałów i systematycznie prowadzoną kontrolą jakości [9,25]. Z pomocą przychodzą nowe opracowania tanich materiałów, do których należy ostatnio wprowadzony rodzaj folii oporowej. Jest to stop niklu i miedzi, który w przeciwieństwie do tradycyjnie stosowanego czystego niklu bardzo łatwo się trawi /podobnie jak miedź/, dobrze się zgrzewa i kosztuje o 90% mniej od niklu.

Zakres spotykanych grubości materiałów stosowanych w mikroelektronice rozciąga się od drutów o średnicy 0,025 mm zgrzewanych do cienkich warstw niklowo-chromowych o grubości 50 Å do dwu zgrzewanych ze sobą drutów lub taśm o grubości 0,635 mm. Dlatego laboratoryjne urządzenia do zgrzewania materiałów w całym zakresie grubości są bardzo drogie.

W produkcji stosuje się bardziej wygodne i tańsze urządzenia specjalizowane. Jedno z nich przystosowane do zgrzewania układów typu Krab ma ustaloną szerokość szczeliny między elektrodami oraz długość impulsu, a zmienia się napięcie kondensatora w zasilaczu, napięcie impulsu i nacisk elektrod /dla układów scalonych firmy T e x a s I n s t r. zgrzewanych do ścieżki o szerokości 0,635 mm odpowiednio 120 V, 0,52 V, 1,2 KG/.

Parametry zgrzewania określa się dla każdego typu połączenia po szczegółowych badaniach metalograficznych.

Większość urządzeń produkują Stany Zjednoczone. Przeciętna wydajność urządzenia działającego automatycznie wynosi kilka tys. połączeń na godzinę.

Połączenia oznaczają się opornością równą oporności metali łączonych /nikiel 5 razy więcej niż miedź/, dużą wytrzymałością mechaniczną /dla połączenia wyprowadzeń układu scalonego ze ścieżką 0,9 kG/, bardzo dużą niezawodnością, niezauważalną ilością ciepła oddziaływującego na elementy i małymi wymiarami zewnętrznymi. Jednak ze względu na to, że należą one do połączeń stapianych, ich własności metalurgiczne nie są najlepsze /kruche połączenia/.

Bardzo przystępną odmianą tej metody jest lutowanie wstępnie pocynowanej płytki obwodu drukowanego i wyprowadzeń układu scalonego. Zastosowanie wstępnego cynowania części łączonych lub też pokrycie ich pastą lutowniczą zmodyfikowało proces łączenia na tyle, że stało się łatwiejsze regulowanie ciepła wydzielanego w połączeniu przez zmianę czasu przepływu i wielkości prądu w adaptowanym do tego celu urządzeniu do zgrzewania oporowego.

Połączenia, jako że lutowane, są kilkakrotnie mniej wytrzymałe mechanicznie, nie znoszą zbyt wysokich temperatur i są nieco większe.

4.3. Zgrzewanie ultradźwiękowe

Bardziej uniwersalną od zgrzewania oporowego jest metoda ultradźwiękowa [13], która nadaje się do różnych metali o dowolnej grubości i niezbyt czystej i gładkiej powierzchni. Niemniej ważna jest neutralność w stosunku do podłoża i części łączonych.

Drgania elektryczne o częstotliwości rzędu kilkudziesięciu KHz przekształca się na wibracje mechaniczne odpowiednio ukształtowanego młoteczka, który niejako wciera wyprowadzenie układu scalonego, aby przez plastyczną deformację otrzymać połączenie drutu ze ścieżką. Mechanizm łączenia nie jest jeszcze kompletnie zrozumiany.

Czynnikami, które przyczyniają się do osiągnięcia dobrego połączenia są: nacisk, amplituda i czas oscylacji. Nacisk musi być skuteczny, aby zapewnić właściwy kontakt pomiędzy powierzchniami łączonymi. Jeśli jest on za duży, wówczas na drut działa zbyt duża siła, co może spowodować jego zerwanie. Odpowiednia amplituda zapewnia skuteczne przebicie utlenionej lub nieznacznie zanieczyszczonej powierzchni, odsłaniając czysty metal. Zbyt duża amplituda będzie przyczyną powierzchniowego utwardzania się metali oraz zniszczenia operującego narzędzia. Zbyt długi czas trwania impulsu będzie przyczyną zniszczenia połączenia, które było już właściwie gotowe.

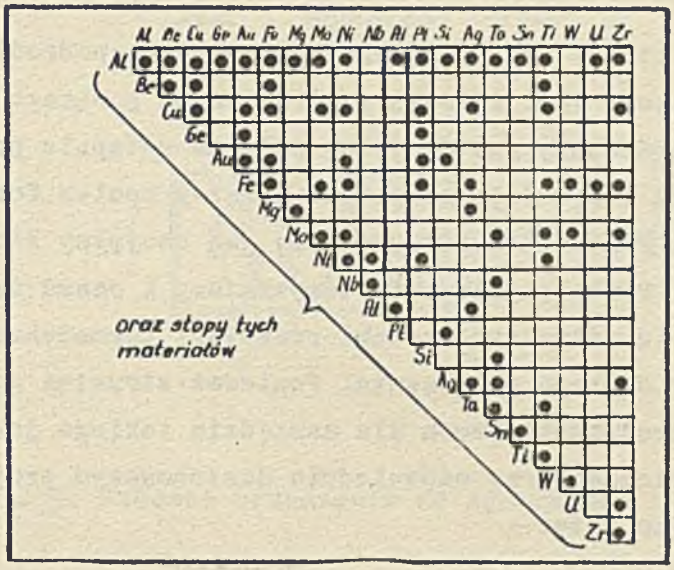
Metoda ta nadaje się do łączenia ze sobą tak cienkich folii jak 0,0043 mm oraz drutów o średnicy mniejszej od 0,076 mm.

Niemniej szeroki jest zakres łączonych materiałów /rys. 7/. Wśród zgrzewanych kombinacji jest złoto, aluminium, miedź, srebro, nikiel i wiele innych materiałów. Można łączyć złoto, nikiel i Kovar do płytek obwodów drukowanych lub innych konstrukcji łączących układy mikroelektroniczne. Jeszcze jedną istotną zaletą jest możliwość łączenia materiałów niskooporowych i wysokooporowych, metali i niemetali o różnej, dowolnie kombinowanej grubości, co było niemożliwe przy zgrzewaniu oporowym.

Łączenie odbywa się z wydzieleniem niezauważalnej ilości ciepła, przy minimalnym nacisku, bez przepływu prądu oraz powstawania łuków, trza trzasków i iskrzenia.

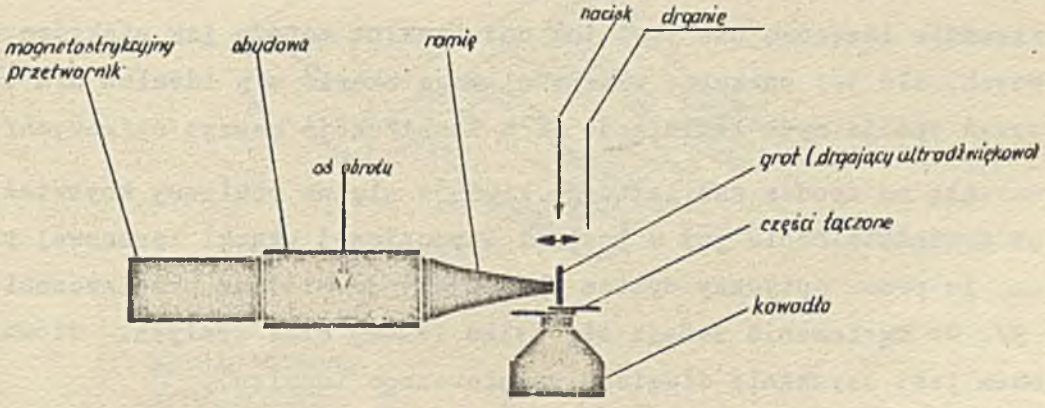
Takie połączenie typu metalurgicznego jest wolne od obcych materiałów i związków międzymetalicznych, a więc odznacza się dużą wytrzymałoś-

cią mechaniczną i temperaturową. Pod względem elektrycznym jest ono połączeniem niskoszumowym i niskooporowym.



Rys. 7. Możliwości łączenia różnych materiałów metodą zgrzewania ultradźwiękowego

Dostępne na rynku europejskim urządzenia pokrywają duży zakres zastosowań. W jednym z nich stosowanym do łączenia przewodów do płytki krzemowej układu scalonego /rys. 8/ nacisk zmienia się od 0 do 170 G, czas od 0 do 1,5 sek., a częstotliwość drgań wynosi 60 KHz. Dokładność kalibracji ze względu na małe wymagania metody, co jest niewątpliwie jeszcze jedną jej zaletą, wynosi około 5%.



Rys. 8. Schemat urządzenia do zgrzewania ultradźwiękowego

4.4. Lutowanie strumieniem gorącego gazu

Jedną z metod, w której wyeliminowano narzędzie bezpośrednio oddziaływujące na części łączone jest lutowanie strumieniem gorącego gazu. Ogrzewany do wysokiej temperatury gaz obojętny pada na wstępnie pocynowaną powierzchnię lutowniczą i powoduje stopienie warstwy spoiwa formującego połączenie lutownicze. Ponieważ wykorzystujemy gaz obojętny stosowanie topnika jest zbyteczne. Przez regulowanie temperatury i czasu trwania fali gazu można z łatwością odtworzyć warunki równowagi termodynamicznej niezbędne dla wykonania dobrego połączenia. Ponieważ strumień gazu z łatwością dosięga miejsc trudnodostępnych dla narzędzia takiego jak np. grot lutowniczy, gęstość połączeń przy odpowiednio dostosowanym projekcie urządzenia może być bardzo duża.

Jedno z urządzeń przystosowane do lutowania układów w płaskiej obudowie z wyprowadzeniami osiowymi oraz w płaskiej obudowie z pionowymi wyprowadzeniami umieszczonymi w metalizowanych otworach posiada 8 głowic ze strumieniem wodoru ogrzanego do temperatury 370°C, które mogą wykonać 20.000 połączeń dziennie [24].

W niedługim czasie ukaże się na rynku europejskim urządzenie produkcji angielskiej [19].

Połączenia takie posiadają wszystkie cechy połączeń lutowanych.

4.5. Zgrzewanie laserowe

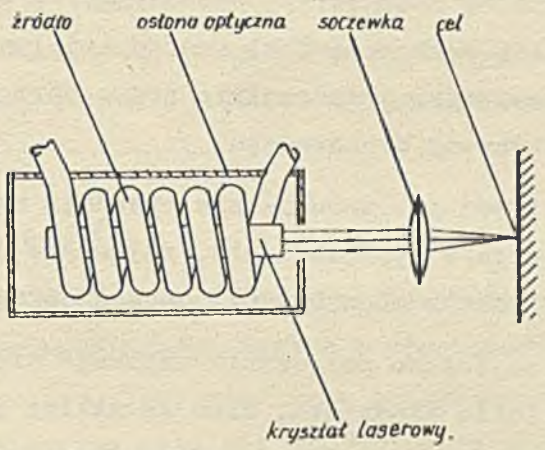
Zgrzewanie laserowe nie jest tak uniwersalną metodą jak niektóre z wymienionych, ale jej unikalne własności mogą okazać się idealne dla wielu zastosowań specjalnych /należą do nich konstrukcje maszyn cyfrowych/.

Promienie ze źródła zasilającego kieruje się na rubinowy kryształ, skąd po przekształceniu już w postaci wzmocnionej wiązki laserowej przesyła się je przez optyczny system skupiający do miejsca przeznaczenia /rys. 9/. Do zgrzewania nadają się tylko lasery ciał stałych. Głównym problemem jest uzyskanie długiego regulowanego impulsu.

W przeciwieństwie do zgrzewania oporowego najlepszą konstrukcją jest połączenie płaskie w postaci dwu taśm umieszczonych jedna na drugiej.

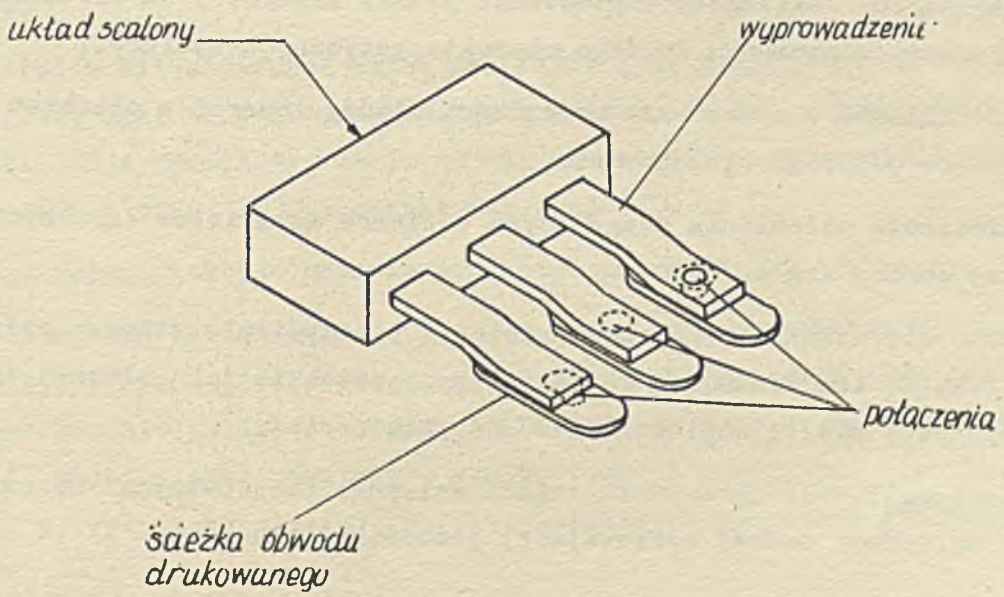
Optymalne wykorzystanie energii wiązki laserowej zależy od fizycznych własności metali tzn. absorpcji, przewodności cieplnej, gęstości, pojem-

ności cieplnej i punktu topliwości oraz od długości impulsu laserowego. Parametry zgrzewania ustala się na drodze doświadczalnej 22 .



Rys. 9. Schemat urządzenia do zgrzewania laserowego

Bardzo dobre połączenia uzyskuje się przy zgrzewaniu układów scalonych do płyt obwodów drukowanych z folią miedzianą, niklową, kovarową itp. Istnieją trzy wersje połączeń skutecznie spawanych przez strumień laserowy: krańcowa, centryczna i z otworem w wyprowadzeniu elementu /rys. 10/.



Rys. 10. Usytuowanie połączeń zgrzewanych strumieniem laserowym

W wersji krańcowej strumień pada jednocześnie za ścieżkę obwodu drukowanego i koniec wyprowadzenia. W wersji centrycznej strumień musi całkowicie przedostać się przez materiał wyprowadzenia. W trzeciej wersji otwór w wyprowadzeniu układu scalonego jest trochę mniejszy od średnicy wiązki laserowej, która przechodzi jednocześnie przez obrzeże otworu i położoną niżej folię ścieżki obwodu drukowanego.

Ze względu na łatwość przeprowadzania operacji zgrzewania najlepsze są połączenia z otworem w wyprowadzeniu, najgorsze połączenia centryczne, co jest związane z ilością absorbowanej energii przez części łączone.

Stwierdzono, że najlepsze połączenia uzyskuje się na laminatach epoksydowo-szklanych z folią miedzianą, mimo że nikiel jest najłatwiejszy do połączenia. Dla zabezpieczenia folii przed utlenieniem w czasie magazynowania należy ją pokryć złotem.

Do zalet tego typu łączenia należy:

- Bardzo małe ilości generowanego ciepła,
- zdolność wykonywania bardzo precyzyjnych spawów o średnicy mniejszej od 0,025 mm,
- możliwość skutecznego zgrzewania różnych materiałów.

Do kłopotów związanych z tego typu łączeniem należą:

- Wymagana 20% dokładność usytuowania plamki laserowej na obiekcie zgrzewanym w stosunku do jego charakterystycznych wymiarów,
- 2% dokładność w ustawieniu odległości między laserem a obiektem w stosunku do drogi ogniskowania,
- konieczność osłaniania reagujących z tlenem materiałów łączonych /przy pomocy argonu/.

Uzyskuje się bardzo dobre, powtarzalne i rzeczywiście stopowe połączenia, tyle że ich własności metalurgiczne, podobnie jak połączeń zgrzewanych oporowo, nie są najlepsze /kruche połączenia/.

Istnieje możliwość zgrzewania części wstępnie pocynowanych. Obecnie nie ma w sprzedaży żadnej zgrzewającej jednostki laserowej.

4.6. Zgrzewanie elektronowe

Ta metoda jest bardzo podobna do poprzedniej z tą różnicą, że operacja musi być przeprowadzona w próżni.

Obie metody nadają się do szybkiego zgrzewania różnych materiałów. Zgrzewanie strumieniem elektronowym jest bardziej odpowiednie dla szerokiego zakresu grubości materiałów, a zgrzewanie strumieniem laserowym jest bardziej ograniczone. Pomimo stadium rozwojowego obie techniki mają bardzo dużą przyszłość.

4.7. Badanie jakości połączeń

Ocenę jakości połączeń każdego typu przeprowadza się przez wykonanie odpowiednich testów. Pierwszym bardzo ważnym i niesłusznie lekceważonym etapem jest kontrola wizualna. Przy właściwym przygotowaniu teoretycznym i praktycznym kontrola wizualna może dać bardzo dobre wyniki. Następnie przeprowadza się pomiary oporności elektrycznej oraz wytrzymałości mechanicznej.

Jednak żadna z tych metod nie daje właściwego obrazu o jakości połączeń. Dopiero kontrola metalograficzna daje odpowiednie wyniki. Ocena metalograficzna jest podstawowym rodzajem kontroli produkcyjnej powszechnie uznawanym przez wytwórców [2].

Po tych badaniach /elektrycznych, wytrzymałościowych i metalurgicznych/ należy poddać połączenia testom klimatycznym i po nich ponowić wymienione badania.

Do testów klimatycznych można zaliczyć: cykl temp. - 55°C + 125°C; długotrwały cykl temperaturowo-wilgotnościowy, 28 dni w temp. 40°C i wilgotności 95% z przepływem prądu 10 mA; długotrwałą wytrzymałość temperaturowa 1000 h przy 125°C z przepływem prądu 100 mA.

Oczywiście, niektóre badania mogą mieć character statystyczny.

Do wykrywalnych wad należą: zbyt gruba lub cienka warstwa kleju pod folią płytki obwodu drukowanego, niewłaściwe pokrycie lub jego grubość na wyprowadzeniu układu scalonego, pęknięcie na granicy fazy międzymetalicznej, obce związki powstałe w wyniku reakcji cieplnych, korozja, szczeliny, wpływy za niskiej lub zbyt wysokiej temperatury łączenia i inne.

Wielu wytwórców systemów posiada własne laboratoria z wyposażeniem takim jak mikroskopy do 500 x, akcesoria fotograficzne, sondy, zasilacze i różnego rodzaju mierniki lub bardziej doskonałe wyposażenie jak przystosowane do badania próbek w wysokiej temperaturze metalografy, mikro-

analizator elektronowy oraz rentgenograf. Oczywiście wiąże się z tym odpowiednio rozszerzony zakres prac i nakładów.

Literatura

[1] D. CHRISTIANSEN: Packing ICs into Equipment Can Be a Critical Task. Electronics, Oct. 17, 1966, pp. 78-84.

[2] K.I. CLEWS, I.G. IOUNG: Metallurgical Evolution of Microcircuit Interconnection Made by the Parallel-gap Process. Microelectronics and Reliability, August 1966.

[3] E.U. COHLER: Computers Design the Layout. Electronics, Nov. 1, 1965, pp. 88-89.

[4] E.C. GARTH, I. COTT: Ultrahigh-Speed ICs Require Shorter, Faster Interconnections. Electronics, July 11, 1966.

[5] A. GODFREY: Welding Integrated Circuits into Electronic Equipments. Microelectronics and Reliability, August 1966.

[6] R.C. GRIMMER: Interconnecting High-Speed Integrated Circuits with Multilayer Boards. Computer Design, Feb. 1967.

[7] S.A. HAYS: The Packaging Revolution, Part IV: Bigger, Better Multilayer Boards. Electronics, Nov. 29, 1965.

[8] I.E. JACKSON: Packaging with Laser Welding. SCP and Solid State Technology, May 1966.

[9] N.A. JAROSCIK, H. SHAPIRO: Gold-plated Nickel Wiring Debugs Parallel-Gap Welding of IC's. Electronics, July 25, 1966.

[10] P.M. KNOWLSON: Fundamentals of Parallel-gap Joning. Microelectron ics and Reliability, August 1966.

[11] R.W. KORB: Multilayer Circuits Boards: Scharpening an Imperfect Art. Electronics, August 8, 1966, pp. 141-147.

[12] A.T. LAWTON: An Equipment Manufacturer's Problems in the Application of Microelectronics. Microelectronics and Reliability, September 1966.

[13] G.E. Littleford: Welding Electronic Devices by Ultrasonics. Industrial Electronics, March 1967.

[14] R.B. Lomerson: Batch Packaging Speeds System Assemblies of IC's. Electronic, April 3, 1967.

[15] H.G. MANFIELD: Microbonding Aspects of the Interconnection Problem. Microelectronics and Reliability, August 1966.

[16] G. MESSNER: Ten Steps to Multilayer Board Design. Electronics Design 3, Feb. 1, 1967.

[17] G.A. NEEDHAM: Advanced Integrated Circuit Packaging. SCP and Solid State Technology, June 1965.

[18] G.A. NEEDHAM: The Transistor Type Integrated Circuit Package. SCP and Solid State Technology, May 1965.

[19] L.S. PHILIPS: Microbonding Techniques. Microelectronics and Reliability, August 1966.

[20] R. RICE: The Packaging Revolution, Part V: Simpler Designs for Complex Systems. Electronics, Feb. 7, 1966.

[21] S. SCHWARTZ: Integrated Circuits in Action, Part IV: Postmortems Prevent Future Failures. Electronics, Jan. 23, 1967.

[22] I.R. SHACKLETON: Laser Microwelds. SCP and Solid State Technology, May 1965.

[23] J.J. STALLER: The Packaging Revolution, Part I: Form and Function Interact. Electronics, Oct. 18, 1965, pp. 72-87.

[24] J.J. STALLER, G. SIDERIS: The Packaging Revolution, Part II: Design and Manufacturing Overlap. Electronics, Nov. 1, 1965.

[25] L.V.M. WHITE: Strips of Nickel Foil Automate Welding of Flatpack Assemblies. Electronics, June 13, 1966, pp. 128-133.

[26] L.M. WHITE, R.W. JONES: Infrared Fault-finder Pinpoints Defects in Multflayer Boards. Electronics, Nov. 29, 1965.

[27] B. ZONBARG: Pick the Proper Prepreg for Multilayer Board Assemblies. Electronic Design 5, March 1, 1967.

Dr inż. ZYGMUNT SAWICKI
Zakład Technologii Specjalnych IMM

621.317:681.327.66-426.2

KRYTERIA OCENY I DYNAMICZNE METODY POMIAROWE PARAMETRÓW DRUTÓW MAGNETYCZNYCH

1. Wstęp

Wybór odpowiednich kryteriów oceny parametrów drutów magnetycznych ma duży wpływ na jakość i koszty wytwarzania pamięci. Opanowanie skutecznych metod pomiarowych może mieć również znaczny wpływ na dalszy rozwój tego typu pamięci.

W pracy niniejszej przedstawiono podstawowe zagadnienia miernictwa drutów magnetycznych w warunkach zbliżonych do rzeczywistych warunków pracy pamięci. Przyjęto założenie, że czytelnik posiada znajomość podstawowych właściwości cienkich warstw magnetycznych oraz zasad działania pamięci z tymi warstwami. W innym przypadku należy przeczytać między innymi poz. [1] i [2].

Odróżniamy dwa główne rodzaje pomiarów drutów magnetycznych. Pierwszy, obejmuje kontrolę dynamicznych parametrów drutu poprzez pomiar sygnału przy odpowiednich warunkach wzbudzania. Ten rodzaj pomiarów jest przedmiotem niniejszego opracowania. Drugi, obejmuje kontrolę podstawowych parametrów magnetycznych określających przydatność drutu magnetycznego do budowy pamięci i oparty jest na metodach quasi-statycznych.

Metody pomiarowe drutów magnetycznych rozwijano w ciągu szeregu lat [4] [5] i [6]. Początkowo różne ośrodki badawcze stosowały różne metody, natomiast obecnie obserwujemy znaczne ujednolicenie tych metod.

Potwierdziły się wcześniejsze przypuszczenia, że w celu osiągnięcia wiarygodnych wyników pomiarów parametrów drutów magnetycznych trzeba przeprowadzać te pomiary w warunkach możliwie zbliżonych do rzeczywistych warunków pracy.

Rzeczywiste warunki pracy drutu magnetycznego w płacie pamięci są określone przez:

- parametry magnetyczne warstwy,
- przestrzenną konfigurację komórki pamięciowej wyznaczoną przez wymiary i wzajemne usytuowanie linii bitowych i linii słów,
- parametry dynamiczne impulsów posyłanych w linie słów i linie sygnałów,
- warunki otoczenia.

Pomiędzy zespołami tych parametrów, aczkolwiek mogą one być wyodrębnione i rozpatrywane oddzielnie, istnieją określone zależności wzajemne, które trzeba brać pod uwagę przy projektowaniu pamięci. Dla różnych typów pamięci rzeczywiste warunki pracy drutów magnetycznych będą różne. Z tego wynika, że aparatura kontrolno-pomiarowa powinna umożliwiać dobranie takich warunków pomiaru, jakie narzuca pamięć o ściśle określonych danych technicznych.

Ogólne charakterystyki dynamicznych zależności parametrów drutów magnetycznych mają jednak podobny kształt, niezależnie od typu pamięci. Dlatego też w dalszych rozważaniach mówić będziemy o kryteriach oceny i metodach kontroli wspólnych dla wszystkich typów pamięci na drutach magnetycznych.

2. Charakterystyki dynamiczne drutów magnetycznych

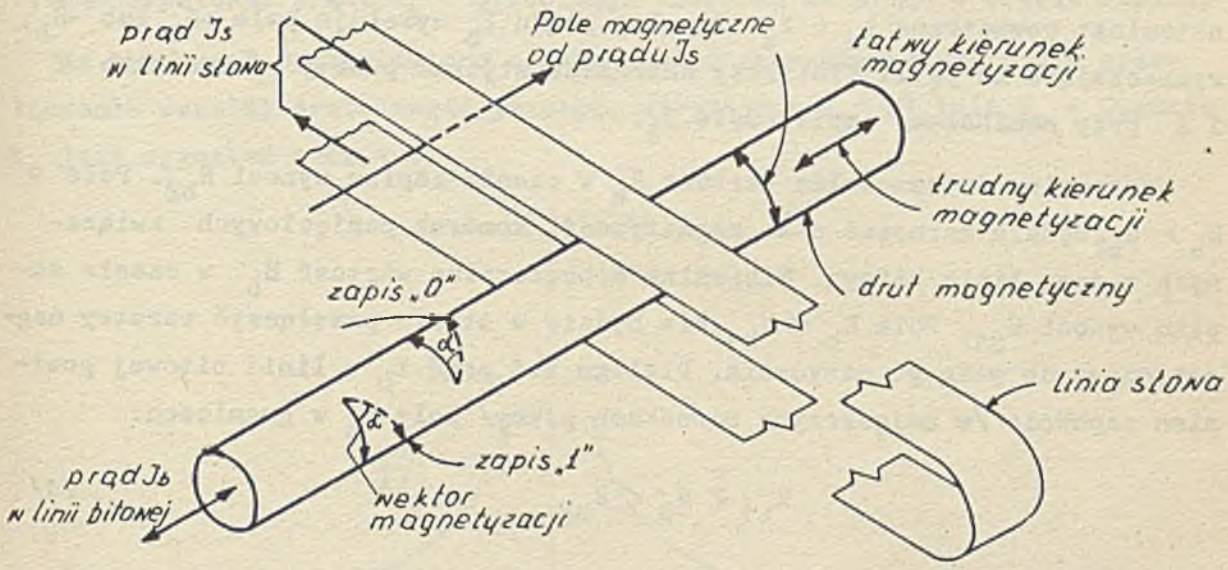
Będziemy rozpatrywać parametry magnetyczne warstwy o anizotropii obwodowej. Warstwa ta jest osadzona na drutach metodą elektrochemiczną. Dla warstw tych $H_k > H_c$, gdzie:

H_k - pole anizotropii,
H_c - pole koercji.

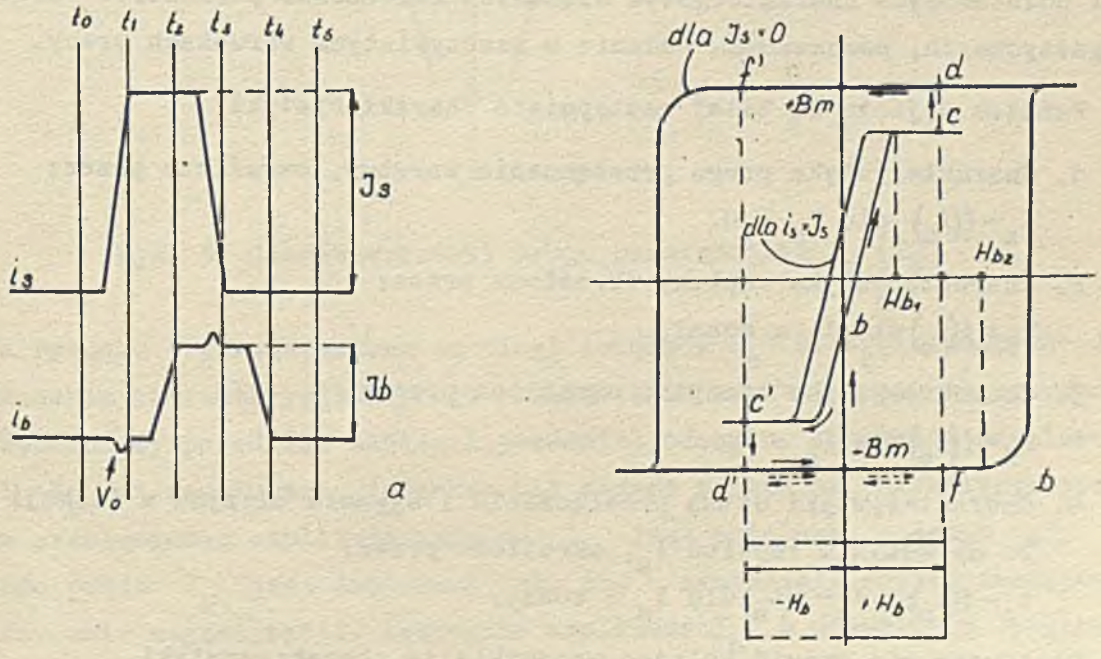
Ideowy układ przestrzeni komórki pamięci przedstawia rys. 1. Rozważania nasze będą dotyczyły właściwości pojedynczych komórek jednobitowych bez uwzględnienia szkodliwego wpływu sąsiednich komórek. Zagadnienia te będą przedmiotem analizy dotyczącej płatu pamięci.

Zależność parametrów magnetycznych warstwy od pól magnetycznych wytwarzanych przez impulsy elektryczne I_s posyłane w linię słowa i impulsy I_b posyłane w linię bitową /linię bitową stanowi drut, na którym osadzono

badaną warstwę/ ilustrują charakterystyki przedstawione na rysunku 2. Impuls oznaczony przez V_o wyraża amplitudę sygnału odczytu w czasie przełączania warstwy magnetycznej, tj. w czasie obrotu wektora magnetyzacji o pewien kąt α , jak na rys. 1.



Rys. 1. Idealny układ przestrzeni bitowej komorki pamięci na drucie magnetycznym



Rys. 2. Typowe charakterystyki pętli histerezy elementu drutu

Na rysunku 2a pokazano typowe przebiegi impulsów I_s posyłanych w linię słowa i impulsów I_b posyłanych w linię bitową, a na rysunku 2b typowe charakterystyki pętli histerezy, określające indukcję w kierunku obwodowym w funkcji natężenia pola wywołanego prądem I_b przy stałej wartości natężenia pola wywołanego prądem i_s. Zewnętrzna charakterystyka odpowiada $I_s = 0$, natomiast wewnętrzna $i_s = I_s$. Impulsy prądu I_b wywołują pola $+H_b$ lub $-H_b$, wyznaczające na pętli histerezy charakterystyczne punkty d i d' lub f i f' przy nominalnej amplitudzie I_b.

Maksymalna dopuszczalna wartość H_b w czasie zapisu wynosi H_{b2}. Pole $H_b > H_{b2}$ będzie naruszać stan magnetyzacji komórek pamięciowych związanych z daną linią bitową. Minimalna dopuszczalna wartość H_b w czasie zapisu wynosi H_{b1}. Pole $H_b < H_{b1}$ nie będzie w stanie przełączyć warstwy magnetycznej do pełnego nasycenia. Dlatego też prąd I_b w linii bitowej powinien zapewnić /w najgorszych warunkach pracy/ pole H_b w granicach:

$$H_{b1} < H_b < H_{b2} \qquad /1/$$

Granice te wyrażają w pewnej mierze zakres poprawnej pracy przy ustalonej amplitudzie prądu I_s. Właściwe kryteria oceny drutów magnetycznych mogą być wyznaczone na drodze analityczno-doświadczalnej w oparciu o zespół podstawowych charakterystyk wiążących zależności parametrów drutów magnetycznych, poddawanych badaniu w rzeczywistych warunkach pracy.

Punktem wyjścia są tutaj następujące charakterystyki:

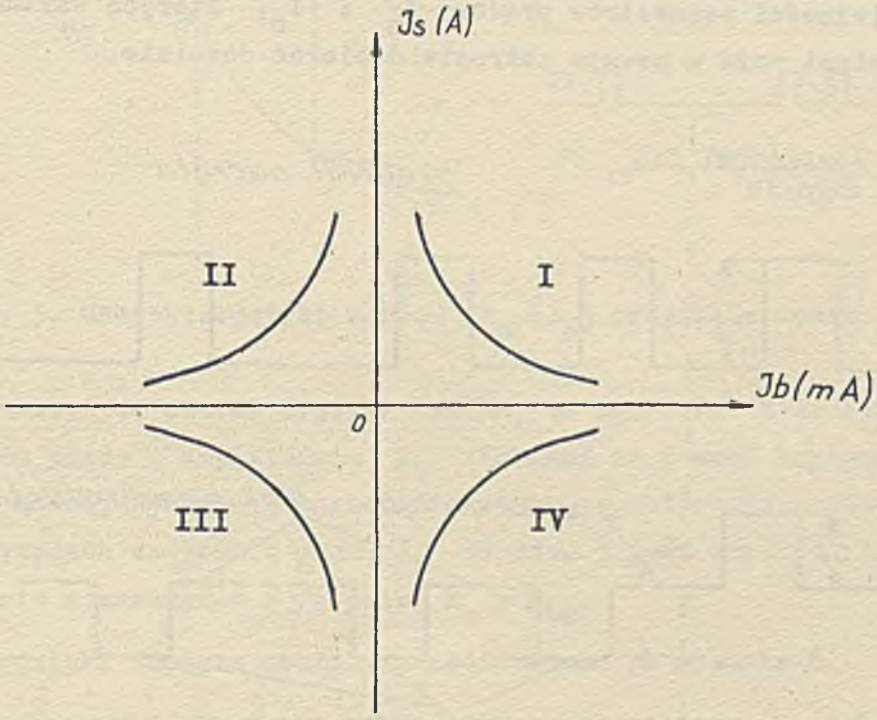
1. Charakterystyka progu przełączania warstwy, określona przez:
   $I_s = f(I_b)$ dla $V_o = 0$.
2. Charakterystyka zapisu, określona przez:
   $V_o = f(I_b)$ dla $I_s = \text{const}$.
3. Charakterystyka odczytu, określona przez:
   $V_o = f(I_s)$ dla $I_b = 0$.
4. Charakterystyki czasu przełączania i sygnału odczytu w funkcji czasu narastania impulsu I_s, określone przez:
   $t_p = f(t_n)$; $V_o = f(t_n)$ dla $I_s = \text{const}$.

Postaramy się omówić kolejno wszystkie te charakterystyki.

Celem określenia wzajemnej zależności prądów I_s i I_b wprowadza się tzw. charakterystykę progu przełączania warstwy z danego stanu na stan

przeciwny. Punkt krytyczny progu przełączania występuje w chwili, gdy sygnał odczytywany po uprzednim zapisie osiąga wartość równą zeru /$V_o = 0$/. Odczyt impulsem prądu I_s poprzedzony jest aktywnym naruszaniem danego stanu magnetyzacji badanej komórki pamięciowej. Naruszanie to jest wywoływane impulsami prądu I_b o przeciwnym kierunku do prądu w czasie zapisu.

Na rysunku 3 przedstawiono ogólną postać charakterystyki progu przełączania warstwy drutu magnetycznego, wyrażającą $I_s = f(I_b)$ dla $V_o = 0$, gdzie V_o jest sygnałem odczytu.

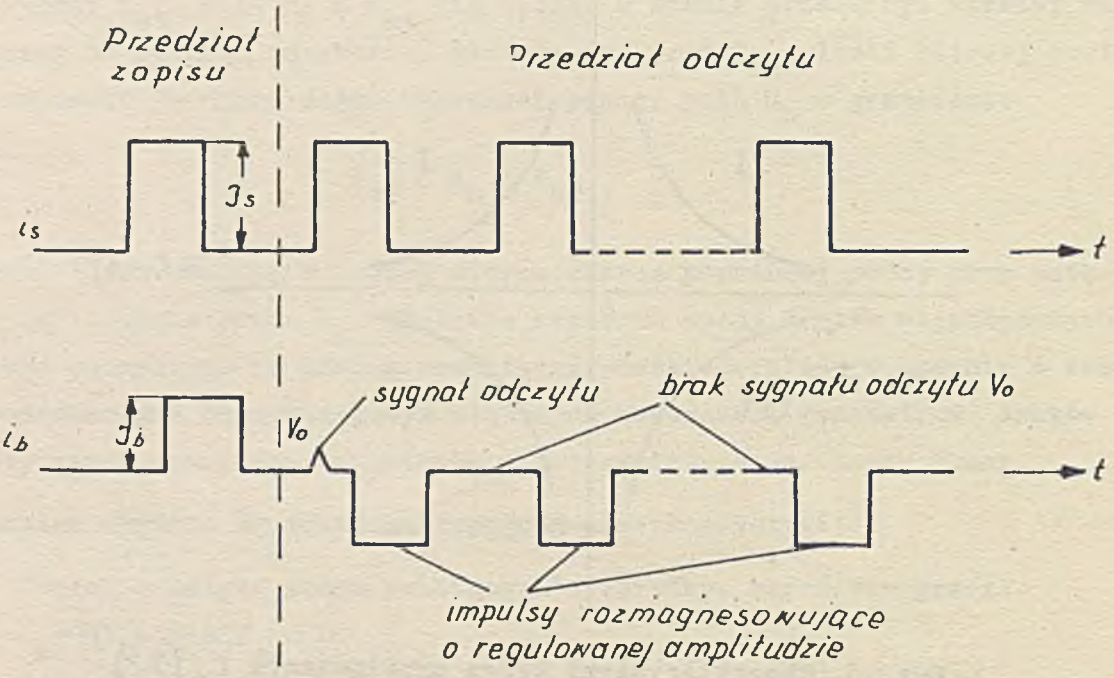


Rys. 3. Charakterystyki progu przełączania I_s (I_b)

Na rysunku 4 przedstawione są ciągi impulsów I_s i I_b, służącycn do zdejmowania charakterystyki progu przełączania warstwy drutu magnetycznego. Wyróżniamy przedział zapisu i przedział odczytu. Amplitudy impulsów I_s i I_b są regulowane, jednakże dla danego punktu na charakterystyce progu przełączania amplituda impulsu I_s jest ustalona, podczas gdy wartość prądu I_b jest dobierana tak, aby w przedziale zapisu następowało nasycenie magnetyzacji. Następnie amplituda I_b o przeciwnej polaryzacji jest dobierana tak, aby sygnał odczytu V_o osiągał wartość równą zero.

Przedstawiony na rys. 4 układ impulsów pozwala określić charakterystyk w drugiej ćwiartce. Dla pierwszej ćwiartki trzeba zmienić polaryzację wszystkich impulsów ciągu I_b. Podobnie trzeba dopasować kierunki prądó impulsów I_s i I_b dla ćwiartek trzeciej i czwartej.

Charakterystyka ta jest bardzo przydatna w praktyce i - co jest szczególnie ważne - może być uzyskana przy pomocy klasycznych urządzeń pomiarowych, tzn. programowanego generatora impulsów i oscyloskopu wyposażonego we wzmacniacz o odpowiednim wzmocnieniu i paśmie częstotliwości. Użyteczność tej charakterystyki wynika z uzyskania bezpośrednich wzajemnych zależności parametrów prądów I_s i I_b, których wartości projektant pamięci może w pewnym zakresie dobierać dowolnie.

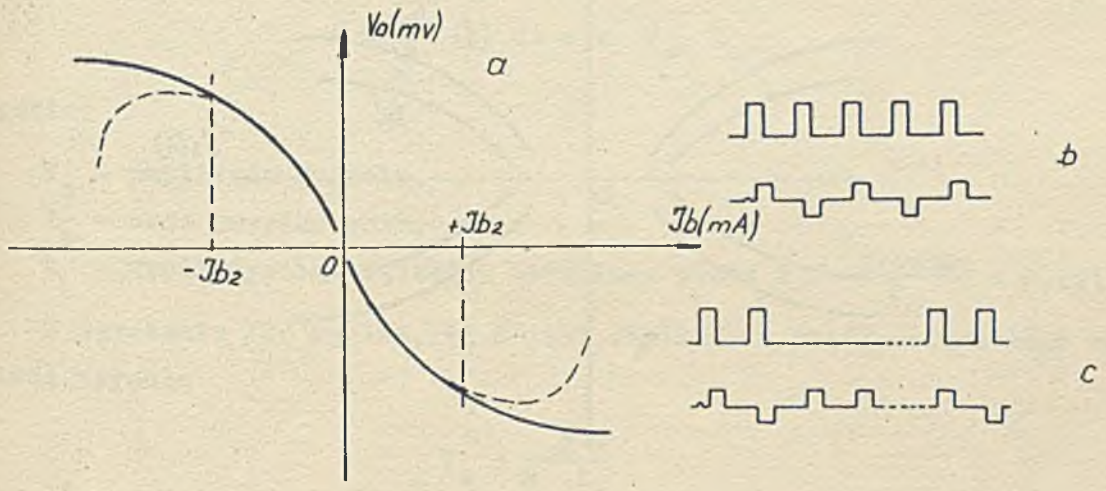


Rys. 4. Ciągi impulsów I_s i I_b w czasie zdejmowania charakterystyki progu przełączania

Charakterystyka progu przełączania zawiera istotne informacje, jednakże są one niewystarczające do określenia optymalnych kryteriów kontroli. Dalszych niezbędnych informacji dostarcza zespół charakterystyk, określający zależność sygnału wyjściowego V_o od parametrów przełączania warstwy w czasie zapisu i odczytu.

Typowe charakterystyki zapisu są przedstawione na rysunku 5a. Linia ciągła odpowiada warunkom zapisu określonym przez ciągi impulsów według

rys. 5b, natomiast linia przerywana odpowiada warunkom zapisu, kiedy występuje naruszanie wcześniej "zapisanego" stanu magnetyzacji warstwy przez określoną serię impulsów działających w kierunku przeciwnym, rys. 5c.



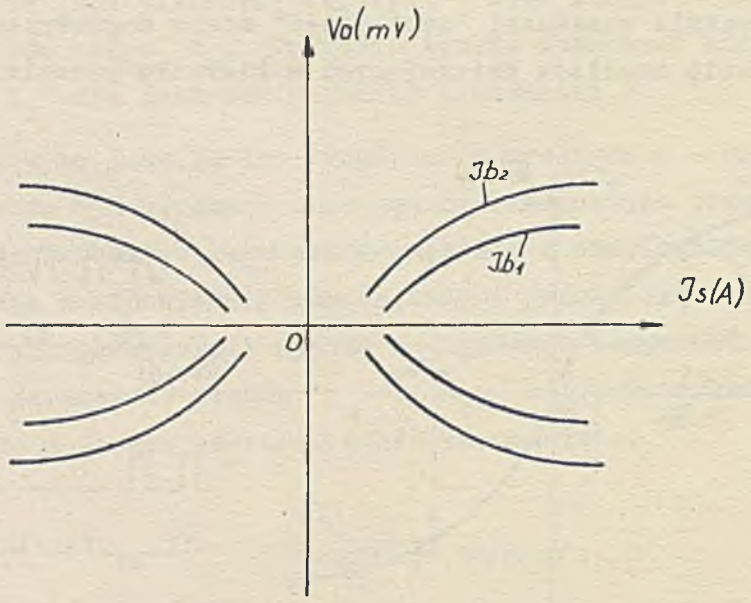
Rys. 5. Charakterystyki zapisu V_o (I_b) przy I_s = const.

Z przedstawionych charakterystyk wynika, że zapis nie może być dokonywany impulsem prądu o amplitudzie I_b większej od pewnej krytycznej wartości, przy której występuje naruszenie stanu magnetyzacji zapisanego bitu. Taką krytyczną wartością prądu I_b na rys. 5 jest wartość I_{b2}, przy której zgodnie z warunkiem /1/, pole $H_b > H_{b2}$.

Charakterystyki odczytu zostały przedstawione na rysunku 6.

Z charakterystyk tych wynika także, że sygnał odczytu V zależy zarówno od wartości prądu I_s jak i prądu I_b. Wartość górna I_{b2} jest wartością krytyczną odczytaną z wykresu na rys. 5. Wartość dolna I_{B1} jest wartością krytyczną ze względu na dopuszczalną minimalną amplitudę sygnału odczytu V_o. Taki zespół charakterystyk pozwala na określenie optymalnego punktu pracy, przy określonym i stałym nachyleniu czoła impulsu I_s.

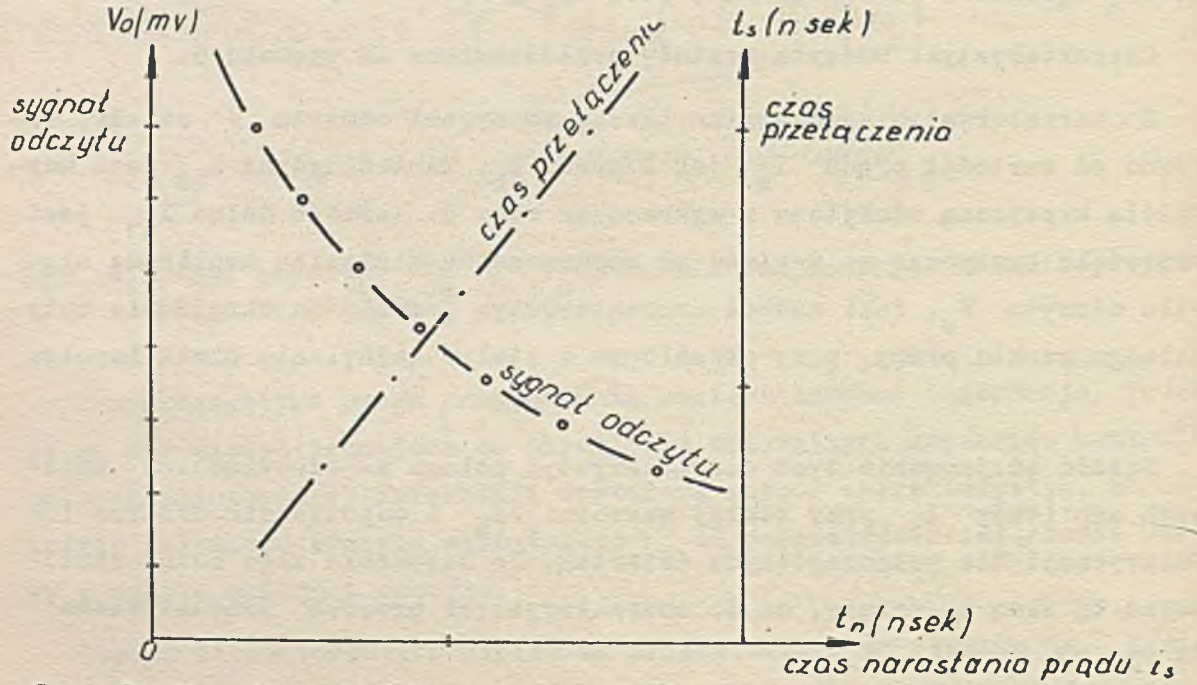
Zasada zdejmowania tych charakterystyk polega na odpowiednich zmianach amplitudy I_s przy stałej wartości I_b i odpowiednim doborze ich polaryzacji dla poszczególnych ćwiartek. Do uzyskania tego można zastosować tę samą aparaturę, co do charakterystyki progowej przedstawionej na rys. 3.



Rys. 6. Charakterystyki odczytu V_o (I_s) przy I_b = const.

Impuls odczytu V_o zależy nie tylko od amplitudy impulsów prądów zapisu i odczytu I_s i I_b, lecz także od czasu przełączania t_p. Ten zaś z kolei zależy od czasu narastania t_n impulsu prądu I_s.

Na rys. 7 przedstawiono typowe zależności tych wielkości przy stałej amplitudzie pradu I_s.



Rys. 7. Charakterystyki czasu przyłączania t_p i sygnału odczytu V_o

Sygnał odczytu określony jest zależnością:

$$V(t) = \frac{d\,\emptyset}{dt}$$

stąd

$$\emptyset = \int_{o}^{t_p} V(t)\,dt = k\ V_o\ t_p, \qquad /2/$$

gdzie:

V_o - amplituda sygnału,

t_p - czas przełączania,

k - współczynnik kształtu, określony przez $k = \frac{V_{średnie}}{V_o} = const.$

Z wyrażenia /2/ wynika, że $\emptyset$ jest wielkością stałą, wobec tego amplituda sygnału

$$V_o = \frac{\emptyset}{k\ t_p}$$

posiada hiberpoliczną zależność od t_p, co ilustruje rys. 7.

Z przedstawionych na rys. 7 charakterystyk wynika, że proces przełączania warstwy nadąża za wzrostem natężenia prądu impulsu I_s w czasie narastania czoła tego impulsu. Zależność taka jest słuszna w zakresie czasów przełączania określonych przez czas narastania prądu I_s, które są znacznie większe od minimalnych możliwych czasów przełączania warstwy. Dzięki temu możemy wpływać na czas trwania przełączania t_p warstwy magnetycznej poprzez zmianę czasu narastania t_n prądu I_s /do amplitudy I_s = const./.

Z tego wynika, że czas narastania prądu I_s ma istotny wpływ na amplitudę sygnału odczytu V_o. Oznacza to, że w rzeczywistych urządzeniach przypadkowe zmiany czasu narastania impulsu I_s mogą powodować niedopuszczalne zmiany sygnału V_o. Dlatego też aparatura pomiarowa i układy pamięci muszą być tak zbudowane, aby zapewniały stabilność czasu narastania prądu odczytu I_s.

Omawiane charakterystyki dobrze określają typowe zależności dynamicznych parametrów drutu magnetycznego w warunkach zbliżonych do rzeczywistych warunków pracy pamięci. Na ich podstawie można określić kryteria oceny podstawowych parametrów drutów magnetycznych.

Kryteria oceny drutów powinny uwzględniać najgorsze przypadki zmian parametrów prądów I_b i I_s, które mogą występować podczas pracy pamięci, przede wszystkim zaś powinny określać punkty krytyczne, tzn. granice zakresu poprawnej pracy. Takie warunki oceny pozwolą na wybór drutów magnetycznych, których parametry będą wykazywały rozrzut wynikający tylko z dokładności i stabilności aparatury kontrolno-pomiarowej. Na podstawie charakterystyk przedstawionych na rys. 5 i rys. 6 ustalamy nominalny punkt pracy, następnie punkty krytyczne dla I_b oraz kilku wartości I_s /jako parametru/, przy których będą realizowane pomiary, Należy jeszcze raz podkreślić, że parametry impulsów I_b i I_s muszą wykazywać cechy dużej stabilności. W szczególności czas narastania t_n impulsu I_s musi być utrzymany w bardzo wąskich granicach dopuszczalnych odchyleń. W praktyce granice te wynoszą $\pm$ /0,5 - 1,0/% wartości t_n w urządzeniach pomiarowych, natomiast w pamięciach mogą być łagodniejsze.

3. Definicje kryteriów oceny

Ogólne kryterium oceny przydatności jednobitowego odcinka drutu magnetycznego określimy jako dopuszczalne zmniejszenie amplitudy sygnału V_n o wartośc ΔV. Sygnał V_n jest nominalnym sygnałem, odczytanym w nominalnych warunkach z wzorcowego odcinka drutu magnetycznego. Dopuszczalną dolną graniczną wartość sygnału V_d określimy za pomocą współczynników ε_1 i ε_2. Mamy więc:

$$V_d = V_n \cdot \frac{V_n - \Delta V}{V_n} = V_n \cdot \varepsilon_1 \cdot \varepsilon_2 \qquad /3/$$

Zakładamy, że w warunkach nominalnych zachodzi 100% przemagnesowanie badanego odcinka warstwy magnetycznej.

Najgorszy przypadek warunków pracy danej komórki jednobitowej zachodzi wtedy, gdy kolejno zachodzą po sobie dwa następujące zdarzenia:

- Strumień magnetyczny przełączanej warstwy w czasie odczytu będzie mniejszy od nominalnego. Może to być spowodowane mniejszym przemagnesowaniem warstwy w czasie zapisu, albo mniejszą wartością magnetyzacji danej komórki, albo też mniejszym przekrojem warstwy itp.
- Ten mniejszy od nominalnego strumień przełączanej warstwy będzie dodatkowo obniżony. Może to być spowodowane działaniem impulsów prądu zapisu o przeciwnej polaryzacji i o zwiększonej amplitudzie.

Na podstawie tych zdarzeń zdefiniujemy dwa zasadnicze k r y t e r i a oceny:

1. Wartość prądu zapisu niezbędna do wywołania sygnału odczytu o wartości $\varepsilon_1 V_n$ jest jak najmniejsza I_{bmin}. Impuls prądu I_{bmin} występuje w koincydencji z impulsem prądu I_s. Pomiar prądu I_{bmin} może być przeprowadzany dla kilku określonych wartości I_s. Badany odcinek drutu magnetycznego posiada tym lepsze właściwości im prąd I_{bmin} jest mniejszy. Dlatego też odcinki drutu magnetycznego, dla których

$$I_{bmin} > I_{b1} \qquad /4/$$

muszą być odrzucone. I_{b1} jest maksymalną wartością dolnej granicy prądu zapisu I_{bmin}.

2. Wartość impulsów prądu, które naruszają w sposób określony stan magnetyzacji warstwy "zapisany" prądami I_{bmin} i I_s jest jak największa I_{bmax}. Pomiar I_{bmax} odbywa się w chwili, gdy amplituda sygnału odczytanego spadnie do wartości $\varepsilon_2 \cdot \varepsilon_1 V_n$. Impuls prądu I_{bmax} posiada przeciwną polaryzację jak w czasie zapisu. Badany odcinek drutu magnetycznego posiada tym lepsze parametry im prąd I_{bmax} jest większy. Dlatego też odcinki drutu magnetycznego, dla których

$$I_{bmax} < I_{b2} \qquad /5/$$

muszą być odrzucone. I_{b2} jest minimalną wartością górnej granicy prądu zapisu I_{bmax}.

Na podstawie warunków /4/ i /5/ można określić oczekiwany margines pracy pamięci, który wyrazi się różnicą wartości prądu zapisu dla górnej i dolnej granicy.

$$I_b = I_{b2} - I_{b1} \qquad /6/$$

Wartość nominalna prądu I_b będzie określona przez:

$$I_b = \frac{I_{b2} + I_{b1}}{2} \qquad /7/$$

Jeżeli charakterystyki przedstawione na rys. 5 i rys. 6 uwzględniają powyższe kryteria, są typowe i dobrze określają parametry drutów magnetycznych.

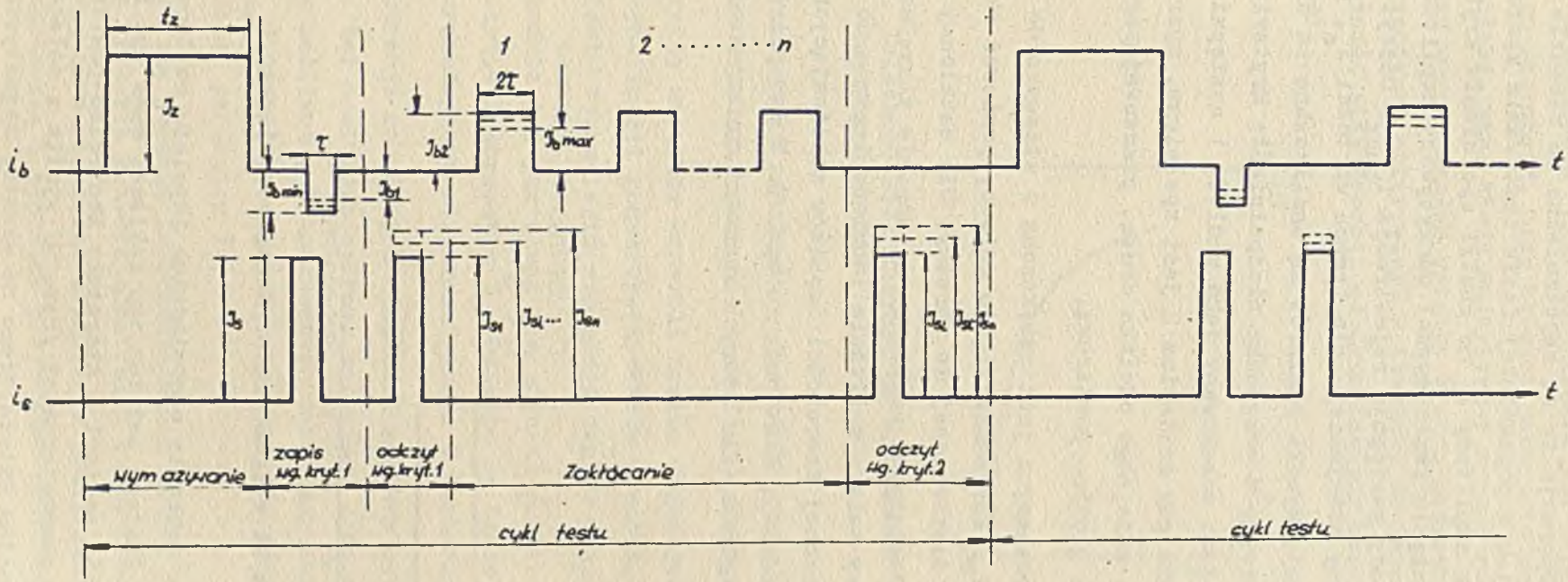
4. Metoda pomiaru

Metoda pomiaru jest oparta o kryteria zdefiniowane w rozdziale 3. Polega ona na pobudzaniu badanego odcinka drutu magnetycznego ciągami impulsów zapisujących, zakłócających i odczytujących, następujących po sobie w odpowiedniej sekwencji, ilości i o dokładnie określonych parametrach. Harmonogram testowania drutu jest ściśle ustalony /rys. 8/. Występują w nim przedziały czasu, w których następuje przygotowanie drutu magnetycznego do pomiaru, i przedziały czasu, w których przeprowadza się pomiar i ocenę przydatności badanego drutu.

Impulsy I_b posyłane są do badanego drutu magnetycznego, impulsy I_s do linii słowa, usytuowanej prostopadle i obejmującej drut jednym lub kilkoma zwojami. Układ impulsów tworzy cykl testu, który jest powtarzany do czasu wykonania pomiarów według kryterium 1, a następnie według kryterium 2.

W cyklu tym wyróżniamy kilka kolejnych faz:

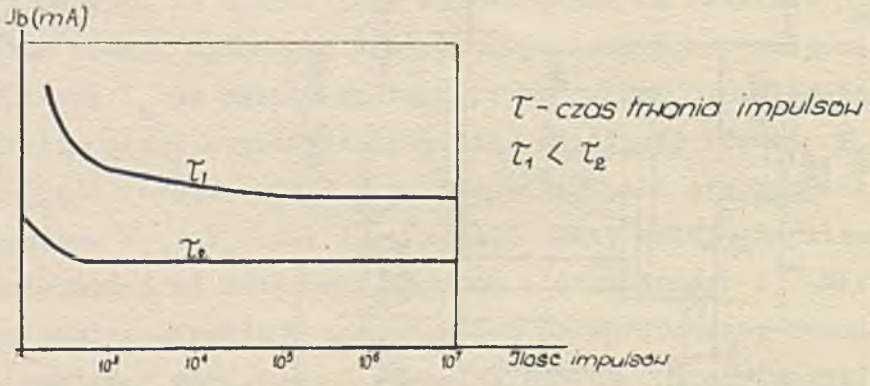
1. Wymazywanie /czyszczenie, zerowanie/ impulsem I_z o czasie trwania t_z. Parametry tego impulsu są tak dobrane, aby w wyniku jego działania następowało pełne przemagnesowanie warstwy magnetycznej badanego odcinka drutu. Dlatego należy założyć, że $I_z \gg I_{bmin}$ i $t_z \gg \tau$.
2. Zapis i odczyt. Te operacje są współzależne na zasadzie pierwszego kryterium oceny. Aparatura kontrolno-pomiarowa automatycznie reguluje amplitudę I_b do chwili spełnienia kryterium 1, polegającego na osiągnięciu przez sygnał odczytywanej amplitudy równej $\varepsilon_1 V_n$. Wtedy następuje rejestracja I_{bmin}. Odczyt sygnału przeprowadza się serią impulsów I_{s1}, I_{s2}..., I_{sk}, w celu oceny zmiany amplitudy sygnału w zależności od natężenia impulsu odczytującego I_s. Należy założyć, że pomiar sygnału V powinien być przeprowadzany conajmniej dla trzech wartości amplitudy I_s: nominalnej I_{sn}, minimalnej dopuszczalnej I_{smin} i maksymalnej możliwej I_{smax}. Wyniki pomiarów są analizowane przez aparaturę pomiarową, która automatycznie kwalifikuje badany element. W przypadku niespełnienia kryterium 1, badany odcinek drutu jest ocechowany jako wadliwy i następnie zostaje odcięty. W przypadku gdy kryterium 1 jest spełnione, aparatura kontynuuje testowanie, przechodząc do następnej fazy.



Rys. 8. Harmonogram impulsów testujących

3. Zakłócenia i odczyt. Operacje te są współzależne na zasadzie kryterium 2. Analogicznie jak w poprzedniej fazie testowania aparatura automatycznie reguluje amplitudę I_b do chwili spełnienia kryterium 2, polegającego na osiągnięciu przez sygnał odczytywany amplitudy równej $\varepsilon_2 \cdot \varepsilon_1 \cdot V_n$. Wtedy następuje rejestracja I_{bmax}. Odczyt sygnału powinien być dokonywany conajmniej dla trzech wartości I_g, podobnie jak w fazie poprzedniej. Wyniki pomiarów są analizowane na bieżąco przez aparaturę pomiarową. W przypadku niespełnienia kryterium 2, badany odcinek drutu jest ocechowany jako wadliwy i następnie zostaje odcięty. W przypadku gdy kryterium 2 jest spełnione, aparatura kontynuuje testowanie kolejnego odcinka drutu, przesuwającego się w sposób ciągły przez głowicę pomiarową.

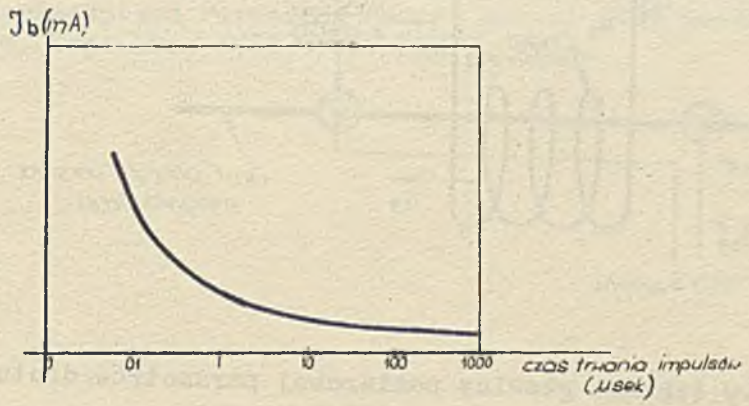
Szerokość impulsów zakłócających jest zwiększona w stosunku do szerokości impulsu nominalnego I_b. Wprowadza się to celowo, ponieważ skuteczność zakłóceń zależy od energii dostarczonej do warstwy. Dla ustalonej amplitudy i szerokości impulsów zakłócających, efektywne naruszanie stanu magnetyzacji badanego odcinka warstwy maleje z ilością impulsów i zdąża do pewnej stałej granicy. Przy zwiększonej szerokości impulsów zakłócających granica ta zostanie osiągnięta wcześniej. Ilustruje to rys. 9. W tym przypadku charakterystyki odpowiadają punktom, dla których zachodzi zmniejszenie magnetyzacji warstwy o połowę.



Rys. 9. Charakterystyki wpływu prądu rozmagnesowującego I_b

Z powyższego wynika, że dobranie odpowiednich impulsów zakłócających jest sprawą szczególnie ważną. Dlatego też ten typ zakłóceń musi być brany pod uwagę w czasie projektowania pamięci i aparatury kontrolno-pomiarowej. Zjawisko opisane przy pomocy charakterystyk /rys. 9/ wynika z zależności siły koercji od szerokości impulsu działającego na warstwę magnetyczną. Zależ-

ność tę przedstawia rys. 10. Określa ona amplitudę pojedynczego impulsu I_b i jego szerokość, przy których następuje zmiana stanu magnetyzacji warstwy o połowę, tzn. kiedy odczytywany sygnał równa się zero. Odczyt jest poprzedzony wprowadzeniem warstwy w stan nasycenia przy pomocy dużego impulsu o przeciwnej polaryzacji w stosunku do impulsu zapisu. Widzimy tu waraźną zależność siły koercji od dynamicznych warunków sterowania warstwy magnetycznej.



Rys. 10. Zależność siły koercji od szerokości impulsu działającego na warstwę magnetyczną

Opisana metoda umożliwia rejestrację wartości I_{bmin} i I_{bmax}. Dane te mogą mieć duże znaczenie dla formułowania wniosków, co do:

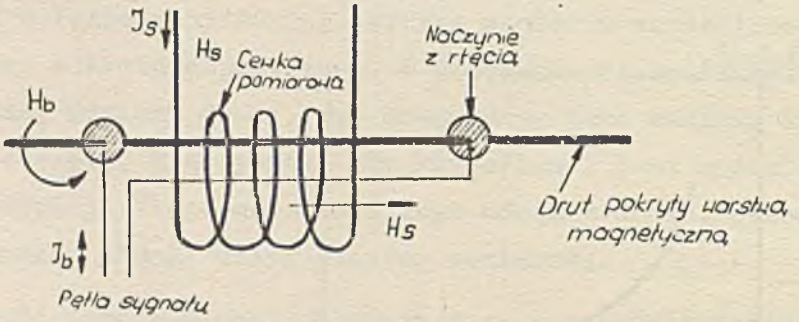
- korekty warunków kontroli drutów magnetycznych,
- optymalizacji konstrukcji płatów pamięci,
- korekty technologii produkcji drutów magnetycznych.

W metodzie tej postuluje się również pomiar sygnałów V_o dla conajmniej trzech wartości parametru I_s. Dane te są także pomocne dla ewentualnej korekty warunków kontroli, konstrukcji i technologii.

Rejestracja omawianych wielkości mogłaby być dokonywana przez odpowiedni konwerter analogowo-cyfrowy sprzężony z dalekopisem. Oczywiście jest to tylko możliwość. W konkretnych warunkach można zrezygnować z rejestracji I_{bmin} i I_{bmax} a ponadto ograniczyć pomiar tylko do jednej optymalnej wartości parametru I_s.

Wydaje się jednak, że przedstawiona metoda jest stosunkowo prosta, a przyjęte kryteria pozwalają na pełną ocenę drutów w warunkach dynamicznych, przy czym niezbędne urządzenia pomiarowe nie są skomplikowane i mogą być zbudowane z typowych przyrządów pomiarowych, używanych do automa-

tycznej selekcji mikroferrytów. Oczywiście konieczna będzie pewna modyfikacja niektórych członów tej aparatury, jak generatory prądów I_s i I_b oraz wzmacniacz sygnałów V_o. Zamiast skomplikowanego, precyzyjnego, automatycznego podajnika rdzeni mikroferrytowych wystąpi tu stosunkowo prosta głowica pomiarowa, której schemat ideowy przedstawia rysunek 11.



Rys. 11. Ideowy schemat głowicy pomiarowej parametrów drutu magnetycznego

Kontakt elektryczny z drutem magnetycznym jest zapewniony za pośrednictwem naczynia z rtęcią, przez które w sposób ciągły przesuwa się badany drut magnetyczny. W pętli sygnału indukuje się tylko napięcie pochodzące od strumienia przełączanej warstwy. Wpływ pola magnetycznego wywołanego prądem I_s oraz sprzężenia pojemnościowe są skompensowane.

Oczywiście przedstawiona metoda pomiaru drutów magnetycznych nie jest jedyną. Mogą być również przyjęte inne kryteria kontroli, np. ocena według granicznych dynamicznych wartości pętli histerezy B i H.

Literatura

[1] Z. SAWICKI, J. RYŻKO: Podstawowe właściwości cienkich warstw magnetycznych stosowanych w pamięciach EMC. Biul. Inf. IMM. Nowości Techniczne nr 4, 1967 r.

[2] Z. SAWICKI: Bloki nośnika informacji na drutach magnetycznych pamięci EMC. Biul. Inf. IMM. Nowości Techniczne IMM, nr 3, 1967 r.

[3] H.S. BELSON: Measurement of Skew, Dispersion and Creep in Plated Wires. Proceedings of Intermag Conf. 1963, p. 12-4-1 do 5-4-6.

[4] I. DANYLCHUK, A.I. PERNESKI and M.W. SAGAL: Plated Wire Magnetic Film Memories. Proceedings of Intermag Conf. 1964, p. 5-4-1 do 5-4-6.

[5] H. MAEDAR and MATSUSHITA: Woven Thin-film Memories. Proceedings of Intermag Conf. 1964, p. 8-1-1 do 8-1-6.

[6] K. FULAMI, S. OSHIMA and T. KAMIBYASHI: The Plated-wire Memory Matrix. Proceedings of Intermag, Conf. 1964, p. 5-1-1 do 5-1-6.

[7] E.W. PUGH, V.T. SHAHAM, W.T. SIEGLE: Device and Array Design for a 120-Nanosecond Magnetic Film Main Memory. IBM Journal, March 1967.

[8] I.T. CHANG, U.F. GIANOLA and M.W. SAGAL: Aging and Stabilization of Electrodeposited Cylindrical Magnetic Films. Journal of Applied Physics, vol. 35, nr 3, March 1964.

[9] B. RABINOVICI and C.A. RENTON: Some Aging Properties of Cylindrical Electrodeposited Permalloy Memory Wires. IEEE Transactions of Magnetics, September 1966, vol. MAG 2, nr 3.

Mgr inż. TADEUSZ WERNER
Zakład Organizacji Maszyn Cyfrowych IMM

681.3.06:681.327.6

PAMIĘC WIRTUALNA MASZYNY CYFROWEJ I ALGORYTMY ZAMIANY

1. Wstęp

Niniejsze opracowanie jest poświęcone pewnemu systemowi organizacji współpracy urządzeń pamięciowych wchodzących w skład zestawu maszyny cyfrowej, zwanemu pamięcią wirtualną. Główna część pracy jest oparta na artykule L.A. Belady'ego pt. A S t u d y o f R e p l a c e m e n t A l g o r i t h m s f o r a V i r t u a l - s t o r a g e C o m p u t e r. Punkty 2 i 3 zawierają wprowadzenie w zagadnienie organizacji pamięci wirtualnej, w punktach 4 i 5 znajduje się opis ważniejszych znanych algorytmów zamiany, a w punkcie 6 definicja i omówienie sprawności tych algorytmów. Punkt 7 jest oparty na artykule T. Kilburne'a pt. O n e - l e v e l S t o r a g e S y s t e m, a punkt 8 na opracowaniu E n g l i s h E l e c t r i c C o m p. S y s t e m 4-75 /patrz spis literatury/. Zawierają one informacje o pamięciach wirtualnych maszyn Atlas i System 4-75, podając dodatkowo dla tej drugiej pewne wiadomości o systemie stronicowania i segmentacji.

2. Wprowadzenie

Jednym z głównych parametrów decydujących o możliwościach maszyny cyfrowej jest pojemność szybkiej pamięci operacyjnej. Budowa szybkich pamięci o wielkich pojemnościach jest droga i trudna technicznie, zachodzi więc konieczność stosowania pamięci o ograniczonych pojemnościach i przechowywania większej części programu i danych w pamięci zewnętrznej. Szczególnie dotyczy to dużych systemów przeznaczonych do pracy wieloprogramowej. Taka praca maszyny wymaga od programistów znacznego wysiłku przy rozplanowaniu i kodowaniu przesyłania informacji między pamięcią operacyjną a zewnętrzną. Częste przesyłanie komplikuje i wydłuża program.

W ostatnich latach coraz częściej powstają maszyny wyposażone w pamięć technicznie niejednolitą, a organizacyjnie jednopoziomową, czyli maszyny o pamięci wirtualnej. Automatyczna wymiana bloków informacji między pamięcią operacyjną a pamięcią zewnętrzną o stosunkowo krótkim czasie dostępu /bębny, dyski/ umożliwia jednolite i bezpośrednie adresowanie wszystkich miejsc pamięci zewnętrznej. Taka organizacja dostarcza efektywnie znacznie rozszerzonego zakresu pamięci operacyjnej.

3. Logika pamięci wirtualnej

Maszyna o pamięci wirtualnej przechowuje w pamięci operacyjnej w każdej chwili tylko bloki najpotrzebniejsze, zachowując pozostałe bloki w pamięci zewnętrznej. W pamięci zewnętrznej mieści się dużo więcej bloków niż w pamięci operacyjnej. Pomiędzy pamięcią operacyjną a zewnętrzną zachodzi automatyczna wymiana bloków. Przez "blok" rozumie się tutaj ustalonej wielkości fragment programu lub danych. Każdy blok ma swoje określone miejsce w pamięci zewnętrznej. Miejsce na blok w pamięci operacyjnej nazywa się stroną. Każda strona może zawierać dowolny blok. Blok z reguły zawiera 2^k słów lub znaków /b y t e´ ów/, gdzie k jest liczbą naturalną.

Jeśli zachodzi potrzeba skorzystania z informacji zawartej w bloku znajdującym się w pamięci operacyjnej, następuje odwołanie się do tej pamięci. Jeżeli potrzebna informacja znajduje się w pamięci zewnętrznej, następuje przepisanie odpowiedniego bloku z tej pamięci do pamięci operacyjnej /wciągnięcie/, poprzedzone opróżnieniem jednej ze stron, polegającym na przepisaniu bloku z tej strony na swoje miejsce w pamięci zewnętrznej /wypchnięcie/. Dopiero po wciągnięciu bloku można skorzystać z zawartej w nim informacji. Strony do opróżnienia, a więc i bloki do wypchnięcia, wyznacza się przy pomocy algorytmu zamiany, realizowanego środkami technicznymi lub programowymi /system operacyjny/. Znany jest cały szereg takich algorytmów.

Podawany przez jednostkę centralną adres odnosi się do pełnego zakresu pamięci wirtualnej i nazywa się adresem wirtualnym. Dzieli się on na dwie części, starszą - wirtualny adres bloku i młodszą - adres wewnątrz bloku. Wirtualny adres bloku określa miejsce bloku w pamięci zewnętrznej. Każdej stronie przyporządkowany jest jej adres w pamięci operacyjnej. Jest to pamięciowy adres strony, nie będący adresem wirtualnym. Jest on zarazem adresem pamięciowym bloku wpisanego do tej strony. Dołączenie do pamięciowe-

go adresu strony adresu wewnątrz bloku tworzy pamięciowy adres słowa, używany do określania miejsca pamięci operacyjnej, w którym znajduje się potrzebne słowo. Ponadto każdej stronie przyporządkowany jest rejestr adresu wirtualnego, przechowujący wirtualny adres bloku. Adres ten zostaje wykorzystany przy wypchnięciu bloku z danej strony. Rejestry adresów wirtualnych z przyporządkowanymi im pamięciowymi adresami stron tworzą tabelę odwzorowań.

Kiedy jednostka centralna, odwołując się do pamięci, podaje adres wirtualny, następuje badanie tabeli odwzorowań - porównywanie wirtualnego adresu bloku z jej pozycjami. W ten sposób bada się, czy żądany blok znajduje się w pamięci operacyjnej. Wskazanie równoważności adresów /wirtualny adres bloku zapisany w pozycji tabeli i wirtualny adres bloku podany z części centralnej/ oznacza obecność bloku. W takim przypadku dołącza się przyporządkowany pozycji tabeli pamięciowy adres strony do adresu wewnątrz bloku i tak otrzymany pamięciowy adres słowa /lub znaku/ posyła się do pamięci operacyjnej.

Jeśli po zbadaniu całej tabeli stwierdza się brak równoważności adresów, przystępuje się do operacji zamiany bloków. Algorytm zamiany określa blok do wypchnięcia. Następuje przepisanie tego bloku do jego miejsca w pamięci zewnętrznej - według zawartości odpowiedniego rejestru adresu wirtualnego. Następnie, wykorzystując podany wirtualny adres bloku, wciąga się potrzebny blok do opóźnionej strony pamięci operacyjnej i wpisuje nowy wirtualny adres bloku do rejestru adresu wirtualnego tej strony, po czym z pamięciowego adresu strony i adresu wewnątrz bloku tworzy się pamięciowy adres słowa.

Ponieważ powyższe czynności wymagają czasu, stosuje się też system, w którym wypchnięcie wykonuje się bezpośrednio po wciągnięciu. W ten sposób uwalnia się jedną stronę dla potrzeb następnego wciągnięcia bloku. W maszynie stosującej podział czasu pamięci operacyjnej można wprowadzić przy tym systemie dużą oszczędność czasu jednostki centralnej, jednak kosztem gorszego wykorzystania pamięci operacyjnej, która ma stale jedną stronę pustą.

4. Algorytmy zamiany

Wybór algorytmu zamiany jest sprawą bardzo ważną. Algorytmy znacznie różnią się między sobą co do sprawności, która zależy od wielkości pamię-

ci operacyjnej i bloku. Dobry algorytm minimalizuje ilość przesłań bloków. Zadaniem jego jest wybrać do wypchnięcia taki blok, który nie będzie już więcej używany, bądź nie będzie używany przez czas najdłuższy, a unikać wyboru bloku, który w najbliższym czasie ponownie będzie potrzebny do odczytu lub zapisu któregokolwiek z jego słów. Niefortunny wybór bloku powoduje z reguły konieczność ponownego jego wciągnięcia w najbliższym czasie, co oznacza dodatkową zamianę do wykonania.

Algorytmy zamiany można podzielić na trzy klasy. W algorytmach klasy I zakłada się zupełną przypadkowość używania bloków przez program i nie korzysta się z informacji o przebiegu wykorzystywania pamięci operacyjnej. Algorytmy wyższych klas posługują się takimi informacjami celem lepszego przeprowadzania wyboru bloków do wypchnięcia. W algorytmach klasy II klasyfikuje się bloki na podstawie przebiegu ich dotychczasowego używania w pamięci operacyjnej, a w klasie III na podstawie przebiegu ich obecności i nieobecności w tej pamięci.

Najprostszym algorytmem klasy I jest RAND, polegający na przypadkowym wyborze bloków z pamięci operacyjnej. Chociaż dla celów praktycznych lepsze są inne algorytmy, ma on znaczenie porównawcze i jest ciekawy z tego względu, że łatwo jest podać dla niego wzór na prawdopodobieństwo dobrego wyboru bloku, czyli teoretyczną sprawność algorytmu. Uważa się, że stwierdzenie wystąpienia odwołania się do tego bloku, który znajduje się w pamięci zewnętrznej i ostatnio doznał wypchnięcia, czyli należy do tzw. zbioru R, jest równoważne z wykryciem uprzedniego złego wyboru bloku. Blok zalicza się do zbioru R, jeśli blok, który go wypchnął, znajduje się aktualnie w pamięci operacyjnej. Zbiór R składa się z bloków w ilości od 1 do c, gdzie c jest ilością stron w pamięci operacyjnej. Przynależność do R może być wielokrotna. Prawdopodobieństwo dobrego wyboru bloku jest teoretyczną sprawnością algorytmu i w przypadku algorytmu RAND wzrasta przy wzroście stosunku wielkości programu do wielkości pamięci operacyjnej. Założenia przypadkowości na ogół nie bywają spełniane i sprawność teoretyczna najczęściej znacznie różni się od rzeczywistej.

Inny algorytm klasy I, FIFO, wybiera ten blok, który spędził w pamięci operacyjnej najwięcej czasu. W pewnych typach programów istnieje duże prawdopodobieństwo, że taki blok nie będzie już więcej używany. Do realizacji FIFO każda strona pamięci musi mieć swój licznik czasu, przy czym miarą czasu powinna być ilość rozkazów wykonanych przez jednostkę centralną.

W klasie II wyróżnia się następujące algorytmy:

. Algorytm LT, w którym stosuje się dynamiczne porządkowanie wszystkich bloków obecnych w pamięci operacyjnej i wybiera ten, do którego od najdłuższego czasu nie było odwołania się. Potrzebne tu są liczniki czasu dla stron pamięci.

. Algorytm S-3, w którym używa się par bitów stanu /P, A/, przyporządkowanych każdej stronie pamięci. Wykorzystywanie ich stanowi uproszczone porządkowanie dynamiczne. Bit P zostaje wyjedynkowany przy każdym odwołaniu się do bloku, bit A tylko przy zapisie do bloku. Bity P zeruje się, wszystkie jednocześnie, w momencie zamiany bloków, o ile przedtem wszystkie zostały wyjedynkowane. Przy wypchnięciu bloku z danej strony zeruje się jej bit A. Algorytm przegląda podzbiory bloków po kolei, według wartości /P, A/ : /0,0/, /0,1/, /1,0/, /1,1/ i z najniższego niepustego podzbioru wybiera blok do zamiany. Metody wyboru w ramach podzbioru mogą być różne, np. wybór przypadkowy.

. Algorytm AR-1, który różni się od S-3 tylko tym, że zerowanie bitów P wykonuje się automatycznie, natychmiast po wyjedynkowaniu ostatniego z nich. W ten sposób podzbiory /0,0/ i /0,1/ nigdy nie są jednocześnie puste.

. Algorytm T, czyli AR-1 z dodaną pewną zasadą wyboru bloku z podzbioru. Wyboru dokonuje się według kolejności, w której bloki zostały wciągnięte.

. Algorytm ML, będący uproszczonym algorytmem AR-1, w którym nie używa się bitu A.

W klasie III znany jest tylko algorytm zastosowany w systemie ATLAS. Dla każdej strony w każdej chwili znane jest t, będące wartością czasu, jaki upłynął od momentu ostatniego odczytu z bloku umieszczonego w tej stronie. Jednocześnie każdy blok dysponuje wartością T, czyli długością czasu od ostatniego zapisu do tego bloku. Przy wciągnięciu bloku zeruje się wartość t, a za T wstawia się tę jego wartość, którą miało ono w końcu poprzedniej obecności bloku w pamięci operacyjnej. Przy wyborze bloków wyszukuje się te, dla których $t > T + k$. W maszynie ATLAS $k = 1$. Jednostką jest tu czas wykonania kolejnych 1024 rozkazów. Tak wybrane bloki mają na ogół małe prawdopodobieństwo ponownego ich użycia. Gdy takich

bloków brak, szuka się bloku, gdzie $t \neq 0$ i $/T - t/$ ma wartość maksymalną. Taki blok przy nie zmienionym charakterze pracy programu zapewne najdłużej nie będzie potrzebny. Jeśli i takiego bloku brak, bierze się ten, dla którego $t = 0$ i T ma wartość maksymalną, co nie zabezpiecza przed złym wyborem, ale zapewnia, że w przypadku ponownej obecności tego bloku w pamięci operacyjnej nie będzie on powtórnie obiektem złego wyboru.

5. Algorytm optymalny

Z powyższymi algorytmami wiąże się częste popełnianie błędów w wyborze, co powoduje w następstwie, że praca programu, pozostając prawidłową, trwa dłużej. Błędy wyboru zdarzają się dlatego, że informacje o uprzednich wydarzeniach w pracy programu tylko z pewnym prawdopodobieństwem mogą określić przebieg przyszłych odwołań się.

Możliwe jest zbudowanie tak zwanego algorytmu optymalnego, takiego, który zawsze wybiera do zamiany ten blok, o którym wiadomo na pewno, że najdłużej nie będzie używany. Zgromadzenie takich informacji wymaga jednak wykonania próbnego przebiegu pracy programu i odpowiednich badań.

Wielkość bloku, pojemność pamięci operacyjnej i wszystkie dane muszą być przy tym takie same, jak przy pracy użytkowej, aby wszystkie skoki odbyły się tak samo i ilości obiegów pętli programowych nie różniły się.

Gdy wszystkie strony pamięci operacyjnej są wypełnione blokami i następuje odwołanie się do bloku znajdującego się w pamięci zewnętrznej, zaczyna się opóźnianie decyzji wyboru bloku. Wypchnięcie, które podczas przebiegu użytkowego zostałoby w tym momencie wykonane, teraz określone zostanie z opóźnieniem. Te z bloków znajdujących się aktualnie w pamięci operacyjnej, do których następuje odwołanie się w okresie opóźniania decyzji, eliminuje się ze zbioru kandydatur do wypchnięcia. Ostatni niewyeliminowany blok zostaje z opóźnieniem określony jako optymalny do użycia podczas danej zamiany. Na ogół prowadzi się jednocześnie kilka opóźniań decyzji, gdyż przeważnie przed momentem stwierdzenia konieczności nowego wciągnięcia nie występuje wystarczająco wiele odwołań się do bloków w pamięci operacyjnej, by móc zakończyć poprzednie opóźnianie decyzji. W takim przypadku czeka się do momentu, w którym pozostanie tyle nie wyeliminowanych bloków, ile decyzji jest opóźnianych i wszystkie je umieszcza się w sekwencji optymalnych wypchnięć. Jeśli opóźnianie decyzji sięg-

nie końca pracy programu, wykorzystuje się wszystkie nie wyeliminowane do tego momentu bloki w sposób dowolny. Notowanie adresów kolejnych bloków wyznaczanych z opóźnieniem do zamiany daje optymalną sekwencję wypchnięć, zapamiętywaną na taśmach magnetycznych.

Gdyby użytkowy przebieg programu wykonać przy użyciu innych danych, wyznaczona kolejność bloków mogłaby okazać się daleką od optymalnej.

Powyższa metoda wymaga próbnego przebiegu i dużej ilości taśmy magnetycznej, nie posiada więc praktycznego znaczenia jako algorytm nadający się do zastosowań użytkowych. Służy natomiast do określania sprawności innych algorytmów.

Podczas przebiegu próbnego pobyt w pamięci operacyjnej może być sprawą umowną, opartą na odpowiednim oznaczaniu bloków.

6. Wyniki podziału na bloki i sprawności algorytmów

Ogólna sprawność pracy programu zależy od jego budowy - a w konsekwencji i od minimalnej ilości zamian, możliwej do osiągnięcia przez algorytm optymalny, od parametrów maszyny i sprawności użytego algorytmu zamiany. Sprawność algorytmu jest to stosunek ilości zamian wykonanych przy zastosowaniu algorytmu optymalnego do ilości zamian wykonanych przez algorytm badany - przy tym samym programie i identycznych danych.

Firma IBM wykonała program algorytmu optymalnego MIN i zestaw programów badawczych SIM. Taki program śledzi pracę badanego programu, posługując się nim jak danymi, i zapamiętuje sekwencję jego odwołań się do pamięci. Służy to następnie do porównania ilości zamian wykonanych przez algorytm optymalny i algorytm rozważany, a także do innych badań.

Dla pewnego typowego programu zależność stosunku ilości słów, które wraz ze swoimi blokami doznały wciągnięcia, do ilości wszystkich słów programu wygląda następująco: dla bloku o 32 słowach - 0,58, dla 1024 - 0,88. W skali logarytmicznej zależność powyższa zbliżona jest do liniowej i nie zależy od pojemności pamięci ani od użytego algorytmu. Natomiast zależy w niewielkim stopniu od rodzaju programu. Niższa wartość tego stosunku oznacza oszczędność miejsc w pamięci. Dwukrotne zmniejszenie wielkości bloku daje przeciętnie 10-15% oszczędności pamięci.

Częstość całkowitych wymian zawartości pamięci operacyjnej szybko maleje przy wzroście ilości stron pamięci operacyjnej, a także przy zastosowa-

niu większej pamięci. Jest ona proporcjonalna do ilości zamian, która z kolei, jak wykazuje doświadczenie, jest proporcjonalna do a/c^k, gdzie a i k to pewne stałe charakteryzujące dany program /1,5 < k < 2,5/, a c jest ilością stron w pamięci operacyjnej.

Zastosowanie większych bloków zmniejsza czas oczekiwania przypadający na jedno słowo w związku ze ściąganiem bloku z bębna lub dysku. Czas ten maleje w przybliżeniu proporcjonalnie do wielkości bloku. W przypadku bloków zawierających mniej niż 64 słowa czas ten jest już stanowczo zbyt długi.

Przy projektowaniu maszyny wybiera się wielkość bloku. Decyzja opierać się musi na parametrach odnoszących się do pamięci i całej maszyny, uwzględniać powyższe zależności oraz zależność sprawności wybranego algorytmu od wielkosci bloku i pojemności pamięci operacyjnej. Należy też pamiętać o utrzymaniu rozsądnych wymiarów tabeli odwzorowań.

Sprawności algorytmów FIFO i RAND są sobie bliskie. Dla pewnego przykładowego programu są one wielkościami rzędu 10-15% przy małych blokach i 35-40% przy większych blokach i większych pojemnościach pamięci, choć przebiegi zależności są mało regularne. Dla tego samego programu sprawność LT waha się, przy różnych wielkościach bloku i pamięci, bardzo nieregularnie między 10% a 80%. Sprawności S-3 i AR-1 nie różnią się między sobą i zawierają się w granicach 40% i 65%. Sprawności ML i T są nieznacznie niższe od S-3; tylko przy małej pamięci i jednocześnie małych blokach różnica ta dochodzi do 15-20%. Sprawności ATLAS'a wahają się nieregularnie między 15% a 50%, a przy dużych wielkościach bloku spadają do kilku procent. Na ogół ze wzrostem wielkości bloku sprawność nieco wzrasta.

Zbadano też granice sprawności FIFO, RAND i AR-1 dla sześciu różnych programów i szerokiego zakresu wielkości pamięci. Dolne granice mają tendencję wzrostu ze wzrostem wielkości bloku - od 15% do 20% dla RAND, od 10% do 40% dla FIFO, od 30% do 50% dla AR-1. Górne granice wynoszą - dla AR-1 100%, dla RAND i FIFO przy blokach nie większych od 512 słów - 100%, powyżej - spadają do 70%. Wartości te odnoszą się do wielkości bloku zmieniających się w granicach od 64 do 2048 słów.

Używanie bitów A pozwala opuszczać wypchnięcia tych bloków, do których nie było zapisu. Oparta na tym specjalna wersja MIN pozwala określić sprawność poprawioną, odnoszącą się do zminimalizowanej sumy wciągnięć i wypchnięć. Dla tego samego przykładowego programu /I n t e g e r p r o -

g r a m m i n g z klasy programowania liniowego: wielkość 3,2 K,1,1.10^7 wykonanych rozkazów, FORTRAN/ i specjalnego MIN stosunek ilości wypchnięć i wciągnięć waha się między 35% dla małych bloków o 90% dla dużych bloków. Zwiększenie pamięci zmniejsza ten stosunek o kilkanaście procent.Dla AR-1 stosunek ten jest przy małych blokach wyższy niż dla MIN, a przy dużych blokach niższy. Różnice dochodzą do 30-40%. Dla większych pamięci stosunek ten maleje.

Przy danym algorytmie, danej wielkości pamięci i stron, przy tym samym programie i tych samych danych ilość wykonanych zamian może być bardzo różna - w zależności od podziału programu i danych na bloki. Zależy to od listy rozkazów maszyny, organizacji adresowania, umiejętności programisty, budowy translatora itd.

Omówiony system pamięci wirtualnej rozszerza zakres pamięci operacyjnej, a w maszynach przystosowanych do pracy wieloprogramowej jest wykorzystywany do automatycznego podziału pamięci między jednocześnie pracujące programy. W maszynach takich jednocześnie z pracą kanału wykonującego zamianę może odbywać się praca innego programu.

7. Pamięć wirtualna maszyny ATLAS

ATLAS ma rdzeniową pamięć operacyjną złożoną z 4 modułów po 4096 słów. Pamięć wirtualna jest realizowana przy pomocy 4 jednostek pamięci bębnowej po 24 576 słów. W pamięci wirtualnej używa się bloków po 512 słów.Słowo rozkazowe ma 24 bity adresu, w tym 1 bit wskazujący, czy jest to adres wirtualny /pamięć stała, pomocnicza i adresowalne rejestry wejścia/wyjścia są poza systemem pamięci wirtualnej/, i 3 bity pozwalające adresować 6-bitowe znaki wewnątrz słowa. Adres wirtualny zawiera 11 bitów wirtualnego adresu bloku i 9 bitów adresu wewnątrz bloku.

Istnieją 32 strony pamięci operacyjnej i tyleż pozycji tabeli odwzorowań, zbudowanej z rejestrów adresu wirtualnego. Jeśli porównanie badanego wirtualnego adresu bloku z pozycjami tabeli nie daje wskazania równoważności, co oznacza brak potrzebnego bloku w pamięci operacyjnej, następuje przerwanie programu i przejście do podprogramu zamiany bloków, a właściwie wciągnięcia bloku. Wciągnięcie odbywa się bowiem do strony opróżnionej bezpośrednio po poprzedniej czynności wciągnięcia.

Przy zapisie bloku informacji na bęben wykorzystuje się pierwsze napotkane puste pole /miejsce na blok/. W pamięci pomocniczej znajduje się sta-

le modyfikowany spis adresów, podający aktualny adres bębnowy żądanego bloku. W czasie wciągania bloku część centralna wykonuje inny program.

Każdej stronie pamięci operacyjnej przyporządkowany jest klucz,który w stanie "0" dopuszcza każdy dostęp do strony, a w stanie "1" pozwala tylko na dostęp spowodowany zgłoszeniem z bębna /także z taśmy i sterowania przerwaniowego/. Klucz ten chroni przed dostępem do przepisywanego właśnie bloku, a także przed kolizją programów głównych z przesyłaniem informacji związanym z urządzeniami peryferyjnymi. Przesyłanie to jest prowadzone przez sterowanie przerwaniowe.

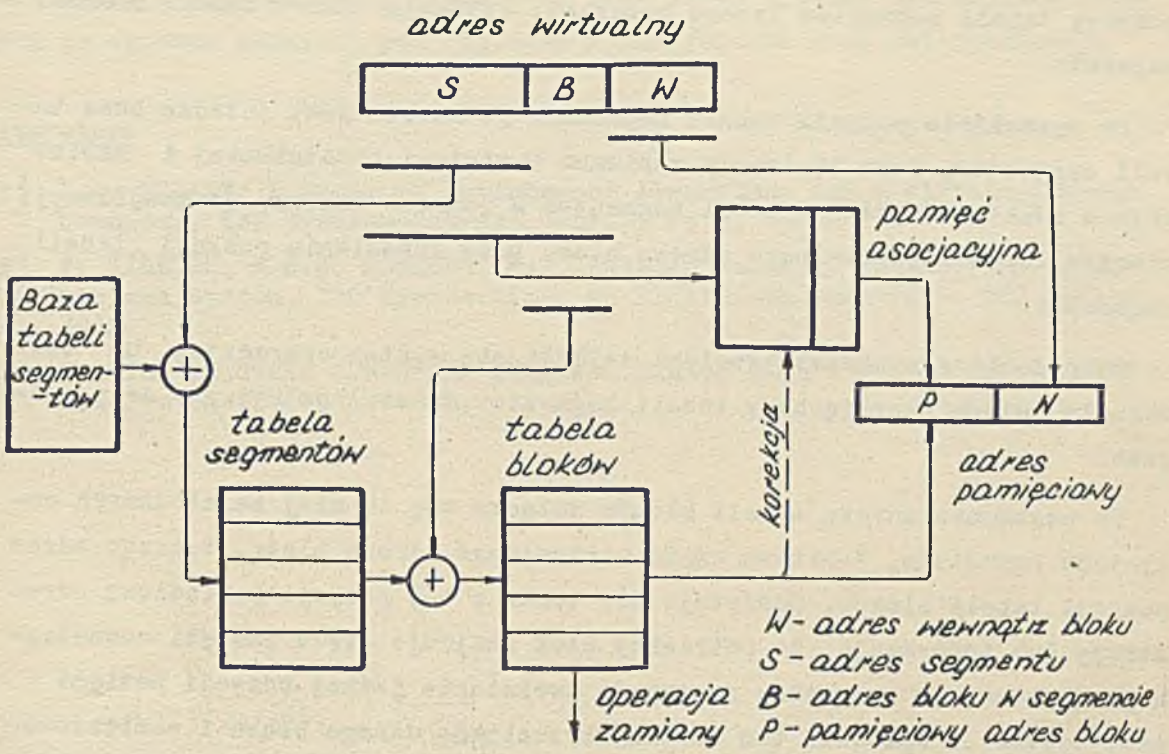
W maszynie ATLAS stosuje się omówiony już algorytm ATLAS.

8. Pamięć wirtualna maszyny 4-75 English Electric Comp.

System 4-75 jest przedstawicielem rodziny System 4 /modele 4-50, 4-70, 4-75 i inne/ o pełnej wymienności programów. Jest to system wieloprogramowy o pamięci wirtualnej i bogatym wyposażeniu peryferyjnym.

Część centralna tego systemu posiada jednostkę sterowania programem, pamięć główną /operacyjną/, notatnikową /s c r a t c h - p a d s t o r e/, arytmometr i jednostkę stronicującą. Pamięć operacyjna zawiera od 2^{16} do 2^{20} znaków /b y t e/ po 8 bitów. Grupa 512 znaków posiada 6-bitową etykietę dla celów ochrony pamięci i gromadzenia informacji o przebiegu używania bloków. Bloki wirtualne mają po 4096 znaków. Znaki tworzą półsłowa, słowa i słowa podwójne. Z 72 słów pamięci notatnikowej 24 użyte są na rejestry dostępne programiście, pozostałe służą celom organizacyjnym. Jednostka stronicująca służy do podziału pamięci operacyjnej na strony.

Początkowy adres każdego programu zostaje automatycznie przesunięty tak, by stanowił wielokrotność liczby 4096. 8 pozycji tabeli odwzorowań mieści się w bardzo szybkiej pomocniczej pamięci asocjacyjnej, a reszta znajduje się w pamięci rdzeniowej. W pamięci asocjacyjnej znajdują się zawsze pozycje odpowiadające ośmiu aktualnie najczęściej używanym blokom spośród wszystkich obecnych w pamięci operacyjnej. Między pamięcią asocjacyjną a rdzeniową prowadzone są zamiany pozycji tabeli odwzorowań według pomocniczego algorytmu, natomiast zamiany bloków prowadzone są przez system operacyjny według pewnego algorytmu należącego do klasy II.



Rys. 1. Adresowanie wirtualne w Systemie 4-75

W Systemie 4-75 prowadzi się segmentację programów. Każdy program budowany jest z segmentów, których liczba może dochodzić do 256. Segment zawiera do 16 bloków. Każdy program posiada w pamięci operacyjnej swoją tabelę segmentów. Każdej pozycji tej tabeli odpowiada jeden segment i adres tabeli bloków. Tabela bloków znajduje się w pamięci operacyjnej, stanowi fragment ogólnej tabeli odwzorowań i posiada 16 pozycji, w których znajdują się pamięciowe adresy stron zawierających bloki wchodzące w skład danego segmentu /w skrócie: pamięciowe adresy bloków/.

Jeśli blok znajduje się w pamięci zewnętrznej, wówczas w pozycji tabeli bloków znajduje się informacja o tym fakcie. Jeżeli następuje odwołanie się do pamięci wirtualnej i okazuje się, że odpowiednia pozycja tabeli odwzorowań /zawierająca adres równy podanemu właśnie wirtualnemu adresowi bloku/ znajduje się w pamięci asocjacyjnej, wówczas pamięciowy adres strony i słowa określa się bezpośrednio i następuje dostęp do pamięci operacyjnej. Jeśli pozycja ta jest w pamięci asocjacyjnej nieobecna, następuje podział wirtualnego adresu bloku. Starsza, 8-bitowa część zostaje wyko-

rzystana do określenia segmentu programu. Przy jej pomocy odszukuje się pozycję tabeli segmentów danego programu, uzyskując adres tabeli bloków segmentu.

Do wyszukania pozycji tabeli segmentów potrzebna jest jeszcze baza tabeli segmentów. Jest to tabela zapisana w pamięci notatnikowej i zawierająca adresy wszystkich tabel segmentów a ponadto używana do modyfikacji starszej części wirtualnego adresu bloku przy określaniu pozycji tabeli segmentów.

Gospodarką wszystkimi tabelami zajmuje się system operacyjny. On też określa, którą pozycją bazy tabeli segmentów ma się posługiwać dany program.

Po uzyskaniu adresu tabeli bloków dołącza się do niej na młodszych pozycjach pozostałą, 4-bitową część wirtualnego adresu bloku, tworząc adres pozycji tabeli bloków. Odczytuje się teraz z tej pozycji pamięciowy adres strony lub informację, że potrzebny blok znajduje się w pamięci zewnętrznej. W pierwszym przypadku następuje zwolnienie jednej pozycji pamięci asocjacyjnej i wpisanie tam adresu wirtualnego danego bloku i pamięciowego adresu strony. W drugim przypadku przechodzi się do podprogramu zamiany bloków, korekcji tabeli bloków, a następnie do korekcji pamięci asocjacyjnej, jak w przypadku pierwszym. Na zakończenie realizuje się dostęp do pamięci operacyjnej.

Pamięć operacyjna Systemu 4-75 pracuje z jedną stroną pustą. Każdy program obecny w pamięci wirtualnej posiada swoją tabelę segmentów, a jeśli jakiś segment jest używany przez więcej niż jeden program, to posiada swoje odpowiednie pozycje w wielu tabelach.

Segmentacja ułatwia budowę programów przeznaczonych do pracy równoczesnej i równoczesne używanie przez nie części wspólnych /kompilatory, podprogramy, pola danych/ i ogromnie przyspiesza przeglądanie tabeli odwzorowań. Koncepcja tabeli odwzorowań jest tu nieco inna, niż przedstawiona na początku niniejszego opracowania. Tabela zawiera tyle pozycji, ile bloków mieści się w pamięci zewnętrznej.

9. Zakończenie

Idea pamięci wirtualnej oraz systemy stronicowania i segmentacji znajdują obecnie coraz szersze zastosowanie w cyfronice. Przeważająca część

związanych z tym czynności wykonywana jest przez systemy operacyjne. Przynosi to ogromne korzyści pod względem wykorzystania mocy obliczeniowej.

Literatura

[1] L.A. BELADY: A Study of Replacement Algorithms for a Virtual-storage Computer, IBM Systems Journal, vol. 5 No 2, 1966.

[2] T. KILBURN, D.B.G. EDWARDS, M.J. LANIGAN and F.H. SUMNER: One-Level Storage System. IRE Transactions on Electronic Computers, vol. EC-11, 1962.

[3] ENGLISH ELECTRIC COMPUTERS LIMITED: System 4-75 a Member of the Micro-integrated System 4 Family. DP 241/2.

Mgr inż. ZOFIA KLAUZNICER
I Pracownia Pamięci Ferrytowych
Zakład Techniki Cyfrowej

681.327.66

ANALIZA KOSZTÓW FERRYTOWYCH PAMIĘCI OPERACYJNYCH W SYSTEMACH 2,5D I 3D

1. Wstęp

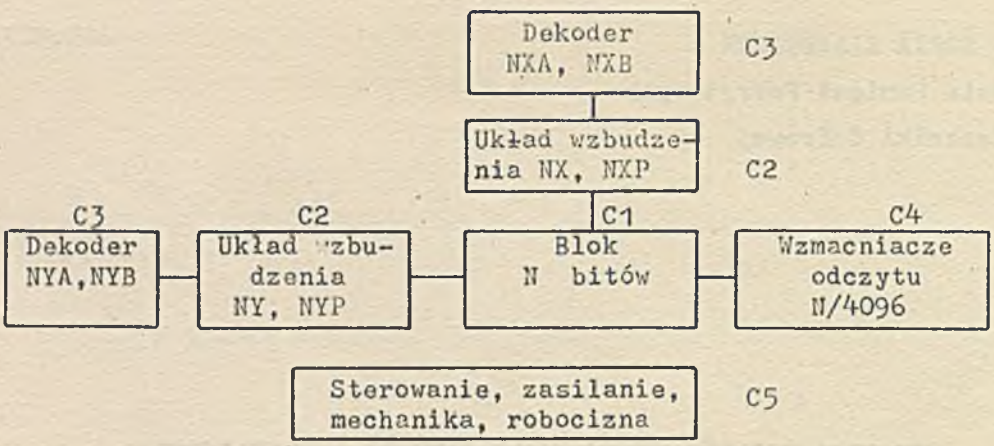
Celem niniejszego artykułu jest analiza kosztów ferrytowych pamięci operacyjnych, realizowanych w systemach 2,5 D oraz 3 D w dużym zakresie ich pojemności. Warto zaznaczyć na wstępie, że z uwagi na różnorodność rozwiązań układów pamięciowych istnieje niebezpieczeństwo uzyskania wyników zarówno nierealnych - w przypadku przyjęcia zbyt dużych uogólnień, jak i trywialnych - przy oparciu się o konkretne rozwiązania.

W rozważaniach przyjęto pewne uproszczone, uogólnione modele pamięci obu systemów, uwzględniające zasadnicze podzespoły podlegające zmianom ilościowym zależnie od pojemności pamięci. Ceny jednostkowe tych podzespołów, jak i całych pamięci w funkcji ilości słów i bitów w słowie wyznaczono przy pomocy maszyny cyfrowej ZAM 41. Artykuł w kolejnych rozdziałach podaje opisy i rysunki przyjętych modeli, omawia program obliczeń napisany w ZAM ALGOL'u, analizuje uzyskane wyniki oraz ocenia zalety i wady przyjętej metody obliczeń.

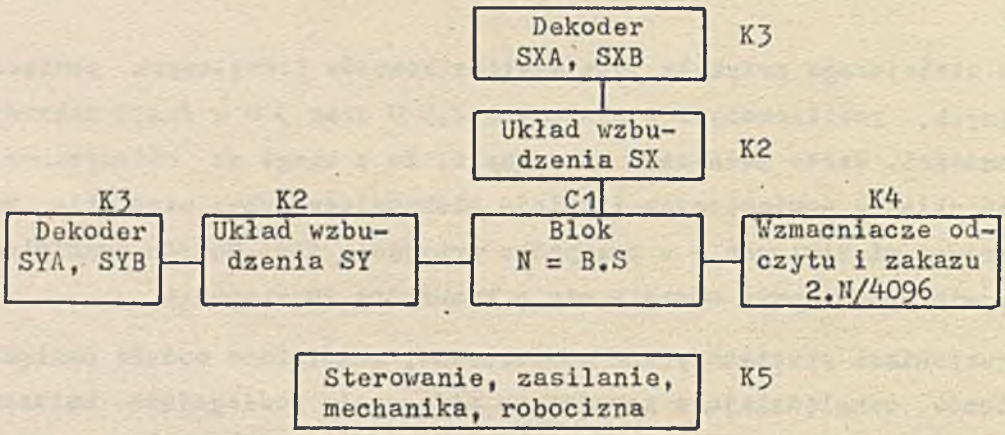
Zasadnicze dane dotyczące systemów ferrytowych pamięci operacyjnych znajdzie czytelnik w pozycjach literatury [1] i [2].

2. Modele pamięci przyjęte do obliczeń

Modele pamięci operacyjnych systemów 2,5 D oraz 3 D przyjęte do obliczeń zostały przedstawione na rysunkach 1 i 2. Są to schematy blokowe uwzględniające zasadnicze podzespoły, jakie zawsze występują w układach pamięciowych, choć mogą się znacznie różnić sposobem realizacji, a więc dekodery, układy wzbudzenia, bloki nośników pamięci, wzmacniacze odczytu, a dla systemu 3D także wzmacniacze zakazu.



Rys. 1. Schemat blokowy dla systemu 2,5 D



Rys. 2. Schemat blokowy dla systemu 3 D

Elementom każdego z tych podzespołów przypisano symbole będące w programie synonimami ilości tych elementów oraz symbole ich cen jednostkowych. Nieco inaczej potraktowano układy sterowania i zasilania, urządzenia mechaniczne oraz udział kosztów robocizny w kosztach całości, przypisując im łącznie pewien współczynnik cen będący w trakcie obliczeń mnożnikiem cen układów zasadniczych. Poniżej przedstawiono wykaz oznaczeń, stosowanych na schematach i w programie dla obu systemów. Przyjęte symbole bardziej odpowiadają specyfice ZAM ALGOL'u - związanej z możliwościami drukarki, niż zapisowi powszechnie stosowanemu, gdyż modele systemów opracowano z myślą o układanym programie.

2,5 D

S - ilość słów w pamięci,
B - ilość bitów w słowie,
N - pojemność pamięci = ilość rdzeni = S.B,
NX - ilość wejść słów /multi-słów/ w bloku nośnika informacji,
NY - ilość wejść bitowych w bloku nośnika informacji,
NXP - ilość elementów wzbudzenia słów /=NX/2, z uwagi na ich podwójne wykorzystanie/,
NYP - iloraz ilości wejść bitowych przez ilość bitów w słowie /NY/B/ - potrzebny przy obliczaniu elementów dekodera,
NXA, NXB - ilość elementów we współrzędnych matrycy dekodera słów,
NYA, NYB - ilość elementów we współrzędnych matrycy dekodera bitów,
C1 - cena rdzenia,
C2 - cena elementu wzbudzenia,
C3 - cena elementu dekodera,
C4 - cena wzmacniacza odczytu,
C5 - współczynnik uwzględniający koszty sterowania, zasilania, urządzeń mechanicznych i robocizny.

3 D

S, B, N, C1 - jak dla 2,5 D,
SX - ilość elementów wzbudzenia współrzędnej adresowej X,
SY - ilość elementów wzbudzenia współrzędnej adresowej Y,
SXA, SXB - ilość elementów we współrzędnych matrycy dekodującej wejścia adresowe X,
SYA, SYB - ilość elementów we współrzędnych matrycy dekodującej wejścia adresowe Y,
K2 - koszt elementu wzbudzenia,
K3 - koszt elementu dekodera,
K4 - koszt wzmacniacza odczytu, a zarazem wzmacniacza zakazu,
K5 - współczynnik uwzględniający koszty sterowania, zasilania, urządzeń mechanicznych i robocizny.

3. Szkic programu liczącego koszty pamięci

Program ułożony jest w taki sposób, że kolejno dla różnych pojemności pamięci obu systemów wyznaczane są ceny zasadniczych podzespołów; na podstawie podanych liczb słów i bitów w słowie obliczana jest liczba elemen-

tów wchodzących w skład poszczególnych podzespołów,a następnie mnożona przez podane ceny jednostkowe. Koszty podzespołów sumowane są, a następnie mnożone przez współczynnik uwzględniający koszty sterowania, zasilania, mechaniki i robocizny, dzięki czemu uzyskuje się koszt całkowity pamięci. Dzieląc go przez ilość bitów w pamięci otrzymuje się "koszt na bit", często stosowany we wszelkich zestawieniach.

Część wydawnicza programu powoduje drukowanie wyników pośrednich i końcowych w formie tabel. Kolejne pozycje to: liczba słów i liczba bitów w słowie oraz koszt rdzeni, koszt wzbudzenia, koszt elementów odczytu i zapisu lub odczytu i zakazu, koszt całkowity i koszt na bit w złotych dla wszystkich wariantów obu systemów.

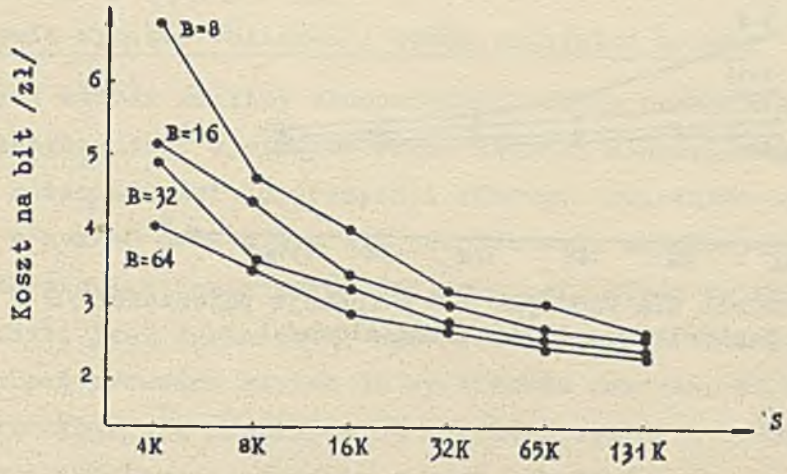
Wyznaczenie ilości elementów poszczególnych podzespołów sprowadza się do tworzenia matryc kwadratowych lub matryc o liczbach elementów współrzędnych maksymalnie zbliżonych do kwadratu, obliczanych każdorazowo tym samym fragmentem programu, zwanym procedurą. W ten sposób dla systemu 2,5 D z ilości rdzeni N wyznacza się ilość wejść słów NX i ilość wejść bitowych NY, z NXP otrzymuje się NXA i NXB, a z NYP-NYA i NYB, co daje optymalny rozkład rdzeni w ramce oraz liczbę elementów wybierania i wzbudzenia. Analogicznie dla systemu 3 D z ilości słów oblicza się SX i SY, a z nich SXA, SXB oraz SYA, SYB.

Konstrukcja programu jest prosta. Początkowo deklaruje się występujące w obliczeniach wielkości, a także wspomnianą wyżej procedurę, a następnie w dwu częściach program prowadzi do obliczenia i druku poszczególnych kosztów dla każdej kombinacji pary S i B kolejno dla obu systemów.

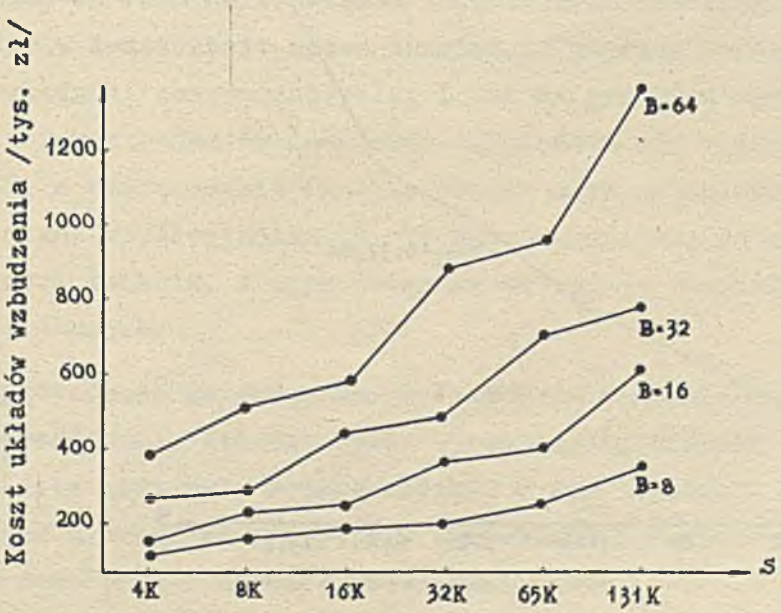
Do niniejszego opracowania dołączono program i uzyskane zestawienie kosztów pamięci. Ciekawsze wyniki, tj. koszt elementów wzbudzenia i koszt na bit w funkcji pojemności, przedstawiono na wykresach /rys. 3-6/.

Dla dołączonych tu wyników przyjęto pewien ciąg liczb słów w pamięci od 4 do 130 tysięcy /w zaokrągleniu/, przy czym każda następna liczba stanowi podwojenie poprzedniej. Przyjęte ilości bitów w słowie to: 8, 16, 32 i 64.

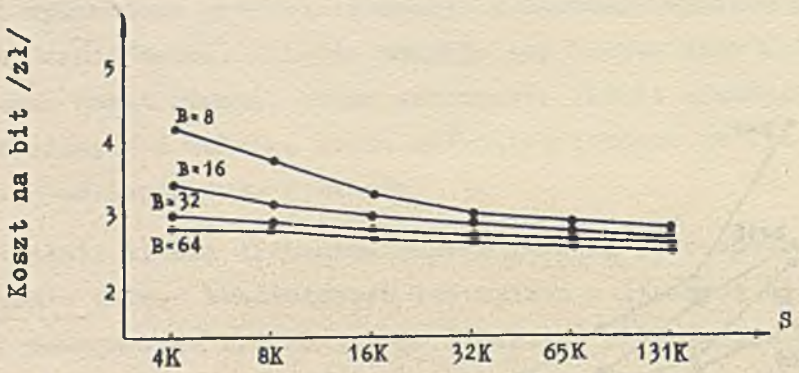
Koszty poszczególnych elementów przyjęto na podstawie danych dla PAO-5 /pamięć operacyjna z wybieraniem liniowym dla EMC ZAM-41/ z zestawień cen pakietów i podzespołów udostępnionych przez Dział Ekonomiczny ZDINM w grudniu 1966 r., wprowadzając nieznaczne zmiany lub zaokrąglenia w oparciu o istniejące lub prawdopodobne tendencje.



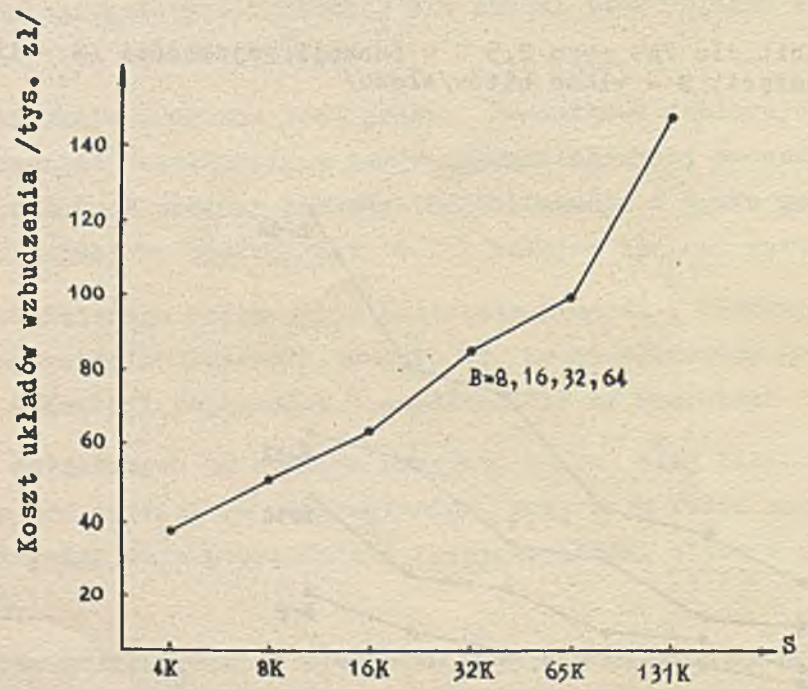
Rys. 3. Koszt na bit dla PAO typu 2,5 D w funkcji pojemności /S - ilość słów w pamięci; B - ilość bitów/słowo/



Rys. 4. Koszt układów wzbudzenia dla PAO 2,5 D w funkcji pojemności /S - ilość słów w pamięci; B - ilość bitów/słowo/



Rys. 5. Koszt na bit dla PAO typu 3 D w funkcji pojemności /S - ilość słów w pamięci; B - ilość bitów/słowo/



Rys. 6. Koszt układów wzbudzenia dla PAO typu 3 D w funkcji pojemności /S - ilość słów w pamięci; B - ilość bitów/słowo/

Poniżej wymienione są w złotych ceny przyjęte do obliczeń, zgodnie z podanymi uprzednio oznaczeniami:

C1 = 1,4

C2 = 60 K2 = 60

C3 = 700 K3 = 700

C4 = 1100 K4 = 1100

Współczynnik C5 = 1,3, a K5 = 1,35.

4. Omówienie wyników obliczeń i ocena przyjętej metody

Przyjęta metoda analizy ekonomicznej jest w pewnym stopniu jednostronna. Nie uwzględniono w niej na razie kwestii niezawodności, szybkości i wielkiej różnorodności logicznych i ideowych rozwiązań układów pamięciowych. Na przykład nasuwający się na podstawie uzyskanych wyników wniosek maksymalnego zwiększania pojemności pamięci, kiedy to osiąga się minimalny koszt na bit, jest ograniczony wymogiem niezawodności. Powiększanie pojemności pamięci prowadzi bowiem do wydłużania uzwojeń, co powoduje, że czasy propagacji stają się porównywalne z czasem cyklu, a rosnące sprzężenia pojemnościowo-indukcyjne powodują wzrost zakłóceń.

Tym niemniej przyjęte uogólnienie zapewnia dużą elastyczność w stosowaniu programu do różnych rozwiązań technicznych poszczególnych podzespołów, gdyż zmiany w konstrukcji można uwzględnić poprzez wprowadzenie do obliczeń odpowiedniej ceny podzespołu. I tak na przykład występowanie w pamięciach różnych rodzajów dekoderów uwzględnia się w odpowiednich cenach C2, czy K2, a zastosowanie dwu lub trzech uzwojeń na rdzeń w pamięci 2,5 D poprzez zmianę współczynnika C5. To samo odnosi się do przypadku stosowania szybszych technik, z czym związane są wyższe ceny rdzeni i elementów półprzewodnikowych.

Dane wprowadzone do programu, zwłaszcza w zakresie cen, są dość elastyczne, ograniczenie stanowi tylko - zrozumiały zresztą - warunek, aby kolejne liczby słów były potęgą dwójki. W tym wypadku dotyczy to także liczby bitów z uwagi na przyjęty w systemie 2,5 D sposób liczenia ilości elementów dekoderów i układów wzbudzenia bitów.

Wyniki uzyskane w oparciu o realne opracowanie w IMM /na podstawie aktualnie znajdujących się w dyspozycji układów podstawowych i rdzeni/ pozwalają się zorientować, a także wysunąć bardziej ogólne wnioski co do tego, jak kształtowałyby się ceny układów pamięciowych w wypadku zwiększania ich pojemności.

Podsumowując: koszt pamięci rośnie wraz z pojemnością, ale koszt na bit maleje, przy czym wyraźniej widać to w systemie 2,5 D, który jest szczególnie korzystny dla dużych pojemności. Dla mniejszych pojemności natomiast bardziej opłacalne jest stosowanie systemu 3D. Ilustrują to wyraźnie wykresy zależności kosztu na bit od ilości słów w pamięci dla różnych ilości bitów w słowie /rys. 3 i 5/. Pokazują one ponadto, że im większa ilość bitów w słowie, tym bardziej płasko przebiega krzywa kosztów na bit, osiągając mniejsze wartości. Oznacza to szczególnie przy małych pojemnościach pamięci, lepsze wykorzystanie sprzętu elektronicznego. Koszt elementów wzbudzenia w funkcji ilości słów i bitów /rys. 4 i 6/ rośnie nierównomiernie i dla konkretnych zastosowań ważne jest odpowiednie wyważenie S i B. Przy dużych pojemnościach pamięci 2,5 D koszt elementów wzbudzenia jest znaczny, a mimo to opłacalny, gdyż wypadkowy koszt na bit dla takich właśnie pamięci jest najmniejszy. Pokrywanie się krzywych zależności kosztów wzbudzenia od ilości słów dla różnej ilości bitów w systemie 3 D jest oczywiste i wyjaśnia zmniejszanie się kosztów na bit przy wzroście ilości bitów.

Uzyskane wyniki potwierdzają ogólne tendencje stosowania obu typów pamięci, tj. systemu 3 D dla mniejszych, a 2,5 D dla większych pojemności oraz wykazują opłacalność budowy dużych pamięci, na które zapotrzebowanie wyraźnie rośnie.

Proponowany program może być wykorzystany także dla innych kombinacji danych, a także stosunkowo prosto wzbogacony, szczególnie przy dokładniej opisanych rozwiązaniach pamięci, o obliczenie niezawodności katastroficznej. Pozwoliłoby to z kolei uzyskać dane co do optymalnej pojemności pamięci, zależnie od łącznych wymogów zarówno kosztów, jak i niezawodności.

Literatura

[1] Cyfrowe zastosowania układów magnetycznych. Praca zbiorowa pod red. A.J. Meyerhoffa, Warszawa 1964.

[2] J. RYŻKO: System 2,5 D na tle dotychczasowego rozwoju pamięci ferrytowych. Nowości Techniczne, 3/1967.

[3] T.J. GILLIGAN: 2,5 D High-Speed Memory Systems - Past, Present and Future. IEEE Trans. on El. Comp., vol. EC-15, No 4, August 1966.

[4] A. ELOVIC: Wiring Design Helps Core Memory Work at Rapid Cycle Time. Electronics, Oct. 31, 1966.

[5] L.C. HOBBS: Present and Future-state-of-the-Art in Computer Memories. IEEE Trans. on EC. Comp., vol. EC-15, No 4, August 1966.

```
PROG: Z.KL. TC2
ALG:


BEGIN

COMMENT OBLICZENIE KOSZTOW PAMIECI 2,5D I 3D ;
REAL N,S,B,C1,C2,C3,C4,C5,K2,K3,K4,K5,S1,S2,S3,S4,S5,S6,B1,
     B2,B3,B4;
COMMENT UZYSKANIE ROZKLADU ELEMENTOW W PLASZCZYZNIE ZBLIZONEJ
DO KWADRATU;
PROCEDURE KW (K,AX,AY);
VALUE K;
REAL K,AX,AY;

BEGIN
REAL A,M ; A:=K; M:=SQRT(K) ;
IF ENTIER (M) EQUAL M THEN GOTO G ELSE
F:A:=A/2 ;
IF A GREATER M THEN GOTO F ELSE
AX:=A ; AY:=K/AX ;
GOTO D;
G: AX:=M ; AY:=M ;
D:END;
INP (1,C1,C2,C3,C4,C5,K2,K3,K4,K5 );
INP (1,S1,S2,S3,S4,S5,S6,B1,B2,B3,B4 );

BEGIN

COMMENT KOSZTY PAMIECI 2,5 D ;
REAL NX,NXP,NY,NYP,NXA,NXB,NYA,NYB,KRD,KWB,KOZ,KCK,KNB ;
OUT (1,'T','ZESTAWIENIE :B KOSZTOW :B PAMIECI :B 2,5 D :/:/:/');
OUT (1,'T','IL.SLOW:B IL.BT:B:B KST.RDZ.:B:B:B KST.WZB.:B:B:B
KST.O.Z.:B:B:B KST.CALK.:B KST.BT:/:/');
FOR S:=S1,S2,S3,S4,S5,S6 DO
FOR B:=B1,B2,B3,B4 DO

BEGIN
N:=S*B;
KW (N,NX,NY);
NXP:=NX/2;  NYP:=NY/B;
KW (NXP,NXA,NXB );
KW (NYP,NYA,NYB );

KRD:=N*C1 ;
KWB:=2*(NXP+NY)*C2+(NXA+NXB+B*(NYA+NYB ))*C3;
KOZ:= N/4096*C4;
KCK:= (KRD+KWB+KOZ)*C5;
KNB:=KCK/N;
COMMENT DRUKOWANIE WYNIKOW;
OUT (1,'Z3DB3D',S);
OUT (1,'Z2B2D',B);
OUT (1,'ZB2DB3DB3D',KRD,KWB,KOZ);
OUT (1,'ZB3DB3DB3D',KCK);
OUT (1,'ZB2D.2D',KNB);
OUT (1,'T',':/:/');
END
END;
```

```
BEGIN
COMMENT KOSZTY PAMIECI 3D;
REAL SX,SY,SXA,SXB,SYA,SYB,KRD,KWB,KOZ,KCK,KNB ;
OUT (1,'T','=* ZESTAWIENIE :B KOSZTOW :B PAMIECI:B  3D :/:/:/');
OUT (1,'T','IL.SLOW:B IL.BT:B:B KST.RDZ.:B:B:B KST.WZB.:B:B:B
KST.O.Z.:B:B:B KST.CALK.:B KST.BT:/:/');
 FOR S:=S1,S2,S3,S4,S5,S6 DO
 FOR B:=B1,B2,B3,B4 DO

BEGIN
N:=S*B;
KW (S,SX,SY) ;
KW (SX,SXA,SXB);
KW (SY,SYA,SYB) ;
KRD:=N*C1;
KWB:=2*(SX+SY)*K2+(SXA+SXB+SYA+SYB)*K3;
KOZ:=2*N/4096*K4;
KCK:=(KRD+KWB+KOZ)*K5;
KNB:=KCK/N;
COMMENT DRUKOWANIE WYNIKOW;
OUT (1,'Z3DB3D',S);
OUT (1,'Z2B2D',B);
OUT (1,'ZB2DB3DB3D',KRD,KWB,KOZ);
OUT (1,'ZB3DB3DB3D',KCK);
OUT (1,'ZB2D.2D',KNB);
OUT (1,'T',':/:/');
END
END

END
```

```
DANE:
1.4 ; 60 ; 700 ; 1100 ; 1.3 ; 60 ; 700 ; 1100 ; 1.35 ;
4096 ; 8192 ; 16384 ; 32768 ; 65536 ; 131072 ; 8 ; 16 ; 32 ; 64 ;
```

ZESTAWIENIE KOSZTÓW PAMIECI 2,5 D.

IL.SŁÓW	IL.BT	KST.RDZ.	KST.WZB.	KST.O.Z.	KST.CAŁK.	KST.BT
4 096	8	45 875	116 800	8 800	222 918	6.80
4 096	16	91 750	152 480	17 600	340 380	5.19
4 096	32	183 501	272 800	35 200	638 951	4.87
4 096	64	367 002	383 360	70 400	1 066 990	4.07
8 192	8	91 750	130 080	17 600	311 260	4.75
8 192	16	183 501	228 000	35 200	580 711	4.43
8 192	32	367 002	293 760	70 400	950 510	3.63
8 192	64	734 003	534 400	140 800	1 831 964	3.49
16 384	8	183 501	183 200	35 200	522 471	3.99
16 384	16	367 002	248 960	70 400	892 270	3.40
16 384	32	734 003	444 800	140 800	1 715 484	3.27
16 384	64	1 468 006	576 320	281 600	3 023 704	2.88
32 768	8	367 002	204 160	70 400	834 030	3.18
32 768	16	734 003	355 200	140 800	1 599 004	3.05
32 768	32	1 468 006	486 720	281 600	2 907 224	2.77
32 768	64	2 936 013	878 400	563 200	5 690 897	2.71
65 536	8	734 003	310 400	140 800	1 540 764	2.94
65 536	16	1 468 006	397 120	281 600	2 790 744	2.66
65 536	32	2 936 013	699 200	563 200	5 457 937	2.60
65 536	64	5 872 026	951 040	1 126 400	10 334 305	2.46
131 072	8	1 468 006	352 320	281 600	2 732 504	2.61
131 072	16	2 936 013	609 600	563 200	5 341 457	2.55
131 072	32	5 872 026	771 840	1 126 400	10 101 345	2.41
131 072	64	11 744 051	1 376 000	2 252 800	19 984 707	2.38

ZESTAWIENIE KOSZTÓW PAMIECI 3 D.

IL.SŁÓW	IL.BT	KST.RDZ.	KST.WZB.	KST.O.Z.	KST.CAŁK.	KST.BT
4 096	8	45 875	37 760	17 600	136 668	4.17
4 096	16	91 750	37 760	35 200	222 359	3.39
4 096	32	183 501	37 760	70 400	393 742	3.00
4 096	64	367 002	37 760	140 800	736 508	2.81
8 192	8	91 750	51 040	135 200	240 287	3.67
8 192	16	183 501	51 040	70 400	411 670	3.14
8 192	32	367 002	51 040	140 800	754 436	2.88
8 192	64	734 003	51 040	281 600	1 439 968	2.75
16 384	8	183 501	64 320	70 400	429 598	3.28
16 384	16	367 002	64 320	140 800	772 364	2.95
16 384	32	734 003	64 320	281 600	1 457 896	2.78
16 384	64	1 468 006	64 320	563 200	2 828 961	2.70
32 768	8	367 002	85 280	140 800	800 660	3.05
32 768	16	734 003	85 280	281 600	1 486 192	2.83
32 768	32	1 468 006	85 280	563 200	2 857 257	2.72
32 768	64	2 936 013	85 280	1 126 400	5 599 385	2.67
65 636	8	734 003	106 240	281 600	1 514 488	2.89
65 636	16	1 468 006	106 240	563 200	2 885 553	2.75
65 636	32	2 936 013	106 240	1 126 400	5 627 681	2.68
65 636	64	5 872 026	106 240	2 252 800	11 111 939	2.65
131 072	8	1 468 006	148 160	563 200	2 942 145	2.81
131 072	16	2 936 013	148 160	1 126 400	5 684 273	2.71
131 072	32	5 872 026	148 160	2 252 800	11 168 531	2.66
131 072	64	11 744 051	148 160	4 505 600	22 137 045	2.64

PRZEWIDYWANA WARTOŚĆ PRODUKCJI MASZYN MATEMATYCZNYCH I URZĄDZEŃ W USA W LATACH 1967, 1968 i 1971 W MILIONACH DOLARÓW

	1967	1968	1971
1. Maszyny cyfrowe /oprócz EMC do sterowania/	2420,0	2855,0	3980,0
2. Maszyny analogowe /oprócz EMA do sterowania/	46,5	60,3	100,5
3. Maszyny hybrydowe /oprócz EMH do sterowania/	42,0	47,0	65,0
4. Maszyny do sterowania procesami przemysłowymi oraz urządzenia	229,0	290,1	406,8
z tego: a/ maszyny cyfrowe	130,0	171,6	234,9
b/ maszyny analogowe	22,0	25,1	41,0
c/ inne urządzenia	77,0	93,4	130,9
5. Urządzenia do sterowania numerycznego	192,5	231,3	386,3
z tego: a/ maszyny cyfrowe	64,0	89,0	134,0
6. Urządzenia zewnętrzne EMC, ogółem	450,3	556,8	844,5
z tego: a/ konwertery a-c i c-a	39,0	43,6	65,5
b/ konwertery karta/taśma	7,1	7,3	3,1
c/ czytniki taśmy papierowej	32,0	33,5	57,0
d/ urządzenia wejścia i wyjścia /drukarki, pisaki, urządzenia wideograficzne/	164,4	221,5	318,6
e/ pamięci magnetyczne, ogółem	207,8	250,9	400,3
z tego: A/ pamięci rdzeniowe	31,3	41,2	39,5
B/ pamięci taśmowe	62,3	67,5	83,7
C/ pamięci bębnowe	17,5	24,0	48,0
D/ pamięci dyskowe	96,7	118,2	229,1
7. Urządzenia transmisji danych	160,1	221,5	550,0
8. Rdzenie ferrytowe dla pamięci	9,0	18,0	28,8
9. Taśmy magnetyczne dla pamięci	66,0	75,5	107,7
RAZEM MASZYNY MATEMATYCZNE I URZĄDZENIA	3615,4	4355,5	6469,6

Opracowano na podstawie: Electronics nr 1/1968

Warunki prenumeraty:

Cena prenumeraty krajowej:

rocznie - zł 100,-

Prenumeraty przyjmowane są do dnia 10 grudnia na rok następny.

Prenumeratę na kraj dla czytelników indywidualnych przyjmują urzęd pocztowe oraz listonosze.

Czytelnicy indywidualni mogą dokonywać wpłat również na konto PKO Nr 1-6-100020 - Centrala Kolportażu Prasy i Wydawnictw "Ruch" Warszawa, ul. Wronia 23.

Wszystkie instytucje państwowe i społeczne mogą zamawiać prenumeratę wyłącznie za pośrednictwem Oddziałów i Delegatur "Ruch".

Prenumeratę ze zleceniem wysyłki za granicę, która jest o 40% droższa od krajowej, przyjmuje Biuro Kolportażu Wydawnictw Zagranicznych "Ruch" Warszawa, ul. Wronia 23 konto PKO Nr 1-6-100024 tel. 20-46-88.

Egzemplarze zdezaktualizowane można nabyć w Punkcie Wysyłkowym Prasy Archiwalnej "Ruch" - Warszawa, ul. Nowomiejska 15/17, konto PKO Nr 114-6-700041 VII O/M Warszawa.