

Paměti a paměťový subsystém

FYZIKÁLNÍ PRINCIPY ČINNOSTI, ZÁKLADNÍ TYPY A ORGANIZACE PAMĚŤOVÝCH BUNĚK.....	2
PAMĚŤOVÝ SUBSYSTÉM, HIERARCHIE, OPERAČNÍ PAMĚŤ, PAMĚTI CACHE, VIRTUÁLNÍ PAMĚŤ.....	2
TECHNOLOGIE OPERAČNÍCH SYSTÉMŮ, SPRÁVY PAMĚTI (EVIDENCE, PŘIDĚLOVÁNÍ, OCHRANA).....	2
1.FYZIKÁLNÍ PRINCIPY ČINNOSTI, ZÁKLADNÍ TYPY A ORGANIZACE PAMĚŤOVÝCH BUNĚK.....	3
Co je to paměť?.....	3
PAMĚŤOVÁ BUŇKA.....	3
POLOVODIČOVÁ PAMĚŤ.....	4
DĚLENÍ PAMĚTI.....	4
ZÁKLADNÍ PARAMETRY.....	5
ZÁVISLOST NA NAPÁJENÍ.....	5
PAMĚŤOVÝ SUBSYSTÉM, HIERARCHIE, OPERAČNÍ PAMĚŤ, PAMĚTI CACHE, VIRTUÁLNÍ PAMĚŤ.....	6
HIERARCHICKÉ USPOŘÁDÁNÍ PAMĚTI.....	6
OPERAČNÍ PAMĚŤ.....	6
CACHE.....	6
VIRTUÁLNÍ PAMĚŤ.....	7
OCHRANA PAMĚTI.....	7
TECHNOLOGIE OPERAČNÍCH SYSTÉMŮ, SPRÁVY PAMĚTI.....	8
SPRÁVA PAMĚTI.....	8
OCHRANA PAMĚTI.....	8
METODY EVIDENCE POUŽITÍ OPERAČNÍ PAMĚTI.....	8
METODY ALOKACE PAMĚTI.....	9

Otázky

Fyzikální principy činnosti, základní typy a organizace paměťových buněk

Co je to paměť
Paměťová buňka
Polovodičová paměť
Dělení paměti
Základní parametry
Závislost na napájení

Paměťový subsystém, hierarchie, operační paměť, paměti cache, virtuální paměť

Hierarchické uspořádání paměti
Operační paměť
Cache
Virtuální paměť
Ochrana paměti

Technologie operačních systémů, správy paměti (evidence, přidělování, ochrana)

Správa paměti
Ochrana paměti
Metody evidence použití operační paměti
Metody alokace paměti

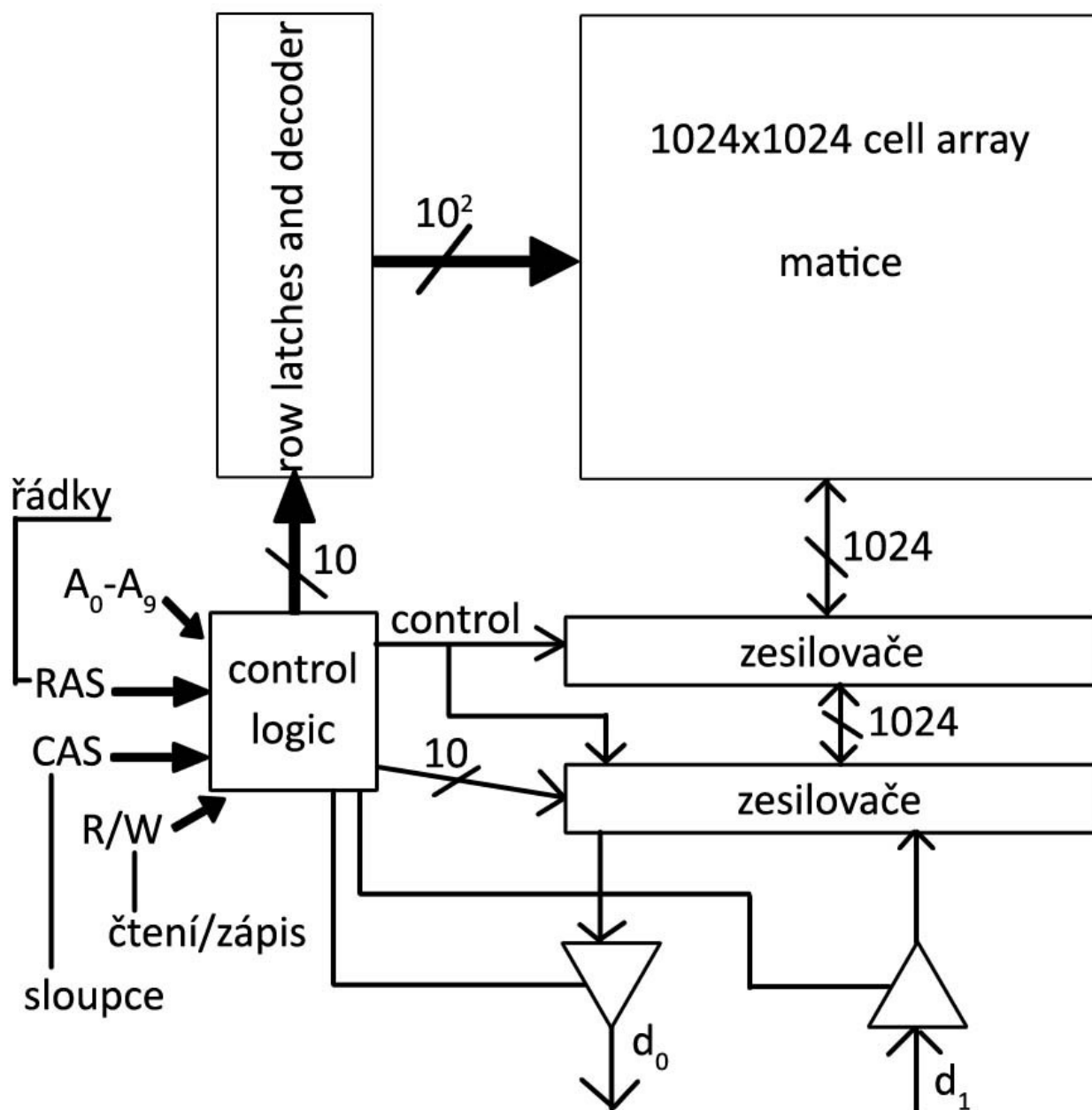
1. Fyzikální principy činnosti, základní typy a organizace paměťových buněk

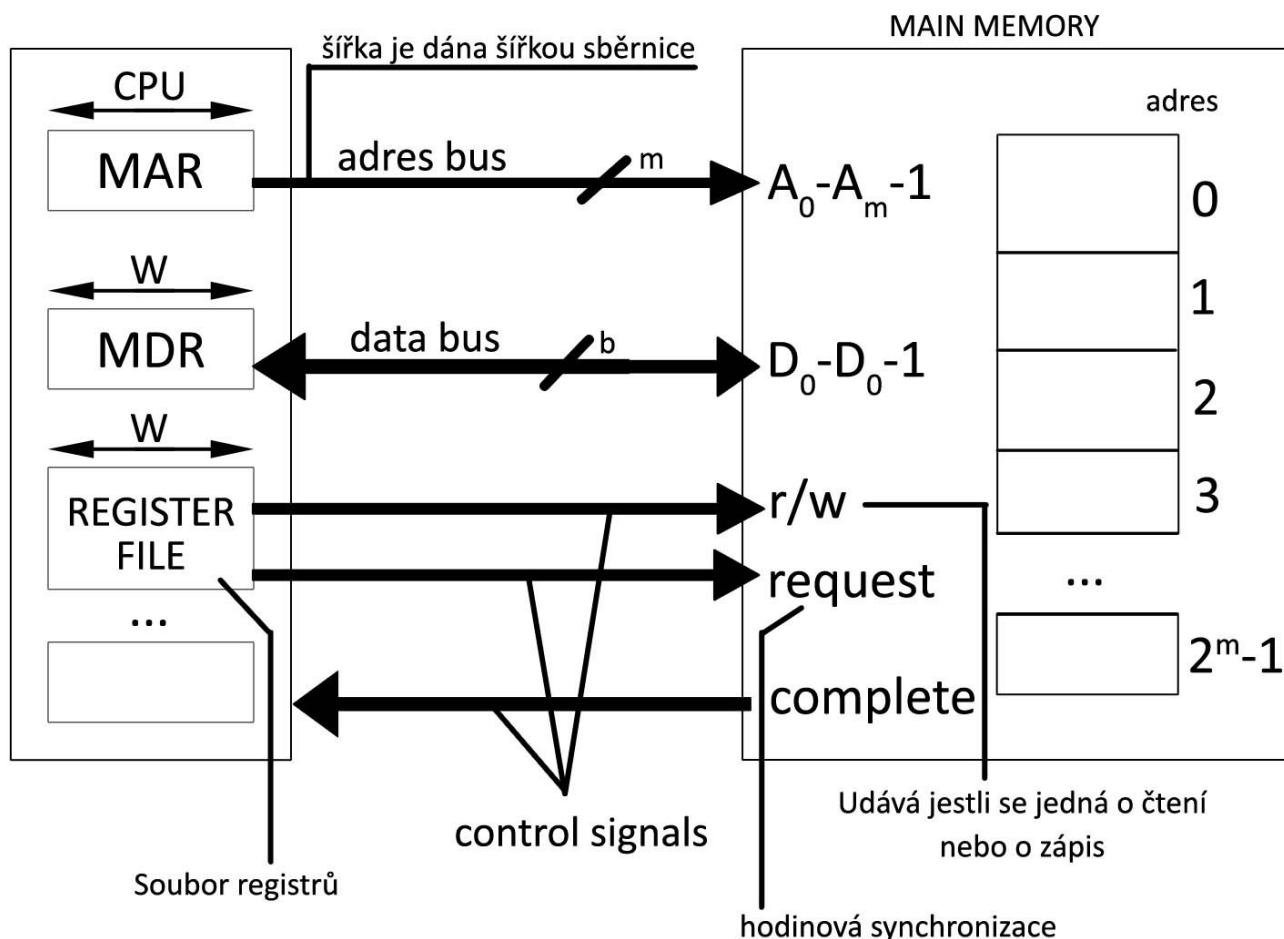
Co je to paměť?

Paměť je zařízení, které je schopné přijmout informace, uchovat je po danou dobu a na požádání je vydat.

Paměťová buňka

Paměťová buňka je množství paměti, které může být najednou adresováno. Šířka paměti musí vždy odpovídat šířce procesorové sběrnice.





Polovodičová paměť

Využívá vlastností polovodičových tranzistorů, buď se realizují klopnými obvody (technologie TTL), nebo obnovováním elektrického náboje (CMOS). Jsou to paměti s nejrychlejším přístupem, jsou součástí procesorů, nebo jsou používány pro cache procesorů.

Režim činnosti polovodičových pamětí

dynamické - informace se musí periodicky obnovovat cyklem čtení, náročnější na řídicí logiku

statické - informace zůstává uchována i bez obnovování, mají vyšší cenu za bit

Dělení pamětí

Podle přístupu k datům

RAM (random access memory) – paměť s libovolným přístupem. Doba odezvy je konstantní a nezávisí na adrese, která je použita pro čtení a zápis. Data jsou po vypnutí napájení vymazána.

SAM (serial access memory) – paměť se sekvenčním přístupem – zpožďovací linky

DAM (direct access memory) – paměť s přímým přístupem (disk)

CAM (content addressable memory) – paměť adresovaná obsahem

Podle implementace

RWM (read/write memory) – paměti pro čtení a zápis na libovolnou adresu

ROM (read only memory) – paměti schopné pouze číst

SRAM (static random access memory) – statické paměti RAM

DRAM (dynamic random access memory) – dynamické paměti RAM

PROM (programmable read only memory) – jednou programovatelné

Základní parametry

SIMM (Single Inline Memory Module) - 72pin, 30pin

DIMM (Dual Inline Memory Module) - 3,3V a 5V. Jedná se o dva moduly SIMM integrované na jedné desce.

Důvodem je obsazení celé šířky sběrnice.

- SDR (Single Data Rate) – paměť typu DIMM (3,3, nebo 5 V), 168 pinů, kapacity od 32 MB do 512 MB, rychlost od 66 MHz do 133 MHz, dva zářezy jako pojistka.
- DDR (Double Data Rate) – paměť typu SDR, 3,3 V, 184pinů (ale jiné umístění zářezů, místo dvou jen jeden), kapacity od 128 do 2048 MB. Vylepšení je v tom, že přenáší data na náběžné i koncové hraně taktovacího impulsu.
- DDR2 – podobné jako DDR, mají vyšší frekvence, stávají se v současné době standardem. Nevýhodou DDR2 jsou vyšší časy latence, než u DDR.
- DDR3 - kvůli vysoké ceně a malé podpoře u motherboardů zatím příliš nerozšířené.

Závislost na napájení

napěťově závislé - pro uchování a přístup k informacím potřebuje paměť napájecí napětí, při jeho zániku zaniká i informace

napěťově nezávislé - potřebuje napětí pro činnost (čtení / zápis), ale při jeho zániku uchovává informaci

Paměťový subsystém, hierarchie, operační paměť, paměti cache, virtuální paměť

Hierarchické uspořádání paměti

vychází ze skutečnosti, že přístupy do paměti mají tendenci shlukovat se do skupin

lokalita referencí:

časová – reference, která byla právě použita a bude v krátké době použita znovu

prostorová – reference následující bude pravděpodobně ležet poblíž právě použité

REGISTRY – vysoká rychlost přístupu – realizace přímo jakou součástí procesoru

CACHE – vyrovnávací paměť – co nejbližší procesoru (10 nanosekund)

HLAVNÍ – fyzická paměť počítače pro instrukce a data (10 – 100 nanosekund)

ODKLÁDACÍ – ukládání stránek hlavní paměti – disk (milisekundy)

SEKUNDÁRNÍ – vnější paměti s náhodným přístupem (disky, diskety)

ARCHIVNÍ – trvalé ukládání velkých objemů dat, ke kterým se přistupuje zřídka

Operační paměť

Operační paměť je nestálá vnitřní paměť, určená pro dočasné uložení zpracovávaných dat a spouštěného programového kódu. Tato paměť má obvykle rychlejší přístup, než vnější paměť. Operační paměť může procesor adresovat přímo. Strojové instrukce jsou adresovány pomocí instrukčního ukazatele a k datům se obvykle přistupuje pomocí adresace prvku paměti hodnotou uloženou v registru procesoru nebo je adresa dat součástí strojové instrukce. Operační paměť je spojena s procesorem pomocí sběrnice, obvykle se mezi procesor a operační paměť vkládá rychlá vyrovnávací paměť typu cache.

Architektura operační paměti

Johnn von Neumannova architektura - paměť je společná pro data i instrukce

Harwardská architektura - existují dvě nezávislé paměti pro data a instrukce

Cache

pro urychlení komunikace s pamětí

rychlá statická paměť

u novějších procesorů velikost stovky kB až MB

více úrovní, přičemž číslo určuje vzdálenost od procesoru

L1 – typicky přímo na procesoru

L2 – například na destičce s procesorem (tzv. boxované procesory)

L3 – na základní desce

write through – data se zapisují ihned (čeká se na dokončení zápisu)

write back – data se zapisují později (na dokončení zápisu se nečeká)

Softwarová cache

Obecně se používá jako vyrovnávací paměť pro pomalé vnější paměti, typickým takovým zařízením je pevný disk počítače. Operační systém se snaží informace, se kterými právě pracuje, uchovávat v rychlé operační paměti a na disk je ukládat co nejpozději. Nemusí tak provádět zbytečné pomalé čtecí a zápisové operace na disku, které jsou o několik řádů pomalejší. Nevýhodou používání cache je riziko spojené s neočekávaným výpadkem napětí - stav datových souborů na disku není vždy aktuální a musí se synchronizovat s obsahem cache.

Hardwarová cache

Paměť cache realizovaná specializovanými paměťovými obvody se používá v některých řídicích jednotkách vnějších pamětí (disky) a v procesorech a jejich podpůrných obvodech.

Cache v řídicích jednotkách vyrovnává rozdíl mezi nepravidelným předáváním dat počítačem (sběrnici) a pravidelným tokem dat do/z magnetických hlav, jehož rytmus je dán rychlostí otáčení disku.

Virtuální paměť

Umožňuje operačnímu systému využívat vnitřní paměť, která je významně větší, než je skutečná fyzická velikost paměti. V danou chvíli se nepotřebná paměť odkládá na pevný disk (nebo na jiný odkládací prostor) a v případě potřeby se opět zavádí do vnitřní paměti.

Výhody

Paměť, kterou má běžící proces k dispozici, není omezena fyzickou velikostí instalované paměti

Každý běžící proces má k dispozici svou vlastní paměťovou oblast, ke které má přístup pouze on sám a nikdo jiný

Paměť jednotlivým procesům lze tak organizovat, že se paměť z hlediska procesu jeví jako lineární, přestože ve skutečnosti může být umístěna na různých místech vnitřní paměti i odkládacího prostoru.

Nevýhody

Při špatném dimenzování reálné operační paměti může dojít ke ztrátě výpočetního výkonu (probíhá neustálá výměna dat mezi pamětí a diskem namísto výpočtů).

Základní principy

Všechny adresy, které proces používá, jsou spravovány pouze jako virtuální - transformaci na fyzické adresy provádí správa virtuální paměti

Existují dvě základní metody implementace virtuální paměti - stránkování a segmentace

- Při stránkování je paměť rozdělena na větší úseky stejné velikosti, které se nazývají stránky. Správa virtuální paměti rozhoduje samostatně o tom, která paměťová stránka bude zavedena do vnitřní paměti a která bude odložena do odkládacího prostoru (swapu).
- Při segmentaci je paměť rozdělena na úseky různé velikosti nazývané segmenty.

Strategie řízení přesunu

FIFO (First in, first out) – nahraď nejstarší stránku. Ta ovšem může být stále používána, tato strategie proto není příliš efektivní. Je ale jednoduchá na implementaci.

LRU (Least Recently Used) – nahraď nejdéle nevyužívanou stránku. Při přesné implementaci vyžaduje náročné udržování informací o používaných stránkách.

LFU (Least Frequently Used) – nahraď nejméně využívanou stránku. Při přesné implementaci vyžaduje náročné udržování informací o používaných stránkách.

Ochrana paměti

Mechanismus řízení přístupu ke specifickým částem operační paměti. Narušení ochrany se projeví jako přerušení, které obsluhuje operační systém. Užívá se v jednouživatelském i víceuživatelském režimu práce.

může být realizováno takto

- mezními registry - obsah MAR se porovnává s limitními hodnotami
- maskou - paměť je rozdělena na stránky, **maska** = číslo stránky
- klíče - prakticky totéž co maska. Srovnání provádí software
- metabity - pomocné bity přidáné k datovému obsahu paměti

Technologie operačních systémů, správa paměti

Správa paměti

Překryvné segmenty

Program se rozdělí na části, které nemusí být najednou v operační paměti a kořenový segment, který je v paměti trvale natahuje překryvné segmenty z vnějšího média dle potřeby.

Rozdělení programu na samostatné nezávislé části je záležitostí programátora

Mechanismus provedení této techniky musí být podporován prostředky programovacího jazyka (*OS tento proces v žádném případě neřídí*)

Relokace adres

Instrukce procesoru identifikují operandy v operační paměti pomocí adres. Číslo, které je přeneseno na adresní část systémové sběrnice v momentě vykonání instrukce je fyzická adresa.

Paměťový adresový prostor je lineární, proces získání fyzické (výpočet efektivní adresy) může být velmi komplikovaný

V momentě překladu programu nejsou hodnoty adres známy, protože se neví, na jaké adrese bude program skutečně vykonáván => veškeré v programu uvedené adresy musí být zapsány takovým způsobem, aby bylo možné při zavedení (vykonání) snadno a rychle vykonávat

Ochrana paměti

Nežádoucí zásah, se rozumí jakákoli neautorizovaná změna paměti, a nezáleží na úmyslném či neúmyslném zásahu.

Mechanismy ochrany paměti

Použití registrů báze a limit

- Při inicializaci programu (procesu) se do registru báze zaneše počáteční adresa přiděleného adresového prostranství a do registru limit délka (velikost) prostranství.
- Výhodou tohoto způsobu ochrany je, že použití báze řeší i problém relokace adres – finální adresa se získá jako součet relativní adresy (od začátku bloku – offset) a báze. Báze + [relativní adresa v programu]
- Nevýhodou je nutnost provést jedno sčítání a jedno porovnání při každém přístupu do paměti

Metoda zámeků a klíčů

- Použita v System360
- Paměť byla rozdělena na bloky po 2kB a každému bloku byl přidělen čtyřbitový ochranný klíč (kód).
- V registru stavového slova programu (PSW) byla čtyřbitová kombinace (zámek). Pokud se program pokoušel přistupovat k bloku, jehož klíč neodpovídal zámku, byl HW cestou zablokován.

Metody evidence použití operační paměti

Metoda bitové mapy

Předpokládá rozdělení paměti na bloky konstantní délky – alokační jednotky.

S každou jednotkou je spojen jeden bit v bitové mapě.

Nastavení bitu signalizuje, že blok je použit, nulová hodnota udává, že je volný.

Volba velikosti alokační jednotky je velmi důležitý parametr. Je-li to jen několik bajtů je bitová mapa velká (spotřebuje hodně paměti). Pokud je alokační jednotka několik kilobajtů je bitová mapa poměrně malá, ale přidělují se často bloky zbytečně velké, což znamená plýtvání paměti.

Jistou nevýhodou této metody je poměrně složitý postup při vyhledávání volné bloku paměti určité velikosti.

Spojový seznam

Je seznam, jehož datové položky jsou v paměti rozptýleny, avšak každá nese informaci, kde je následující položka. Používá se seznam (většinou) tříděný podle adresy.

každá položka obsahuje alespoň následující pole:

- charakteristika bloku
- délka bloku
- adresa následující položky seznamu

třídění podle adresy je výhodné při uvolnění bloku

třídění podle velikosti bloku je výhodné při přidělování podle požadavků

uvolnění bloku realizuje složitější algoritmus – každý blok má dva sousedy (s výjimkou prvního a posledního) a jsou tedy čtyři možné varianty:

- obsazeno na obou stranách, tedy díra
- volno vlevo
- volno vpravo
- volno na obou stranách, tedy vznikne jeden souvislý blok místo tří původních

Metody alokace paměti

Paměť je procesům přidělována podle určitého algoritmu. Správce paměti musí mít přehled o obsazení paměti (bitová mapa, svázané seznamy).

- nejjednodušším postupem je metoda prvního vhodného výběru (First – fit)
- modifikací je následný vhodný výběr (Next – fit)
- jinými postupy jsou metody
 - nejlepšího výběru (Best – fit)
 - nejhoršího výběru (Worst – fit)

Cíl: dosáhnout co nejmenší fragmentace paměti a co nejlepšího využití paměťového prostoru

Jednoduché operační systémy, které zpracovávají v daný moment pouze jednu úlohu, mají i nekomplikovanou správu operační paměti.

Základní funkce všech systémů správy paměti jsou následující:

- udržování přehledu o volné a obsazené paměti
- přidělování paměti podle požadavků
- uvolnění bloku paměti
- poskytování informací o stavu paměti (kolik celkem a kolik volné)