

ULOTNE PAMIĘCI PÓŁPRZEWODNIKOWE

Pamięci półprzewodnikowe są układami służącymi do przechowywania informacji cyfrowych.

Pamięci są zorganizowane w taki sposób, aby zapamiętywać N słów, przy czym każde słowo zawiera S bitów.

Całkowita pojemność pamięci wynosi zatem $N \times S$ bitów.

Termin słowo może mieć dwa znaczenia w terminologii komputerowej.

W pamięciach słowo jest definiowane jako grupa bitów lub bajtów, która działa jako pojedyncza jednostka i która może być przechowywana w jednym miejscu pamięci.

W języku assemblerowym słowo jest definiowane jako dwa bajty.

Jednostki pamięci

1 bajt [B] = 8 bitów [b]

1 kilobajt [KB] = 2^{10} bajtów = 1024 bajty [B]

1 megabajt [MB] = 2^{10} kilobajtów = 1024 kilobajty [KB]

1 gigabajt [GB] = 2^{10} megabajtów = 1024 megabajty [MB]

1 terabajt [TB] = 2^{10} gigabajtów = 1024 gigabajty [GB]

lub inaczej

1 B = 8 b

1 KB = 2^{10} B

1 MB = 2^{10} KB = 2^{20} B

1 GB = 2^{10} MB = 2^{20} KB = 2^{30} B

1 TB = 2^{10} GB = 2^{20} MB = 2^{30} KB = 2^{40} B

Parametry dynamiczne pamięci półprzewodnikowych

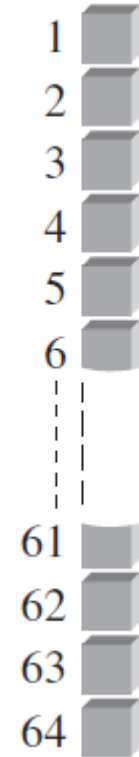
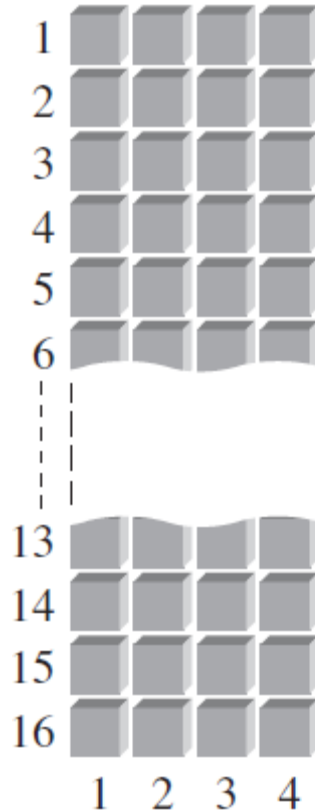
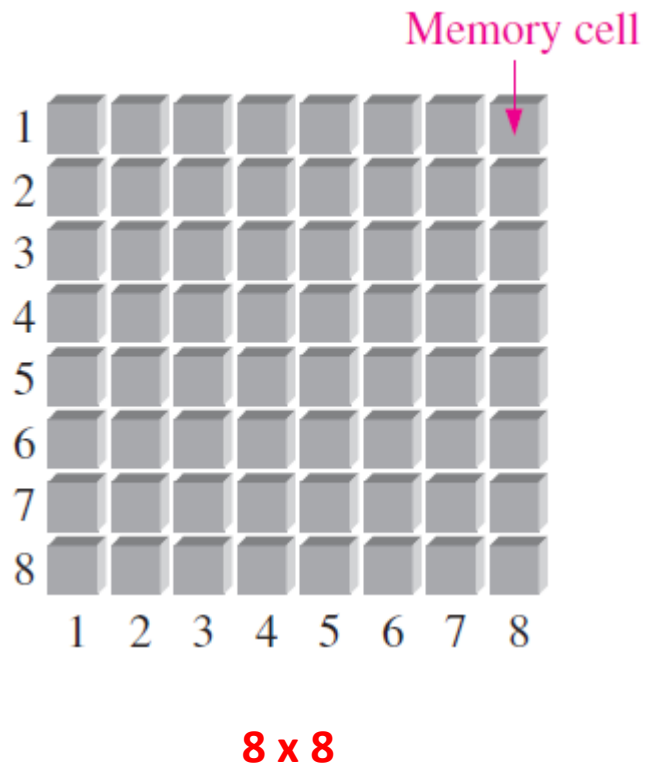
Czas dostępu jest to czas między podaniem sygnału na wejścia adresowe a pojawieniem się danych na wyjściu.

Czas cyklu jest to najmniejszy czas między kolejnymi wywołaniami informacji z pamięci.

Odwrotność czasu cyklu jest **szybkością pracy pamięci** i wyrażana jest w megahercach (MHz).

organizacja pamięci

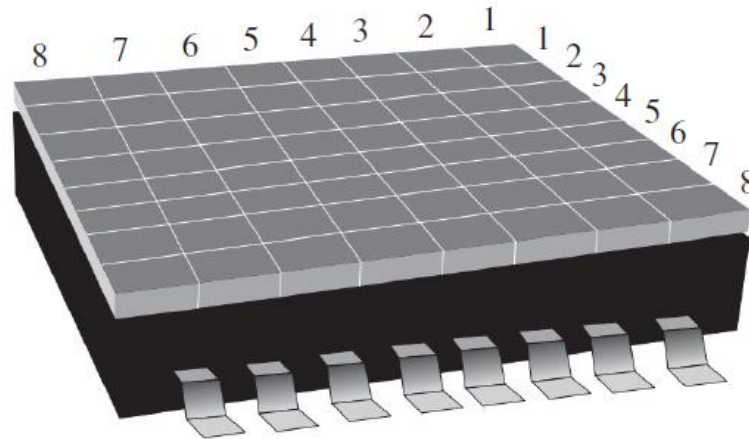
Pamięci półprzewodnikowe są pamięciami matrycowymi, gdyż tworzą tzw. matrycę pamięciową, czyli strukturę prostokątną, w której na przecięciu każdej kolumny i każdego wiersza jest umiejscowiona komórka pamięci.



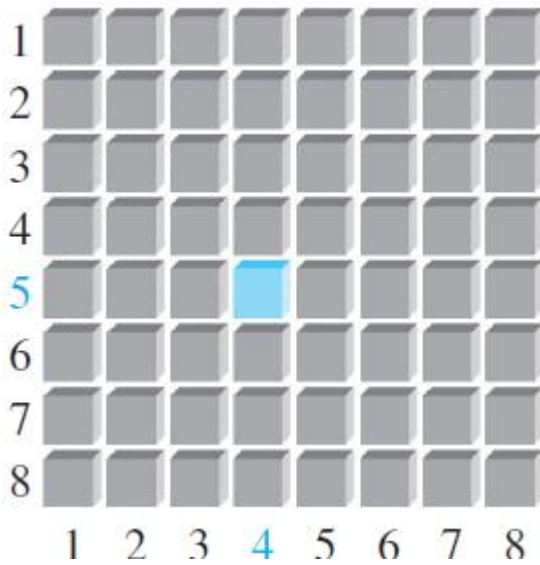
[*]

Różne organizacje 64 bitowej matrycy pamięci

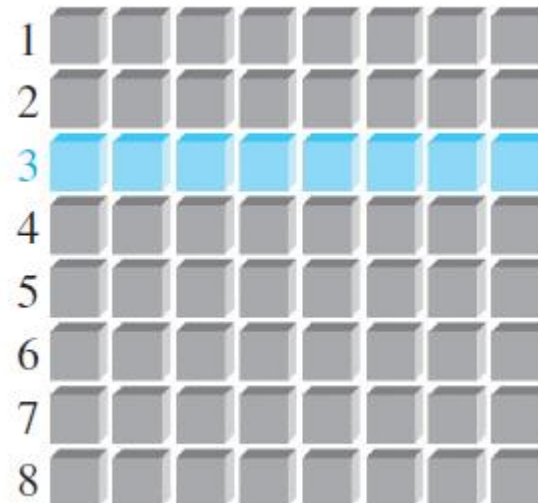
przykłady adresowania w dwuwymiarowej macierzy pamięci



fizyczna struktura 64 bitowej pamięci



adres komórki pamięci: 5 rząd 4 kolumna

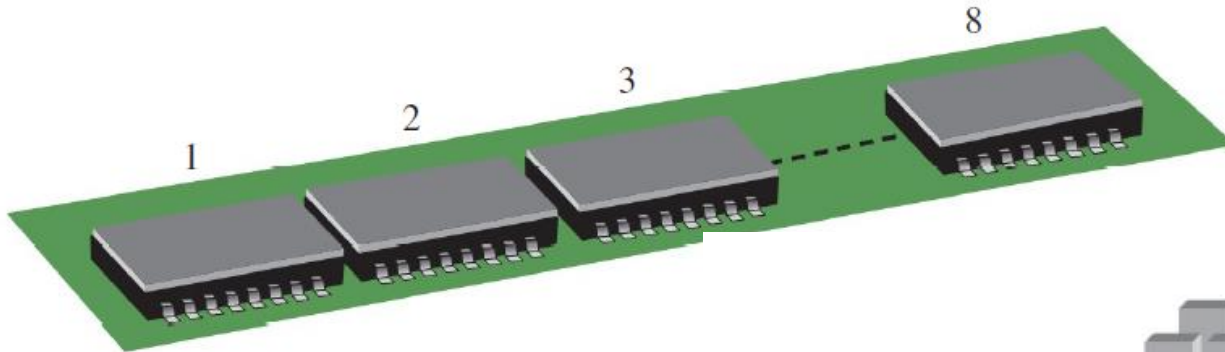


adres komórek pamięci: 3 rząd

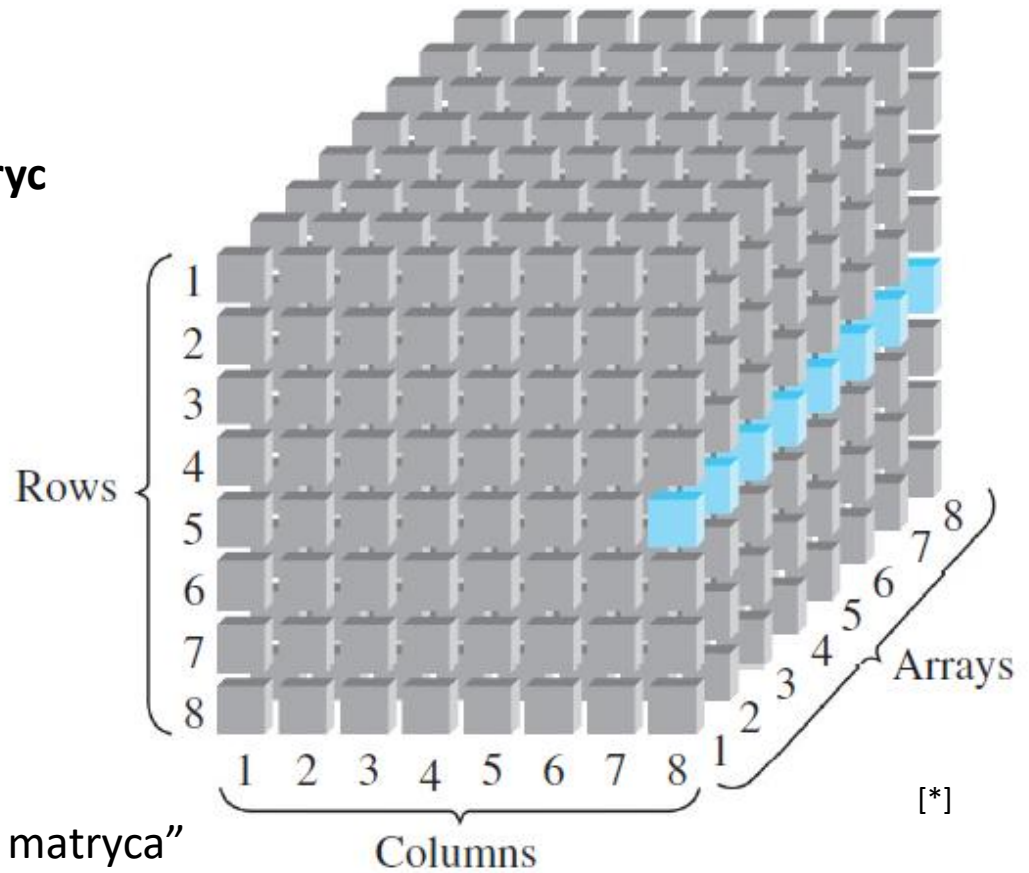
Komputery osobiste mają pamięć zorganizowaną w bajtach

[*]

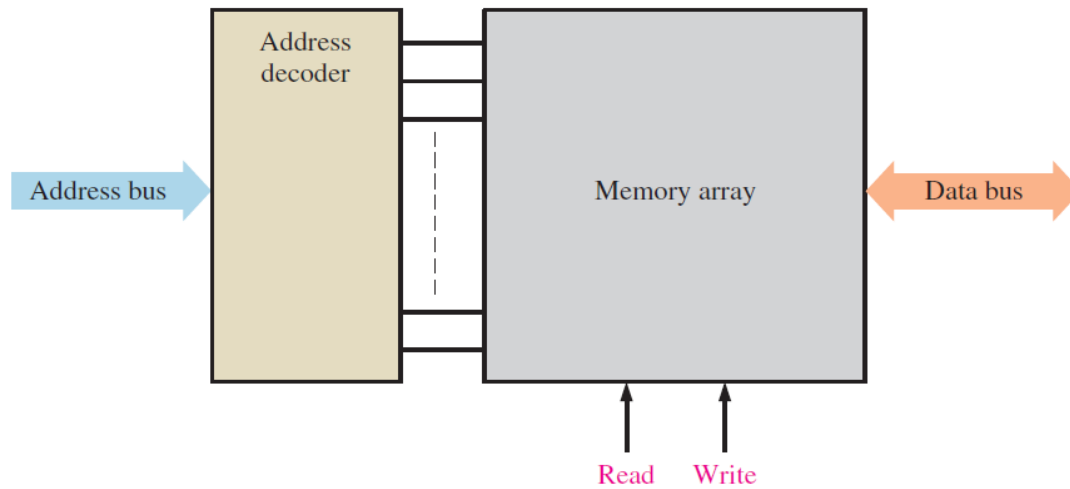
przykłady adresowania w rozszerzonej pamięci



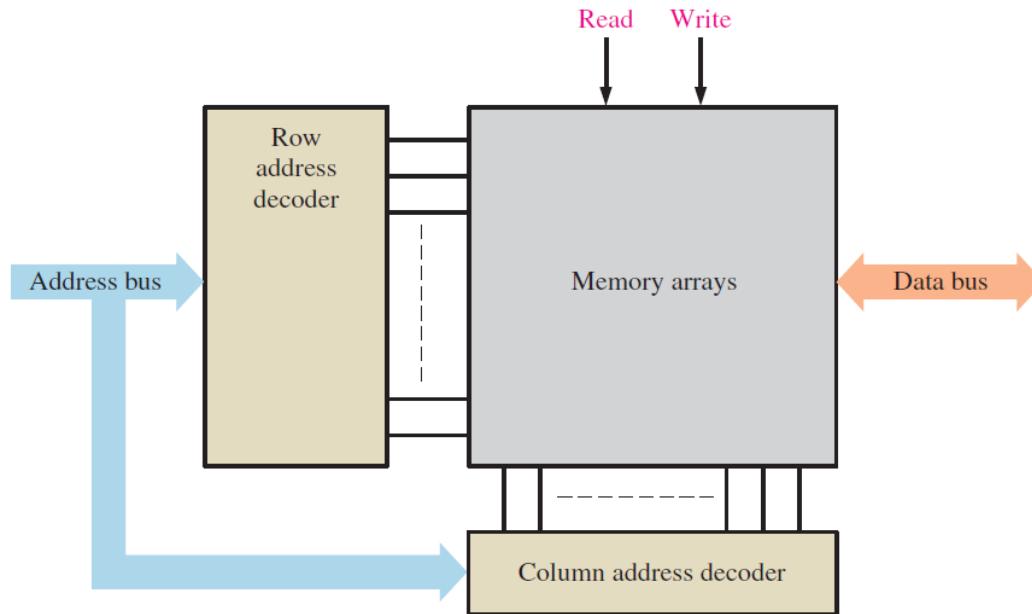
rozszerzanie pamięci – kilka matryc



powstaje „trójwymiarowa matryca”



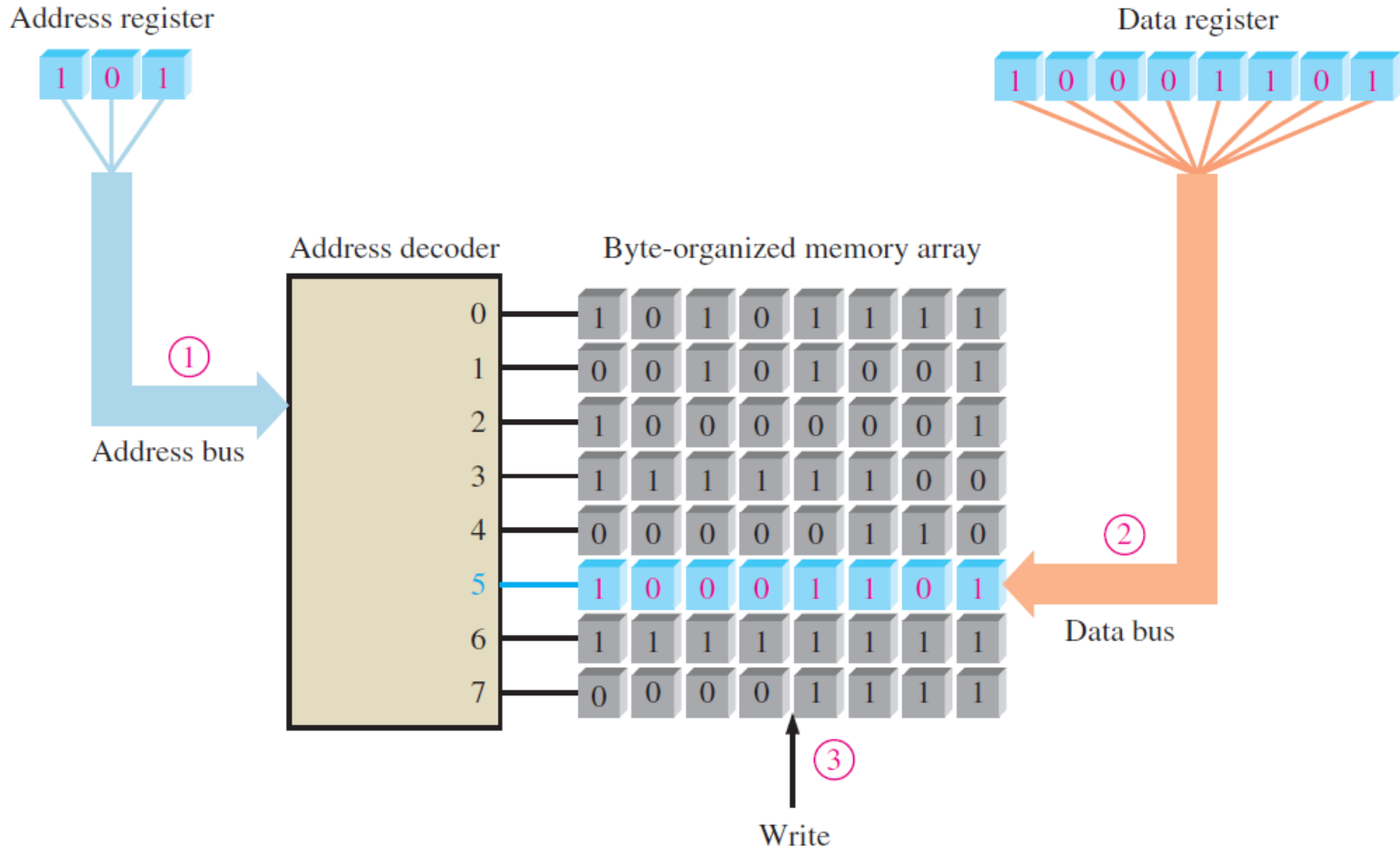
Schemat blokowy pamięci z pojedynczą matrycą



Schemat blokowy pamięci z wieloma matrycami

[*]

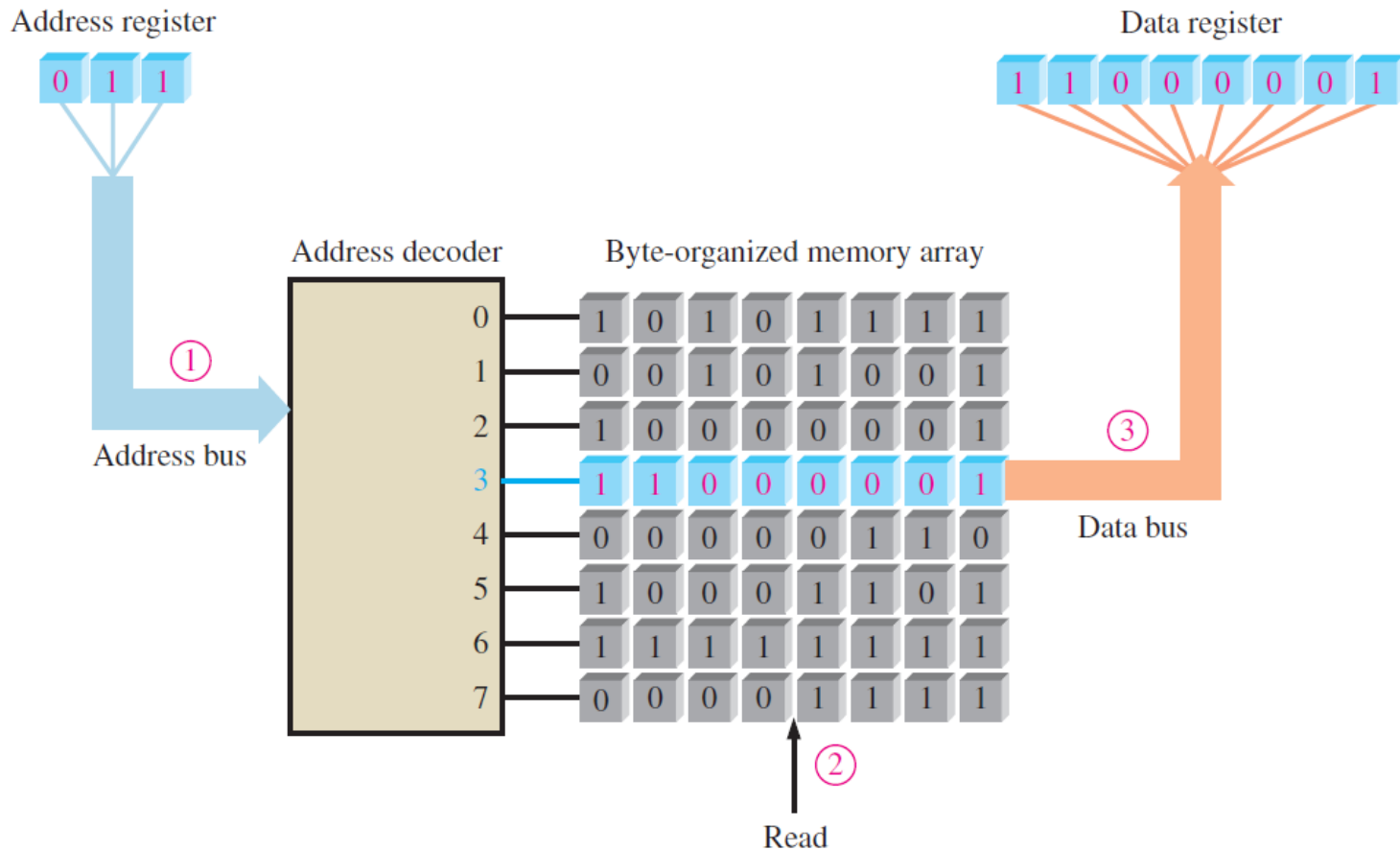
operacja zapisu



[*]

- 1 Kod adresowy 101 jest umieszczony na szynie adresowej i wybierany jest adres 5
- 2 Bajt danych jest umieszczany na szynie danych.
- 3 Polecenie zapisu powoduje zapisanie bajtu danych pod adresem 5, zastępując poprzednie dane

operacja odczytu



[*]

- 1 Kod adresowy 011 jest umieszczony na szynie adresowej i wybierany jest adres 3
- 2 Uruchomiono polecenie odczytu
- 3 Treść danych z adresu 3 jest umieszczana na szynie danych i przenoszona do rejestru danych. Zawartość danych z adresu 3 nie jest kasowana podczas operacji odczytu.

PAMIĘCI O DOSTĘPIE SWOBODNYM RAM

Najszerzej stosowane pamięci ulotne.

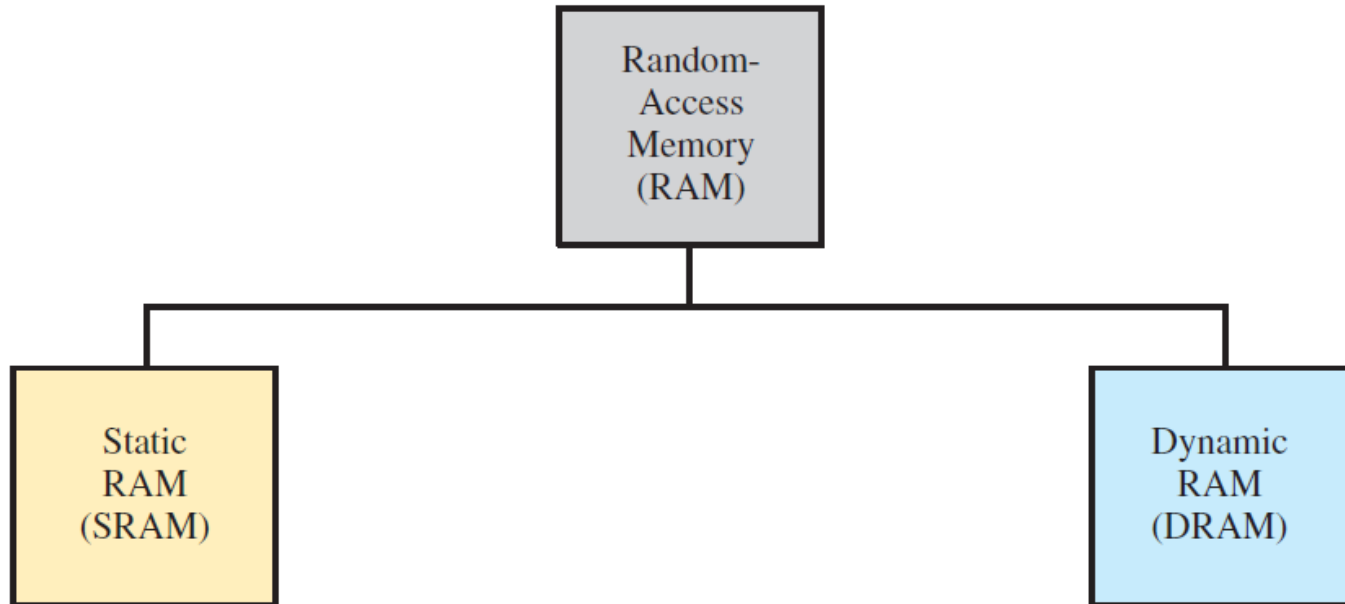
Umożliwiają zarówno szybki zapis, jak i szybki odczyt słowa danych.

Tradycyjnie określane terminem „pamięci o dostępie swobodnym” RAM (*Random-Access Memory*).

Ta tradycyjna nazwa nie jest najtrafniejsza, gdyż cecha swobodnego dostępu występuje też w innych rodzajach pamięci.

Właściwa nazwa „pamięć czytaj/pisz” RWM (*Read/Write Memory*) ale niestosowana.

Pamięci RAM dzieli się na dwie grupy: pamięci statyczne SRAM i pamięci dynamiczne DRAM.



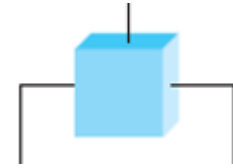
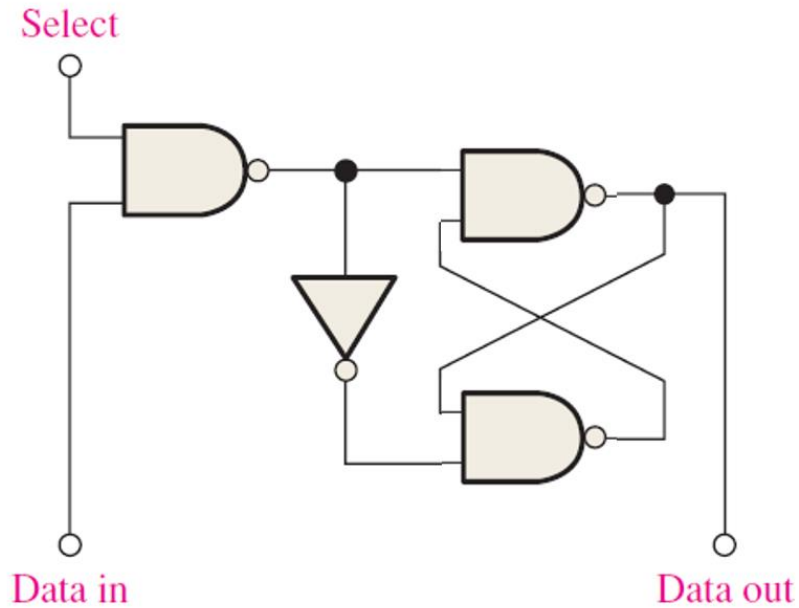
[*]

pamięci statyczne o dostępie swobodnym

SRAM

podstawowa komórka pamięci zatraskowej SRAM

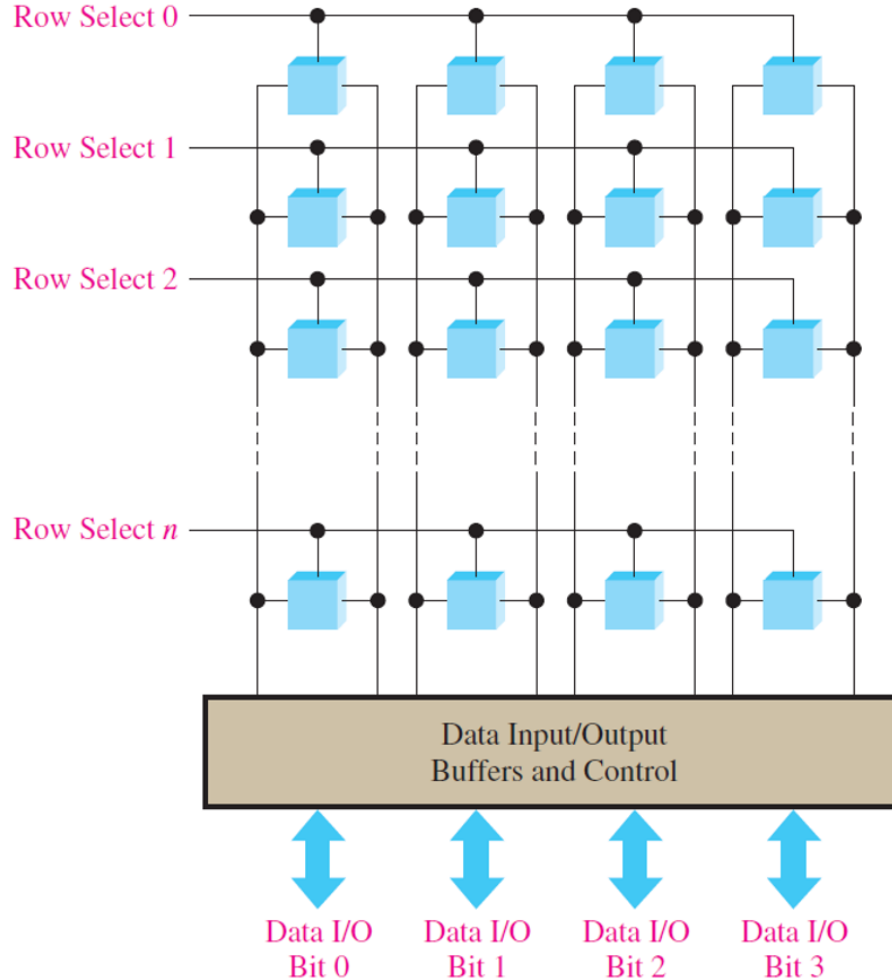
PAMIĘĆ SRAM



[*]

bramkowany zatrask D

Date in	Select	Date out	Komentarz
0	1	0	wpisanie 0
1	1	1	wpisanie 1
x	0	pamiętanie stanu poprzedniego	



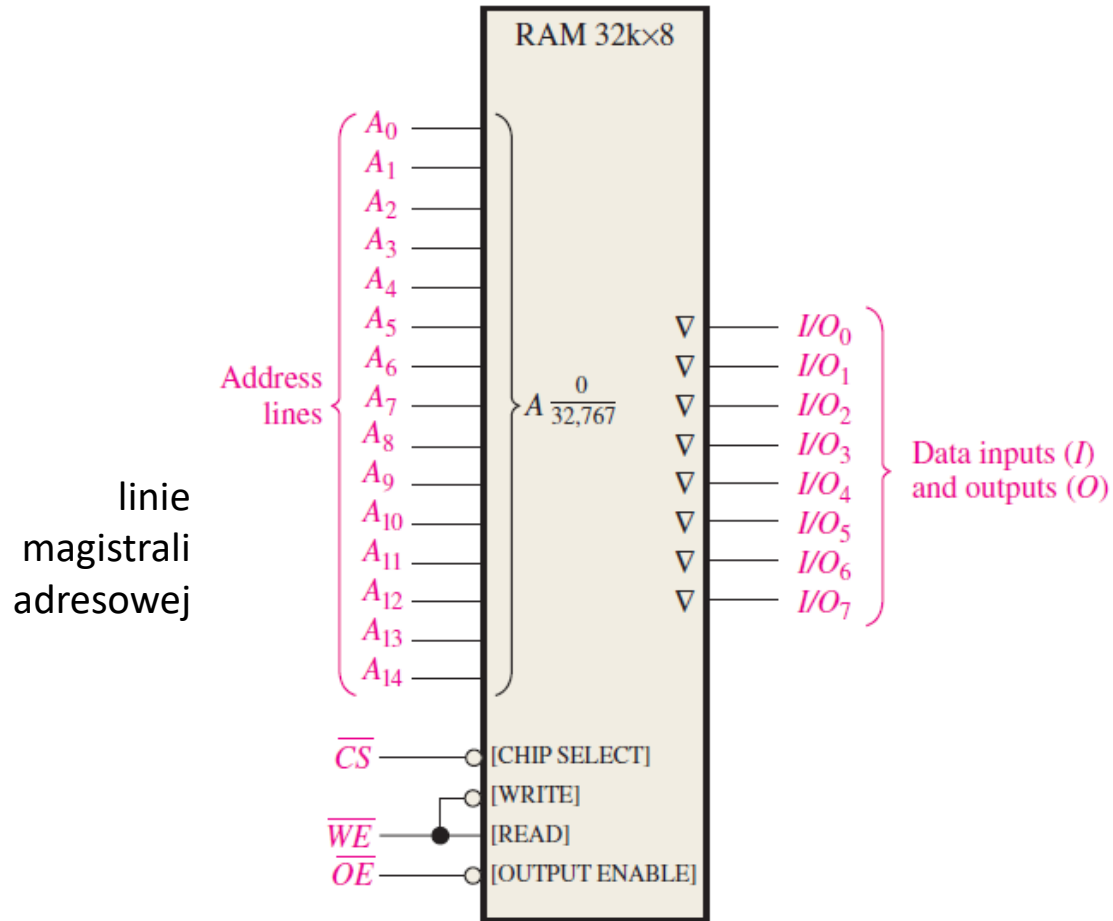
Komórki pamięci są zorganizowane w wiersze i kolumny.

Wszystkie komórki w wierszu mają tę samą linię wyboru wiersza.

bufor danych we/wy i sterowanie

[*]

diagram logiczny pamięci SRAM 32k x 8



Bufory trójstanowe (symbol - odwrócony trójkąt) pamięci pozwalają liniom danych magistrali danych działać jako linie wejściowe lub wyjściowe i łączyć pamięć z szyną danych w komputerze.

[*]

W trybie READ osiem bitów danych, które są przechowywane pod wybranym adresem, pojawia się w kontaktach wyjściowych.

W trybie WRITE osiem bitów danych, które są na kontaktach wejściowych jest wprowadzanych do pamięci pod wybrany adres.

Pamięci statyczne RAM zajmują stosunkowo dużo miejsca w układzie scalonym, zatem osiągnięcie dużych pojemności tych pamięci okupione jest dużymi rozmiarami układu.

Pamięci SRAM są najszybsze, więc istnieje dylemat godzenia niewielkiej możliwej do uzyskania skali integracji z możliwymi osiągalnymi parametrami dynamicznymi.

pamięci dynamiczne o dostępie swobodnym

DRAM

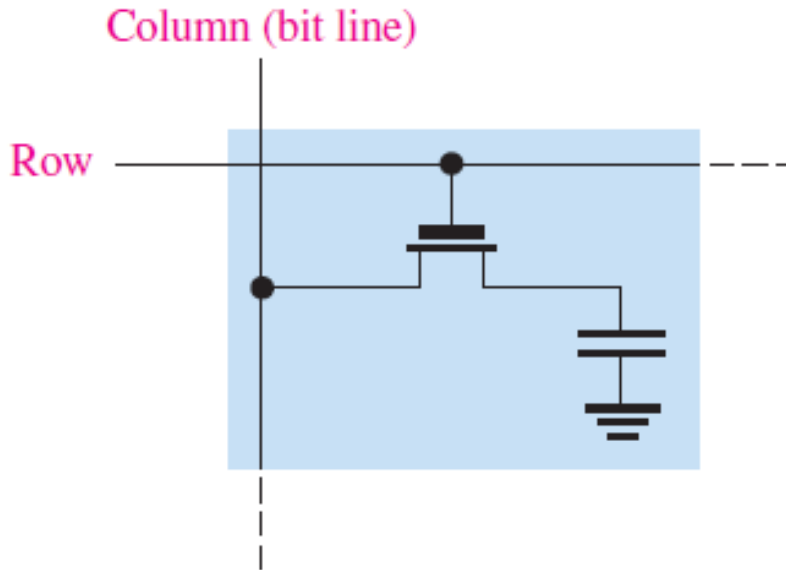
W przeliczeniu na jeden bit pamięci najmniejszą powierzchnię układu scalonego zajmują komórki pamięci dynamicznej RAM.

Mają one najmniejszą moc strat i charakteryzują się małym kosztem wytwarzania.

Komórka pamięci dynamicznej DRAM

Zawiera tylko dwa elementy: tranzystor MOS i kondensator.
 Pojemność kondensatora jest bardzo mała (ok. $50 \text{ fF} = 50 \times 10^{-15} \text{ F}$)

Tranzystor MOS stanowi przełącznik łączący lub odłączający kondensator od linii bitu(kolumny).



[*]

Pamiętanie bitu informacji wynika z dwóch stanów ładunku zmagazynowanego w kondensatorze.

Pamiętanie bitu informacji wynika z dwóch stanów ładunku zmagazynowanego w kondensatorze.

Stan 0 (kondensator rozładowany) jest stanem trwałym w czasie, a stan 1 (kondensator naładowany) jest stanem nietrwałym, gdyż kondensator ulega stopniowemu rozładowaniu wskutek nieuniknionych upływności.

Aby nie dopuścić do rozładowania kondensatora i tym samym do przejścia ze stanu 1 w stan 0, ładunek istniejący w kondensatorze w stanie 1 musi być co jakiś czas (zazwyczaj co kilka milisekund) regenerowany w procesie zwanym odświeżaniem.

Odświeżanie wymaga dodatkowych obwodów pamięci i komplikuje działanie pamięci DRAM.

Główne ograniczenia szybkości działania układów DRAM:

- odczyt stanu kondensatora w komórce DRAM jest destrukcyjny, co wymaga czasu na powtórny zapis z odświeżeniem po odczycie;
- potrzebny jest czas na proces ładowania wstępnego linii bitu;
- adres w pamięciach DRAM jest ustalany w czasie kolejno w dwu etapach (multipleksowanie adresu);
- niezbędne jest periodyczne odświeżanie zawartości całej matrycy pamięciowej, co wymaga pewnego czasu.

Pamięci DRAM są wolniejsze od pamięci statycznych, ale koszt ich wytworzenia (w przeliczeniu na jeden bit pojemności) jest niższy.

Maksymalna pojemność pamięci dynamicznych (wytwarzanych przy tym samym poziomie technologicznym) jest kilkakrotnie wyższa w porównaniu do statycznych.

Dlatego pamięci DRAM są chętnie stosowane w typowych zastosowaniach komputerowych.

organizacja systemów cyfrowych w kontekście umiejscowienia wykorzystania pamięci SRAM i DRAM

PAMIĘĆ PODRĘCZNA

Procesor komputera jest bardzo szybki i cały czas przetwarza dane pobierane z pamięci. Z uwagi na różnice szybkości działania często oczekuje na dostarczenie danych ponieważ czasy dostępu do pamięci są dłuższe niż możliwość ich pobierania przez procesor.

Celem zminimalizowania tego niekorzystnego zjawiska opracowano technikę łączenia szybkiej pamięci o małej pojemności z wolniejszą pamięcią o dużej pojemności.

Udało się w ten sposób pogodzić względnie dużą wydajność pamięci z jej względnie dużą pojemnością i to wszystko za umiarkowaną cenę.

Powstała w ten sposób **pamięć podręczna** (*cache'owa*)

(*cache – rzecz. kryjówka, czas. chować*)

zasada lokalności odwołań

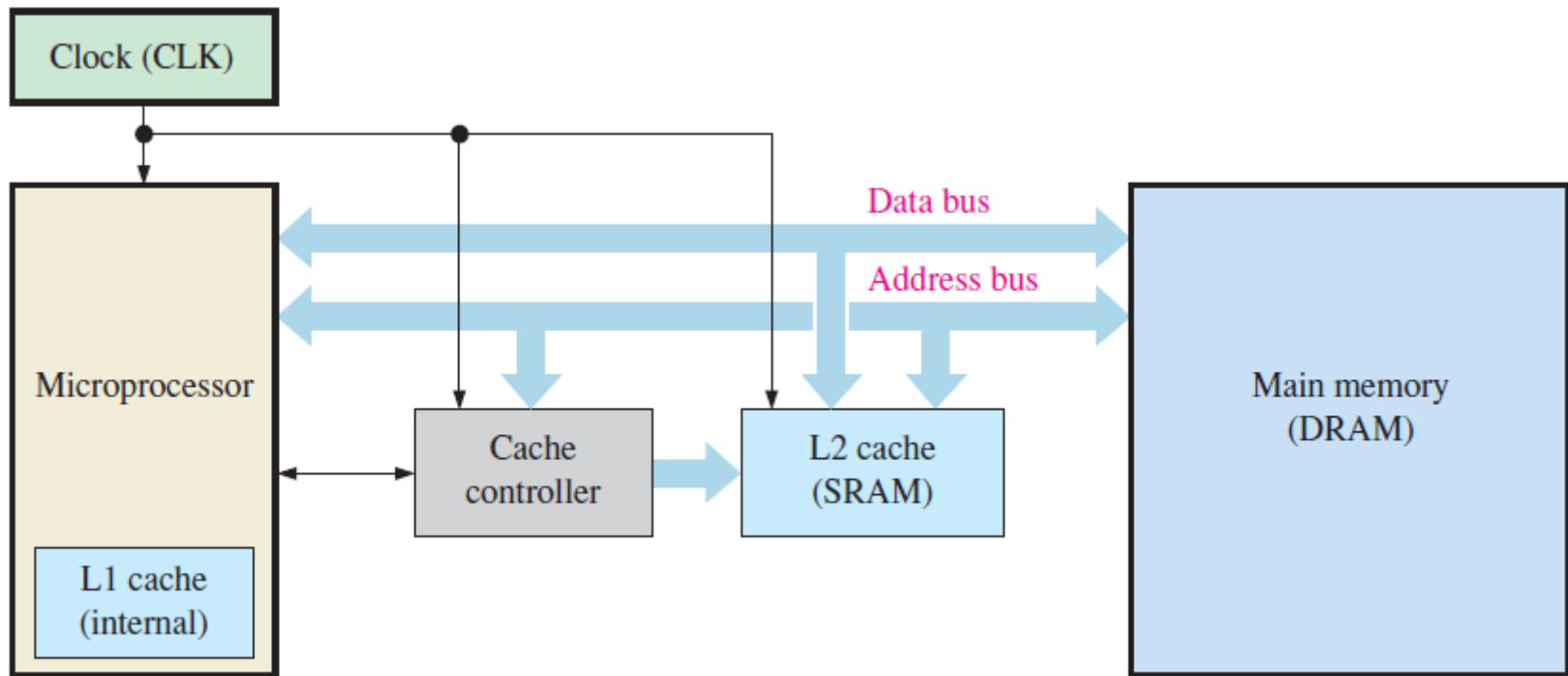
Odwołania procesora do pamięci pojawiające się w krótkim czasie wykorzystują niewielką część przestrzeni adresowej. Istnieje duże prawdopodobieństwo, że każda kolejna operacja dostępu do danych w pamięci będzie operacją poboru danych z lokalizacji sąsiadującej z poprzednio wykorzystywaną.

Jeżeli pamięć podręczna zawiera więcej słów z otoczenia poprzednio wykorzystywanego adresu to można oczekiwać, że kolejne wywołanie będzie dotyczyło informacji przebywającej już w pamięci podręcznej.

Pojemność pamięci podręcznej pierwszego poziomu L1 jest mniejsza. Pamięć ta znajduje się we wnętrzu układu scalonego procesora. Pamięć podręczna drugiego poziomu L2 stanowi osobny układ scalony, ma większą pojemność i pośredniczy pomiędzy procesorem a pamięcią główną.

Pamięć podręczna wykonana jest w technice SRAM.

Pamięć główna wykonana jest w technice DRAM.



[*]

umieszczenie pamięci podręcznej w komputerze

etapy dystrybucja żywności jako analogia wykorzystania pamięci podręcznej

operacja poboru danych a operacja poboru żywności



pamięć podręczna L1



pamięć podręczna L2



pamięć główna

ULOTNE PAMIĘCI PÓŁPRZEWODNIKOWE

KONIEC

[*] Floyd T. L.: Digital Fundamentals. PEARSON