

# Budowa i sposób działania pamięci półprzewodnikowych

## *Rodzaje pamięci*

**Pamięci półprzewodnikowe** są jednym z ważniejszych elementów systemów cyfrowych. Są to układy scalone służące do przechowywania większych ilości informacji w postaci cyfrowej (binarnej). Ilość informacji, które mogą przechowywać pojedyncze układy scalone pamięci, zawiera się w zakresie od kilobajtów do dziesiątków megabajtów. Podstawowymi parametrami charakteryzującymi dany rodzaj pamięci są pojemność oraz czas dostępu.

**Pojemnością pamięci** nazywamy maksymalną ilość informacji, jaką możemy przechowywać w danej pamięci. Pojemność pamięci podaje się w bitach (b) lub bajtach (B).

**Czasem dostępu** do pamięci w przypadku operacji odczytu nazywamy czas, jaki musi upłynąć od momentu podania poprawnego adresu odczytywanego słowa w pamięci do czasu ustalenia się poprawnej wartości tego słowa na wyjściu z pamięci, lub w przypadku operacji zapisu - czas jaki upłynie do momentu zapisania wartości do tego słowa z wejścia pamięci.

W technice komputerowej praktycznie używa się pamięci półprzewodnikowych o dostępie swobodnym (w odróżnieniu np. od dostępu sekwencyjnego w pamięciach recyrkulacyjnych).

**Pamięcią o dostępie swobodnym** nazywamy pamięć, dla której czas dostępu praktycznie nie zależy od adresu słowa w pamięci, czyli od miejsca w którym przechowywana jest informacja. Ze względu na własności użytkowe pamięci możemy podzielić na **RAM** i **ROM**.

**Pamięcią RAM** nazywamy pamięć półprzewodnikową o dostępie swobodnym przeznaczoną do zapisu i odczytu. Pamięć Ram jest pamięcią ulotną, co oznacza, że po wyłączeniu zasilania informacja przechowywana w pamięci jest tracona.

**Pamięcią ROM** nazywamy pamięć półprzewodnikową o dostępie swobodnym przeznaczoną tylko do odczytu. Pamięć ROM jest pamięcią nieulotną

Z podanych własności pamięci wynikają ich zastosowania w technice komputerowej. Z pamięci RAM buduje się tzw. pamięć operacyjną komputera, przeznaczoną do przechowywania w trakcie pracy systemu danych oraz programów. W pamięci ROM przechowuje się programy inicjujące pracę komputera, gdyż muszą być one przechowywane w pamięci nieulotnej.

Ze względu na technologie wykonania, pamięci RAM dzielimy na dwie podstawowe grupy:

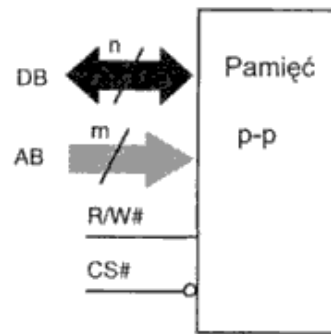
- **pamięci dynamiczne - DRAM,**
- **pamięci statyczne - SRAM.**

Pomiędzy tymi dwoma grupami pamięci występują istotne różnice w ich parametrach i własnościach użytkowych. Pamięci dynamiczne są pamięciami wolniejszymi od pamięci statycznych, są natomiast znacznie tańsze. Ponadto pamięci dynamiczne znacznie łatwiej podlegają scalaniu, co oznacza, że dla porównywalnej wielkości układu uzyskujemy w nich znacznie większe pojemności. Wadą pamięci dynamicznych jest fakt, że dla ich poprawnego funkcjonowania konieczny jest tak zwany proces **odświeżania**. Polega ona na cyklicznym, ponownym zapisie przechowywanej informacji do komórek tej pamięci. Z porównania własności tych pamięci wynika także miejsce ich zastosowania w technice komputerowej. Pamięci dynamiczne stosowane są do budowy głównej pamięci operacyjnej komputera, co wynika z ich niskiej ceny i dużych pojemności układów scalonych tej pamięci. Wadą tych pamięci w porównaniu z pamięciami statycznymi jest przede wszystkim ich szybkość działania. Jednakże ze względów ekonomicznych technologicznych (mniejszy stopień scalenia) nie buduje się pamięci operacyjnej z pamięci statycznych. Stosuje się natomiast w komputerach tzw. pamięć podręczną cache, o znacznie mniejszej pojemności w porównaniu do pamięci operacyjnej, zbudowaną z szybkich układów pamięci statycznych.

## Organizacja pamięci

Podstawowe wyprowadzenia układu pamięci półprzewodnikowej pokazane są na rysunku poniżej.

Szyna wejścia/wyjścia danych (DB) służy do wprowadzania i wyprowadzania informacji do i z pamięci. Wejście adresowe AB służy do dokonania wyboru, na którym z wielu słów w pamięci zostanie wykonana operacja (zapisu lub odczytu). Wejście sterujące R/W# informuje układ pamięci, jakiego rodzaju operacja będzie wykonywana: odczyt lub zapis. Wreszcie wejście CS# służy do uaktywnienia układu pamięci. Wejście to jest używane przy budowie zespołów pamięci metodą łączenia dwóch lub więcej układów scalonych pamięci.



Rysunek 1. Podstawowe wyprowadzenia układu scalonego pamięci.

**Adresem** nazywamy niepowtarzalną liczbę

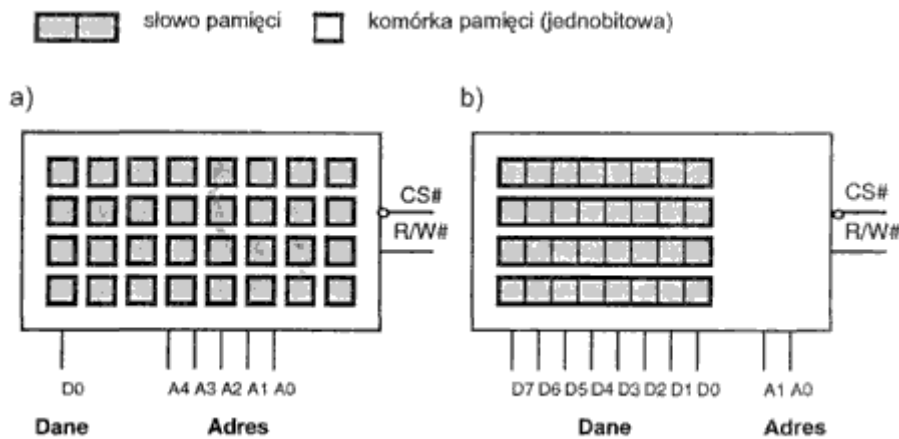
(numer) przypisaną danemu miejscu (słowu) w pamięci w celu jego identyfikacji.

**Słowem w pamięci** nazywamy zestaw pojedynczych adresów komórek pamięci, do którego odwołujemy się pojedynczym adresem.

Ilość bitów w pojedynczym słowie pamięci będziemy nazywać **długością słowa pamięci**. Warto zauważyć, że długość słowa pamięci musi być równa ilości wyprowadzeń szyny wejścia/wyjścia, gdyż słowa są wprowadzane i wyprowadzane z pamięci równolegle.

**Organizacją pamięci** nazywamy sposób podziału obszaru pamięci na słowa.

Pojęcie organizacji pamięci najprościej wyjaśnić na przykładzie. Pamięci narysowane symbolicznie na rysunku 2 a i b mają tę samą pojemność wynoszącą 32b, różnią się natomiast organizacją. Pamięć z rysunku a ma organizację bitową (32 x 1b). Pamięć z rysunku b ma organizację bajtową (4 x 8b lub inaczej 4 x 1B). Przy okazji należy zwrócić uwagę na ilość linii danych i adresowych dla każdej z tych pamięci.



Rysunek 2. Interpretacja organizacji pamięci.

## Łączenie układów pamięci

Budowa bloków (banków) pamięci polega na łączeniu układów scalonych pamięci o określonej pojemności i organizacji w ten sposób, aby uzyskać zespoły pamięci o większej pojemności i/lub o zmienionej długości słowa. Stąd problem rozbudowy pamięci możemy podzielić na dwa podstawowe przypadki:

- zwiększanie (rozszerzanie) długości słowa przy niezmienionej ilości słów,
- zwiększanie ilości słów przy niezmienionej długości słowa.

### *Zwiększanie długości słowa*

W celu zwiększenia długości słowa pamięci szerszą magistralę danych budujemy z bitów linii danych kolejnych układów scalonych pamięci, natomiast magistralę adresową i sygnały sterujące łączymy równolegle. Połączenie równoległe wejść adresowych oznacza, że we wszystkich układach, z których budujemy blok o większej długości słowa, wybieramy słowa położone w takim samym miejscu. Nie ma żadnego powodu, aby robić inaczej, gdyż jest to rozwiązanie najprostsze. Podobnie z sygnałami sterującymi. Musimy uaktywnić wszystkie układy scalone przechowujące słowa składowe tworzące słowo o zwiększonej długości, stąd równoległe połączenie sygnałów CS#. I wreszcie na wszystkich słowach składowych wykonujemy tę samą operację, zapis lub odczyt, co wymaga równoległego połączenia sygnałów R/W#. *Zwiększanie ilości słów w pamięci*

Zwiększenie ilości słów pamięci oznacza zwiększenie ilości potrzebnych adresów, a co za tym idzie - rozbudowę szyny adresowej o dodatkowe bity potrzebne do uzyskania tych adresów. Przy niezmiętej długości słowa szyna danych pozostaje bez zmian. Dodatkowe bity adresu służą, przy wykorzystaniu dekodera, do wyboru jednego z łączonych układów pamięci, z którego odczytamy lub do którego zapiszemy informację. Wyboru dokonujemy przy użyciu wejścia CS# uaktywniającego układy scalone pamięci. Magistrale adresowe, danych i sygnały sterujące układów, z których budujemy nowy blok pamięci, łączymy równolegle.

### *Rozbudowa pamięci w komputerze*

Moduły pamięci instalowane są w systemie w tzw. bankach.

**Bank pamięci** jest to układ pamięci, których łączna długość słowa wyrażona w bitach jest równa szerokości magistrali danych procesora. Wszystkie moduły wchodzące w skład banku muszą być jednakowego typu.

### *Wykorzystanie pamięci operacyjnej przez komputer*

„Oryginalny” komputer PC posiadał 1 MB pamięci podstawowej, z czego górna część tej pamięci (pomiędzy 640 KB a 1024 KB) były zarezerwowane dla sprzętu. W związku z tym wyznaczono niejako granicę **pamięci podstawowej** na poziomie **640 KB**. Komputer wyposażony w pamięć o większej pojemności niż 1 MB traktuje tą „nadmiarową” pamięć jako **pamięć dodatkową (extended memory)**.

Wraz z pojawieniem się komputerów opartych na procesorze 80286 stało się możliwe wykorzystanie obszaru pamięci o adresach wyższych niż 1MB. W początkowym okresie niektóre programy wykorzystujące ten rodzaj pamięci opierały swoje działanie na nowej funkcji BIOS, tzw. INT 15h. Nie było to jednak rozwiązanie doskonałe, dlatego rozpowszechnił się protokół dostępu do **pamięci Extended** zwany **XMS (eXtended Memory Specification)**. Obecnie terminu XMS używa się w odniesieniu do całej pamięci Extended, choć nie jest to formalnie poprawne.

Pierwszy 64-kilobajtowy (bez 16 bajtów) blok pamięci Extended stanowi tzw. obszar **pamięci wysokiej HMA** i jest dostępny dla systemów operacyjnych opartych na jądrze DOS'a, gdy uruchomiony jest sterownik HIMEM.SYS. Obecnie ten rodzaj pamięci ma zdecydowanie drugorzędne znaczenie.

**Pamięć dodatkowa (Expanded, EMS)** jest przeznaczona głównie dla przechowywania danych. W początkowym okresie pamięć ta była dostępna tylko w postaci dodatkowych kart i dla komputerów XT stanowiła jedyną możliwość wykorzystania większej niż 640 kB wielkość pamięci. Ponieważ zasady dostępu do pamięci Expanded sformułowane zostały w specyfikacji **Expanded Memory Specification (EMS)** więc określenie EMS stanowi synonim pamięci Expanded. Obecnie pamięć EMS ma drugorzędne znaczenie.

**Pamięć zarezerwowana UMB** jest to pamięć adresów pomiędzy 640 kB a 1MB. Jest to rodzaj pamięci ściśle związany z procesorami 386/486. Warunkiem wykorzystania tego rodzaju pamięci jest załadowanie tzw. sterownika pamięci wirtualnej (np. dla DOS'a jest to EMM386).

## **Pamięć ROM**

Pamięć **ROM (Read Only Memory)** jest pamięcią nieulotną, przeznaczona tylko do odczytu. Nieulotność oznacza, że po wyłączeniu napięcia zasilania tej pamięci, informacja w niej przechowywana nie jest tracona (zapominana). Określenie, że jest to pamięć 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. Określenie „tylko do odczytu” oznacza, że do pamięci tej nie możemy zapisywać danych w trakcie jej normalnej pracy w systemie.

Podział pamięci ROM oparty jest przede wszystkim na własnościach użytkowych tych pamięci, choć niewątpliwie ma to związek z zasadą ich działania i technologią wykonania. Podstawowymi typami pamięci ROM są:

- **MROM (Mascable ROM)** - pamięci, których zawartość jest ustalana w procesie produkcji przez wykonanie odpowiednich masek (stąd nazwa) i nie może być zmieniana. Przy założeniu realizacji długich serii produkcyjnych jest to najstarszy rodzaj pamięci ROM. W technice komputerowej dobrym przykładem zastosowania tego typu pamięci jest BIOS obsługujący klawiaturę.
- **PROM (Programmable ROM)** - pamięć jednokrotnie programowalna. Oznacza to, że użytkownik może sam wprowadzić zawartość tej pamięci, jednakże potem nie można jej już zmieniać. Cecha ta wynika z faktu, że programowanie tej pamięci polega na nieodwracalnym niszczeniu niektórych połączeń wewnątrz niej. Obecnie ten typ pamięci nie jest już używany.
- **EPROM** - pamięć wielokrotnie programowalna, przy czym kasowanie poprzedniej zawartości tej pamięci odbywa się drogą naświetlania promieniami UV. Programowanie i kasowanie zawartości tej pamięci odbywa się poza systemem, w urządzeniach zwanych odpowiednio kasownikami i programatorami pamięci EPROM. Pamięć ta wychodzi już z użycia.
- **EEPROM** - pamięć kasowana i programowana na drodze czysto elektrycznej. Istnieje możliwość wprowadzenia zawartości tego typu pamięci bez wymontowywania jej z systemu, choć czas zapisu informacji jest nieporównywalnie dłuższy niż czas zapisu do pamięci RAM. W tego typu pamięci przechowywany jest tak zwany Flash-BIOS czyli oprogramowanie BIOS, które może być uaktualniane przez wprowadzanie jego nowej wersji.

Pewną odmianą pamięci związaną z pamięciami ROM, choć nie należącą ściśle do tej grupy, jest pamięć **NVRAM**. Stanowi ona połączenie pamięci SRAM z pamięcią EEPROM. Pamięć NVRAM może być odczytywana i zapisywana. Wprowadzona informacja może jednak zostać przepisana do pamięci typu EEPROM. Zapewnia to zachowanie zawartości tej pamięci po wyłączeniu napięcia zasilania. Czas zapisu do pamięci EEPROM jest oczywiście znacznie dłuższy niż czas dostępu do pamięci SRAM (rzędu kilkunastu ms). Nie jest to jednak istotne, gdyż przepisanie zawartości pamięci SRAM do EEPROM nie następuje po każdym zapisie do pamięci SRAM, a jedynie na „żądanie”. Przykładem zastosowania tych pamięci może być przechowywanie parametrów konfiguracji urządzeń wprowadzonych w trakcie danej sesji pracy z urządzeniem, które chcemy zachować w celu ich użycia w kolejnych sesjach.