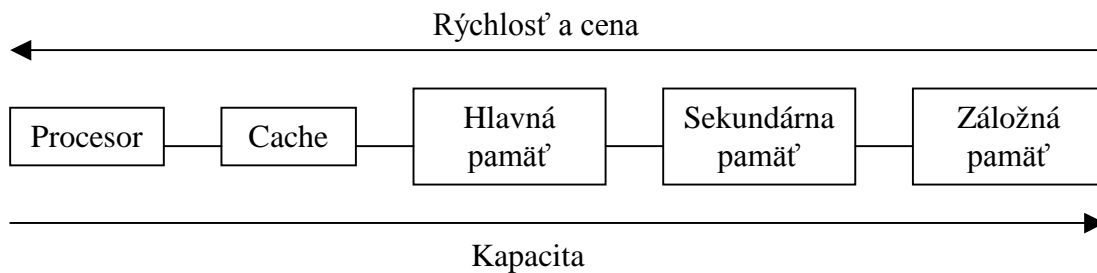


Pamäť

Pamäť je jednou zo základných častí číslicového počítača. Slúži na uchovávanie programov a údajov, s ktorými programy pracujú. Ideálna pamäť má vysokú kapacitu, je rýchla (čo najrýchlejšia – aby zbytočne nespomaľovala prácu procesora) a pritom je cenovo dostupná. Tieto požiadavky sú navzájom protichodné a zatiaľ ich nie je možné všetky splniť súčasne. Preto sa v číslicových počítačoch používa niekoľko druhov pamätí, ktoré tvoria hierarchický systém:



Pamäť v procesore je tvorená registrami, ktoré procesor obsahuje. Je to najrýchlejšia pamäť, pretože je možné s ňou pracovať tak rýchlo, ako pracuje procesor. Okrem toho odpadá pri práci s ňou potreba použiť zbernicu počítača na prenos údajov. Samozrejme táto pamäť má veľmi malú kapacitu – maximálne niekoľko desiatok registrov.

Cache pamäť slúži ako *vyrovnávací pamäť*, pretože umožňuje vyrovnávanie rýchlosti prenosu údajov medzi procesorom a hlavnou pamäťou. Je teda podstatne rýchlejšia ako hlavná pamäť, ale zároveň je rádovo menšia. Princíp vyrovnávacej pamäte spočíva v tom, že časť údajov sa presunie z hlavnej pamäte do vyrovnávacej a tým sa zvýši rýchlosť prístupu ku informáciám zo strany procesora. Je možné použiť viacero pamätí typu cache, pričom sa líšia rýchlosťou a veľkosťou. Potom hovoríme o vyrovnávacej pamäti prvej, druhej, prípadne tretej úrovne (L1, L2 resp. L3 cache).

Hlavná (primárna) pamäť počítača sa tiež nazýva *operačná pamäť* („ramka“). Tvorí základ pamäťového podsystemu. Obsahuje práve vykonávaný program a spracúvané údaje. Kapacita operačnej pamäte sa v súčasnosti bežne pohybuje v oblasti stoviek megabajtov, prípadne až v oblasti gigabajtov. Tiež je možné sledovať neustálu snahu o zrýchľovanie hlavnej pamäte.

Sekundárna pamäť slúži na uchovanie informácií, ktoré sa momentálne nepoužívajú. Do tejto oblasti patria predovšetkým pevné disky. Táto pamäť tiež slúži na permanentné uloženie údajov, keďže predošlé typy pamätí pri výpadku napájania zvyčajne svoj obsah strácajú.

Záložná pamäť – ako napovedá už jej názov, slúži predovšetkým na archiváciu informácií, prípadne na ich prenos. Sem patria USB kľúče, CD-ROM/RW, DVD-RW/ROM, diskety, magneto-optické disky, v minulosti sa dosť využívali kazetovo-páskové jednotky.

Medzi jednotlivými pamäťami sa prenášajú údaje po blokoch. Veľkosť blokov je v jednotlivých prípadoch rôzna. Pri prenose procesor – cache je to niekoľko bajtov (1 – 4 B), cache – hlavná pamäť: 8 – 128 B, hlavná pamäť – sekundárna pamäť: 512 B. Tieto hodnoty sú samozrejme odlišné pre rôzne počítače.

Celý pamäťový podsystem je riadený pomerne zložitými riadiacimi obvodmi. Ich úlohou je zabezpečiť, aby bol potrebný údaj vždy „poruke“, čiže v cache pamäti (aby sa obmedzilo čakanie procesora na minimum). Ak požadovaný údaj v cache nie je, dochádza ku *chybe bloku* a je potrebné načítať do cache pamäte ten správny blok z hlavnej pamäte.

Podobne, ak sa nenájde žiadaná informácia ani v hlavnej pamäti, dochádza ku *chybe stránky* alebo *chybe segmentu* (podľa toho či sa používa stránkovanie alebo segmentovanie). Ďalšou úlohou, ktorú musí riadiaci systém vykonávať je zabezpečenie zhody dát v rôznych úrovniach pamätí. To znamená, že ak procesor zmení údaj v cache pamäti, je potrebné aby sa daná zmena prejavila aj v hlavnej pamäti (keďže v cache je vlastne len kópia údajov z hlavnej pamäte). Hovoríme, že je potrebné zabezpečiť *koherenciu dát*.

Rozdelenie pamätí

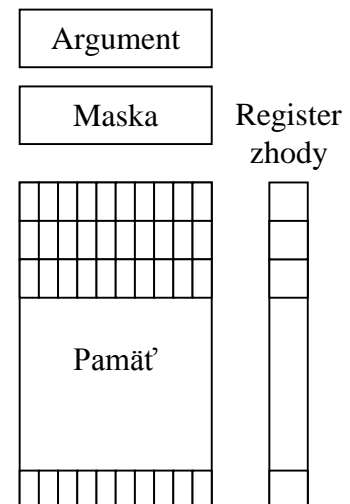
Z historického hľadiska

- kondenzátorové
- feritové
- polovodičové

Podľa spôsobu prístupu k údajom

- pamäte s ľubovoľným (náhodným) prístupom (Random Access Memory – RAM): doba prístupu k údajom nie je závislá na mieste uloženia (na adrese). Sem patria bežné pamäte používané ako cache alebo hlavná pamäť.
- pamäte so sekvenčným (postupným) prístupom (Sequential Access Memory – SAM): hľadaný údaj je možné získať až po systematickom prehladaní pamäte. Doba prístupu k údajom je závislá na mieste uloženia v pamäti. Sekvenčný prístup je typický pre magnetické páskové pamäte.
- pamäte s priamym prístupom (Direct Access Storage Media – DASM): Do tejto skupiny patria pevné disky (HDD), CD-ROM resp. DVD-ROM, magneto-optické disky ... Každé pamäťové miesto je priamo prístupné, ale doba prístupu je závislá na fyzickom umiestnení informácie a na momentálnej polohe čítacieho zariadenia. Pre tieto zariadenia sa udáva ako ich parameter priemerná doba prístupu k údajom (podľa typu pamäte sa pohybuje v jednotkách až stovkách milisekúnd).
- pamäte s asociatívnym prístupom (Content Access Memory – CAM): sprístupnenie pamäťového miesta sa uskutočňuje na základe porovnania časti jeho informačného obsahu so zadaným argumentom. Pamäte CAM je možné považovať za opak pamätí RAM: RAM dáva údaj pre zadanú adresu a CAM poskytuje adresu pre daný údaj. V registri pre argument je uložený hľadaný údaj. Pomocou masky je možné určiť, ktoré bity sa majú brať do úvahy pri prehladávaní pamäte. V registri zhody je zaznačený výsledok hľadania, napr. na tých miestach, ktoré obsahujú údaj vyhovujúci zadaniu bude "1". Register zhody musí obsahovať toľko bitov, koľko slov obsahuje celá pamäť.

CAM pamäť je možné použiť hlavne v aplikáciách, ktoré vyžadujú vysokorýchlostné prehladávanie.



Podľa smeru prístupu

- RWM (Read Write Memory) – pamäte určené pre čítanie a zápis, pričom čítanie aj zápis je rovnako náročné (resp. jednoduché).
- xROM (Read Only Memory) – pamäte určené len na čítanie. Niektoré pamäte označované ako ROM v skutočnosti umožňujú aj zápis, ale operácia zápisu sa v porovnaní s čítaním používa zriedka a okrem toho je značne náročnejšia (či už technicky alebo časovo).

Veľmi často sa pamäte typu RWM označujú ako RAM, hoci skratka RAM nehovorí nič o možnostiach čítania či zápisu do pamäti, ale len o rýchlosti prístupu k danej pamäťovej bunke. V skutočnosti aj pamäte ROM sú väčšinou typu RAM!

xROM pamäte je možné rozdeliť na:

- ROM: tieto pamäte sú určené len na čítanie, ich obsah je určený pri procese výroby.
- PROM (Programmable ROM): túto pamäť je možné jednorázovo naprogramovať, potom je možné už len čítanie. Informácia je uložená v tenkých kovových drôtikoch. Ak je drôtik nezmenený, je tam log.0, ak sa prepáli (a to doslova), je tam 1. Výhodou je stálosť a odolnosť voči vonkajším vplyvom, nevýhodou veľmi veľké rozmery "bunky" a nemožnosť preprogramovania.
- EPROM (Erasable PROM): patrí medzi pamäte, do ktorých je možné opakovane zapisovať. Zápis sa uskutočňuje elektricky pomocou riadeného prierazu (náboj prekoná bariéru izolantu). Náboj je potom uväznený na izolovanom hradle. Mazanie je možné pomocou ultrafialového žiarenia, ktorým sa zvýši vodivosť okolia a náboj vyprchá. EPROM pamäte majú v púzdre okienko umožňujúce prienik UV žiarenia priamo na čip. Počet cyklov mazanie – zápis je obmedzený. Ďalšou nevýhodou je veľká plocha zabraná jednou bunkou a tiež chýlostivosť na žiarenie a teplotu.
- EEPROM (Electrically Erasable PROM): na rozdiel od EPROM pamätí sa aj ich mazanie uskutočňuje pomocou elektrických impulzov. Čas potrebný na zápis informácie je mnohonásobne vyšší ako čas potrebný na jej prečítanie.
- FLASH PROM (alebo FLASH EPROM): je najnovším typom ROM. Je rýchlejšia ako predchádzajúce typy (Flash – blesk). Používa sa napríklad na uchovanie BIOSu v modernejších základných doskách počítačov. Tým je umožnené prepísanie BIOSu inou verziou. Často sa s nimi môžeme stretnúť v podobe USB kľúčov, alebo rôznych pamäťových kariet.

Podľa spôsobu udržania informácie

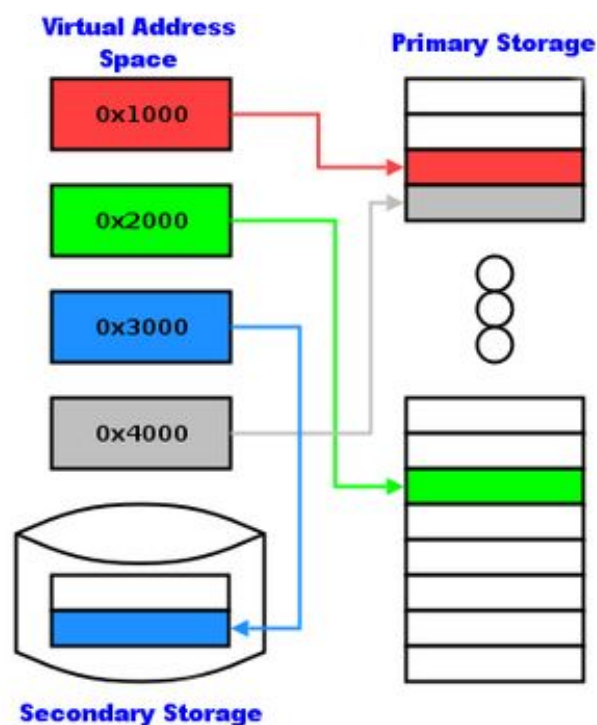
- statické (SRAM): informácia je po zápise zachovaná až do jej prepísania. Ako základný stavebný prvok sa používajú bistabilné klopné obvody, napríklad typu D (pre každý bit jeden). Tieto pamäte sú rýchle, ale pomerne komplikované a tým aj drahé. Výhodou je, že nie je potrebné starať sa o obnovovanie obsahu pamäte. Používajú sa napríklad v cache pamätiach.
- dynamické (DRAM): tieto pamäte využívajú na uloženie jedného bitu informácie parazitné kapacity, v ktorých buď je alebo nie je uložený náboj. Tento náboj sa postupne vybíja, preto je potrebné pravidelne obnovovať obsah pamäte (refresh). Obnovovanie sa uskutočňuje čítaním pamäte. DRAM pamäte sa najčastejšie využívajú ako hlavná pamäť, pretože nie sú príliš drahé a môžu mať pomerne veľkú kapacitu. Ak sú tieto pamäte synchronne, označujú sa SDRAM. Novšie typy, umožňujúce dvojnásobne rýchly prenos informácií sa označujú ako DDR SDRAM (Double Data Rate) alebo len DDRAM.

Podľa energetickej závislosti

- energeticky závislé pamäte: ich obsah sa po prerušení napájania stráca. Sem patria najmä SRAM a DRAM (cache a hlavná pamäť).
- energeticky nezávislé pamäte: obsah pamätí je nezávislý na napájaní. Sú to všetky typy ROM pamätí, pevné disky, CD/DVD-ROM ...

Virtuálna pamäť

V číslicových počítačoch rozlišujeme pamäť na *fyzickú* a *virtuálnu*. Pod fyzickou pamäťou sa myslí operačná (hlavná) pamäť – „ramka“. Pod pojmom „*virtuálna pamäť*“ sa často chápe len „*swapovanie*“ pamäte. V skutočnosti virtuálna pamäť predstavuje taký spôsob menežovania pamäte, kedy operačný systém prezentuje nespojitú časť pamäte ako jeden spojitý celok. To znamená, že aplikácia „*má dojem*“, že jej údaje sú uložené v jednom súvislom bloku pamäte, pričom fyzicky môžu byť údaje uložené na rôznych miestach v pamäti, prípadne dokonca aj v rôznych druhoch pamäti (RAM, disk ...). Swapovanie je len jednou z techník používaných pri práci s virtuálnou pamäťou. Na nasledovnom obrázku je znázornený princíp virtuálnej pamäte – pamäťové bloky s adresami 0x1000, 0x2000, 0x3000 a 0x4000 sú v skutočnosti uložené vo fyzickej pamäti v úplne inom poradí a blok 0x3000 dokonca aj v inom médiu (na disku).



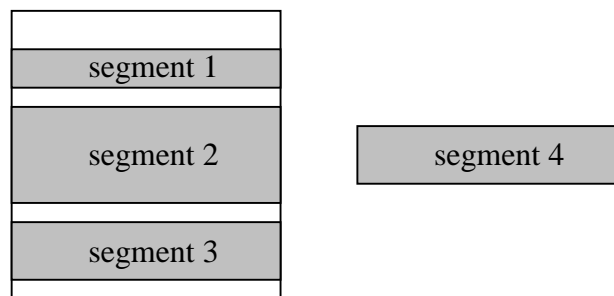
Princíp swapovania je založený na tom, že aktuálny program (resp. dáta, s ktorými pracuje) nemusí byť uložený v hlavnej pamäti celý, ale môže tam byť len jeho časť. Keď je potrebné pracovať s časťou programu, ktorá nie je v hlavnej pamäti, načíta sa zo sekundárnej pamäte (z disku) a pokračuje sa ďalej. Momentálne nepotrebná časť programu sa naopak zapíše na disk. Tento proces je pritom z hľadiska vykonávaného programu transparentný (neviditeľný), takže program si môže „myslieť“, že má k dispozícii veľkú operačnú pamäť (napr. 512 MB), ale pritom jej môže mať relatívne málo (napr. 128 MB). Samozrejme, používanie swapovania spomaľuje prácu počítača, keďže sekundárna pamäť (disk) je podstatne pomalšia než hlavná pamäť.

Organizácia pamäte je v moderných číslicových počítačoch realizovaná pomocou špeciálneho hardvéru (ktorý je často súčasťou procesora) a softvéru nazývaného *menežér pamäte*. V operačnom systéme DOS boli menežéry pamäte samostatné programy ako napr. EMM386, prípadne QEMM. V prípade multitaskingových systémov je menežment pamäte súčasťou operačného systému. Hardvérovo je práca s virtuálnou pamäťou podporovaná napr.

prostredníctvom jednotky MMU (Memory Management Unit), ktorá má na starosti preklad virtuálnych adries na fyzické.

Pri práci s virtuálnou pamäťou sa využívajú dva spôsoby prenosu údajov medzi fyzickou a virtuálnou pamäťou:

- *stránkovanie*: fyzická aj virtuálna pamäť je rozdelená na rovnako veľké oblasti – stránky, pričom stránka je súvislý pamäťový priestor s veľkosťou najčastejšie 4 kB. Výhoda použitia stránkovania spočíva v pevnej veľkosti stránok. Ak je potrebné načítať do hlavnej pamäte novú stránku, stačí zabezpečiť aby bolo pre danú stránku voľné miesto.
- *segmentovanie*: medzi fyzickou a virtuálnou pamäťou sa presúvajú celé segmenty. Segment je súvislá oblasť pamäti, ktorej obsah má určitý logický súvis. Môže sa jednať napríklad o segment programový, dátový, zásobníkový ... Veľkosť segmentov je vo všeobecnosti rôzna. To je aj hlavnou nevýhodou segmentovania. Je potrebné používať určitú stratégiu umiestňovania segmentov do pamäti, aby bola využitá čo najefektívnejšie. Môže sa totiž ľahko stať, že nový segment sa do hlavnej pamäte už nedá načítať, pretože nie je voľný dostatočne veľký súvislý blok pamäte, hoci súčet všetkých voľných miest v pamäti je dostačujúci. Umiestnenie nového segmentu potom vyžaduje odstránenie iného segmentu z pamäte alebo usporiadanie segmentov tak, aby bolo vytvorené súvislé voľné miesto (defragmentácia).



S problémami, ktoré vznikajú pri segmentovaní úzko súvisí alokácia blokov pamäte (alokácia – vyhradenie miesta).

Medzi najčastejšie spôsoby alokácie blokov pamäti patria:

- **BEST FIT** (najpresnejšie umiestnenie): Pri umiestňovaní bloku sa hľadá v pamäti voľné miesto, ktoré je rovnako veľké ako daný blok, alebo najbližšie väčšie. Nevýhodou tejto metódy je, že po umiestnení bloku zvyčajne ostane malé nevyužitú miesto, ktoré sa zvyčajne už nedá využiť na nič rozumné. Táto metóda je dosť náročná na čas, pretože pri každej požiadavke na alokáciu pamäte je potrebné prehľadať všetky voľné bloky a vybrať ten najlepší. Prehľadávanie končí, ak sa nájde rovnako veľké voľné miesto ako je požadované. Samozrejme sa neprehľadáva celá fyzická pamäť, ale len zoznam alokovaných a voľných miest v pamäti.
- **WORST FIT**: Daný blok sa snažíme umiestniť čo „nejpresnejšie“ – hľadáme najväčší voľný blok v pamäti a tam ho umiestnime. Táto metóda využíva predpoklad, že po umiestnení bloku ostane v pamäti ešte dosť voľného miesta na ďalší blok. Nevýhodou je zmenšenie veľkých voľných blokov. Pri požiadavke na umiestnenie veľkého bloku je potrebná defragmentácia. Najvýhodnejšie je udržiavať zoznam voľných miest v pamäti usporiadaný podľa veľkosti.
- **FIRST FIT**: blok uložíme na prvé voľné miesto, ktoré je dostatočne veľké. Táto metóda nemusí prehľadávať celú pamäť a teda je pomerne rýchla.
- **BUDDY FIT**: Pri tomto spôsobe je umožnené alokovanie len blokov určitých povolených veľkostí. Povolená veľkosť je zvyčajne mocninou dvojky. Pre každú povolenú veľkosť

existuje zoznam voľných miest v pamäti. Pri umiestňovaní nového bloku sa hľadá najprv najbližší väčší voľný blok. Ak sa nenájde, blok sa rozdelí na dve časti a hľadá sa umiestnenie pre ne atď. Väčší blok teda môže byť fyzicky rozdelený na viac menších častí. Avšak pri práci s ním sa musí daný blok „tváriť“, akoby bol vcelku. Tým sa ale komplikuje menežment pamäte.

Momentálne najčastejšie využívanou stratégiou alokácie pamäte je CIRCULAR FIRST FIT (tiež nazývaná Next Fit). Je to modifikovaná metóda First Fit tak, že hľadanie voľného bloku začína tam, kde bol naposledy umiestnený blok. Pri metóde First Fit sa najprv použijú prvé voľné bloky. Pri každej požiadavke sa prehľadáva pamäť od začiatku a opakovane sa kontrolujú malé voľné miesta, ktoré zostali po predchádzajúcich alokáciach. Tento nedostatok odstraňuje práve použitie Circular First Fit. Podobne ako Worst Fit má táto metóda tendenciu zaplňovať veľké voľné bloky, čo môže byť problém.

V prípade, že už nie je možné do pamäte vložiť nový blok, je nutné buď pamäť defragmentovať, alebo z pamäte niektorý blok vylúčiť. To je možné tiež urobiť rôznymi spôsobmi:

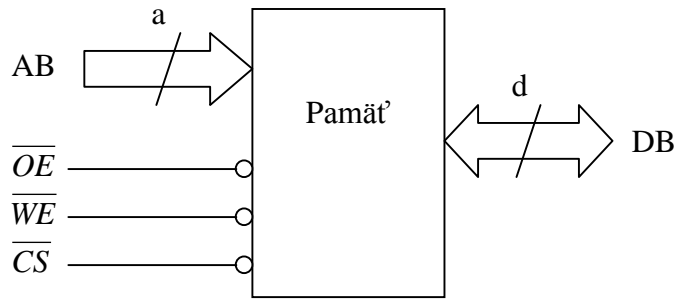
- LRU (Least Recently Used): z pamäte sa vyhodí najdávnejšie používaný blok. Každý blok (či už segment, alebo stránka) musí mať pridelený údaj o poslednom použití. Pri vyhadzovaní je potrebné nájsť blok, ktorý najdlhšie nebol použitý. Táto metóda je najpoužívanejšia.
- LFU (Least Frequently Used): z pamäte sa vylúči blok, ktorý bol používaný najmenej krát. Každý blok pamäte potrebuje počítadlo, koľkokrát bol daný blok použitý.
- FIFO: vyradovanie na princípe zásobníka – prvý dnu, prvý von. Z pamäte sa vyraduje najstarší blok. Pri tomto spôsobe vyradovania stránok z pamäti môže dôjsť ku tzv. Beladyho anomálii. Táto anomália znamená, že zvýšením veľkosti pamäte (a teda aj počtu stránok, ktoré sa do nej zmestia) sa paradoxne môže zvýšiť aj počet chýb, teda dochádza častejšie ku chybe stránky.
- RADOM: vyradí sa náhodne vybraný blok.
- OPT (jasnovidecký algoritmus): z pamäte sa vylúči blok, ktorý bude potrebný najneskôr. Keďže zatiaľ počítače nevedia predpovedať budúcnosť, je táto metóda čisto teoretická a slúži len na porovnanie ostatných metód.

Fyzické usporiadanie pamäte

Pamäte určené pre PC sú v súčasnosti usporiadané do formy pamäťových modulov, ktoré sa skladajú z viacerých pamäťových čipov. Je to z toho dôvodu, že pre výrobcov pamäťových čipov by bolo nemožné vyrábať pamäte všetkých možných kombinácií, najmä čo sa týka kapacity a šírky dátového slova. Preto vyrábajú len niekoľko druhov pamäťových čipov, z ktorých je možné poskladať pamäť takmer ľubovoľných parametrov.

Kapacita pamäťových čipov sa najčastejšie udáva v megabitoch (Mb, $M = 2^{20}$), prípadne gigabitoch (Gb, $G = 2^{30}$). Okrem kapacity sa jednotlivé čipy líšia aj vnútorným usporiadaním, teda šírkou slova a počtom slov. Napríklad 512 Mb čip môže byť usporiadaný ako 128 Mega štvorbitových slov ($128M \times 4$), 64 Mega osembitových slov ($64M \times 8$), prípadne 32 M šestnásťbitových slov ($32M \times 16$). Na vytvorenie pamäťového modulu s kapacitou 512 MB je potom potrebných 8 takýchto čipov.

Hlavné vstupy a výstupy pamäťového čipu sú znázornené na nasledovnom obrázku.



DB (Data Bus) – obojsmerná údajová zbernica so šírkou d bitov

AB (Address Bus) – adresná zbernica so šírkou a bitov

\overline{OE} (Output Enable) – povolenie výstupu údajov z pamäte. Využíva sa pri čítaní z pamäte.

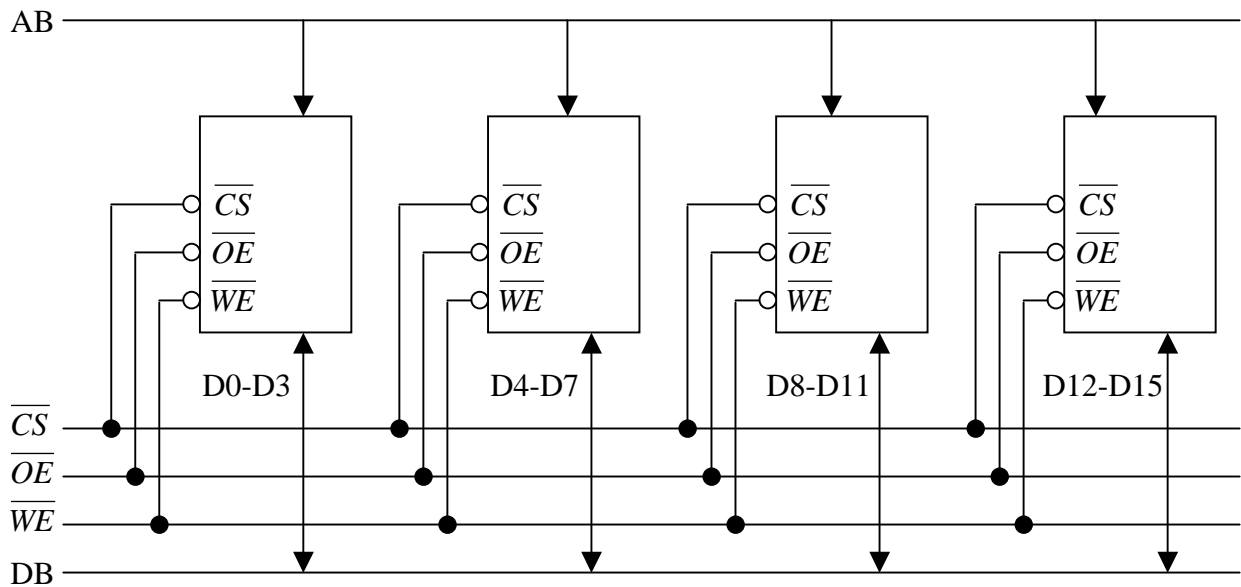
\overline{WE} (Write Enable) – povolenie zápisu do pamäte. Keď je tento signál aktívny (v log.0), zapíšu sa údaje z dátovej zbernice na miesto v pamäti, ktoré je určené adresnou zbernicou.

\overline{CS} (Chip Select) – vstup povolujúci prácu čipu. Čítanie aj zápis do pamäte sa môže uskutočniť iba vtedy, ak je tento vstup v aktívnej úrovni (log.0). Tento vstup umožňuje spájať niekoľko čipov do väčších pamäťových celkov.

Vytváranie pamäťových modulov

Existujú dva spôsoby spájania pamäťových čipov do väčších celkov. Buď sú zapojené čipy tak, aby sa zvýšila kapacita - teda počet slov (sériové zapojenie) alebo sa zvýši šírka slova (paralelné zapojenie). Samozrejme je možné aj kombinovať obidva spôsoby a súčasne zväčšiť šírku slova aj kapacitu.

Zväčšenie šírky slova – vytvorenie modulu 2Kx16 z čipov 2Kx4:



Zväčšenie kapacity – vytvorenie modulu 8Kx4 z čipov 2Kx4:

