

M14 Zvyšování výkonu a kapacity operační paměti

#technicke_vybaveni_pocitacu

- neustále se nacházejí nové způsoby, jak zvýšit výkon paměti a jiných součástek
- lze dosáhnout zvýšením paměti RAM, upgradu z HDD na SSD, optimalizací systémových nastavení nebo použitím Google Disku jako rozšířenou RAM

Skládání paměťových buněk do celků

- **paměťová buňka**
 - uchovává jednobitovou informaci (0 nebo 1)
 - jsou organizovány do větších struktur
 - malý el. obvod

vyšší kapacitou

- paralelní spojení
 - spojeny do celku pracujícího současně
 - buňky jsou přístupné nezávisle na sebe
 - např. DRAM
- sériové spojení- data jsou přístupná postupně
- Multilevel Cell (MLC) - technologie umožňující uložit více než jeden bit v buňce
- vrstvení
- shlukování - buňky jsou spojeny do celků, které jsou pak adresovány jako větší paměťové zařízení

větším adresovým rozsahem

- zapojením do paralelních sběrnic
- spojením do bank - každá banka má svůj adresový rozsah
- adresováním pomocí více bitů

Paměť cache

- součást, která uchovává často používaná data a tím zrychluje přístup k nim
- od bufferu se liší tím, že data uchovává (buffer je jen přestupní bod)
- je tvořena rychlejší a dražší pamětí → menší velikost (než úložný prostor ke kterému zrychluje přístup)
- lze ji najít
 - hardwarově - v mikroprocesorech, pevných discích; tvořena paměťovými obvody
 - softwarově - v operační paměti; řízena jádrem OS; vytvořená programově
- vynalezena v 1. pol. 60. let 20. st.
- př.: cache webového prohlížeče uchovává objekty (obrázky aj.; neměnné) pro rychlejší načtení při otevření stránky - nestahují se znovu z internetu

softwarová

- obvykle jako vyrovnávací paměť pro pomalé vnější paměti (pevný disk počítače)
- OS se snaží často používané informace ukládat do cache v co nejvýhodnějším pořadí
- je přidělena dynamicky - podle množství volné paměti a potřeb systému
- rizikem je nepředvídatelný výpadek napájení
 - stav datových souborů na disku není vždy aktuální a musí se synchronizovat s obsahem cache
 - proto OS vyžadují před vypnutím proces **shutdown** který korektně ukončí procesy systému a uloží obsah diskového cache do souborů na disku
 - před odpojením je důležité odmountovat vyměnitelná média jinak může dojít k poškození souborového systému

- moderní systém se snaží problém eliminovat zapomocí žurnálů

hardwarová

- v řídicích jednotkách vyrovnává rozdíl mezi nepravidelným předáváním/přebíráním dat sběrnici a pravidelným tokem dat do/z magnetických hlav
- obvod je tvořen z tranzistorů a její funkcí je vyrovnávat rozdílnou rychlost mezi procesorem a operační pamětí
- vyšší rychlostí lze dosáhnout použitím kvalitnějších tranzistorů a položením blíže k procesoru

Organizace

- asociací
 - plně asociativní cache umožňuje každému bloku paměti uložit se na jakékoliv místo v mezipaměti
 - set-associative cache rozděluje mezipaměť na sady, přičemž každý blok paměti může být uložen v jakékoli sadě
 - direktivní cache přiřazuje každý blok paměti do jednoho konkrétního umístění v mezipaměti
- velikostí
 - bloku
 - určuje kolik dat se uloží do jednoho bloku
 - větší bloky mohou zlepšit pravděpodobnost úspěšného vyrovnávání; mohou zvýšit pravděpodobnost kolizí
 - sad
 - určují kolik sad paměti cache bude k dispozici
 - může ovlivnit pravděpodobnost cache hit a miss
- výměna - pokud je cache plná, je určen blok který bude nahrazen novými daty

Modularizace paměti do čipu

- umožňuje vytvořit paměťové systémy s větší kapacitou bez nutnosti externích modulů
- menší latence protože je paměť připojena přímo k procesoru (kratší vzdálenost)
- menší spotřeba energie
- každý modul může být optimalizován pro svůj vlastní účel a mít vlastní správu odvádění tepla

Šířka

slova

- jaké největší číslo dokáže procesor zpracovat během jediné operace/hodinového kmitu
- v počtech bitů
- určuje kolikabitový daný procesor je (8bitový dokáže zpracovat max číslo 0-255; $< 0; 2^n - 1 >$)
- větší čísla musí být rozdělena na menší a musí být zpracována po částech
- vnější - jak velké číslo je procesor schopen vyslat/přijmout na systémovou sběrnici

adresové sběrnice

- určuje kolik bitů může být přeneseno najednou mezi různými částmi systému za jednu přenosovou operaci nebo hodinový cyklus
- např 32bitová sběrnice umožňuje přenos 32 bitů najednou
- širší sběrnice obvykle umožňují rychlejší přenos dat, protože mohou přenášet více bitů najednou
- klíčové při přenosu dat mezi procesorem a pamětí
- paralelní sběrnice přenášejí více bitů současně (například 8, 16, 32 nebo 64 bitů)
- seriálová sběrnice přenáší bity postupně po jednom
- širší sběrnice mohou vyžadovat více fyzických vodičů - větší nároky na fyzický design systému
- přenos více bitů naráz může vyžadovat více energie
- standardní šířky: 8 bitů (*byte*), 16 bitů (*word*), 32 bitů (*double word*) nebo 64 bitů
- výběr šířky sběrnice závisí na konkrétních požadavcích aplikace a vyvážení mezi rychlostí

Latence

- doba kterou trvá procesoru získat data z paměti; interval mezi požadavkem a doručení
- statická paměť má menší latenci než dynamická
- vyšší frekvence a menší přenosová vzdálenost dokážou snížit latenci
- typ paměti má také dopad na latenci (cache má menší než RAM)
- CAS (*Column Address Strobe*) - doba kterou je potřeba čekat před následujícím čtením z operační paměti (asynchronní paměti DRAM v nanosekundách, synchronní SDRAM paměti v cyklech taktovacích hodin)