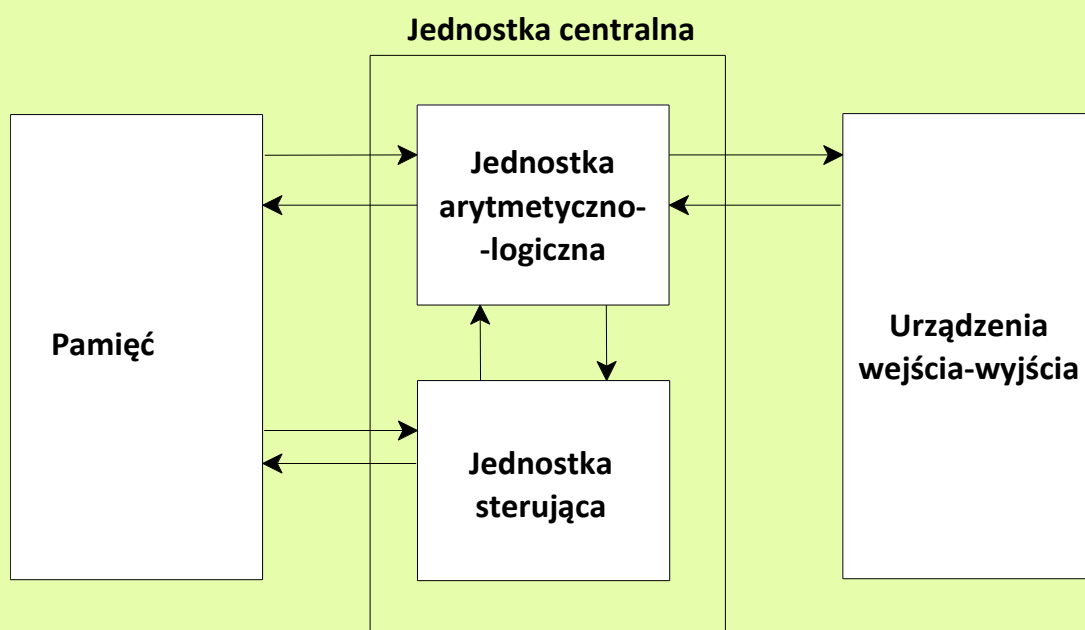


# Systemy komputerowe

dr Zbigniew Bonikowski  
Instytut Informatyki

## Architektura von Neumanna



# Pamięć komputera

---

- Pamięć odgrywa bardzo ważną rolę w komputerach opartych na modelu von Neumanna.
- Pamięć służy do przechowywania programów oraz danych wykorzystywanych przez te programy.

3

## Cechy pamięci - położenie

---

- Z uwagi na lokalizację dzielimy pamięć na:
  - pamięć wewnętrzną - pamięć umieszczona na płycie głównej i w procesorze (np. pamięć operacyjna),
  - pamięć zewnętrzną – pamięć umieszczoną na osobnym urządzeniu i dostępną dla procesora poprzez sterowniki wejścia-wyjścia (np. twarde dyski).

4

## Cechy pamięci - pojemność

- Pojemność pamięci określa ilość informacji, jaką można w niej przechować.
- Pojemność jest mierzona w jednostkach pamięci takich, jak:
  - bit (b) – najmniejsza jednostka pamięci,
  - bajt (B) - 8 bitów,
  - słowo - 8, 16, 32, 64,... bitów.
- Pojemność pamięci wewnętrznej wyrażana jest w bajtach lub słowach, pamięci zewnętrznej w bajtach.

5

## Adresowanie pamięci

- Pamięć składa się z komórek.
- Każda komórka przechowuje pewien element danych.
- Każda komórka posiada swój numer (**adres**), poprzez który programy odwołują się do niej.
- Każda komórka zawiera tyle samo bitów.
- Sąsiadujące komórki mają kolejne adresy.
- Komórka pamięci jest najmniejszą adresowalną jednostką pamięci.
- Słowo maszynowe składa się z kilku przylegających do siebie komórek, które mają często taki sam adres i które prawie zawsze traktujemy jako niepodzielną całość. Długość słowa odpowiada wielkości danych, jakie są najwydajniej przetwarzane w danej architekturze.

6

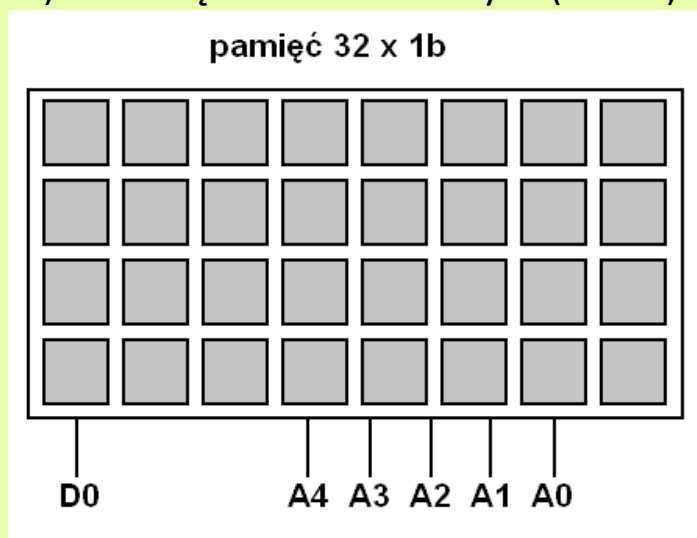
## Organizacja pamięci

- Adres komórki jest liczbą binarną. Długość adresu (liczba bitów) decyduje o wielkości pamięci.
- Jeżeli adres ma  $n$  bitów, to największa liczba komórek w pamięci wynosi  $2^n$ .
- Jeżeli mamy  $m$ -bitową szynę danych i  $n$ -bitową szynę adresową, to możemy w pamięci przechować  $2^n$  słów  $m$ -bitowych. Zatem pojemność pamięci wynosi  $m * 2^n$  bitów.
- Pamięci o tej samej wielkości mogą mieć zatem różną liczbę komórek (np. pamięć 32-bitowa może składać się z 32 komórek jednobitowych lub 4 komórek o długości 8 bitów).

7

## Bitowa organizacja pamięci - przykład

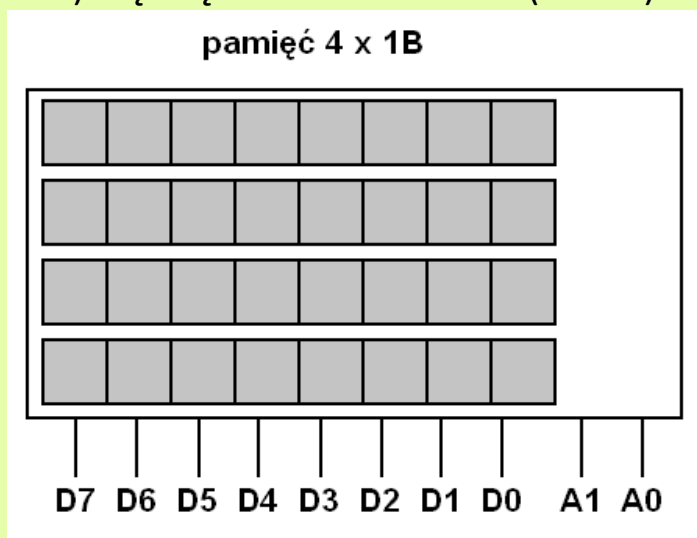
- Komórki pamięci składają się z jednego bitu.
- Pamięć służy do przechowywania słów o długości 1 bita, zatem jest tylko jedno wejście (D0).
- Pamięć składa się z 32 komórek, zatem adresy komórek są 5-bitowe ( $2^5=32$ ). Jest więc 5 linii adresowych (A0-A4).



8

## Bajtowa organizacja pamięci - przykład

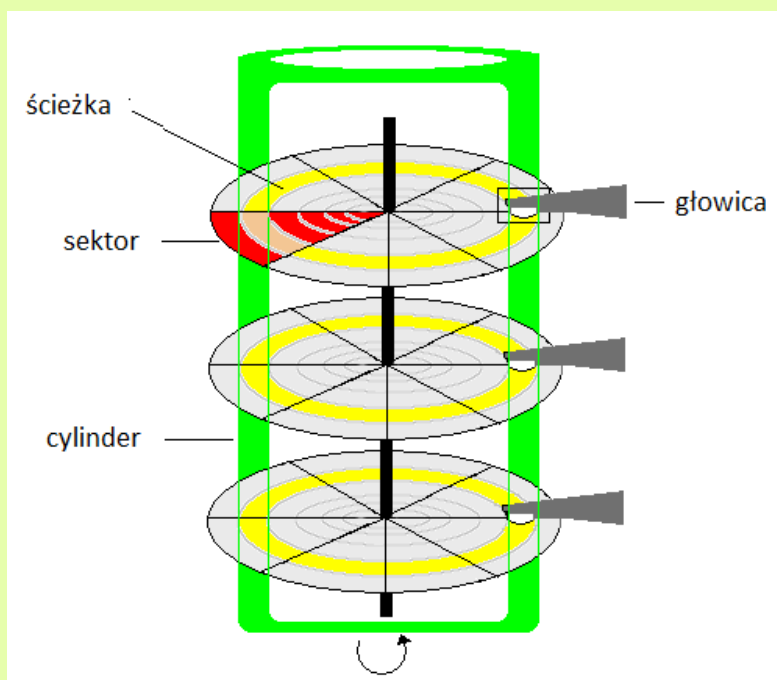
- Komórki pamięci składają się z 8 bitów.
- Pamięć służy do przechowywania słów o długości 8 bitów (1 bajta), zatem jest 8 linii wejściowych (D7-D0).
- Pamięć składa się z 4 komórek, zatem adresy komórek są dwubitowe ( $2^2=4$ ). Są więc 2 linie adresowe (A1-A0).



9

## Adresowanie pamięci

- W pamięciach masowych (np. w magnetycznych twardych dyskach, HDD) adresowanymi fragmentami nie są pojedyncze słowa, lecz większe obszary. Obszary te nazywane są sektorami.



10

## Cechy pamięci – jednostka transferu

---

- **Jednostka transferu** jest liczbą bitów jednocześnie odczytywanych lub zapisywanych w pamięci.
- Jest ona równa liczbie linii danych doprowadzonych do pamięci i wychodzących z niej.
- Jest ona często równa długości słowa, ale to nie jest zasadą.

11

## Cechy pamięci – wydajność

---

- Wydajność opisuje szybkość pracy pamięci, tzn. jak często procesor lub inne urządzenie może z niej korzystać.
- Na wydajność składają się:
  - czas dostępu,
  - czas cyklu,
  - szybkość transferu.

12

## Czas dostępu

---

- **Czas dostępu** to czas niezbędny do zrealizowania operacji odczytu lub zapisu.
- Jest on mierzony od chwili dostarczenia adresu do momentu odczytania danych lub ich zapisania.

13

## Czas cyklu

---

- **Czas cyklu** to najkrótszy możliwy czas, jaki musi upłynąć pomiędzy dwoma kolejnymi żądaniami dostępu do pamięci (kolejnymi dwoma odczytami lub zapisami).
- Jest większy od czasu dostępu z uwagi na opóźnienia wnoszone przez układy elektroniczne.

14

## Szybkość transferu

- **Szybkość transferu** to szybkość, z jaką dane mogą być odczytywane z pamięci lub w niej zapisywane.
- Najczęściej szybkość podaje się w bitach lub bajtach na sekundę.
- Szybkość transferu ma szczególne znaczenie w przypadku pamięci, w których adresowane są fragmenty większe niż słowo (np. pamięci dyskowych).

15

## Cechy pamięci – trwałość zawartości

- Pamięć
  - ulotna – zawartość pamięci jest tracona po wyłączeniu zasilania pamięci,
  - nieulotna – zawartość pamięci jest zachowywana po wyłączeniu zasilania pamięci.
- Pamięć
  - do zapisu i odczytu – zawartość pamięci może być zmieniana,
  - tylko do odczytu – zawartość pamięci nie może być zmieniana.
- Pamięć
  - statyczna – zawartość pamięci jest trwała w czasie utrzymywania zasilania,
  - dynamiczna – zawartość pamięci zanika w czasie utrzymywania zasilania.

16



## Cechy pamięci – sposób dostępu do danych

---

Wyróżnia się cztery sposoby dostępu do danych:

- dostęp sekwencyjny,
- dostęp bezpośredni,
- dostęp swobodny,
- dostęp skojarzeniowy.

17

## Dostęp sekwencyjny

---

- Dostęp do określonej jednostki danych (określonej pozycji) z pozycji bieżącej odbywa się poprzez przejście przez wszystkie pozycje pośrednie.
- Czas dostępu do poszczególnych jednostek jest różny i zależy od odległości danej jednostki od bieżącej.
- Przykładem pamięci o dostępie sekwencyjnym jest pamięć taśmowa.

18

## Dostęp bezpośredni

---

- Bloki pamięci są adresowane.
- Dostęp do jednostki danych odbywa się poprzez odszukanie bloku pamięci (bezpośrednio przez adres bloku) a następnie wykonanie niezbędnego przesunięcia.
- Czas dostępu do poszczególnych jednostek jest zmienny.
- Przykładem pamięci o dostępie bezpośrednim jest pamięć dyskowa.

19

## Dostęp swobodny

---

- Każda jednostka danych (adresowalna) ma fizycznie wbudowany mechanizm adresowania.
- Czas dostępu do poszczególnych jednostek jest stały i nie zależy od adresu.
- Przykładem pamięci o dostępie swobodnym jest pamięć operacyjna.

20

## Dostęp skojarzeniowy

---

- Rodzaj dostępu swobodnego.
- Umożliwia dodatkowo porównywanie i specyficzne badanie zgodności części bitów wszystkich słów pamięci.
- Na podstawie zgodności zostaje wyprowadzona zawartość całego słowa.
- Czas dostępu do poszczególnych jednostek jest stały.
- Przykładem pamięci o dostępie skojarzeniowym jest asocjacyjna pamięć podręczna.

21

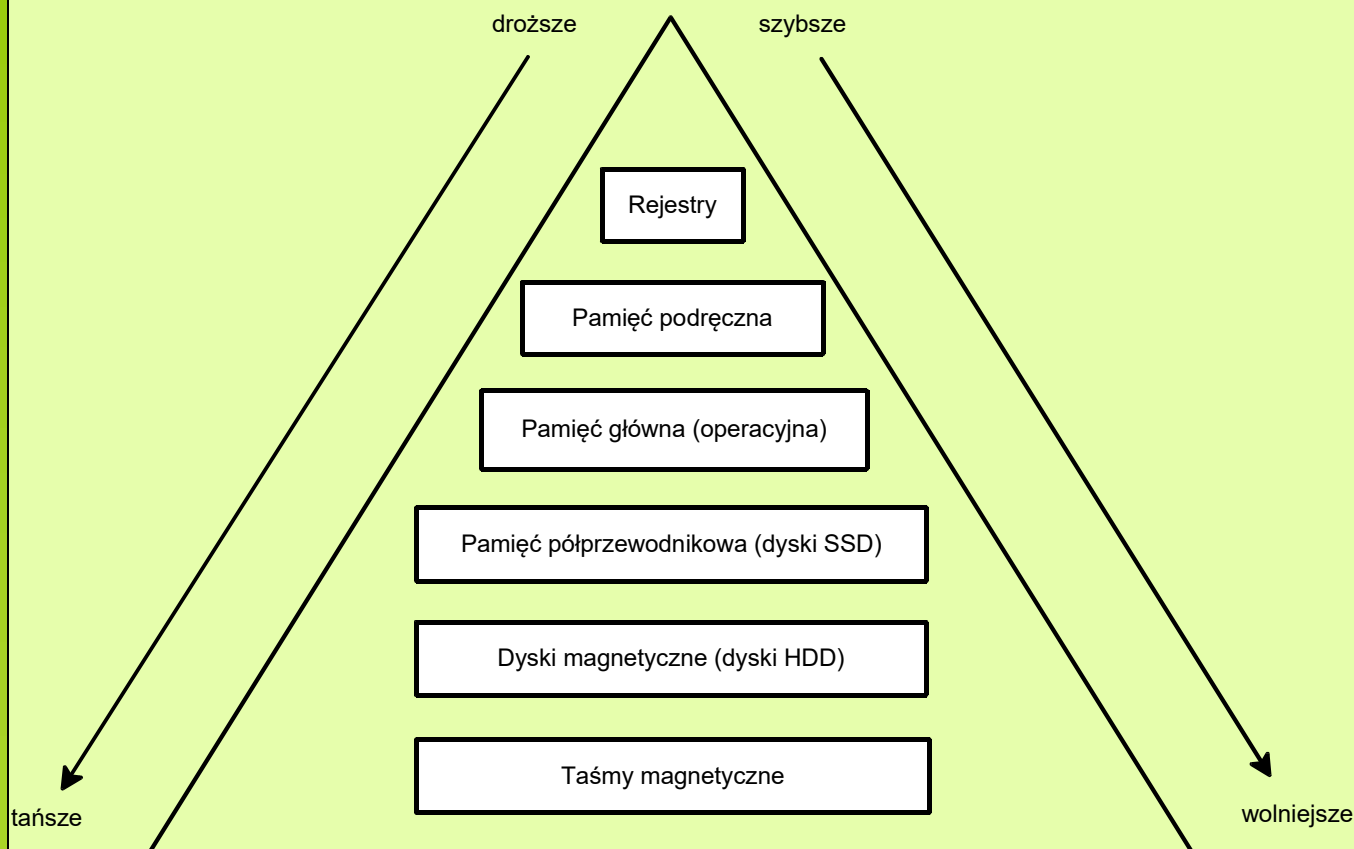
## Hierarchia pamięci

---

- System komputerowy zawiera kilka rodzajów pamięci.
- Różnią się one parametrami funkcjonalnymi oraz związanym z nimi kosztem na bit.
- Zachodzą następujące zależności:
  - krótszy czas dostępu – większy koszt na bit,
  - większa pojemność – mniejszy koszt na bit,
  - większa pojemność – większy czas dostępu.
- Ponieważ szybsze pamięci są droższe, stosuje się je w mniejszych ilościach tam, gdzie częstość odwoływania się przez system jest większa.

22

# Hierarchia pamięci



23

## Hierarchia pamięci a dostęp do danych

- W momencie pobierania dowolnych danych procesor wysyła odpowiednie żądania do najszybszej pamięci (z reguły rozpoczyna od pamięci podręcznej).
- Jeżeli dane znajdują się w pamięci podręcznej, zostają udostępnione procesorowi.
- Jeżeli danych nie ma w pamięci podręcznej, żądanie jest przesyłane na kolejny, niższy poziom.
- Jeżeli dane zostaną znalezione, do pamięci podręcznej zostają przekazane nie tylko żądane dane, ale cały blok je zawierający.
- Jeżeli dane nie zostaną znalezione, to żądanie zostaje przekazane na niższy poziom itd.

24

## Trafienia i chybień

- **Trafienie** – żądane dane zostały znalezione w pamięci danego poziomu.
- **Chybień** – żądane dane nie zostały znalezione w pamięci danego poziomu.
- **Współczynnik trafień** – procentowa wartość liczby trafień w stosunku do wszystkich operacji dostępu do pamięci na danym poziomie.
- **Współczynnik chybień** – procentowa wartość liczby chybień w stosunku do wszystkich operacji dostępu do pamięci na danym poziomie.
- **Czas trafienia** – czas potrzebny do pobrania żądanych danych z pamięci na określonym poziomie.
- **Czas chybień** – czas niezbędny na przetworzenie chybień, czyli na zastąpienie bloku pamięci znajdującej się na wyższym poziomie i dostarczenie danych do procesora.

25

## Rejestry

- Pamięć rejestrowa to zbiór wszystkich dostępnych rejestrów znajdujących się w procesorze i przechowujących argumenty operacji wykonywanych przez procesor.
- Dostęp do nich jest możliwy poprzez układ sterowania, którego zadaniem jest wybranie odpowiedniego rejestru na podstawie kodu maszynowego.
- Jest to pamięć statyczna, złożona z przerzutników.
- Jest to pamięć ulotna.

26

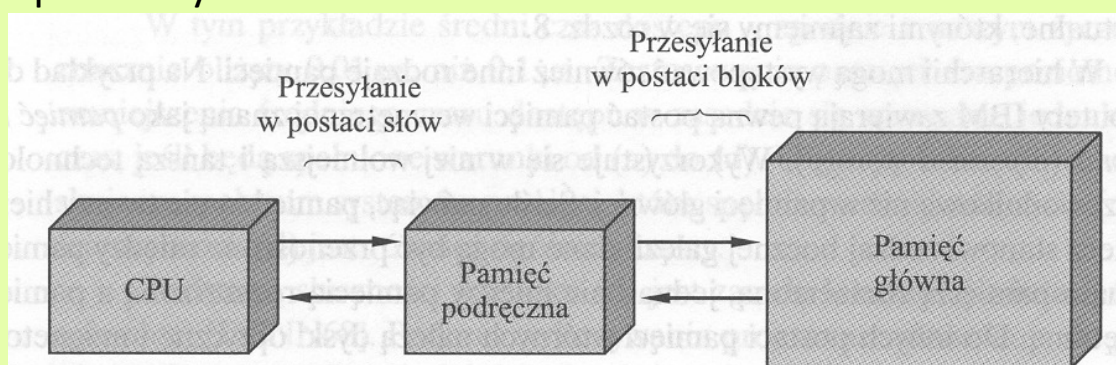
## Pamięć podręczna

- Pamięć podręczna jest mało pojemną lecz szybką pamięcią, z której korzysta procesor do przechowywania danych, których prawdopodobnie w najbliższym czasie będzie ponownie potrzebował.
- Zadaniem pamięci podręcznej jest skrócenie czasu dostępu do pamięci, co osiągane jest poprzez przechowywanie ostatnio pobranych danych, zamiast w pamięci głównej, w pamięci bliżej położonej procesora.
- Stosuje się do czterech poziomów pamięci podręcznej (L1, L2, L3, L4)

27

## Lokalność odniesień

- W czasie wykonywania programu procesor często żąda danych i instrukcji znajdujących się w pamięci blisko siebie. Własność ta nosi nazwę **lokalności odniesień**.
- Aby móc ją wykorzystać, przetwarzanie chybienia polega na przesłaniu do pamięci znajdującej się na wyższym poziomie nie tylko żądanych danych, ale całego bloku zawierającego te dane. Z uwagi na lokalność odniesień jest bardzo prawdopodobne, że w najbliższym czasie pojawi się żądanie danych znajdujących się w przesłanym bloku.



28

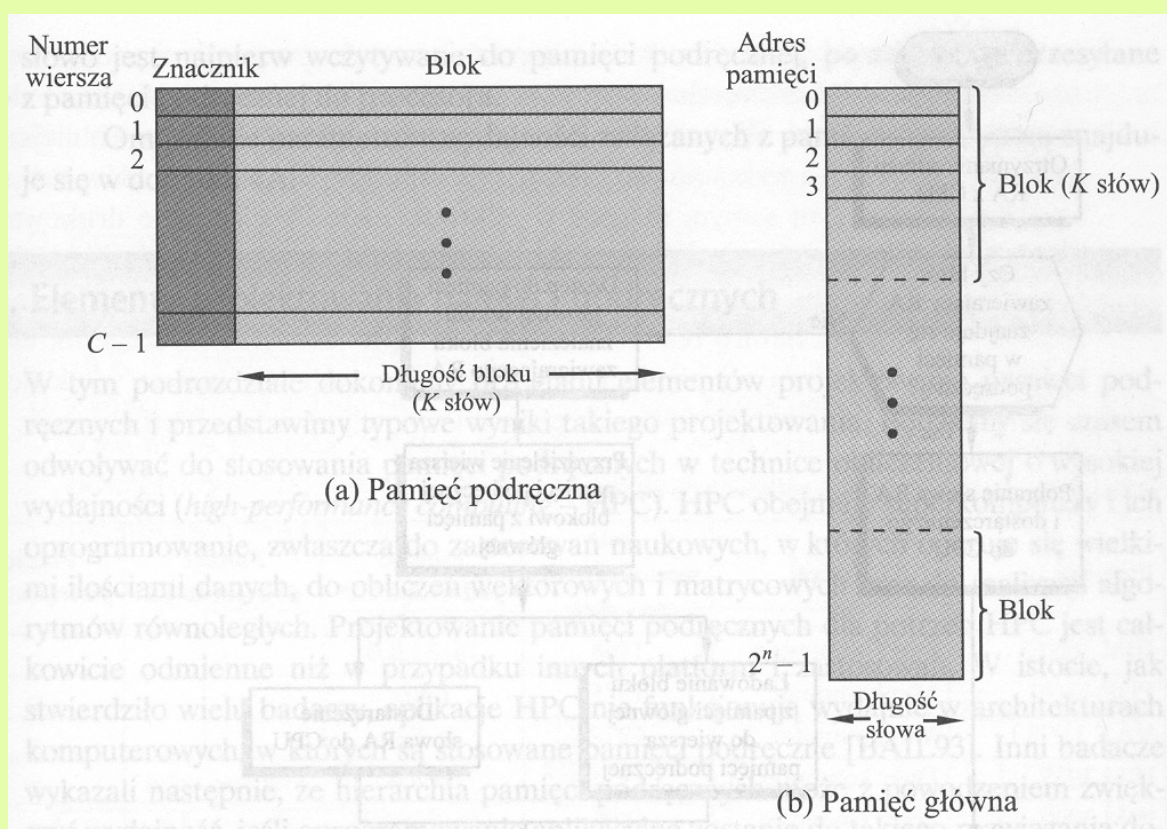


## Rodzaje lokalności odniesień

- **Lokalność czasowa** – jest bardzo możliwe, że w najbliższym czasie pojawi się żądanie danych ostatnio pobranych, np. podczas wykonywania pętli.
- **Lokalność przestrzenna** – programy z reguły zawierają pętle i wywołują podprogramy, zatem występują w nim odniesienia do tego samego zestawu rozkazów; podobnie operacje np. na tablicach polegają na dostępie do danych umieszczonych w zwartym obszarze.

29

## Struktura pamięci podręcznej i pamięci głównej



30

# Struktura pamięci podręcznej i pamięci głównej

Pamięć główna składa się z  $2^n$  adresowalnych słów, przy czym każde słowo ma jednoznaczny adres  $n$ -bitowy. Aby było możliwe odwzorowywanie, pamięć ta składa się z pewnej liczby bloków o stałej długości, zawierających  $K$  słów każdy. Tak więc występuje  $M = 2^n/K$  bloków.

Pamięć podręczna zawiera  $C$  wierszy zawierających  $K$  słów każdy, a liczba wierszy jest znacząco mniejsza od liczby bloków w pamięci głównej.

W każdej chwili pewien zespół bloków pamięci pozostaje w wierszach pamięci podręcznej. Jeśli słowo w bloku pamięci jest odczytywane, to odpowiedni blok jest przenoszony do jednego z wierszy pamięci podręcznej. Ponieważ bloków jest więcej niż wierszy, określony wiersz nie może być jednoznacznie i trwale przypisany określonemu blokowi.

Każdy wiersz zawiera w związku z tym **znacznik** określający, który blok jest właśnie zapisywany. Znacznik jest zwykle częścią adresu pamięci głównej.

31

## Odwzorowywanie pamięci głównej do pamięci podręcznej

- Odwzorowywanie bezpośrednie - każdy blok pamięci głównej jest odwzorowywany na dokładnie jeden wiersz pamięci podręcznej (gdy w pamięci podręcznej jest  $m$  wierszy, to na  $i$ -ty wiersz jest odwzorowywany co  $m$ -ty blok (blok  $i, m+i, 2m+i, \dots$ )).
- Metoda prosta i tania we wdrażaniu
- Ponieważ dla każdego bloku istnieje stała lokalizacja w pamięci podręcznej (stały wiersz), to w przypadku, gdy program będzie się często odwoływał do słów znajdujących się w różnych blokach przypisanych do tego samego wiersza, to bloki te będą musiały być ciągle przenoszone do pamięci podręcznej, co obniża ogólna szybkość działania.

32



# Odwzorowywanie pamięci głównej do pamięci podręcznej

- Odwzorowywanie skojarzeniowe - blok pamięci głównej może być odwzorowany do dowolnego wiersza. Pole znacznika jednoznacznie określa blok pamięci głównej (bo jest częścią adresu).
- Konieczność stosowania algorytmów zastępowania umożliwiające maksymalizację współczynnika trafień.
- W celu stwierdzenia, czy blok znajduje się w pamięci podręcznej, sterujące układy logiczne pamięci podręcznej muszą jednocześnie zbadać zgodność znacznika każdego wiersza.
- Odwzorowywanie sekcyjno-skojarzeniowe – pamięć podręczna dzielona jest na  $t$  sekcji, z których każda składa się z  $k$  wierszy (w pamięci podręcznej jest w sumie  $k*t$  wierszy). Wewnątrz sekcji blok może być odwzorowany na dowolny wiersz tej sekcji.

33

## Algorytmy zastępowania

Gdy do pamięci podręcznej jest wprowadzany nowy blok, jeden z istniejących bloków musi być zastąpiony. W przypadku odwzorowania bezpośredniego istnieje tylko jeden możliwy wiersz dla każdego bloku i to zawartość tego wiersza zostanie wymieniona.

W pozostałych przypadkach (odwzorowanie skojarzeniowe, sekcyjno-skojarzeniowe) stosowane są algorytmy zastępowania:

1. Najbardziej efektywny to algorytm „najdalej używany” (least-recently used – LRU) – określa, że należy zastąpić ten blok, który pozostawał w pamięci podręcznej najdłużej bez odwoływania się do niego.
2. algorytm „pierwszy wchodzi – pierwszy wychodzi” (first-in-first-out – FIFO) – należy zastąpić ten blok, który pozostawał najdłużej w pamięci podręcznej.
3. algorytm „najrzadziej używany” (least frequently used – LFU) – zastępowany jest blok, którego dotyczyło najmniej odniesień.

34

## Pamięć RAM

---

Pamięć RAM (*Random Access Memory*) jest:

- pamięcią o dostępie swobodnym,
- pamięcią przeznaczoną do odczytu i zapisu,
- pamięcią ulotną.

Z uwagi na technologię wykonania dzieli się na:

- pamięć statyczną SRAM,
- pamięć dynamiczną DRAM.

35

## Pamięć statyczna SRAM

---

- Jest zbudowana z przerzutników.
- Pojedyncza komórka (służąca do przechowania jednego bitu) składa się z 6 tranzystorów.
- Jej cechą charakterystyczną jest utrzymywanie zawartości tak długo, jak długo dostarczane jest zasilanie.
- Jest bardzo szybka – czas dostępu jest rzędu kilku nanosekund.
- Charakteryzuje się dużym poborem mocy.
- Jest stosowana w pamięci podręcznej

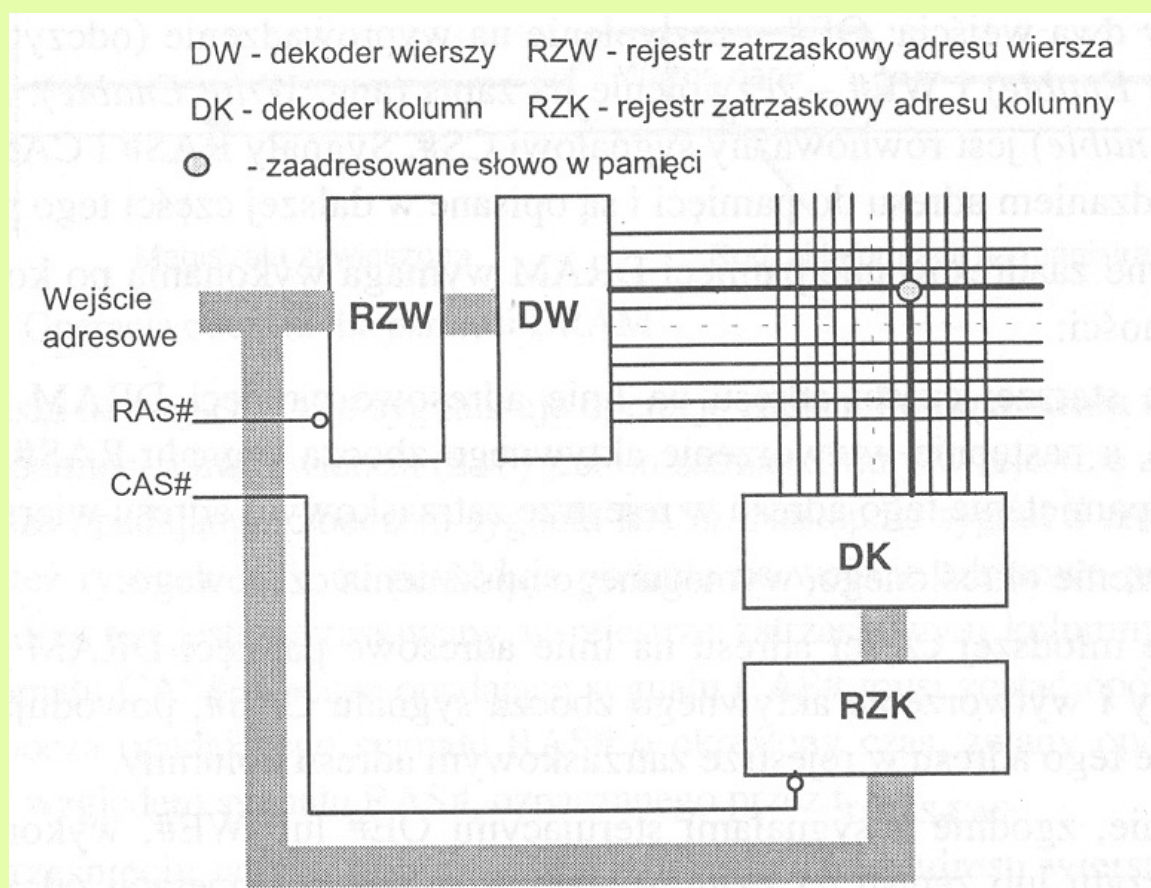
36

# Pamięć dynamiczna DRAM

- Nie jest zbudowana z przerzutników.
- Jest to macierz komórek, z których każda składa się z jednego tranzystora i jednego małego kondensatora.
- Kondensatory można ładować i rozładowywać, co odpowiada zapisywaniu jedynek i zer.
- Z uwagi na nieuchronny wypływ ładunku z kondensatora (naturalną tendencję do rozładowywania się) musi być regularnie odświeżana.
- Jest wolniejsza od pamięci SRAM – jej czas dostępu jest rzędu dziesiątek nanosekund.
- Jest tańsza od pamięci SRAM i zużywa mniej energii.
- Umożliwia uzyskanie większej gęstości (liczby bitów w jednym układzie).
- Wykorzystywana w pamięci operacyjnej (głównej).

37

# Pamięć dynamiczna DRAM



38

## Pamięć ROM

---

Pamięć ROM (*Read Only Memory*) jest

- pamięcią o dostępie swobodnym,
- pamięcią przeznaczoną tylko do odczytu,
- pamięcią nieulotną.

Przechowuje ważne dane, od których zależy funkcjonowanie systemu, takie jak program wymagany do przeprowadzenia inicjalizacji komputera.

Określenie, że ROM jest pamięcią tylko do odczytu, nie jest równoznaczne z tym, że zawartości tej pamięci w określonych warunkach nie można zmieniać. Dla niektórych typów technologicznych pamięci ROM jest to możliwe.

39

## Rodzaje pamięci ROM

---

- ROM (*Read-Only Memory*) – zawartość pamięci ustalana raz w czasie produkcji.
- PROM (*Programmable ROM*) – programowalna pamięć ROM, umożliwia jednorazowy zapis.
- EPROM (*Erasable PROM*) – wymazywalna pamięć PROM, umożliwia wielokrotne wymazywanie poprzez długotrwałe naświetlanie promieniowaniem ultrafioletowym.
- EEPROM (*Electrically Erasable PROM*) – pamięć PROM wymazywalna elektrycznie, umożliwia wielokrotne wymazywanie poprzez doprowadzenie impulsu elektrycznego, dużo wolniejsza, droższa i mniej pojemna od pamięci RAM.
- Flash – odmiana pamięci EEPROM umożliwiająca wielokrotny zapis; używana np. w kartach SD, dyskach SSD; wadą jest mała dopuszczalna liczba skasowań (ok. 100 tysięcy).

40

## Pamięć zewnętrzna

---

- Dyski magnetyczne
- Pamięć optyczna (CD, DVD)
- Pamięć półprzewodnikowa (dyski SSD, karty SD)
- Taśmy magnetyczne

41

## Organizacja jednostki centralnej

---

Głównymi składnikami jednostki centralnej (CPU, procesora) są:

- **Jednostka arytmetyczno-logiczna** (ALU), która realizuje funkcję przetwarzania danych przez komputer,
- **Jednostka sterująca**, która steruje działaniem procesora,
- **Rejestry**, które służą do wewnętrznego przechowywania danych w procesorze,
- **Połączenia wewnętrzne**, które zapewniają łączność między elementami CPU.

42

## Zadania jednostki centralnej

---

- **Pobieranie rozkazów.** Procesor ma odczytywać rozkazy z pamięci.
- **Interpretowanie rozkazów.** Pobrane rozkazy należy zdekodować, aby określić, jakie działania będą wymagane.
- **Pobieranie danych.** W celu wykonania rozkazu może zachodzić potrzeba pobrania danych z pamięci lub modułu wejścia-wyjścia.
- **Przetwarzanie danych.** Wykonanie rozkazu może wymagać przeprowadzenia na danych operacji arytmetycznych lub logicznych.
- **Zapisanie danych.** Wyniki przeprowadzonych operacji mogą być przekazywane do pamięci lub urządzeń wejścia-wyjścia.

43

## Rozkazy maszynowe

---

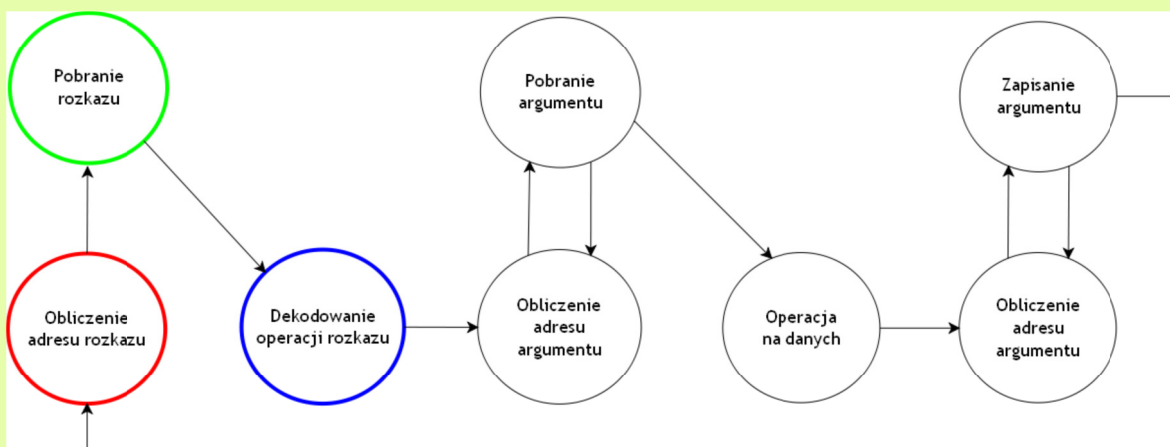
- Rozkazy wykonywane przez procesor nazywane są **rozkazami maszynowymi**.
- Zbiór rozkazów, które może wykonać dany procesor, nazywa się **listą rozkazów** tego procesora.

44



## Cykl wykonania rozkazu

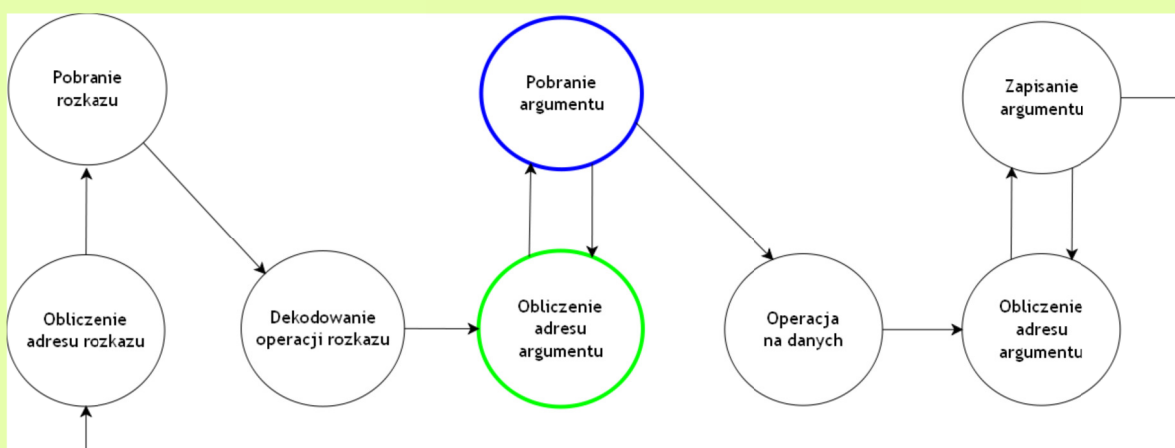
1. **Obliczenie adresu rozkładu.** Określenie adresu następnego rozkazu przeznaczonego do wykonania.
2. **Pobieranie rozkazu.** Wczytanie rozkazu do procesora z pamięci.
3. **Dekodowanie operacji rozkazu.** Interpretacja rozkazu w celu określenia operacji, która ma być przeprowadzona.



45

## Cykl wykonania rozkazu

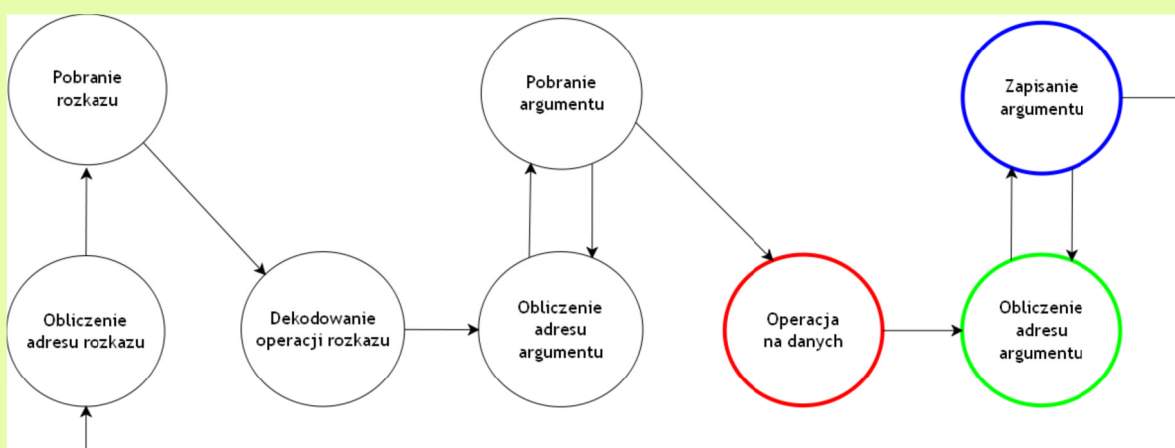
4. **Obliczenie adresu argumentu.** Określenie adresu argumentu, jeśli operacja wymaga argumentu i znajduje się on w pamięci lub jest dostępny przez urządzenia wejścia-wyjścia.
5. **Pobieranie argumentu.** Wczytanie argumentu z pamięci lub urządzenia wejścia-wyjścia.



46

## Cykl wykonania rozkazu

6. **Operacja na danych**. Wykonanie operacji wskazanej w rozkazie.
7. **Obliczenie adresu argumentu**. Określenie adresu, gdzie zostanie skierowany wynik operacji.
8. **Zapisanie argumentu**. Zapisanie wyniku w pamięci lub skierowanie go do urządzenia wejścia-wyjścia.



47

## Formaty rozkazów

- Format rozkazu określa sposób interpretacji poszczególnych bitów w rozkazie.
- Rozkaz zawiera kod operacji oraz ewentualne argumenty tej operacji.
- W większości list rozkazów występuje więcej niż jeden format rozkazu.
- Jednym z kluczowych parametrów rozkazu jest jego długość.
- Zaletą krótszych rozkazów jest to, że zajmują mniej miejsca w pamięci i mogą być szybciej wykonywane. Z kolei dłuższe rozkazy umożliwiają np. zwiększenie liczby operacji, dzięki czemu programy mogą być krótsze.

48



## Elementy rozkazów maszynowych

---

- **Kod operacji.** Określa operację, które ma być przeprowadzona (np. dodawanie).
- **Odniesienie do argumentów źródłowych.** Operacja może wymagać danych wejściowych.
- **Odniesienie do wyniku.** Operacja może zwracać wynik.
- **Odniesienie do następnego rozkazu.** Wskazuje położenie następnego rozkazu.

49

## Tryby adresowania

---

Argumenty w rozkazie mogą być adresowane w różny sposób:

- **Adresowanie natychmiastowe.** Wartość, do której ma nastąpić odwołanie, znajduje się w rozkazie.
- **Adresowanie bezpośrednie.** Wartość, do której ma nastąpić odwołanie, jest wczytywana bezpośrednio z komórki pamięci, której adres znajduje się w kodzie rozkazu.
- **Adresowanie rejestrowe.** Wartość, do której ma nastąpić odwołanie, jest wczytywana nie z pamięci, lecz z rejestru, który wskazywany jest w kodzie rozkazu.

50

## Tryby adresowania

---

Argumenty w rozkazie mogą być adresowane w różny sposób:

- **Adresowanie pośrednie.** W kodzie rozkazu znajduje się adres komórki w pamięci, w której znajduje się adres komórki przechowującej wartość argumentu.
- **Adresowanie rejestrowe pośrednie.** W kodzie rozkazu znajduje się wskazanie na rejestr, w którym znajduje się adres komórki przechowującej wartość argumentu.

51

## Tryby adresowania

---

Argumenty w rozkazie mogą być adresowane w różny sposób:

- **Adresowanie z przesunięciem.** Adresowanie łączące możliwości adresowania bezpośredniego z pośrednim adresowaniem rejestrowym. Wymaga dwóch pól adresowych w rozkazie. W pierwszym polu bezpośrednio jest podany adres komórki w pamięci. W drugim polu znajduje się wskazanie do rejestru, którego zawartość po dodaniu do adresu z pierwszego pola daje adres komórki przechowującej wartość argumentu.

52

## Źródła

---

- Część ilustracji pochodzi z książki:  
Stallings W., *Organizacja i architektura systemu komputerowego*, WNT 2000.