

Architektura komputerów

dr inż. Piotr Zielniewicz
e-mail: piotr.zielniewicz@cs.put.poznan.pl

Literatura

- P. Metzger, *Anatomia PC*, Helion, 2007
- J. Biernat, *Architektura komputerów*, Oficyna Wydawnicza Politechniki Wrocławskiej, 1999
- W. Stallings, *Organizacja i architektura systemu komputerowego*, WNT Warszawa, 2003
- R. Pełka, *Mikrokontrolery*, WKŁ Warszawa, 2000
- J. Sibigtroth, *Zrozumieć małe mikrokontrolery*, BTC Warszawa, 2003
- C. Petzold, *Kod*, WNT Warszawa, 2002

2

Zarys historii komputerów

- 1945: ENIAC (*Electronic Numerical Integrator And Computer*)
 - konstrukcja oparta na lampach próżniowych
 - arytmetyka w systemie dziesiętnym
 - pamięć dwudziestu 10-cyfrowych liczb
 - ręczne programowanie z użyciem przełączników i kabli krosujących
 - szybkość przetwarzania: 5000 operacji dodawania na sekundę
- 1946: EDVAC (*Electronic Discrete Variable Automatic Computer*)
 - zastosowanie systemu binarnego
 - zastosowanie pamięci do przechowywania danych oraz kodu programu

3

Zarys historii komputerów

- 1951: UNIVAC (*Universal Automatic Computer*)
 - pierwszy komputer dostępny w handlu (określany terminem „mózg elektronowy”)
 - szybkość przetwarzania: ponad 20 000 operacji na sekundę
 - zastosowania: obsługa powszechnego spisu ludności z 1950 roku, prognozowanie wyników wyborów prezydenckich w 1952 roku
- Koniec lat 50-tych: DEC PDP-1, IBM seria 7000
 - konstrukcje oparte na tranzystorach
 - bardziej złożone jednostki arytmetyczno-logiczne i sterujące
 - pamięci ferrytowe
 - rozbudowane urządzenia peryferyjne
 - 50-krotny wzrost prędkości w stosunku do komputerów pierwszej generacji

4

Zarys historii komputerów

- Połowa lat 60-tych: DEC PDP-8, IBM System 360
 - konstrukcja oparta na układach scalonych
 - pamięci półprzewodnikowe
 - koncepcja minikomputera
 - struktura magistralowa
- 1971: Mikroprocesor Intel 4004
 - wszystkie elementy procesora zintegrowane w jednym układzie (2300 tranzystorów)
 - 4-bitowa szerokość szyny danych
- 1974: Mikroprocesory Intel 8080, Motorola 6800
 - mikroprocesory ogólnego przeznaczenia
 - 8-bitowa szerokość szyny danych
 - 16-bitowa szerokość szyny adresowej

5

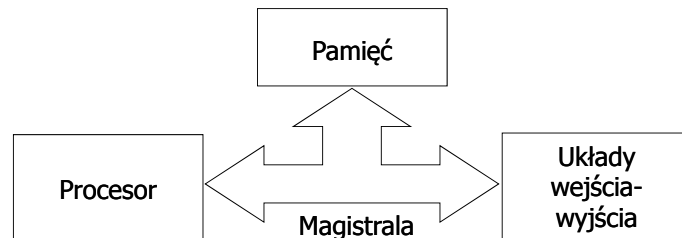
Zarys historii komputerów

- 1977: Mikrokomputer Apple II
- 1978, 1979: Mikroprocesory Intel 8086, 8088
 - 16-bitowa szerokość szyny danych
 - 20-bitowa szerokość szyny adresowej
 - wbudowane rozkazy mnożenia i dzielenia
- 1981: Mikrokomputer IBM PC
- 1984: Mikrokomputer Apple Macintosh
- 1985: Mikroprocesor Intel 80386
 - 32-bitowa szerokość szyny danych
 - 32-bitowa szerokość szyny adresowej
 - wsparcie dla systemów wielozadaniowych

6

Klasyczna koncepcja komputera

- Elementy funkcjonalne komputera:
 - Procesor, jednostka centralna, CPU (*Central Processing Unit*)
 - Pamięć operacyjna
 - Układy wejścia-wyjścia
- Schemat architektury:



7

Klasyczna koncepcja komputera

- Elementy komputera komunikują się ze sobą poprzez magistrale: danych, adresową i sterującą
- Magistrale pracują w sposób synchroniczny, co oznacza, że sygnały na nich występujące, mogą się pojawiać wyłącznie w ściśle określonych momentach wyznaczonych przez tzw. cykl zegarowy
- Przykład pobierania zawartości komórki pamięci operacyjnej przez procesor:
 - procesor wysyła adres żądanej komórki na magistralę adresową, a po odczekaniu odpowiedniej liczby cykli zegarowych odczytuje z magistrali dane
 - pamięć dekoduje przesyłany adres po czym umieszcza zawartość zaadresowanej komórki na magistrali danych w odpowiednim momencie i na odpowiednio długi czas

8

Klasyczna koncepcja komputera

- jeżeli pamięć operacyjna jest zbyt wolna w stosunku do szybkości pracy procesora i nie może wysłać danej w odpowiednio krótkim czasie, to informuje o tym procesor, wysyłając odpowiedni sygnał (WAIT) przez magistralę sterującą
- procesor wydłuża wówczas okres oczekiwania na daną o dodatkowy cykl zegarowy i w tak określonym momencie odczytuje ją z magistrali danych
- Opisany cykl pracy procesora nazywa się cyklem maszynowym i jest on związany ze zmianą adresu przesyłanego przez magistralę adresową
- Cykle maszynowe nie są jednakowej długości, choć zawsze są wielokrotnością cyklu zegarowego

9

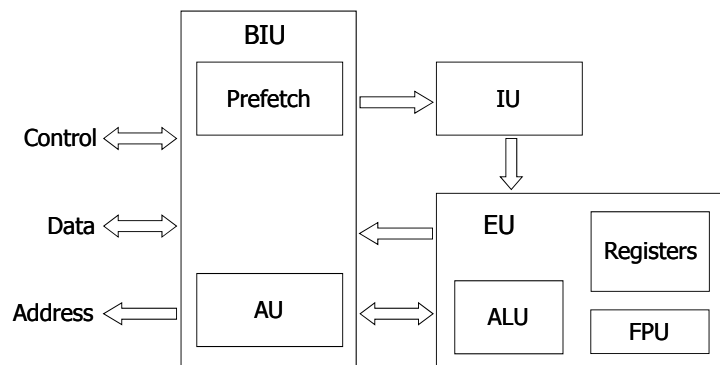
Klasyczna koncepcja komputera

- Komunikacja między procesorem a układami wejścia-wyjścia odbywa się w postaci analogicznych cykli maszynowych
- Komunikacja między urządzeniami wejścia-wyjścia a pamięcią operacyjną także przebiega w postaci podobnych cykli maszynowych w przypadku transmisji przez kanał bezpośredniego dostępu do pamięci
- Cykl pracy magistrali komputera składający się z pewnej liczby cykli maszynowych, w którym następuje wykonanie pełnego rozkazu procesora nazywa się cyklem rozkazowym

10

Ogólna zasada działania procesora

- Schemat blokowy procesora:



11

Ogólna zasada działania procesora

- Komunikacja procesora ze światem zewnętrznym odbywa się za pośrednictwem układu sterowania magistralami BIU (*Bus Interface Unit*), w którego skład wchodzi jednostka adresowania AU (*Addressing Unit*) i podręczna kolejka rozkazów (*Prefetch*)
- Jednostka adresowania ma za zadanie wytworzenie na magistrali adresowej adresu, pod którym znajdują się wymienne z procesorem dane lub pobierane rozkazy
- Pobierane rozkazy umieszczane są z wyprzedzeniem w podręcznej kolejce rozkazów, której uzupełnianie odbywa się automatycznie

12

Ogólna zasada działania procesora

- Dekodowanie rozkazów pobieranych z kolejki rozkazów odbywa się w bloku dekodera instrukcji IU (*Instruction Unit*)
- Rozkodowane instrukcje przekazywane są do układu wykonawczego EU (*Execution Unit*), który w zależności od rodzaju rozkazu wydaje odpowiednie polecenia do jednostki adresującej oraz jednostki arytmetyczno-logicznej ALU (*Arithmetic-Logic Unit*) lub jednostki zmiennoprzecinkowej FPU (*Floating Point Unit*)
- Wykonywane rozkazy mogą dotyczyć manipulacji na zawartości wewnętrznych rejestrów procesora (*Registers*), jak też komórek pamięci operacyjnej

13

Ogólna zasada działania procesora

- Zadaniem jednostki arytmetyczno-logicznej jest wykonywanie bezpośrednich operacji na danych tj. dodawanie, odejmowanie, porównywanie itd.
- W przypadku rozkazów operujących na danych zmiennoprzecinkowych ich przetwarzaniem zajmuje się wyspecjalizowana jednostka zmiennoprzecinkowa
- Nie istnieje ścisły podział pamięci operacyjnej na obszary danych i rozkazów – w każdej komórce pamięci operacyjnej mogą znajdować się zarówno podlegające przetwarzaniu dane, jak i przeznaczone do wykonania rozkazy

14

Ogólna zasada działania procesora

- Ponieważ większość działań procesora sprowadza się do operacji na stosunkowo wolnej pamięci RAM (*Random Access Memory*), to w celu zwiększenia efektywności procesor wyposażony jest w rejestry – szybką pamięć wewnętrzną
- Wykonanie jakiegokolwiek operacji na pamięci sprowadza się do przesłania zawartości komórki pamięci do rejestru, wykonania operacji i odesłania wyniku do pamięci
- Rejestry można podzielić na trzy funkcjonalne grupy:
 - rejestry ogólnego przeznaczenia (np. AX, BX, CX, DX)
 - rejestry segmentowe (np. CS, DS, SS)
 - rejestry statusowe i sterujące (np. FLAGS, IP)

15

Ogólna zasada działania procesora

- Zadaniem rejestrów ogólnego przeznaczenia jest przechowywanie danych
- W praktyce każdy z tych rejestrów jest przeznaczony do wykonywania określonych zadań np. rejestr AX (akumulator) w operacjach arytmetycznych i logicznych, BX do adresowania pamięci, CX jest licznikiem w pętlach, natomiast DX umożliwia przekazywanie i odbieranie danych z portów wejścia-wyjścia
- Rejestry segmentowe służą do przechowywania:
 - CS (*Code Segment*) - adresu segmentu kodu do wykonania
 - DS (*Data Segment*) - adresu segmentu danych
 - SS (*Stack Segment*) - adresu segmentu stosu

16

Ogólna zasada działania procesora

- Zadaniem rejestru stanu FLAGS jest kontrolowanie pracy procesora
- W rejestrze tym znajdują się m.in. następujące znaczniki stanu (flagi):
 - Znak (*Sign Flag*) - bit znaku wyniku ostatniej operacji arytmetycznej
 - Zero (*Zero Flag*) - flaga ustawiana jeżeli wynik operacji jest równy zero
 - Przeniesienie (*Carry Flag*) - flaga ustawiana jeżeli wynikiem operacji arytmetycznej jest przeniesienie
 - Przepełnienie (*Overflow*) - flaga ustawiana jeżeli w czasie wykonywania operacji arytmetycznej wystąpiło przepełnienie

17

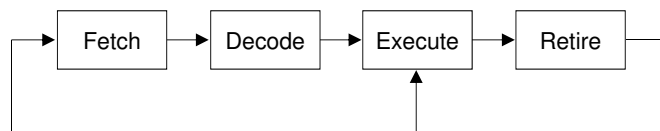
Ogólna zasada działania procesora

- Rejestr wskaźnika rozkazów IP (*Instruction Pointer*) zawiera przemieszczenie aktualnie wykonywanej instrukcji względem początku segmentu kodu
- O tym że zawartość komórki pamięci operacyjnej zostanie przez procesor zinterpretowana jako rozkaz do wykonania decyduje fakt jej zaadresowania przez wskaźnik (licznik) rozkazów
- W cyklu rozkazowym procesora możemy wyróżnić następujące cztery fazy:
 - Faza pobrania rozkazu (*Fetch*), w której następuje pobranie rozkazu z podręcznej kolejki rozkazów i jeżeli zachodzi taka potrzeba, uruchomienie procedury uzupełniania kolejki

18

Ogólna zasada działania procesora

- Faza dekodowania rozkazu (*Decode*), w której określone są operacje do wykonania oraz obliczane adresy efektywne argumentów (jeżeli występują)
- Faza wykonania (*Execute*), w której pobierane są z pamięci argumenty rozkazu oraz wykonywane operacje ustalone w fazie dekodowania
- Faza zapisania wyników (*Retire, Write-Back*), w której następuje zapisanie wyniku w miejscu określonym w kodzie rozkazu oraz ustawienie znaczników rejestru statusowego



19

Obsługa urządzeń zewnętrznych

- Jednym z zadań procesora jest wymiana informacji z urządzeniami zewnętrznymi (stacja dyskieta, karta graficzna, klawiatura itd.)
- Urządzenie zewnętrzne komunikuje się z układem wejścia-wyjścia za pośrednictwem interfejsu, tworząc wspólnie tzw. urządzenie peryferyjne
- Podstawowymi funkcjami realizowanymi przez układ wejścia-wyjścia jest:
 - komunikacja z urządzeniem zewnętrznym (głównie przez interfejs równoległy lub szeregowy), polegająca na przesyłaniu danych, wysyłaniu do niego rozkazów sterujących i odczytywaniu informacji o jego stanie

20

Obsługa urządzeń zewnętrznych

- komunikacja z procesorem, obejmująca dekodowanie adresu układu, wymianę danych, przesyłanie rozkazów i informacji o stanie układu
- buforowanie i formatowanie danych
- wykrywanie i informowanie o błędach
- **Przestrzeń adresowa układów wejścia-wyjścia może być:**
 - odwzorowana w pamięci (układy wejścia-wyjścia i komórki pamięci współdzielą tę samą przestrzeń adresową)
 - izolowana (układy wejścia-wyjścia i komórki pamięci korzystają z niezależnych przestrzeni adresowych)

21

Obsługa urządzeń zewnętrznych

- **Do typowych układów wejścia-wyjścia zaliczamy:**
 - układy szeregowej i równoległej transmisji danych
 - timery (układy pomiaru czasu, zliczania lub generowania zdarzeń)
 - zegary czasu rzeczywistego
 - układy przetworników analogowo-cyfrowych i cyfrowo-analogowych
- **Sterowanie pracą urządzenia zewnętrznego polega na zaprogramowaniu sterującego nim układu wejścia-wyjścia, czyli wpisaniu do odpowiadającego mu portu określonej wartości, która wymusza takie a nie inne zachowanie urządzenia**

22

Obsługa urządzeń zewnętrznych

- **Przesyłanie danych między urządzeniem peryferyjnym a pamięcią operacyjną może odbywać się w trybie programowanego wejścia-wyjścia PIO (*Programmed Input/Output*) lub w trybie bezpośredniego dostępu do pamięci DMA (*Direct Memory Access*)**
- **W trybie programowanego wejścia-wyjścia całym procesem przesyłania danych zajmuje się procesor**
- **W trybie bezpośredniego dostępu do pamięci procesor przechodzi w stan wysokiej impedancji, a na czas transmisji danych kontrolę nad magistralami komputera przejmuje układ DMA**

23

Obsługa urządzeń zewnętrznych

- **Urządzenie może poinformować procesor o zmianie swojego stanu zgłaszając tzw. przerwanie (*interrupt*)**
- **Typową sytuację powodującą zgłoszenie przerwania jest zajście zdarzenia zewnętrznego lub zakończenie zadania „zleconego” urządzeniu przez procesor**
- **Zgłoszenie przerwania przez urządzenie powoduje zawieszenie wykonywania bieżącego programu, wykonanie procedury obsługi przerwania a następnie powrót do wykonywania bieżącego programu**
- **Alternatywną metodą sprawdzania stanu urządzenia jest jego cykliczne przepytanie przez procesor w oczekiwaniu na spełnienie warunków koniecznych do realizacji transmisji (*polling*)**

24

Podstawowe własności i klasyfikacje pamięci

- W typowym systemie komputerowym można zauważyć hierarchię podsystemów pamięci – niektóre z nich umiejscowione są bezpośrednio w procesorze, inne tworzą główną pamięć operacyjną, jeszcze inne są podłączane do urządzeń wejścia-wyjścia stanowiąc pamięć pomocniczą
- Podstawową własnością pamięci jest jej pojemność wyrażana najczęściej w bajtach lub w słowach
- Innym parametrem związanym z pojemnością jest tzw. jednostka transferu (*transfer unit*) będąca liczbą bitów jednocześnie odczytywanych z pamięci lub do niej zapisywanych

25

Podstawowe własności i klasyfikacje pamięci

- W zależności od sposobu dostępu pamięci możemy podzielić na:
 - pamięci o dostępie sekwencyjnym, w których dostęp do danych (najczęściej w postaci rekordów) jest możliwy w określonej sekwencji liniowej (przykład: pamięć taśmowa)
 - pamięci o dostępie bezpośrednim, w których możliwy jest bezpośredni dostęp do bloku danych (dzięki mechanizmowi adresacji bloku), natomiast dostęp do konkretnej danej w bloku odbywa się sekwencyjnie (przykład: pamięć dyskowa)
 - pamięci o dostępie swobodnym, w których poszczególne dane mają swój unikatowy adres i dostęp do nich nie zależy od sekwencji poprzednio wykonanych operacji dostępu (przykład: pamięć operacyjna)

26

Podstawowe własności i klasyfikacje pamięci

- pamięci o dostępie skojarzeniowym, będący rodzajem dostępu swobodnego, w których adresowanie danych odbywa się na zasadzie badania zgodności (porównywania) wybranych bitów słowa z zawartością pamięci (przykład: pamięć podręczna)
- Kolejną istotną własnością pamięci jest jej wydajność, określana m.in. przez następujące parametry:
 - czas dostępu – czas od chwili doprowadzenia adresu do momentu udostępnienia danych
 - czas cyklu pamięci – czas jaki musi upłynąć między dwiema operacjami dostępu do pamięci
 - szybkość przesyłania – szybkość z jaką dane mogą być wprowadzane do pamięci lub z niej wyprowadzane

27

Podstawowe własności i klasyfikacje pamięci

- W zależności od technologii wykonania pamięci możemy podzielić na półprzewodnikowe, optyczne i magnetyczne
- Pamięci półprzewodnikowe możemy podzielić na następujące typy:
 - pamięci o dostępie swobodnym RAM (*Random Access Memory*)
 - nieulotne pamięci o dostępie swobodnym NVRAM (*Non-Volatile RAM*)
 - pamięci stałe ROM (*Read Only Memory*)
 - programowalne pamięci stałe PROM (*Programmable ROM*)
 - wymazywalne pamięci stałe EPROM (*Erasable and Programmable ROM*)

28

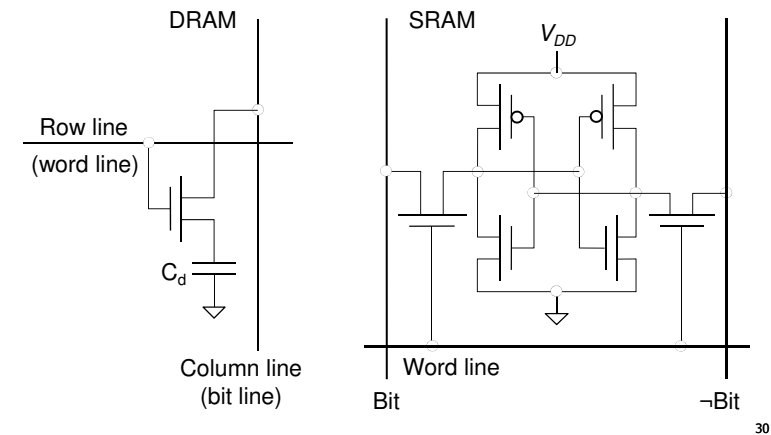
Podstawowe własności i klasyfikacje pamięci

- elektrycznie wymazywalne pamięci stałe EEPROM (*Electrically Erasable and Programmable ROM*)
- pamięci błyskawiczne (błyskowe) FLASH
- Pamięci o dostępie swobodnym RAM można podzielić na:
 - pamięci statyczne SRAM (*Static RAM*), w których komórki pamięci zbudowane są z dwustanowych przerzutników (konstruowanych z bramek logicznych)
 - pamięci dynamiczne DRAM (*Dynamic RAM*), w których komórki pamięci zbudowane są z układów przechowujących ładunek elektryczny (konstruowanych z kondensatora i tranzystora) i wymagających okresowego odświeżania ładunku w celu zachowania danych

29

Podstawowe własności i klasyfikacje pamięci

- Schematyczny diagram komórek pamięci RAM:



30

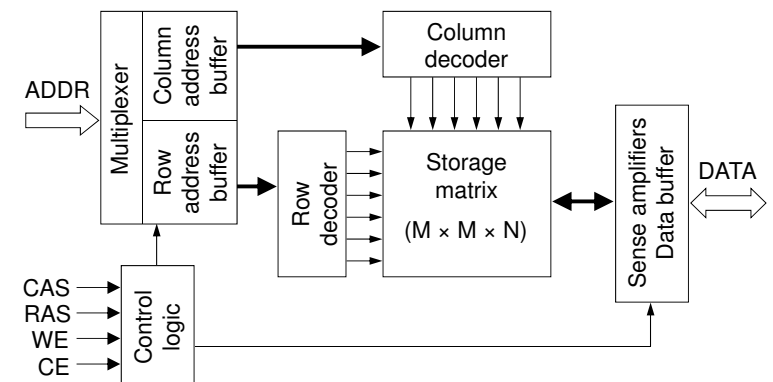
Pamięć operacyjna

- Pamięć operacyjna typowego komputera zbudowana jest z układów pamięci DRAM
- W układzie pamięci DRAM można wyróżnić m.in. następujące linie sygnałowe:
 - adresy słowa $A_0 \div A_{m-1}$ (m linii adresowych pozwalających na zaadresowanie maksymalnie 2^{2^m} słów)
 - dane słowa $D_0 \div D_{n-1}$ (n linii danych)
 - uaktywnienie układu CE (*Chip Enable*)
 - zezwolenie na zapis WE (*Write Enable*)
 - zezwolenie na odczyt OE (*Output Enable*)
 - wybór adresu wiersza RAS (*Row Address Strobe*)
 - wybór adresu kolumny CAS (*Column Address Strobe*)

31

Pamięć operacyjna

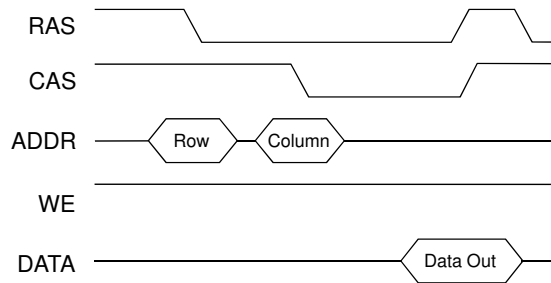
- Organizacja wewnętrzna układu pamięci DRAM:



32

Pamięć operacyjna

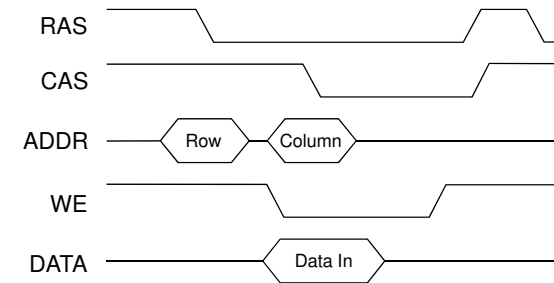
- Cykl odczytu pamięci DRAM działającej w trybie PM (*Page Mode*):



33

Pamięć operacyjna

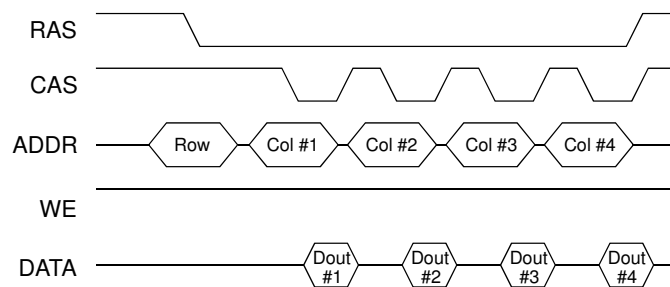
- Cykl zapisu pamięci DRAM działającej w trybie PM (*Page Mode*):



34

Pamięć operacyjna

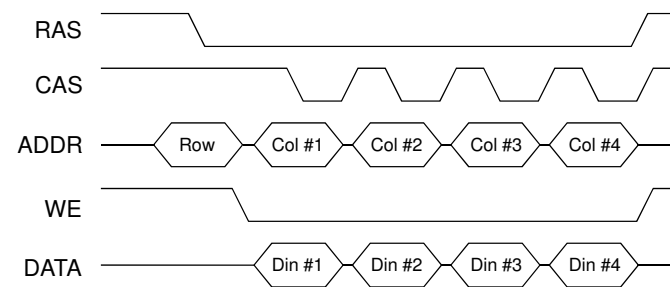
- Cykl odczytu pamięci DRAM działającej w trybie FPM (*Fast Page Mode*):



35

Pamięć operacyjna

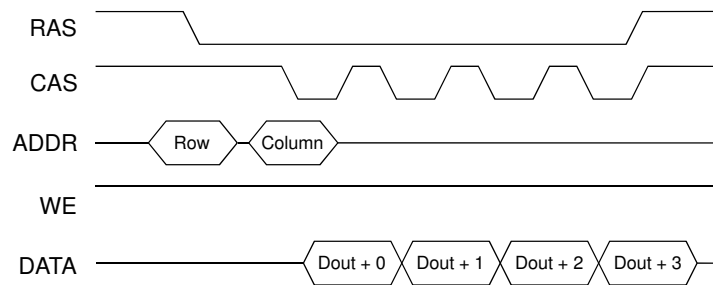
- Cykl zapisu pamięci DRAM działającej w trybie FPM (*Fast Page Mode*):



36

Pamięć operacyjna

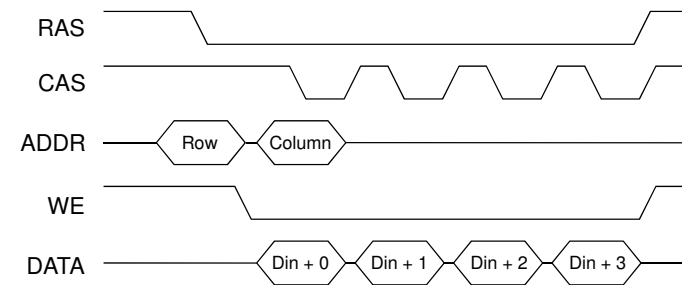
- Cykl odczytu pamięci DRAM działającej w trybie pakietowym (*Burst Mode*):



37

Pamięć operacyjna

- Cykl zapisu pamięci DRAM działającej w trybie pakietowym (*Burst Mode*):



38

Pamięć operacyjna

- Cechą charakterystyczną układów pamięci DRAM, wynikającą z zasady jej działania jest konieczność okresowego odświeżania komórek pamięci tj. uzupełniania ładunków elektrycznych w tych komórkach pamięci, które przechowują wartość logiczną „1”
- W celu prawidłowego funkcjonowania mechanizmu odświeżania konieczne jest cykliczne pobudzenie (adresowanie) poszczególnych wierszy matrycy pamięci z częstotliwością nie mniejszą, niż to wynika z tzw. czasu cyklu odświeżania (*Refresh Time*)

39

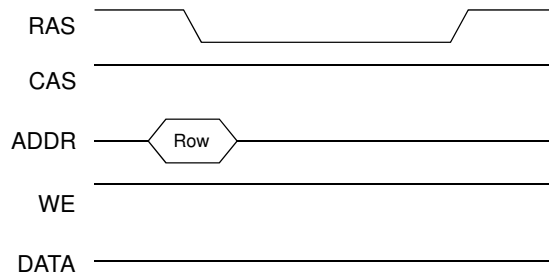
Pamięć operacyjna

- Operacja odświeżania pamięci DRAM realizowana jest przez układy logiczne będące albo elementem systemu komputerowego (kontroler pamięci) albo stanowiące integralną część samego układu pamięci
- Istnieją cztery podstawowe sposoby odświeżania pamięci DRAM:
 - odświeżanie sygnałem RAS (*RAS Only*)
 - odświeżanie CAS przed RAS (*CBR, CAS-before-RAS*)
 - odświeżanie ukryte (*Hidden*)
 - autoodświeżanie (*Self-refresh*)

40

Pamięć operacyjna

- Cykl odświeżania sygnałem RAS (*RAS-Only*) pamięci DRAM:



41

Pamięć operacyjna

- Podstawowym, stosowanym obecnie typem pamięci dynamicznej jest pamięć SDRAM (*Synchronous DRAM*), która różni się od typowej pamięci DRAM sposobem sterowania matrycą pamięci oraz techniką dostępu
- Cechą charakterystyczną układu pamięci DRAM jest synchronizowanie wszystkich sygnałów sterujących specjalnym przebiegiem zegarowym (CK), co ułatwia współpracę z magistralami systemu komputerowego i pozwala na zwiększenie szybkości jej działania
- Pamięć SDRAM jest optymalizowana do realizacji przesyłania pakietowego, ale dostęp do niej jest także możliwy w trybie pojedynczych cykli

42

Pamięć operacyjna

- Układ pamięci SDRAM zbudowany jest z niezależnych od siebie banków (najczęściej dwóch lub czterech), które adresowane są naprzemiennie, co przyspiesza proces dostępu do danych
- W układ pamięci SDRAM wbudowany jest licznik odświeżania i układy logiczne, które pozwalają na jego autonomiczne odświeżanie
- Przed użyciem pamięć SDRAM wymaga zaprogramowania przez wpisanie do specjalnego rejestru (*Mode Register*) informacji konfiguracyjnych m.in. parametrów czasowych, liczby przesyłanych słów w pojedynczym cyklu pakietowym

43

Pamięć operacyjna

- W celu uzyskania określonej liczby jednorazowo przesyłanych bitów danych układy pamięci łączone są w tzw. moduły (np. SIMM, DIMM)
- W przypadku modułów DIMM informacje o jego strukturze wewnętrznej niezbędne do prawidłowego zaprogramowania układów SDRAM pamiętane są w niewielkiej (256 B) pamięci stałej SPD (*Serial Presence Detect*)
- W przypadku konstrukcji modułów złożonych z dużej liczby układów pamięci może okazać się konieczne buforowanie sygnałów sterujących modułem

44

Pamięć operacyjna

- Odmianą pamięci SDRAM są pamięci DDR (*Double Data Rate*), które w każdym taktie zegara przekazują dwa słowa danych
- Do układu pamięci DDR przekazywane są dwa komplementarne przebiegi zegarowe (CLK/ \neg CLK), których przecięcie wyznacza moment kluczowania
- Pamięć DDR wykorzystuje dodatkowo dwukierunkowy sygnał strobujący DQS, który w cyklu odczytu generowany jest przez układ pamięci, natomiast w cyklu zapisu przez kontroler pamięci
- Dane są odczytywane z pamięci w środku impulsu DQS, a zapisywane do niej na zboczach tego impulsu

45

Pamięć operacyjna

- Innym typem stosowanych obecnie układów pamięci są pamięci RDRAM (*Rambus DRAM*), które do komunikacji z otoczeniem wykorzystują dedykowaną magistralę Rambus przesyłającą dane w trybie pakietowym
- Pojedynczy kanał magistrali Rambus pozwala na dołączenie maksymalnie 32 układów pamięci
- Magistrala Rambus jest bardzo szybka (taktowanie 400 MHz) magistralą o szerokości 16 bitów danych (podzielonych na dwie 8-bitowe szyny: DQA i DQB), co pozwala na transmisję danych z szybkością 1.6 GB/s

46

Pamięć operacyjna

- Informacja w pamięci RDRAM przechowywana jest w typowych blokach pamięci SDRAM zorganizowanych w niezależne banki (zwykle od 4 do 32) o szerokości najczęściej 128 bitów
- Układy logiczne pamięci RDRAM przesyłają takie porcje danych z/do magistrali Rambus wykorzystując bardzo szybkie multipleksery i demultipleksery
- Do synchronizacji przesyłanych danych magistrala Rambus wykorzystuje dwie pary komplementarnych sygnałów zegarowych CTM (*Clock to Master*) i CFM (*Clock from Master*)

47

Pamięć operacyjna

- Dane są odczytywane z układu pamięci w trakcie narastającego i opadającego zbocza sygnału CTM, a zapisywane do niego w trakcie narastającego i opadającego zbocza sygnału CFM
- Układ pamięci RDRAM wyposażony jest w 8-bitową magistralę adresową, w której 3 bity przypadają na adres wiersza, a 5 bitów na adres kolumny
- Przed użyciem układ pamięci RDRAM wymaga zaprogramowania tj. wpisania do jego wewnętrznych rejestrów sterujących odpowiednich informacji konfiguracyjnych (przesłanych z wykorzystaniem linii sterujących magistrali Rambus)

48

Pamięć operacyjna

- Z uwagi na możliwość wystąpienia przekłamań układy pamięci i ich kontrolery wyposaża się dodatkowo w mechanizmy, których celem jest wykrycie albo wykrycie i korekcja błędów
- Najprostszym stosowanym mechanizmem wykrywającym błędy jest tzw. kontrola parzystości
- Realizacja kontroli parzystości wymaga zapamiętania dodatkowego bitu, którego wartość dobierana jest tak, aby liczba „jedynek” w słowie danych była parzysta (operacja XOR)
- Kontrola parzystości pozwala na wykrycie dowolnej nieparzystej liczby błędnych bitów danych

49

Pamięć operacyjna

- Bardziej zaawansowanym mechanizmem kontrolnym są tzw. kody korekcyjne ECC (*Error Correction Code*) bazujące głównie na tzw. kodowaniu Hamminga
- Najprostsza wersja kodu Hamminga polega na generowaniu bitów parzystości dla odpowiednio dobranych grup bitów danych
- Przykładowo korekcja jednego bitu w słowie 64-bitowym wymaga dodatkowych 7 bitów kontrolnych
- Stosowane obecnie 8-bitowy kod ECC w modułach pamięci DIMM pozwala na wykrycie przekłamań dowolnych dwóch bitów oraz na korekcję dowolnego pojedynczego bitu w 64-bitowym słowie danych

50

Pamięć dyskowa

- Nośnikiem w pamięciach dyskowych jest krążek (dysk) z tworzywa pokrytego warstwą magnetyczną
- W niewielkiej odległości nad/pod obracającym się dyskiem umieszczona jest na ruchomym ramieniu głowica odczytująco-zapisująca
- W trakcie zapisu przepływający przez głowicę prąd elektryczny wytwarza pole magnetyczne trwale namagnesowujące przemieszczający się nad/pod głowicą nośnik
- W trakcie odczytu przemieszczający się nad/pod głowicą nośnik wytwarza zmienne pole magnetyczne wzbudzające w głowicy prąd elektryczny

51

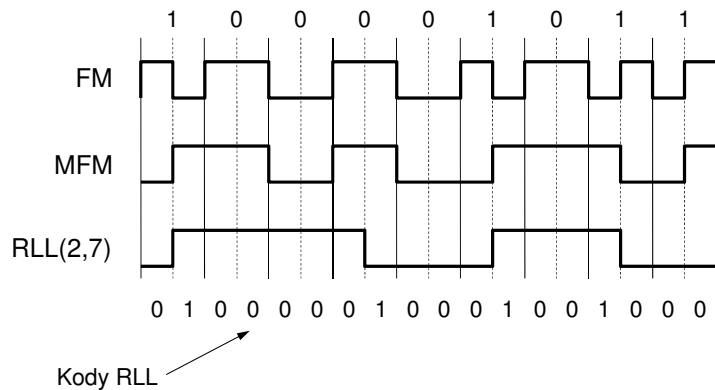
Pamięć dyskowa

- Ruch obrotowy dysku pozwala na dostęp do informacji zlokalizowanej w dowolnym miejscu koncentrycznego okręgu czyli pierścienia na dysku
- Dostęp do kolejnej „porcji danych” (kolejnego pierścienia na dysku) wymaga przesunięcia ramienia, na którym zamocowana jest głowica czyli tzw. pozycjonowania głowicy
- Aby w prawidłowy sposób odtworzyć zapisane dane (zinterpretować impulsy z głowicy odczytującej) zapisywana na nośniku informacja użyteczna musi zostać „wzbogacona” o informację synchronizującą wynikającą z przyjętego sposobu kodowania

52

Pamięć dyskowa

- Przegląd sposobów kodowania informacji:



53

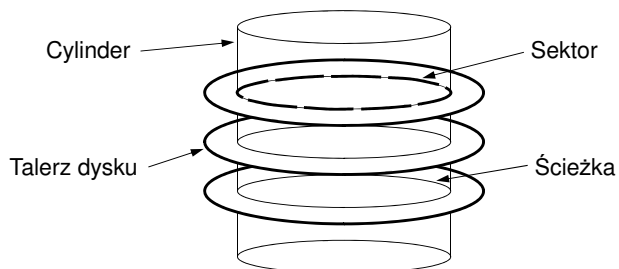
Pamięć dyskowa

- Nośnik w pamięci dyskowej (dysk) może być niewymienny (np. w dysku twardym) lub wymienny (*removable*) (np. w stacji dyskietek)
- Warstwa magnetyczna może być umieszczona po obu stronach nośnika (dysk dwustronny) lub tylko po jednej stronie (dysk jednostronny)
- W pamięci dyskowej może znajdować się jeden dysk lub ich większa liczba zamocowana na wspólnej osi
- Mechanizm głowicy może być kontaktowy, mieć ustaloną odległość (przerwę) od nośnika lub też odległość ta może być zmienna w wyniku oddziaływań aerodynamicznych

54

Pamięć dyskowa

- Przyjęta konstrukcja pamięci dyskowej powoduje, że dane organizowane są cyndry będące zbiorem koncentrycznych okręgów o tym samym promieniu zwanych ścieżkami, które dzielone są na mniejsze fragmenty (bloki danych) zwane sektorami



55

Pamięć dyskowa

- Początek każdej ścieżki oznaczony jest:
 - ciągiem bajtów synchronizujących (*SYNC*)
 - znacznikiem adresowym indeksu (*IAM*)
- W każdym sektorze można wyszczególnić:
 - ciąg bajtów synchronizujących (*SYNC*)
 - znacznik adresowy (*IDAM*)
 - identyfikator (adres) sektora (*ID*)
 - znacznik adresowy danych (*DAM*)
 - blok danych sektora
 - pola kodu ECC lub CRC (*Cyclic Redundancy Check*)
- Znacznik początku ścieżki oraz poszczególne sektory rozdzielone są od siebie przerwami (*GAP*)

56

Pamięć dyskowa

- Adres fizyczny sektora na dysku (tzw. adres CHS) określają trzy parametry:
 - numer cylindra (*Cylinder*)
 - numer głowicy (*Head*)
 - numer sektora (*Sector*)
- Czas dostępu do danych na dysku zależy od:
 - czasu przeszukiwania (*seek time*) związanego z pozycjonowaniem głowicy nad ścieżką
 - czasu opóźnienia obrotowego (*rotational latency*) związanego z przemieszczeniem się głowicy nad sektor
- Gęstość zapisu danych na dysku zależy od:
 - szerokości ścieżki
 - przyjęty sposób kodowania danych

57

Pamięć dyskowa

- W celu poprawy wydajności podsystemu dyskowego, a w szczególności zwiększenia jego niezawodności i szybkości działania stosuje się macierze dyskowe RAID (*Redundant Array of Independent Disks*)
- Macierz RAID zbudowana jest z wielu połączonych ze sobą dysków fizycznych, które widoczne są przez system operacyjny jako jeden dysk logiczny
- Dane zapisywane w macierzy rozpraszane są na wiele dysków, co gwarantuje odzyskanie danych w przypadku uszkodzenia dysku
- W praktyce stosuje się różne poziomy RAID oznaczane zwykle jako RAID 0 do RAID 5

58

Pamięć dyskowa

- Poziom RAID 0 (*striping*)
 - dane dzielone są na tzw. paski, które rozmieszczane są cyklicznie na kolejnych dyskach macierzy
 - brak redundancji danych
 - duża szybkość odczytu i zapisu danych
 - pojemność macierzy jest równa sumie pojemności wszystkich dysków wchodzących w jej skład
- Poziom RAID 1 (*mirroring*)
 - dane są paskowane i następnie odwzorowywane na dwóch odrębnych dyskach fizycznych
 - średnia szybkość odczytu, mała szybkość zapisu danych
 - pojemność macierzy jest równa połowie pojemności wszystkich dysków wchodzących w jej skład

59

Pamięć dyskowa

- Poziom RAID 2
 - dane są paskowane (na paski o rozmiarze mniejszym niż w RAID 0 i RAID 1) i rozmieszczane cyklicznie na kolejnych dyskach danych macierzy
 - na pozostałych dyskach (tzw. dyskach redundancyjnych) umieszczone zostają informacje korekcyjne, utworzone w oparciu o kod Hamminga na podstawie odpowiednich bitów z każdego dysku danych
 - liczba dysków redundancyjnych jest proporcjonalna do logarytmu liczby dysków danych
 - duża szybkość odczytu i zapisu (równoległy dostęp do wielu dysków)
 - bardzo wysoki koszt

60

Pamięć dyskowa

- Poziom RAID 3
 - dane są paskowane (na paski o podobnym rozmiarze jak w RAID 2) i rozmieszczane cyklicznie na kolejnych dyskach danych macierzy
 - na dysku redundancyjnym (tylko jednym niezależnie od rozmiarów macierzy) umieszczone zostają informacje korekcyjne, utworzone w oparciu o bity parzystości wyliczone dla odpowiednich bitów z każdego dysku danych
 - duża szybkość odczytu i zapisu (równoległy dostęp do wielu dysków)
- Poziom RAID 4
 - dane są paskowane (na paski o dużym rozmiarze) i rozmieszczane cyklicznie na kolejnych dyskach danych macierzy

61

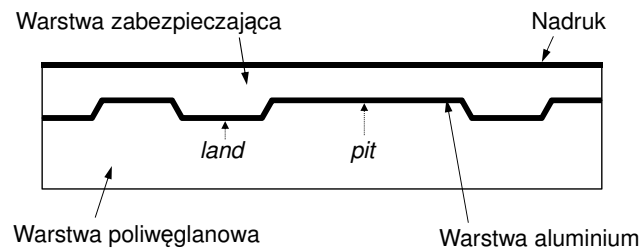
Pamięć dyskowa

- na dysku redundancyjnym (tylko jednym niezależnie od rozmiarów macierzy) umieszczone zostają paski informacji korekcyjnych, utworzone w oparciu o bity parzystości wyliczone dla odpowiednich pasków z każdego dysku danych
- dyski działają niezależnie od siebie, co pozwala na obsługę wielu żądań wejścia-wyjścia równoległe
- średnia szybkość odczytu, mała szybkość zapisu
- Poziom RAID 5 (*striping with parity*)
 - organizacja zbliżona do RAID 4
 - paski parzystości są rozproszone na wszystkie dyski w macierzy
 - średnia szybkość odczytu i zapisu, ale lepsza niż w RAID 1

62

Pamięć optyczna

- W dyskach CD-ROM (*Compact Disk ROM*) nośnikiem jest przezroczysta warstwa poliwęglanowa z wytłoczonymi wgłębieniami, która pokryta jest cienką warstwą aluminium i zabezpieczona od góry powłoką lakieru
- Przekrój poprzeczny nośnika CD-ROM:



63

Pamięć optyczna

- Promień lasera odbijając się od warstwy aluminium w obszarze zwanym landem ulega mniejszemu rozproszeniu niż odbijając się w obszarze zwanym pitem
- Po odbiciu od nośnika promień lasera pada przez półprzepuszczalny pryzmat i układ soczewek na światłoczułe detektory, których zadaniem jest:
 - przetworzenie zmian natężenia padającego światła na sygnał elektryczny (sygnał danych)
 - wytworzenie sygnałów sterujących układem ogniskowania wiązki
 - wytworzenie sygnałów sterujących układem pozycjonowania wiązki

64

Pamięć optyczna

- Podczas odczytu promień lasera przemieszcza się nad spiralną ścieżką z szybkością zależną od prędkości wirowania płyty
- W praktyce stosuje się dwie metody odczytu danych:
 - ze stałą prędkością liniową CLV (Constant Linear Velocity)
 - ze stałą prędkością kątową CAV (Constant Angular Velocity)
- Informacje na płycie CD-ROM kodowane są metodą EFM (*Eight to Fourteen Modulation*), w której każdy zapisywany bajt zastępowany jest ciągiem 14 bitów dobranym w taki sposób, aby dwie kolejne jedyńki były rozdzielone nie mniej niż dwoma i nie więcej niż dziesięcioma zerami (co oznacza 2 pity na bajt)

65

Pamięć optyczna

- Dane na płycie CD-ROM zorganizowane są fizycznie w ramki o następującej strukturze:
 - ciąg 27 bitów synchronizujących
 - bajt sterujący
 - dwa 16-bajtowe bloki (każdy złożony z ciągu 12 bajtów danych i 4 bajtów kontrolnych)
- Grupa połączonych 98 ramek tworzy logiczny sektor, na który składa się:
 - ciąg 12 bajtów synchronizujących
 - 4 bajtowy nagłówek
 - blok 2048 bajtów danych
 - blok 288 bajtów kontrolno-korekcyjnych

66

Pamięć optyczna

- Pojedynczy sektor zawiera 2048 bajtów danych zapisanych w trybie 1 (z korekcją błędów) lub 2336 bajtów danych zapisanych w trybie 2 (bez korekcji błędów)
- Bloki bajtów kontrolno-korekcyjnych połączone z odpowiednim rozlokowaniem (przeplotem) kolejnych ramek umożliwiają korekcję błędów metodą CIRC (*Cross-Interleave Reed-Solomon Coding*), która pozwala na skorygowanie ciągu około 450 błędnych bajtów danych
- Typowa płyta CD-ROM o średnicy 12 centymetrów może pomieścić 650-700MB danych

67

Pamięć optyczna

- Poza płytami CD-ROM przeznaczonymi wyłącznie do odczytu coraz większe znaczenie mają nośniki pozwalające na jednokrotny (CD-R) lub wielokrotny (CD-RW) zapis danych
- Warstwa poliwęglanowa w tego typu płytach pokryta jest specjalną substancją, która mętnieje (zmienia postać z krystalicznej na amorficzną) pod wpływem naświetlania promieniem laserem o zwiększonej mocy (co w konsekwencji powoduje zmniejszenie natężenia wiązki odbitej w trakcie odczytu)
- Substancja stosowana w płytach CD-RW może na powrót zmienić postać na krystaliczną

68

Pamięć optyczna

- Powszechnie stosowanym rodzajem pamięci optycznej o dużej pojemności przeznaczonym do odczytu są płyty DVD-ROM (*Digital Versatile Disk ROM*)
- Dzięki zastosowaniu lasera o mniejszej długości fali (co pozwoliło dwukrotnie zwiększyć gęstość zapisu i zmniejszyć odległości pomiędzy ścieżkami) oraz dwuwarstwowej struktury nośnika (pierwsza warstwa nośnika jest półprzezroczysta) uzyskano możliwość zapisania wielokrotnie większej ilości informacji niż na płytach CD-ROM, pomimo zastosowania nośnika tej samej wielkości

69

Pamięć optyczna

- W trakcie odczytu wiązka lasera jest ogniskowana odpowiednio na pierwszej lub drugiej warstwie nośnika
- Specyfikacja DVD-ROM dopuszcza stosowanie płyt:
 - jednostronnych jednowarstwowych o pojemności 4.7GB
 - jednostronnych dwuwarstwowych o pojemności 8.5GB
 - dwustronnych jednowarstwowych o pojemności 9.4GB
 - dwustronnych dwuwarstwowych o pojemności 17GB
- Format danych zapisanych na płycie DVD-ROM jest zbliżony do formatu danych na płycie CD-ROM, lecz z uwagi na zwiększoną gęstość zapisu został ulepszony mechanizm korekcji błędów

70

Pamięć magnetoptyczna

- Nośnikiem informacji w dyskach magnetoptycznych jest materiał, w którym następuje zmiana polaryzacji światła pod wpływem pola magnetycznego
- Odczyt danych polega na oświetleniu nośnika wiązką spolaryzowanego światła o odpowiednio dobranej płaszczyźnie polaryzacji
- Po odbiciu od nośnika płaszczyzna ta ulega skręceniu o kąt zależny od stopnia jego namagnesowania, a następnie przez półprzepuszczalny pryzmat i filtr polaryzacyjny (którego zadaniem jest tłumienie wiązki odbitej od obszaru nienamagnesowanego) trafia na układ detektora

71

Pamięć magnetoptyczna

- Zapis danych polega na wstępnym namagnesowaniu nośnika, a następnie na jego oświetleniu wiązką światła laserowego o dużej energii, co powoduje miejscowe podgrzanie nośnika powyżej tzw. punktu Curie i w konsekwencji jego rozmagnesowanie
- Typowe dyski magnetoptyczne są najczęściej wymienne i mają pojemność rzędu kilkuset MB
- Wadą tego typu pamięci jest wysoki koszt zarówno nośnika jak i samego napędu

72

Interfejs IDE (EIDE)

- Zadaniem interfejsu EIDE (*Enhanced Integrated Device Electronics*) będącego rozszerzeniem interfejsu IDE jest podłączenie do systemu komputerowego takich urządzeń zewnętrznych jak dyski twarde, napędy optyczne i magnetoptyczne
- Cechą charakterystyczną obsługiwanych przez ten interfejs urządzeń jest to, że zintegrowano w nich mechanizm napędu wraz z układem sterującym (kontrolerem) odpowiedzialnym m.in. za pozycjonowanie głowic, kodowanie/dekodowanie sygnału, kontrolę i korekcję błędów, translację adresów oraz przesyłanie danych

73

Interfejs IDE (EIDE)

- Standardy interfejsu IDE (AT-BUS):
 - 1994: Standard ATA-1 (*Advanced Technology Attachment, AT Attachment*)
 - transmisja danych w trybie PIO
 - prędkość transmisji do 4 MB/s
 - obsługa najwyżej dwóch dysków twardech
 - 1996: Standard ATA-2
 - transmisja danych w trybach PIO i DMA
 - prędkość transmisji do 16.6 MB/s
 - 1997: Standard ATA-3
 - system autodiagnostyki S.M.A.R.T (*Self Monitoring, Analysis and Reporting Technology*)

74

Interfejs IDE (EIDE)

- 1998: Interfejs EIDE (Fast ATA), standard ATA/ATAPI-4
 - transmisja danych w trybach PIO, DMA i Ultra-DMA
 - prędkość transmisji do 33 MB/s
 - interfejs logiczny ATAPI (*AT Attachment Packet Interface*) pozwalający na obsługę takich urządzeń jak napędy CD-ROM, streamery itp.
 - obsługa do czterech urządzeń zewnętrznych
- 2000: Standard ATA/ATAPI-5 (Ultra ATA/100)
 - prędkość transmisji do 100 MB/s
- Standard ATA/ATAPI-6 (Ultra ATA/133)
 - prędkość transmisji do 133 MB/s
 - 48-bitowy model adresowania LBA
 - funkcje AAM (*Automatic Acoustic Management*)

75

Interfejs IDE (EIDE)

- Magistrala interfejsu IDE połączona jest z magistralą systemową przez kontroler IDE (*IDE Host-Adapter*), którego zadaniem jest dekodowanie adresu oraz buforowanie sygnałów na magistralach
- Standardowy interfejs IDE pozwala na dołączenie do kontrolera dwóch urządzeń: nadrzędnego (*Master*) i podrzędnego (*Slave*), wykorzystując w tym celu 40-, 50- lub 80-żyłowy przewód o długości do 46 cm
- W przypadku interfejsu EIDE wykorzystuje się dwa kanały IDE: IDE 0 (*Primary*) oraz IDE 1 (*Secondary*), dzięki czemu możliwe jest podłączenie do systemu czterech urządzeń

76

Interfejs IDE (EIDE)

- Kanały kontrolera EIDE są od siebie całkowicie niezależne tzn. do komunikacji z systemem wykorzystują oddzielne przerwania i adresy wejścia-wyjścia rejestrów kontrolno-sterujących CBR (*Control Block Registers, Command Block Registers*):
 - IDE 0: IRQ 14, Porty 1F0÷1F7, 3F6
 - IDE 1: IRQ 15, Porty 170÷177, 376
- Przesyłanie danych pomiędzy pamięcią operacyjną a urządzeniem zewnętrznym może odbywać się w trybach:
 - programowanego wejścia-wyjścia: PIO-0 (3.33 MB/s), PIO-1 (5.22 MB/s), PIO-2 (8.33 MB/s), PIO-3 (11.11 MB/s) lub PIO-4 (16.66 MB/s)

77

Interfejs IDE (EIDE)

- bezpośredniego dostępu do pamięci:
 - Singleword DMA 0 (2.08 MB/s), Singleword DMA 1 (4.16 MB/s), Singleword DMA 2 (8.33 MB/s)
 - Multiword DMA 0 (4.16 MB/s), Multiword DMA 1 (13.33 MB/s), Multiword DMA 2 (16.66 MB/s)
 - Ultra DMA 1 (25.00 MB/s), Ultra DMA 2 (33.33 MB/s), Ultra DMA 3 (44.44 MB/s), Ultra DMA 4 (66.64 MB/s), Ultra DMA 5 (99.96 MB/s), Ultra DMA 6 (133.28 MB/s)
- Komunikacja z urządzeniami IDE realizowana jest za pomocą rozkazów należących do:
 - zbioru podstawowego (*General Feature Set*)
 - zbioru rozkazów trybu pakietowego (*Packet Command Feature Set*)

78

Interfejs IDE (EIDE)

- Specyfikacja ATA/ATAPI pozwala na obsługę urządzeń implementujących podstawowe (*Power Management*) lub zaawansowane (*Advanced Power Management*) funkcje zarządzania poborem mocy
- Podstawowy mechanizm zarządzania poborem mocy definiuje cztery możliwe tryby pracy urządzenia:
 - tryb gotowości (*Active*)
 - tryb jałowy (*Idle*)
 - tryb czuwania (*Standby*)
 - tryb uśpienia (*Sleep*)
- W trybie zaawansowanym możliwe jest ustawienie jednego z 254 możliwych poziomów poboru energii

79

Interfejs IDE (EIDE)

- Urządzenia IDE obsługujące protokół ATAPI mogą implementować następujące mechanizmy:
 - System zabezpieczeń (Security Mode) chroniący dostęp do urządzenia mechanizmem dwóch haseł: administratora (*master password*) i użytkownika (*user password*)
 - System autodiagnostyki S.M.A.R.T (*Self Monitoring, Analysis and Reporting Technology*) pozwalający na bieżącą kontrolę krytycznych parametrów pracy dysku, co zapewnia wczesne wykrywanie problemów i minimalizowanie skutków awarii
 - System zarządzania parametrami akustycznymi (*Automatic Acoustic Management*) pozwalający na wybór „kompromisu” pomiędzy szybkością pracy napędu, a emitowanym przez niego poziomem hałasu

80

Interfejs Serial ATA (SATA)

- Interfejs szeregowy Serial ATA służy do podłączenia do systemu komputerowego takiej samej klasy urządzeń zewnętrznych jak interfejs EIDE
- Cechy interfejsu Serial ATA:
 - przesyłanie danych z szybkością od 1.5 Gb/s (SATA I)
 - programowa zgodność ze standardem ATA/ATAPI
 - niższy pobór mocy (niższe napięcie magistrali)
 - topologia gwiazdy (każde urządzenie połączone jest z kontrolerem własnym przewodem)
 - zwiększona odporność na błędy (transmisja dwukierunkowa w systemie różnicowym)
 - możliwość zmiany konfiguracji przyłączonych urządzeń w trakcie pracy (*Hot Plugging*)

81

Interfejs Serial ATA (SATA)

- Interfejs SATA wykorzystuje 7-żyłowy przewód połączeniowy (cztery linii sygnałowe plus trzy linie masy) o długości mniejszej niż 1m
- Dane fizycznie przesyłane interfejsem Serial ATA zorganizowane są w 32 bitowe słowa, kodowane regułą 8b/10b (NRZ), co zapobiega jednostronnej polaryzacji kabla
- Wymiana informacji pomiędzy kontrolerem Serial ATA i urządzeniem zewnętrznym odbywa się za pomocą:
 - symboli (*Primitives*)
 - ramek FIS (*Frame Information Structure*)

82

Interfejs Serial ATA (SATA)

- Symbol jest pojedynczym słowem, którego zadaniem jest przesłanie:
 - informacji o stanie magistrali i urządzenia
 - sygnatury oznaczającej początek (SOF) lub koniec (EOF) przesyłanej ramki
 - znaku synchronizującego (SYNC) w stanie spoczynkowym (*Idle*) magistrali
- Ramka FIS ma zmienną długość i jest zbudowana z:
 - nagłówka (*Header*) o długości 1 słowa zawierającego m.in. sygnaturę (*FIS Type*) i długość ramki
 - bloku danych użytkowych o długości do 2048 słów
 - kodu CRC o długości 1 słowa

83

Interfejs Serial ATA (SATA)

- Ramki FIS służą do przesyłania danych użytkowych, diagnostycznych i rozkazów programujących rejestry trybu pracy SRB (*Shadow Register Block*)
- Specyfikacja Serial ATA definiuje symetryczny (z potwierdzeniem) protokół transmisji danych (zestaw reguł opisujących proces przesyłania danych), według którego, każda odebrana ramka wymaga przesłania zwrotnie symbolu potwierdzającego poprawny odbiór ramki (ROK) lub sygnalizującego błąd transmisji (RERR)
- Transfer danych po magistrali Serial ATA może odbywać się zarówno w trybie PIO, jak i DMA

84

Interfejs Serial ATA (SATA)

- Rozwój interfejsu SATA:
 - SATA I: przepustowość 1.5 Gb/s (ok. 180 MB/s)
 - SATA II: przepustowość 3 Gb/s (ok. 375 MB/s), kolejowanie zadań NCQ (*Native Command Queuing*), możliwość przyłączenia wielu urządzeń do jednego portu dzięki powielaczom portów (*Port Multipliers*), redundancja połączeń dzięki selektorom portów (*Port Selectors*)
 - SATA III: przepustowość 6 Gb/s (ok. 750 MB/s), ulepszony mechanizm NCQ, tryb izochorycznego transferu danych
 - eSATA (external SATA): zewnętrzny port SATA 3Gb/s do podłączania urządzeń zewnętrznych na odległość do 2 m
 - xSATA: rozwinięcie eSATA do transmisji na odległość do 8 m

85

Interfejs SCSI

- Standard SCSI (*Small Computer System Interface*) powstał w 1982 roku dla potrzeb obsługi urządzeń zewnętrznych (m.in. dysków twardych, napędów taśmowych, czytników CD-ROM, skanerów) przez minikomputery i wydajne stacje robocze
- Magistrala SCSI została opracowana jako szybka, uniwersalna, elastyczna i niezawodna magistrala równoległa obsługująca wiele urządzeń (także różnych producentów) przyłączonych do jednego kontrolera i montowanych zarówno wewnątrz, jak i na zewnątrz obudowy komputera

86

Interfejs SCSI

- Standardy interfejsu SCSI:
 - Standard SCSI-1 (1986)
 - magistrala danych o szerokości 8 bitów
 - adresowanie do 8 urządzeń
 - prędkość transmisji do 5 MB/s
 - długość przewodu połączeniowego do 6m
 - Standard SCSI-2 (1990)
 - magistrala danych o szerokości 16 lub 32 bitów (Wide SCSI)
 - adresowanie do 16 urządzeń
 - szybkość transmisji do 10 MB/s (tryb Fast SCSI), 20 MB/s (tryby Fast Wide SCSI i Ultra SCSI) lub 40 MB/s (tryby Wide Ultra SCSI i Ultra2 SCSI)

87

Interfejs SCSI

- długość przewodu połączeniowego do 25m
- Standard Ultra3 (Ultra160) SCSI , Ultra4 (Ultra320) SCSI
 - wykorzystanie obydwu zboczy sygnału taktującego pozwala na maksymalny transfer 160 MB/s (320 MB/s)
 - możliwość korzystania z różnych interfejsów fizycznych pełna kontrola integralności przesyłanych danych kodami CRC (Ultra320 SCSI)
 - obsługa protokołu pakietowego (Ultra320 SCSI)
 - adresowanie do 32 urządzeń
 - wprowadzenie protokołu diagnostycznego jakości łącza
 - przyspieszenie arbitrażu w dostępie do magistrali
- Standard Ultra640 SCSI
 - maksymalny transfer 640 MB/s

88

Interfejs SCSI

- Każde urządzenie podłączone do magistrali SCSI ma przydzielony unikalny identyfikator ID (liczba z zakresu 0÷7 lub 0÷15) określający adres i priorytet urządzenia
- Najczęściej urządzeniem o najwyższym priorytecie (identyfikatorze) jest kontroler SCSI (*Host Adapter*)
- Długość przewodu połączeniowego elektrycznej wersji magistrali SCSI wynosi:
 - do 6m w przypadku magistrali pojedynczej (*single-ended*)
 - do 25m w przypadku magistrali różnicowej (*differential*)
 - do 12m w przypadku magistrali niskonapięciowej-różnicowej LVD (*low-voltage differential*)

89

Interfejs SCSI

- Urządzenia do magistrali przyłączone są łańcuchowo (równolegle), przy czym końce magistrali muszą być wyposażone w terminatory tj. mostki rezystorowe zapewniające dopasowanie falowe
- Szerokość magistrali SCSI zależy od jej realizacji – w przypadku jej elektrycznej wersji (SCSI-2) wynosi odpowiednio:
 - 50 linii sygnałowych dla 8-bitowej magistrali danych
 - 68 linii sygnałowych dla 16-bitowej magistrali danych
- Magistrala SCSI pozwala na niezależną komunikację między różnymi parami dołączonych do niej urządzeń w ramach tzw. kanałów logicznych

90

Interfejs SCSI

- Specyfikacja SCSI definiuje protokół wymiany informacji przez magistralę, obejmujący następujące fazy pracy:
 - Faza wolnej magistrali (*Bus Free*), w której żadne urządzenie nie korzysta z magistrali
 - Faza arbitrażu (*Arbitration*), w której dokonywany jest wybór urządzenia (inicjującego lub docelowego), które przejmie kontrolę nad magistralą
 - Faza selekcji (*Selection*), w której urządzenie inicjujące dokonuje wyboru urządzenia docelowego (przesyłając na magistralę identyfikator swój i urządzenia docelowego)
 - Faza rozkazowa (*Command*), w której przesyłane są rozkazy do urządzenia docelowego

91

Interfejs SCSI

- Faza reSelekcji (*Reselection*), w której następuje nawiązanie połączenia przez urządzenie docelowe z urządzeniem inicjującym (faza taka ma miejsce najczęściej po wykonaniu przez urządzenie docelowe rozkazu, w trakcie wykonania którego magistrala była zwolniona)
- Faza przesyłania danych (*Data*), w której następuje przesyłanie danych między urządzeniami
- Faza przesyłania raportu (*Status*), w której przesyłana jest informacja o sposobie zakończenia wykonywanej operacji
- Faza przesyłania wiadomości (*Message*), w której przesyłane są (dwukierunkowo) informacje pomocnicze
- Faza zgłaszania przesłania wiadomości (*Attention*), w której urządzenie inicjujące zgłasza żądanie wysłania wiadomości
- Faza zerowania (*Reset*), w której następuje zerowanie wszystkich urządzeń na magistrali

92

Interfejs SCSI

- Przesyłanie danych magistralą SCSI może odbywać się w trybie asynchronicznym lub synchronicznym
- Rozkazy SCSI przesyłane są w tzw. blokach CDB (*Command Data Block*) o długości 6, 10 lub 12 bajtów, zawierających odpowiednio:
 - kod operacji
 - parametry rozkazu
 - bajt kontrolny
- Urządzenie docelowe wyposażone jest najczęściej w kolejkę rozkazów, dzięki której możliwe jest bardziej efektywne wykorzystanie magistrali

93

Interfejs IEEE-1284

- Interfejs IEEE-1284 jest standardem przesyłania informacji między urządzeniami za pośrednictwem łącza równoległego
- Specyfikacja standardu IEEE-1284 definiuje pięć trybów pracy łącza równoległego:
 - tryb kompatybilny (*Compatibility Mode*)
 - tryb półbajtowy (*Nibble Mode*)
 - tryb bajtowy PS/2 (*Byte Mode*)
 - tryb EPP (*Enhanced Parallel Port*)
 - tryb ECP (*Extended Capabilities Port*)
- Pierwsze trzy z wymienionych trybów pracy określa się także jako tryb SPP (*Standard Parallel Port*)

94

Interfejs IEEE-1284

- Interfejs IEEE-1284 pracujący w trybie SPP wykorzystuje oprócz 8-bitowej magistrali danych, następujące linie sygnałowe:
 - wejściowy sygnał błędu drukarki: ERR (*Error*)
 - wyjściowy sygnał strobujujący dane: STR (*Strobe*)
 - wejściowy sygnał potwierdzenia: ACK (*Acknowledge*)
 - wejściowy sygnał zajętości: BSY (*Busy*)
 - wejściowy sygnał braku papieru: PERR (*Paper Error*)
 - wejściowy sygnał gotowości (stanu „on-line”): SEL (*Select*)
 - wyjściowy sygnał wysuwu wiersza: ALF (*Auto Line Feed*)
 - wyjściowy sygnał zerowania (inicjowania) drukarki: INI (*Initialization*)
 - wyjściowy sygnał wyboru drukarki: DSL (*Device Select*)

95

Interfejs IEEE-1284

- Tryb kompatybilny zwany także standardowym lub trybem Centronics realizuje jednokierunkową transmisję danych z prędkością do 200 kB/s
- Tryb półbajtowy pozwala na dwukierunkową transmisję danych, w której do odbierania danych wykorzystuje się linie BSY, ACK, PERR i SEL połączone z magistralą danych przez multiplekser sterowany sygnałem STR
- Tryb bajtowy realizuje dwukierunkową transmisję danych opartą na wykorzystaniu dwukierunkowego portu danych (*Bi-Directional Data Port*)

96

Interfejs IEEE-1284

- Dostęp do łącza równoległego pracującego w trybie SPP odbywa się za pośrednictwem:
 - rejestru danych (*Data Register*), umożliwiającego zapis/odczyt danych z magistrali danych interfejsu
 - rejestru stanu (*Status Register*), pozwalającego m.in. na odczyt stanu linii BSY, ACK, PERR, SEL i ERR
 - rejestru sterującego (*Control Register*), pozwalającego m.in. na aktywowanie dwukierunkowej pracy portu danych oraz na ustawianie stanu linii DSL, INI, ALF i STR
- Implementacja protokołów transmisji danych w trybach SPP (polegająca m.in. na wytworzeniu właściwych sekwencji sygnałów sterujących) musi być zrealizowana programowo

97

Interfejs IEEE-1284

- Tryb EPP pozwala na realizację dwukierunkowej transmisji danych z prędkością do 2 MB/s
- Protokół wymiany danych w trybie EPP wykorzystuje następujące linie sygnałowe:
 - sygnał kierunku transmisji: WRITE (linia STR trybu SPP)
 - sygnał wyzwolenia przerwania: INTERRUPT (linia ACK trybu SPP)
 - sygnał zakończenia cyklu EPP: WAIT (linia BSY trybu SPP)
 - sygnał strobujący dane: DATA STROBE (linia ALF trybu SPP)
 - sygnał strobujący adres: ADDR STROBE (linia DSL trybu SPP)
 - sygnał zerowania (inicjowania): RESET (linia INI trybu SPP)

98

Interfejs IEEE-1284

- Dostęp do łącza równoległego pracującego w trybie EPP odbywa się za pośrednictwem zestawu rejestrów SPP, rozszerzonego o:
 - dwukierunkowy rejestr danych (*Data Register*)
 - dwukierunkowy rejestr adresowy (*Address Register*)
- Wytworzenie właściwych sekwencji sygnałów sterujących związanych z realizacją protokołu transmisji danych w trybie EPP realizowane jest automatycznie przez kontroler łącza równoległego
- Transmisję danych i/lub adresów realizuje się przez ich bezpośrednie wpisanie do odpowiednich rejestrów EPP

99

Interfejs IEEE-1284

- Tryb ECP pozwala na realizację dwukierunkowej transmisji danych z prędkością do 2 MB/s
- Protokół wymiany danych w trybie ECP wykorzystuje m.in. następujące linie sygnałowe:
 - sygnał zegarowy: HostCLK (linia STR trybu SPP)
 - sygnał zegarowy: PeriphCLK (linia ACK trybu SPP)
 - sygnał sterujący przesyłaniem danych/rozkazów: PeriphACK (linia BSY trybu SPP)
 - sygnał sterujący przesyłaniem danych/rozkazów: HostACK (linia ALF trybu SPP)
 - sygnały zmiany kierunku transmisji: RevREQ (linia INI trybu SPP) oraz AckREV (linia PERR trybu SPP)

100

Interfejs IEEE-1284

- Tryb ECP pozwala na korzystanie z:
 - przesyłania danych z kontrolera w trybie DMA
 - dodatkowego (najczęściej 16-stopniowego) rejestru buforowanego FIFO
 - kompresji przesyłanych danych algorytmem RLE (*Run Length Encoding*)
 - definiowania i adresowania do 128 kanałów logicznych
- Dostęp do łącza równoległego pracującego w trybie ECP odbywa się za pośrednictwem zestawu rejestrów SPP, rozszerzonego m.in. o:
 - rejestr adresowy FIFO (*ECP Address FIFO*)
 - rejestry danych FIFO (*Data FIFO*)
 - rejestr sterujący ECR (*Extended Control Register*)

101

Interfejs IEEE-1284

- Wytworzenie właściwych sekwencji sygnałów sterujących związanych z realizacją protokołu transmisji danych w trybie ECP realizowane jest automatycznie przez kontroler łącza równoległego
- Interfejs IEEE-1284 korzysta z poziomów logicznych TTL, a długość kabla połączeniowego (niezależnie od trybu pracy) nie powinna przekraczać 5 m
- Gniazdo przyłączeniowe interfejsu IEEE-1284 może mieć 25 pinów (DB-25) lub 36 pinów (Centronics)

102

Interfejs RS-232C

- Interfejs RS-232C (V24) jest standardem przesyłania informacji między dwoma urządzeniami za pośrednictwem asynchronicznego łącza szeregowego
- Dane przesyłane tego typu łączem są zorganizowane w ramki SDU (*Serial Data Unit*), składające się z:
 - znacznika początku ramki (bitu startu o wartości 0)
 - 5, 6, 7 lub 8 bitów danych
 - opcjonalnego bitu kontroli parzystości
 - znacznika końca ramki (1, 1½ lub 2 bity stopu o wartości 1)
- Przed nawiązaniem połączenia oba urządzenia muszą mieć ustawione takie same parametry transmisji tj. częstotliwość taktowania i format ramki

103

Interfejs RS-232C

- Odbiornik w celu zsynchronizowania swojego zegara z zegarem nadajnika próbkuje stan linii odbiorczej z częstotliwością wielokrotnie wyższą od częstotliwości z jaką przesyłane są dane
- Przekształcaniem danych z postaci równoległej na szeregową i odwrotnie zajmuje się układ UART (*Universal Asynchronous Receiver-Transmitter*)
- Transmisja danych może odbywać się z maksymalną prędkością 56 kb/s (UART 8250) lub 115200 kb/s (UART 16450) w trybie:
 - jednokierunkowym (simpleksowym)
 - dwukierunkowym (półdupleksowym i duplexowym)

104

Interfejs RS-232C

- Standard RS-232C definiuje protokół przesyłania danych między terminalem DTE (*Data Terminal Equipment*) i modemem DCE (*Data Communication Equipment*), wykorzystujący następujące linie sygnałowe:
 - dane wysyłane: TxD (*Transmitted Data*)
 - dane odbierane: RxD (*Received Data*)
 - żądanie rozpoczęcia nadawania: RTS (*Request To Send*)
 - potwierdzenie sygnału RTS (zgłoszenie gotowości odbioru danych z terminala): CTS (*Clear To Send*)
 - gotowość modemu na przyjęcie danych: DSR (*Data Set Ready*)
 - gotowość terminala: DTR (*Data Terminal Ready*)

105

Interfejs RS-232C

- odbiór fali nośnej: DCD (*Data Carrier Detect*)
- sygnał wywołania abonenta: RI (*Ring Indicator*)
- Para sygnałów RTS/CTS jest odpowiedzialna za sterowanie przepływem danych (oraz za sterowanie kierunkiem transmisji w trybie półduplexowym)
- Para sygnałów gotowości DTR/DSR odpowiada za utrzymanie nawiązanego połączenia
- Interfejs RS-232C do reprezentowania stanów logicznych wykorzystuje napięcia z zakresu od -15 do +15 V (poziom logiczny „0” odpowiada napięciu na linii sygnałowej od +3 do +15 V, natomiast poziom logiczny „1” odpowiada napięciu od -3 do -15 V)

106

Interfejs RS-232C

- Typowe gniazdo przyłączające interfejsu RS-232C ma 25 (DB-25) lub 9 pinów (DB-9)
- Maksymalna odległość między nadajnikiem i odbiornikiem nie przekracza 200 m
- Istnieje możliwość wymiany danych pomiędzy dwoma urządzeniami typu DTE połączonych tzw. kablem modemu zerowego (*null modem cable*)
- Obsługa łącza szeregowego może być realizowana przez:
 - przerwania zgłaszane przez układ UART
 - okresowe testowanie przez procesor rejestru stanu transmisji układu UART

107

Magistrala USB

- Magistrala USB (*Universal Serial Bus*) jest standardem łącza szeregowego przeznaczonym do podłączania szerokiej gamy urządzeń peryferyjnych
- Cechy magistrali USB:
 - prędkość transmisji: 1.5 Mb/s (*Low Speed*), 12 Mb/s (*Full Speed*), 480 Mb/s (*High Speed*) lub 4.8 Gb/s (*Super Speed*)
 - obsługa przez jeden kontroler (*Host Adapter*) do 127 urządzeń USB
 - topologia połączeń typu gwiazda (o maksymalnie siedmiu poziomach) realizowana przez koncentratory USB (*Hub*)
 - maksymalna długość kabla połączeniowego wynosi 5 m (maksymalna odległość między urządzeniem i kontrolerem nie może przekroczyć 35 m)

108

Magistrala USB

- możliwość zmiany konfiguracji podłączonych urządzeń w trakcie pracy (*Hot-Plugging*)
- automatyczny przydział identyfikatorów (adresów) do urządzeń (*Auto-ID*)
- możliwość zasilania podłączonych urządzeń kablem USB (obciążenie do 500 mA dla USB 2.0, 900mA dla USB 3.0)
- Dane kodowane metodą NRZI (*Non Return to Zero Invert*) przesyłane są parą przewodów sygnałowych: D+ i D- w sposób różnicowy (napięcie od 0 do 3.6 V)
- Kabel połączeniowy magistrali USB to czterożyłowa skrętka (w USB 3.0 – ośmiożyłowa) zakończona złączami standardowymi (płaskim - A i kwadratowym - B) lub złączami miniaturowymi (mini, micro)

109

Magistrala USB

- Wymiana informacji magistralą USB odbywa się w pakietach przez wirtualne kanały przesyłowe (*Pipes*) tworzone na żądanie przez kontroler USB
- Każdy kanał przesyłowy może wykorzystywać jeden z czterech trybów transmisji:
 - tryb asynchroniczny nie gwarantujący określonej przepustowości (*Bulk Transfer*), który przeznaczony jest do obsługi podstawowych urządzeń zewnętrznych (np. drukarka, skaner)
 - tryb synchroniczny o gwarantowanej szerokości pasma (*Isochronous Transfer*), który przeznaczony jest do obsługi urządzeń przesyłających duże porcje danych z określoną prędkością (np. kamera)

110

Magistrala USB

- tryb natychmiastowy (*Interrupt Transfer*), który przeznaczony jest do obsługi urządzeń asynchronicznych przesyłających niewielkie porcje danych (np. myszka, klawiatura)
- tryb sterujący (*Control Transfer*), który przeznaczony jest do konfigurowania nowo podłączonych do magistrali urządzeń
- Wszystkie przesyłane magistralą USB dane transmitowane są w kolejności od najmłodszego do najstarszego bitu (bajta)
- Każdy przesyłany pakiet rozpoczyna się 8 lub 32-bitową sekwencją synchronizującą (*SYNC*), po której przekazywany jest 8-bitowy identyfikator pakietu (*PID*) określający m.in. rodzaj pakietu

111

Magistrala USB

- Specyfikacja USB definiuje następujące rodzaje pakietów:
 - żeton (*Token*) służący do określania adresu urządzenia USB (7 bitów ADDR + 4 bity ENDP), kierunku przepływu pakietów danych oraz testowania aktywności urządzenia
 - pakiet danych (*Data*) służący do przesyłania danych w blokach o długości 0÷1024 bajtów zabezpieczonych 16-bitowym kodem CRC
 - pakiet potwierdzający (*Handshake*) służący do przekazania raportu o przesłanych pakietach danych czy informacji o aktywności urządzenia
 - pakiet specjalny (*Special*) służący do sterowania przepływem danych w magistrali, do której podłączone są urządzenia o różnej prędkości transmisji

112

Magistrala FireWire

- Magistrala FireWire (i.Link, IEEE-1394) jest odmianą łącza szeregowego służącą do podłączania szerokiej gamy urządzeń peryferyjnych (przede wszystkim audiowizualnych)
- Cechy magistrali FireWire:
 - transmisja w trybie asynchronicznym z prędkością 100 Mb/s (S100), 200 Mb/s (S200) lub 400 Mb/s (S400)
 - negocjowana minimalna gwarantowana przepustowość pasma (*Guaranteed Bandwidth*)
 - automatyczny przydział numerów (adresów) do urządzeń (*Auto-ID*)
 - możliwość zmiany konfiguracji podłączonych urządzeń w trakcie pracy (*Hot-Plugging*)

113

Magistrala FireWire

- automatyczne dopasowywanie końców magistrali impedancją falową
- fizyczna topologia połączeń typu drzewo (logiczna magistrala) złożona z kaskadowo połączonych łańcuchów (drzew) węzłów (*Nodes*) spajanych pomostami (*Bridges*)
- maksymalna głębokość pojedynczego drzewa wynosi 16
- 16-bitowe adresowanie węzłów (pojedynczy pomost pozwala na dołączenie do 63 węzłów)
- 10-bitowe adresowanie pomostów
- maksymalna długość kabla połączeniowego wynosi 4.5 m (maksymalna odległość między urządzeniem i kontrolerem nie może przekroczyć 72 m)
- możliwość zasilania urządzeń izolowanych (maksymalne obciążenie do 1.5 A)

114

Magistrala FireWire

- Sześćżyłowy kabel połączeniowy magistrali FireWire (2 pary ekranowanej skrętki + 2 linie zasilające) zakończony jest jednym z dwóch typów gniazd: 6 (1394-1995) lub 4 pinowym (1394a-2000)
- Pierwsza para skrętek służy do przesyłania strumienia danych, druga do przesyłania impulsów synchronizujących (transmisja typu simplex)
- Wymiana danych odbywa się poprzez kanały wirtualne służące zarówno do realizacji klasycznej transmisji asynchronicznej jak i transmisji z gwarantowaną szerokością pasma (*Isochronous Data Transport*)

115

Magistrala FireWire

- Najnowsza wersja standardu IEEE-1394b wprowadza następujące rozszerzenia:
 - tryby prędkości transmisji: 400 Mb/s, 800 Mb/s, 1600 Mb/s i 3200 Mb/s
 - kodowania danych regułą 8/10, co pozwala na równoczesne korzystanie z obu par skrętek sygnałowych do transmisji danych
 - zwiększenie maksymalnej długości kabla połączeniowego do 100 m wykorzystując okablowanie strukturalne kat. 5
 - możliwość realizacji połączeń na bazie innych mediów (np. światłowodu)
 - wprowadzenie nowych typów kabli połączeniowych (9-żyłowych) i nowych złączy
 - optymalizacja protokołu transmisji

116

Architektura procesora

- Począwszy od lat 60-tych procesory budowane były w myśl zasady CISC (*Complex Instruction Set Computer*) cechującej się tendencją do rozszerzania listy rozkazów o nowe rozkazy, realizujące coraz bardziej złożone zadania (np. przesłania blokowe, operacje na łańcuchach)
- Lista rozkazów typowego procesora CISC obejmuje nawet kilkaset rozkazów cechujących się zmienną długością i czasem wykonania od jednego do kilkudziesięciu taktów zegara
- Na pojedynczy rozkaz składa się sekwencja kilku do kilkunastu tzw. mikrooperacji

117

Architektura procesora

- Analiza kodu generowanego przez kompilatory języków wysokiego poziomu pozwoliła na wyciągnięcie następujących wniosków:
 - za 80% czasu wykonywania programu odpowiada 20% instrukcji (reguła 80/20)
 - dominującą grupą rozkazów stanowią proste instrukcje przypisania, instrukcje warunkowe (IF, LOOP) oraz wywołania i powroty z podprogramów
 - większość używanych argumentów stanowią zmienne skalarne (głównie zmienne lokalne w procedurach)
- Powyższe obserwacje doprowadziły do powstania koncepcji procesora o zredukowanej liczbie rozkazów RISC (*Reduced Instruction Set Computer*)

118

Architektura procesora

- Kluczowymi elementami koncepcji RISC jest:
 - nieduży zestaw stosunkowo prostych rozkazów o stałej długości, pozwalający na znaczące uproszczenie konstrukcji jednostki sterującej i układu dekodera instrukcji oraz na zwiększenie szybkości przetwarzania rozkazów
 - wielopoziomowa pamięć podręczna usprawniająca pobieranie zwiększonej w stosunku do procesorów CISC liczby rozkazów
 - duża liczba rejestrów wewnętrznych umożliwiająca optymalizację odwołań do argumentów (zwłaszcza lokalnych)
 - optymalizacja potoków wykonawczych z uwagi na duży udział w kodzie rozkazów rozgałęzień warunkowych oraz wywołań i powrotów z podprogramów

119

Architektura procesora

- Procesory oparte na architekturze RISC (np. Intel i860, Sun SPARC, IBM RS/6000, Motorola MC88000) cechują się pewnymi wspólnymi własnościami:
 - wykonywanie większości rozkazów w jednym cyklu zegara
 - prosty format i stała długość rozkazu (dostosowana do szerokości magistrali danych)
 - jawne określanie rejestrów jako argumentów w kodzie rozkazu
 - mała liczba prostych trybów adresowania (głównie adresowanie natychmiastowe, bezpośrednie i rejestrowe)
- Wiele obecnie konstruowanych procesorów łączy w sobie cechy architektur CISC i RISC (np. Intel 80960, Pentium 4, PowerPC 601)

120

Przetwarzanie potokowe

- Przetwarzanie pojedynczego rozkazu można dekomponować na szereg etapów (faz), które realizowane są przez niezależne obwody procesora
- Strategia przetwarzania potokowego (*Pipeline*) polega na zorganizowaniu procesu przetwarzania rozkazów w sposób zbliżony do linii montażowej w zakładzie produkcyjnym
- Podobnie jak linia montażowa potok zawiera rozkazy znajdujące się w różnym stadium przetwarzania, co prowadzi do zwiększenia szybkości ich przetwarzania
- Podstawowym parametrem mającym wpływ na efektywność potoku jest jego głębokość

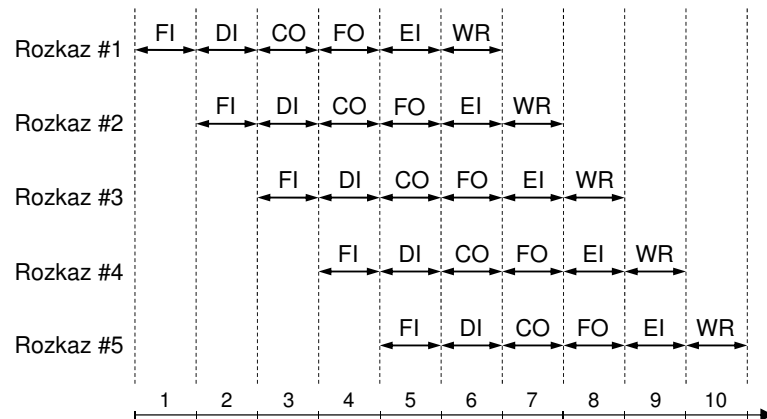
121

Przetwarzanie potokowe

- Dla przykładu 6-etapowy potok (z jednostkowymi czasami trwania poszczególnych etapów), który wykorzystuje następującą dekompozycję przetwarzania rozkazu:
 - Faza pobierania rozkazu (FI)
 - Faza dekodowania rozkazu (DI)
 - Faza obliczania argumentów (CO)
 - Faza pobierania argumentów (FO)
 - Faza wykonania rozkazu (EI)
 - Faza zapisania wyniku (WR)
- pozwała na skrócenie czasu przetwarzania pięciu rozkazów z 30 do 10 cykli zegara

122

Przetwarzanie potokowe



123

Przetwarzanie potokowe

- Czynniki obniżające wydajność przetwarzania potokowego:
 - różny czas trwania poszczególnych etapów prowadzi do powstawania opóźnień i najczęściej wymusza stosowanie dodatkowych buforów
 - niewszystkie etapy mogą być zawsze realizowane równoległe, z uwagi na możliwość wystąpienia konfliktów w dostępie do zasobów (np. fazy FI, FO, WR z poprzedniego przykładu mogą wymagać dostępu do pamięci)
 - wystąpienie przerwania, błędu strony lub naruszenie ochrony zaburza sekwencję przetwarzanych rozkazów i wymusza czyszczenie całego potoku
 - wystąpienie instrukcji rozgałęzienia warunkowego może prowadzić do unieważnienia rozkazów wewnątrz potoku

124

Przetwarzanie potokowe

- Sposoby postępowania w przypadku rozgałęzień:
 - pobieranie rozkazu docelowego z wyprzedzeniem (pobieranie rozkazu docelowego razem z rozkazem następującym po rozgałęzieniu)
 - zwielokrotnienie strumienia (jednoczesne przetwarzanie kilku ścieżek wykonawczych)
 - stosowanie bufora pętli (zapamiętywanie w niewielkiej, szybkiej pamięci ostatnio pobranych rozkazów i sprawdzanie czy cel rozgałęzienia znajduje się wewnątrz bufora)
 - opóźnianie rozgałęzienia (zmiana porządku wykonywanych rozkazów tak, aby rozgałęzienie występowało później)
 - przewidywanie rozgałęzień (predykcja wyniku rozgałęzienia w oparciu o kod operacji, zapamiętaną historię rozkazu lub tablicę historii rozgałęzień)

125

Przetwarzanie potokowe

- Najczęściej stosowanym obecnie rozwiązaniem jest technika przewidywania biegu programu (*Branch Prediction*) w oparciu o bufor historii rozgałęzień BTB (*Branch Target Buffer*)
- Tablica BTB jest szybką pamięcią asocjacyjną zawierającą od kilkuset do kilku tysięcy rekordów przechowujących: adres instrukcji rozgałęzienia, adres instrukcji docelowej oraz pewną liczbę bitów rejestrujących przebieg (historię) przetwarzania
- Efektywność przewidywania zależy od stosowanego algorytmu predykcji oraz wielkości bufora BTB i waha się od ok. 65% do prawie 90%

126

Przetwarzanie superskalarne

- Technika przetwarzania superskalarne polega na zastosowaniu kilku działających równolegle potoków przetwarzających dane skalarne, co pozwala na wykonywanie więcej niż jednego rozkazu w jednym cyklu zegara
- Podstawową trudnością w realizacji przetwarzania superskalarne jest wzajemna współzależność rozkazów wykonywanych w poszczególnych potokach, wynikająca m.in. z rozgałęzień oraz uzależnień i konfliktów w dostępie do zasobów
- Dla przykładu procesor Pentium wykorzystuje dwa potoki wykonawcze, oznaczone odpowiednio U i V

127

Przetwarzanie superskalarne

- Dzięki równoległemu wykonywaniu rozkazów, możliwe jest wykonywanie instrukcji całkowito-liczbowych w obu potokach w jednym cyklu rozkazowym
- Potoki skojarzone z obiema kolejkami rozkazów nie są równouprawnione:
 - potok U pozwala na wykonywanie dowolnej instrukcji z zestawu rozkazów procesorów 80x86
 - potok V umożliwia realizację jedynie tzw. instrukcji prostych (rozказы operujące na danych całkowitych oraz rozказы zamiany danych w rejestrach), które realizowane są całkowicie sprzętowo, bez udziału mikroprogramu wbudowanego w pamięć ROM procesora

128

Przetwarzanie superskalarne

- Rozkazy przeznaczone dla potoku U i V przesuują się synchronicznie przez poszczególne stopnie przetwarzania
- Jeżeli rozkaz w jednym z potoków przejdzie określony stopień przetwarzania przed zakończeniem tej samej fazy w drugim potoku, to wówczas oczekuje on na drugi rozkaz, tak aby oba rozkazy razem mogły przejść do kolejnej fazy przetwarzania
- Procesor jest w stanie z każdego potoku równocześnie odwoływać się do danych umieszczonych w wewnętrznej pamięci podręcznej

129

Przetwarzanie superskalarne

- Uzyskanie zrównoleglenia przetwarzania wymaga tzw. „parowania rozkazów” tj. łączenia ich w pary w taki sposób, aby mogły być wykonywane równolegle przez oba potoki
- Instrukcje podlegające łączeniu w pary można podzielić na trzy grupy:
 - grupę UV stanowią instrukcje, które połączone w pary mogą być wykonywane w obu potokach
 - grupę PU stanowią instrukcje wykonywane tylko w potoku U połączone w pary z odpowiednimi instrukcjami z potoku V
 - grupę PV stanowią instrukcje (wykonywane w obu potokach), które połączone w pary z instrukcjami ze zbioru UV i PU będą wykonywane w potoku V

130

Przetwarzanie superskalarne

- Próba wykonania pewnych par instrukcji w sposób równoległy może prowadzić do pojawienia się konfliktów lub uzależnień (które w konsekwencji utrudniają lub wręcz uniemożliwiają łączenie takich instrukcji w pary):
 - RAR (*Read After Read*) np. `add bx, ax`
`mov cx, ax`
 - RAW (*Read After Write*) np. `mov ax, bx`
`mov [mem], ax`
 - WAR (*Write After Read*) np. `mov bx, ax`
`add ax, cx`
 - WAW (*Write After Write*) np. `mov ax, [mem]`
`mov ax, bx`

131

Przetwarzanie superskalarne

- Uzależnienie typu RAR jest wynikiem konfliktu podczas próby jednoczesnego odczytu rejestru lecz można je stosunkowo łatwo usunąć poprzez zdublowanie portu odczytu rejestru
- Wynikiem uzależnienia typu RAW jest wystąpienie opóźnienia związanego z oczekiwaniem na wynik wcześniejszej operacji
- Uzależnienia typu WAR i WAW wynikają z konfliktu podczas próby jednoczesnego zapisu (lub odczytu i zapisu) do rejestru i można im zapobiec stosując technikę przemianowywania rejestrów (*Register Renaming*)

132

Przetwarzanie superskalarne

- Technika przemianowywanie rejestrów polega na wykorzystaniu puli dodatkowych rejestrów i ich dynamicznym mapowaniu do rejestrów dostępnych programowo
- Po zdekodowaniu rozkazu, którego wynik ma być docelowo umieszczony w rejestrze, następuje przydzielenie z puli rejestru, na którym będzie wykonywana operacja i zaktualizowanie tablicy zamienników RAT (*Register Allocation Table*)
- Ponieważ rejestr dostępny programowo może być mapowany na kilka rejestrów z puli, pozwala to na zrównoleglenie wykonywania rozkazów

133

Przetwarzanie superskalarne

- Dla przykładu sekwencja rozkazów (WAR):

mov bx, ax

add ax, cx

po podstawieniu:

ax -> R0

bx -> R1

cx -> R2

i przemianowaniu:

ax -> R3

zostaje przekształcona w sekwencję:

R1 := R0

R3 := R0 + R2

134

Dynamiczne wykonywanie instrukcji

- Dynamiczne wykonywanie instrukcji jest techniką pozwalającą na wykonywanie wejściowej sekwencji instrukcji w dowolnej kolejności (*out-of-order execution*) w celu uniknięcia konfliktów w dostępie do zasobów oraz powstawania opóźnień podczas potokowego przetwarzania instrukcji
- Dynamiczne wykonywanie instrukcji oparte jest na:
 - zaawansowanej predykcji rozgałęzień (*deep branch prediction*)
 - dynamicznej analizie przepływu danych (*dynamic data flow analysis*)
 - spekulatywnym wykonywaniu rozkazów (*speculative execution*)

135

Dynamiczne wykonywanie instrukcji

- Dynamiczna analiza przepływu danych polega na analizowaniu w czasie rzeczywistym sekwencji instrukcji przetwarzanych przez procesor w celu wykrycia i określenia zależności pozwalających na ich optymalne uszeregowanie w jednostkach wykonawczych
- Spekulatywne wykonywanie instrukcji polega na wykonywaniu instrukcji znajdujących się za jeszcze nierozstrzygniętą instrukcją warunkową, przy czym wyniki tychże instrukcji są zapamiętywane w specjalnym bloku rejestrów tymczasowych

136

Dynamiczne wykonywanie instrukcji

- Typowa sekwencja dynamicznego wykonania instrukcji przebiega następująco:
 - utworzenie przez układ MIS (*Microcode Instruction Sequencer*) sekwencji mikrooperacji do wykonania
 - uzupełnienie sekwencji mikrooperacji informacją o ich oryginalnej kolejności
 - przydzielenie do mikrooperacji rejestrów tymczasowych
 - pobranie mikrooperacji do wykonania przez układ przydziału zasobów RS (*Reservation Station*) w takiej kolejności, która zapewnia najbardziej optymalne ich wykonanie
 - odtworzenie przez układ RU (*Retirement Unit*) właściwej kolejności zrealizowanych mikrooperacji i przepisanie wyników z rejestrów tymczasowych do rejestrów dostępnych programowo

137

Pamięć podręczna

- Celem stosowania pamięci podręcznej (*Cache*) jest zwiększenie wydajności podsystemu pamięci wynikające z zasady lokalności, w myśl której:
 - zdecydowana większość kodu przetwarzana jest w sposób sekwencyjny (za wyjątkiem instrukcji skoków oraz wywołań i powrotów z procedur), co oznacza, że zajmuje on spójny i ciągły blok w pamięci
 - znaczna część obliczeń dotyczy przetwarzania zmiennych prostych, tablic i rekordów, które położone są w pamięci stosunkowo blisko siebie
- Cechą charakterystyczną pamięci podręcznej jest bardzo krótki czas dostępu, co pozwala na efektywne buforowanie wolniejszej pamięci operacyjnej RAM

138

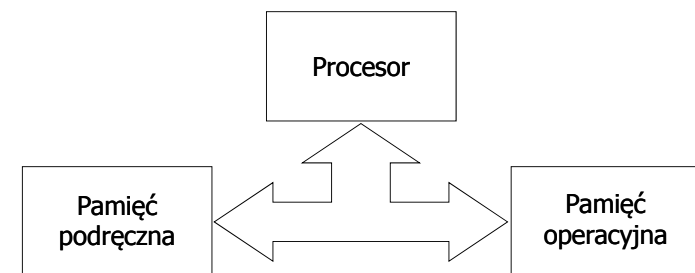
Pamięć podręczna

- Typowa pamięć podręczna jest systemem wielopoziomowym złożonym najczęściej z dwóch (L1+L2) lub trzech (L1+L2+L3) kaskadowo połączonych układów pamięci statycznych o coraz większym czasie dostępu i coraz większej pojemności
- Pamięć podręczna L1 ma zwykle pojemność od kilku do kilkudziesięciu KB i na ogół służy do rozdzielnego buforowania danych i instrukcji
- Rodzaje architektur systemów z pamięcią podręczną:
 - konwencjonalna (*Look-Aside*)
 - szeregową (*Look-Through*)
 - dwuszynową (*Backside*)

139

Pamięć podręczna

- Umieszczenie pamięci podręcznej w architekturze konwencjonalnej (*Look-Aside*):



140

Pamięć podręczna

- Pamięć podręczna typu *Look-Aside* dołączana jest do systemu przez magistralę systemową i pracuje z wyznaczoną przez nią częstotliwością taktowania
- Kontroler pamięci podręcznej monitoruje wszystkie odwołania procesora do pamięci, aby sprawdzić, czy żądana dana znajduje się w pamięci podręcznej
- W przypadku „trafienia” (*cache hit*) tj. znalezienia żądanej danej, procesor przerywa cykl magistrali i informuje podsystem sterujący pamięcią operacyjną, aby zignorował żądanie
- W przypadku „chybienia” (*cache miss*), cykle magistrali są wykonywane w normalny sposób

141

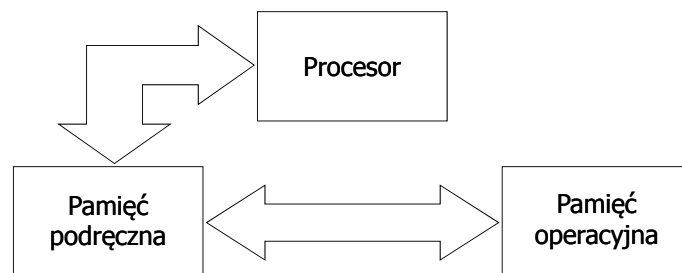
Pamięć podręczna

- Zalety stosowania pamięci typu *Look-Aside*:
 - niski koszt wynikający z prostej konstrukcji
 - polepszony czas dostępu do danych, które nie znajdują się w pamięci podręcznej
- Podstawową wadą stosowania pamięci typu *Look-Aside* jest brak izolowania magistrali procesora od magistrali systemowej, co prowadzi do:
 - niemożności zredukowania stopnia wykorzystania magistrali systemowej
 - braku możliwości współbieżnego wykonywania operacji (dostęp do pamięci podręcznej realizowany jest przez magistralę systemową)

142

Pamięć podręczna

- Umieszczenie pamięci podręcznej w architekturze szeregowej (*Look-Through*):



143

Pamięć podręczna

- Po zainicjowaniu cyklu odczytu pamięć typu *Look-Through* przeszukuje swój katalog w celu rozstrzygnięcia, czy zawiera kopię żądanej przez procesor danej
- W przypadku „trafienia” dana wysyłana jest do procesora z zerowymi cyklami oczekiwania magistrali, a magistrala systemowa pozostaje dostępna dla innych urządzeń (np. dla danych przesyłanych w trybie DMA)
- W przypadku „chybienia” następuje odczyt danej z pamięci operacyjnej (przez magistralę systemową)

144

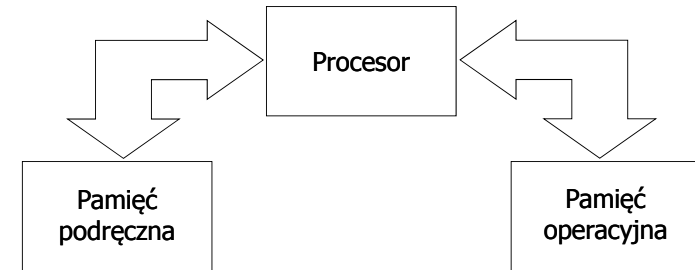
Pamięć podręczna

- Zalety stosowania pamięci typu *Look-Through*:
 - zredukowanie stopnia wykorzystania magistrali systemowej
 - możliwość równoczesnego (współbieżnego) wykonywania operacji na dwóch magistralach znajdujących się w jednym systemie
 - wykonywanie operacji zapisu w zerowych cyklach oczekiwania magistrali
 - możliwość taktowania magistrali lokalnej inną częstotliwością niż magistrali systemowej
- Wady stosowania pamięci typu *Look-Through*:
 - wyższy koszt wynikający z większej złożoności układu
 - wolniejszy niż w przypadku pamięci *Look-aside* odczyt danych, które nie znajdują się w pamięci podręcznej

145

Pamięć podręczna

- Umiejscowienie pamięci podręcznej w architekturze dwuszynowej (*Backside*):



146

Pamięć podręczna

- Pamięć podręczna typu *Backside* podłączona jest do procesora dedykowaną magistralą BSB (*Back Side Bus*), która jest całkowicie niezależna od magistrali pamięciowej FSB (*Front Side Bus*)
- Rozwiązanie to pozwala na swobodny wybór częstotliwości pracy pamięci podręcznej
- Podobnie jak w przypadku pamięci typu *Look-Through* obie magistrale mogą działać współbieżnie
- Pamięć typu *Backside* może być wkomponowana w strukturę samego procesora (Pentium Pro) lub umieszczona w jego pobliżu (Pentium III)

147

Pamięć podręczna

- Jeżeli dane w pamięci operacyjnej różnią się od odpowiadających im danych w pamięci podręcznej mamy do czynienia z utratą spójności pamięci podręcznej
- Przyczyną takiej sytuacji może być „trafienie” zapisu w pamięć podręczną lub zapisanie danych do pamięci operacyjnej przez urządzenie zewnętrzne
- Strategie zapisu zapewniające spójność pamięci podręcznej:
 - Zapis jednoczesny (*Write-Through*)
 - Zapis buforowany (*Buffered Write-Through*)
 - Zapis opóźniony (*Write-Back*)

148

Pamięć podręczna

- W strategii *Write-Through* każdy zapis do pamięci podręcznej powoduje natychmiastowe wykonanie cyklu zapisu do pamięci operacyjnej, co gwarantuje stałe utrzymywanie aktualnych danych w pamięci operacyjnej
- Operacja zapisu w pamięci operacyjnej realizowana jest niezależnie od tego, czy nastąpiło „trafienie” czy „chybienie” z zapisem do pamięci podręcznej
- Implementacja tej metody jest bardzo efektywna i prosta, cechuje się jednak małą wydajnością, gdyż każda operacja zapisu powoduje odwołanie się do wolnej pamięci operacyjnej

149

Pamięć podręczna

- Strategia *Buffered Write-Through* polega na zapamiętywaniu przez kontroler pamięci podręcznej realizowanych operacji zapisu w specjalnym buforze, co umożliwia ich późniejsze zrealizowanie w pamięci operacyjnej bez wpływu na pracę procesora
- Podstawową zaletą tej metody jest brak cykli oczekiwania magistrali zarówno w przypadku „trafienia” jak i „chybienia” zapisu
- Ten rodzaj zapisu powoduje, że inne urządzenie wymagające dostępu do magistrali nie może przeprowadzać na niej żadnych operacji do chwili zakończenia operacji zapisu

150

Pamięć podręczna

- Strategia *Write-Back* zwiększa całkowitą wydajność systemu, gdyż dokonuje aktualizacji pamięci operacyjnej tylko wtedy, gdy jest to konieczne
- Aktualizacja pamięci operacyjnej ma miejsce w przypadku gdy:
 - urządzenie zewnętrzne inicjuje operację odczytu lub zapisu tych słów pamięci, które zawierają „przeterminowane” dane
 - słowo pamięci podręcznej, które zawiera zmodyfikowane dane, ma zostać nadpisane danymi z pamięci operacyjnej
- Słowa pamięci podręcznej są oznaczane w lokalnym katalogu w momencie kiedy są aktualizowane przez procesor

151

Pamięć podręczna

- W momencie zapisu lub odczytu do lub z pamięci operacyjnej, podsystem pamięci podręcznej musi monitorować magistralę systemową aby sprawdzić, czy są to odwołania do słów pamięci oznaczonych jako zmodyfikowane
- Zachowanie spójności pamięci podręcznej (zwłaszcza w systemach wieloprocesorowych) oparte jest na protokole MESI, który określa stan każdego wiersza pamięci podręcznej danych jako:
 - zmodyfikowany (*Modified*)
 - wyłączony (*Exclusive*)
 - współdzielony (*Shared*)
 - nieważny (*Invalid*)

152

Pamięć podręczna

- Pamięć podręczna zorganizowana jest w wiersze, których szerokość odpowiada najczęściej szerokości magistrali danych
- Pamięć operacyjna dzielona jest na strony o rozmiarze równym liczbie wierszy pamięci podręcznej
- Informacja o tym, które wiersze pamięci operacyjnej są odwzorowane w pamięci podręcznej znajduje się w katalogu pamięci podręcznej TAG-RAM
- Strategie odwzorowywania pamięci operacyjnej:
 - Odwzorowanie bezpośrednie (*Direct Mapped*)
 - Odwzorowanie skojarzeniowe (*Fully Associative*)
 - Odwzorowanie sekcyjno-skojarzeniowe (*Set Associative*)

153

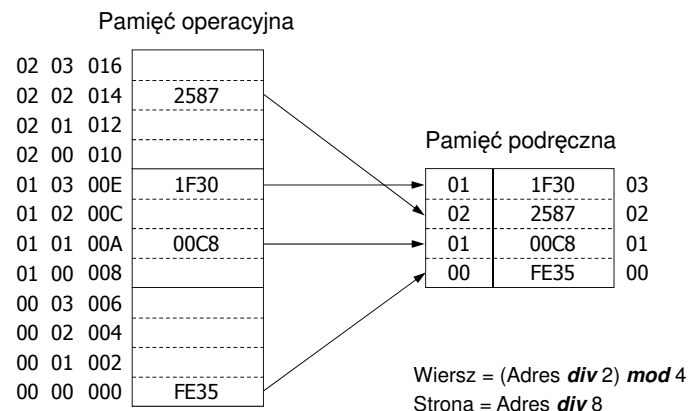
Pamięć podręczna

- Odwzorowanie bezpośrednie polega na takim sposobie odwzorowywania pamięci operacyjnej na pamięć podręczną, że n-ty wiersz każdej strony pamięci operacyjnej może zostać odwzorowany tylko w n-tym wierszu pamięci podręcznej
- Zaletą tej metody jest duża szybkość wyszukiwania informacji (tylko jedna operacja porównania) oraz prosta konstrukcja układu
- Ponieważ każdy wiersz pamięci operacyjnej ma stałą lokalizację w pamięci podręcznej, prowadzi to do ciągłej wymiany zawartości pamięci podręcznej, a tym samym do obniżenia wydajności systemu

154

Pamięć podręczna

- Przykład odwzorowania bezpośredniego:



155

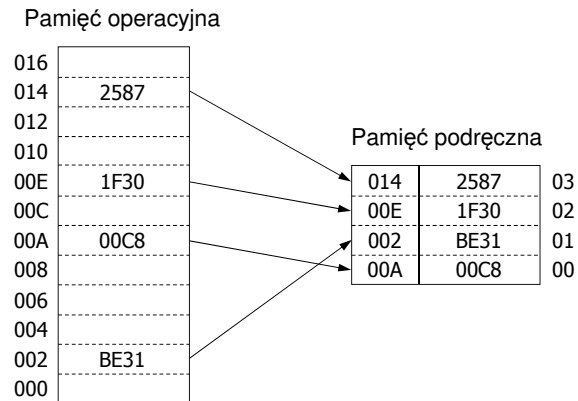
Pamięć podręczna

- Odwzorowanie skojarzeniowe pozwala na ładowanie każdego wiersza pamięci operacyjnej do dowolnego wiersza pamięci podręcznej, co eliminuje największą wadę odwzorowania bezpośredniego
- W tego typu odwzorowaniu nie występuje podział pamięci operacyjnej na strony – wszelkie operacje wykonywane są na wierszach identyfikowanych przez adresy
- Podstawową wadą tej metody jest konieczność przeszukania całego katalogu TAG-RAM aby zbadać, czy żądany wiersz pamięci operacyjnej został odwzorowany w pamięci podręcznej

156

Pamięć podręczna

- Przykład odwzorowania skojarzeniowego:



157

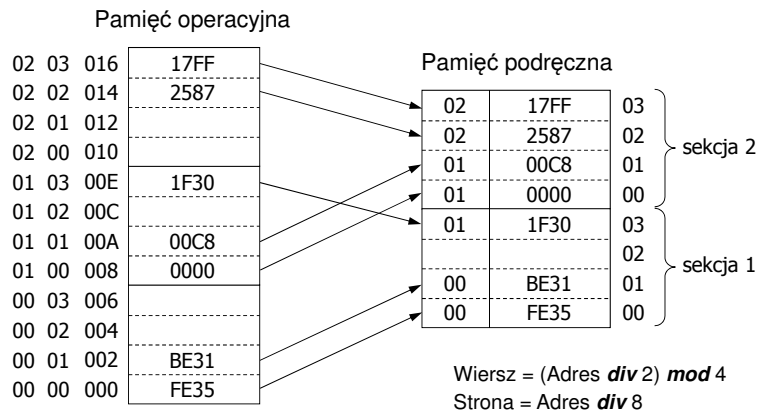
Pamięć podręczna

- Odwzorowanie sekcyjno-skojarzeniowe jest metodą, która łączy zalety odwzorowania bezpośredniego i skojarzeniowego
- Cechą charakterystyczną tego typu podejścia jest podział pamięci podręcznej na 2, 4 lub 8 sekcji (*2-Way-Set, 4-Way-Set, 8-Way-Set*) o rozmiarze równym wielkości strony pamięci operacyjnej
- Ponieważ pamięć operacyjna odwzorowywana jest na każdą sekcję w sposób bezpośredni, wyszukanie odpowiedniej informacji w pamięci podręcznej wymaga przeprowadzenia co najwyżej 8 porównań (jedno porównanie na sekcję)

158

Pamięć podręczna

- Przykład dwudroźnego odwzorowania sekcyjno-skojarzeniowego (*2-Way-Set*):



159

Pamięć podręczna

- Wspólną cechą odwzorowania skojarzeniowego i sekcyjno-skojarzeniowego jest konieczność wyboru tego wiersza pamięci podręcznej, który zostanie zastąpiony przez nowo ładowany wiersz z pamięci operacyjnej
- W tym celu kontroler pamięci podręcznej implementuje jeden z następujących algorytmów:
 - FIFO (*First In First Out*), który do zastąpienia wybiera wiersz najdłużej pozostający w pamięci podręcznej
 - LFU (*Least Frequently Used*), który do zastąpienia wybiera najrzadziej używany wiersz pamięci podręcznej
 - LRU (*Least Recently Used*), który do zastąpienia wybiera najmniej ostatnio używany wiersz pamięci podręcznej

160

Przegląd architektury IA-32

- Architektura IA-32 jest 32-bitową architekturą procesorów firmy Intel, będącą rozwinięciem architektury procesorów rodziny 80x86
- Każdy procesor z rodziny IA-32 cechuje zachowanie wstecznej zgodności programowej (nowszy typ procesora może wykonywać kod utworzony dla starszego typu procesora)
- Typowymi przedstawicielami linii produktów opartej na architekturze IA-32 są procesory, które wyewoluowały z procesorów Pentium II, takie jak: Pentium III, Pentium 4, Pentium 4 HT (*Hyper Threading*), Intel Xeon, Pentium M

161

Przegląd architektury IA-32

- Ewolucja procesorów linii IA-32:
 - Intel 8086 (1978):
 - liczba tranzystorów: 29 tys.
 - częstotliwość taktowania: 8 MHz
 - magistrala danych: 16 bitów
 - przestrzeń adresowa: 1 MB
 - podręczna kolejka rozkazów
 - Intel 80286 (1982):
 - liczba tranzystorów: 134 tys.
 - częstotliwość taktowania: 12.5 MHz
 - przestrzeń adresowa: 16 MB
 - tryb chroniony (*protected mode*)

162

Przegląd architektury IA-32

- Intel 80386 (1985):
 - liczba tranzystorów: 275 tys.
 - częstotliwość taktowania: 20 MHz
 - magistrala danych: 32 bity
 - przestrzeń adresowa: 4 GB
 - tryb wirtualny procesora 8086 (*virtual 8086 mode*)
 - jednostka stronicowania pamięci (*paging unit*)
- Intel 80486 (1989):
 - liczba tranzystorów: 1.2 mln
 - częstotliwość taktowania: 25 MHz
 - pamięć podręczna: 8 KB
 - jednostka zmiennoprzecinkowa FPU

163

Przegląd architektury IA-32

- Intel Pentium (1993):
 - liczba tranzystorów: 3.1 mln
 - częstotliwość taktowania: 60 MHz
 - magistrala danych: 64 bity (wewn. 128/256 bitów)
 - pamięć podręczna: 8 KB kod + 8 KB dane (MESI)
 - dwa potoki wykonawcze całkowitoliczbowe
 - wsparcie dla systemów dwuprocesorowych
 - technologia MMX
- Intel Pentium Pro (1995):
 - liczba tranzystorów: 5.5 mln
 - częstotliwość taktowania: 200 MHz
 - przestrzeń adresowa: 64 GB
 - pamięć podręczna: 16 KB (L1), 256/512 KB (L2)

164

Przegląd architektury IA-32

- mikro-architektura superskalarna P6 (3 instrukcje/takt zegara; 5 jednostek wykonawczych: 2 x Integer, 2 x FPU, 1 x MIU)
- Intel Pentium II (1997):
 - liczba tranzystorów: 7 mln
 - częstotliwość taktowania: 266 MHz
 - pamięć podręczna: 32 KB (L1), 256/512 KB (L2)
 - technologia MMX
- Intel Pentium III (1999):
 - liczba tranzystorów: 8.2÷28 mln
 - częstotliwość taktowania: 500 MHz
 - pamięć podręczna: 32 KB (L1), 256/512/1024 KB (L2)
 - technologia SSE (*Streaming SIMD Extensions*)

165

Przegląd architektury IA-32

- Intel Pentium 4 (2000):
 - liczba tranzystorów: 42 mln
 - częstotliwość taktowania: 1.5 GHz
 - pamięć podręczna: 12.000 μ ops (ETC), 8 KB (L1), 256 KB (L2)
 - Intel NetBurst Micro-Architecture
 - technologia SSE2
- Intel Xeon (2002):
 - liczba tranzystorów: 55 mln
 - częstotliwość taktowania: 2.2 GHz
 - pamięć podręczna: 12.000 μ ops (ETC), 8 KB (L1), 512 KB (L2)
 - technologia HT (*Hyper Threading*)

166

Przegląd architektury IA-32

- Intel Pentium 4 HT (2002):
 - liczba tranzystorów: 55 mln
 - częstotliwość taktowania: 3.06 GHz
 - pamięć podręczna: 12.000 μ ops (ETC), 8 KB (L1), 512 KB (L2)
 - technologia HT
- Intel Pentium M (2003):
 - liczba tranzystorów: 77 mln
 - częstotliwość taktowania: 1.6 GHz
 - pamięć podręczna: 32 KB kod + 32 KB dane (L1), 1 MB (L2)
 - technologia SSE2
 - technologia APM (*Advanced Power Management*)

167

Przegląd architektury IA-32

- Podstawowe środowisko wykonawcze procesorów IA-32 tworzy:
 - Zestaw rejestrów podstawowych, który zawiera:
 - osiem 32-bitowych rejestrów ogólnego przeznaczenia (EAX, EBX, ECX, EDX, EBP, ESI, EDI, ESP), których młodsze 16-bitów jest odwzorowywanych na rejestry 16-bitowe procesorów 8086 i 80286 (AX, BX, CX, DX, BP, SI, DI, SP); ponadto rejestry AX, BX, CX i DX są dodatkowo mapowane na pary rejestrów 8-bitowych (AH, AL, BH, BL, CH, CL, DH, DL)
 - sześć 16-bitowych rejestrów segmentowych (CS, DS, ES, FS, GS, SS)
 - jeden 32-bitowy rejestr statusowy (EFLAGS)
 - jeden 32-bitowy rejestr wskaźnika rozkazów (EIP)

168

Przegląd architektury IA-32

- Zestaw rejestrów jednostki zmiennie-przecinkowej (FPU), który zawiera m.in.:
 - osiem 80-bitowych rejestrów danych zorganizowanych w postaci stosu: $ST(0) \div ST(7)$
 - jeden 16-bitowy rejestr sterujący (FPU Control Word)
 - jeden 16-bitowy rejestr stanu (FPU Status Word)
 - jeden 16-bitowy rejestr słowa wyróżników (FPU Tag Word)
 - jeden 48-bitowy rejestr wskaźnika rozkazu (FPU Instruction Pointer)
 - jeden 48-bitowy rejestr wskaźnika argumentu (FPU Operand Pointer)
- Zestaw ośmiu 64-bitowych rejestrów technologii MMX ($MM0 \div MM7$), współdzielonych z rejestrami danych FPU

169

Przegląd architektury IA-32

- Zestaw rejestrów technologii SSE i SSE2, który zawiera:
 - osiem 128-bitowych rejestrów danych ($XMM0 \div XMM7$)
 - jeden 32-bitowy rejestr kontrolno-statusowy (MXCSR)
- Zestaw rejestrów pomocniczych, który zawiera m.in.:
 - rejestry sterujące pracą procesora ($CR0 \div CR4$)
 - rejestry podsystemu zarządzania pamięcią (GDTR, LDTR, IDTR, TR)
 - rejestry uruchomieniowe ($DR0 \div DR7$)
 - rejestry zaawansowanego kontrolera przerwań APIC (ISR, TMR, IRR, ESR)
 - rejestry wspomagające testowanie procesora ($TR3 \div TR7$)

170

IA-32: Tryby pracy procesora

- Procesor może pracować w jednym z następujących trybów pracy:
 - trybie rzeczywistym (*Real Mode, Real-address Mode*)
 - trybie chronionym (*Protected Mode*)
 - trybie zarządzania SMM (*System Management Mode*)
- Procesor działając w trybie rzeczywistym (np. po włączeniu zasilania czy aktywacji sygnału Reset) odpowiada funkcjonalnie procesorowi 8086
- W tym trybie pracy 20-bitowy adres fizyczny tworzony jest na podstawie 16-bitowego adresu segmentu i 16-bitowego przesunięcia w segmencie ($Address = 16 * Segment + Offset$)

171

IA-32: Tryby pracy procesora

- Tryb chroniony jest podstawowym trybem pracy procesora, zapewniającym ochronę niezależnych zadań wykonywanych współbieżnie w środowisku wielozadaniowym (*multi-tasking environment*)
- Jedną z możliwości oferowanych przez tryb chroniony jest udostępnianie dla każdego z wykonywanych zadań niezależnego środowiska procesora 8086 (*Virtual-8086 Mode*)
- Tryb wirtualny procesora 8086 pozwala na uruchamianie programów pisanych dla trybu rzeczywistego w środowisku wielozadaniowym (chronionym)

172

IA-32: Tryby pracy procesora

- Tryb zarządzania SMM pozwala na realizację funkcji systemowych wysokiego poziomu, takich jak zarządzanie energią czy ochrona zasobów przed nieuprawnionym dostępem
- Funkcjonowanie procesora w trybie zarządzania SMM kontrolowane jest wyłącznie przez oprogramowanie systemowe zawarte w pamięci stałej w sposób transparentny dla innych programów
- Typowym zastosowaniem tego trybu pracy jest implementacja systemu zarządzania zasilaniem, który może wprowadzać procesor i wybrane urządzenia zewnętrzne w stan „uśpienia” w okresie bezczynności

173

IA-32: Zarządzanie pamięcią

- Procesory linii IA-32 mogą adresować maksymalnie do 64 GB (2^{36}) pamięci fizycznej (najmniejszą adresowalną jednostką pamięci fizycznej jest bajt)
- Mechanizm zarządzania pamięcią wirtualną procesora (*virtual memory management mechanism*) odwzorowuje przestrzeń adresową wykonywanego programu w przestrzeń pamięci fizycznej wykorzystując jeden z trzech modeli pamięci:
 - model płaski (*flat model*)
 - model segmentowy (*segment model*)
 - model trybu rzeczywistego procesora 8086 (*real-address model*)

174

IA-32: Zarządzanie pamięcią

- W modelu płaskim pamięć widoczna dla programu jest ciągła i liniowa (kod, dane oraz stos programu rezydują w jednej przestrzeni adresowej)
- Maksymalny rozmiar udostępnianej programowi przestrzeni liniowej wynosi 4 GB (2^{32})
- W modelu segmentowym pamięć widoczna jest dla programu jako grupa niezależnych obszarów pamięci zwanych segmentami (najczęściej kod, dane oraz stos programu rezydują w niezależnych segmentach)
- W tym modelu program może zaadresować maksymalnie 16383 segmenty różnego typu i o różnym rozmiarze (nie większym jednak niż 4 GB)

175

IA-32: Zarządzanie pamięcią

- Adres logiczny (*far pointer*) w modelu segmentowym składa się z dwóch elementów: 16-bitowego selektora segmentu (*segment selector*) i 32-bitowego przesunięcia w segmencie (*offset*)
- Selektor segmentu stanowi indeks do globalnej (GDT) lub lokalnej (LDT) tablicy deskryptorów opisujących poszczególne segmenty
- Deskryptor segmentu (*segment descriptor*) jest strukturą przechowującą jego typ, granularność, rozmiar, lokalizację w pamięci oraz prawa dostępu
- Adresacja wewnątrz pojedynczego segmentu jest liniowa

176

IA-32: Zarządzanie pamięcią

- Selektory segmentów przechowywane są w rejestrach segmentowych (CS, DS, ES, FS, GS i SS)
- Przesunięcia wewnątrz segmentów przechowywane są najczęściej w rejestrach ogólnego przeznaczenia (np. EBX i ESI dla DS, EDI dla ES, EBP i ESP dla SS)
- Zaletą stosowania modelu segmentowego jest możliwość poprawienia niezawodności programu (systemu operacyjnego) dzięki:
 - odseparowaniu od siebie programów wykonywanych wspólnie w środowisku wielozadaniowym
 - rozdzieleniu logicznie niezależnych „fragmentów” programu (np. stosu od danych)

177

IA-32: Zarządzanie pamięcią

- Model trybu rzeczywistego odwzorowuje mechanizm adresowania pamięci procesora 8086
- Tryb ten wykorzystuje specjalną implementację modelu segmentowego polegającą na podziale pamięci na nakładające się wzajemnie i przesunięte o 16 B segmenty o stałym rozmiarze 64 KB
- Maksymalny rozmiar udostępnianej programowi przestrzeni adresowej wynosi 1 MB (2^{20})
- Fizyczny adres w pamięci wyznaczany jest na podstawie 16-bitowego adresu segmentu (*segment*) i 16-bitowego przesunięcia w segmencie (*offset*)

178

IA-32: Zarządzanie pamięcią

- Począwszy od procesora 80386 odwzorowanie adresu logicznego na adres fizyczny (w modelach płaskim i segmentowym) może być realizowane z wykorzystaniem mechanizmu stronicowania pamięci
- Jednostka stronicowania (*Paging Unit*) dzieli liniową przestrzeń adresową na strony (*pages*) o stałym rozmiarze (4 KB, 2 MB lub 4 MB), które są następnie odwzorowywane w pamięci wirtualnej złożonej z pamięci operacyjnej i pamięci dyskowej
- Maksymalny rozmiar adresowalnej w ten sposób pamięci wirtualnej wynosi 64 TB

179

IA-32: Zarządzanie pamięcią

- W przypadku strony o rozmiarze 4 KB interpretacja adresu liniowego jest następująca:
 - Najstarsze 10 bitów adresu stanowi indeks w katalogu stron (*Page Directory*) przechowującym wskaźniki do tablic stron; adres fizyczny katalogu stron znajduje się w rejestrze CR3
 - Kolejnych 10 bitów adresu stanowi indeks w tablicy stron (*Page Table*) przechowującej fizyczny adres początku strony
 - Pozostałe 12 bitów adresu traktowane jest jako przesunięcie wewnątrz wybranej strony
- W przypadku strony o rozmiarze 4 MB katalog stron (indeksowany 10 bitami adresu) przechowuje adres fizyczny początku strony, pozostałe 22 bity adresu liniowego stanowi przesunięcie wewnątrz strony

180

IA-32: Zarządzanie pamięcią

- Jeżeli wyznaczona w powyższy sposób strona nie znajduje się w pamięci operacyjnej, to jej adres zostaje zapisany w rejestrze CR2 i generowany jest wyjątek błędu strony (*page fault*)
- Zadanie wczytania brakującej strony z dysku spoczywa na systemie operacyjnym (procedurze obsługi tego wyjątku)
- W celu zwiększenia efektywności mechanizmu stronicowania procesor wyposażony jest w pamięci podręczne TLB (*Translation Lookaside Buffer*) przechowującą zawartość ostatnio używanych (maksymalnie 128) tablic stron

181

IA-32: Technologia MMX

- Technologia MMX (*MultiMedia eXtension*) ma celu zwiększenie efektywności wykonywania programów (głównie multimedialnych), które cechują się powtarzaniem takich samych, prostych operacji na długich ciągach krótkich danych
- Idea leżąca u podstaw technologii MMX stanowi krok w stronę przetwarzania równoległego SIMD (*Single Instruction Multiple Data*) – zamiast kolejnego przetwarzania pojedynczych jednostek informacji, można te same dane przetworzyć szybciej, pobierając i wykonując rozkazy na kilku porcjach informacji na raz

182

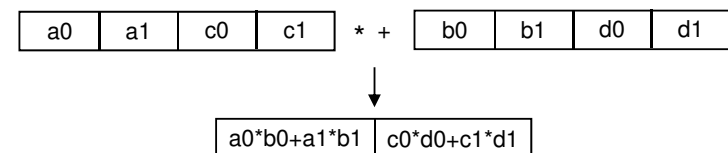
IA-32: Technologia MMX

- Procesor implementujący technologię MMX wyposażony jest w osiem 64-bitowych rejestrów: MM0÷MM7, pozwalających na przechowywanie nowych typów danych (tzw. danych spakowanych):
 - Packet Bytes (8 x 8 bitów)
 - Packed Words (4 x 16 bitów)
 - Packet Doublewords (2 x 32 bity)
 - Quadword (1 x 64 bity)
- Jednoczesne wykonywanie operacji na danych spakowanych umożliwia zestaw dodatkowych 57 rozkazów typu SIMD wykorzystujących tzw. arytmetykę nasycenia

183

IA-32: Technologia MMX

- W arytmetyce nasycenia nie występuje sygnalizacja wystąpienia przekroczenia zakresu (w wyniku pakowania lub wykonywania operacji) lecz zastąpienie wyniku wartością maksymalną lub minimalną dla danego typu danych
- Dla przykładu obliczanie dwóch iloczynów skalarnych wektorów może zostać zastąpione pojedynczym rozkazem PMADDWD:



184

IA-32: Technologia SSE (SSE2, ...)

- Technologia SSE (*Streaming SIMD Extensions*) rozszerza ideę jednoczesnego przetwarzania grupy danych pojedynczym rozkazem (znaną z technologii MMX) na dane w formacie zmiennoprzecinkowym
- Procesor implementujący technologię SSE (SSE 2) wyposażony jest w osiem 128-bitowych rejestrów XMM0÷XMM7 oraz rejestr kontrolno-sterujący MXCSR
- Zaletą takiego rozwiązania jest niezależnienie się od rejestrów zmiennoprzecinkowych FPU
- Największą wadą tej technologii jest brak zgodności z wcześniejszymi typami procesorów

185

IA-32: Technologia SSE (SSE2, ...)

- Lista rozkazów SSE zawiera 72 rozkazy obejmujące:
 - rozkazy przetwarzające grupy danych zmiennoprzecinkowych (*SIMD Floating Point*)
 - rozkazy przetwarzające pojedyncze dane zmiennoprzecinkowe (*Single Scalar*)
 - rozkazy sterujące przepływem danych (*Memory Streaming Instructions*)
 - rozkazy rozszerzające technologię MMX (*SIMD-INT*)
 - rozkazy sterujące (*Control*)
- Typowym zastosowaniem technologii SSE jest obróbka grafiki 3D oraz pełnoekranowa kompresja (dekompresja) MPEG-2 w czasie rzeczywistym

186

Magistrale systemowe

- Magistrala jest zbiorem pewnych linii sygnałowych umożliwiającym takie połączenie współpracujących elementów, aby każdy z nich mógł komunikować się ze wszystkimi pozostałymi
- Podstawowe cechy magistral:
 - Rodzaj (specjalistyczna, multipleksowana)
 - Metoda arbitrażu (centralny, rozproszony)
 - Koordynacja czasowa (synchroniczna, asynchroniczna)
 - Szerokość magistrali (adres, dane)
 - Rodzaj transferu danych (zapis, odczyt, odczyt-modyfikacja-zapis, odczyt-po-zapisie, blokowy)

187

Magistrale systemowe

- Magistrala systemowa jest komponentem zapewniającym komunikację między podstawowymi podzespołami systemu komputerowego (procesor, pamięć, układy wejścia-wyjścia)
- W stosowanych obecnie magistralach systemowych możemy wyróżnić trzy grupy linii sygnałowych:
 - linie adresowe (tworzące tzw. magistralę adresową)
 - linie danych (tworzące tzw. magistralę danych)
 - linie sterujące (tworzące tzw. magistralę sterującą)
- Typowym rozwiązaniem zwiększającym wydajność systemu komputerowego jest stosowanie struktur wielomagistralowych o określonej hierarchii

188

Magistrale systemowe

- W zależności od przeznaczenia można wyróżnić wiele rodzajów magistral:
 - magistrala pamięciowa
 - magistrala międzyukładowa (wewnętrzna)
 - magistrala rozszerzająca (zewnętrzna)
 - magistrala lokalna
- Zadaniem magistrali pamięciowej (*memory bus*) jest przesyłanie informacji między procesorem a pamięcią operacyjną i/lub pamięcią podręczną
- Magistrala międzyukładowa (*interconnect bus*) służy do wymiany informacji między poszczególnymi układami otoczenia procesora (*chipsets*)

189

Magistrale systemowe

- Magistrala rozszerzająca (*expansion bus*) służy do wymiany informacji między urządzeniami wejścia-wyjścia a resztą systemu komputerowego
- Najczęściej konstrukcja tego typu magistrali musi spełniać warunki zgodności z wcześniejszymi rozwiązaniami i zapewnić poprawną współpracę z szerokim zakresem urządzeń peryferyjnych
- Magistrala lokalna (*local bus*) jest odmianą magistrali rozszerzającej przeznaczoną do obsługi urządzeń wymagających bardzo szybkiej transmisji danych (np. dysków twardych, kart graficznych, kontrolerów SCSI)

190

Magistrale systemowe

- Cechą charakterystyczną magistrali lokalnej jest bezpośrednie korzystanie z sygnałów sterujących procesora oraz praca z prędkością wyznaczoną przez zegar systemowy
- Przegląd standardów magistral rozszerzających:
 - ISA (*Industry Standard Architecture*)
 - standard magistrali oparty na rozwiązaniu z IBM PC/AT
 - przepustowość 8.33 MB/s (taktowanie 8.33 MHz)
 - 16-bitowa szyna danych
 - kontroler DMA oparty na dwóch połączonych kaskadowo układach 8237A
 - kontroler przerw sprzętowych oparty na dwóch połączonych kaskadowo układach 8259A

191

Magistrale systemowe

- MCA (*Micro Channel Architecture*)
 - brak zgodności ze standardem ISA
 - 32-bitowa szyna danych
 - przepustowość 20 MB/s (taktowanie 10 MHz)
 - dostęp do pamięci taktowany częstotliwością zegara procesora
 - wsparcie dla systemów wieloprocesorowych (centralny procesor na płycie głównej tzw. *Host CPU* może być wspomagany przez 16 procesorów umieszczonych na kartach rozszerzających; dla potrzeb komunikacji między procesorami, uwzględnienia ich priorytetów, przydziału odcinków czasowych dla dysponowania magistralami, dedykowano specjalną 4-bitową szynę sterującą)

192

Magistrale systemowe

- kontroler DMA obsługujący osiem 32-bitowych kanałów (brak emulacji 8- i 16-bitowych trybów pracy układu 8237A; wszystkie kanały mogą być jednocześnie aktywne)
- kontroler przerw sprzętowych wyzwalanych poziomem pozwalający dzięki przydziałowi wielu urządzeń do jednego kanału na obsługę do 255 urządzeń
- kontroler przerw sprzętowych obsługujący do 255 urządzeń (przydział jednego kanału wielu urządzeniom; przerwanie wyzwalane poziomem)
- EISA (*Extended Industry Standard Architecture*)
 - zgodność ze standardem ISA
 - 32-bitowa szyna danych
 - przepustowość 33 MB/s (taktowanie 8.33 MHz)

193

Magistrale systemowe

- dostęp do pamięci realizowany jest z pełną częstotliwością zegara procesora
- możliwość symulowania przez kontroler magistrali starego, 16-bitowego trybu pracy
- wsparcie dla systemów wieloprocesorowych (kontroler magistrali umożliwia hierarchiczno-priorytetowy sposób dostępu wielu procesorów do magistrali systemowej)
- kontroler DMA obsługujący siedem 8-, 16- lub 32-bitowych kanałów DMA (rotacyjny system przydziału kanałów; pełne 32-bitowe adresowanie umożliwiające transfer bloków o wielkości do 4 GB)
- kontroler przerw sprzętowych obsługujący przerwanie wyzwalane poziomem

194

Magistrale systemowe

- VESA Local Bus (*Video Electronics Standards Association*)
 - 32-bitowa szyna danych
 - przepustowość do 120 MB/s
 - praktycznie obsługa tylko dwóch urządzeń: karty graficznej i kontrolera dysków
 - silna zależność od sprzętu (magistrala wykorzystuje część niebuforowanych sygnałów procesora 80486)
- PCI (*Peripheral Component Interconnect*)
 - 32/64-bitowa szyna danych
 - przepustowość 133/266/533 MB/s (taktowanie 33/66 MHz)
 - odczyt i zapis w trybie pakietowym (*burst*)
 - obsługa do 32 urządzeń (wielofunkcyjnych)

195

Magistrale systemowe

- możliwość współpracy do 256 magistral PCI w jednym systemie komputerowym (*PCI-to-PCI bridges*)
- możliwość obsługi magistrali ISA (*PCI-to-ISA bridge*)
- niezależność od typu procesora (komunikacja z procesorem i pamięcią realizowana jest za pośrednictwem specjalnego układu - *PCI bridge*)
- wbudowany mechanizm autokonfiguracji
- urządzenia dołączane do magistrali mogą należeć do jednej z dwóch kategorii: urządzeń inicjujących transmisję – inicjatorów (*initiators*), które zarządzają magistralą oraz urządzeń podporządkowanych (*slaves*) potrafiących jedynie transmitować dane; transmisja danych może odbywać się między dwoma inicjatorami lub inicjatorem a urządzeniem podporządkowanym

196

Magistrale systemowe

- Magistrala umożliwia współdzielenie każdej z czterech linii zgłoszenia przerwania (INTA#÷INTD#) przez kilka urządzeń, lecz nie określa sposobu ich podłączenia do kontrolera przerwania systemu
- urządzenia PCI mogą pracować z napięciem 5V lub 3.3V
- AGP (*Accelerated Graphics Port*)
 - magistrala AGP dedykowana do obsługi układu graficznego jest rozszerzeniem standardu PCI 2.1
 - magistrala AGP może pracować w jednym z czterech trybów:
 - 1x (częstotliwość 66MHz, przepustowość 266 MB/s)
 - 2x (wykorzystanie obu zbroczy sygnału zegarowego 66MHz, przepustowość 532 MB/s, obniżenie napięcia z 5V do 3.3V)

197

Magistrale systemowe

- 4x (dwie transakcje na jeden cykl zegara, przepustowość 1064 MB/s, obniżenie napięcia do 1.5V) – specyfikacja AGP 2.0
- 8x (zmiany w protokole transmisji, przepustowość 2.1 GB/s, obniżenie napięcia do 0.8V, zgodność z trybem 4x) – specyfikacja AGP 3.0
- Aby wykorzystać pełne pasmo przenoszenia 32-bitowej magistrali AGP do przesyłania danych kontroler graficzny może opcjonalnie wykorzystać pomocniczą, 8-bitową magistralę SBA (*Sideband Addressing*) do niezależnego przesyłania adresów i rozkazów
- PCI Express (PCIe, PCI-E)
 - magistrala połączeniowa typu punkt-punkt (kontroler – urządzenie PCI-E, najczęściej karta graficzna)

198

Magistrale systemowe

- magistrala PCI-E wykorzystuje od 1 do 32 linii sygnałowych determinujących jej przepustowość (500 MB/s na jedną linię w trybie *full-duplex* v2.0) oraz rozmiar złącza połączeniowego
- przepustowość magistrali PCI-E w zależności od jej wariantu wynosi od 250 MB/s (x1) do 16 GB/s (x16)
- maksymalny pobór mocy urządzenia podłączonego do magistrali wynosi: 75 W (v1.0), 150 W (v2.0) lub 300 W (v3.0)

199