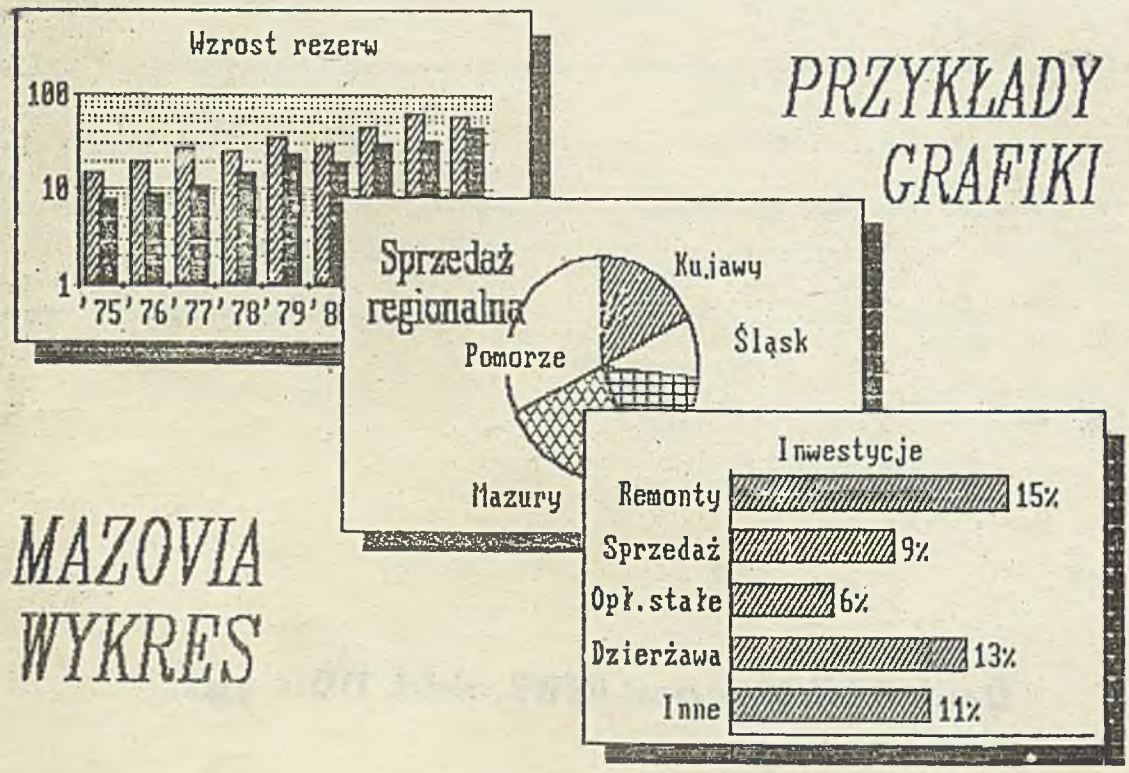
POLITECHNIKA ŚLĄSKA Biblioteka Główna
P.3057/88

techniki komputerowe

5-6 '88

IMM

BIULETYN INFORMACYJNY



PRZYKŁADY GRAFIKI

MAZOVIA WYKRES

INSTYTUT MASZYN MATEMATYCZNYCH
BRANŻOWY OŚRODEK INTE

Rysunek na okładce: zob. opracowanie Danuty PYZEL
"Program WYKRES - grafika użytkowa" s.61

Druk IMM zam. 4/89 nakł. 1100 egz.

TECHNIKI KOMPUTEROWE

P. 3057/88

Rok XXVI Nr 5-6 1988

Spis treści

str.

BOTT Cz., KOPROŃ A., KWIATKOWSKI J., DZIK K., KOSIEL W., WARDA A.: Napędy dysków elastycznych..3

CHOLEWA W., MOCZULSKI W.: Reprezentacja wiedzy w diagnostycznych systemach doradczych.....37

PYZEL D.: Program WYKRES - grafika użytkowa.....61

GRZEGOREK M.: Charakterystyka zintegrowanego systemu komputerowego wspomagania produkcji z użyciem stacji roboczej SIEMENS WS10.....73

NOWOŚCI TECHNICZNE - opracował Jan RYŻKO.....85

DWUMIESIĘCZNIK

Wydaje:

INSTYTUT MASZYN MATEMATYCZNYCH
ul. Krzywickiego 34 02-078 WARSZAWA tel. 28-37-29

BRANŻOWY OŚRODEK INFORMACJI NAUKOWEJ, TECHNICZNEJ I EKONOMICZNEJ

Komitet Redakcyjny:

dr inż. Stanisława BONKOWICZ-SITTAUER (redaktor naczelny), mgr Hanna DROZDOWSKA-STRZEMIŃSKA (sekretarz redakcji), mgr inż. Zdzisław GROCHOWSKI, mgr inż. Jan KLIMOWICZ, dr inż. Piotr PERKOWSKI, mgr inż. Romuald SYNAK

TECHNIKI KOMPUTEROWE 5-6/88

inż. Czesław Bott - ZWM Unitra - Lubartów
mgr inż. Andrzej Koproń - ZWM Unitra - Lubartów
mgr inż. Jerzy Kwiatkowski - ZWM Unitra - Lubartów
mgr inż. Krzysztof Dzik - Instytut Maszyn Matematycznych
mgr inż. Włodzimierz Kosiel - Instytut Maszyn Matematycznych
mgr inż. Andrzej Warda - Instytut Maszyn Matematycznych

Napędy dysków elastycznych

Napędy dysków elastycznych

Wstęp

Napędy dysków elastycznych są najszerzej stosowanymi pamięciami masowymi w mikrokomputerach personalnych. Są na ogół tanie (w porównaniu z napędami dysków stałych), a jednocześnie w porównaniu z pamięciami kasetowymi i taśmą papierową są wygodniejsze w użytkowaniu.

W niniejszym artykule opisana jest budowa i działanie napędów dysków elastycznych, a także przedstawione są ich podstawowe parametry na przykładzie dysków 5.25". W dalszej części analizujemy rozwiązania konstrukcyjne stosowane w napędach produkowanych przez wybrane firmy zagraniczne. Analizie poddano zarówno część mechaniczną napędów dyskowych, jak i ich elektroniczne układy sterujące. Artykuł jest zaadresowany do użytkowników pamięci dyskowych, ale także może być pomocny dla konstruktorów napędów dysków elastycznych i konstruktorów sterowników tych napędów.

Charakterystyka techniczna napędów dysków elastycznych

Przeznaczenie napędów dysków elastycznych

Głównym przeznaczeniem dysków elastycznych jest przechowywanie programów użytkownika, programów narzędziowych (np.: kompilatory, debuggery, edytory, itd.), systemów operacyjnych i plików danych. Na dysku przechowywane są również wyniki pracy komputera.

Komputer jest na ogół wyposażony w niewielką pamięć stałą, w której zapisany jest program ładujący system operacyjny z dyskietki lub dysku stałego oraz niektóre procedury sterujące urządzeniami komputera. Pozostałe procedury sterujące są również ładowane z dysku stałego lub elastycznego, w związku z tym istnieje łatwość modyfikowania i rozszerzania tych procedur, jak i systemu operacyjnego.

Dyski elastyczne mają pewne zalety w porównaniu z dyskami stałymi i w związku z tym są nadal wykorzystywane. Te zalety to:

- możliwość przenoszenia programów z komputera na komputer,
- łatwość dystrybucji oprogramowania na dyskietkach,
- stosowanie dysków elastycznych nie ogranicza pojemności instalowanego oprogramowania i zbiorów roboczych,
- możliwość zabezpieczenia (ukrycia) ważnych zbiorów zapisanych na dyskietkach przed niepożądanym dostępem; nie ma konieczności chronienia całego komputera.

Pierwsze napędy dysków elastycznych wykorzystywały dyskietki o średnicy 8 cali. Napędy te były na ogół duże i nie nadawały się do stosowania w komputerach personalnych. Najczęściej były montowane w innej obudowie niż jednostka centralna i stanowiły autonomiczne urządzenie, stację napędów dysków elastycznych. Obecnie napędy te są rzadko wykorzystywane.

Wraz z rozwojem technologicznym, stało się możliwe zmniejszenie średnicy dyskietek. Spotyka się napędy o następujących średnicach stosowanych dyskietek: 5.25", 3.5", 3" (rzadko) i ostatnio słychać o opracowywaniu napędów o średnicy dyskietek 2".

Innymi cechami charakteryzującymi rozwój napędów dysków elastycznych są:

- zmniejszenie wysokości napędów (typowe wysokości dla napędu 5.25": 80, 40 i 25 mm),
- zmniejszenie pobieranej mocy,
- zwiększenie pojemności dyskietek,
- zwiększenie szybkości działania,
- stosowanie dyskietek w sztywnych obudowach,

- wyposażenie w coraz większą "inteligencję" przez stosowanie wyspecjalizowanych układów VLSI i przejmowanie części funkcji spełnianych dotychczas przez sterownik.

Obecnie najszerzej stosowanymi napędami są napędy 5.25", ze względu na bardzo duże rozpowszechnienie komputerów IBM XT/AT i kompatybilnych.

Standardowa pojemność dyskietki stosowanej w komputerach IBM XT/AT, Mazovii 1016 i innych kompatybilnych wynosi 360 kB, a jej średnica jest równa 5.25". Kolejnym typem napędów dysków elastycznych stosowanym przez firmę IBM jest napęd o pojemności dyskietki 1.2 MB.

W nowej rodzinie mikrokomputerów personalnych o nazwie PS/2, produkowanej przez firmę IBM, zastosowano napędy dysków elastycznych o średnicy 3.5" i pojemności 1.44 MB. Prawdopodobnie będzie to nowy standard dla napędów dysków elastycznych stosowanych w mikrokomputerach personalnych. Napęd ten charakteryzuje się innym interfejsem, wykorzystywaniem sztywnych dyskietek, w sztywnych obudowach i znacznym zaawansowaniem technicznym w porównaniu z napędami 5.25".

Ogólna zasada budowy (schemat blokowy)

W pamięci dyskowej można wyróżnić następujące zasadnicze elementy: korpus, zespół napędowy, zespół głowicy, układ elektroniki i współpracującą z nimi dyskietkę.

Korpus najczęściej jest wykonany jako odlew, choć niekiedy jest też w formie ukształtowanego wykroju z blachy.

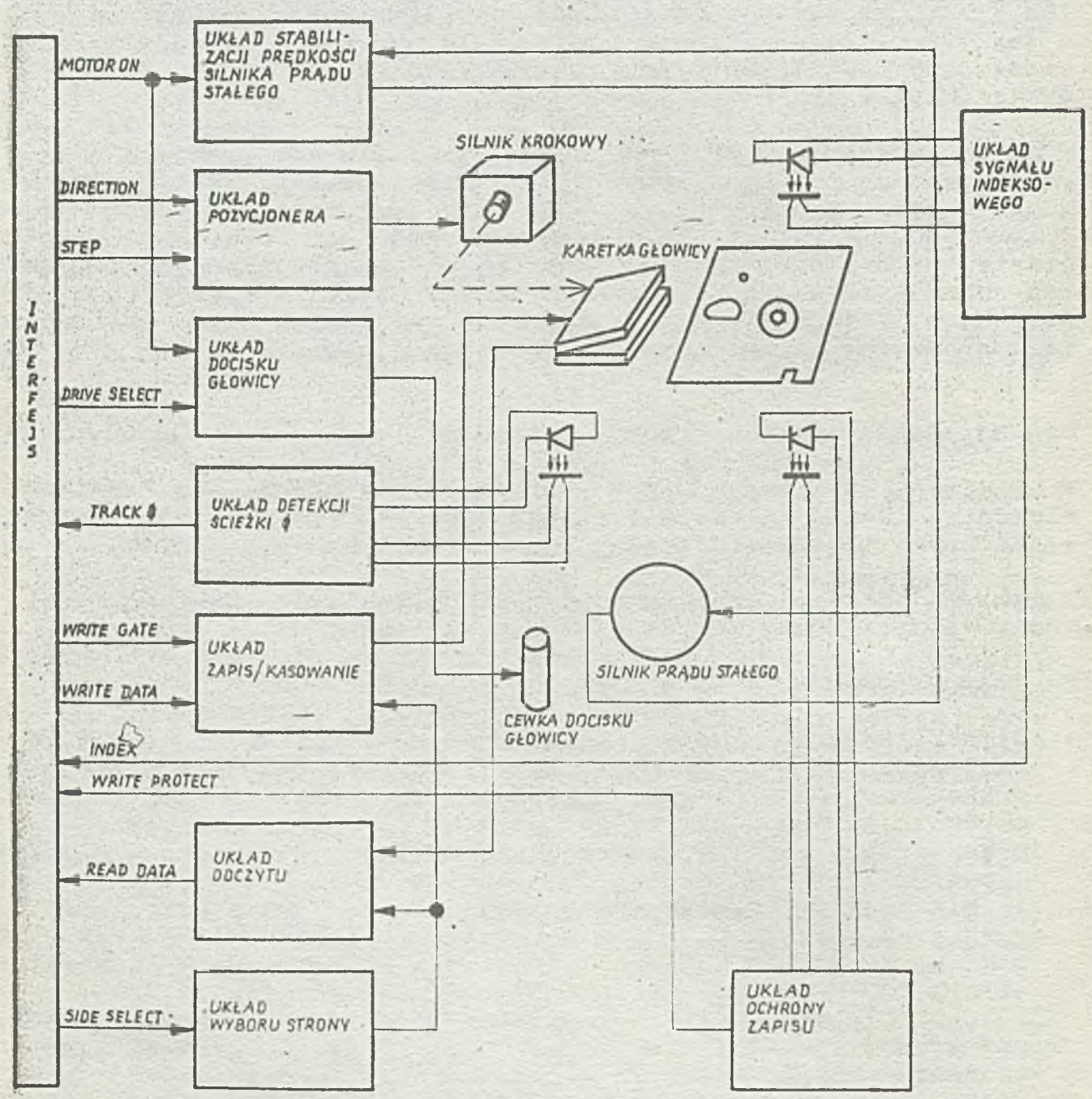
Na *zespół napędowy* składają się:

- silnik napędu stałego służący do napędu dyskietki; najczęściej jest wykonany jako zintegrowany zespół z płytką drukowaną, na której znajdują się układy sterowania prędkością obrotową tego silnika i układ pomiaru tej prędkości, połączone ze sobą w układzie ze sprzężeniem zwrotnym,
- silnik krokowy służący do przesuwania zespołu głowic wzdłuż dyskietki; dla napędu 360 kB jest to silnik 100-krokowy o kącie 3.6° dla każdego kroku, dla napędu 1.2 MB jest to silnik 200-krokowy o kącie 1.8° dla każdego kroku.

Zespół głowic składa się z karetki i dwóch głowic, gdyż z założenia napęd ma być dwustronny. Głowice są tak przymocowane w karetce, że po włożeniu do napędu dyskietki przylegają do obydwu powierzchni dyskietki. Oczywiście w danym momencie może działać tylko jedna głowica, uaktywniana sygnałem wyboru głowicy (SELECT HEAD), co jest uwidocznione na rys. 1.

Najczęściej po włożeniu dyskietki i zamknięciu napędu, głowice są opuszczane na powierzchnie dyskietki, choć niekiedy, zdarza

się, że napęd jest zaopatrzony w elektromagnes dociskający głowicę do dyskietki w odpowiednim momencie. W zależności od konstrukcji napędu sygnałami uaktywniającymi elektromagnes mogą być: sygnał włączenia silnika (**MOTOR ON**), sygnał wyboru napędu (**DRIVE SELECT**) lub sygnał opuszczenia głowicy (**HEAD LOAD**), który nie występuje we wszystkich interfejsach.



Rys. 1. Schemat blokowy napędu dysków elastycznych

Układ elektroniki znajduje się najczęściej na jednej głównej płytce, która skupia podstawowe układy elektroniczne. Poza tą płytką są realizowane układ stabilizacji prędkości obrotowej silnika prądu stałego i zespoły czujników elektronicznych. Napęd dysków elastycznych jest wyposażony standardowo w następujące czujniki:

- czujnik ścieżki zerowej, który wykrywa obecność głowicy na ścieżce zerowej,
- czujnik otworu indeksowego, który generuje impuls, gdy na jego wysokości znajduje się otwór indeksowy dyskietki określający początek ścieżki,
- czujnik blokujący układ zapisu w napędzie dysku elastycznego po wykryciu zabezpieczenia zapisu na dyskietce.

Na płytce elektroniki znajdują się zwory, które pozwalają ustawić niektóre warunki pracy. Najczęściej wybór dotyczy:

- numeru, pod jakim ma być widoczny napęd,
- rodzaju sygnału, który inicjuje ładowanie głowicy,
- sygnału, który uaktywnia wskaźnik na przedniej ściance napędu sygnalizując pracę.

Podstawowe układy elektroniczne i zespoły mechaniczne przedstawiono na rys. 1.

Budowa dyskietki

Opisywane napędy dysków elastycznych wykorzystują dyskietki elastyczne o rozmiarach 5.25 cala (133,4 mm). Zasadniczą częścią dyskietki jest krążek z tworzywa sztucznego (mylaru), pokryty nośnikiem magnetycznym i zapakowany w kopertę ochronną.

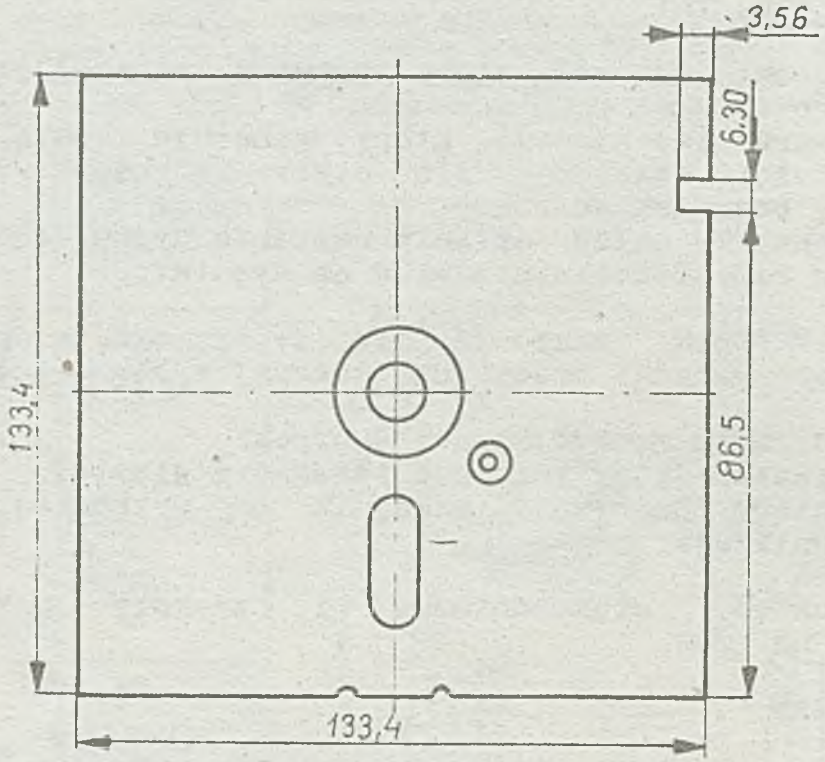
W kopercie są wykonane następujące otwory:

- centralny otwór, służący do napędzania dyskietki,
- otwór umożliwiający dostęp głowicy do nośnika magnetycznego,
- otwór indeksowy, służący do wykrywania otworu indeksowego w nośniku magnetycznym, określającego początek ścieżki,
- otwór ochrony przed zapisem; zalepienie tego otworu uniemożliwia dokonanie zapisu na dyskietce.

Rys. 2. przedstawia wygląd zewnętrzny dyskietki wraz z rozmieszczeniem otworów i ważniejszymi wymiarami. Dokładne wymiarowanie dyskietki zawiera norma BN 87/3104.

W napędach dysków elastycznych istnieje możliwość wykorzystywania zarówno dyskietek z formatem miękkim (naniesionym programowo), jak i twardym (naznaczonym mechanicznie). Dyskietka z formatem miękkim ma tylko jeden otwór indeksowy, określający początek ścieżki, natomiast rozmieszczenie sektorów na ścieżce jest warunkowane wstępnym zapisem magnetycznym na ścieżce podczas procesu formatowania dyskietki.

Dyskietka z formatem twardym ma tyle otworów indeksowych, ile jest sektorów na ścieżce. Każdy otwór indeksowy jest znacznikiem początku sektora. Dyskietki tego typu są wykorzystywane raczej rzadko.



Rys. 2. Wygląd zewnętrzny dyskietki

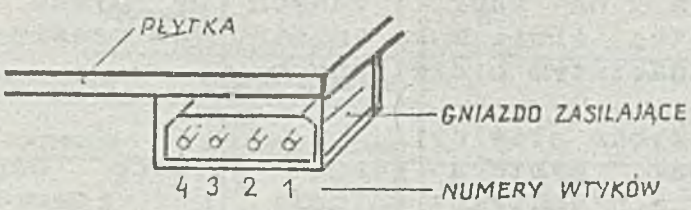
Poniżej przytoczono niektóre dane z projektu normy BN-87/3104-, dotyczące rozmiarów ścieżek i poszczególnych bitów na ścieżce.

- Liczba ścieżek na jednej stronie dysku powinna wynosić 40 lub 80. Dla 40 ścieżek, ich gęstość na 1 milimetr wynosi 1.88976, a dla 80 odpowiednio 3.77952. Wartości liczbowe występujące w normie podane są w układzie SI z zaokrągleniem, wynikającym z przeliczenia cala angielskiego na milimetry.
- Tolerancja lokalizacji ścieżki powinna wynosić dla dysku z 40 ścieżkami ± 0.085 mm, a dla dysku z 80 ścieżkami ± 0.0425 mm.
- Tolerancja kąta zapisu: odchyłka osi zmiany strumienia magnetycznego na ścieżce od promienia dysku elastycznego powinna mieścić się w granicach ± 0.005236 rad.
- Znamionowa długość komórki bitowej powinna wynosić dla metody zapisu FM 251.33 µrad, a dla metody MFM 125.66 µrad, co odpowiada znamionowej długości wyrażonej w jednostkach czasu odpowiednio 8 µs i 4 µs.
- Długookresowa średnia długość komórki bitowej, mierzona jako wartość średnia na długości 2500 komórek bitowych, powinna mieścić się w tolerancji ± 3.5% znamionowej długości komórki bitowej.

- Krótkookresowa średnia długość komórki bitowej, mierzona jako wartość średnia na długości 8 komórek bitowych, powinna mieścić się w tolerancji ± 8% długookresowej średniej komórki bitowej mierzonej w otoczeniu.

Interfejs

Na rys. 3 są przedstawione sygnały i odpowiadające im numery kontaktów interfejsu pomiędzy kontrolerem, a napędami dysków elastycznych.



Rys. 3. Interfejs napędu dysku elastycznego

Do jednego interfejsu kontrolera dysków elastycznych można przyłączyć maksymalnie dwa napędy dysków elastycznych, połączonych jednym płaskim kablem wielożyłowym z dwiema łączówkami. Do drugiej łączówki (na końcu kabla) jest dołączony napęd wybierany jako pierwszy (czyli A). Między łączówkami jest wykonany przeplot na odpowiednich liniach, aby zapewnić odpowiednie sterowanie napędami, bez konieczności zmiany ustawienia zwieraczy na samym napędzie.

Znaczenie poszczególnych sygnałów z interfejsu jest następujące:

- **DS0, 1, 2 i 3 (Drive Select 0, 1, 2 i 3).**
  Sygnał wyboru napędu - gdy jest aktywny, wybiera dany napęd i sygnały tylko tego napędu są dołączone do interfejsu. Gdy napęd nie jest wybrany, wszystkie jego sygnały wyjściowe są wprowadzone w trzeci stan logiczny, a sygnały wejściowe są ignorowane. Jedynie sygnał włączenia silnika (MOTOR ON) działa niezależnie od sygnału DS. Napęd może być wybrany jedynie wtedy, gdy jego silnik jest włączony.

- **Motor On (włączenie silnika).**
  Gdy jest aktywny, włączony jest silnik napędu dyskietki i po odczekaniu określonego czasu (typowo 0.3 do 1s), możliwe jest dokonywanie operacji zapisu lub odczytu, po uprzednim wybraniu napędu.

- **Direction Select (wybór kierunku).**
  Określa kierunek, w którym jest przesuwana głowica, po wystąpieniu impulsu na linii Step. Stan aktywny powoduje przesuwanie głowicy w kierunku osi silnika napędu dyskietki. Przez minimum 1 µs przed i po opadającym zboczu sygnału Step, stan wyboru kierunku musi być stabilny.

- **Step (krok).**
  Impuls na tej linii powoduje przesunięcie głowicy odczytu/zapisu na sąsiednią ścieżkę. O kierunku przesuwu decyduje sygnał Direction. Czas trwania impulsu wynosi 1 µs.

- **Write Data (zapis danych).**
  Linia zapisu danych na dysk elastyczny - dane mogą się pojawić na tej linii tylko wtedy, gdy aktywna jest linia Write Enable. Czas impulsów zapisu wynosi 150 ns.

- **Write Enable (zezwolenie na zapis danych).**
  Linia zezwala na zapis danych na dysku elastycznym. Uaktywnianie tej linii musi nastąpić w czasie od 4 do 8 µs po ostatnim znaczącym bicie danych.

- **Select Head (wybór głowicy).**
  Aktywny sygnał wybiera górną głowicę napędu dysków elastycznych, a nieaktywny dolną.

- **Index (Impuls indeksowy).**
  Na linii tej przy każdym obrocie dyskietki pojawia się impuls indeksowy wskazujący początek ścieżki (rys. 9).

- **Track 0 (ścieżka zerowa).**
  Sygnał ten jest aktywny, gdy górna głowica znajduje się nad ścieżką zerową dyskietki.

- **Write Protect (ochrona przed zapisem).**
  Sygnał ten jest aktywny, gdy do napędu jest włożona dyskietka zabezpieczona przed zapisem poprzez zaklejenie otworu.

- **Read Data (odczyt danych).**
  Na linii tej pojawiają się 150 ns impulsy odczytywanych danych .

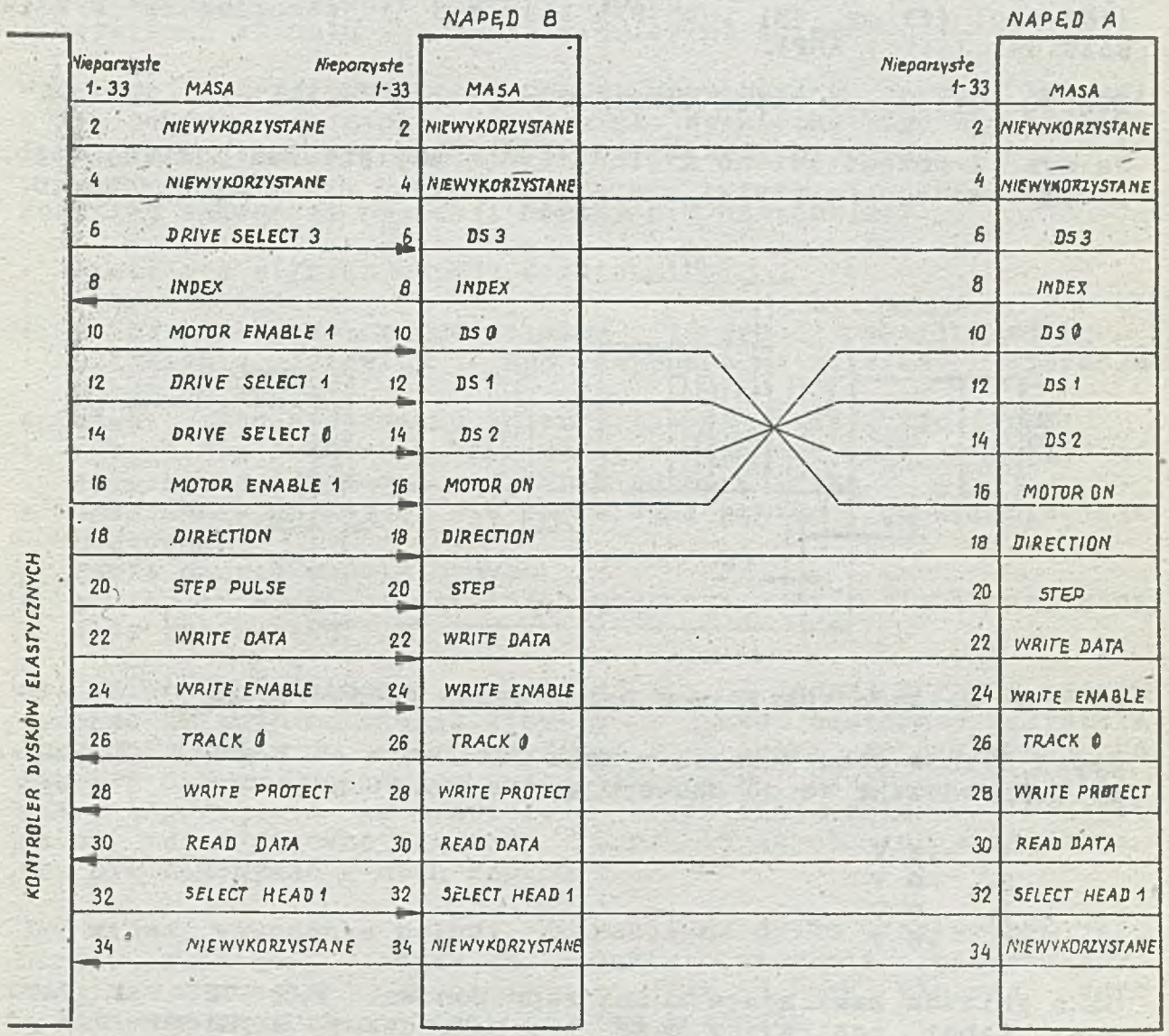
Uwaga:

W niektórych rodzajach dysków (w szczególności w napędach dysków o dużej gęstości) są wykorzystywane następujące linie:

- **Kontakt 2 - Reduced Write.**
  W stanie nieaktywnym sygnał wskazuje obecność dyskietki o dużej gęstości w napędzie dysków dużej gęstości, co wymaga normalnego prądu zapisu. Stan aktywny tego sygnału wskazuje obecność w napędzie dyskietki o małej gęstości, wymagającej zmniejszenia prądu zapisu.

- **Kontakt 34 - Diskette Change.**
  Sygnał ten jest aktywny, chyba że w napędzie jest obecna dyskietka i po wybraniu napędu, został odebrany sygnał indeksu. Niekiedy linia ta jest oznaczana jako READY.

Kontakt 4 - Head Load.
Sygnał aktywny powoduje opuszczenie głowicy na powierzchnię dyskietki. Sygnał ten nie jest wykorzystywany zarówno w komputerze IBM PC/XT, jak i w IBM PC/AT.



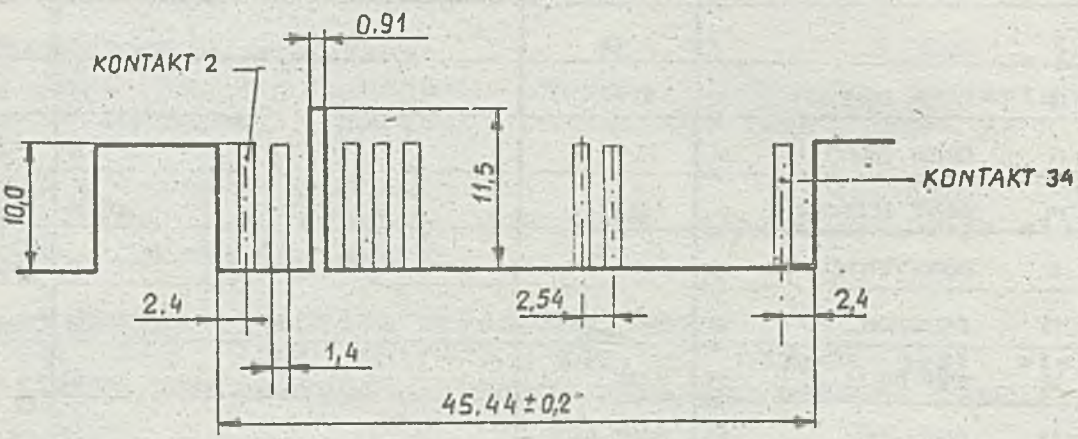
Rys. 4. Łącze interfejsu napędu dysku elastycznego

Na rys. 4 przedstawiono wymiary złącza krawędziowego napędu dysku elastycznego, do którego podłącza się płaski kabel sterujący. Na rysunku tym widoczne jest wycięcie kluczujące (między kontaktami 4 i 6), które zapobiega błędnemu włożeniu kabla sterującego. Przy braku klucza w łączówce kabla, należy uwzględnić wyróżnioną innym kolorem skrajną linię. Wyróżniona linia powinna znajdować się po stronie kontaktu 2.

Jako łączówkę wykorzystywaną w kablu płaskim, pasującą do wyżej wyspecyfikowanego złącza krawędziowego można zaproponować: P/N 3463-0001 (firmy 3M) lub P/N 583717-5 i jednocześnie P/N 1-583616-1 (firmy AMP).

Złącze zasilające napęd dysku elastycznego

Na rys. 5 przedstawiono wygląd zewnętrzny gniazda zasilającego, przymocowanego do płytki sterującej napędu dysku elastycznego.



Rys. 5. Gniazdo zasilające napęd dysku elastycznego

Napęd dysków elastycznych wymaga zasilania +5 i +12 V. Napięcia te doprowadzane są do następujących wtyków gniazda:

1. +12 V
2. 0 V
3. 0 V
4. +5 V

Jako gniazdo zasilające można zaproponować: P/N 172349-1 (firmy AMP). Kabel zasilający może być zakończony wtykiem: P/N 1-480424-0 i stykami 60617-1 lub 60619-1 (firmy AMP).

Zasada działania

Zasada działania napędu dysków elastycznych zostanie omówiona na podstawie typowego napędu 5.25 cala. Jeśli wystepować będą elementy, które w innych napędach są inaczej rozwiązane, zostanie dodany odpowiedni komentarz.

Napęd dysków elastycznych służy do zapisywania i odczytywania informacji z dyskietek. Napęd jest sterowany poprzez linie interfejsu i potrafi wykonywać jedynie bardzo proste, elementarne operacje, wśród których można wyróżnić:

- zapis danych,
- odczyt danych,
- włączenie lub wyłączenie silnika liniowego,
- przesuw głowicy na sąsiednią ścieżkę,
- wytworzenie sygnału indeksowego,
- wytworzenie sygnału ochrony zapisu,
- wysłanie sygnału położenia głowicy na ścieżce 0.

W dalszej części artykułu został przytoczony format ścieżki i aby dokonć np. zapisu zbioru na dyskietkę zgodnie z tym formatem, należy wykonać wiele elementarnych operacji.

Przykład sekwencji operacji mógłby być następujący:

- włączenie silnika napędu dyskietki,
- wybranie napędu,
- ustawienie głowicy nad ścieżką 0 (operacja rekalibracji),
- ustawienie głowicy nad wybraną ścieżką (operacja pozycjonowania),
- identyfikacja początku ścieżki (za pomocą sygnału INDEX),
- odczyt nagłówka sektora i na tej podstawie potwierdzenie, że głowica znajduje się nad odpowiednią ścieżką,
- odczekanie aż pojawi się odpowiedni sektor i odczytanie jego nagłówka,
- zapis danych w polu danych,
- powtórzenie dwóch ostatnich operacji, aż zostanie zapisany cały zbiór, itd.

Gdy zbiór jest długi i nie mieści się na jednej ścieżce należy dokonać zmiany położenia głowicy i znowu powtarzać wszystkie operacje związane z identyfikacją ścieżki i sektorów. Po zapisie danych należy dokonać zapisów w obszarach dyskietki, które odpowiadają za prawidłową organizację zbiorów. Ponadto należy identyfikować sektory, które są zajęte lub uszkodzone, aby nie dokonywać w nich zapisu.

Jak widać wykonanie pełnej operacji na dysku elastycznym jest bardzo złożone. Jednak z punktu widzenia programisty, realizacja dowolnej operacji nie jest aż tak złożona, gdyż wspomagają programistę trzy warstwy logiczne.

Na najniższym poziomie, napędem dysków elastycznych steruje kontroler, który jest zbudowany na ogół na układzie dużej skali integracji. Kontroler potrafi wykonywać wszystkie operacje elementarne, interpretuje sygnały wychodzące z napędu, identyfikuje nagłówki i dokonuje porównania zawartych w nich wartości z wartościami zaprogramowanymi.

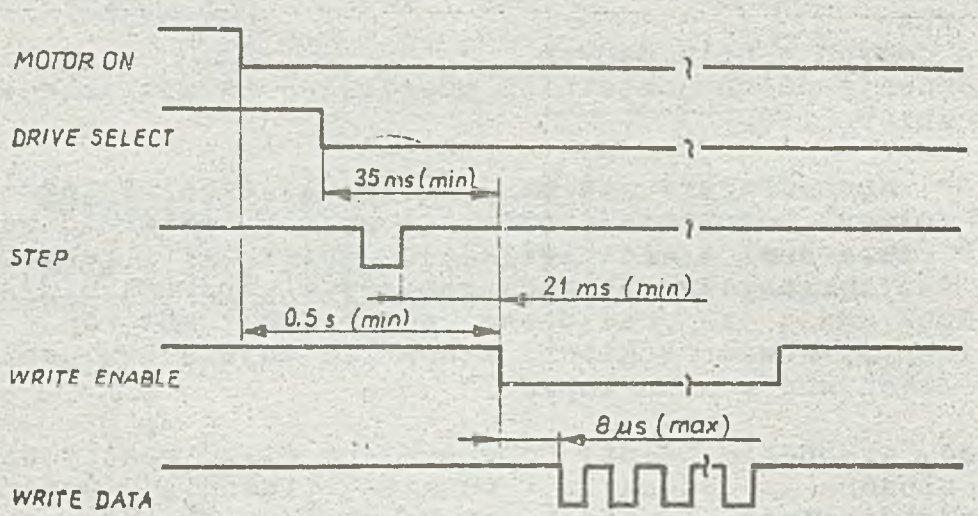
Kolejną warstwą logiczną jest znajdujący się w komputerze program sterujący BIOS (podstawowy system wejścia/wyjścia),

który między innymi zawiera procedury sterujące urządzeniami wejścia/wyjścia komputera. Procedury dotyczące obsługi napędów dysków elastycznych wykonują operacje znacznie bardziej złożone (ogólne), bez wnikania w sprawy, którymi zajmuje się kontroler. Przykładami procedur mogą być: zapis/odczyt określonego sektora, zaformatowanie zadanej ścieżki, itd. W szczególnych sytuacjach programista może wykorzystywać te procedury.

Najwyższym poziomem separującym programistę od sprzętu jest system operacyjny, którego przykładowymi zleceniami mogą być: formatowanie całej dyskietki, kopiowanie całej dyskietki, przepisanie całego zbioru, itd.

Po włożeniu dyskietki i zamknięciu napędu, silnik liniowy w niektórych napędach jest automatycznie włączany na kilka sekund. Można w tym czasie stwierdzić czy dyskietka jest włożona do napędu centrycznie. Po włożeniu dyskietki i zamknięciu napędu obsługa manualna operatora jest zakończona, sterowanie przejmuje komputer, napęd jest gotowy do działania.

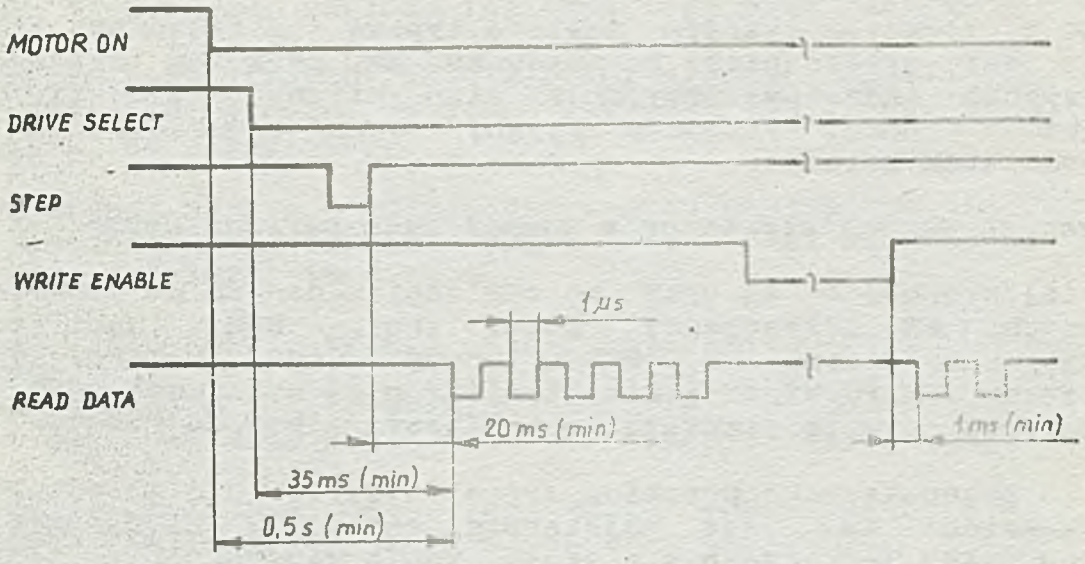
Przed rozpoczęciem jakiejkolwiek operacji na dyskietce należy włączyć programowo silnik napędzający dyskietkę, za pomocą sygnału **MOTOR ON**, a następnie odczekać określony czas (około 0.5 s), aby silnik nabrał nominalnej prędkości obrotowej. Następnie należy dokonać wyboru napędu za pomocą sygnału **DRIVE SELECT** i można rozpocząć operację. Na rys. 6 przedstawiono przykład przebiegu czasowego inicjacji zapisu.



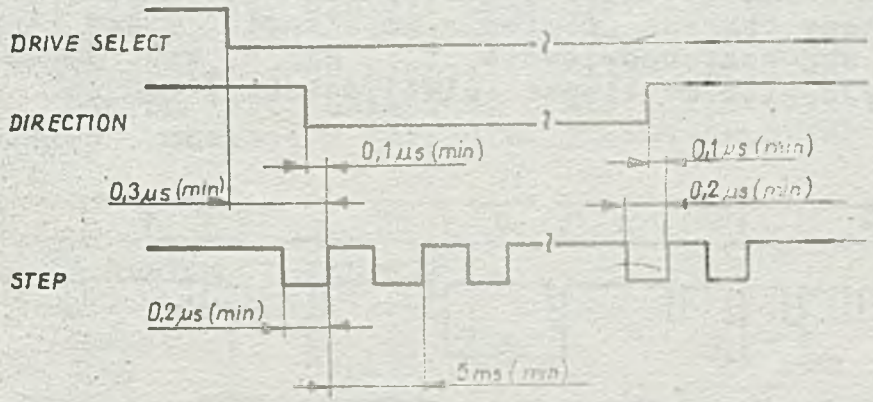
Rys. 6. Przebieg czasowy inicjacji operacji zapisu

Czas 35 ms między aktywnym sygnałem **DRIVE SELECT**, a aktywnym sygnałem **WRITE ENABLE** jest związany z czasem ładowania głowicy (dla tych napędów, w których głowica jest ładowana za pomocą elektromagnesu). Czas 21 ms między aktywnym sygnałam **STEP**, a aktywnym sygnałem **WRITE ENABLE** jest związany z czasem przejścia głowicy ze ścieżki na sąsiednią ścieżkę (typowo 6 ms) i z czasem uspokojenia głowicy (15 ms).

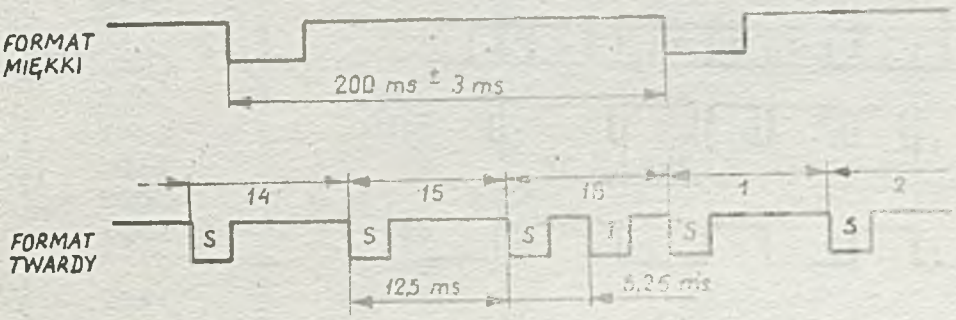
Na rys. 7 przedstawiono przebieg inicjacji odczytu.



Rys. 7. Przebieg inicjacji operacji odczytu



Rys. 8. Przebieg operacji pozycjonowania głowicy



Rys. 9. Przebiegi czasowe sygnału indeksowego

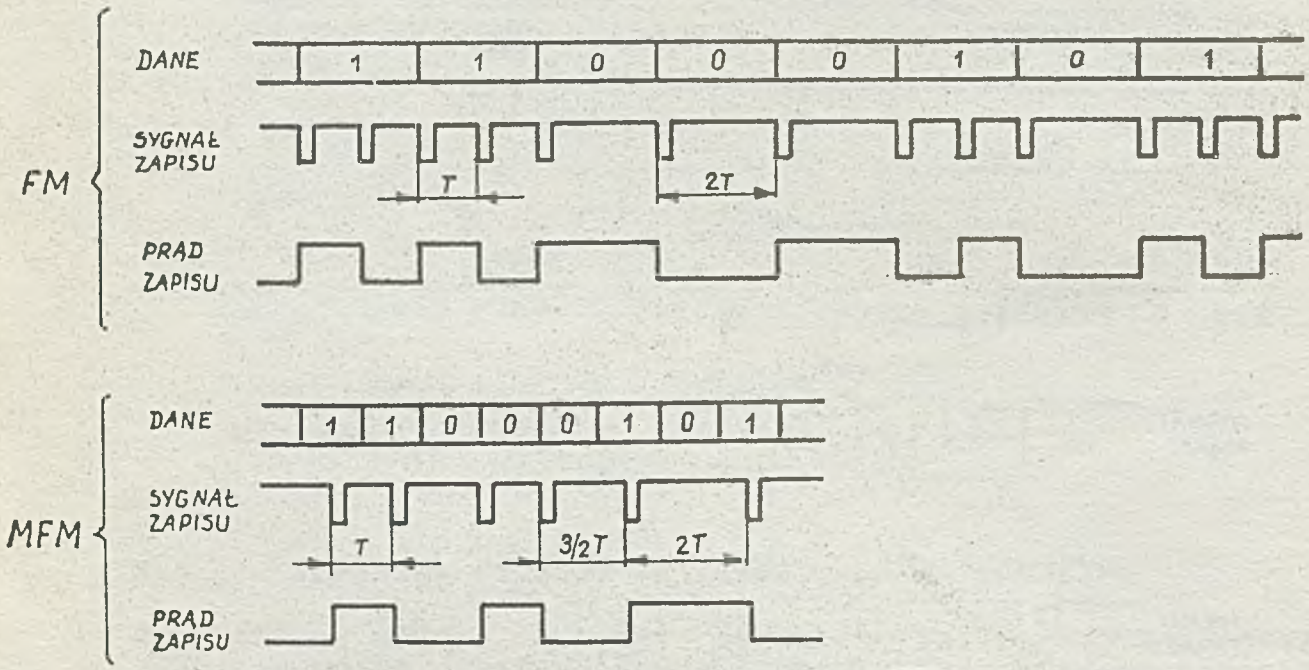
Na rys. 8 przedstawiono przebieg operacji pozycjonowania głowicy, a na rys. 9 przebiegi czasowe sygnału indeksowego dla dyskietki z formatem miękkim i twardym.

Dyskietka z formatem twardym zawiera typowo 16 sektorów; ich numeracja jest uwzględniona na rysunku. Na początku każdego sektora sygnał indeksowy przyjmuje stan aktywny, a w sektorze 16 również w środku występuje aktywny sygnał indeksowy, aby wskazać początek ścieżki.

Metody zapisu danych stosowane w komputerach personalnych

Najczęściej stosowane są trzy metody zapisu danych na dyskach elastycznych: **FM** (Frequency Modulation), **MFM** (Modified Frequency Modulation) i GCR (Group Code Recording). W komputerach IBM PC/XT/AT i kompatybilnych są stosowane metody **FM** i **MFM**, przy czym ta pierwsza jest stosowana bardzo rzadko.

Na rys. 10 przedstawiono przebieg sygnałów zapisu dla metody **FM** i **MFM**. Metoda **FM** jest najstarszą metodą i najbardziej niezawodną, ale jednocześnie gęstość zapisywanych danych jest najmniejsza. W metodzie **FM** każda komórka danych jest zaopatrywana na początku okresu trwania w impuls taktujący, przez co istnieje łatwość synchronizacji odczytywanych danych, nawet w wypadku zapisu długiej serii samych 0. Można wyodrębnić dwa różne odstępy czasowe pomiędzy impulsami. Czasy trwania tych odstępów wynoszą odpowiednio, T i 2T, a więc są znaczne różnice, istnieje więc możliwość łatwej detekcji danych.



Rys. 10. Przebiegi sygnałów zapisu dla metody FM i MFM

W metodzie MFM zwiększono dwukrotnie gęstość zapisu ponieważ nie wtrąca się impulsu taktującego w każdą komórkę danych. Aktywny impuls występuje w środku nominalnego okresu trwania bitu danych, gdy wartość logiczna tego bitu jest 1.

Ponadto jest wtrącany impuls pełniący rolę impulsu taktującego, na początku komórki danych, gdy występują co najmniej dwa kolejne zera logiczne.

Jak widać na rys. 10 gęstość zapisu danych została dwukrotnie zwiększona, przy zachowaniu nie zmienionego minimalnego odstępu między kolejnymi impulsami.

Jednakże doszedł trzeci odstęp czasowy o czasie trwania 3/2 T. Zwiększenie gęstości zapisu okupiono mniejszą niezawodnością i koniecznością stosowania bardziej rozbudowanych układów dekodujących dane w kontrolerze dysków elastycznych.

Trzecią stosowaną metodą kodowania danych jest GCR. Jest to metoda najbardziej zaawansowana technicznie, ale nie jest stosowana przez firmę IBM i dlatego nie będzie tak dokładnie omawiana.

GCR charakteryzuje się tym, że każdy bit danych zajmuje jednakową fizycznie długość ścieżki, a więc na zewnętrznych ścieżkach mieści się więcej sektorów. Nazwa tej metody zapisu, zapis z kodowaniem grup bierze się stąd, że do każdej sekwencji 4 bitów danych jest dołączany piąty bit i cała sekwencja jest odpowiednio kodowana, zgodnie z zależnością przedstawioną w tabeli:

BITY DANYCH	KOD GCR
0000	01010
0001	01011
0010	10010
0011	10011
0100	01110
0101	01111
0110	10110
0111	10111
1000	01001
1001	11001
1010	11010
1011	11011
1100	01101
1101	11101
1110	11110
1111	10101

Jak widać z powyższej tabeli, w strumieniu danych mogą wystąpić co najwyżej dwa kolejne zera logiczne, co powoduje że nie ma kłopotów z synchronizacją.

Format zapisu danych dla napędów elastycznych 5.25 cala

Standardowo wykorzystuje się dyskietki z miękkim formatowaniem i opis będzie dotyczył tego typu dyskietki.

Informacje na dyskach 5.25 calowych są zapisywane na koncentrycznych ścieżkach. Początek ścieżki wykrywa się wykorzystując otwór indeksowy.

Położenie pozostałych sektorów jest określone zapisem magnetycznym w procesie formatowania. Na dyskietce o pojemności 360 kB znajduje się 40 ścieżek na każdej powierzchni (razem 80 ścieżek). Na każdej ścieżce jest 9 sektorów, każdy o pojemności 512 bajtów.

Na dyskietce o pojemności 1.2 MB znajduje się 80 ścieżek na każdej powierzchni (razem 160 ścieżek). Na każdej ścieżce jest 15 sektorów, każdy o pojemności 512 bajtów.

Są również stosowane dyskietki o pojemności 720 kB, aczkolwiek nie przez firmę IBM. Znajduje się tam 80 ścieżek na każdej powierzchni (razem 160 ścieżek). Na każdej ścieżce jest 9 sektorów o pojemności 512 bajtów każdy.

Organizacja ścieżek dla dyskietki 360 kB jest następująca:

- Odstęp indeksowy.
- 9 kolejnych sektorów, których budowa jest następująca:
  a) odstęp nagłówka sektora: 12 bajtów 00, 3 bajty A1,
  b) pole identyfikacyjne,
     - bajt identyfikacyjny: FE,
     - numer ścieżki,
     - numer strony,
     - numer sektora,
     - długość sektora,
     - 2 bajty CRC,
  c) odstęp danych: 22 bajty 4E,
  d) blok danych,
     - odstęp bloku danych: 12 bajtów 00, 3 bajty A1,
     - bajt identyfikacji: FB,
     - dane: 512 bajtów,
     - 2 bajty CRC.

  Odstęp końcowy: zapisane bajty 4E, aż do napotkania otworu indeksowego.

Podstawowe parametry techniczne

Parametry charakteryzujące dyski elastyczne można podzielić tematycznie na kilka grup: pojemność pamięci, parametry funkcjonalne, parametry czasowe, wymiary napędu, parametry

zasilania, warunki środowiskowe, parametry eksploatacyjne. Poniżej przedstawiono poszczególne grupy parametrów wraz z typowymi wartościami dla napędu o średnicy dyskietki 5.25 cala i pojemności 360 kB.

- Pojemność pamięci

Metoda kodowania		FM	MFM
Pojemność formatowana (kB)	sektora	0.25	0.5
	ścieżki	2.25	4.5
	dysku	180	360

- Parametry funkcjonalne

Metoda kodowania	FM	MFM
Szybkość transmisji (kb/s)	125	250
Gęstość danych na wewnętrznej ścieżce (bpi)	2938	5876
Gęstość przemagnesowań na wewnętrznej ścieżce (frpi)	5876	5876

Liczba ścieżek 80
Gęstość ścieżek (ścieżki/cal) 48
Prędkość obrotowa dysku (obr/min) 300 ± 1.5%

- Parametry czasowe
Czas przejścia na sąsiednią ścieżkę (ms) 6
Czas docisku głowicy (ms) 35
Czas uspokojenia głowicy (ms) 15
Czas rozruchu (s) 0.5

- Wymiary napędu
Wysokość (mm) 41.3
Szerokość (mm) 146
Głębokość (mm) 203
Masa (kg) 1.3

- Parametry zasilania
+12 Vdc ± 5%, 0.22 A (średnio), 0.54 A (maksymalnie),
+5 Vdc ± 5%, 0.3 A (średnio), 0.38 A (maksymalnie).

Pobór mocy
4.1 W - moc średnia podczas działania napędu,
1.5 W - moc spoczynkowa.

- Parametry eksploatacyjne
Czas pracy bezawaryjnej MTBF 5000 h
Średni czas naprawy MTTR 0.5 h

Czas eksploatacji 5 lat
Czas eksploatacji dyskietki 3×10^6 (przejścia/ścieżkę)
Stopa błędu:
- korygowalny błąd $1/10^9$ bitów odczytanych
- niekorygowalny błąd $1/10^{12}$ bitów odczytanych
- błąd pozycjonowania $1/10^6$ szukanych ścieżek

Warunki środowiskowe

Zakres temperatur:
- pracy od 4° do 46° C
- transportu od -40° do 65° C
- składowania od -22° do 60° C

Wilgotność względna:
- pracy od 20 do 80%
- transportu od 5 do 95%
- składowania od 10 do 90%

Zmiana temperatury:
- pracy mniej niż 15° C/h
- transportu i składowania mniej niż 30° C/h

Wibracje:
- pracy mniej niż 0.5 G (0 do 55 Hz)
  mniej niż 0.25 G (55 do 500 Hz)
- transportu mniej niż 2 G (0 do 100 Hz)

Udary:
- pracy mniej niż 10 G (0 do 10 ms)
- transportu mniej niż 40 G (0 do 10 ms)

Wysokość:
- pracy mniej niż 3000 m
- transportu mniej niż 12000 m

ANALIZA ROZWIĄZAŃ UKŁADOWYCH WZORCÓW FIRM ZAGRANICZNYCH

HD-55 (HOSHIN)

Jest to stacja dysków elastycznych 5.25" typu "slim line" z zapisem dwustronnym o gęstości ścieżek 48 TPI i pojemności 0.5 MB. Stacja przeznaczona jest do współpracy z mikrokomputerem funkcjonalnie odpowiadającym IBM PC/XT.

Parametry techniczne stacji

- pojemność nieformatowana/formatowana	500 kB/360 kB
- szybkość transmisji	250 kbit/s
- czas rozruchu silnika	0.5 s
- prędkość obrotowa dyskietki	300 obr/min
- gęstość ścieżek	48 TPI
- liczba ścieżek na stronie	40
- liczba ścieżek na dysku	80
- metody zapisu	FM/MFM
- zasilanie	+12 V; +5 V

Ważniejsze podzespoły elektroniczne w układzie sterującym stacji dysków

Układ scalony	MC 3470
- " -	MC 3471
- " -	74 LS 38
- " -	74 LS 08
- " -	74 LS 14
- " -	74 LS 123
- " -	74 LS 74
- " -	74 LS 175
- " -	74 LS 86
- " -	ULN2003

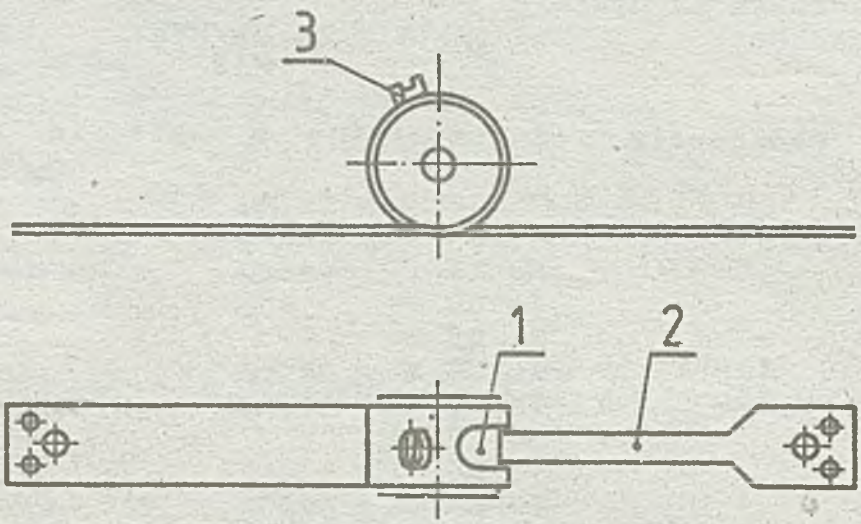
W układzie zapisu/odczytu pracują dwa wyspecjalizowane układy MC 3470 i MC 3471. Układ MC 3470 pracuje w torze odczytu i pozwala na prawidłowe odczytanie informacji zapisanej na dyskietce i przesłanie jej w postaci sygnału cyfrowego do kontrolera stacji. Układ 3471 pracuje w torze zapisu. Odbiera on sygnały z kontrolera, przeznaczone do zapisu na dyskietce oraz steruje pracą głowicy. Sterowanie funkcjami stacji zrealizowane jest na układach TTL serii LS. W układzie sterowania silnika krokowego do napędu karetki głowicy zastosowano również układ ULN2003, który zawiera tranzystory połączone w układzie Darlingtona o napięciu wyjściowym V_{CE} 50 V i prądzie wyjściowym I_C 500 mA.

Płytka sterująca wykonana jest na druku dwustronnym. Znajdujące się na niej złącza pozwalają na przyłączenie do niej

podzespołów współpracujących (silnik napędu dyskietki, silnik napędu karetki, głowice, czujnik ścieżki "00").

Dioda LED sygnalizująca stan gotowości stacji do pracy, fototranzystor wysyłający sygnały otworu indeksowego oraz fototranzystor wysyłający sygnał ochrony zapisu umieszczone są na osobnej płytce przyłączonej również do płytki sterującej.

Do wykonania omawianej stacji wykorzystuje się korpus odlewany wtryskowo, którego określone powierzchnie obrobiono metodą skrawania (frezowanie, wiercenie, rozwiercanie, gwintowanie). W prawym górnym rogu korpusu umieszczono silnik krokowy napędzający karetkę z dwiema głowicami. Oś silnika jest prostopadła do kierunku przemieszczania się karetki i równoległa do płaszczyzny dyskietki. Napęd z silnika przenoszony jest przez stalową taśmę związaną na stałe z karetką. Obrót silnika o kąt 3.6° powoduje przesunięcie głowic na sąsiednie ścieżki.



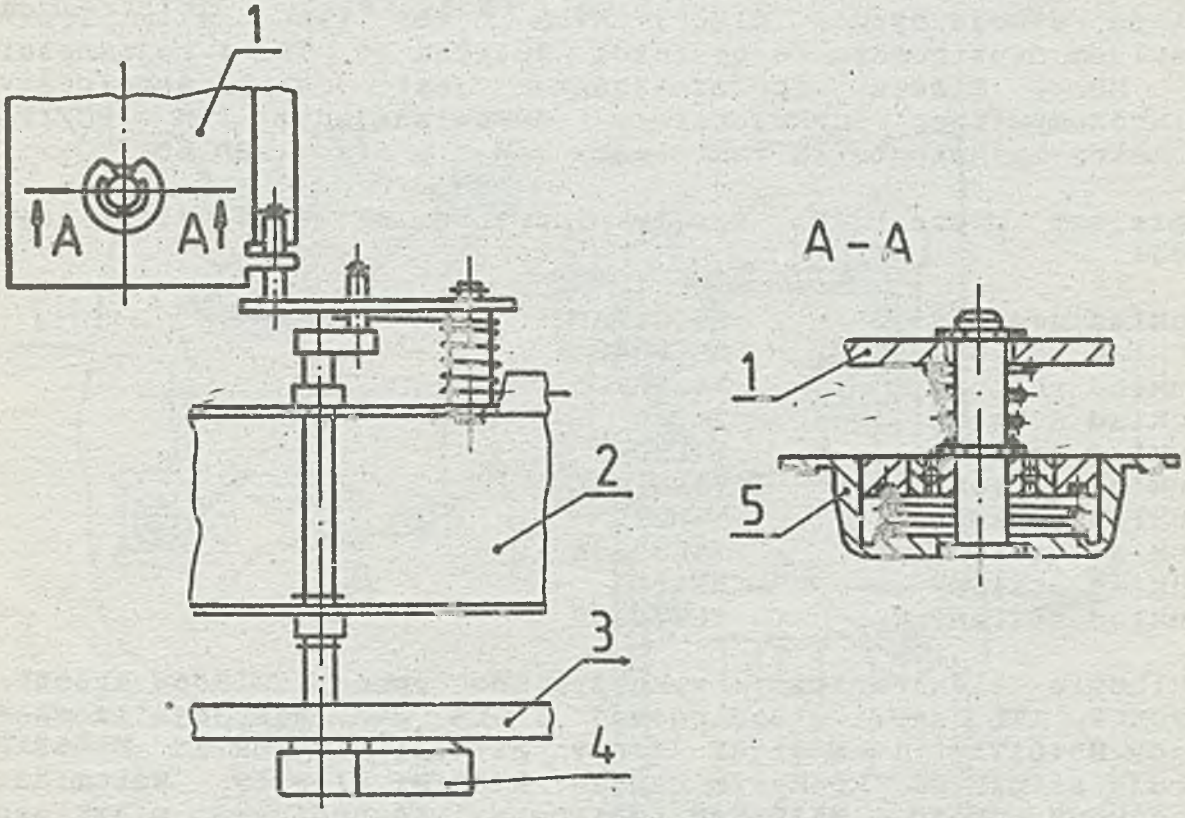
Rys.11. Układ przeniesienia napędu
1. Koło zdawcze silnika
2. Taśma napędowa
3. Wkręt mocujący taśmę

Karetka przemieszcza się wzdłuż walcowych prowadnic. Jedna z nich jest "odpowiedzialna" za właściwe prowadzenie, druga zaś uniemożliwia obrót karetki względem osi pierwszej prowadnicy. Wielkość siły docisku głowic można w pewnym zakresie regulować przez umieszczenie ramienia sprężyny dociskającej w odpowiednim wycięciu wspornika.

Z lewej strony karetki znajduje się układ zabezpieczający czoła głowic przed zetknięciem bez dyskietki. W jego skład wchodzi wspornik oraz dwie dźwigienki, które zajmując określone

położenia zależne od obecności dyskietki umożliwiają lub uniemożliwiają opuszczenie górnej głowicy. Na układ ten składają się także osie, sprężynki i pierścienie zabezpieczające.

Układ sprzęgający dyskietkę z silnikiem napędowym przedstawiono na rys. 12.



Rys.12. Układ sprzęgający dyskietkę z silnikiem

1. Ramię dociskowe
2. Płyta łożyskowa
3. Ścianka przednia
4. Dźwignia
5. Tarcza centrująco-dociskająca

Układ ten zawiera dźwignię 4, którą obracając zgodnie z ruchem wskazówek zegara o kąt 90° łączymy dyskietkę z kołem napędowym silnika. Dźwignia ta za pomocą przedstawionych na rysunku elementów pośredniczących powoduje opuszczenie ramienia dociskowego i oraz tarczy centrująco - dociskającej 5.

Stacja wyposażona jest także w mechaniczną blokadę uniemożliwiającą wycofanie głowicy poza ścieżkę "00".

W przedniej ściance zamontowano świetlik, za którym znajduje się dioda elektroluminescencyjna (LED) sygnalizująca stan gotowości stacji do pracy.

Do napędu dyskietki zastosowano silnik TM 600 PCB.

NFD-52B (SANKYO PEARL-LINE)

Jest to stacja dysków elastycznych 5.25" typu "slim line" z zapisem dwustronnym o gęstości ścieżek 48 TPI i pojemności 0.5 MB. Stacja przeznaczona jest do współpracy z mikrokomputerem funkcjonalnie odpowiadającym IBM PC/XT. Parametry techniczne są takie same jak dla stacji HD-55.

Ważniejsze podzespoły elektroniczne w układzie sterującym stacji

- układ scalony M51017AP
- układ scalony M54534P
- układ scalony 74LS27
- układ scalony 7438N
- układ scalony 74LS00
- układ scalony 74LS14
- układ scalony 74LS74
- układ scalony 74LS423N
- układ scalony NE555P
- układ scalony TL7705CP

W układzie elektronicznym wykorzystano oprócz układów **NE555P, TL7705CP, TTL** serii standardowej i LS dwa wyspecjalizowane układy **M51017AP** i **M54534P** firmy Mitsubishi. Układ **M54534P** steruje silnikiem krokowym napędu karetki głowicy. Natomiast 60-nóżkowy układ **M51017AP** montowany technologią płaskiego montażu pracuje w torze zapisu i odczytu danych z dyskietki realizując jednocześnie funkcje sterowania głowicami.

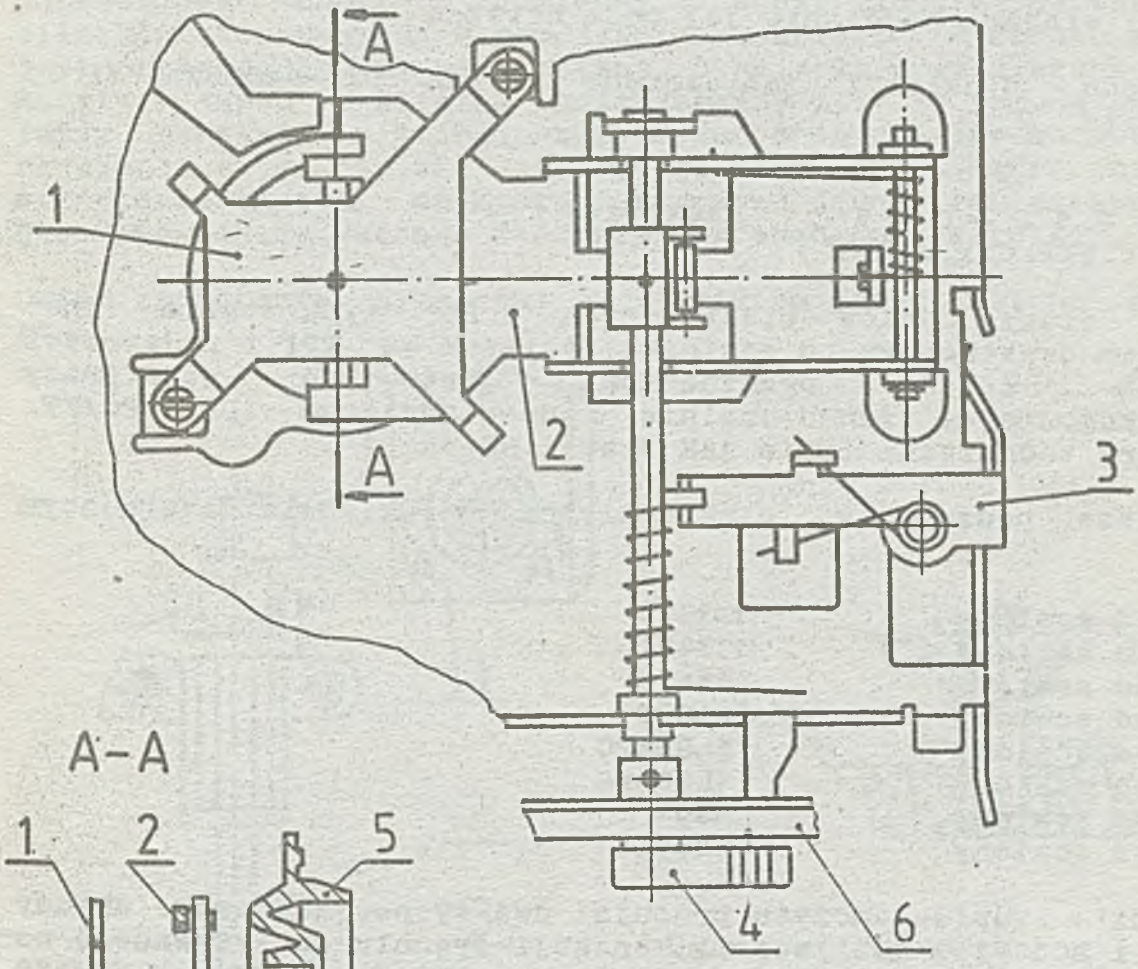
Płytka sterująca wykonana jest na druku dwustronnym z zastosowaniem technologii montażu tradycyjnego i płaskiego. Na osobnej płytce są umieszczone dioda LED sygnalizująca gotowość stacji do pracy, fototranzystor wysyłający sygnały otworu indeksowego oraz fototranzystor wysyłający sygnał ochrony zapisu.

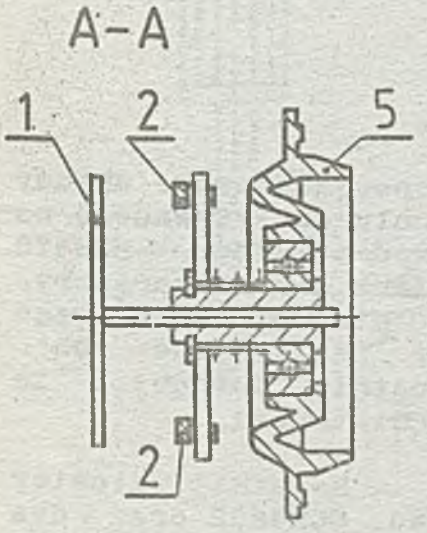
Stacja ma korpus wykonany z blachy stalowej przez wykrawanie, gięcie i ciągnienie. W celu usztywnienia korpusu w miejscach dużych płaszczyzn wykonano przetłoczenia. Ścianki boczne po dwukrotnym zagięciu pod kątem 90° przybrały postać ceowników. Ponadto przynitowano dwa kołki z gwintowanymi otworami, służącymi do zamocowania płytki sterowania.

Stacja wyposażona została w silnik krokowy 39SH-29BAG firmy Matsushita o kroku 1.8° umieszczony w prawym górnym rogu korpusu. Napęd karetki przenoszony jest z silnika za pomocą stalowej taśmy w sposób identyczny jak w stacji HD-55.

Karetka prowadzona jest przez dwie prowadnice walcowe. Należy jednak dodać, że jedna z prowadnic współpracuje ze sprężyną kasującą luz między karetką, a prowadnicą.

Z lewej strony karetki znajduje się wspornik z umieszczoną na nim dźwignią, która współpracując z ramieniem dociskowym podnosi lub opuszcza górną głowicę.





1. Prowadnik tarczy
2. Ramię dociskowe
3. Dźwignia blokady
4. Dźwignia
5. Tarcza centrująco-dociskająca
6. Ścianka przednia

Rys.13. Układ sprzęgający dyskietkę z silnikiem

Zabezpieczenie przed zetknięciem czół głowic bez dyskietki zrealizowano w nieco odmienny sposób. Polega on na tym, że przy braku dyskietki w stacji, dźwignia sprzęgająca dyskietkę z silnikiem napędowym jest zablokowana. Po usunięciu dyskietki, dźwignia blokady 3 zmienia swoje położenie i umożliwia obrót dźwigni sprzęgającej.

W ściance przedniej, podobnie jak w stacji HD-55, znajduje się świetlik, za którym LED sygnalizuje stan gotowości stacji do pracy.

Stacja zawiera również mechaniczną blokadę karetki uniemożliwiającą wycofanie jej poza ścieżkę "00".

Do napędu dyskietki zastosowano silnik DFX-80C2BRA firmy Matsushita.

FDD 212-5-10 (WONG'S TECHNOLOGY)

Jest to stacja dysków elastycznych 5.25" typu "slim line" z zapisem dwustronnym o gęstości ścieżek 48 TPI i pojemności 0.5 MB. Stacja przeznaczona jest do współpracy z mikrokomputerem funkcjonalnie odpowiadającym IBM PC/XT. Parametry techniczne takie jak stacji HD-55

Ważniejsze podzespoły elektroniczne w układzie sterującym stacji

Układ scalony	MC3470
Układ scalony	MC3471
Układ scalony	7438PC
Układ scalony	7416PC
Układ scalony	74LS04PC
Układ scalony	74LS08PC
Układ scalony	LM393
Układ scalony	UDN5711

W układzie zapisu/odczytu pracują dwa wyspecjalizowane układy MC3470 i MC3471 z możliwością korekcji sygnału odczytywanego na wyjściu komparatora włączając w obwód układu MC3470 potencjometr regulowany. Układ LM393 zawiera dwa komparatory, do których wejść podawane są sygnały z otworu indeksowego dyskietki oraz sygnał ustawienia głowicy na ścieżce "00". W komparatorach tych słabe sprzężenie dodatnie powoduje, że pracują one jako dyskryminatory napięcia z histerezą.

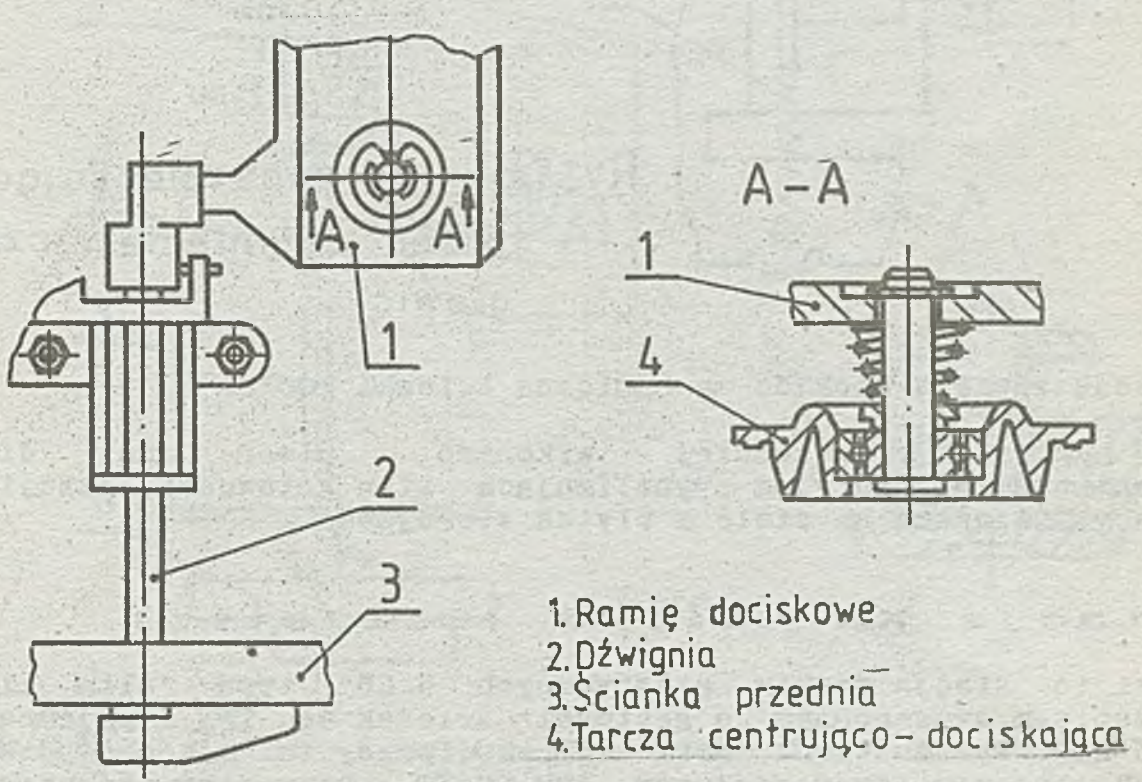
Sterowanie silnika krokowego napędu karetki głowicy zrealizowano stosując specjalistyczny układ UCN4202 oraz dwa układy UDN5711. Wykorzystano również w obwodach sterowania układy TTL serii standardowej i LS.

Dioda LED, fototranzystor ochrony zapisu i fototranzystor otworu indeksowego umieszczone są na tej samej płytce sterującej wykonanej na druku dwustronnym.

Korpus jest odlewany wtryskowo ze stopów lekkich. Ścianki boczne są pochylone i nie mają otworów, które musiałyby być formowane z suwaków. Przystosowany jest również do pośredniego napędu dyskietki. Jednak w tym modelu zastosowano napęd bezpośredni wykorzystując silnik F-80DEZSLR firmy Sankyo.

Silnik napędu karetki to krokowy silnik MSHF2000B71 produkcji Sankyo. Jest on nieco odmienny w konstrukcji w porównaniu do silników stosowanych w poprzednich modelach. Oś obrotu silnika jest prostopadła do płaszczyzny dyskietki oraz kierunku ruchu karetki. Napęd z silnika przekazywany jest za pomocą stalowej taśmy. Karetka, w której zamocowano dwie głowice ma możliwość regulacji siły nacisku górnej głowicy przez wkręcenie lub wykręcenie wkręta, na którym spoczywa jedno ramię sprężyny. Dwie prowadnice walcowe zapewniają prowadzenie karetki.

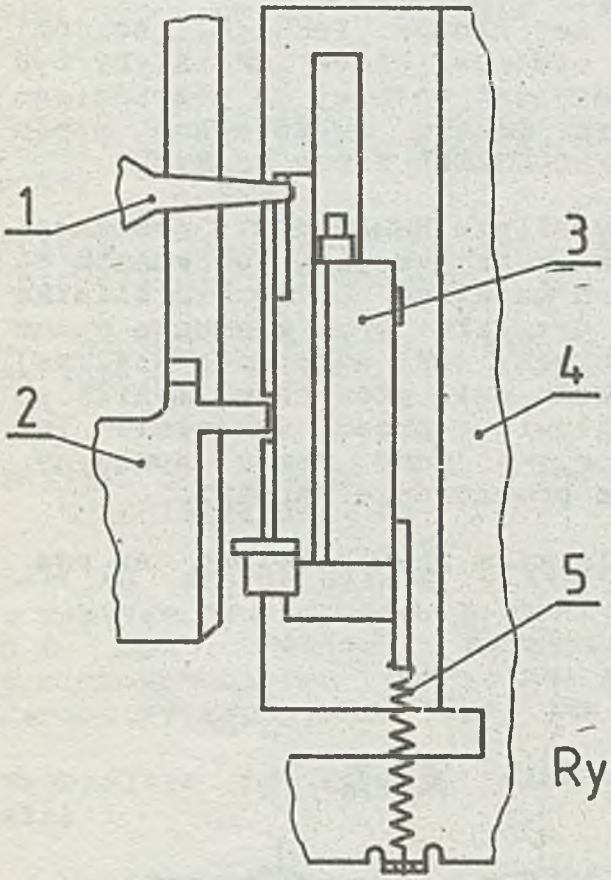
Układ sprzęgania dyskietki z silnikiem przedstawiono na rys. 14.



Rys.14. Układ sprzęgający dyskietkę z silnikiem

Zastosowano także układ zabezpieczający przed zetknięciem czół głowic (rys. 15), który pełni dodatkową funkcję wysuwania dyskietki po uniesieniu ramienia dociskającego. Ułatwia to wymianę dyskietek. Na uwagę zasługuje bardzo prosta konstrukcja

układu sprowadzająca się do suwaka, sprężyny oraz odpowiednich wycięć w płycie.



1.Ramię karetki
2.Ramię dociskowe
3.Suwak
4.Płyta
5.Sprężyna

Rys.15. Układ zabezpieczający przed zetknięciem czół głowic

Stacja zawiera blokadę mechaniczną ścieżki "00".

W ściance przedniej wykonano otwór na diodę elektroluminescencyjną sygnalizującą stan gotowości stacji do pracy, związaną na stałe z płytką sterującą.

Stacja XM 4311/3 (Olivetti)

Jest to stacja dysków elastycznych 5.25" typu "slim line" z zapisem dwustronnym o gęstości ścieżek 48 TPI i pojemności 0.5 MB. Przeznaczona jest do współpracy z mikrokomputerem funkcjonalnie odpowiadającym IBM PC/XT. Parametry techniczne takie jak stacji HD-55.

Ważniejsze podzespoły elektroniczne w układzie sterującym stacji

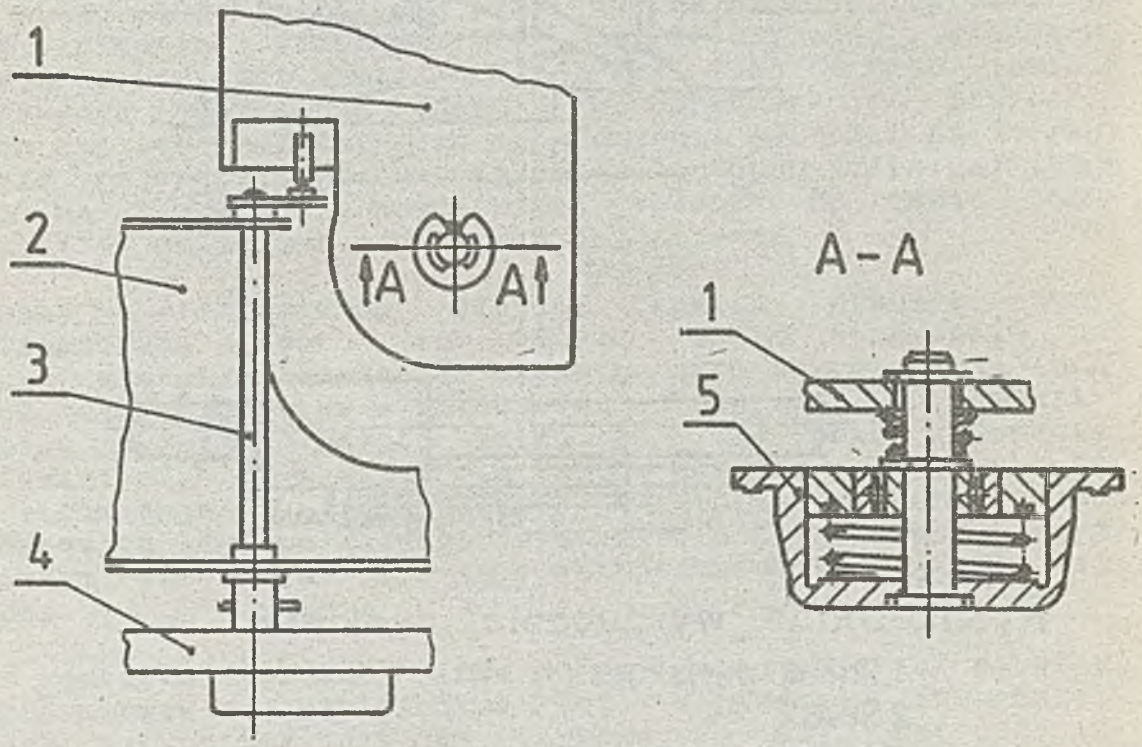
Układ scalony	MC3470
Układ scalony	SN94510CN
Układ scalony	HA13007
Układ scalony	LM339

Układ scalony	74LS04
Układ scalony	7438
Układ scalony	7407
Układ scalony	NE555P
Układ scalony	ULA9RB015E
Układ scalony	DAN403

Układ sterowania oprócz układów TTL zawiera cztery układy wyspecjalizowane: MC3470, SN94510CH, HA13007 oraz 40-nóżkowy układ ULA9RB015E firmy Ferranti, który stanowi integralną część układu sterującego.

W torze zapisu/odczytu pracują dwa układy MC3470 i SN94510CH. Układ odczytu nie zawiera elementów regulacyjnych.

Uzwojenia silnika krokowego napędu karetki połączone są z układem HA13007 firmy Hitachi z wysokonapięciowymi i wysokoprądowymi obwodami wyjściowymi.



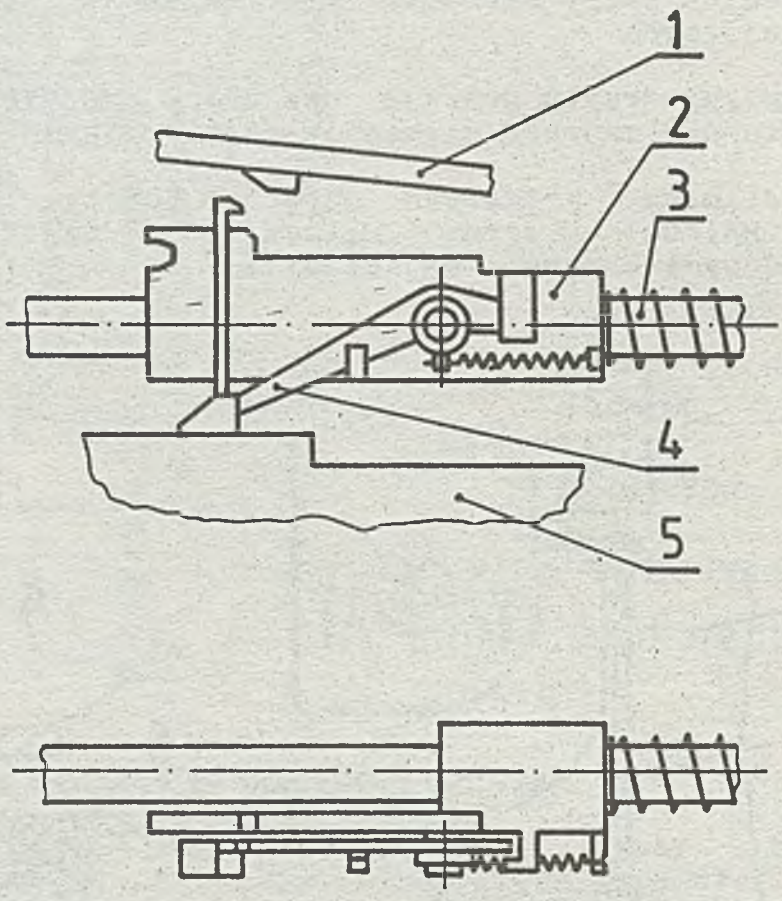
Rys.16. Układ sprzęgający dyskietkę z silnikiem

1. Ramię dociskowe
2. Płyta łożyskowa
3. Dźwignia
4. Ścianka przednia
5. Tarcza centrująco-dociskająca

Korpus stacji jest odlewany wtryskowo ze stopów lekkich.

W lewym górnym rogu znajduje się silnik krokowy KP39HM4-016 o kroku 3.6°, napędzający przez stalową taśmę karetkę z dwiema głowicami. Prowadzenie karetki zapewniają dwie prowadnice walcowe.

Omawiana stacja nie ma zabezpieczenia przed zetknięciem czół głowic bez dyskietki. Wyposażona jest jednak w układ 'wyrzucania' dyskietki po uniesieniu ramienia dociskowego. Budowę tego układu przedstawia rys. 17.



Rys.17. Układ „wyrzucania" dyskietki

1. Ramię dociskowe
2. Suwak
3. Prowadnica karetki
4. Dźwignia
5. Korpus

W ściance przedniej umieszczono diodę LED sygnalizującą stan gotowości stacji do pracy. Kołnierz diody mocowany jest w uformowanych w ściance zatrzaskach.

CHARAKTERYSTYKA GŁOWNYCH PODZESPOŁOW

Głowica

Zespół dwóch głowic ferrytowych (rejestracja po obu stronach dyskietki) umocowanych w specjalnych ramkach, przystosowanych do jednoczesnego pozycjonowania głowic po obu stronach dyskietki przez odpowiednią konstrukcję mechaniczną, zapewnia właściwy docisk i opasanie głowicy przez część elastycznej dyskietki. Głowice te są typu uniwersalnego (zapisująco/odczytujące) z kasowaniem tunelowym przy użyciu prądu stałego.

Istotne tu jest przesunięcie szczelin roboczych głowic o ok. 2.3 mm, tj. dokładnie o 4 odległości pomiędzy ścieżkami zapisanymi na dyskietce.

Głowice muszą umożliwiać zapis i odczyt w standardzie 48 TPI (gęstość 48 koncentrycznych ścieżek na cal), z szerokością zapisu śladu 300 ± 25 µm przy nominalnych odległościach pomiędzy sąsiednimi ścieżkami 529 µm. Obowiązujący standard zapisu wzdłuż ścieżki wymaga rejestracji 7958 ftprad (zmian namagnesowania nośnika na radian), co prowadzi do gęstości liniowych zapisów uzależnionych od promienia ścieżki, na której odbywa się rejestracja do wartości około 5900 zmian/cal (230 zmian/mm) na najbardziej wewnętrznej ścieżce, do około 3600 zmian/cal (140 zmian/mm) na zewnętrznej ścieżce.

Od strony elektrycznej głowica zapisu pracuje przy częstotliwościach 125000 lub 250000 zmian strumienia na sekundę, a właściwie w odcinku czasu będącym odwrotnością tych wielkości, tj. 8 lub 4 µs w zależności od zapisanej informacji. Prąd zapisu wynosi z reguły kilka mA, zaś prąd kasowania kilkadziesiąt mA. Ze względu na to, że głowica pracuje z dość dużym strumieniem (indukcjami), istnieje konieczność rozmagnesowania jej po każdorazowym użyciu do pisania, gdyż w przeciwnym razie pojawiłby się efekt podkasowywania zapisanej informacji.

W czasie odczytu zapisanej informacji na uzwojeniu głowicy powstaje napięcie uzależnione od zapisanej informacji o wartości kilkudziesięciu mV (typowo 25 mV). Wartość ta uzależniona jest również od właściwości nośnika i napędu, na którym odbyła się rejestracja, jednak ze względu na poprawność odczytu, znacznie bardziej niż rozrzuty napięcia odczytu, istotne jest rozsuwanie wierzchołków (tzw. peak shift) oglądane na oscyloskopie, związane z charakterystyką częstotliwościową (aplitudowo-fazową) głowicy, gdyż na obecnym etapie rozwoju technologii układów scalonych nie występują trudności ze wzmacnianiem takich sygnałów.

Elektronika odczytu i zapisu

Najczęściej stosowanymi układami w torze odczytu/zapisu są układy firmy Motorola typu MC3470 i MC3471.

Mogą one współpracować z głowicą dwustronną, realizując wszystkie funkcje związane z zapisem i odczytem danych z dyskietki. Różnice występują jedynie w sposobie korekcji sygnałów, podlegających formowaniu w tych obwodach. Element regulacyjny (potencjometr regulowany) w stacji FDD pozwala na ustawienie chwili, w której zmiana sygnału wyjściowego komparatora (w układzie MC3470), pracującego jako detektor przejścia przez zero, następuje dokładnie w momencie przejścia przez zero sygnału wejściowego, co wpływa na poprawność odczytu sygnału zapisanego na dyskietce. Potencjometrem tym, włączonym w obwód wzmacniacza różniczkującego, regulować należy do momentu uzyskania na wyjściu 10 układu MC3470 symetrycznego sygnału cyfrowego.

Silnik pozycjonera głowic

Do pozycjonowania głowic magnetycznych powszechnie stosuje się silniki skokowe, spotyka się ich dwa wykonania, tj. płaskie i prostopadłościenne, o 100 lub 200 stabilnych położeniach wirnika na obrót. Wykonanie silnika wpływa na kształt i wielkość płytki drukowanej z elektroniką.

Od silnika, poza gabarytami zewnętrznymi wymaga się:

- pracy przy napięciu zasilania 12 V,
- skoku 3.6^{o} lub 1.8^{o} (uwarunkowanie związane z przeniesieniem napędu za pomocą taśmy metalowej nawiniętej na koło o średnicy ok. 17 mm),
- poprawnej pracy przy zasilaniu w układzie otwartym z bezwładnością ok. 3 gcm^2 przy czasach pobudzeń 5 do 6 ms,
- odpowiedniej dokładności uzależnionej od:
  - wykonania samego silnika - niekumulowany błąd położenia stabilnych stanów wirnika nie obciążonego momentem,
  - sztywności charakterystyki momentu funkcji kąta obrotu wirnika - zwykle sinusoida z zerem w stabilnym punkcie i wartościami ekstremalnymi w sąsiednich stabilnych punktach.

Faktyczne zatrzymanie wirnika (położenie głowicy) następuje w obszarze uzależnionym od relacji momentu tarcia głowica - nośnik i momentu maksymalnego statycznego silnika, który to moment liniowo maleje wraz ze wzrostem temperatury rezystancji uzwojenia. Wydaje się, że łączny maksymalny uchyb powinien być niższy niż 10 %.

Układ elektroniczny sterujący silnikiem pozycjonera głowic

W napędzie HDD-55 do sterowania silnika krokowego zastosowano układy TTL serii LS oraz układ ULN2003.

Układ ULN2003 steruje silnikiem krokowym napędu karetki i zawiera siedem par tranzystorów połączonych w układzie Darlingtona z diodami zabezpieczającymi te tranzystory od skutków przepięć, przy pracy tego układu z obciążeniem indukcyjnym.

Ponieważ w stacji HD-55 zastosowano silnik o skoku znamionowym 1.8^{o} układ musi zapewnić dwukrotny skok silnika pod wpływem pojedynczego impulsu STEP, aby obrót jego wynosił 3.6^{o}, co jest wymagane dla stacji o gęstości zapisu ścieżek 48 TPI. Takie rozwiązanie zapewnia możliwość sterowania silnikiem o skoku znamionowym 1.8^{o} lub 3.6^{o} w stacjach, gdzie wymagany jest skok 3.6^{o}, jak również sterowanie silnikiem o skoku znamionowym 1.8^{o} w stacji o gęstości zapisu ścieżek 96 TPI (dla stacji przeznaczonych do współpracy z mikrokomputerem IBM AT). Zasilanie tego silnika zrealizowano w sposób dwustopniowy, który zapewnia w okresach, gdy ma on ustaloną pozycję, mniejszy pobór mocy przez ten silnik, a jednocześnie mniejsze obciążenie układu sterującego tym silnikiem. Pełne napięcie zasilające (12 V) jest podawane na silniki tylko w chwilach jego obrotu, w celu uzyskania określonego momentu napędowego. Gdy osiągnie on żądaną ścieżkę napięcie spada do 5 V, co wystarcza do utrzymania silnika w stałym położeniu.

Silnik do napędu dyskietki zintegrowany z elektroniką

Jest to silnik o specjalnej konstrukcji, płaski, o grubości ok. 1 cm i średnicy 9 cm, rozwijający dość duży moment napędowy przy prędkości wirowania 300 obr/min (5 obr/s), z dużą stabilnością prędkości chwilowej. Ze względu na wymiary konstrukcji nie ma możliwości zastąpienia go silnikiem innego typu.

Wirnik zbudowany jest w formie pierścienia ferrytowego o średnicy zewnętrznej 90 mm, wewnętrznej ≈ 40 mm i grubości L = 5 mm, co pozwala oszacować jego moment bezwładności jako $J = 1500\ gcm^2 = 1.5\ Gcms^2$. Wirnik ferrytowy namagnesowany jest w 8 segmentach wzdłuż obwodu (8×45°).

Stojan silnika zbudowany jest z 6 cewek otwartych (elektromagnesów) umocowanych do płytki z laminatu, z obwodem drukowanym, do którego również przymocowane są trzy elementy Halla - czujniki położenia kątowego wirnika dostarczające sygnałów do układu elektronicznego, pełniącego funkcje komutatora prądów w fazach uzwojeń stojana; do zasilania silnika używa się napięcia + 12 V. Na płytce laminatowej pod wirnikiem nadrukowane jest dodatkowe uzwojenie - prawdopodobnie dla celów sterowania w stanach prędkości ustalonej.

Dla całego tego węzła (w warunkach obciążenia dyskietką oraz ze składnikiem starzeniowym) najczęściej producenci określają prędkość wirowania 5 obr/s ± 1.5 %; stabilizacja prędkości jest realizowana elektronicznie. Od prędkości zależy bezpośrednio liniowa gęstość zapisu oraz częstotliwość, amplituda i przesunięcia wierzchołków sygnału odczytu. Moc mechaniczna potrzebna na wałku silnika jest iloczynem prędkości kątowej (wyrażonej w radianach) i momentem sprowadzonego na wałek silnika, a pochodzącego od siły tarcia w układzie głowica - dyskietka (pozostałe składniki tarcia można zaniedbać).

Norma branżowa podaje wartość momentu obciążenia w nominalnym punkcie pracy jako M_t = 0.03 Nm. Zatem niezbędna moc mechaniczna wynosi ≈ 1 W. Moment (średni) konieczny do rozpędzenia dyskietki obciążonej głowicą w czasie T ≈ 0.25s można oszacować (iloczyn momentu bezwładności i przyspieszenia kątowego wirnika) łącznie na 0.05 Nm (500 Gcm).

Czujniki

W konstrukcji napędu konieczne jest zastosowanie trzech czujników dwustanowych typu fotooptycznego i jednego typu mechanicznego (kontaktron lub zestyk mechaniczny).

Konstrukcja mechaniki napędu

W korpusach wykonywanych jako odlewy wtryskowe ze stopów lekkich, wzajemne położenie silnika napędzającego dyskietkę i karetki z głowicami, zależy od dokładności uzyskanych na drodze obróbki skrawaniem. W korpusie blaszanym położenie wymienionych podzespołów zależy od dokładności uzyskanych w procesie wykrawania i gięcia.

Na uwagę zasługuje także układ sprzęgający dyskietkę z silnikiem stosowany w stacji FDD. Składa się on z na ogół niewielu niezbyt trudnych technologicznie części. Dosyć łatwy wydaje się także montaż tego układu.

W stacji FDD interesujące jest także rozwiązanie układu zabezpieczającego przed zetknięciem czół głowic. Przy użyciu zdecydowanie najmniejszej liczby części, układ ten spełnia także dodatkową funkcję wysuwania dyskietki. Jest to jedyny model stacji (spośród omawianych), w którym realizowane są obie z wymienionych funkcji jednocześnie.

Układ przeniesienia napędu karetki rozwiązany jest podobnie we wszystkich modelach. Pewne różnice wynikają jedynie z konstrukcji karetek. Sposób prowadzenia karetki w poszczególnych modelach stacji jest również bardzo podobny. Wypada jednak zwrócić uwagę na kasowanie luzu między prowadnicą a karetką w stacji NFD.

Literatura

[1] Technical Reference Personal Computer XT
[2] Technical Reference Personal Computer AT
[3] Nota aplikacyjna AN917 (firmy Motorola)
[4] Karty katalogowe układów MC3470 i MC3471P (firmy Motorola)
[5] Norma ISO 7487/1 (wydana 1985.03.15)
[6] Norma ISO 6596/1 (wydana 1985.04.01)
[7] Norma ISO 8378/1 (wydana 1986.12.15)
[8] TEAC FD-55BV-16 Mini Flexible Disk Drive Specification

Wojciech CHOLEWA, Wojciech MOCZULSKI
Instytut Mechaniki i Podstaw Konstrukcji Maszyn
Politechnika Śląska, Gliwice

REPREZENTACJA WIEDZY W DIAGNOSTYCZNYCH SYSTEMACH DORADCZYCH*

1. Wprowadzenie**

Możliwości stworzone przez technikę pomiarową pozwalają na ciągłe obserwowanie i gromadzenie licznych sygnałów związanych z działaniem wybranych maszyn. Sygnały te doprowadzane są najczęściej do prostych układów zabezpieczających i ostrzegających, które wykrywają przekroczenie przez obserwowane wielkości zadanych wartości progowych. Jednocześnie prowadzone są prace nad wdrożeniem układów umożliwiających analizę rejestrowanych sygnałów w celu estymacji ich cech, mogących być symptomami aktualnego stanu technicznego badanej maszyny.

Interpretacja wyników takich analiz jest czynnością skomplikowaną i wymagającą odpowiedniego doświadczenia. Z reguły nie jest więc zasadne udostępnianie kompletnych wyników personelowi, bezpośrednio nadzorującemu działanie maszyn. Najczęściej ze zbioru cech sygnałów wybiera się nieliczne wielkości i przekazuje je (okresowo lub w sposób ciągły) dyspozytorowi maszyny. Wszystkie wyniki analiz mogą być także dostępne specjaliście z zakresu diagnostyki maszyn. Specjalista ten winien posiadać obszerną wiedzę interdyscyplinarną oraz umiejętności, będące wynikiem długotrwałego i indywidualnego gromadzenia doświadczeń. Określenie sposobu i programu szkolenia specjalistów w zakresie diagnostyki (o ile w ogóle możliwe) jest więc poważnym problemem.

Dostępny obecnie w Polsce sprzęt mini- i mikrokomputerowy umożliwia wspomaganie działań związanych z ocenianiem stanu technicznego maszyn. Celowe jest więc prowadzenie badań związanych z opracowaniem odpowiedniego oprogramowania. W pierwszym etapie wdrażania oprogramowanie byłoby stosowane wyłącznie do wspomagania działań służb diagnostycznych. Po odpowiedniej weryfikacji oprogramowania możliwe będzie jego udostępnienie również dyspozytorom maszyn.

Opracowywanie programów wspomagających proces wnioskowania w diagnostyce technicznej znacznie utrudnia brak ogólnych metod formalnego zapisu wiedzy diagnostycznej specjalisty (eksperta) oraz brak możliwości pełnej algorytmizacji procesu wnioskowania. Podejmowane próby zastosowania modeli statystycznych nie prowadzą i nie mogą prowadzić do zadowalających wyników, ponieważ w większości wypadków zgromadzone dane dotyczą obiektów jednostkowych (co uniemożliwia wyznaczenie wiarogodnych statystyk dla populacji rozpatrywanej grupy maszyn). Analizując różne sposoby przygotowywania omawianych programów należy zwrócić uwagę na tzw. systemy doradcze czyli "programy komputerowe stosujące wiedzę i procedury wnioskowania do rozwiązywania takich złożonych problemów, które wymagają udziału eksperta wtedy, gdy rozwiązywane są przez człowieka" [10]. Na badania związane z systemami doradczymi przeznaczane są w skali światowej ogromne nakłady finansowe. Szczególną cechą systemów doradczych jest to, że umożliwiają one gromadzenie wiedzy, na przykład w postaci zbiorów izolowanych stwierdzeń oraz reguł, bez konieczności jednoznacznego formułowania ogólnych twierdzeń lub hipotez. Systemy doradcze określane są w tekstach angielskich terminem "expert systems". Poszukując adekwatnego terminu w języku polskim rozpatrywano [7], [9] trzy nazwy: "systemy ekspertowe", "systemy rzeczoznawcze", "systemy doradcze". Ponieważ autorzy interesują się przede wszystkim zastosowaniem systemów doradczych do rozwiązywania problemów diagnostyki technicznej oraz wspomagania prac inżynierskich, postanowiono stosować nazwę "systemy doradcze".

Celem niniejszego artykułu jest zwięzłe przedstawienie istoty działania systemów doradczych oraz zwrócenie uwagi na problemy występujące podczas ich konstruowania i eksploatacji, ze szczególnym uwzględnieniem problemów reprezentacji wiedzy dla potrzeb diagnostycznych systemów doradczych.

---

* Artykuł publikujemy bez żadnych zmian, w formie nadesłanej przez Autorów

** Fragmenty opracowania związane są z pracami badawczymi, objętymi programami:
CPBP 02.15, CPBR 8.8, CPBR 3.14, CPBP 02.03

2. Istota działania systemów doradczych

Prace dotyczące systemów doradczych zainicjowane zostały w Stanford University i rozwijane były początkowo głównie w takich ośrodkach jak Carnegie-Mellon University, Massachusetts Institute of Technology oraz Stanford Research Institute. Obecnie szybko rośnie zainteresowanie tymi systemami, przy czym znajdują one zastosowanie głównie w diagnostyce medycznej, interpretacji wyników pomiarów, nadzorowaniu działania złożonych obiektów, wspomaganiu prac projektowo-konstrukcyjnych, kompletowaniu złożonych obiektów, przeprowadzaniu analiz ekonomicznych i wspomaganiu procesów podejmowania decyzji w dziedzinie ekonomii, automatycznym dowodzeniu twierdzeń itp.

Początkowo (lata sześćdziesiąte) wspomaganie procesu podejmowania decyzji formułowano jako zadanie poszukiwania rozwiązania optymalnego w zadanym zbiorze rozwiązań dopuszczalnych (możliwych). Większość badań dotyczyła poprawy numerycznej efektywności algorytmów wyszukiwania w rozpatrywanym zbiorze elementów, dla których określona funkcja kryterialna osiąga wartości ekstremalne. Przykładem znaczących osiągnięć w zakresie optymalizacji takich algorytmów są liczne prace związane z automatycznym grupowaniem i klasyfikacją obiektów. Stwierdzono, że postępowanie takie napotyka na wiele ograniczeń wynikających głównie z dopuszczalnych wielkości pola możliwych rozwiązań oraz z dopuszczalnych czasów wykonywania obliczeń.

Poważnym przełomem w sposobie rozwiązywania omawianych zadań było zastosowanie algorytmów nazywanych heurystykami, prowadzących do rozwiązania za pomocą ciągu działań ustalonych przy pomocy reguł, wynikających z wiedzy o obiektach badań oraz uwzględniających umiejętności specjalistów (ekspertów) z rozpatrywanej dziedziny. Termin "heurystyka" określa niekompletną metodę rozwiązywania zadania, dla którego to zadania nie jest znana zdeterminowana strategia postępowania. Niejednokrotnie przeprowadzenie pełnego dowodu poprawności heurystyk nie jest możliwe. Celowość ich użycia uzasadnia się skutecznością algorytmu sprawdzonego dla wybranych przykładów z jednoczesnym zastrzeżeniem, że dopuszcza się możliwość istnienia przykładu, dla którego heurystyka będzie działać nieskutecznie lub nawet błędnie.

Spotkać można różne konstrukcje systemów doradczych stosujących rozmaite sposoby reprezentacji wiedzy (rachunek predykatów, reguły, sieci semantyczne, sieci bayesowskie, scenariusze, ramy). Najczęściej stosowane są obecnie systemy, których działanie określa odpowiedni zbiór reguł i które składają się z następujących głównych elementów:

- zbiory:
  - baza wiedzy (np. zbiór reguł),
  - baza danych stałych (dane o obiekcie),
  - baza danych zmiennych (np. wyniki pomiarów, hipotezy),
  - stos zadań (zawierający wykaz działań programu zawieszonych ze względu na brak określonych stwierdzeń w bazach danych),
- procedury wnioskowania,
- procedury objaśniania,
- procedury sterowania dialogiem (procedury wejścia/wyjścia) umożliwiające formułowanie zadania przez użytkownika i przekazywanie rozwiązania przez program.

Zakłada się, że system doradczy powinien posiadać ponadto zbiór procedur umożliwiających rozszerzanie oraz modyfikację bazy wiedzy (co pozwala na występowanie użytkownika w roli "nauczyciela" systemu doradczego).

2.1. Bazy danych

Elementami baz danych są stwierdzenia opisujące to co zaszło lub zachodzi w rzeczywistości. Baza danych "stałych" zawiera ogólne informacje np. o strukturze obiektu i jego stanie. Może być ona aktualizowana na podstawie wartości zapisywanych w bazie danych "zmiennych". Baza danych "zmiennych" zawiera wyniki pomiarów i odpowiedzi użytkownika wprowadzane za pomocą procedur sterowania dialogiem oraz wyniki pośrednie (wnioski) wyznaczane przez procedury wnioskowania na podstawie znanych programowi danych "stałych" i danych "zmiennych".

Obie bazy danych są zbiorami stwierdzeń, zapisywanych na przykład w postaci następującej trójki

uporządkowanej (<ATRYBUT>, <OBIEKT>, <WARTOSC>) interpretowanej jako skrót zdania (stwierdzenia):

"cecha <ATRYBUT> obiektu <OBIEKT> przyjmuje wartość <WARTOSC>".

Obiektem wskazywanym przez nazwę <OBIEKT> może być konkret (rzecz) lub dowolna kategoria abstrakcyjna (pojęcie). Wartość cechy może być elementem dowolnego typu (np. 13, tak, biały, ...).

W licznych publikacjach dotyczących systemów doradczych spotkać można stosowanie nazwy "fakt" jako synonimu nazwy "stwierdzenie". Chcemy zwrócić uwagę na potrzebę wyraźnego rozróżniania tych dwóch pojęć. W spotykanych obecnie rozwiązaniach systemów doradczych stosowane są często zbiory reguł określających relacje między stwierdzeniami. Relacje te można interpretować jako model dziedziny, dla której opracowywana jest baza. Dziedzina ta, czyli określony fragment "rzeczywistości" może być przedmiotem poznania, którego celem będzie rozpoznanie (identyfikacja) występujących w niej "faktów". Pod pojęciem faktu rozumiemy to,co zaszło lub zachodzi w rzeczywistości, niezależnie od realizowanego procesu poznania. Stwierdzenia są wynikiem poznania i stanowią orzeczenia (subiektywne sądy) o faktach.

Z powyższych uwag wynika, że relacje między stwierdzeniami nie są równoważne relacjom między faktami. Celowe jest także zwracanie uwagi na to, że stwierdzenie orzeka o faktach istniejących niezależnie od stwierdzeń.

Zapisywanie stwierdzenia bez informacji o relacjach występujących między obiektami utrudnia, a w pewnych przypadkach uniemożliwia przeprowadzenie skutecznego wnioskowania. Dla opisu takich relacji stosowane są np. sieci semantyczne. Inną dogodną formą zapisu są tzw. ramy, umożliwiające gromadzenie zarówno wiedzy o charakterze deklaratywnym jak również proceduralnym. Sposoby reprezentacji wiedzy opisano w rozdz. 3.

2.2. Baza wiedzy

Zbiory stwierdzeń (zbiorami tymi są bazy danych) nie wystarczają do zapisania wiedzy. Istotnym składnikiem systemu doradczego jest baza wiedzy, zawierająca informacje o określonej dziedzinie. Wiedza diagnostyczna zapisywana jest często w postaci tzw. reguł, mających następującą postać:

REGUŁA xxx:	jeżeli	PRZESŁANKA
	to	KONKLUZJA

gdzie PRZESŁANKA jest wyrażeniem złożonym z prostych zdań logicznych połączonych funktorami "i", "lub". Określa ona warunki (wyznaczone na podstawie stwierdzeń), których spełnienie pozwala na przyjęcie konkluzji (czyli wpisanie do bazy danych "zmiennych" stwierdzenia odpowiadającego tej konkluzji).

Przykłady reguł (fragment bazy wiedzy systemu doradczego wspomagającego badania diagnostyczne sprężarek):

REGUŁA 19	Jeżeli	zostały wykryte objawy drgań wirowych w łożysku ślizgowym,
	i	temperatura łożyska jest obniżona,
	to	luz promieniowy w łożysku jest za duży.
REGUŁA 20	Jeżeli	zostały wykryte objawy drgań wirowych w łożysku ślizgowym,
	i	temperatura łożyska jest normalna,
	i	przekrój poprzeczny panewki łożyska jest kołowy,
	to	wskazana jest modyfikacja kształtu panewki łożyska.
REGUŁA 21	Jeżeli	poziom drgań korpusu łożyska ślizgowego jest podwyższony,
	i	nie zostały wykryte podstawowe objawy niewyrównoważenia wirnika,
	i	w widmach prędkości drgań korpusu łożyska występują składowe o częstotliwości od 0,2 do 0,45 częstotliwości wirowania wału,
	to	zostały wykryte objawy drgań wirowych w łożysku ślizgowym.

Znane są przykłady działających systemów doradczych (np. INTERNIST), w których wiedza o dzie-

dzinie została uwzględniona już w fazie układania programu i w których zbiór reguł nie występuje jako oddzielny, niezależny element systemu. Nie kwestionując znaczenia i skuteczności takich rozwiązań należy stwierdzić, że w chwili obecnej zapisywanie informacji o dziedzinie w postaci oddzielnej bazy wiedzy jest postępowaniem standardowym.

Aby w pełni oddzielić bank wiedzy danej dziedziny i uniezależnić procedury wnioskujące od informacji związanych z tą dziedziną, dokonano uogólnienia pojęcia reguły ([24], [6]). Uogólnienie to polega na przyjęciu, że konkluzja reguły jest opisem pewnego działania (które ma być wykonane przez program), a nie stwierdzeniem. Tak określoną regułę nazywa się regułą działania. Należy ją traktować jako przepis poprawnego postępowania. To pozornie drobne uogólnienie pozwala na znaczne rozszerzenie zakresu możliwych zastosowań reguł. W szczególności umożliwia ono formułowanie reguł dotyczących przebiegu procesu wnioskowania, tzn. umożliwia zapis strategii tego procesu (tzw. meta-reguły).

W rozdz. 3 omówiono inne sposoby zapisu wiedzy, które mają zastosowanie w diagnostycznych systemach doradczych. Stosuje się w tym celu tzw. ramy i sieci semantyczne. Oba te środki ułatwiają opisanie (zapis) informacji o strukturze diagnozowanego obiektu lub konstruowanego środka technicznego.

Wybór sposobu zapisywania wiedzy uzależniony jest od wielu czynników, z których do najważniejszych należy zaliczyć rodzaj wiedzy wymaganej dla poprawnego działania systemu oraz niezbędną wielkość bazy wiedzy. Szczególne znaczenie ma unikanie redundancji (nadmiarowości) informacji w tej bazie. Najczęściej stosowane są różne mieszane sposoby zapisu wiedzy, wywodzące się ze sposobów wyżej opisanych.

Należy zwrócic uwagę, że przyjęta struktura systemu doradczego stanowi jakościową zmianę w zakresie programowania komputerów. We wcześniej opracowywanych programach informacje o dziedzinie były uwzględniane w fazie określania algorytmu wnioskowania. Praktycznie nie była więc możliwa prosta adaptacja programu opracowanego w ramach jednej dziedziny do zastosowań w innej (odmiennej) dziedzinie, gdyż wymagałaby ona daleko idących zmian w tym algorytmie, których przeprowadzenie byłoby możliwe (w zasadzie) jedynie przez autora programu.

2.3. Informacje niepewne i niedokładne

Systemy doradcze przeznaczone są głównie do wspomagania prac prowadzonych w dziedzinach, dla których nie opracowano zwięzłych, ścisłych teorii. Przykładem takiej dziedziny jest diagnostyka techniczna maszyn, w której występuje potrzeba bazowania na niejednoznacznych (subiektywnych) opiniach specjalistów, wynikająca z braku możliwości zbudowania kompletnego modelu opisującego działanie badanych obiektów i określającego jednoznacznie relacje między stanem technicznym i cechami obserwowanych sygnałów diagnostycznych.

Dostępne dla systemu doradczego informacje są wynikiem działalności myślowej człowieka i/lub wynikiem procesu poznania, są więc niepewne, ważone, niedokładne i/lub rozmyte. Proces rozumowania realizowany przez człowieka ma zazwyczaj charakter przybliżony, nie zawsze dający się opisać modelami klasycznej logiki formalnej. Reguły określające działanie systemu doradczego obowiązują jedynie w większości znanych lub zaobserwowanych przypadków. Sposoby uwzględniania wymienionych "niepewności" oraz "niedokładności" są tematem licznych prac dotyczących tzw. rozumowania przybliżonego, w których wykorzystuje się najczęściej metody teorii zbiorów rozmytych i rozpatruje między innymi:

- wartości logiczne (np. prawda, fałsz) określane jako liczby rozmyte na odcinku [0,1],
- lingwistyczne wartości logiczne należące do przeliczalnego zbioru tzw. zmiennych lingwistycznych, takich jak "bezwzględnie prawdziwe", "bardzo prawdziwe", "dosyć prawdziwe", "nieokreślone", "dość fałszywe",
- modyfikacje klasycznych reguł wnioskowania.

Ze względu na potrzeby systemów doradczych rozumowanie przybliżone rozpatrywane jest jako postępowanie polegające np. na zastosowaniu:

- dokładnych reguł i przybliżonych przesłanek (stwierdzeń),
- przybliżonych reguł i dokładnych przesłanek,
- przybliżonych reguł i przybliżonych przesłanek.

Podejmując decyzję o zastosowaniu rozumowania przybliżonego w konstruowanym systemie doradczym, należy ściśle ustalić sposoby określania:

- niepewności stwierdzeń przybliżonych,
- niepewności przybliżonych reguł wnioskowania,
- niepewności przesłanek złożonych z kilku stwierdzeń przybliżonych,
- niepewności konkluzji wynikających z przybliżonych przesłanek i przybliżonych reguł,
- niepewności konkluzji wyznaczanej za pomocą kilku niezależnych reguł przybliżonych.

Nie opracowano jednak ogólnej metody, która byłaby przyjmowana powszechnie przez konstruktorów systemów doradczych. Często stosowanym rozwiązaniem jest wprowadzenie stopnia pewności [32], [36], będącego liczbą rzeczywistą z przedziału [0,1] przypisaną każdemu stwierdzeniu. Wartość 1 oznacza stwierdzenia w pełni prawdziwe, a wartość 0 oznacza stwierdzenia w pełni fałszywe. W podobny sposób stopień pewności może być przyporządkowany regułom.

Opierając się wyłącznie na rozważaniach heurystycznych [36] można przyjąć następujące zasady wykonywania działań na stopniach pewności:

- Stopień pewności przesłanki będącej koniunkcją kilku stwierdzeń jest równy minimalnej wartości stopni pewności tych stwierdzeń. Przyjęcie takiej zasady tłumaczone jest za pomocą modelu łańcucha, którego wytrzymałość jest zależna od wytrzymałości najsłabszego ogniwa.
- Reguły rozpatrywane są jako "układy wzmacniające" o wzmocnieniu (nie większym niż 1) równym stopniowi pewności reguły. Na wejściu tych "układów" występuje stopień pewności przesłanki, a na wyjściu stopień pewności konkluzji.
- Jeżeli jakąś konkluzję otrzymano w wyniku zastosowania kilku niezależnych reguł, to jej stopień pewności jest maksymalną wartością stopni pewności konkluzji tych reguł.

Opisane zasady wykonywania działań na stopniach pewności mogą być przyczyną wielu intuicyjnych wątpliwości. Główna wątpliwość związana jest z utożsamianiem w tym podejściu wartości logicznej stwierdzenia (prawda, fałsz) z jego stopniem pewności. Liczne wątpliwości związane są z przyjmowanymi modelami propagacji niepewności i niedokładności w układzie wnioskującym [20]. Opisany sposób uwzględniania propagacji niepewności przez reguły prowadzi do utożsamiania pojęć "niepewny" i "fałszywy".

Zdaniem autorów, do bardziej poprawnych metod jednoczesnego uwzględniania niepewności i niedokładności stwierdzeń oraz reguł należy zaliczyć metody rozpatrujące stwierdzenia jako zbiory lub wartości rozmyte, a reguły jako relacje rozmyte [27], [15], [19]. Metody te umożliwiają między innymi rozpatrywanie dolnych i górnych ograniczeń stopni wiarogodności stwierdzeń [31]. Uwzględniając definicje przyjęte w logice modalnej określające, że uznanie stwierdzenia "A" jest konieczne wtedy i tylko wtedy, gdy nie jest możliwe uznanie stwierdzenia "nie A", wprowadzono miary konieczności i możliwości stwierdzenia A. Postępując podobnie można przyjąć [22], że dla opisu reguły przybliżonej wymagana jest znajomość jej stopnia konieczności oraz stopnia dostateczności. Miara dostateczności określa stopień, w jakim przesłanka w pełni prawdziwa wystarcza na to, by konkluzja mogła być uznana jako prawdziwa. Analogicznie miara konieczności reguły określa stopień, w jakim będzie konieczna prawdziwa przesłanka, wtedy gdy konkluzja jest prawdziwa. Znane są sposoby (np. [22], [8]) modyfikacji klasycznych schematów wnioskowania (modus ponens, modus tollens) do postaci pozwalającej na uwzględnianie miar możliwości i konieczności stwierdzeń oraz miar dostateczności i konieczności reguł. Dokładniejszy opis metod rozumowania przybliżonego (dla potrzeb systemów doradczych) zamieszczono w [9].

2.4. Procedury wnioskowania

Zbiory reguł są w większości systemów doradczych zbiorami licznymi. Ich stosowanie do przekształcania bazy danych wymaga odpowiedniej strategii postępowania. W celu uniknięcia nadmiernego wzrostu liczby elementów bazy danych "zmiennych" stosowane są różne sposoby ograniczania liczności zbioru reguł przeznaczonych do analizy.

Najprostszym sposobem wnioskowania (wnioskowanie progresywne) jest postępowanie polegające na wyznaczaniu konkluzji ze znanych przesłanek za pomocą uwzględnianej reguły, a następnie rozpatrywanie tej konkluzji jako przesłanki dla kolejnej reguły itd. W przypadku, gdy informacje zgromadzone w bazach danych nie pozwalają na określenie wartości logicznej przesłanek, system doradczy zadaje odpowiednie pytania użytkownikowi. Niedogodnością takiego postępowania jest konieczność badania w kolejnych krokach konkluzji wszystkich kroków poprzednich, co prowadzić może do "lawinowego" wzrostu koniecznych do rozpatrzenia przesłanek.

Działaniem odwrotnym (wnioskowanie regresywne, wstecz) jest postępowanie polegające na wyznaczaniu przesłanek z zakładanych konkluzji (mających wtedy znaczenie hipotez) za pomocą uwzględnianej reguły, a następnie rozpatrywanie tych przesłanek jako konkluzji kolejnych reguł. Niedogodnością takiego postępowania jest to, że w przypadku rozpatrywania przesłanek złożonych będących iloczynem logicznym przesłanek prostych, następuje szybki wzrost liczby analizowanych przesłanek, z których każda musi być badana indywidualnie. Działający regresywnie interpretator reguł (w systemie doradczym) musi uwzględniać możliwość występowania badanej hipotezy jako konkluzji kilku niezależnych reguł. Poważnym dodatkowym utrudnieniem jest konieczność zabezpieczenia realizowanego przez system doradczy procesu wnioskowania przed skutkami konfliktów wywołanych tym, że wynikiem działań mogą być między innymi zmiany wartości logicznych uwzględnianych wcześniej przesłanek reguł. Obie opisane metody postępowania należy uznać za nieodpowiednie wtedy, gdy łańcuch relacji niezbędnych do przekształcenia znanej przesłanki w poszukiwaną (badaną) konkluzję końcową jest długi. W takich przypadkach proponuje się zastosowanie metody mieszanej, czyli tzw. kroczenia do przodu i wstecz.

Brak jest ogólnej teorii umożliwiającej optymalizację wyboru sposobu działania (progresywny/regresywny) procedur wnioskujących. Stwierdza się jedynie, że działanie regresywne jest zalecane ze względu na czas analizy wtedy, gdy zadanie polega na weryfikacji jednej określonej hipotezy. Dla zwiększenia efektywności procedur wnioskujących celowym jest dokonywanie podziału rozpatrywanego zadania na podzadania. Jest to jednak związane z dodatkowymi utrudnieniami wynikającymi z istniejących w praktyce powiązań między takimi podzadaniami.

2.5. Procedury objaśniania

Należy zaznaczyć, że ze względu na brak w chwili obecnej odpowiednich ustaleń prawnych, nie został rozwiązany problem odpowiedzialności za decyzje podjęte wskutek konsultacji użytkownika z systemem doradczym. Obecnie zakłada się, iż odpowiedzialność za takie decyzje ponosi wyłącznie użytkownik systemu - osoba podejmująca decyzję. Dostawca systemu doradczego powinien poinformować o tym użytkownika. Oznacza to jednocześnie, że w systemie doradczym należy przewidzieć środki, które pozwolą na uzasadnianie poprawności realizowanego przez ten system procesu wnioskowania. W praktyce uzasadnianie poprawności działania realizowane jest przez procedury objaśniające. Przeznaczone są one do udzielania odpowiedzi na:

- pytania ogólne nie związane z przebiegiem dialogu, dotyczące treści określonych reguł (np. treści reguł spełniających podany warunek),
- pytania (np. dlaczego?, z czego wynika, że ... ?) związane z przebiegiem dialogu, dotyczące reguł używanych przez program w celu uzyskania wniosków,
- pytania o występowanie określonych relacji między danymi,
- pytania dotyczące znaczenia terminów stosowanych przez system doradczy w dialogu z użytkownikiem,
- pytania o wyjaśnienie przyczyn zadawania określonych pytań przez system doradczy,
- pytania (np. co będzie jeżeli ...?) o skutki jakie wynikną z udzielenia przez użytkownika określonej odpowiedzi na pytanie zadane wcześniej przez system,
- pytania o wskazanie odpowiedzi (udzielonych wcześniej przez użytkownika), których zmiana może spowodować znaczącą zmianę otrzymanej konkluzji.

Odpowiedź ustalana jest w wyniku przeszukiwania bazy wiedzy i/lub baz danych według zadanych słów kluczowych, a następnie składana z pewnych standardowych fragmentów tekstów. Aby system doradczy mógł szybko i skutecznie uzasadnić poprawność uzyskanych wyników oraz celowość zadawanych pytań, wskazane jest zapisywanie w bazie danych "zmiennych" tzw. śladu procesu wnioskowania, czyli kolejności w jakiej zastosowano reguły, odpowiedzi udzielonych przez użytkownika, danych pobranych z innych źródeł itp.

Należy zwrócić uwagę na to, że system doradczy może objaśniać swoje działanie wyłącznie w zakresie informowania o przebiegu realizowanego procesu wnioskowania. System taki w zasadzie nie sprawdza poprawności reguł zapisanych w bazie wiedzy (zadania tej klasy rozwiązywane są przy użyciu pewnych specjalnych narzędzi programowych w stadium weryfikacji działania danego systemu doradczego, opisanym w dalszej części). Aby umożliwić użytkownikowi ocenę poprawności zastosowanych reguł lub zwiększyć jego zaufanie do rozwiązania sugerowanego przez regułę zaleca się modyfikację reguł do

następującej postaci:

REGUŁA xxx	jeżeli	PRZESŁANKA
	to	KONKLUZJA
	ponieważ	UZASADNIENIE

gdzie UZASADNIENIE jest tekstem informującym na przykład o źródle (autor, publikacja itp.), na podstawie którego sformułowano regułę.

3. Reprezentacja wiedzy

Zakres znaczeniowy terminu "reprezentacja wiedzy" przyjmowany jest dość dowolnie przez różnych autorów. Termin ten używany będzie do określania niezależnego od rozpatrywanej informacji, ogólnego formalizmu przekazywania, zapisywania i gromadzenia dowolnego zasobu wiedzy. Wymaga się, aby ten zapis był prosty, kompletny (wyczerpujący), zwięzły, zrozumiały i wyraźny (tzn. nie zawierający elementów domyślnych i wieloznacznych). Kryterium jednoznaczności wynika np. z potrzeby podobnego pojmowania użytych pojęć oraz ich zakresów znaczeniowych przez autorów bazy wiedzy i użytkowników systemu doradczego wykorzystującego tę bazę. Dla ułatwienia sposobu korzystania z zapisu powinien on ujawniać występujące ograniczenia.

Upraszczając zagadnienie można stwierdzić, że omawiany zapis wiedzy winien umożliwić identyfikację (nazwanie) określonych obiektów rozpatrywanej dziedziny oraz umożliwić identyfikację (nazwanie) relacji między tymi obiektami. Rozpatrywanymi obiektami mogą być zarówno obiekty proste jak i obiekty złożone, które mogą być konkretami (rzeczami) lub pojęciami abstrakcyjnymi.

Formalizacja sposobów reprezentacji wiedzy jest problemem trudnym, który do chwili obecnej nie znalazł dostatecznie ogólnego rozwiązania. Zwięzłe opisy problemów występujących w tym zakresie można znaleźć np. w [2], [9]. Stwierdzono [2], że przyjęcie określonej techniki zapisu wiedzy należy poprzedzić ustaleniem przedmiotu zapisu, którym mogą być: obiekty indywidualne, własności obiektów, relacje między obiektami, klasy obiektów, rodzaje zdarzeń, konkretne zdarzenia, motywy i przyczyny zdarzeń, skutki zdarzeń, twierdzenia itp. Zazwyczaj poszczególne "jednostki wiedzy" z danej dziedziny są powiązane ze sobą na różne sposoby, odzwierciedlające związki tematyczne, zakresy pojęć itp. Występuje również potrzeba zapisywania informacji warunkowych, informacji sprzecznych oraz informacji dotyczących wyjątków.

Uogólniając stosowane obecnie techniki, można wyróżnić następujące dwa typy reprezentacji wiedzy:

- reprezentacja proceduralna, polegająca na określeniu zbioru procedur, których działanie reprezentuje wiedzę o dziedzinie (np. procedura obliczania objętości bryły, program wyznaczania naprężeń kryterialnych w połączeniu kołnierzowym metodą elementów skończonych),
- reprezentacja deklaratywna, polegająca na określeniu zbioru specyficznych dla rozpatrywanej dziedziny faktów, stwierdzeń i reguł (np. katalog łożysk tocznych zapisany w postaci bazy danych).

Zaletą reprezentacji proceduralnej jest wysoka efektywność reprezentowania procesów. Z kolei reprezentacja deklaratywna jest bardziej "oszczędna" (każdy fakt lub reguła zapisywane mogą być tylko jeden raz) i łatwiejsza jest jej formalizacja. Wiedza o dziedzinie zapisana w postaci proceduralnej uwzględniana jest często już w fazie układania programu - wówczas baza wiedzy nie występuje jako oddzielny, niezależny element systemu doradczego. Zapis informacji o dziedzinie w postaci deklaratywnej wiąże się z wyodrębnieniem bazy wiedzy jako oddzielnej struktury, co ułatwia późniejsze modyfikacje. Jeśli baza wiedzy nie została wydzielona jako niezależny od programu element systemu doradczego, dokonanie jakiejkolwiek modyfikacji wymaga najczęściej wprowadzenia zmian do programu, czyli jest zadaniem praktycznie niewykonalnym przez użytkownika programu. Jako rozwiązanie optymalne można uznać reprezentację łączącą w sobie cechy reprezentacji proceduralnej i deklaratywnej. Przykładami takiej reprezentacji są opisane dalej ramy oraz języki obiektowe.

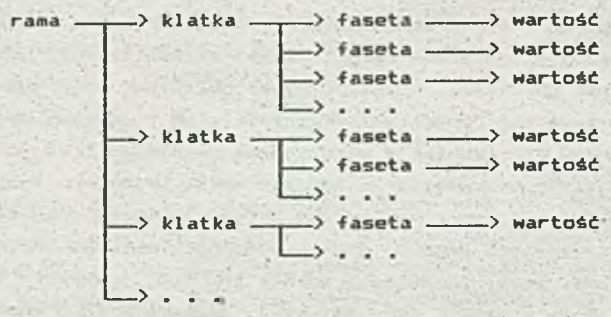
3.1. Ramy i klasy

Koncepcję ram jako sposobu zapisu wiedzy zaproponował Minsky [23]. Znane są liczne wersje zastosowania ram, różniące się szczegółami i używaną terminologią (np. [14], [16], [12]). Pierwszym większym systemem programów stosujących ramy był FRL (Frame Representation Language) [28]. Wprowa-

dzono w nim nową technikę zapisywania własności zastępującą stosowane wcześniej pary <atrybut, wartość>. Prace nad tym sposobem reprezentacji wiedzy prowadzone są dwoma drogami. Część autorów prac wychodzi z założenia, że proponowany język zapisu wiedzy powinien spełniać szereg jednoznacznie określonych założeń oraz powinien się cechować możliwością pełnej jego formalizacji (np. w kategoriach logiki). Wynikiem takich prac jest język KL-ONE [5] (i języki podobne, takie jak NIKL, KRYPTON [4]). Większość prac prowadzonych w tym zakresie nie obejmuje jednak szczegółowej, formalnej analizy semantycznej i wychodzi wyłącznie z przesłanek heurystycznych [11]. Ich wynikiem są na przykład takie rozbudowane systemy jak KEE [16], HP-RL [21] i PC-Plus [26].

3.1.1. Ramy

Rama stanowi semantyczną definicję wybranego pojęcia, nazywanego dalej obiektem (którym może być abstrakt lub konkret). Rama jest strukturą posiadającą nazwę. Jest ona złożoną z dowolnej liczby klatek nazywanych również szczelinami (ang. slot), które z kolei są zbiorami elementów klatek (angielski termin "facet" oznaczający element klatki nie jest tłumaczony jednoznacznie na język polski). Dalej używane będą pojęcia "klatka" i "faseta".



Klatki i fasety identyfikowane są za pomocą ich nazw. Nazwa klatki określa wybrany atrybut (nazwę cechy) obiektu. Własności obiektu (wartości cech) określane są przez elementy klatki (czyli fasety) oraz przysługujące im wartości. Nie wprowadza się żadnych ograniczeń dla typów tych wartości - mogą być nimi liczby, teksty, piktogramy lub na przykład kolejne ramy. Każdej klatce przyporządkowany jest określony (stały) zbiór faset, które mogą być dowolnie uzupełniane fasetami wynikającymi z potrzeb specyficznego zastosowania.

Przykład fragmentu ramy opisującej stalową śrubę dokładną o długości 70 mm, z łbem sześciokątnym, gwintem metrycznym M12, o własnościach mechanicznych klasy 5.6, wykonaną zgodnie z PN-74/M-82105 , która jest elementem oznaczonym (w wykazie elementów danego zespołu) numerem 123:

```
(OBIEKT   'SRUBA-123'
          (TYP
                    (VALUE          'PN-74/M-82105' ) )
          (DŁUGOSC
                    (VALUE          70.0 ) )
          (GWINT
                    (VALUE          'M' ) )
          (SREDNICA
                    (VALUE          12.0 ) )
          (TWORZYWO
                    (VALUE          5.6 ) ) )
```

Rama ta posiada nazwę "SRUBA_123" i zawiera klatki "TYP", "DŁUGOSC", "GWINT", "SREDNICA" oraz "TWORZYWO". Występujące w przykładzie fasety o nazwach VALUE są jednymi z wielu możliwych. Ich zawartości są interpretowane jako wartości cech określanych przez klatki. Przykładami nazw innych faset, które mogą występować w klatkach ramy są [29], [16]:

- COMMENT określająca fasetę zawierającą tekst będący opisem klatki (np. dla potrzeb dokumentacyjnych),
- VALUE_CLASS określająca fasetę zawierającą definicję (identyfikator) typu wartości fasety VALUE,
- CARDINALITY określająca fasetę zawierającą liczbę faset typu VALUE występujących w klatce (np. z wartością domyślną = 1),

- DEFAULT określająca fasetę zawierającą wartość domyślną przyjmowaną wtedy, gdy faseta VALUE nie jest określona,
- IF_ADDED oraz IF_REMOVED określająca fasetę "zawierającą" procedury uruchamiane w chwili modyfikowania zawartości fasety VALUE; procedury te posiadają co najmniej trzy argumenty (nazwa ramy, nazwa klatki, wartość),
- REQUIRE określająca fasetę "zawierającą" procedurę testującą poprawność określenia odpowiedniej fasety o nazwie VALUE (zabezpiecza ona przed wprowadzaniem do ramy wartości niedopuszczalnych),
- IF_NEEDED określająca fasetę "zawierającą" listę procedur, które należy wywołać w celu uzyskania wartości cechy definiowanej przez klatkę (argumentami tych procedur są co najmniej: nazwa ramy i nazwa klatki); procedury te uwzględniają stosowany mechanizm dziedziczenia wartości klatek (por. rozdz.3.1.3); są one wywoływane w kolejności ustalonej przez tę listę aż do uzyskania wartości określonej (uzyskanie wartości nieokreślonej nie kończy ciągu wywołań).

Przykładem zastosowania fasety o nazwie IF_NEEDED są kolejne klatki ramy "SRUBA_123":

```
(NAZWA
        (VALUE_CLASS    String )
        (IF_NEEDED      funkcja1 ) )
(SREDNICA_ZEWNETRZNA
        (VALUE_CLASS    Real )
        (IF_NEEDED      funkcja2 ) )
```

gdzie

- "funkcja1" jest procedurą określającą nazwę śruby na podstawie wartości zapisanej w klatce "TYP" (w omawianym przykładzie nazwą tą będzie tekst 'śruba z łbem sześciokątnym'),
- "funkcja2" jest procedurą określającą (zgodnie z odpowiednią normą) średnicę zewnętrzną gwintu; parametrami wejściowymi tej procedury są wartości zapisane w klatkach "GWINT" oraz "SREDNICA" ramy "SRUBA_123".

Zauważmy, iż wprowadzenie fasety "IF_NEEDED" pozwala na eliminację zbędnych obliczeń. Przykładowo masa elementu będzie obliczana dopiero wtedy, gdy wystąpi potrzeba jej odczytu. Dla ograniczenia zbędnych powtórzeń wykonywanych działań (np. obliczanie masy na podstawie objętości i gęstości przy każdorazowym jej odczycie) wynik działań procedur fasety "IF_NEEDED" może być zapisywany w odpowiedniej fasecie "VALUE" - wymaga to jednak kasowania tej fasety przy każdej zmianie wartości gęstości lub objętości. Zadania takie wykonywane są przez procedury występujące w fasetach "IF_REMOVED" klatek zawierzających wartość gęstości i wartość objętości. Zakres ważności definicji tych procedur określony jest w taki sposób, że nie są one bezpośrednio dostępne programowi, wykorzystującemu ramę. Dla podkreślenia tej szczególnej własności procedury te nazywane są często demonami. Optymalizując sposoby definiowania klatek można przyjąć rozwiązanie polegające na tym, że demony kasujące wartość masy będą wstawiane dynamicznie do odpowiednich klatek przez procedurę obliczającą masę (procedurę fasety "IF_NEEDED") oraz, że demony te będą się same kasowały po ich wykonaniu.

Koncepcja ram wiążąca dane z procedurami przeznaczonymi do określania lub modyfikacji wybranych wartości upraszcza programy korzystające z takich struktur. Np. szczegółowa weryfikacja wprowadzanych danych dokonywana jest wówczas przez procedury faset "IF_NEEDED".

Podstawowymi działaniami wykonywanymi przez programy korzystające z reprezentacji wiedzy w postaci ram są:

- egzemplifikacja (omówiona w rozdz.3.1.2) i usuwanie ram,
- podstawianie, odczytywanie i usuwanie elementów klatek (wartości i demonów).

Omawiana struktura (rama) umożliwia deklaratywną i proceduralną reprezentację wiedzy. Wprowadzając pojęcie ramy stwierdzono, że stanowi ona semantyczną definicję wybranego pojęcia określanego jako obiekt, gdzie pojęciem tym może być abstrakt lub konkret. Obiektami abstrakcyjnymi mogą być na przykład reguły i relacje [13], [17], [16]. Dowodem wysokiego stopnia ogólności omawianej struktury jest to, że szczególnym przypadkiem obiektu abstrakcyjnego mogą być działania wykonywane przez program (np. rama ZADANIE).

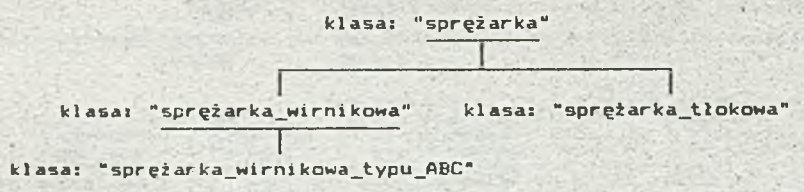
3.1.2. Klasy

Przedstawiony przykład dotyczył ramy opisującej jedną śrubę o szczegółowo określonych cechach konstrukcyjnych. Należy podkreślić, że rama ta jest elementem licznego zbioru ram opisujących różne śruby. We wszystkich tych ramach powtarzać się będą wybrane ich elementy (np. procedura obliczania średnicy zewnętrznej gwintu, oznaczenie normy). Dla ograniczenia redundancji w zbiorze ram wprowadzono pojęcie klasy obiektów. Klasa określona jest ramą-wzorcem, na podstawie której w wyniku działania nazywanego egzemplifikacją generowane mogą być ramy opisujące obiekty należące do klasy. W tak generowanych ramach wprowadza się dodatkową klatkę o nazwie "AKO" (is a kind of) przeznaczoną do przechowywania wskazania ramy-wzorca (ramy opisującej klasę). Dla umożliwienia dynamicznych modyfikacji zbioru ram, w ramie opisującej klasę powinny być zapisywane nazwy (identyfikatory) wszystkich ram, które powstały w wyniku jej egzemplifikacji. Nazwy takie zapisywane są w klatce "INSTANCES". Zapis nazw realizowany jest przez procedurę fasety "IF_ADDED" klatki "AKO". Przykładem fragmentów ramy-wzorca dla pokazanej wcześniej ramy opisującej śrubę jest:

```
(KLASA    SRUBY
          (TYP
                    (VALUE_CLASS    String )
                    (VALUE          unknown )
                    (COMMENT        'Symbol normy') )
          (DŁUGOSC
                    (VALUE_CLASS    Real )
                    (VALUE          unknown )
                    (COMMENT        'Długość w [mm]' ) )
          (GWINT
                    (VALUE_CLASS    String )
                    (IF_NEEDED      funkcja3 ) )
          (SREDNICA
                    (VALUE_CLASS    Real )
                    (VALUE          unknown ) )
          (TWORZYWO
                    (VALUE_CLASS    Real )
                    (VALUE          unknown )
                    (COMMENT        'Klasa własności' ) )
          (NAZWA
                    (VALUE_CLASS    String )
                    (IF_NEEDED      funkcja1 ) )
          (SREDNICA_ZEWNETRZNA
                    (VALUE_CLASS    Real )
                    (IF_NEEDED      funkcja2 ) ) )
```

gdzie
- "unknown" oznacza element nieokreślony,
- "funkcja3" jest procedurą określającą rodzaj gwintu na podstawie wartości zapisanej w klatce "TYP".

Klasy mogą być porządkowane, tworząc struktury hierarchiczne, przedstawiane na przykład za pomocą grafów. Oto przykład hierarchicznej struktury klas:



W hierarchicznych strukturach klas wyróżnia się klasy nadrzędne (tzw. rodziców) oraz klasy podrzędne (tzw. dzieci). W pokazanym przykładzie "sprężarka_wirnikowa" posiada klasę nadrzędną "sprężarka" i klasę podrzędną "sprężarka_wirnikowa_typu_ABC". Obiekty klasy "sprężarka_wirnikowa" będą posiadały między innymi własności określone dla odpowiednich obiektów nadrzędnych z klasy "sprężarka". Klasa może posiadać wiele klas nadrzędnych, może także posiadać wiele klas podrzędnych, z czego wynika, że graf reprezentujący hierarchiczną strukturę klas może mieć wiele korzeni. Dla określenia

położenia klasy w takim grafie wprowadza się w ramie opisującej klasę klatkę o nazwie "VIEW" [14], w której zapisywana jest lista nazw bezpośrednich klas nadrzędnych. Podczas egzemplifikacji takiej ramy lista ta może być zastępowana listą nazw obiektów nadrzędnych. Wprowadzenie takich struktur pozwala na definiowanie mechanizmów zagnieżdżania i dziedziczenia klas. Są to niezwykle ważne i ciekawe narzędzia.

3.1.3. Dziedziczenie

Wprowadzenie mechanizmów dziedziczenia pozwala na określanie hierarchii klas, która implikuje odpowiadającą jej hierarchię obiektów. Umożliwiają one znaczne ograniczenie redundancji baz wiedzy. Zachowana zostaje jednocześnie możliwość uwzględniania wyjątków, co jest na przykład zadaniem bardzo trudnym wtedy gdy reprezentacja wiedzy bazuje wyłącznie na koncepcji struktur hierarchicznych (bez indywidualnych mechanizmów dziedziczenia). W najprostszym podejściu zakłada się, że obiekt dziedziczy wszystkie te własności, które zostały określone dla obiektów nadrzędnych. Oznacza to, że rama opisująca obiekt posiada (domyślnie) wszystkie klatki ram opisujących obiekty nadrzędne. Dziedziczone są zarówno atrybuty (nazwy klatek) jak i własności (wartości faset oraz demony). W celu umożliwienia uwzględniania przypadków szczególnych wprowadza się różne wersje dziedziczenia. Identyfikator typu aktualnej wersji mechanizmu dziedziczenia zapisywany jest indywidualnie dla każdej klatki w fasecie o nazwie INHERITANCE. Faseta ta zawierać może na przykład jedną z następujących wartości [16]:

- OVERRIDE oznacza, że w rozpatrywanej klatce kopiowane są (domyślnie) listy wartości klatek dziedziczonych (rodziców), wtedy gdy lista wartości określonych lokalnie jest pusta (wartości nie zostały lokalnie określone); oznacza to, że lokalne określenie fasety "VALUE" pozwala na "zasłonięcie" odpowiedniej fasety dziedziczonej klatki,
- UNION oznacza, że do listy wartości określonych dla lokalnej klatki dopisywane są elementy list wartości klatek dziedziczonych,
- RUNION prowadzi do działań takich jak dla wartości UNION, z tą różnicą, że wynikowa lista wartości posiada odwrócony porządek elementów,
- SAME oznacza, że wartość określana lokalnie dla klatki musi być równa wartości dziedziczonej, a jeśli jest różna od wartości dziedziczonej to jest interpretowana jako wartość nieokreślona,
- UNIQUE blokuje dziedziczenie wartości klatki i oznacza, że wartość ta musi być określana lokalnie, a w przypadku jej braku jest ona interpretowana jako nieokreślona.

Rozbudowane mechanizmy dziedziczenia pozwalają również na określanie zakresu ważności każdej klatki. Rodzaj zakresu identyfikowany może być następującymi wartościami fasety o nazwie "SCOPE":

- CLASS oznacza, że wszystkie elementy klatki mogą być dziedziczone przez ramy podrzędne,
- OWN oznacza, że klatka nie może być dziedziczona (jest ona niewidoczna dla ram podrzędnych).

3.1.4. Zagnieżdżanie

Należy zwrócić uwagę na fakt, że ramy mogą opisywać obiekty definiowane na różnych stopniach szczegółowości. Oznacza to, że rama może na przykład opisywać obiekt "przekładnia zębata". Definicja takiej ramy będzie zawierała odwołania do innych ram ("korpus","łożysko","wał","uzębienie" itd). Skutkiem wprowadzenia takiej ramy będzie możliwość odwoływania się do niej innych ram, opisujących obiekty w których przekładnia jest wyłacznie jednym z wielu elementów. Prowadzi to do tzw. zagnieżdżania ram [18]. Mechanizm zagnieżdżania różni się od opisanego wcześniej mechanizmu dziedziczenia. Główne różnice objawiają się w procesie egzemplifikacji obiektów.

3.1.5. Języki obiektowe

Przyczyną podjęcia prac nad tzw. językami obiektowymi (ang. object-oriented programming) był wzrastający stopień złożoności programów, opracowywanych w sposób klasyczny, a także związane z tym trudności wynikające z potrzeby odpowiedniego pielęgnowania oprogramowania.

Podstawowymi pojęciami występującymi w językach obiektowych są: "obiekt", "klasa obiektów", "komunikat", "metoda". Istotna zmiana wprowadzona przez języki tej grupy polega na innym sposobie formułowania zadań i rezygnacji z "klasycznego" rozpatrywania oprogramowania jako połączenia okre-

ślonego zbioru algorytmów i struktur danych. Zmiana polega na tym, że zadania określane są przy założeniu iż dziedzina, której one dotyczą, może być modelowana za pomocą komunikujących się obiektów (obiektów wysyłających i odbierających komunikaty). Pozwala to na taką organizację pracy twórczej autorów oprogramowania, że proces ich rozumowania przebiega od zagadnień ogólnych do szczegółów.

Obiekty występujące w omawianym modelu są zdolne do odbierania komunikatów i mogą być interpretowane jako lokalny (niedostępny bezpośrednio z zewnątrz) obszar pamięci wyposażony w odpowiednie "łącza" z otoczeniem. Wynikiem odebrania przez obiekt zrozumiałego komunikatu może być zmiana stanu tego obiektu, jak również wysłanie przez ten obiekt komunikatu do innego obiektu. Odbieranym komunikatom przyporządkowane są tzw."metody" określające sposób reagowania obiektu. "Metoda" jest procedurą wywoływaną (przez obiekt) po odebraniu określonego komunikatu.

Protokoły przekazywania komunikatów definiowane są globalnie. Dostęp do lokalnych danych zawartych w obiekcie może następować wyłącznie za pośrednictwem komunikatów. Szczególnym przypadkiem obiektu może być rama. Obiekty charakteryzujące się podobnym działaniem mogą być grupowane w klasy. Na zasadach podobnych jak to ma miejsce w przypadku ram, klasy takie mogą być porządkowane w struktury hierarchiczne z odpowiednimi mechanizmami dziedziczenia "metod" co pozwala na unikanie wielokrotnego występowania takich samych procedur. Wprowadzenie klas wymaga określenia odpowiednich mechanizmów egzemplifikacji obiektów będących elementami tych klas.

W rodzinie języków obiektowych należy wymienić w pierwszym rzędzie interpreter języka SmallTalk. Ciekawymi z punktu widzenia ewentualnych zastosowań, są uzupełnienia języka C (np.[33]). Należy zwrócić uwagę na to, że stosowanie języków obiektowych napotyka na trudności w systemach wielodostępnych [34].

3.2. Sieci stwierdzeń i sieci neuronowe

Ciekawą techniką reprezentacji wiedzy są losowe sieci stwierdzeń, nazywane również sieciami bayesowskimi. Stanowią one uogólnienie sieci semantycznych (por. np. [3], [9]) stosowanych do definiowania pojęć i będących zbiorami węzłów i gałęzi ilustrujących związki między węzłami. Sieci stwierdzeń można rozpatrywać także jako rozwinięcie koncepcji perceptronu [30]. Sieci te zostały zaproponowane [25] jako model procesu rozumowania realizowanego przez człowieka, wtedy gdy dokonuje on uogólnień i wyciąga wnioski z dużej liczby danych, tworząc określoną, spójną interpretację tych danych.

3.2.1. Sieci stwierdzeń

Załóżmy, że wiedza o wybranej dziedzinie jest reprezentowana za pomocą zbioru stwierdzeń i relacji między nimi. Zbiór taki może być przedstawiany za pomocą tzw. sieci stwierdzeń czyli grafu, którego węzłami są stwierdzenia. W grafie tym występują skierowane gałęzie łączące tylko te węzły, które odpowiadają stwierdzeniom pozostającym w bezpośredniej relacji przyczynowo-skutkowej. Węzłom-stwierdzeniom są przypisywane wagi. Mogą one określać np. stopień przekonania o słuszności danego stwierdzenia. Wagi przypisywane są również gałęziom. Określają one np. stopień przekonania o tym, że między węzłem (stwierdzeniem) poprzednim i węzłem (stwierdzeniem) następnym zachodzi relacja przyczynowo-skutkowa.

Definiowanie sieci realizowane jest w procesie pozyskiwania wiedzy i polega na zapisaniu pełnego zbioru rozpatrywanych stwierdzeń oraz określeniu relacji między tymi stwierdzeniami. Stwierdzeniom przypisywane są wyjściowe wartości wag (np. odpowiadające brakowi przekonania o słuszności stwierdzenia).

Wprowadzanie wiadomości do sieci (podczas działania systemu doradczego) polega na zmianie wagi węzła odpowiadającego stwierdzeniu, którego dotyczy wiadomość. Kolejne wiadomości mogą zwiększać lub zmniejszać stopnie przekonania o słuszności wybranych stwierdzeń. Jeżeli węzeł, którego waga ulega zmianie posiada następnika (w sieci), to ze względu na istniejącą między nimi relację przyczynowo-skutkową waga tego następnika może zostać zmieniona. Zmiana taka dokonywana jest zgodnie z przyjętymi dla danej sieci zasadami propagacji (aktualizacji) wag. Proces aktualizacji wag po wprowadzeniu nowej wiadomości obejmować może liczny zbiór węzłów, ponieważ następnik węzła, którego waga ulega zmianie może posiadać swojego następnika itd. Ze względu na hierarchiczną strukturę sieci proces ten

jest procesem skończonym. Można zauważyć, że sieć stwierdzeń jest szczególną formą zapisu zbioru reguł przybliżonych (por. rozdz. 2.3).

Wprowadzenie odpowiednich zasad propagacji wag w sieci pozwala na rozpatrywanie całej sieci jako układu dynamicznego. Jeżeli układ taki znajduje się w stanie równowagi, to wprowadzenie nowej wiadomości polegające na zmianie wagi wybranego węzła, pociąga za sobą zakłócenie równowagi sieci. Uwzględnienie tej nowej wiadomości polegać będzie na zmianie stanu równowagi sieci.

Określanie zasad propagacji wag w sieci przy założeniu, że przedmiotem rozważań jest cała sieć prowadzi do mało efektywnych algorytmów. Dla zwiększenia tej efektywności podejmowane są próby wyodrębniania w sieci autonomicznych izolowanych obszarów, pozwalających na ograniczenie liczby węzłów rozpatrywanych w kolejnych fazach poszukiwania stanu równowagi.

Sieć stwierdzeń może być weryfikowana za pomocą następujących kryterów:

- w sieci występują wyłącznie niezbędne gałęzie; nadmierne zwiększanie liczby gałęzi prowadzi do lawinowego zwiększania liczby działań niezbędnych podczas wykonywania operacji na sieci oraz może być przyczyną występowania w tej sieci nieoczekiwanych sprzeczności,
- liczba gałęzi występujących w sieci jest dostateczna dla opisania relacji zachodzących między stwierdzeniami.

Przedmiotem optymalizacji może być struktura sieci. Z formalnego punktu widzenia dopuszczalne jest dowolne uporządkowanie węzłów generujące różne struktury sieci. Zakładając jednak, że podstawą uporządkowania węzłów są relacje przyczynowo-skutkowe identyfikowane w procesie pozyskiwania wiedzy, należy oczekiwać iż struktura końcowa sieci nie będzie przypadkowa.

3.2.2. Losowe sieci stwierdzeń

Szczególnym przypadkiem sieci stwierdzeń są sieci, w których wagi węzłów i gałęzi interpretowane są jako prawdopodobieństwa. Oznacza to, że węzły rozpatrywane są jako zmienne losowe. Wcześniej założono, że węzły odpowiadają stwierdzeniom. Każde stwierdzenie orzeka o posiadaniu jakiegoś atrybutu co oznacza, że węzeł może przyjmować wartości odpowiadające wartościom tego atrybutu. Wyznaczają one zbiór wartości rozpatrywanej zmiennej losowej. Dla większości zastosowań można przyjąć założenie, że zbiór ten jest skończony, czyli rozpatrywane zmienne są zmiennymi dyskretnymi.

Przyjmujemy następującą notację:

- duże litery oznaczają węzły (identyfikują określone atrybuty, cechy obiektów) i przyporządkowane im dyskretne zmienne losowe; np. X,
- małe litery oznaczają określone stałe wartości zmiennych losowych (wartości atrybutów, wartości cech obiektów); np. x_1, x_2,
- duże podkreślone litery oznaczają zbiory możliwych wartości zmiennych losowych; np. $\underline{X} = (x_1, x_2, \ldots)$;
  zakłada się, że wartości $x_i \in \underline{X}$ są wartościami wzajemnie wykluczającymi się czyli zmienna X może przyjmować wyłącznie jedną wartość ze zbioru $\underline{X}$,
- stopień przekonania (czyli np. prawdopodobieństwo) o tym, że węzeł (zmienna losowa) X przyjmuje wartość x_i oznaczany będzie $BEL(x_i)$.

Każdej skierowanej gałęzi A-->B, gdzie A interpretowane jest jako przyczyna i B jako skutek, przyporządkowana jest tablica wag H w postaci prawdopodobieństw warunkowych

$$H_{B|A}[i,j] = P(B=b_i|A=a_j).$$

W pracy [25] stwierdzono, że wartości wag mają spełniać założenia rachunku prawdopodobieństwa, pozostawiając otwartą sprawę sposobu definiowania wartości prawdopodobieństw. Interpretowanie wag jako prawdopodobieństw jest źródłem dodatkowych warunków formalnych które mogą być zastosowane podczas weryfikowania sieci.

Prawdopodobieństwo złożonego zdarzenia polegającego na tym, że zmienne A,B,C,... przyjmują odpowiednio wartości a,b,c,...

(A=a) & (B=b) & (C=c) & ...

jest równe iloczynowi prawdopodobieństw odpowiedniego ciągu zdarzeń warunkowych

(A=a), (B=b dla A=a), (C=c dla (A=a & B=b)), ...

czyli prawdopodobieństwo łączne (dla niezerowych prawdopodobieństw warunkowych)

$$P(a,b,c,\ldots) = \ldots * P(c|a,b) * P(b|a) * P(a).$$

Dla uzyskania izolowanych obszarów w sieci wykorzystuje się uporządkowanie węzłów wyznaczone skierowanymi gałęziami. Dla każdego węzła X określa się zbiór "bezpośrednich rodziców", którego elementami są węzły posiadające tę własność, że znajomość ich wag (prawdopodobieństw) pozwala na określenie skutków wynikających z wpływu całej sieci na węzeł X. Oznacza to, że zbiór bezpośrednich rodziców izoluje rozpatrywany węzeł od jawnych wpływów pozostałych węzłów sieci. Wynika stąd, że dla opisu takiej sieci wystarczy określenie relacji każdego węzła do zbioru S jego bezpośrednich rodziców w postaci prawdopodobieństw warunkowych

$P(x_i|s_j) = P(x_i|a_j,b_j,\ldots)$

spełniających warunek $\Sigma_{x\in X} P(x_i|s_j) = 1$ gdzie sumowanie przebiega po wszystkich wartościach zmiennej X (niezależnie dla każdej wartości $s_j \in S$).

Możliwość "izolowania" węzła upraszcza procedury aktualizacji sieci w procesie pozyskiwania wiedzy, który polegać będzie na modyfikowaniu wartości określonych wcześniej prawdopodobieństw warunkowych oraz na wprowadzaniu nowych węzłów do sieci. Wprowadzenie nowego węzła wymaga wyłącznie arbitralnego określenia zbioru jego bezpośrednich rodziców (czyli tych węzłów, które wpływają bezpośrednio na węzeł wprowadzany) i wag ich wpływów na nowy węzeł. Lokalne wprowadzanie wag $P(x_i|s_j)$ nie naruszy spójności całej sieci. Wartości prawdopodobieństw $P(x_i|s_j)$ zapisywane mogą być w odpowiedniej tablicy. Należy zwrócić uwagę na liczbę elementów tej tablicy. Na przykład dla węzła X i 10-elementowego zbioru rodziców, z których każdy może przyjmować 3 różne wartości, tablica ta zawiera 177147 elementów.

Dla uproszczenia zadania określania wag $P(x_i|s_j)$, które może być trudne wtedy, gdy liczba bezpośrednich rodziców jest duża, wprowadza się hierarchiczną strukturę tego zbioru (przez wyróżnienie w nim podzbiorów w wyniku odpowiednich grupowań węzłów).

3.2.2.1. Niezależność węzłów

Wcześniej przyjęto, że w sieci stwierdzeń występują wyłącznie gałęzie łączące te węzły, które odpowiadają stwierdzeniom pozostającym w bezpośredniej relacji przyczynowo-skutkowej. Węzły niezależne nie są połączone bezpośrednio żadną gałęzią. Nie wynika stąd jednak iż brak gałęzi między dwoma węzłami oznacza ich niezależność. Na przykład dla sieci złożonej z trzech węzłów mogą występować następujące wersje rozdzielenia węzłów A i B przez węzeł X

A <---- X ----> B (1)

A ----> X ----> B (2)

A <---- X <---- B (3)

A ----> X <---- B (4)

gdzie wyłącznie w wersjach (1),(2),(3) występuje warunkowa niezależność węzłów A i B dla ustalonej wartości węzła X. W wersji (4) przyjęcie ustalonej (stałej) wartości w węźle X prowadzi do braku niezależności węzłów A i B. Opis rodzajów niezależności węzłów zamieszczono w [25]. Został on uzupełniony w [20], [17].

3.2.2.2. Stany sieci

Wynikiem procesu pozyskiwania wiedzy będzie sieć stanowiąca model określonej dziedziny. Wiedza o dziedzinie zapisana będzie w postaci struktury sieci i odpowiednich wartości liczbowych określających wagi gałęzi (prawdopodobieństwa warunkowe jako miary relacji przyczynowo-skutkowych) oraz prawdopodobieństwa 'a priori' przypisane korzeniom sieci (węzłom nie posiadającym rodziców).

Sieć zawiera węzły (dyskretne zmienne losowe) należące do zbioru W={A,B,C, ...} i przyjmujące wartości będące elementami zbiorów

$\underline{A}=\{a_1,a_2,\ldots\}$, $\underline{B}=\{b_1,b_2,\ldots\}$, ...

Przestrzeń możliwych stanów sieci (przestrzeń hipotez) Ω jest iloczynem kartezjańskim zbiorów wartości węzłów

$\Omega = \underline{A} * \underline{B} * \underline{C} * \ldots$

Stosowanie modelu zapisanego w postaci sieci polega na uwzględnianiu danych będących wynikiem obserwacji określonych obiektów lub wprowadzanych przez użytkownika, które ustalają wartości wybranych węzłów (zmiennych losowych). Najczęściej użytkownik dokonuje wyboru jednej wartości dla rozpa-

trywanego węzła (stwierdza jaką wartość posiada rozpatrywany atrybut). Wprowadzenie takich danych ogranicza przestrzeń możliwych stanów sieci i wpływa na prawdopodobieństwa warunkowe wystąpienia wartości przyporządkowanych tym węzłom, które nie są określone . Jeżeli bowiem stwierdzono, że węzeł X posiada wartość

$$x_i \in X$$

to przestrzeń Ω zostaje ograniczona do

$$\Gamma_i = \{ w \in \Omega \mid X(w) = x_i \}$$

oraz prawdopodobieństwa $P(.)$ zostają zastąpione prawdopodobieństwami $P(.|X=x_i)$. Dialog prowadzony z użytkownikiem jest źródłem kolejnych stwierdzeń ograniczających przestrzeń możliwych stanów sieci.

Należy zauważyć, że w sieci stwierdzeń zapisana jest wiedza o dziedzinie (będąca wynikiem procesu pozyskiwania wiedzy) oraz racje użytkownika (wynikające z odpowiedzi użytkownika lub bezpośrednich obserwacji). Wiedza o dziedzinie reprezentowana jest strukturą sieci oraz początkowymi wartościami wag. Wyznacza ona przestrzeń możliwych stanów sieci Ω. Racje użytkownika są zbiorami wiadomości określających zbiory wartości węzłów i wyznaczających ograniczone przestrzenie możliwych stanów sieci Γ_X. Aktualna przestrzeń możliwych stanów sieci jest częścią wspólną tak wyznaczanych przestrzeni Γ.

Po rozwiązaniu określonego zadania racje użytkownika mogą być kasowane. Prowadzi to do inicjalizacji początkowego stanu sieci warunkującego pełną przestrzeń jej możliwych stanów $\Gamma=\Omega$.

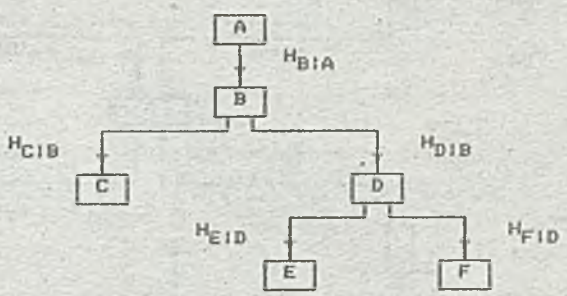
3.2.2.3. Propagacja stopni przekonania

Ograniczamy rozważania do sieci w postaci drzewa tzn. takiej, w której każdy węzeł (za wyjątkiem korzenia) może posiadać wyłącznie jednego rodzica. Ponadto zakładamy, że gałęziom sieci przyporządkowano wagi o stałej wartości. Każdemu węzłowi X przyporządkowany jest zbiór stopni przekonania

$$\{ BEL(x_i) = P(x_i|\Gamma) : x_i \in X \}$$

o dynamicznie zmiennych wartościach, gdzie Γ jest ograniczoną przestrzenią stanów sieci wynikającą z aktualnego zbioru ustalonych wartości węzłów (racje użytkownika).

Przykład sieci:



Rozpatrzmy [25] węzeł D sieci pokazanej na rysunku. Stopnie przekonania $BEL(d_i)$ określane dla każdej wartości $d_i \in D$ zależą od:

1. danych zawartych w drzewie, którego korzeniem jest węzeł D,
2. danych zawartych w drzewie, którego korzeniem jest węzeł C,
3. danych zawartych w drzewie nad węzłem B.

Ograniczona przestrzeń stanów sieci wynikająca z danych zawartych w drzewie o korzeniu X oznaczana będzie Γ_X^-. Ograniczona przestrzeń stanów wynikająca z pozostałych danych zapisanych w sieci oznaczana będzie Γ_X^+. Wynika stąd

$$\Gamma = \Gamma_X^- \cap \Gamma_X^+ .$$

Wpływ źródeł 2 i 3 na węzeł D jest określony przez parametry węzła B, ponieważ węzeł B oddziela węzeł D od pozostałych węzłów, za wyjątkiem potomków węzła D (węzły E i F). Oznacza to, że:

$$P(d_i | b_j, \Gamma_D^+) = P(d_i | b_j)$$

oraz

$$P(d_i, c_j | b_k) = P(d_i | b_k) \cdot P(c_j | b_k) .$$

Dla zbioru wzajemnie wykluczających się zdarzeń losowych (hipotez) x_1, x_2, ... zachodzących ze znanymi prawdopodobieństwami a priori $P(x_1)$, $P(x_2)$, ... oraz dla dowolnego zdarzenia losowego y, dla którego znane są prawdopodobieństwa warunkowe $P(y|x_1)$, $P(y|x_2)$, ... określać można prawdopodobieństwa a posteriori zdarzeń x_1, x_2, ... wg zależności Bayesa

$$P(x_i|y) = \alpha * P(y|x_i) * P(x_i)$$

gdzie α jest czynnikiem skalującym takim, że

$$1/\alpha = \Sigma_{x_i \in \underline{X}} P(y|x_i) * P(x_i) .$$

Wpływ $\Gamma = \Gamma_D^- \cap \Gamma_D^+$ na stopnie przekonania przypisane wartościom węzła D określany jest na podstawie zależności Bayesa [20]

$$BEL(d_i) = P(d_i|\Gamma) = P(d_i|\Gamma_D^+ \cap \Gamma_D^-) = \alpha * \sigma(d_i) * \pi(d_i) .$$

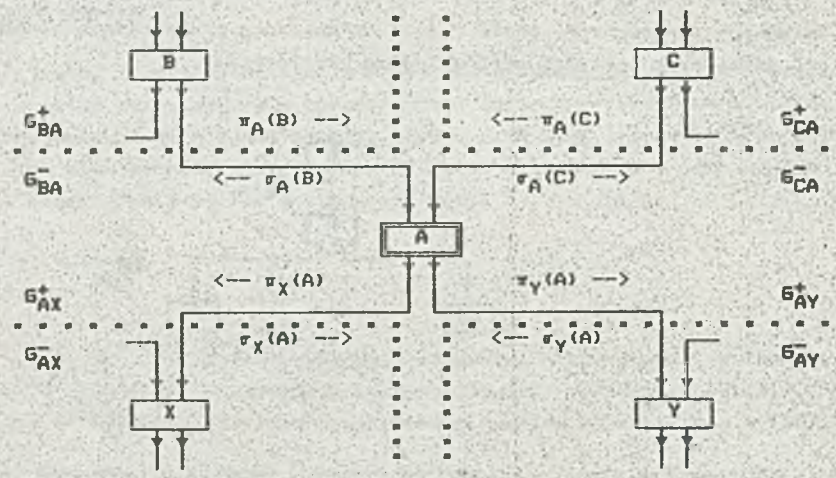
dla

$$\sigma(d_i) = P(\Gamma_D^-|d_i)$$

$$\pi(d_i) = P(d_i|\Gamma_D^+)$$

gdzie $\pi(d_i)$ opisuje przyczynowy wpływ "przodków" na rozkład zmiennej D oraz $\sigma(d_i)$ opisuje wpływ ograniczeń wnoszonych przez "potomków". Rozkład zmiennej D może być wyznaczony wtedy gdy dla węzła odpowiadającego tej zmiennej znane są tablice $\sigma(D)$, $\pi(D)$ zawierające odpowiednio elementy $\sigma(d_1)$, $\sigma(d_2)$, ... oraz $\pi(d_1)$, $\pi(d_2)$, ...

Rozpatrywana dotychczas sieć była drzewem, w którym każdy węzeł mógł mieć wyłącznie jednego przodka. W zadaniach reprezentacji wiedzy występuje potrzeba stosowania bardziej ogólnych wersji sieci, a w szczególności takich sieci, w których każdy węzeł może mieć kilku przodków. Uwzględniając możliwości grupowania węzłów, wymaganie to sprowadza się do potrzeby rozpatrywania sieci, w której dopuszczalne będzie następująca struktura otoczenia każdego węzła



Gałąź B-->A dzieli graf G na dwie części: górną G^+_{BA} i dolną G^-_{BA}. Dane zawarte w tych dwóch grafach wyznaczają ograniczone przestrzenie stanów Γ^+_{BA} i Γ^-_{BA} oraz ograniczoną przestrzeń stanów całego grafu $\Gamma = \Gamma^+_{BA} \cap \Gamma^-_{BA}$. Podobnie gałęzie

C-->A, A-->X, A-->Y

definiują odpowiednie pary grafów oraz pary ograniczonych przestrzeni stanów sieci.

Ponieważ węzeł A jest wspólnym potomkiem węzłów B i C, nie rozdziela on grafu G^+_{BA} od grafu G^+_{CA}. Rozdziela on jedynie grafy $G^+_{BA} \cup G^+_{CA}$, G^-_{AX}, G^-_{AY}, co prowadzi do zapisu

$$P(\Gamma^-_{AX}, \Gamma^-_{AY}|a_i, \Gamma^+_{BA}, \Gamma^+_{CA}) = P(\Gamma^-_{AX}|a_i) * P(\Gamma^-_{AY}|a_i) .$$

Zastosowanie reguły Bayesa prowadzi do [25]

$$BEL(a_i) = P(a_i|\Gamma^+_{BA}, \Gamma^+_{CA}, \Gamma^-_{AX}, \Gamma^-_{AY}) =$$
$$= \alpha * P(a_i|\Gamma^+_{BA}, \Gamma^+_{CA}) * P(\Gamma^-_{AX}|a_i) * P(\Gamma^-_{AY}|a_i) =$$
$$= \alpha * \sigma_X(a_i) * \sigma_Y(a_i) * \Sigma_{b_j \in \underline{B}, c_k \in \underline{C}} P(a_i|b_j, c_k) * \pi_A(b_j) * \pi_A(c_k)$$

gdzie α jest czynnikiem skalującym oraz

$\sigma_X(a_i) = P(\Gamma^-_{AX}|a_i)$,
$\sigma_Y(a_i) = P(\Gamma^-_{AY}|a_i)$,
$\pi_A(b_j) = P(b_j|\Gamma^+_{BA})$,
$\pi_A(c_k) = P(c_k|\Gamma^+_{CA})$.

Wprowadzenie nowych danych a_i do sieci prowadzić będzie do aktualizacji wartości $\sigma(.)$ i $\pi(.)$

$\sigma_A(b_i) = \alpha * \Sigma_{c_j \in C} [\pi_A(c_j) * \Sigma_{a_k \in A} \sigma_X(a_k) * \sigma_Y(a_k) * P(a_k|b_i,c_j)]$
$\sigma_A(c_i) = . . .$
$\pi_X(a_i) = \alpha * \sigma_Y(a_i) * [\Sigma_{b_j \in B, c_k \in C} P(a_i|b_j,c_k) * \pi_A(b_j) * \pi_A(c_k)]$
$\pi_Y(a_i) = . . .$

3.2.3. Sieci neuronowe

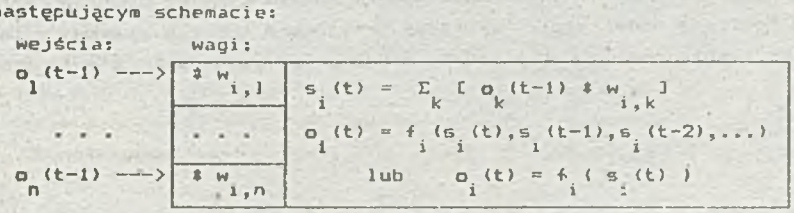
Sieci neuronowe są modelem, stosowanym do opisu działania złożonych układów wieloprocesorowych. Ponieważ wydaje się celowe rozważenie możliwości zastosowania takich układów sieciowych do reprezentacji wiedzy dla potrzeb systemów doradczych, pokrótce opisane zostaną ich najważniejsze własności.

Dynamiczny rozwój techniki komputerowej pozwala na wytwarzanie układów o bardzo złożonych strukturach. Dla zwiększenia szybkości wykonywania działań wprowadzono komputery wyposażone w pewną liczbę (a nawet wiele tysięcy - por. [35]) procesorów działających równolegle, z których każdy realizuje sekwencyjnie określone ciągi rozkazów. Dla synchronizacji wykonywanych zadań, uzależnia się rozpoczęcie wykonywania ciągu rozkazów przez procesor od wcześniejszego zakończenia działań przez inny procesor. Przewiduje się zastosowanie tych układów do automatycznego rozpoznawania obrazów i mowy, do sterowania robotów oraz do tworzenia układów samoorganizujących oraz samouczących się.

Programowanie układów komputerowych złożonych z tak dużej liczby procesorów nie jest możliwe w sposób klasyczny. Są one modelowane jako sieci neuronowe. Dla zdefiniowania sieci neuronowej należy określić [35]:

- komórki sieci i działania realizowane przez te komórki,
- strukturę sieci (listę połączeń między komórkami),
- reguły uczenia określające sposoby zmieniania stanu każdej komórki.

Podstawowym elementem sieci neuronowej jest "komórka" przetwarzająca doprowadzone do niej sygnały wejściowe w jeden sygnał wyjściowy. Doprowadzanymi sygnałami wejściowymi są sygnały wyjściowe innych komórek (lub połączeń sieci z otoczeniem). Komórka sieci posiada własną pamięć przeznaczoną do przechowywania danych określających stan komórki. Danymi takimi są na przykład wagi przyporządkowane sygnałom wejściowym. Suma ważona sygnałów wejściowych jest tzw. łącznym sygnałem wejściowym. Sygnał wyjściowy komórki jest wartością charakterystycznej dla rozpatrywanej komórki funkcji (np. funkcji liniowej, skokowej itp.), której argumentem jest łączny sygnał wejściowy. Dodatkowymi argumentami tej funkcji mogą być wcześniejsze wartości łącznego sygnału wejściowego. Sygnał wyjściowy komórki może być doprowadzony do dowolnej liczby innych komórek. Przekształcanie sygnałów wejściowych w wyjściowe realizowane może być w sposób ciągły (układy analogowe) lub dyskretny (symulacja cyfrowa układu analogowego). Komórkę sieci neuronowej działającej dyskretnie można pokazać w uproszczeniu na następującym schemacie:



gdzie (t-1) i (t) oznaczają kolejne chwile czasu.

Strukturę sieci neuronowej, czyli połączenia między komórkami, określa jej konstruktor. Podczas wyboru takich połączeń należy ustalić rozwiązanie kompromisowe umożliwiające minimalizację liczby połączeń przy jednoczesnym zapewnieniu zachowania wymaganych oddziaływań.

Najciekawszą własnością sieci neuronowych są mechanizmy uczenia. Określa się je za pomocą reguł zmian wartości wag w czasie. Zmiana tych wartości prowadzi do zmiany działania całej sieci. Należy zwrócić uwagę na istotne ograniczenie postaci omawianych reguł wynikające z warunku iż każda komórka realizuje proces uczenia indywidualnie, wyłącznie na podstawie lokalnych danych dostępnych dla tej komórki. Zmiany wartości wag w sieci działającej dyskretnie mogą być dokonywane zgodnie z zależnością:

$$w_{i,j}(t) = w_{i,j}(t-1) + \delta w_{i,j}(t) .$$

Przyjęcie algorytmu wyznaczania poprawek $\delta w_{i,j}$ określa strategię procesu uczenia. Wymaga się aby proces ten był zbieżny, tzn. aby uwzględnianie poprawek prowadziło do określonej wartości granicznej wag. Dla zapewnienia tej zbieżności wprowadza się dodatkowe parametry stanu komórki $v_{i,j}(t)$ ograniczające przedziały czasu w których może następować zmiana wag. Parametry te stają się istotnie różne od zera (czyli "zezwalają" na zmianę wag) po wystąpieniu określonego stanu komórki i utrzymują takie wartości przez określony czas po zmianie stanu. Przykładem algorytmu określania wartości wag może być [1]

$$\begin{aligned} w_{i,j}(t) &= w_{i,j}(t-1) + \delta w_{i,j}(t) \\ \delta w_{i,j}(t) &= \alpha * b(t-1) * v_{i,j}(t-1) \\ v_{i,j}(t) &= \beta * v_{i,j}(t-1) + (1-\beta) * s_i(t-1) * o_j(t-1) \end{aligned}$$

gdzie α, β $(0<\alpha<1,\ 0\leq\beta<1)$ są parametrami procesu uczenia oraz wartość funkcji $b(t) \in \{-1,+1\}$ jest miarą oceniającą poprawność działania sieci w chwili t.

4. Konstruowanie systemu doradczego

Konstruktorem "małego" systemu doradczego może być jedna osoba. Konstruowanie dużych systemów doradczych wymaga jednak współdziałania licznego zespołu złożonego między innymi ze specjalistów z zakresu tzw. "inżynierii wiedzy" (ang. knowledge engineering; pojęcie to obejmuje m.in. reprezentację wiedzy, techniki jej pozyskiwania oraz stosowanie i weryfikację systemów doradczych). Udział w tym zespole specjalisty-eksperta reprezentującego dziedzinę (lub dziedziny), której dotyczy system, jest konieczny ze względu na potrzebę uwzględniania:

- jego doświadczeń osobistych dotyczących wcześniej rozwiązanych zadań,
- jego wiedzy na temat sposobów rozwiązywania określonych zadań,
- jego wiedzy na temat sposobu wybierania odpowiednich metod rozwiązywania zadań.

Należy wyraźnie zaznaczyć, że proces konstruowania systemu doradczego jest procesem twórczym i w związku z tym trudno jest jednoznacznie określić kolejność działań, które powinien wykonać konstruktor systemu doradczego. W ogólnym przypadku są to następujące działania:

- identyfikacja problemu,
- określenie istoty działania systemu,
- określenie zakresu bazy wiedzy, która powinna wystąpić w systemie,
- wybór sposobu reprezentacji wiedzy,
- konstruowanie elementów systemu,
- badania testujące i ocena przydatności systemu.

Proponowana kolejność może ulegać zmianom spowodowanym między innymi tym, że podczas realizacji jakiegoś etapu stwierdza się potrzebę modyfikacji rozwiązań przyjętych podczas etapów wcześniejszych. Ponadto duży wpływ na zakres i kolejność działań wykonywanych przez konstruktora mają stosowane narzędzia programowe.

4.1. Identyfikacja problemu

Wynikiem identyfikacji problemu powinien być nieformalny opis potrzeby jaka ma zostać zaspokojona za pomocą konstruowanego systemu doradczego. Zakres zastosowań systemu oraz krąg użytkowników powinny być określone dokładnie i jednoznacznie. Z założeń ma wynikać co i komu doradzać będzie system.

Określanie zakresu zastosowań konstruowanego systemu jest istotnym zadaniem. Wybór nieodpowied-

niego zakresu może doprowadzić do zadań, dla których nie jest znany sposób ich rozwiązywania. Może on również doprowadzić do powstania systemu rozwiązującego zadania trywialne, gdzie nakłady związane z konstruowaniem i uruchamianiem systemu wielokrotnie przekraczają efekty wynikające z jego stosowania. Pomijając w tym miejscu małe systemy doradcze (zawierające na przykład mniej niż 100 reguł), przeznaczone do wspomagania często wykonywanych działań rutynowych można stwierdzić, że podejmowanie prac związanych z konstruowaniem systemu doradczego jest uzasadnione wtedy, gdy formułowane dla niego zadania nie są "zbyt proste" oraz "zbyt złożone".

Podjęcie decyzji o konstruowaniu systemu doradczego należy poprzedzić próbą rozwiązania zadania za pomocą klasycznych, algorytmicznych metod programowania, zalecanych zwłaszcza wtedy, gdy zadanie ogranicza się do działań numerycznych i wymaga małej oraz niezmiennej bazy danych. Metody algorytmiczne często prowadzą do programów bardziej efektywnych niż systemy doradcze. Zastosowanie systemu doradczego jest wymagane wtedy, gdy zadanie wymaga użycia wiedzy subiektywnej, niejednoznacznej, symbolicznej oraz takiej, która nie może być weryfikowana w sposób jednoznaczny. Przykładem takiej dziedziny jest diagnostyka techniczna.

4.2. Określenie istoty działania systemu

Dla określenia istoty działania systemu należy przeprowadzić analizę sposobu działania specjalisty (eksperta) podczas rozwiązywania zadań podobnych do tych jakie mają być rozwiązywane przez system. Identyfikacja sposobów działania specjalisty jest możliwa np. na podstawie obserwacji jego postępowania przy rozwiązywaniu odpowiedniego zbioru przykładowych zadań. Uzupełnieniem takich badań może być analiza publikacji i podręczników dotyczących wybranych dziedzin.

Istota działania systemu doradczego może być zapisana w postaci ogólnego schematu, określającego:

- przebieg dialogu użytkownika z systemem, np.:
  - pytania zadawane przez system, odpowiedzi użytkownika (o kolejności zadawania pytań decyduje system),
  - pytania zadawane przez użytkownika, odpowiedzi systemu (o kolejności zadawania pytań decyduje użytkownik),
  - pytania zadawane przez system, możliwość bezpośredniego sterowania przez system działaniem obiektu (w określonym zakresie),
- sposób prowadzenia dialogu, np.:
  - za pomocą odpowiedniego menu,
  - za pomocą pytań, na które należy udzielać odpowiedzi "tak", "nie", "nie wiem",
  - za pomocą pytań, na które można udzielać odpowiedzi w języku zbliżonym do języka naturalnego,
- sposób formułowania odpowiedzi systemu, którym może być:
  - wybór odpowiedzi ze znanego systemowi zbioru odpowiedzi określonych przez konstruktora,
  - ograniczenie odpowiedzi do słów TAK, NIE, NIE WIEM,
  - formułowanie odpowiedzi w języku zbliżonym do języka naturalnego.

Określając istotę działania systemu należy uwzględniać między innymi następujące kryteria:

- możliwość właściwego wykorzystywania bazy wiedzy przez system,
- możliwość zmiany i uzupełniania elementów bazy wiedzy,
- możliwość objaśniania przez system przebiegu procesu rozumowania,

oraz należy ustalić:

- przy użyciu jakiego sprzętu będzie eksploatowany system,
- w jaki sposób zostanie przygotowany program dla tego sprzętu,
- podział systemu na niezależne mniejsze systemy (niezbędne zwłaszcza dla dużych systemów), które mogą być konstruowane niezależnie oraz zakres i sposób przekazywania danych między elementami systemu.

Typowymi elementami występującymi w dużej liczbie systemów są: baza wiedzy, baza danych "stałych" i "zmiennych", procedury wnioskowania, objaśniające oraz procedury sterowania. Można zauważyć, że wyłącznie baza wiedzy i bazy danych zależą bezpośrednio od zastosowań systemu, podczas gdy pozostałe elementy mogą być w dużym stopniu niezależne od zakresu zastosowań. Program określający działanie systemu doradczego może być przygotowany w dowolny sposób. Uwzględniając jednak jego

złożoność, celowym jest użycie specjalnych narzędzi programowych, do których zaliczyć można odpowiednie języki programowania (np.PROLOG, LISP), szkieletowe systemy doradcze oraz metasystemy doradcze.

Szkieletowe systemy doradcze są systemami posiadającymi pustą bazę wiedzy oraz puste bazy danych i kompletne pozostałe elementy. Konstruowanie systemu doradczego przy pomocy systemu szkieletowego polega głównie na zapisaniu (w odpowiedni sposób) bazy wiedzy. Możliwość tworzenia szkieletowych systemów doradczych wynika z opisanej struktury systemu doradczego, w którym jako wyraźnie oddzielone elementy występują:

- zbiór informacji o dziedzinie (stwierdzenia, reguły),
- ogólny, niezależny od dziedziny "mechanizm" wnioskowania.

Systemy szkieletowe nie umożliwiają swobodnego wyboru postaci bazy wiedzy. Wybór taki jest możliwy przy zastosowaniu bardziej uniwersalnych narzędzi nazywanych metasystemami doradczymi, które są systemami doradczymi wspomagającymi projektowanie systemów doradczych. Generowanie systemu doradczego za pomocą metasystemu polega na określaniu zbioru informacji o dziedzinie.

Konstruowanie systemu doradczego bez odpowiedniego narzędzia programowego wymaga ogromnego nakładu pracy. Zastosowanie wybranego narzędzia jest jednak celowe wyłącznie wtedy, gdy:

- sposób reprezentacji wiedzy i mechanizmy wnioskowania przyjęte w systemie szkieletowym lub dostępne w metasystemie są odpowiednie dla rozwiązywanych zadań,
- narzędzia posiadają odpowiednią dokumentację i działają w sposób zgodny z powszechnie stosowanymi standardami, co pozwala na szybkie zapoznanie się ze sposobem ich obsługi.

4.3. Wybór sposobu reprezentacji wiedzy

Przed wyborem sposobu reprezentacji (zapisu) wiedzy niezbędnej dla działania systemu, należy rozstrzygnąć:

- czy wymagania stawiane systemowi doradczemu prowadzą do jakiegoś określonego sposobu reprezentacji,
- czy wymagane będzie stosowanie stwierdzeń niepewnych oraz czy dopuszczalny będzie brak odpowiedzi użytkownika na wybrane pytania zadawane przez system,
- czy potrzeba uwzględniania złożonych relacji między stwierdzeniami jest ważniejsza od potrzeby małej złożoności procesu pozyskiwania wiedzy,
- czy możliwe jest zastosowanie dostępnych systemów szkieletowych lub metasystemów, czy też konieczne jest konstruowanie systemu za pomocą standardowych języków programowania,
- czy konstruowany system doradczy będzie współdziałał z innymi programami albo zewnętrznymi bazami danych lub wiedzy,
- czy wymagana będzie możliwość modyfikowania bazy wiedzy przez użytkownika (na podstawie doświadczeń, uzyskanych podczas eksploatacji systemu, a także wynikających z rozwoju dziedziny),
- czy wymagane jest wyposażenie systemu w mechanizm umożliwiający pozyskiwanie wiedzy od użytkownika (adaptacja i uczenie systemu),
- czy wymagana jest możliwość samoczynnego weryfikowania przez system jego bazy wiedzy w celu wykrycia występujących w niej sprzeczności oraz redundancji,
- czy wymagane będzie dołączanie istniejących już, szczegółowo sprawdzonych fragmentów opracowanych w różnych językach proceduralnych (np. analiza częstotliwościowa sygnałów),
- czy wymagana jest wysoka efektywność (szybkość działania) systemu doradczego.

4.4. Wymagania stawiane przez diagnostykę techniczną

Konstruując system doradczy dla potrzeb diagnostyki technicznej należy uwzględniać różne czynniki, nie występujące w innych zastosowaniach tych systemów. W pierwszym rzędzie należy dokonać wyboru właściwej postaci bazy danych i bazy wiedzy. Potrzeba dokonania takiego wyboru jest często nie dostrzegana z powodu pozornego podobieństwa zadań diagnostyki technicznej i diagnostyki medycznej, dla której znanych jest już wiele rozwiązań. Istotna różnica (ze względu na wymaganą postać baz) między tymi zadaniami polega na tym, że przedmiotem zainteresowań diagnostyki medycznej są obiekty należące do jednej klasy (ludzie). Odmienna sytuacja występuje w diagnostyce technicznej, gdzie

struktura obiektów może być bardzo zróżnicowana oraz w wyniku przeprowadzanych modernizacji maszyn może ulegać zmianom w czasie działania obiektu. Inną cechą charakterystyczną zadań z zakresu diagnostyki technicznej jest potrzeba uwzględniania tzw. "sezonowości" w działaniu obiektów, wywołanej występującymi w praktyce cyklami remontowymi.

Drugim specyficznym dla diagnostyki technicznej zadaniem (które nie występuje wyraźnie w diagnostyce medycznej) jest potrzeba wyznaczania złożonych ocen dużej liczby obserwowanych sygnałów diagnostycznych. W chwili obecnej znane są efektywne algorytmy wyznaczania takich ocen i większość ośrodków prowadzących badania w tym zakresie dysponuje odpowiednim oprogramowaniem. Zadaniem trudnym, które musi być rozwiązane przez konstruktora systemu, jest określenie sposobu wprowadzania tych danych do bazy danych systemu i określenie formatu zapisu tych danych w bazie. Trudność określenia odpowiedniego formatu spowodowana jest między innymi dużą licznością tych danych (liczby danych przekraczają o kilka rzędów liczby danych występujących dla każdego obiektu w bazach systemów doradczych konstruowanych dla potrzeb medycyny). Dla właściwego rozwiązania wymienionych zadań niezbędne jest równoległe zastosowanie różnych narzędzi programowych.

4.5. Konstruowanie elementów systemu

Duża liczba czynników, które muszą być brane pod uwagę, jest przyczyną wyróżniania dwóch faz procesu konstruowania elementów złożonych systemów doradczych. W pierwszej fazie konstruowany jest prototyp systemu, zawierający ograniczoną bazę wiedzy. Tak otrzymany prototyp poddawany jest następnie szczegółowej weryfikacji (prowadzonej głównie przez autorów systemu) dotyczącej adekwatności reprezentacji wiedzy, skuteczności mechanizmu wnioskowania oraz organizacji dialogu z użytkownikiem. Szczególna uwaga powinna być zwrócona na właściwe działanie układów objaśniających. Wynikiem weryfikacji prototypu, przeprowadzanej najczęściej na podstawie wcześniej określonych przykładów, mogą być wnioski dotyczące zmian przyjętych założeń (prowadzić to będzie do kolejnego prototypu). Po zakończeniu weryfikacji, prototyp jest rozbudowywany w celu otrzymania kompletnego systemu doradczego. Rozbudowywanie polega głównie na zapisywaniu pełnej bazy wiedzy, czyli dotyczy pozyskiwania wiedzy.

4.6. Pozyskiwanie wiedzy

Proces pozyskiwania wiedzy jest obecnie najtrudniejszym zadaniem podczas konstruowania systemów doradczych. Określono liczne zalecenia i propozycje dotyczące pozyskiwania wiedzy dla systemów charakteryzujących się prostą strukturą i prostym mechanizmem wnioskowania. Zadanie to pozostaje problemem otwartym w odniesieniu do dużych systemów doradczych posiadających złożone mechanizmy wnioskowania i wykorzystujących bazy wiedzy obejmujące rozległe dziedziny.

W procesie pozyskiwania wiedzy uczestniczy najczęściej konstruktor systemu i odpowiednia grupa specjalistów-ekspertów. Możliwe jest wyposażenie systemu w narzędzia umożliwiające realizację procesu pozyskiwania wiedzy bez bezpośredniego udziału w nim konstruktora systemu (np. systemy szkieletowe). Większość twórców istniejących systemów stwierdza, że najtrudniejszym zadaniem jest wyłonienie grupy specjalistów, którzy będą posiadali odpowiednie kwalifikacje i jednocześnie znajdą czas oraz wyrażą zgodę na to by posiadana przez nich wiedza i doświadczenie zostały zapisane w bazie wiedzy systemu doradczego. Z tych względów w praktyce głównym źródłem wiedzy o rozpatrywanej dziedzinie są publikacje.

Dokonując uogólnienia stosowanych metod pozyskiwania wiedzy należy wyróżnić trzy rodzaje postępowania:

- specjalista (sam lub przy pomocy twórców systemu) formułuje reguły i stwierdzenia w postaci nadającej się do bezpośredniego zapisania w bazie wiedzy i bazie danych,
- specjalista określa szczególne przypadki reguł i stwierdzeń, które mogą być zapisywane w bazach systemu dopiero po odpowiednim ich uogólnieniu,
- specjalista na podstawie znanych mu faktów, przedstawia szczegółowe przykłady relacji <STWIERDZENIE, STWIERDZENIE>, które mogą stanowić podstawę do wyznaczenia szczególnych przypadków reguł, wymagających jednak dalszego uogólnienia.

4.7. Weryfikacja i ocena przydatności systemu

Końcową fazą procesu konstruowania systemu jest jego przekazanie użytkownikowi i ostateczne ocenienie jego skuteczności. Wyniki badań testujących skuteczność systemu mogą być podstawą modyfikacji reguł sterujących przebiegiem procesu wnioskowania. Najczęściej wyniki tych badań sugerują potrzebę modyfikacji bazy wiedzy, która powinna dotyczyć wyłącznie treści (zawartości) tej bazy, a nie przyjętego sposobu reprezentacji wiedzy, ponieważ konieczność zmiany sposobu reprezentacji oznacza w praktyce powrót do początkowych etapów procesu konstruowania systemu.

Weryfikację prowadzą twórcy systemu. Ocenie podlegają między innymi: poprawność rozwiązań, zdolność objaśniania (przez system doradczy) otrzymywanych wyników, kolejność pytań zadawanych użytkownikowi itp.

Badania weryfikacyjne prowadzone są głównie na podstawie analizy sposobu rozwiązywania zadań podczas eksploatacji systemu. Niezbędne jest również przeprowadzanie takich badań przed przekazaniem systemu użytkownikowi. Mogą być one realizowane przez porównanie wyników otrzymywanych za pomocą systemu z wynikami rozwiązań hipotetycznych (testowych) zadań, bazujących na danych literaturowych lub na wywodzących się z praktyki przykładach "przypadków" diagnostycznych. Ze względu na potrzebę wielokrotnego używania tych przykładów, wskazany jest ich zapis do bazy danych stałych.

Niewłaściwa (nie oczekiwana przez twórców systemu) kolejność zadawania pytań nie jest równoznaczna z błędnym działaniem systemu. Może być ona skutkiem nieprzewidywanej kolejności stosowania reguł. Należy ją eliminować na przykład przez wprowadzenie odpowiednich reguł lub stwierdzeń pomocniczych, ponieważ prowadzi ona najczęściej do zmniejszenia stopnia przekonania użytkownika o poprawności działania systemu.

Błędne wyniki działania systemu mogą być skutkiem:

- nieoczekiwanego pomijania (przez system) wybranych reguł, spowodowanego zbyt szczegółową treścią przesłanek tych reguł, zawierających nadmierną liczbę warunków,
- tego, że terminy stosowane podczas zadawania pytań użytkownikowi mają nieznany (temu użytkownikowi) zakres znaczeniowy, w wyniku czego udziela on błędnych odpowiedzi.

Mogą występować sytuacje, w których system nie będzie w stanie określić końcowego rozwiązania (końcowej konkluzji), spowodowane tym, że:

- zbiór reguł jest niekompletny,
- w zbiorze reguł występują sprzeczne reguły (lub łańcuchy reguł), np.:

```
REGUŁA xxx      Jeżeli "A" to "B".
REGUŁA yyy      Jeżeli "A" to "C".
REGUŁA zzz      Jeżeli "C" to "nie B".
```

- w zbiorze reguł występują zamknięte łańcuchy reguł, po przerwaniu których konkluzja ostatniej reguły jest elementem przesłanki pierwszej reguły, np.:

```
REGUŁA xxx      Jeżeli "A" i "B" to "C".
REGUŁA yyy      Jeżeli "C" to "D".
REGUŁA zzz      Jeżeli "D" to "B".
```

Celem weryfikacji może być również optymalizacja baz wiedzy polegająca na usuwaniu redundancji wynikającej z istnienia reguł nakładających się, to znaczy takich, które dla różnych podzbiorów elementów przesłanek prowadzą do takiej samej konkluzji, np.:

```
REGUŁA xxx      Jeżeli "A" i "B" to "C".
REGUŁA yyy      Jeżeli "A" to "C".
```

Znane są algorytmy formalnej weryfikacji zbiorów reguł dokładnych. Brak niestety takich algorytmów dla reguł przybliżonych.

5. Wnioski

W przedstawianym opracowaniu opisano istotę działania i zasady konstruowania systemów doradczych. Więcej uwagi poświęcono opisowi ważniejszych metod reprezentacji wiedzy. Wyboru dokonano, mając na uwadze potrzebę utworzenia, weryfikacji i doskonalenia bazy wiedzy wchodzącej w skład diagnostycznego systemu doradczego.

Zwrócono uwagę na to, że przed wyborem sposobu reprezentacji wiedzy należy bardzo wyraźnie

określić zakres problemów, który ma być rozwiązywany przy pomocy konstruowanego diagnostycznego systemu doradczego. Podstawą wyboru odpowiedniej metody powinna być ocena możliwości realizacji procesu pozyskiwania wiedzy dla tworzonego systemu. Główne kryteria oceny dokonanego wyboru (a posteriori!) to: skuteczność działania tego systemu oraz łatwość weryfikacji i modyfikacji bazy wiedzy.

Z omówionymi tu zagadnieniami reprezentacji wiedzy dla potrzeb diagnostycznego systemu doradczego łączy się problem właściwej ochrony praw autorskich osób, których wiedzę zapisano w bazach systemu doradczego. Można wyrazić obawę, że brak ochrony praw autorskich ekspertów może stać się poważną przeszkodą w procesie pozyskiwania wiedzy fachowej dla potrzeb systemów doradczych.

Stopień zaawansowania prowadzonych w Polsce prac dotyczących układów monitorowania stanu maszyn upoważnia do postawienia wniosku o celowości połączenia wysiłków wszystkich zainteresowanych ośrodków zmierzających do opracowania efektywnego systemu doradczego w tej dziedzinie, którego zastosowanie umożliwi gromadzenie doświadczeń specjalistów oraz prowadzić będzie do zwiększenia skuteczności procesu diagnozowania maszyn.

6. Literatura

[1] BARTO A.G., SUTTON R.S., ANDERSON C.W.: Neuronlike adaptive elements that can solve difficult learning control problems. IEEE-SMC 13 (1983) ss.834-846.

[2] BOLC L., CICHY M., ROZAŃSKA L.: Przetwarzanie języka naturalnego. WNT, Warszawa 1982.

[3] BRACHMAN R.J.: On the Epistemological Status of Semantic Networks. [W:] "Associative Networks: Representation and Use of Knowledge by Computers", FINDER N.V. (Ed.), Academic Press, New York 1979, ss.3-50.

[4] BRACHMAN R.J., PIGMAN V.G., LEVESQUE H.J.: An Essential Hybrid Reasoning System - Knowledge and Symbol Level Accounts of KRYPTON. Proc.of the IJCAI, Los Altos CA 1985, ss.532-539.

[5] BRACHMAN R.J., SCHMOLZE J.G.: An Overview of the KL-ONE Knowledge Representation System. "Cognitive Science" 9 (1985), ss.171-216.

[6] BUCHANAN B.G., FEIGENBAUM E.A.: DENDRAL and Meta-DENDRAL. Artificial Intelligence, 11 (1978) s.5-24.

[7] CHOLEWA W.: Automatyczne układy i metaukłady doradcze. Materiały I Krajowej Konferencji Naukowo-Technicznej "Diagnostyka Techniczna Urządzeń i Systemów", Kielce-Sielpia 1986, t.2, ss.47-56.

[8] CHOLEWA W.: Bilateral fuzzy implication. First Conference of the IFCS. Aachen 1987. Abstracts, s.34.

[9] CHOLEWA W., PEDRYCZ W.: Systemy doradcze. Skrypt Politechniki Śląskiej nr 1447. Gliwice 1987.

[10] FEIGENBAUM E.A.: Expert Systems in the 1980. [W:] "State of the Art Report on Machine Intelligence", BOND A. (Ed.). Pergamon-Infotech.

[11] FININ T.: Understanding Frame Languages. "AI Expert" Nov. 1986, ss.44-50.

[12] FOX M., WRIGHT J., ADAM D.: Experiences with SRL - An Analysis of a Frame-based Knowledge Representation. [W:] "Expert Database Systems", KERSCHBERG L. (Ed.). Benjamin-Cummings Publ., Menlo Park CA., 1985.

[13] HACKEMANN P.: Entwurf und Implementierung eines Dialogsystems zur Repräsentation und Propagation von vagem Wissen. Diplomarbeit. TU Braunschweig 1988.

[14] HEUER S., KOCH U., CRYER C.: INVEST - An Expert System for Financial Investments. "IEEE Expert" Summer 1988, ss.60-68.

[15] KACPRZYK J.: Zbiory rozmyte w analizie systemowej. PWN, Warszawa 1986.

[16] KEE Software Development System - User's Manual. El Camino Real West Mountain View CA 1986.

[17] KÖRNER S.: Implementierung eines probabilistischen Ansatzes zur Propagation von Unschärfe. Diplomarbeit. TU Braunschweig 1988.

[18] KRECZMAR A.: Języki obiektowe. "Informatyka" 1,2 (1988).

[19] KRUSE R., MEYER K.D.: Statistics with Vague Data. Reidel Publ.Comp., Braunschweig 1987.

[20] KRUSE R., SCHWECKE E.: Propagation von Unschaerfe. Informatik Fachbericht 87-09. TU Braunschweig 1987.

[21] LANAM D., LETSINGER R., ROSENBERG S., HUYUN P., LEMON M.: Guide to the Heuristic Programming and Representation Language - Part I - Frames. Application and Technology Lab. MEMO-83-3, Computer Research Center, Hewlett Packard Jan. 1984.

[22] LEBAILLY J., MARTIN-CLOUAIRE R., PRADE H.: Use of Fuzzy Logic in a Rule-based System in Petroleum Geology. [W:] Artificial Intelligence - Applications of Quantitative Reasoning. SANCHEZ E., ZADEH L.A. (Eds.), Pergamon Press 1986.

[23] MINSKY M.: A Framework for Representing Knowledge. [W:] The Psychology of Computer Vision. WINSTON P.H. (Ed.). Mc Graw-Hill. New York 1975, ss.211-277.

[24] NEWELL A., SIMON H.A.: Human problem solving. Prentice-Hall, Engelwood Cliffs 1972.

[25] PEARL J.: Fusion, Propagation and Structuring in Bayesian Networks. Artificial Intelligence 29 (3) 1986, ss.241-288.

[26] Personal Consultant Plus. User's & Reference Guide. Texas Instruments. 1986.

[27] PRADE H.: A Synthetic View of Approximate Reasoning Techniques. Proc. of the 8th IJCAI (1983), ss.130-136.

[28] ROBERTS R.B., GOLDSTEIN I.P.: The FRL Manual. MIT AI Laboratory Memo 409, Cambridge MA 1977.

[29] ROSENBERG S., ROBERTS B.: Coreference in a Frame Database. Proc. of the IJCAI-79, ss.729-734.

[30] ROSENBLATT F.: The Perceptron - A Probabilistic Model for Information Storage and Organization

in the Brain. Psychological Review 65 (1958), ss.386-407.
[31] SHAFER G.: A Mathematical Theory of Evidence. Princeton University Press, London 1976.
[32] SHORTLIFFE E.H.: Computer-Based Medical Consultations. American Elsevier, New York 1976.
[33] SKELLY Ch.: Objective-C. "AI Expert" April 1987, ss.81-86.
[34] STEIN J.: Object-Oriented Programming and Databases. "Dr. Dobb's Journal", March 1988, ss.18-34.
[35] WERNTGES H., ECKMILLER R.: Neuronale Computer. Magazin für Computer Technik 10/1988, ss.70-82.
[36] WINSTON P.H.: Artificial Intelligence. Addison-Wesley, Massachusetts 1984.

mgr inż. Danuta PYZEL

Instytut Maszyn Matematycznych

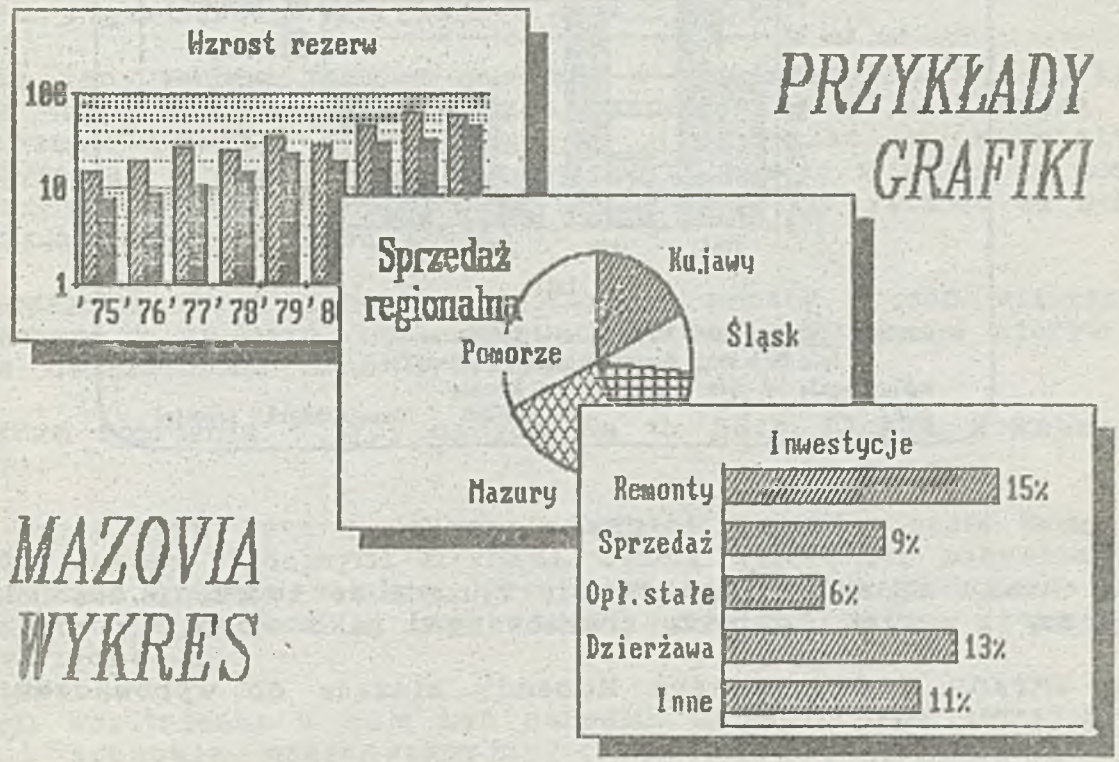
Program WYKRES-grafika użytkowa

Program WYKRES jest wygodnym narzędziem do graficznego prezentowania dużej ilości danych z zakresu zarządzania, statystyki, administracji, handlu itp. na komputerze Mazovia oraz kompatibilnych z IBM PC.

Podstawową zaletą programu jest łatwość obsługi. Z powodzeniem mogą go używać osoby nie znające języków obcych. Wszystkie teksty wyświetlane na ekranie, tj. komendy, komunikaty i wskazówki dla użytkownika są napisane w języku polskim.

Opracowanie programu wychodzi naprzeciw zapotrzebowaniu na wizualne prezentowanie różnych danych omawianych podczas konferencji, sympozjów, kursów, spotkań itp.

Rodzaj prezentowanych danych jest praktycznie nieograniczony: ceny, budżety, koszty, zyski, zapasy, przebiegi, plany, przyrosty, produkcja, nakłady i płace, to tylko niektóre z nich.



rys. 1

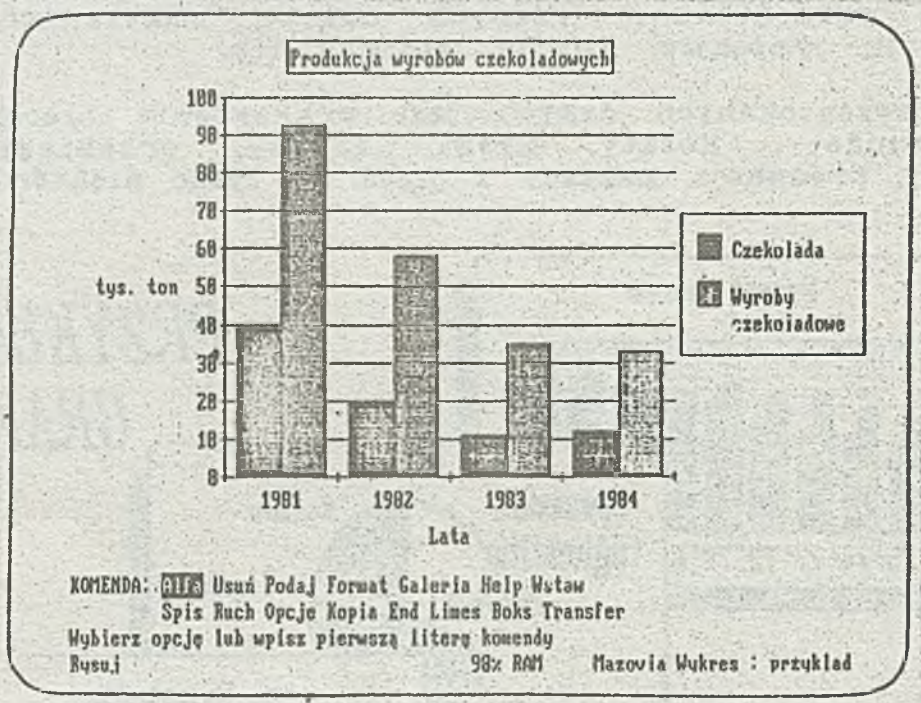
Ogólna charakterystyka programu

Program ma trzy główne ekrany. Są to ekrany: **Spis**, **Podaj** i **Rysuj**. W dolnej części każdego ekranu wyświetlona jest lista zleceń, linia informacyjna i linia stanu.

Lista zleceń przedstawia menu komend możliwych do wykonania w danej chwili.

Linia informacyjna "podpowiada", co powinien w danej chwili zrobić użytkownik lub informuje o popełnionym błędzie, np. *Wybierz opcję lub wpisz pierwszą literę komendy*, *Wprowadź datę*, *Niewłaściwa data*.

Linia stanu pomaga w kontrolowaniu postępów pracy. Podaje ona nazwę aktualnego, głównego ekranu, obszar wolnej pamięci, nazwę pliku, z którym się w danej chwili pracuje i inne dodatkowe informacje, różne dla głównych ekranów.



rys.2

Menu ekranu **Spis** zawiera komendy służące do tworzenia, łączenia i do pracy z seriami danych, traktowanymi jako całość.

Menu ekranu **Podaj** zawiera komendy służące do wprowadzania i edycji serii danych.

Menu ekranu **Rysuj** zawiera komendy służące do tworzenia i poprawiania wykresów.

Z każdego z głównych ekranów można przejść do innego głównego ekranu. Wybór komendy następuje przez podświetlenie nazwy komendy i naciśnięcie klawisza WYKONAJ lub przez naciśnięcie klawisza z pierwszą literą komendy. Większość komend ma jeszcze podkomendy. Ich lista, stanowiąca nowe menu, pojawia się na ekranie po wyborze odpowiedniej komendy, zastępując główne menu. To nowe menu rozpoczyna się nazwą wybranej komendy, napisaną dużymi literami, np. dla komendy Transfer nowe menu wygląda następująco:

TRANSFER: Ładuj, Zapisz, Czyść, Usuń, Fuzja, Opcje

Wybór odpowiedniej podkomendy przebiega podobnie, jak wybór komend podstawowych. Niektóre komendy lub podkomendy mają jeszcze tzw. pola, w które należy wpisać oczekiwane informacje. Istnieją dwa typy pól: "wpisz dane" - wymagające wpisania odpowiedzi i "wybierz z menu" - zawierające listę opcji do wyboru, w której należy podświetlić wybraną odpowiedź.

W trakcie działania programu użytkownik może w każdej chwili skorzystać z informacji pomocniczych, związanych z aktualnie wykonywaną komendą. Chcąc te informacje uzyskać, należy skorzystać z opcji Help, znajdującej się w każdym głównym menu. Można także, po podświetlenu nazwy komendy lub pola, nacisnąć klawisz HELP. Zostanie wtedy wyświetlona dokładna informacja o tej komendzie lub polu. W każdej chwili można powrócić do przerwanej pracy z programem, wybierając z menu opcję Wróć.

Przygotowywanie danych

Seria danych jest zbiorem punktów o współrzędnych x i y. Każda para (x,y) jest punktem danych (współrzędne x nie muszą być liczbami). Serie danych można wprowadzać z klawiatury, korzystając z ekranu Podaj, można je skopiować z pliku za pomocą komendy Transfer lub utworzyć nową serię, korzystając ze starej, używając komendy Zmień Kopiuj.

Poszczególne serie danych mogą być w prosty sposób włączane i wyłączane z wykresu, tzn. można utworzyć wykres, na którym nie będą uwzględnione dane wyłączonej serii.

Program rozróżnia 4 typy danych. Są to: **Data, Liczba, Sekwencja** i **Tekst.**

Dla serii typu Data i Sekwencja wartości x mają stałe przyrosty między dwiema kolejnymi liczbami. Jeżeli wpisze się pierwszą datę lub liczbę i poda wielkość przyrostu (dla Daty jednostka czasu) program sam wyliczy i wprowadzi daty lub liczby stanowiące współrzędne x.

Jeżeli współrzędne x mają być nazwami, należy wybrać typ Tekst, a jeżeli stanowią niesekwencyjny ciąg liczb, należy wybrać typ Liczba. Dla obu tych typów należy, wprowadzając dane, podawać obie współrzędne punktów.

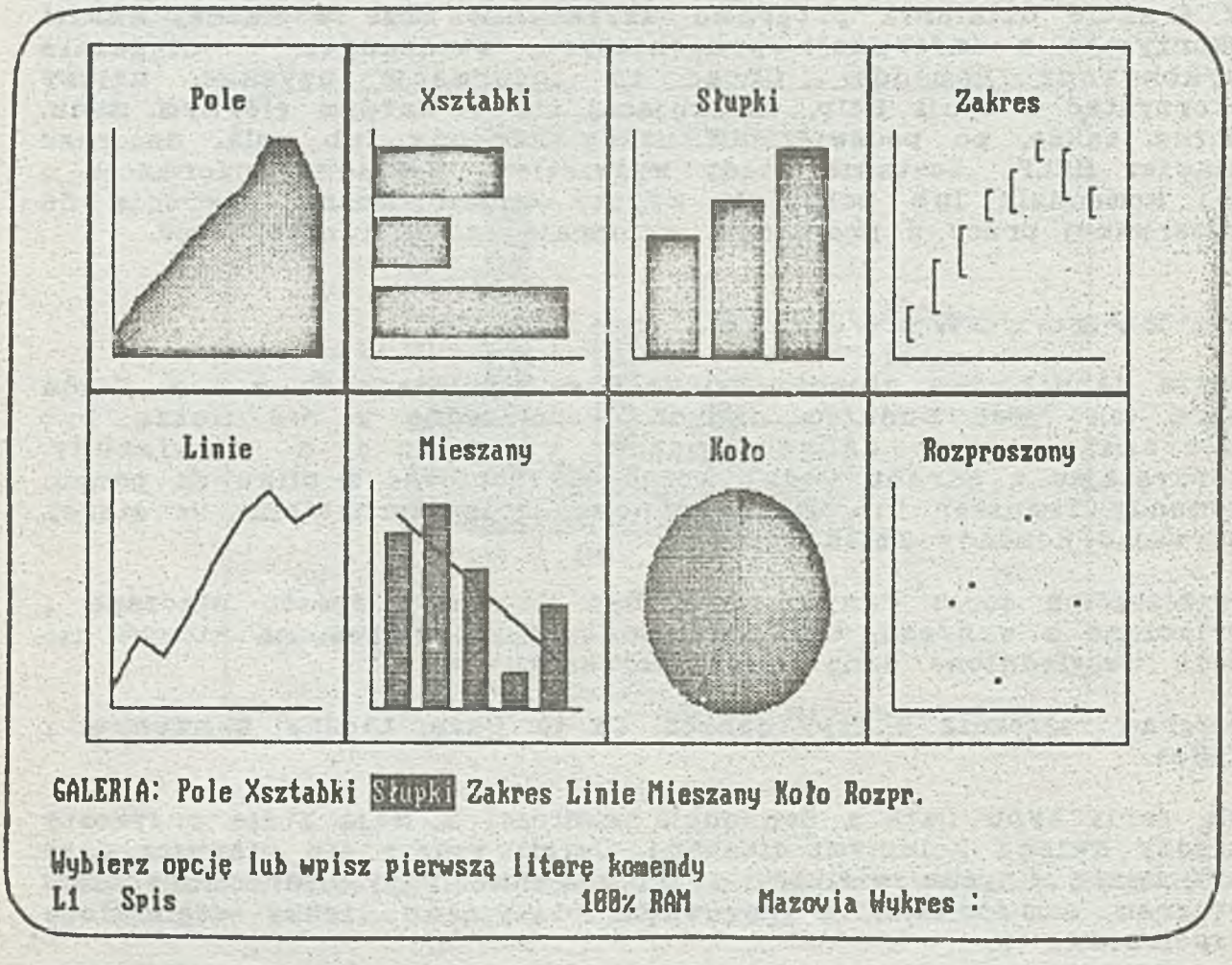
Typy wprowadzanych liczb mogą być całkowite, dziesiętne i zmiennoprzecinkowe. Liczby mogą mieć do 14 cyfr znaczących.

Serie danych mogą zawierać maksimum 64 punkty, a wykres maksimum 512 punktów.

Program może przeprowadzić analizę wprowadzonych danych, tj. dokonać siedmiu różnych analiz: kumulacyjną, procentową, różnicową, statystyczną, średnią, trendu i wzrostu.

Tworzenie wykresów

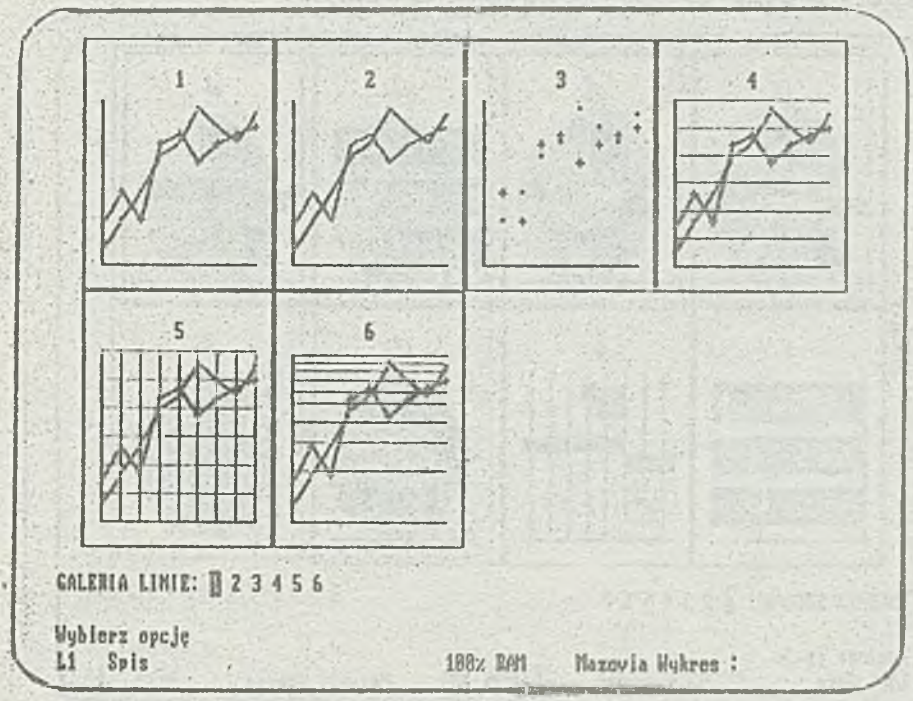
Jeżeli na ekran Spis została wprowadzona przynajmniej jedna seria, można utworzyć już wykres.Do wyboru odpowiedniego typu wykresu służy komenda Galeria. Po wybraniu tej komendy na ekranie zostanie wyświetlone 8 podstawowych typów wykresów (rys.3).



rys.3

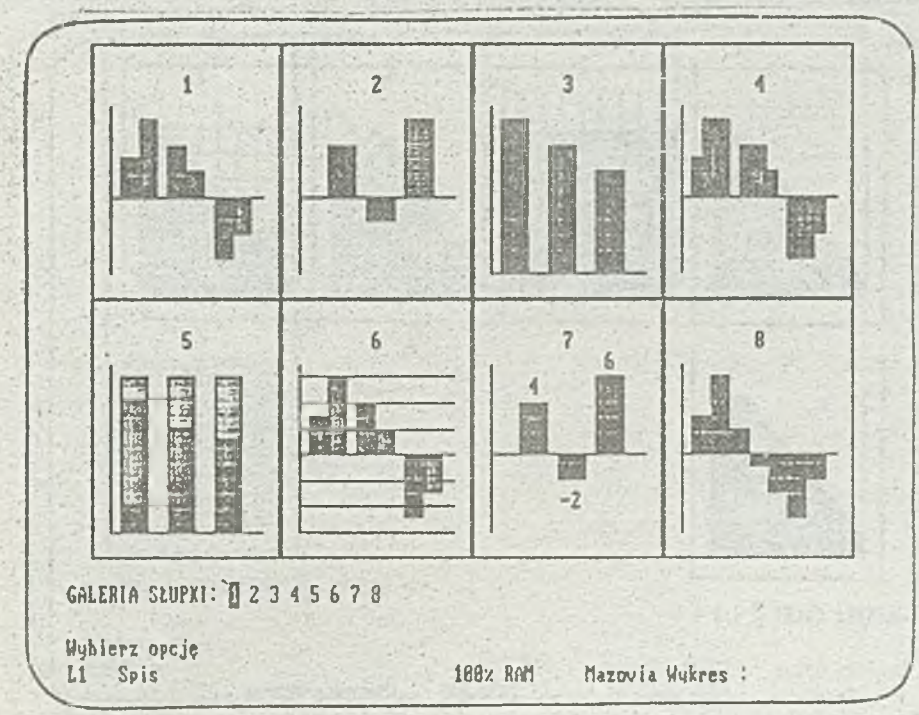
Każdy podstawowy typ ma swoją galerię zawierającą kilka (5 do 8) wariantów danego typu.

Wykresy liniowe mogą być wykorzystywane do ilustrowania tendencji danych oraz w seriach danych z dużą liczbą punktów na przestrzeni pewnego okresu czasu (rys.4).



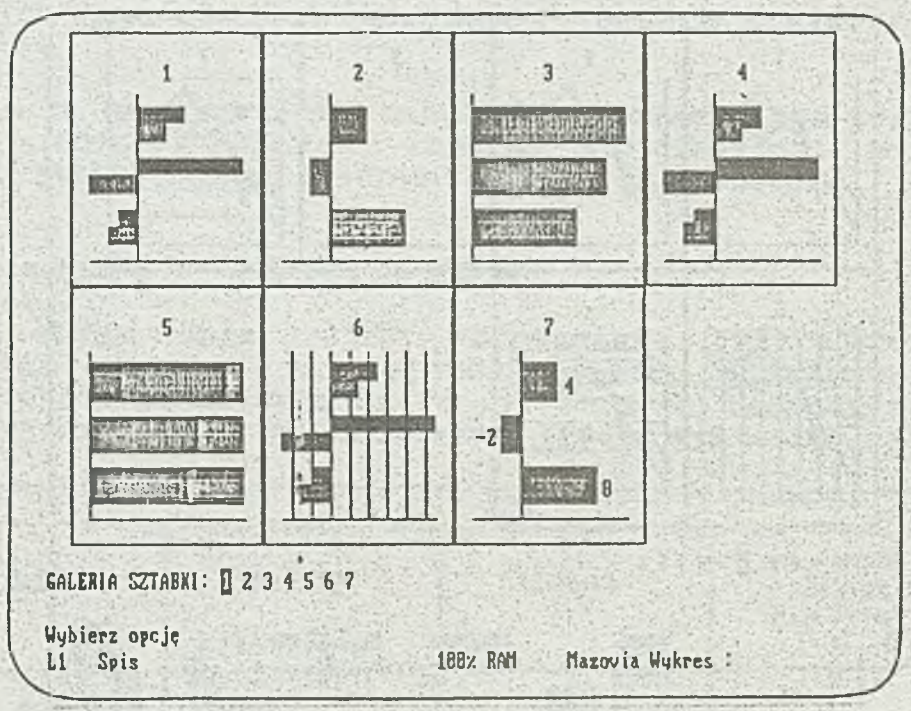
rys.4

Wykresy słupkowe są przydatne do ilustrowania serii danych z małą liczbą punktów w danym okresie czasu lub gdy występuje ekstremalna zmienność danych w serii (rys.5).



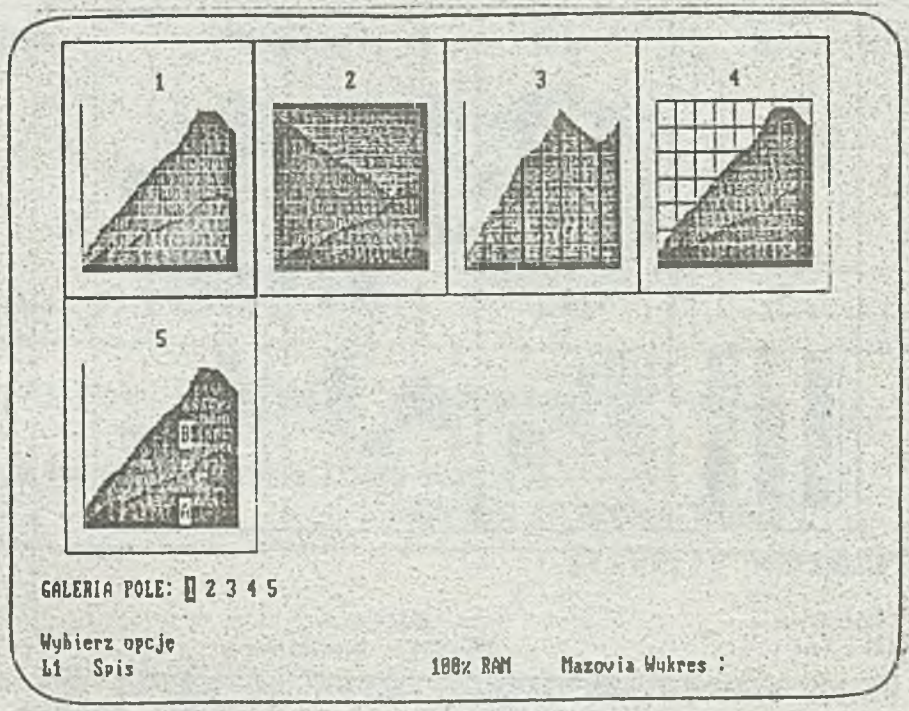
rys.5

Wykresy sztabkowe są najlepsze do porównywania serii danych, których współrzędne x są tekstem (rys.6).



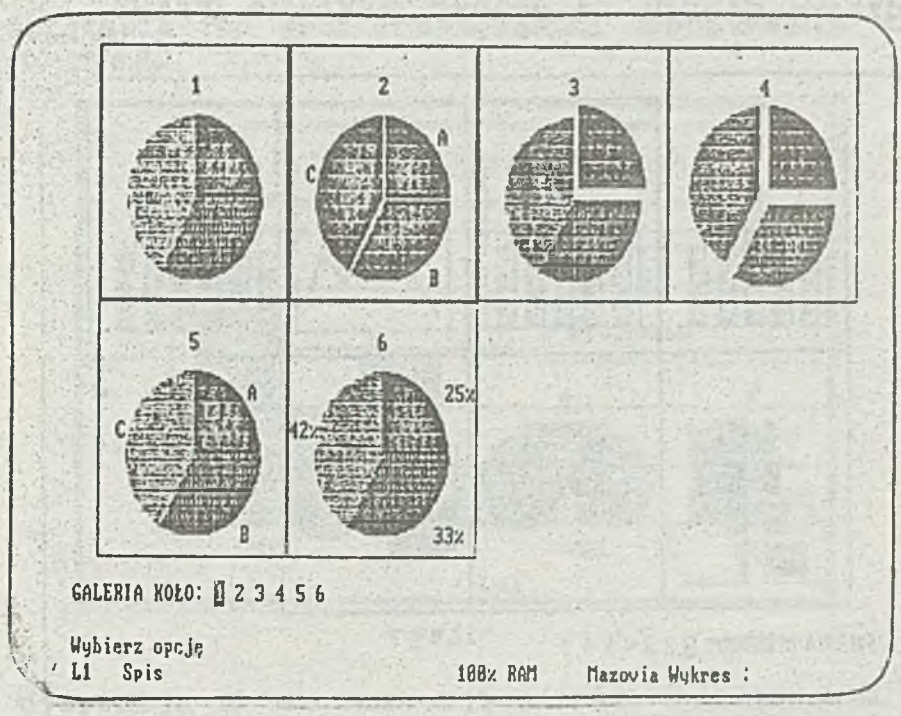
rys.6

Wykresy powierzchniowe są przydatne do efektownego ilustrowania zarówno wartości danych w seriach w danym okresie czasu, jak i tendencji danych w tym okresie (rys.7).



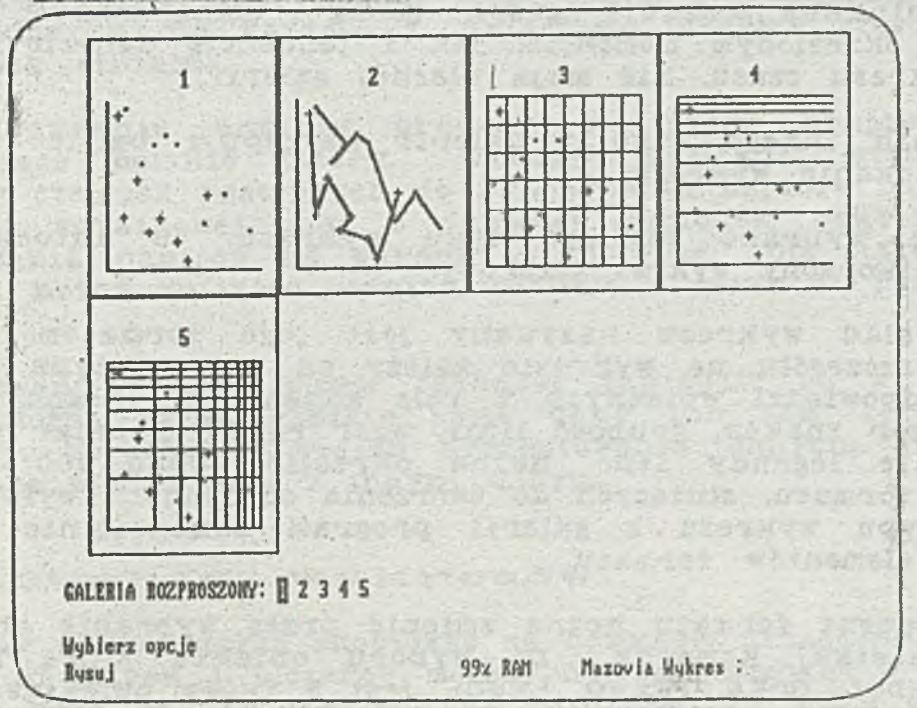
rys.7

Wykresy kołowe są przydatne do ilustrowania względnych proporcji różnych punktów danych w określonym momencie (rys.8).



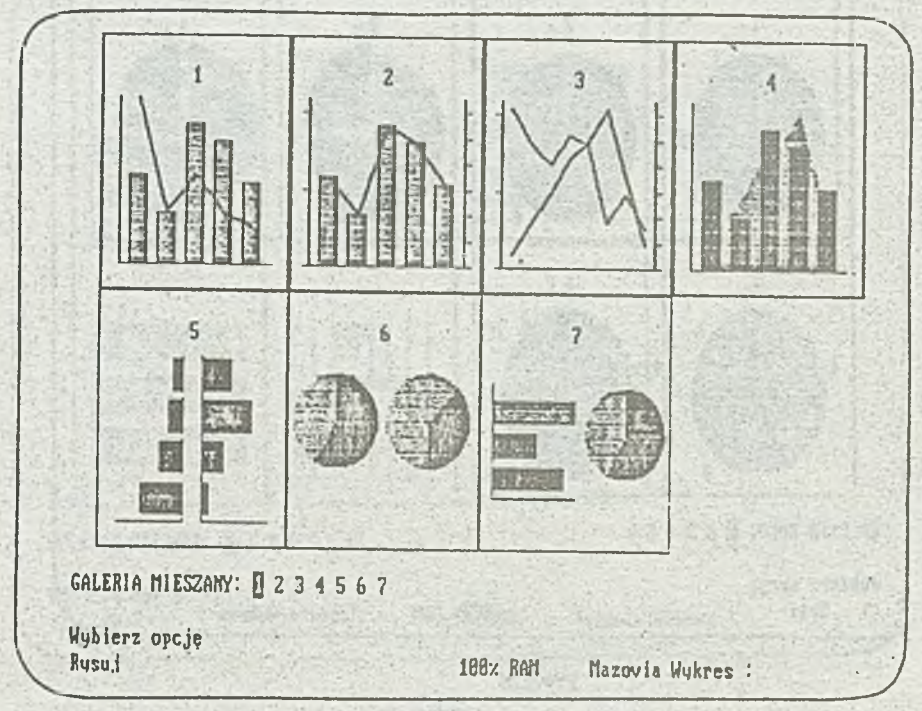
rys.8

Wykresy rozproszone są przydatne do ilustrowania dużej liczby danych statystycznych lub doświadczalnych, ukazujących relacje pomiędzy zmiennymi (rys.9).



rys.9

Wykresy mieszane, czyli jeden wykres nałożony na drugi lub dwa wykresy umieszczone obok siebie, mogą być używane do porównywania różnych rodzajów danych na jednym wykresie (rys.10).



rys.10

Wykresy minmaks pokazują zarówno górną i dolną granicę wartości danych w określonym momencie, jak i tendencję danych w ciągu pewnego okresu czasu. Nie mają własnej galerii.

Po wybraniu innego typu z Galerii następuje bardzo szybkie przeformatowanie wykresu.

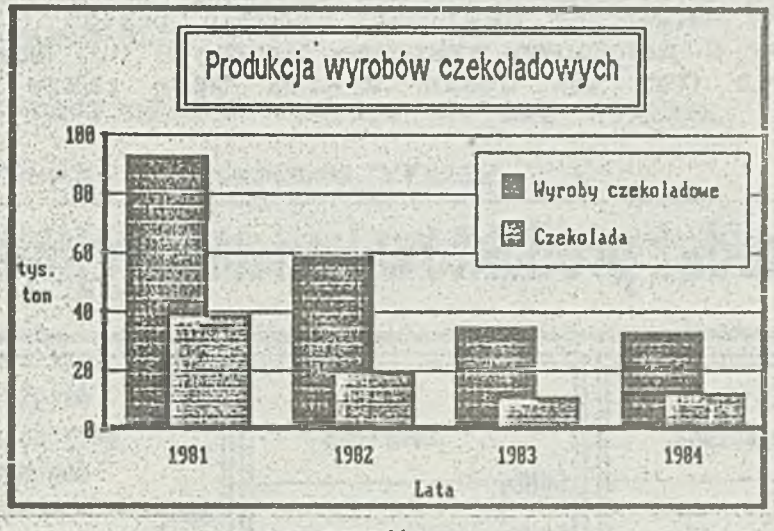
Jeżeli nie wybrano żadnego typu wykresu, to automatycznie zostanie utworzony wykres słupkowy.

Ogólny wygląd wykresu nazywany jest jego formatem. Wygląd dowolnego szczegółu na wykresie zależy od wyboru typu wykresu oraz od odpowiedzi wpisanych w pole komend, np. można określić kolor i wzór znaków, grubość linii, wzór ramki, rozmiar wykresu, umieszczenie legendy itp. Można określić ponad 100 różnych elementów formatu, służących do tworzenia odmiennych wykresów. Po wyborze typu wykresu z galerii program samodzielnie określa większość elementów formatu.

Dowolny element formatu można zmienić przez wybranie obiektu i użycie właściwej komendy. Do wyboru obiektu służą klawisze kierunku (DÓŁ, GÓRA, PRAWO, LEWO). Jest 5 typów obiektów: wykres jako całość, osie, seria lub punkt danych, opisy, legenda. Na wykresie może być więcej niż jeden obiekt każdego z wymienionych

typów. Opisy i legendę można umieszczać w dowolnym miejscu na ekranie.

Rys.11 przedstawia (po przeformatowaniu) zobrazowanie tych samych danych co rys.2.



rys.11

Program dysponuje 15 wzorami wypełnienia powierzchni, 10 wzorami znaczników (do oznaczania punktów), pięcioma rodzajami linii, trzema grubościami linii, czterema rodzajami znaków podziałki, czterema sposobami opisów znaków podziałki, pięcioma rodzajami ramek. Na wykresach można zmieniać punkt przecięcia się osi, prowadzić główne i pomocnicze linie siatki, wprowadzić linie odnośnikowe oraz linie między maksymalną i minimalną wartością współrzędnej y dla każdego y, uwidocznić dowolne części wykresu za pomocą strzałki.

Do wykonywania napisów program ma opisy krojów czcionek zawierające polskie litery. Istnieje możliwość stosowania 13 rodzajów czcionek rastrowych do wykonywania napisów na ekranie i drukarce mozaikowej oraz 8 rodzajów czcionek wektorowych do wykonywania napisów na ekranie i ploterze. Dla każdego kroju czcionek można zmieniać rozmiar, co daje w sumie 64 różne rodzaje pisma.

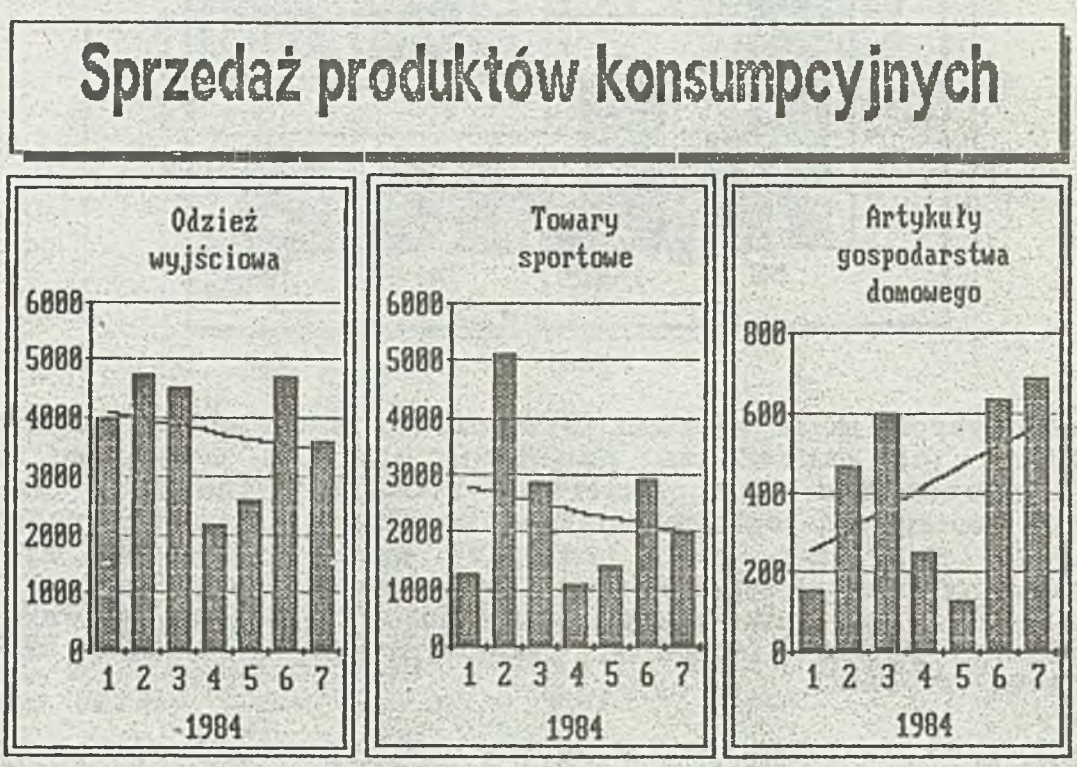
Przygotowany wykres można zachować zapisując go w całości na dysku. Można także zapisać same dane lub sam format, czyli najlepszy, opracowany sposób prezentowania danych, aby później używać go wielokrotnie z innymi danymi.

Kopiowanie wykresów

Utworzony wykres można skopiować na dowolnym urządzeniu wybranym z listy drukarek i ploterów mogących współpracować z programem. Wykresy mogą być sporządzane na papierze lub kalce w dostępnej gamie kolorów.

Praca z większą liczbą wykresów

Program WYKRES pozwala na tworzenie i wyświetlanie do 16 różnych wykresów umieszczonych razem na ekranie lub na tej samej stronie. Każdy wykres może być umieszczony w dowolnym miejscu ekranu lub strony. Wykresy mogą się nakładać, tworząc postać mieszaną (np. wykres słupkowy z nałożonym wykresem liniowym) lub mogą znajdować się obok siebie (rys 12). Jeden wykres może także przesłaniać drugi (rys.1).



rys.12

Współpraca z innymi programami

Program pozwala na wczytywanie plików utworzonych przez programy MULTOPLAN, LOTOS A B C, VisiCalc, BaZAd 2, a także innych programów, które używają formatów DIF, BAZAd 2, ASCII i SILK. Program akceptuje dane z tych programów, przekształcając je na odpowiednią dla siebie postać.

Wymagania sprzętowe

Do pracy z programem WYKRES są potrzebne:

- komputer kompatibilny z IBM PC/XT/AT, z co najmniej 128kB pamięci (zalecane 192kB pamięci),

- przynajmniej jeden dwustronny napęd dyskowy,

- karta CGA lub Herkules,

- 80-kolumnowy monitor graficzny lub 80-kolumnowy czarno-biały monitor, lub 40-kolumnowy graficzny monitor typu telewizyjnego , lub monochromatyczny monitor komputera kompatibilnego z IBM (tylko z karta Herkule-s),

- system operacyjny DOS (wersja 2.0 lub wyższe),

- kopia dyskietki z programem WYKRES,

- kopia dyskietki z krojami czcionek, używanymi przez program WYKRES.

TECHNIKI KOMPUTEROWE 5-6/88

mgr inż Maria GRZEGOREK

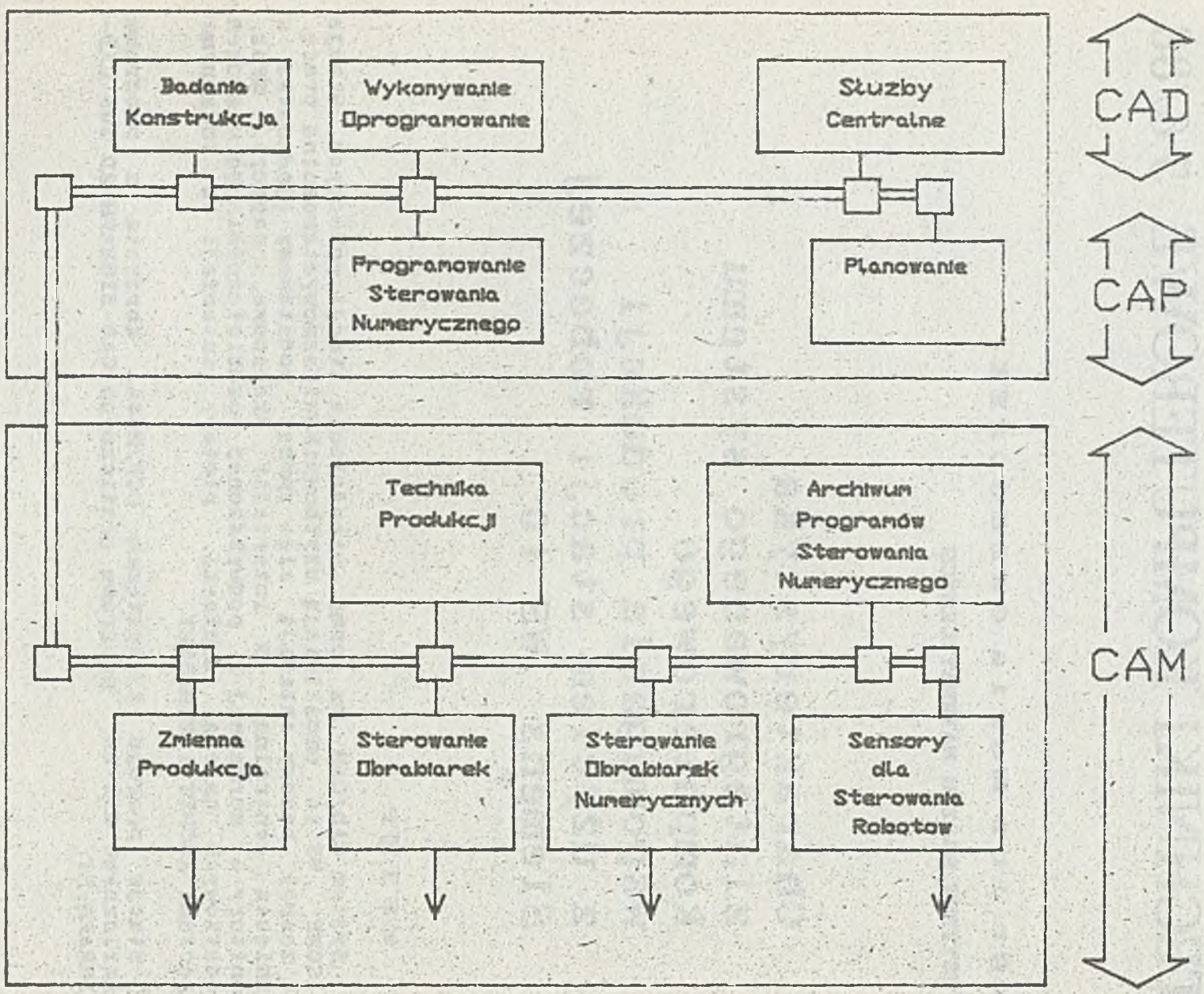
INSTYTUT MASZYN MATEMATYCZNYCH

Charakterystyka zintegrowanego systemu komputerowego wspomagania produkcji z użyciem stacji roboczej Siemens WS 10

WSTĘP

System CAD/CAM ma formę pierścienia stacji roboczych. Stacje SICOMP WS 10 umożliwiają użytkownikowi samowystarczalną pracę. Na zespół bazowy składają się: sprzęt podstawowy (jednostka z pamięcią centralną, 3 interfejsy szeregowe, napędy), mysz, monitory o normalnej i podwyższonej rozdzielczości. Do urządzeń dodatkowych należą drukarki, ploter, interfejs z lokalnymi sieciami ARCNET i ETHERNET.

Stacja pracuje z systemem CCP/M-86, korzysta z programu graficznego GSX-86, programu graficznego CAD-SIGRAPH-2D lub CAD-SIGRAPH-ET.



rys. 1. Koncepcja CAD/CAM

KONCEPCJA CAD/CAM

Komputery wprowadzone są do całego procesu wytwarzania, zaczynając od prac badawczo-rozwojowych i konstrukcyjnych (CAD) przez planowanie aż po produkcję (CAM). W kolejnych etapach wytwarzania (CIM) wspomagają następujące czynności:

* rozwiązywanie zadań konstrukcyjnych zależnych od geometrii,
* produkcja na podstawie danych i planów produkcji,
* przygotowanie sterowania produkcji i automatyzacji przepływu materiałów na podstawie planów.

Przedstawiona dalej koncepcja CAD/CAM, w formie pierścienia stacji roboczych, optymalnie integruje wspomaganie komputerowe w wymienionych etapach (rys. 1).

Ma ona następujące zalety:

* niska cena jednostkowa,
* nieskomplikowana obsługa,
* bogate oprogramowanie (również dla części biurowej),
* możliwość stopniowej rozbudowy (np. w sieci lokalne),
* możliwość przyłączania urządzeń peryferyjnych (np. ploter, drukarka), do których mają dostęp wszystkie stacje robocze,
* możliwość przyłączania wspólnych banków danych,
* różnorodne możliwości sprzężenia z macierzystym komputerem.

STACJA ROBOCZA SICOMP WS 10

Stacja umożliwia użytkownikowi samowystarczalną pracę na urządzeniu liczącym, przystosowanym do jego potrzeb. Ma on do swojej dyspozycji procesor, pamięć i moc obliczeniową. Nie ma uciążliwych czasów oczekiwania występujących zwykle w dużych maszynach. Połączenie pojedyńczych stacji roboczych w wydajne sieci lokalne zapewnia szybki przepływ informacji między użytkownikami zajmującymi się poszczególnymi zakresami prac. Możliwa jest bezpośrednia wymiana danych między sferami konstrukcji, przygotowaniem wytwarzania i sterowaniem produkcją. Tak np. w czasie prac konstrukcyjnych uzyskuje się w systemie CAD dane do programowania obrabiarek i do zestawienia listy detali dla planowania i sterowania produkcją. Z drugiej strony konstruktor rozporządza wynikami produkcji i kontroli jakości.

Stacje WS 10 mogą być łączone ze sobą w dwie sieci lokalne, tj.

ARCNET i ETHERNET

Sprzężenie z obu mechanizmami sieciowymi jest opisane bardziej szczegółowo w dalszej części.

System bazowy stacji roboczej SICOMP WS 10 składa się z monitora czarno-białego 31 cm, monitora kolorowego 36 cm, drukarki PT 88, klawiatury i myszy z podłożem.

Koszt stanowiska pracy CAD - SICOMP WS 10 z programem SIGRAPH jest niższy niż stosowane dotąd tylko w wielkim przemyśle pakiety programów użytkowych CAD i związany z nimi sprzęt, a więc nadają się do wykorzystania w średnich przedsiębiorstwach i biurach konstrukcyjnych.

ELEMENTY SKŁADOWE STACJI ROBOCZEJ SICOMP WS 10

Zespół bazowy

- sprzęt podstawowy

  * jednostka centralna:
    - z procesorem 16-bitowym 8088,
    - z procesorem arytmetycznym 8087,

  * pamięć centralna 2^{10} bajtów,
  * trzy interfejsy szeregowe dla:
    - plotera, drukarki, adaptera komunikacji,

  * napędy pamięci zewnętrznej:
    - napęd dyskietek 720 Kbajtów,
    - napęd dysku stałego 10 lub 40 Mbajtów,

- wejście graficzne

  * mysz

- wejście alfanumeryczne

  * monitor o normalnej rozdzielczości
    - czarno-biały 31 cm (12"), 560x275 rastrów,

- wyjście graficzne

  * monitor o normalnej rozdzielczości
    - 8 kolorów, 36 cm (14"), 640x480 rastrów,

  * monitor graficzny o podwyższonej rozdzielczości
    - czarno-biały lub 8 kolorów 47 cm (19"), 1200x1024 rastrów.

Urządzenia dodatkowe

- drukarka PT88 i PT89 z pamięcią buforową $4x2^{10}$ bajtów

  * drukarka natryskowa 150 znaków/s, głośność 145 dB,
  * drukarka mozaikowa 80 znaków/s, głośność 160 dB,
    obie: 96 znaków (duże i małe litery),
    szerokość strony: PT88 - do 250 mm,
    PT89 - do 400 mm,

  b) ploter (do grafiki kolorowej) format A4 ÷ A0

- interfejs

  do sprzężenia z lokalnymi sieciami ARCNET i ETHERNET,

- interfejsy:

  * do komputerów,
  * do procesu przez specjalne adaptery SIMATIC S5 i szyny IEC.

Oprogramowanie

Do oprogramowania podstawowego sprzętu należą:

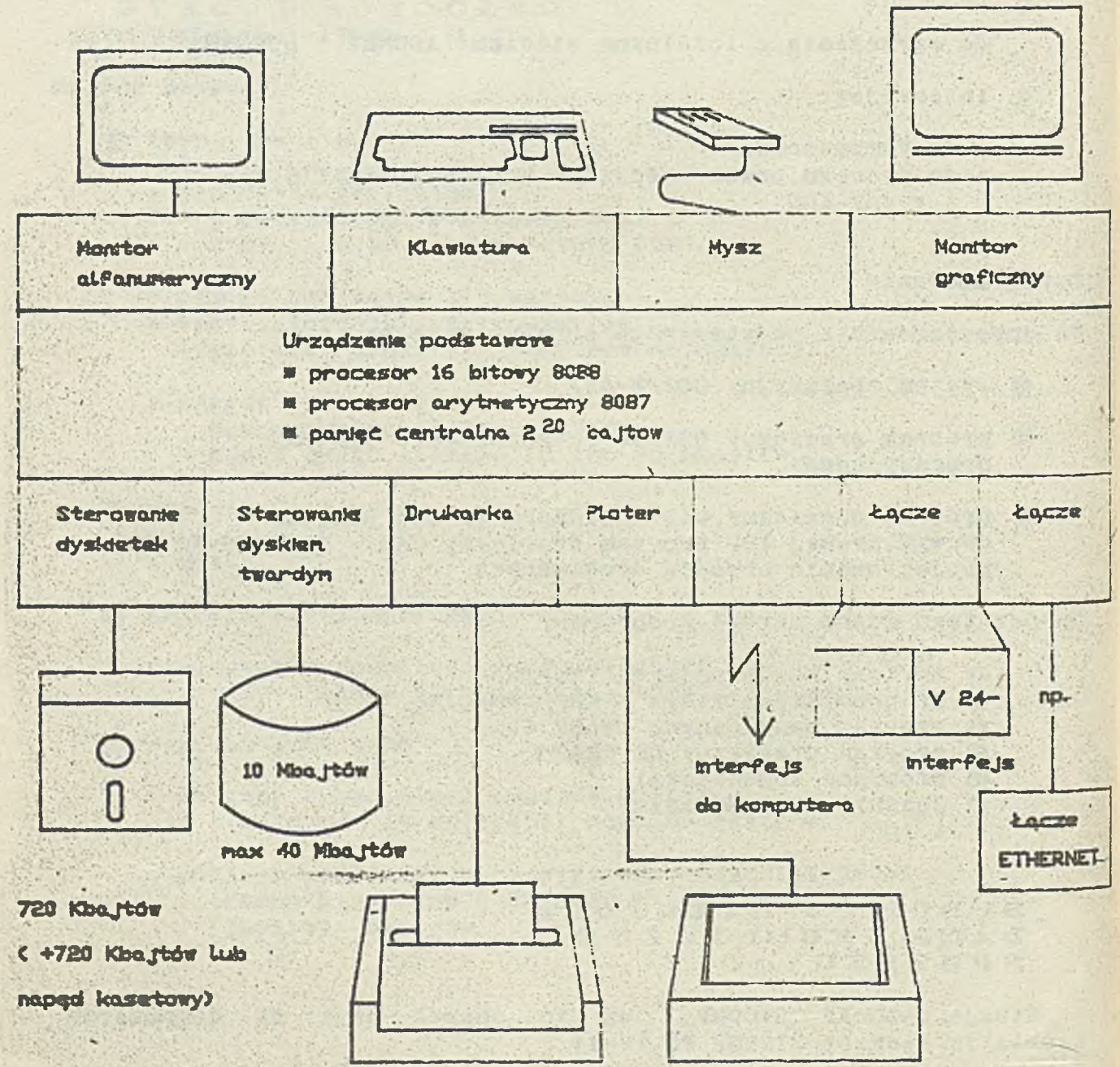
- system operacyjny CCP/M-86,
- program graficzny GSX-86 (rozszerzenie systemu operacyjnego),
- program graficzny CAD - SIGRAPH-2D dla grafiki dwuwymiarowej lub program graficzny CAD - SIGRAPH-ET do projektowania obwodów drukowanych.

Oprócz tego można używać programów:

- edytory WordStar z MailMerge,
- program kalkulatora MIKROPLAN/LINK,
- program bazy danych dBASE II,
- program graficzny DR-GRAPH,
- programy komunikacji,
- języki programowania,
  i in.

BUDOWA I NIEKTÓRE DANE TECHNICZNE SPRZĘTU

Stacja robocza SICOMP WS 10 oparta jest na komputerze osobistym Siemens SICOMP PC 16-11



rys. 2. Konfiguracja sprzętu stacji roboczej SICOMP WS 10

Zbudowana jest z następujących elementów (rys. 2):

- urządzenia podstawowe

  - procesor 16-bitowy 8088 i procesor arytmetyczny 8087,

  - pamięć centralna (RAM) 2 zespoły po $512x2^{10}$ bajtów,

  - interfejsy do drukarki, plotera i komputera (24 V)

    pobór mocy ok. 120 VA,
    napięcie sieci ~220 V +10%-15%,
    częstotliwość 48÷63 Hz,
    wymiary 440x225x340 mm,
    waga ok. 15 kg.

  - napęd dyskietek 5 1/4"

    pojemność 720 kbajtów,
    80 ścieżek na stronie,
    sektory o zmiennej długości, 9 sektorów na ścieżce,
    512 bajtów w sektorze,
    zapis danych MFM,
    gęstość zapisu 2331 bit/mm (5922 bpi),
    obroty wrzeciona 300 obr/min,
    średni czas pozycjonowania 94 ms,

  - napęd pamięci kasetowej

    standardowa kaseta taśmy magnetofonowej 1/4",
    pojemność 45 Mbajtów,
    długość bloków 512 bajty,
    gęstość zapisu 315 bit/mm (8000 bpi),
    szybkość przesuwu taśmy 229 cm/s,

  - napęd dysku twardego

	10 Mbajtów	40 Mbajtów
nośnik	2 talerze 5 1/4"	4 talerze 5 1/4"
liczba ścieżek	306	512
sektorów na ścieżce	16	19
zapis	MFM	MFM
gęstość zapisu	356,2 bit/mm	362 bit/mm
obroty wrzeciona	3600 obr/min	3529 obr/min
śr. czas pozycjonowania	85 ms	45 ms
szybkość przesyłania danych	$5x10^6$ bit/s	$3.856x10^6$ bit/s
wymiary		80x147x204 mm
waga	2.0 kg	2.7 kg

- monitory

z normalną rozdzielczością, do porozumiewania się między użytkownikiem i aktywnym programem oraz przedstawiania prostej grafiki:

- czarno-biały
  wielkość ekranu 310 mm (12"),
  rozdzielczość 560x275 rastrów,
  znaków w wierszu 80,
  liczba znaków - generator ze 128 znakami,
  wielkość pola znaku - matryca 11x7,
  kolor ekranu zielony,
  częstotliwość odnawiania ekranu 50 Hz,
  wymiary 316x280x310 mm,
  waga ok. 6 kg,

- kolorowy
  wielkość ekranu 360 mm (14"),
  rozdzielczość 640x480 rastrów,
  8 kolorów,
  częstotliwość odnawiania obrazu 60 Hz,
  wymiary 362x330x405 mm,
  waga 9 kg,

- stacje graficzne z podwyższoną rozdzielczością dla całej grafiki - czarno-białe i kolorowe

  wielkość ekranu 471 mm (19"),
  rozdzielczość 1280x1024 rastrów,
  częstotliwość odnawiania obrazu 83 Hz,
  rozmiary 480x403x462 mm czaro-biała,
  480x403x475 mm kolorowa,
  waga odpowiednio ok. 21 i 23 kg,

sterownik	czarno-biała	kolorowa
napięcie sieci	220 V~	220 V~
pobór mocy	120 VA	160 VA
rozmiary	480x155x519 mm	480x275x521 mm
waga	10 kg	13 kg

- klawiatura

  międzynarodowy zbiór znaków,
  klawisze alfanumeryczne, klawisze numeryczne, 16 programowalnych klawiszy funkcyjnych, klawisze funkcyjne sterowania kursorem,
  wymiary 500x30x250 mm,
  waga 1,7 kg,

- mysz z podłożem

  3 przyciski sterujące,
  rozdzielczość 0.254 mm,
  rozmiary myszy 66x28x48 mm,
  podłoża 279x228 mm,

Zestaw można rozszerzyć o następne elementy:

- drukarka natryskowa PT88 lub mozaikowa PT89 do edycji tekstu lub grafiki czarno-białej

	PT88	PT89
szybkość druku	150 zn/s	80 zn/s
głośność	≤45 dB	≤60 dB

interfejs szeregowy V 24,
kod 7-bitowy ASCII,
szybkość transmisji 9600 bitów/s,
zbiór znaków - 96 (duże i małe litery),
8 zestawów znaków,
matryca znaków 9x9,
maks. 136 zn. w wierszu (zależnie od odstępu między znakami),
szerokość pisma normalnego 10, 12 lub 17 zn/cal,
szerokiego 17 zn/cal,
papier perforowany składany 124 : 400 mm,
pojedyncze kartki 105 : 400 mm,
na rolce 105 ÷ 216 mm,
przesuw 4.23 lub 3.17 mm programowany w 0.35 mm krokach,
napięcie ~220 V,
częstotliwość 50 Hz,
pobór mocy 30 W,

	PT88	PT89
wymiary	410x140x310 mm	580x148x313 mm
waga	7.5 kg	10 kg

- ploter do edycji grafiki kolorowej C1603

format rysunków maks. 294x400 mm,
format papieru A3 i rolkowy maksymalna szerokość 297 mm,
materiał rysunków: papier, folia matowa i przezroczysta,
mocowanie papieru elektrostatyczne,
pisaki maksymum 8 kolorów,
szybkość maksymalna 70 cm/s,
6 zn/s przy wysokości pisma 3 mm,
najmniejsze adresowalne przesunięcie 0.025 mm,
powtarzalność ± 0.1 mm bez zmiany pisaka,
pamięć buforowa $2x2^{10}$ bajtów,

rozkazy makro: ruch, znak, zbiór znaków (6 zestawów), wybór pisaka, szybkości, koło i łuk koła, segment koła, znacznik, oś, typ linii, współczynnik skali, przesunięcie początku układu współrzędnych,
pobór mocy 40 VA,
wymiary 620x190x530 mm,
waga 20 kg.

PRZESTRZEŃ ROBOCZA

Przestrzeń robocza jest tak ukształtowana, że ułatwia i upraszcza dialog człowieka z komputerem. Rysunek opracowany przez użytkownika jest wyświetlony na monitorze kolorowym lub stacji graficznej. Monitor czarno-biały służy do podania menu i instrukcji dialogowych. Za pomocą myszy rozpoznawane są symbole graficzne lub tekstowe i przesuwany kursor monitora kolorowego lub czarno-białego (zależnie od stanu programu). Z klawiatury wprowadza się dane alfanumeryczne.

OPROGRAMOWANIE

● Funkcje globalne oprogramowania użytkowego CAD SIGRAPH:

- rysowanie

Funkcja może zobrazować wszystkie powstałe rysunki i rysunki przechowywane pod określoną nazwą w pamięci zewnętrznej.

- definiowanie warstw

Istnieje 15 wewnętrznych warstw, na których można określić elementy rysunku, a następnie nałożyć je na siebie w dowolnej kolejności i wykreślić na ploterze lub narysować na ekranie. Funkcja definiuje warstwę, na której będzie się wykonywać następne kroki procesu konstruowania.

- definicje podstawowych elementów

Funkcja pozwala uzyskać informacje np. długość, promień, charakterystykę konstruowanych właśnie prostych, kół, łuków kół itp.

- definicja nowego rodzaju linii

Funkcja definiuje rodzaj linii na powierzchni dla leżących na niej elementów. Dostępne są 4 kolory, 4 grubości linii, 4 rodzaje linii.

- zmiana podziałki

Funkcja ustala związek między podziałką pierwotną, w której powstaje konstrukcja i podziałką odwzorowania określającą wielkość wycinka na ekranie graficznym.

- dołączanie obrazów

Funkcja wymazuje i wstawia rysunki na monitorze.

- zmiana strony

Przy powstawaniu większych rysunków istnieje możliwość zmiany fragmentu, który jest wyświetlany na monitorze. Aktualny fragment rysunku może być przesuwany w ośmiu kierunkach, przy czym zachowane jest ok. 25% przykrywanie stron.

- powrót do poprzedniego menu

Za pomocą tej funkcji można wrócić w każdej chwili do poprzedniego zakresu menu.

● Funkcje lokalne oprogramowania użytkowego CAD SIGRAPH

Oprócz funkcji globalnych, dostępnych w każdej chwili, SIGRAPH dysponuje wieloma funkcjami lokalnymi. Są one umieszczone w hierarchicznie uporządkowanych zakresach menu. W danej fazie konstrukcji są dostępne tylko stosowne dla niej funkcje.

SIGRSPH zawiera następujące zakresy menu:

elementy podstawowe,
funkcje złożone,
wymiarowanie,
funkcje ogólne,
restart programu,
koniec programu.

Funkcje wybiera się za pomocą myszy zaznaczając odpowiednie pole menu na monitorze czarno-białym.

- elementy podstawowe

Szereg geometrycznych elementów podstawowych umożliwia szkicowanie złożonych rysunków. Program dysponuje następującymi elementami podstawowymi:

punkt,
prosta,
koło i łuk koła (istnieje osobna funkcja zaokrąglania naroży),
prostokąt.

- funkcje elementarne

Służą one do zmiany elementów rysunku, a m. in. do:

usuwania elementów,
odbicia lustrzanego elementów,
fazowania elementów,
kreskowania,
zmiany rodzaju linii,
zmiany długości prostych i łuków,
zmiany warstwy,
obliczenia pola przekroju.

- funkcje złożone

Funkcje złożone składają się z wielu funkcji elementarnych. Za pośrednictwem nazwy, pod którą są przechowywane na nośniku zewnętrznym można je wywołać. Funkcja złożona może być zestawiona z innych funkcji złożonych, aby ułatwić posługiwanie się nimi oprogramowanie CAD zawiera następujące funkcje:

generacja funkcji złożonych przez zaznaczenie ramką,
przechowywanie na zewnętrznym nośniku,
kopiowanie na rysunek,
przesuwanie,
obrót,
skalowanie (jednakowa skala w obu kierunkach),
usuwanie,
odbicie lustrzane wg dowolnej prostej,
kopiowanie wielokrotne (przesuwanie o moduł kątowy lub liniowy).

- wymiarowanie

Linie wymiarowe i pomocnicze rysowane są automatycznie. Użytkowenik zaznacz tylko element wymiarowany, jak również format i wysokość liczby wymiarowej. Liczba ta umieszczana jest w środku linii wymiarowej.

- funkcje ogólne

Do funkcji ogólnych należą:

usuwanie funkcji złożonych,
kreślenie,
definiowanie punktu odniesienia (za pomocą kursora graficznego). Do tego punktu odnoszą się wszystkie transformacje i zmiany.

- program restartu

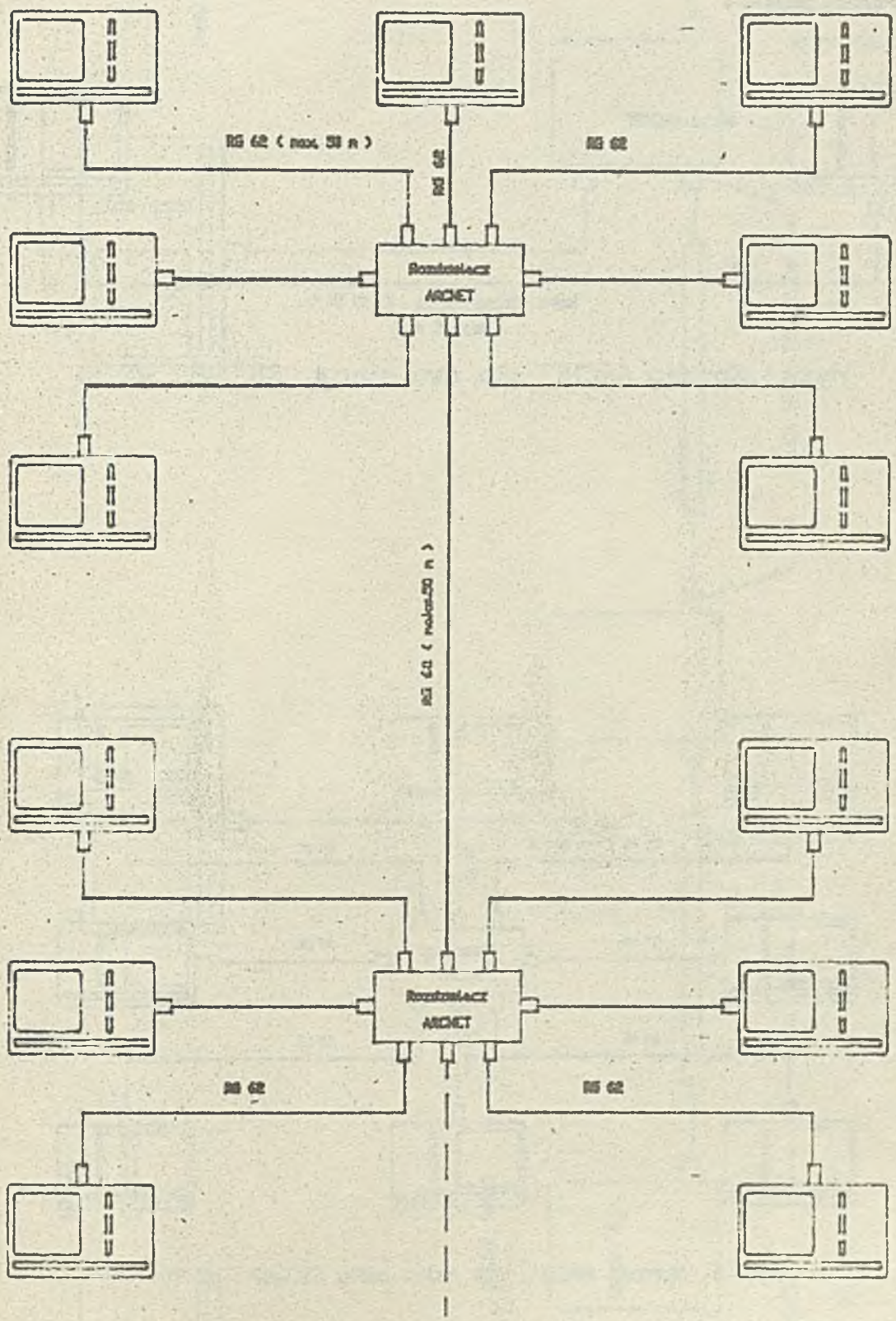
Funkcja restartu programu przywraca programowi stan początkowy. Można więc zapamiętać powstały rysunek i zacząć pracę nad nowym.

INTERFEJS Z SIECIAMI LOKALNYMI

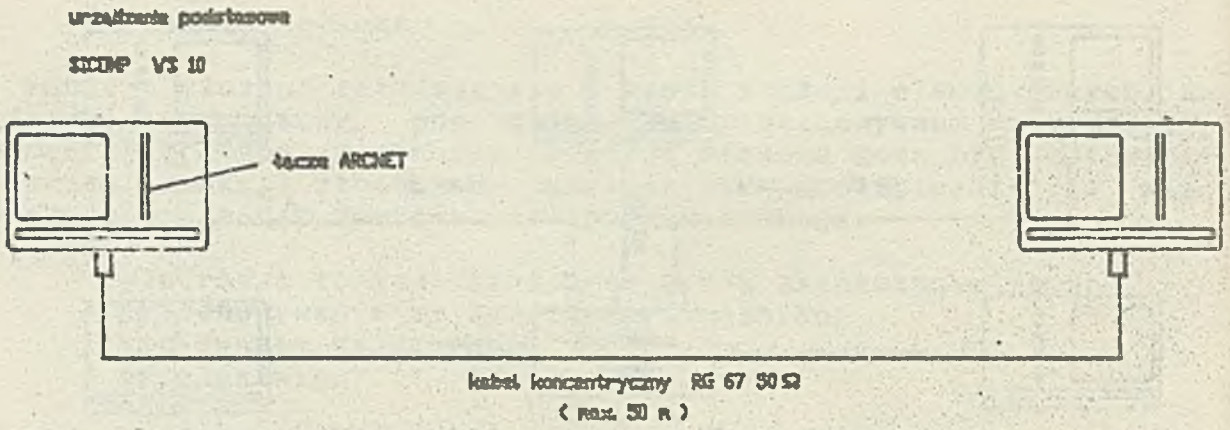
● Elementy sieci

Łącza ARCNET i ETHERNET umożliwiają użytkownikowi WS 10 budowę dwóch rodzajów szeroko rozpowszechnionych sieci lokalnych. W połączeniu z systemem zarządzania siecią DR-NET można:

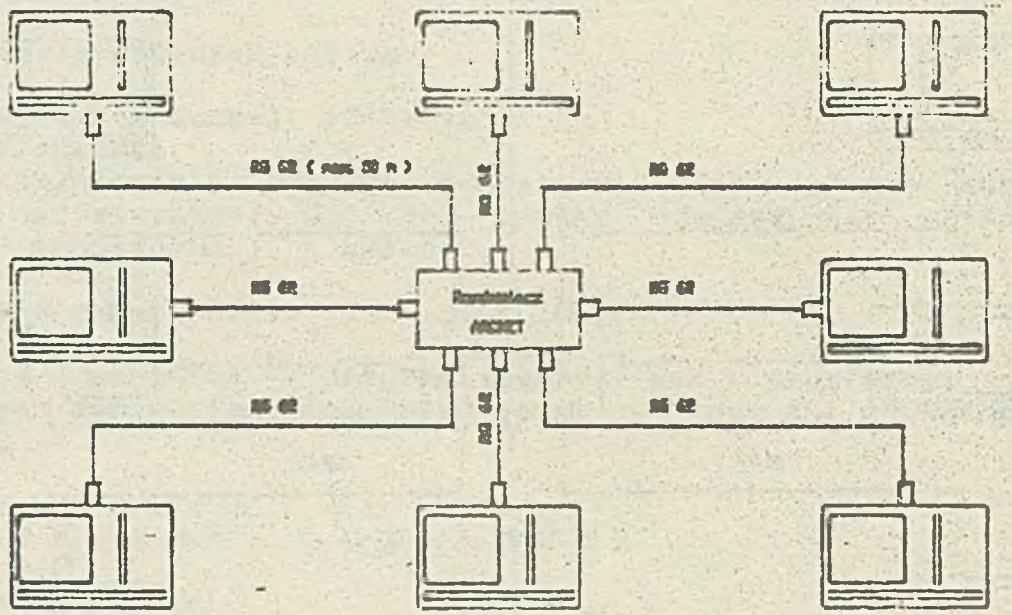
uruchomić jednocześnie do 255 stacji WS 10,
używać urządzenia zewnętrznych takiich, jak drukarka, napędy, programy wspólnie z innymi uczestnikami.



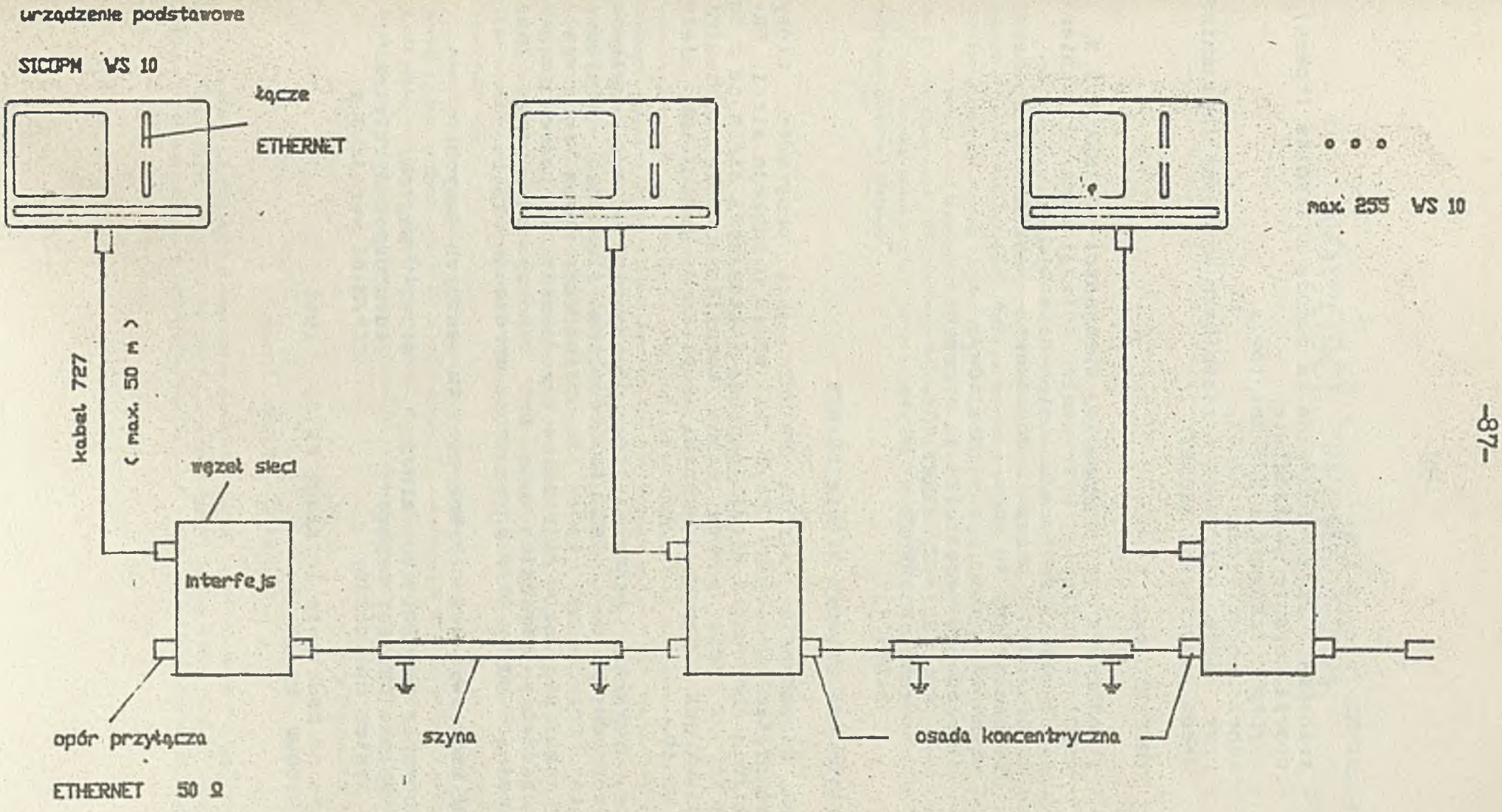
rys, 3.Sprzeg dla 9 do 255 SICOMP WS 10



rys. 4. Sprzęg ARCNET dla dwu stacji SICOMP WS 10



rys. 5. Sprzeg ARCNET dla maks. ośmiu SICOMP WS 10



rys. 6. Interfejs ETHERNETj max. długość wszystkich kabli 500 m

cechy ARCNET (rys. 3,4,5)

- transmisja synchronizowana za pomocą znacznika (token),
- uporządkowanie gwiaździste,
- szybkość transmisji 2.5 Mbajtów/s,
- kabel koncentryczny RG 62,
- przy więcej niż dwóch uczestnikach wymagany conajmniej jeden rozdzielacz ARCNET.

cechy ETHERNET (rys. 6)

- standardowa sieć o wysokiej sprawności wg IEEE 802.3,
- wspólny dostęp z wykrywaniem kolizji CSMA/CD (canier-sense-multiple-acces/colision-detection),
- transmisja przez dodatkowo ekranowany kabel koncentryczny 50 omów,
- szybkość transmisji 10 Mbajtów/s,
- dla jednego uczestnika są wymagane:
  interfejs ETHERNET 755,
  łącze ETHERNET w WS 10,
  kabel 727.

● System zarządzania siecią DR-NET

System zarządzania siecią DR-NET jest programem, który uzupełnia system operacyjny CCP/M dla WS 10 o funkcje sieci. DR-NET umożliwia zbudowanie sieci lokalnej składającej się z od 2 do 255 WS 10. Z jego pomocą można komunikować się z takimi urządzeniami jak: napędy, drukarki i plotery jednej lub wielu stacji WS 10.

Napędy, drukarka, ploter i inne urządzenia, które przyłącza się bezpośrednio do WS 10, oznacza się jako "fizyczne urządzenia". Urządzenia, które są obsługiwane przez inne węzły sieci oznacza się jako "wirtualne urządzenia". Za pośrednictwem DR-NET urządzenia wirtualne mogą być wykorzystane przez inne węzły sieci jak gdyby były przyłączone wprost do WS 10.

DR-NET umożliwia też ochronę odczytu własnych danych.

Zapotrzebowanie pamięci centralnej	512×2^{10} bajtów.
Zapotrzebowanie pamięci napędów	jeden napęd dyskietek.
Wymagany system operacyjny	CCP/M-86 wersja 3.1

Opracowano na podstawie katalogu PR 33 - 1985
SIEMENS SICOMP WS 10.

NOWOŚCI TECHNICZNE

Archimedes 310

Anonsowany przed kilku miesiącami przez firmę Accorn system Archimedes o architekturze RISC pojawił się na rynku wprowadzony przez francuskie przedsiębiorstwo Informanic w cenie 1930 dolarów. System ma bogate oprogramowanie graficzne, szkoleniowe i różne rodzaje emulatorów. Działa pod systemem operacyjnym Arthur 1.2 będącym wersją Unixa. Zawiera procesor 32-bitowy pozwalający na pracę z prędkością 4 miliony operacji na sekundę. Standardowa pojemność pamięci operacyjnej wynosi 1 MB. W pamięci stałej zawarte są sprzęgi użytkownika, edytor stronicowy i silny Basic. System ma wbudowaną stację dyskietek 3,5 cala o pojemności 1 MB i możliwości dołączenia dysku sztywnego 20 MB ze sterownikiem, a także sterownika sprzęgu SCSI (Small Computer System Interface - sprzęg systemów małych komputerów), karty rozszerzenia pamięci itd. Ma też rozległe możliwości graficzne i dźwiękowe.

Micro Systemes nr 89 - wrzesień 1988

Nowe stanowisko pracy firmy Sony

Sony Enterprises oferuje we Francji nowe stanowisko pracy o nazwie View System zgodne z IBM PC, które działa na mikroprocesorze Intel 80286, jest wyposażone w pamięć operacyjną 640 KBajtów i pracuje pod systemem operacyjnym DOS 3.2 firmy Microsoft. Wyposażone jest też w czytnik dysków optycznych w standardzie Laservision (LDP 1500 lub 1550), z których każdy może przechowywać 54 tysiące obrazów, a czas dostępu wynosi odpowiednio 2,5 lub 2 sekundy. Zestaw podstawowy wyposażony jest w stację dyskietek o średnicy 3,5 cala i pojemności 1,44 MB. Opcjonalnie natomiast może dołączyć dysk sztywny o pojemności 40 MB i czytnik pamięci stałej na dyskach typu kompakt.

Stacja wyposażona jest w dwa sterowniki wizyjne, jeden zgodny z kartą CGA, a drugi sterujący rozdzielczością 672x496 elementów w 256 kolorach spośród 4096 odcieni. Sygnały generowane przez stację mogą być nakładane na sygnały wizyjne. Dyski natomiast mogą być wykorzystane do gromadzenia dźwięku w postaci cyfrowej. Cena, wraz z systemem operacyjnym, wynosi 14200 dolarów.

Równocześnie Sony zaprezentowała termiczną drukarkę kolorową UP-5000, która umożliwia wydzielenie składowych podstawowych z sygnału wizyjnego, zamianę ich do postaci cyfrowej i wydruk w formacie A5. Każda składowa zapewnia 256 poziomów, a cała paleta jest rzędu 10 milionów odcieni na piksel. Wyposażona jest ona w sprzęg RS 232C do sterowania z mikrokomputera co pozwala na modyfikację barw i kontrastu przed wydrukiem końcowym. Cena drukarki wynosi 10,5 tys. dolarów.

Micro Systemes nr 89 - wrzesień 1988

Drukarki komputerowe w połowie 1988 r.

Poszczególne rodzaje drukarek przeznaczone są do różnych zastosowań. Drukarki mozaikowe o 24 igłach wykorzystywane są w sekretariatach. Według oceny IDC w 1986 roku 70% sprzedanych drukarek stanowiły drukarki mozaikowe uderzeniowe, wśród których wyróżnić można 2 grupy: drukarki 24-igłowe przeznaczone do wydruku wysokiej jakości oraz drukarki szybkie, o prędkości

wydruku conajmniej 200 zn/s, których produkcja w 1987 roku wzrosła o 58% w stosunku do roku poprzedniego. Dla wszystkich drukarek mozaikowych wskaźnik ten wyniósł 37%, ale w ciągu trzech lat następnych przewidywano, że nie przekroczy 19%. Drukarki 24-igłowe stanowią alternatywę dla drukarek laserowych. Ich jakość wydruku jest taka, jak dla drukarek margaretkowych (ang. daisy wheel), przy znacznie większej szybkości i możliwościach graficznych. Wyposażone są one zwykle w pamięć buforową o dostatecznej pojemności dla zapewnienia szybkiej transmisji i tak jak drukarki laserowe oferują różne emulacje i zestawy czcionek. Są one jednak od nich wolniejsze, a jakość wydruku nieco gorsza. Natomiast cena jest konkurencyjna, jak również znacznie niższe są koszty eksploatacji. Ponadto dla pewnych zastosowań istotna jest możliwość osiągania kilku kopii i wydruku kolorowego, choć w tym ostatnim przypadku pojawiają się ostatnio próby stosowania barwnego tonera w drukarkach laserowych.

Konstruktorzy drukarek mozaikowych poprawiają wciąż jakość wydruku, co w przypadku firmy Epson doprowadziło do pierwszych drukarek 48-igłowych (4x12). Średnica każdej igły wynosi poniżej 0,2 mm, a osiągana rozdzielczość jest rzędu 360 punktów na cal. Drukarki takie zastępują drukarki margaretkowe, są stosowane w systemach projektowych i złożonych systemach graficznych. Drukarka ta miała pojawić się na rynku europejskim jesienią 1988 r i miała być wyposażona w zestawy czcionek dostosowane do rozpoznawania optycznego. Ponadto przewidziany był tu wydruk kolorowy, duży format do 40 cm, możliwość wydruku kilku kopii i podwójnego wykorzystania zarówno papieru składanego, jak i oddzielnych arkuszy. Cena miała być niższa od 3200 dolarów.

Jeśli chodzi o prędkość wydruku to od 1984 roku dominuje tu firma Output Technology Corp., która w marcu 1987 r. wprowadziła na rynek serię drukarek 800, osiągających 850 znaków/sekundę, drukujących normalnie 240 wierszy na minutę (co odpowiada 544 zn/s). Opracowywana jest seria 2100, która ma drukować z prędkością 600 wierszy na minutę, a więc 1360 znaków na sekundę. Osiąga się to przez stosowanie trzech głowic jednocześnie, co wymaga złożonego sterowania, ale rozwiązuje problemy nagrzewania się głowic i ich długowieczności. Ponadto taśma barwiąca nachylona jest do poziomej płaszczyzny głowic, co zapewnia lepsze ich wykorzystanie. Parametry wydruku ustala się łatwo przez dawanie pozytywnych lub negatywnych odpowiedzi na ustalone uprzednio menu. Drukarki OTC stosuje się do etykietek, wydruków obliczeń, fakturowania itp. Niekiedy pracują one po 12 godzin na dobę, co dobrze świadczy o ich niezawodności. Cena serii 800 waha się od 3500 do 6500 dolarów. We Francji OTC jest reprezentowana przez przedsiębiorstwo Omnilogic. Aktualnym (wrzesień 1988) rekordzistą w tym zakresie jest drukarka termiczna ITH-80 opracowana przez firmę Hengstler Contrôle Numérique o prędkości 1052 znaki na sekundę przy 80 kolumnach i papierze zwijanym o szerokości 114 mm. Posiada ona rozdzielczość 6 punktów na mm i jest stosowana w handlu, punktach informacyjnych, medycynie. Wyposażona jest w sprzęg RS 232C, jeden ze sprzęgów równoległych i pętlę prądową 29 mA. Kosztuje 1150 dolarów.

Minis et Micros nr 304 - czerwiec 1988 r

Kolorowa drukarka strumieniowa firmy Hewlett-Packard

Pierwsze drukarki strumieniowe były często zawodne ze względu na zapychanie się dysz i inne problemy techniczne. Opracowana przez firmę Hewlett-Packard technika strumieniowo-cieplna, w której część atramentu odparowywana jest za pomocą energii cieplnej umożliwiając tym samym przejście kropli przez dyszę, daje dobre rezultaty. Już w 1984 r. opracowana była drukarka Think Jet o rozdzielczości 96 elementów na cal, a więc porównywalna do 24-igłowych drukarek mozaikowych, lecz cicha i bardziej niezawodna, prawie nie wymagająca konserwacji. Zastosowana tu była mechaniczna regulacja ciśnienia.

Natomiast opublikowane w sierpniu 1988 r. parametry drukarki kolorowej PaintJet, mówią o postępie, jaki się tu dokonał. Na październik 1988 zapowiedziano publikację o trzeciej drukarce opartej na zasadzie DeskJet, która zapewnia rozdzielczość 300 elementów i ma inne parametry zbliżone do drogich drukarek laserowych. Cena jej jest zbliżona do drukarki 24-igłowej, ale jakość wydruku jest znacznie lepsza.

Drukarka PaintJet dostosowana jest do wykorzystania z komputerami osobistymi i skomputeryzowanymi stanowiskami pracy; daje kolorowe kopie obrazu z monitora. Rozwiązanie termiczno-strumieniowe okazało się korzystne pod względem technicznym i ekonomicznym. Uzyskuje się w ten sposób tekst i rysunki o rozdzielczości 180 elementów na cal przy prędkości wydruku 167 znaków na cal dla jakości NLQ. Przeciętną stronę tekstu uzyskuje się w ciągu około 35 sekund , a kolorową stronę grafiki w ciągu 4 minut. Można też łączyć tekst z grafiką. Drukarka wykorzystuje dwa podzespoły drukujące (print cartridge), w tym jeden do wydruku kolorowego. Głowica jednokolorowa zawiera 30 dysz, a wielokolorowa po 10 dysz na każdy z podstawowych kolorów (niebieski, czerwony i żółty). Głowice te zawierają systemy dostarczania płynnego atramentu i napędu kropli. Rozwiązanie to oparte jest na tworzeniu kropli na żądanie: każda z dysz wysyła w odpowiedniej chwili kroplę atramentu, gdy głowica przesuwa się w stosunku do ośrodka. Krople emitowane są przez elektryczne podgrzewanie oporników cienkowarstwowych. Gęstość kropli wybrano na 100 pl, zbyt duża objętość zmniejsza prędkość wydruku, a zbyt mała jest trudna do uzyskania. Wybrano też raczej dużą prędkość poruszania się kropel - 12 m/s, co powoduje powstanie długiego "ogona" za kroplą, nie

wpływa to jednak na jakość wydruku, a prędkość taka zapewnia poprawne działanie nawet przy zagęszczonym atramencie. Maksymalna częstotliwość wysyłania kropel wynosi 3kHz, a energia na kroplę 17 µJ. Typowa szerokość impulsu pobudzającego ma wartość 5 µs, a typowe napięcie wynosi 11,5V.

Wysoka rozdzielczość uzyskiwana jest między innymi przez to, że do wytwarzania głowic wykorzystuje się konwencjonalną fotolitografię i obróbkę cienkowarstwową. Ułatwia to również znacznie projektowanie głowicy kolorowej z wieloma komorami. Uzyskuje się dobrą zgodność kolorów, co daje obrazy wysokiej jakości. Drukowanie odbywać się może zarówno na papierze, jak i na przezroczystej folii bez zmiany głowicy.

PaintJet stanowi pomost pomiędzy drukarkami a ploterami. Daje ona możliwość wydruku ośmiu kolorami podstawowymi z ponad tysiąca możliwych. Przy projektowaniu poszczególnych podzespołów kierowano się kompromisem pomiędzy wymaganiami - jakości i szybkości wydruku, a możliwością produkcji i ceną. A więc przy projektowaniu generatora kropli należało ustalić odpowiednią pojemność kropli, ich prędkość oraz zapobiegać zniekształceniom wydruku. Źródłem tych ostatnich są niedokładności produkcyjne (np. niezgodność ustawienia grzejnika i otworu wylotowego) oraz odchylenia kropel na skutek zaburzeń w atramencie. Również pęcherzyki powstające w atramencie mogą utrudnić poprawne działanie drukarki. Ograniczeniem prędkości okazało się ponowne wypełnianie dysz. Wymagało to odpowiedniego zaprojektowania ich kształtu, z uwzględnieniem również minimalizacji oddziaływań dysz sąsiednich.

Pojemnik na atrament doprowadza ten płyn do głowicy przy optymalnym ciśnieniu, które musi być niższe od atmosferycznego aby atrament nie wyciekał z dysz. Inaczej niż w ThinkJet przyjęto zjawisko włoskowatości jako podstawę rozprowadzania atramentu, jak to jest w większości współczesnych piór. Wykorzystano do tego celu piankę poliuretanową.

Opracowanie drukarki rozpoczęto od bardzo wnikliwego zebrania życzeń i wymagań klientów, natomiast w trakcie opracowywania poszczególnych zespołów poddawano je, jak również i całą drukarkę, testom głównie ze względu na jakość wydruku i niezawodność. Jeśli chodzi o jakość wydruku symulowano programowo błędy rozmieszczania kropek. Analizowano też wszystkie występujące błędy i starano się określić ich źródła. Jako parametr niezawodnościowy wprowadzono liczbę stron wydrukowanych bezbłędnie (MPBF - mean pages between failures), na przykład piór przebadano 2500 i potem okazało się, że przez pięć miesięcy po wprowadzeniu wyrobu nie było żadnej reklamacji z tego powodu. Dla zespołu drukującego przyjęto minimum milion znaków wydrukowanych (około 600 stron) bez uszkodzenia.

Hewlett-Packard Journal nr 4/88

Nowa, wymazywalna pamięć optyczna

Firma Tandy z Fort Worth w Teksasie ogłosiła informację o opracowaniu wymazywalnego ośrodka pamięciowego, pozwalającego uzyskać pamięć optyczną z wielokrotnym zapisem. Technika ta nosi nazwę THOR (Tandy High-Intensity Optical Recording - zapis optyczny o wysokiej intensywności firmy Tandy) i może być realizowane w standardzie Compact Disk lub innych. Technologia ośrodka zakupiona została od firmy Optical Data z Beaverton. Wiązka laserowa o określonej długości fali podgrzewa dwuwarstwowy ośrodek powodując rozszerzenie warstwy dolnej i wypchnięcie warstwy górnej. Ta ostatnia ochładza się szybciej niż dolna i utrzymuje wybrzuszony kształt aż zostanie ponownie podgrzana wiązka o innej długości fali. Wiązka laserowa padająca na te wybrzuszenia odbijana jest inaczej aniżeli przy padaniu na płaską powierzchnię, co pozwala fotodetektorom rozróżnić bity informacji dokładnie w taki sam sposób, jak w przypadku wypalanych trwale otworków. Badania ustaliły, że nie występują widoczne zmiany ośrodka przy 100 cyklach zapisu-odczytu i że ośrodek jest bardzo odporny na oddziaływanie otoczenia. Firma nie ujawniła jeszcze wszystkich szczegółów jak dokonuje się zapis. Wiadomo, że w technice magnetooptycznej cała ścieżka musi być wymazana przy jednym obrocie dysku, a nowe dane zapisane w następnym obrocie - tak więc dla zmiany zawartości nawet jednego sektora muszą tam zaistnieć dwa obroty. W przypadku ośrodka THOR zmiana stanu zależy od długości fali padającego światła, przy czym światło wymazujące wybrzuszenia nie oddziałują na powierzchnię płaską. Teoretycznie więc zmiana informacji mogłaby się tu odbywać przy pojedynczym obrocie.

Tandy przygotowuje zarówno własne wyroby, oparte na tej technice, jak i szuka partnerów do wspólnych opracowań. Pierwsze ma być urządzenie do zapisywania i odtwarzania dźwiękowych dysków kompaktowych, którego cena nie powinna przekroczyć pięciuset dolarów, przy czym można będzie na nim odtwarzać dyski będące w sprzedaży. Cenę dysku przewiduje się w granicach 12-20 dolarów. Opracowywane jest również urządzenie pamięciowe, które miałoby format CD-ROM (stała pamięć na dyskach kompaktowych), lecz byłoby wymazywalne. Dotychczasowe dyski przechowują 600 Mbajtów na jednej stronie, lecz Tandy uważa, że osiągnie tu większe pojemności, nie podając jednak wartości. Urządzenie dźwiękowe spodziewane jest w sprzedaży w 1990 roku, zaś opracowanie układu pamięciowego zajmie około 3 lat.

Czas dostępu pamięci stałych na dyskach kompaktowych jest bliższy wartości tego parametru dla dysków elastycznych niż dla dysków typu Winchester ze względu na masę głowic laserowych. Problem ten nie występuje w urządzeniach dźwiękowych, gdzie ścieżki przebiegają w sposób ciągły do środka, lecz w pamięciach występuje szybkie przyspieszenie głowic do przypadkowo wybieranych ścieżek, a masa głowic określa możliwość ich przyspieszenia. Nadzieje na opracowanie głowic holograficznych o małej masie łączy się więc z możliwością zmniejszenia czasu dostępu.

Computer Design z 01.06.1988 r.

Nowoczesne drukarki laserowe

Mimo ogromnego upowszechnienia tego typu drukarek są one wciąż drogie. Pierwsze drukarki laserowe pojawiły się w końcu lat siedemdziesiątych. Były to duże systemy, które drukowały ze zmienną prędkością i zastosowanie ich było ograniczone do niewielu użytkowników, których stać było na taką inwestycję (rzędu stu kilkudziesięciu tysięcy dolarów) jak niektóre banki i instytucje rządowe. Producenci tych urządzeń również nie byli zbyt liczni. Musimy tu wymienić IBM, Simensa i Datagraphix.

Druga faza rozwoju drukarek laserowych wiązała się z masowym wprowadzeniem na rynek fotokopiarek. Pierwsze modele były jeszcze dość wolne ze względu na ograniczoną prędkość obrotu bębna. W roku 1983 firma Xerox jako pierwsza zaoferowała model drukujący 12 stron na minutę w cenie jeszcze dosyć wysokiej (32 tysiące dolarów). Era biurowych drukarek laserowych rozpoczęła się od modelu firmy Hewlett-Packard, który kosztował tylko 4600 dolarów, wykorzystywał zespół drukujący firmy Canon, drukował z szybkością 8 stron na minutę i rozdzielczością 300 elementów na cal.

Zasada pracy drukarki laserowej jest taka sama jak fotokopiarki i sprowadza się do uczulenia powierzchni bębna pokrytego specjalną powłoką. Można to zrealizować w różny sposób. W drukarce laserowej osiąga się to za pomocą wiązki laserowej. Wydruk realizowany jest w trzech fazach: 1) tworzenie się obrazu do wydruku, 2) tworzenie matrycy wydruku i 3) przekazanie obrazu do druku.

W pierwszej fazie generowane są znaki tworzące obraz. Są one następnie przekształcane i przenoszone w postaci punktów świetlnych na fotoczułą powłokę bębna. Tworzenie się obrazu jest wynikiem dwóch jednoczesnych ruchów: obrotu bębna ze stałą prędkością i przesuwania się wiązki laserowej po powierzchni bębna. Powłoka bębna zawiera selen i powinna być uprzednio spolaryzowana. Gdy wiązka laserowa pada na bęben, w punkcie padania zanika ładunek dodatni; w ten sposób tworzone są znaki.

W drugiej fazie obraz elektrostatyczny utworzony uprzednio przesuwa się przez stanowisko, gdzie nakładany jest toner (rodzaj atramentu w proszku), który przylega do miejsc o zmniejszonym ładunku dodatnim.

W fazie trzeciej mając bęben odpowiednio pokryty tonerem należy przemieścić ten obraz na papier. Dokonuje się tego przez zetknięcie bębna z papierem i wytworzenie podciśnienia, aby toner przemieścił się na papier. Utrwalenie obrazu dokonuje się przez podgrzanie papieru, a następnie ściskanie go między dwoma walcami. Powoduje to absorbcję tonera przez papier, obraz zostaje utrwalony, a papier skierowany na zewnątrz.

Istnieją inne, podobne technologie jak elektrooptyczna i magnetograficzna. Zasadnicza różnica pomiędzy nimi polega na sposobie uczulenia bębna. Występują tu rozwiązania elektrooptyczne oparte na polaryzacji bębna soczewką z ciekłych kryształów (Epson), działem elektronowym (Philips i Kentek) lub przez jonizację (Bull), gdzie uzyskuje się znaczne szybkości.

Należy zwrócić uwagę, że równolegle do techniki laserowej rozwija się technika strumieniowa; drukarka Deskjet firmy Hewlett-Packard ma rozdzielczość 300 elementów na cal przy szybkości 120-240 znaków na sekundę w zależności od jakości wydruku. Rezultaty te nie dorównują parametrom drukarek laserowych, ale cena (zaledwie 2240 dolarów) jest atrakcyjna.

Najczęściej producenci drukarek laserowych konstruują swe urządzenia wokół modułu drukującego opracowanego przez inne firmy. Najbardziej znanymi wytwórcami tych modułów są: Canon, Ricoh, Konica, Sharp, Toshiba, Xerox, Hitachi, Kyocera, Kentek i Mita. Drugim istotnym zespołem drukarki laserowej jest jej sterownik, który sprzęga jednostkę centralną z modułem drukującym. Współczesna drukarka laserowa traktowana jest jako urządzenie inteligentne i za pomocą swego sterownika może interpretować i przetwarzać język opisu strony. Istnieje kilka takich języków. Ich zasada polega na określeniu "obrazów" stron, które mogą być tworzone przez mechanizm wydruku. Oznacza to określenie milionów białych i czarnych punktów tworzących stronę.

Możliwości każdego języka określone są przez jego zdolność do wykorzystania wszystkich możliwości, jakie daje dana technika. Znamy następujące języki opisu stron: DDL firmy Imagen, Postscript firmy Adobe i Interpress firmy Xerox, z których najpopularniejszym okazał się Postscript utworzony w 1982 roku.

Parametry drukarek laserowych

Konstruktor	Model	Szybkość /str/min/	Rozdzielczość /el/cal/	Format	Pamięć	Cena /w dol/	Uwagi
ACE	ACE SX	8	300	A4	512KB/4MB	3430	Sześć zestawów czcionek
	ACE MX	8	300	A4	512KB	3430	Sterownik z językiem Postscript OTP
	ACE EX	11	300	A4	512KB	4310	Jedenaście zestawów czcionek
ACER	Acer LP 75	6	300	A4,A5,B5	1,5MB/2MB	3880	Dziewięć zestawów czcionek
AGFA	P 3400 PS	18	400	A4	6MB	14650	Postscript, dysk sztywny 20 MB
ALCATEL	L5	8	300	A4,B5	1,5MB	4650	Sterownik z jęz. Postscript,mod.NEC
	TL 410	8	300	A4	512KB	3690	Zestawy czcionek zdalnie ładowane. Moduł drukarki
AST	Turbolaser	8	300	A4	2MB	6190	Zestawy czcionek zdalnie ładowane
	Postscript	8	300	A4	4MB	6190	Sterownik z językiem Postscript.
APPLE	Laserwriter SC	8	300	A4	1MB	3430	
	Laserwriter II NT	8	300	A4	2MB	5000	68000, Postscript
	Laserwriter II NTX	8	300	A4	2MB/12MB	6190	68020, Postscript
BROTHER	HL8	8	300	A4	512KB	3970	Pięć zestawów czcionek
CANON	LBP-8 II	8	300	A4	512KB/2,5MB	3410	Jeden zestaw czcionek
	LBP-8IIR	8	300	A4	512KB/2,5MB	5310	Zestawy czcionek zdalnie ładowane
	LBP-8IIT	8	300	A4	512KB/2,5MB	4280	Zestawy czcionek zdalnie ładowane
CITIZEN	Ouverture 110	10	300	A4	512KB	3950	Zestawy czcionek zdalnie ładowane
	Ouverture 110+	10	240	A4	512KB	3190	Prosty pojemnik
COMDATA	LP 300X	8	300	A4	1,2MB	4300	Zestawy czcionek zdalnie ładowane
	LP 300	8	300	A4	-	3360	Zestawy czcionek zdalnie ładowane
CITOH	Lips 10	10	300	A4	512KB/2MB	5860	Dziewiętnaście zestawów czcionek
	CI-5 Compact	5	300	A4	512KB	8620	Sześć zestawow czcionek
CENTRONICS	PP8	8	300	A4	1,5MB	3430	Trzy zestawy czcionek
DATAPRODUCTS	LZR 1200	12	300	A4	512KB/1,5MB	5520	Postscript
	LZR 2610	26	300	A3	512KB	18620	Postscript
	LZR 2620	26	300	A3	512KB	20000	Postscript
	LZR 2630	26	300	A4	512KB	21550	Postscript
	LZR 2650/55	26	300	A4	512KB/3MB	4830	Postscript
	LZR 2660/65	22	300	A4	512KB/2,5MB	25790	Postscript
DEC	LN03	8	300	A4	256KB	6030	Szesnaście zestawów czcionek
	LN03+	8	300	A4	256KB/1MB	8610	Siedemnaście zestawów czcionek
	Scriptprinter	8	300	A4	2MB	10850	Sterownik z językiem Postscript
EPSON	G03	6	300	A4,A5,B5	512KB/1,2MB	3430	Sterownik z językiem Postscript
FACIT	P2080	8	300	A4	500KB	5160	Cztery zestawy czcionek
	P7150	15	300	A4	500KB	-	Cztery zestawy czcionek
	P6010	6	300	A4	1MB	3430	Dwa zestawy czcionek

Parametry drukarek laserowych /c.d./

Konstruktor	Model	Szybkość /str/min/	Rozdzielczość /el/cal/	Format	Pamięć	Cena /w dol/	Uwagi
FUJITSU	RX 7100	6	300	A4	640KB/4MB	3020	Cztery zestawy czcionek
	RX 7200	12	300	A4	640KB/4MB	5520	Cztery zestawy czcionek
	RX 7300	17	300	A4	2,5MB/8MB	12410	Osiem zestawów czcionek
GENICOM	5010	10	300	A4,B4	256KB/1,8MB	4920	Trzy zestawy czcionek
	PagePrinter 8	8	300	A4,B4	1,8MB	3440	Cztery zestawy czcionek
HP	Laser 500+	8	300	A4	512KB/2MB	6300	Język PCL/DDL
	Laserjet 2000	20	300	A3	1,5MB/6,5MB	27230	Trzydzieści cztery zestawy czcionek
	Laserjet Série 2	8	300	A4	512KB/4,5MB	4830	Dwa zestawy czcionek
HONEYWELL	Laserpage 201	8	300	A4,A5	512KB/2MB	5970	Sterownik z językiem Postscript
IBM	Laser 4216	6	300	A4,B4,B5	512KB/2MB	5830	Sterownik z językiem Postscript
	3812	12	240	A4,B5,A5	256KB	11600	Sterownik z językiem Postscript
KYOCERA	F-3000	18	300	A4	1,5MB/3,5MB	13100	Zestawy czcionek zdalnie ładowane
	F-1000	10	300	A4	512KB/1,5MB	4480	78 zestawów czcionek
	F-2200	10	300	A4	1,5MB/3,5MB	9660	78 zestawów czcionek
	F-1200	10	300	A4	1,5MB/3,5MB	6380	78 zestawów czcionek
MANNESMANN	MT-910	10	300	A4	512KB/1,5MB	5590	Trzy zestawy czcionek
OKI	Laserline	6	300	A4	1,5MB/2,5MB	3880	Piętnaście zestawów czcionek
OLIVETTI	PG 850	8	300	A4	2MB	7980	Dziewięć zestawów czcionek
	PG 208	8	300	A4	1MB	4170	Dziewięć zestawów czcionek
PANASONIC	KX-T 4450	11	300	A3	512KB	4450	Podwójny pojemnik
TRT-TI	LBP 8	8	300	A4	512KB	3410	Cztery zestawy czcionek
QMS	PS 800/2	8	300	A4	2MB/3MB	9240	Sterownik z językiem Postscript
	PS 810	8	300	A4	2MB/3MB	8190	Sterownik z językiem Postscript
	SmartWriter 20	8	300	A4	1MB	4220	Trzydzieści cztery zestawy czcionek
	SmartWriter 30	8	300	A4	2MB	5140	Trzydzieści cztery zestawy czcionek
QUADRAM	Quadlaster	8	300	A4	1MB	5780	Zestawy czcionek zdalnie ładowane
	Quadlaser Plus	8	300	A4	2MB	6310	Zestawy czcionek zdalnie ładowane
QUME	Laser Ten Plus	10	300	A4	512KB	4120	Dziewiętnaście zestawów czcionek Moduł druk.Hitachi
	Scripten	10	300	A4	3MB	8020	Trzydzieści pięć zestawów czcionek Moduł druk. Hitachi
XEROX	2700	12	300	A4	256KB	16240	Dwa zestawy czcionek
	4045/4046	10	300	A4	512KB/1,5MB	7740	Przy cenie 9480 dol. podwójny pojemnik
	3700	24	300	A4	1MB/10MB	41380	Zawiera dysk sztywny
RICOH	4081	8	300	A4	1MB	6210	Osiem zestawów czcionek
	PC 6000	6	300	A4	1MB	4050	Osiem zestawów czcionek
STAR	LaserPrinter 8	8	300	A4	1MB	4290	Cztery zestawy czcionek

Parametry drukarek laserowych /c.d./

Konstruktor	Model	Szybkość /str/min/	Rozdzielczość /el/cal/	Format	Pamięć	Cena /w dol/	Uwagi
TANDY	LP 1000	6	300	A4,A5,B5	1,5MB	3190	Osiem zestawów czcionek
TEXAS	Omnilaser 2115	15	300	A4	3MB	9660	Sterownik z językiem Postscript
	Omnilaser 2015	15	300	A4	512KB	6210	
	Omnilaser 2106	6	300	A4	2MB	6350	Sterownik z językiem Postscript
	Omnilaser 2108	8	300	A4	2MB	6380	Sterownik z językiem Postscript
TOSHIBA	Page Laser 12	12	300	A4	512KB/2MB	5320	Trzy zestawy czcionek
VISA	Visa 600	6	300	A4	1,5MB	3420	Dziewięć zestawów czcionek
WANG	LCS 15	15	300	A4	3MB	11380	Osiemnaście zestawów czcionek
	LDP 8	8	300	A4	512KB	4740	Dwa zestawy czcionek
WENGER	Wenger 8/1	8	300	A4	1,5MB/2MB	4830	Zestawy czcionek zdalnie ładowane

Przytoczone zestawienie parametrów kilkudziesięciu drukarek laserowych oferowanych na rynku francuskim (ceny przeliczono w dolarach USA) pokazuje, że najczęściej spotykaną prędkością jest 8 stron na minutę - prędkość taką miała prawie połowa oferowanych modeli. Średnia prędkość oferty wynosiła 10,4 str/minutę. Zakres występujących prędkości zawiera się między 5 a 26 str/min, tę ostatnią osiąga większość drukarek firmy Dataproducts. Rozdzielczość większości zestawionych drukarek wynosi 300 elementów na cal, tylko drukarka firmy Agfa ma ten parametr korzystniejszy (400 el/cal). Według Systems international z września 1988 roku brytyjska firma Oce oferuje swą drukarkę model 6750 o rozdzielczości 508 elementów na cal. Jest to uniwersalna drukarka laserowa z pamięcią 20MB, drukująca na różnych formatach i nośnikach. Cena jej wynosiła nieco ponad 50000 dolarów. Najczęściej stosowanym formatem jest A4, czasami występuje też A3, A5, B4, B5. Pojemność pamięci drukarek zawiera się od 256KB do 12MB, średnio 1,7 MB, najczęściej występuje pamięć 0,5MB. Cena natomiast nie schodzi poniżej 3 tysięcy dolarów (np. RX 7100 firmy Fujitsu) i sięga 42 tys. dolarów (dla Xerox 3700), średnio 7600 dolarów. Wśród innych informacji podawanych o drukarkach zwracają uwagę zestawy czcionek, które wahają się od jednego dla drukarki LBP-8 II firmy Canon do 78 dla drukarek firmy Kyocera. Inną właściwością drukarek, nie uwzględnioną w podanej tabeli, są emulacje, jakie mogą realizować poszczególne modele. Można tu wymienić tradycyjną drukarkę firmy Diablo, a z nowych Laserjet i Laserwriter.

Minis et Micros nr 304 z 1988.06.08

Braki w zaopatrzeniu w dynamiczne kostki pamięciowe i wzrost ich ceny

Podzespoły elektroniczne oznaczane jako DRAM (Dynamic Random Access Memory - pamięć dynamiczna o przypadkowym dostępie) są przyrównywane do tlenu podsycającego rozwój przemysłu informatycznego. Periodycznie obserwuje się okresy nadmiernych zapasów, a następnie braków w zaopatrzeniu w podzespoły. To jednak co zaczęło dziać się na rynku kostek DRAM w końcu 1987 roku znamionuje zjawisko znacznie bardziej złożone.

Przyczyny jego należy dopatrywać się w gwałtownym wzroście zapotrzebowania na te podzespoły, opóźnieniu w realizacji kostek o dużych pojemnościach (np. 1Mbit), a także podpisanym porozumieniu amerykańsko-japońskim dotyczącym podzespołów półprzewodnikowych. Wszystkie te czynniki spowodowały nie tylko znaczny wzrost cen, ale wprowadzenie sprzedaży wiązanej, tzn. żąda się przy zakupie kostek DRAM innych, mniej "chodliwych" wyrobów.

Poniższa tabela przytacza przewidywania producentów japońskich odnoszących się do produkcji i eksportu podzespołów pamięciowych w drugim i trzecim kwartale 1988 r.

	Pojemność	Kwartał 88	Produkcja (mln szt)	Eksport (mln szt)
Dynamiczne	64 Kbit	2	21,2	12,9
		3	18,9	11,9
	256 Kbit	2	159,3	90,5
		3	155,1	88,6
	1 Mbit	2	43,6	28,3
		3	56	35
Statyczne	16 Kbit	2	40,8	16,2
		3	42,2	16,5
	64 Kbit	2	46,8	21
		3	50,6	22,2
	256 Kbit	2	22,5	10
		3	26	11,8

Widać z niej, że pojemność produkowanych kostek DRAM 64 Kbitowych stanowi tylko 1,5% pojemności wszystkich kostek dynamicznych i produkcja ich maleje. Najwięcej produkuje się kostek DRAM o pojemności 256 Kbitów, przy czym wytwarzanie ich również nie wzrasta. Rośnie natomiast produkcja kostek megabitowych i to w tempie ok 25% kwartalnie. Eksport kostek dynamicznych wynosi ok. 60% i jest najwyższy dla kostek o pojemności 1 Mbit.

Statycznych pamięci produkuje się dwukrotnie mniej aniżeli dynamicznych i są one o jeden stopień mniej zintegrowane. Produkcja kostek o pojemności 64 Kbity niewiele przewyższa produkcję kostek 16-Kbitowych, lecz jest 2-krotnie większa od 256-Kbitowych. Rośnie ona dla wszystkich rodzajów, najszybciej dla ćwierćmegabitowych (ok. 15% kwartalnie). Eksport jest tu niższy (około 40%) i rośnie ze wzrostem pojemności kostek. Nie są to jednak ilości, które mogłyby zaspokoić potrzeby światowe.

Na przełomie lat siedemdziesiątych i osiemdziesiątych liczba amerykańskich dostawców kostek DRAM sięgała dziesiątki. Tani import z Japonii wyeliminował wówczas większość z nich. Na placu boju pozostały dwa przedsiębiorstwa: Texas Instruments z Dallas i Micron Technology z Boise. Firmy japońskie tak tanio sprzedawały swe wyroby, iż ocenia się, że same poniosły straty rzędu 4 mld dolarów. Doprowadziło to do podpisania wspomnianego porozumienia w 1986 roku i w nim wielu użytkowników podzespołów dopatrywało się przyczyny braków i wzrostu cen. W rzeczywistości zostało ono zawarte za późno - amerykańscy konkurenci byli pokonani.

W powstałej sytuacji Micron Technology podwoiła swoją produkcję, a Texas Instruments ściąga czołowych specjalistów ze swych filii japońskich do USA. Jednakże nie zaspokaja to potrzeb rynku. Ocenia się, że nawet przy inwestycjach rzędu 2 mld dolarów, inni wytwórcy amerykańscy będą mogli w 1991 r. dostarczyć tylko 40% żądanych podzespołów. Działania w tym kierunku podjęły takie znane firmy, jak Motorola czy National Semiconductor, lecz wiąże się to z dużym ryzykiem, gdyż wielcy producenci japońscy mogą znów zacząć dostarczać tanie podzespoły. Dlatego myśli się o rozwiązaniach w postaci nowej umowy, a jednocześnie,niektórzy wytwórcy systemów chcą budować własne wytwórnie DRAM.

Na razie (sytuacja z kwietnia 1988 roku) cena kostki pamięci dynamicznej o pojemności 256 Kbitów wzrosła z 3 dolarów w połowie 1987 roku do 12, a kostki o pojemności 1Mbita z 15 do 55 dolarów. Wytwórcy systemów, zagrożeni zupełnym odcięciem dostaw, nie skarżą się nawet zbyt głośno i godzą się na sprzedaż wiązaną, którą poprzednio, przy niskich cenach kostek pamięciowych, stosowano odwrotnie, zobowiązując klientów do zakupu określonej ich ilości.

W rozwiązaniu tego problemu powinna pomóc zmiana przepisów prawnych w USA wobec rodzimych producentów podzespołów i stworzenia takich mechanizmów rynkowych, które pozwoliłyby na szybką reakcję na próby wykorzystywania zaistniałej sytuacji do osiągania nadmiernych zysków.

Computer Design z 1988.06.01

Rynek informatyczny w 1987 roku

Firma CAP Gemini Sogeti dokonała, na podstawie raportów z różnych źródeł, oceny światowego rynku informatycznego w 1987 roku. Rynek ten składa się z czterech zasadniczych działów: 1) sprzęt, 2) specjalne usługi sprzętowe i oprogramowanie, 3) szeroko pojęte usługi informatyczne i 4) komunikacja danych. Poszczególne działy rozbijają się na bardziej szczegółowe pozycje, dla których określono wartość obrotów w 1987 r., procentowy ich udział w całym rynku informatycznym, tempo wzrostu w latach 1987-1990, główne przedsiębiorstwa specjalizujące się w poszczególnych pozycjach oraz przykłady ilustrujące wyroby i usługi, bądź określenia uściślające zakres poszczególnych pozycji.

Cały światowy rynek informatyczny w 1987 roku wyraził się sumą 256 mld dolarów, z czego najwięcej (41%) przypada na pierwszy z wymienionych działów, a więc sprzęt. Obejmuje on komputery, urządzenia peryferyjne i terminale. Przewidywany roczny przyrost w latach 1987-1990 wynosi 9%.

Pierwszą pozycją tego działu są duże maszyny obejmujące superkomputery i komputery, których cena przekracza 700 tysięcy dolarów. Wartość tej pozycji wynosi 37 mld dolarów, co stanowi 15% całego rynku i prawie 37% pierwszego działu. Tempo rocznego wzrostu w wymienionych latach jest tu najniższe w całej informatyce i wynosi zaledwie 6%. Pozycję tę reprezentują takie firmy jak IBM (około 70% udziału), Fujitsu, NEC, Unisys i Bull/HBI.

Drugą pozycją są systemy stosowane w dużych jednostkach administracyjnych (departamental systems) i minikomputery. Są to duże maszyny o cenie niższej niż 700 tysięcy dolarów, tzw. supermini i minikomputery. Stanowią one najmniejszą pozycję tego działu - 26 mld dolarów, to jest 10% całego rynku i ponad 24% działu. Tutaj zarówno IBM, jak i DEC mają ponad 10% udział. Natomiast na dalszych miejscach idą Hewlett-Packard, Bull/HBI, Data General, Fujitsu, NCR, NEC, Prime, Siemens, Tandem, Unisys i Wang. Wszystkie one mają ponad 2% udział.

Ostatnią pozycję pierwszego działu stanowią systemy lokalne, mikrokomputery i terminale. Obejmuje ona skomputeryzowane stanowiska pracy, niezależne lub połączone mikrokomputery, a także nieprogramowane terminale. Stanowiły one w 1987 roku największy 39% sprzętu (42 mld dolarów) i miały najwyższe tempo wzrostu w tym dziale - 12%. Kolejność udziału poszczególnych firm w tej pozycji jest następująca: IBM (jedyna mająca ponad 10% udziału), Apple, Compaq, DEC, Hewlett-Packard, NCR, NEC, Nixdorf, Olivetti, Tandy, Wang i Zenith.

Drugi dział składa się z dwóch pozycji. Pierwsza z nich to konserwacja sprzętu obejmująca wykrywanie błędów, lokalną i zdalną naprawę uszkodzeń oraz konserwację profilaktyczną. Wartość tej pozycji wynosiła w 1987 roku 25 mld dolarów, co stanowiło 10% całego rynku informatycznego i prawie 59% drugiego działu. Pozycja ta wykazuje wyższy od przeciętnego wskaźnik wzrostu (17%). Realizują ją te same firmy co w drugiej pozycji działu pierwszego (minikomputery).

Drugą pozycją drugiego działu jest oprogramowanie podstawowe, które w 1987 r. stanowiło co prawda tylko 7% rynku informatycznego (19 mld dolarów), lecz wykazywało największy wskaźnik wzrostu - aż 24%. Obejmuje ono szeroki zakres tematyczny: systemy operacyjne, bazy danych, zarządzanie, monitorowanie telekomunikacji, planowanie i zarządzanie operacjami,

języki, słowniki danych, niektóre programy użytkowe i programowe narzędzia wspomagające. Łącznie drugi dział stanowił w 1987 r. 17% rynku (44 mld dolarów) i wykazywał wskaźnik wzrostu 20%, najwyższy z wszystkich działów; obejmuje on podstawowo wyroby i usługi dla komputerów. Dominuje tu znów IBM (ponad 10% udziału), a ponadto Unisys, DEC, Computer Associates i Microsoft.

Trzeci dział usług informatycznych obejmujący wyroby i usługi dla użytkowników informatyki stanowił w 1987 r. 29% rynku (74 mld dolarów) i wykazywał drugi z kolei wskaźnik wzrostu (16%). Reprezentowany był przez trzy pozycje, z których największa (31mld dolarów - 13% rynku i prawie 45% działu) to usługi sprzętowe takie, jak zbieranie danych, przetwarzanie na zamówienie, stacje usługowe banków informacyjnych, zarządzanie urządzeniami i usługi sieciowe. Wykazują one umiarkowaną (11%) stopę wzrostu. Występują tu firmy rzadziej spotykane w klasycznej informatyce jak ADP, EDS, Reuters i Dun and Bradsteet.

Druga pozycja usług to oprogramowanie użytkowe stanowiące na razie najmniejszą pozycję rynku (14 mld dolarów - 5% rynku, ponad 17% działu), ale rozwija się dosyć szybko (18% wzrostu) i obejmuje zarówno pakiety programowe prostszych zastosowań, jak listy płac, księgowość, gospodarka magazynowa, przetwarzanie tekstów i arkusze obliczeniowe, a także bardziej złożonych, jak śledzenie notowań giełdowych, zarządzanie szpitalami, projektowanie wspomagane komputerowo. Głównymi dostawcami w tej dziedzinie są IBM i Lotus.

Ostatnią pozycją w tym dziale są najbardziej złożone usługi takie, jak zarządzanie zasobami komputerowymi, opracowanie i rozwój oprogramowania, programowa część integracji systemów, dostosowywanie pakietów do określonych warunków pracy, szkolenie i konwersja danych i programów. Usługi te stanowią 11% rynku i ponad 28% działu (29 mld dolarów) i rozwijają się w szybkim tempie - 20% wzrostu. Pierwsze miejsce, podobnie jak w większości pozycji, zajmuje tu IBM, a poza tym firmy mniej znane z branży informatycznej jak CGS, Arthur Andersen, Computer Sciences, TRW i EDS.

Ostatnim działem jest komunikacja danych, która stanowi tylko 13% rynku (33 mld dolarów) i dzieli się na dwie pozycje: sprzęt i usługi, których wartość pozostaje w stosunku do siebie jak 6:7. Tempo wzrostu sprzętu jest nieco większe (12%) niż usług (10%) i obejmuje on urządzenia transmisji i przełączania danych, sterowniki telekomunikacyjne, modemy, multipleksery, przełączniki pakietów i sprzęt związany z sieciami lokalnymi. IBM ma tu ponad 10% udziału, a dalej idą takie firmy, jak ATT, DEC, Motorola, Recal, Alcatel, Northern Telecom i Siemens. Natomiast usługi obejmują dzierżawę linii transmisyjnych i przesyłania danych po sieciach niższego typu. Pozycję tę reprezentują firmy telekomunikacyjne, czasami wręcz poczty jak ATT, NTT, 7BOC, Deutsche Bundespost, BT, France Télécom i GTE.

Systems International nr 9/88

BIBLIOTEKA GŁÓWNA
Politechniki Śląskiej
P 3057/88

ISSN 0239-8044