

Cache – základní dělení (level 1 atd.), vlastnosti; sběrnice USB – princip funkce a parametry

Cache obecně

Cache paměť je dražší rychlá (vyrovnávací¹) paměť s malou kapacitou mezi rychlým zařízením (procesor) a pomalým zařízením (operační paměť, vnější paměť) sloužící k urychlení přenosu dat právě mezi zmíněnými zařízeními. Cache jsou také v HDD, kde nám umožňují *přednačítání* (vnitřní rychlost v HDD je totiž proměnná a cache tedy zajišťuje konstantní výstupní rychlost). Buffery v operační paměti mohou také být nazývány jako cache¹. Velikost (u novějších procesorů) dosahuje stovek kB až MB (L2 nebo L3).

Každá cache má svůj řadič, který řídí čtení a zápis. Jde o paměť RAM, data se tedy nemusí číst v pořadí, v jakém se zapisovala. Většinou paměť SRAM.

Dělení cache pamětí (asociativnost)

- **asociativní** – vyhledávání podle klíče (podle obsahu)
- **neasociativní** – vyhledávání podle adresy

Dělení cache pamětí (adresace)

- **virtuální** – vyhledávání podle logické adresy
- **fyzická** – adresa se nejprve vypočítá a poté se vyhledává

Dělení podle počtu položek na řád (vícecestnost)

- jednocestná cache
- dvoucestná
- čtyřcestná
- osmicestná
- *vícecestná*

Dělení cache (synchronní × asynchronní)

- **synchronní sběrnice** – může být využita pro jiné účely (paměti typu burst, používá se pro velká data); vždy je nastaveno, jak velký blok se bude přenášet; každý klíč se musí vyplnit celým blokem dat
- **asynchronní sběrnice** – sběrnice obsazena po celou dobu přenosu; u cache pamětí málo používané, protože je to u nich nevhodné (zároveň se u nich většinou čte i zapisuje)

¹ **cache** (mezipaměť) vs. **buffer** (vyrovnávací paměť) – cache často také slouží jako buffer a naopak, nicméně cache pracuje na předpokladu, že z ní budou stejná data čtena několikrát, že zapsaná data budou brzy přečtena a že je vysoká šance na spojení zapisovaných nebo čtených dat do jednoho většího bloku.

Dělení cache (vnitřní × vnější/podle vzdálenosti od procesoru – L1, L2, L3)

- **vnitřní** – integrována přímo v pouzdře procesoru; dělí se dále:
 - o **on die** (na jednom čipu – společně s procesorem)
 - o **off die** (na více čípech – na jednom je procesor, na druhém cache)
- **vnější** (mimo procesor) – tři typy:
 - o **L1** (nejrychlejší; typicky uvnitř procesoru, výjimečně může být i vnější, dělí se na instrukční a datovou, slouží k ukládání právě zpracovávaných instrukcí a dat v procesoru)
 - o **L2** (pomalejší než L1; je mezi L1 a L3 = mezi procesorem a operační pamětí; může být umístěna na destičce s procesorem – tzv. *boxované procesory*)
 - o **L3** (nejpomalejší; nejbliže operační paměti; dříve na základní desce, dnes součástí CPU)

L2 a L3 cache obsahují data a instrukce, které procesor aktuálně nepoužívá, ale pravděpodobně bude potřebovat. Různými typy je možné dosáhnout kompromisů mezi cenou a rychlostí paměti.

Princip vyrovnávací paměti cache L2

Je fyzicky umístěna na datové sběrnici mezi procesorem a operační pamětí. Je rozdělena do bloků konstantní velikosti, které se nazývají *cache line*. Velikosti těchto bloků jsou různé, v praxi se pohybují od 8 B do 1 kB. Např. vyrovnávací paměť o kapacitě 512 kB a velikosti bloku 32 kB obsahuje celkem 16 384 bloků.

Požaduje-li procesor načíst určitý obsah paměťové buňky na určité adrese v operační paměti (může se jednat jak o instrukce, tak o zpracovávaná data), provede se kontrola, zda se tato data nachází ve vyrovnávací paměti. Pokud ano, jsou data z vyrovnávací paměti přečtena a předána procesoru mnohem rychleji než v případě čtení dat přímo z operační paměti. Ovšem ve chvíli, kdy potřebná data nejsou nalezena ve vyrovnávací paměti, musí do ní být přenesena. Dojde k uvolnění bloku ve vyrovnávací paměti a data jsou do tohoto bloku načtena (přenáší se vždy celý blok). Dále se pokračuje v přenosu do procesoru.

Dělení cache (podle přístupu)

- **write through** – současně se zapisuje do cache i do operační paměti
- **write back** – data se zapisují do cache a do operační paměti se zapisují buď když je cache plná nebo v určitých nastavených intervalech (desítky nebo stovky ms)

Dělení cache (transparentnost)

- **transparentní** – sledují dění v operační paměti a v ostatních cache pamětech; když se něco změní, prohlásí svá data za neplatná
- **netransparentní** – nesleduje dění, pouze cache 486

Dělení cache (porty)

- **jednoportová** – napojená na jednu sběrnici
- **víceportová** – zvláštní port pro komunikaci s procesorem a zvláštní pro komunikaci s operační pamětí; současně tedy může dělat dva přenosy, nesmí však být ze stejné adresy

Dělení cache (podle obsahu)

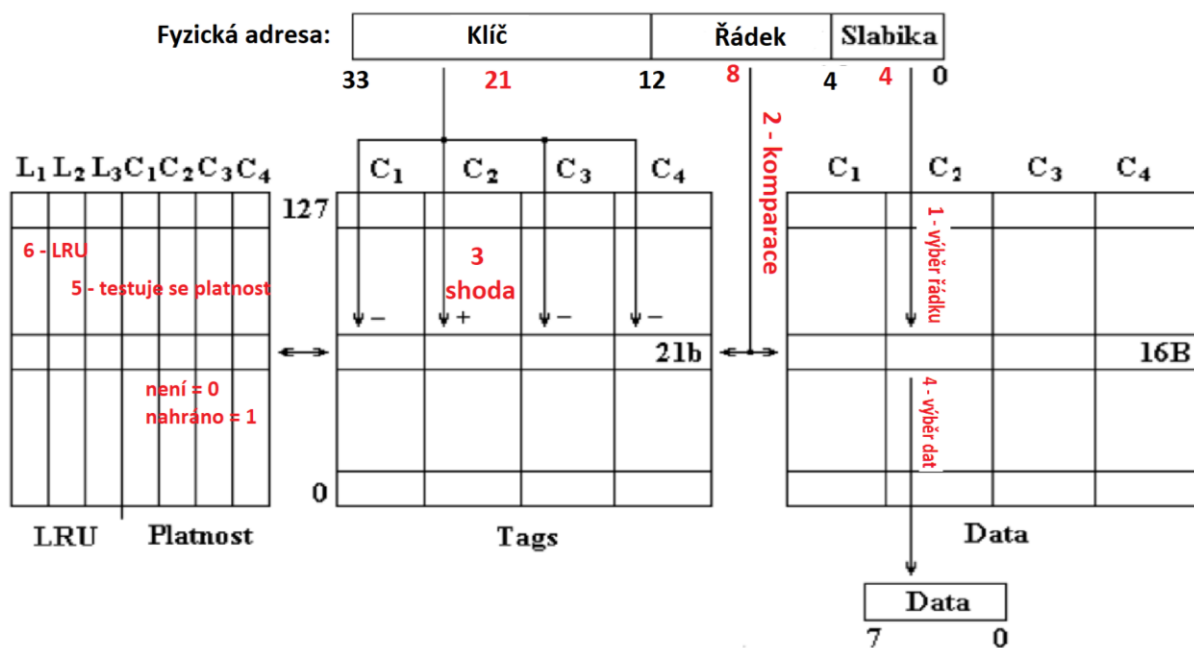
- společná pro data a instrukce
- pro data
- pro instrukce
- pro přerušení
- pro mikroinstrukce

Cache 486

Je čtyřcestně asociativní. Obsahuje 4×128 klíčů. Klíč je obsažen v horních 21 bitech fyzické adresy.

Výběr místa v paměti cache 486 probíhá následovně:

1. adresa řádku je přivedena na dekodér (neasociativní vyhledávání) a vybere se řádek
2. komparace – zadaný klíč je komparátorem porovnán proti čtyřem klíčům na vybraném řádku (asociativní vyhledávání)
3. pokud je prohlášena shoda, provede se výběr dat v datové části paměti (datová část je v každém sloupci 16 B, ze kterých se poté pomocí slabiky – dolních 4 bitů – vybere jeden konkrétní požadovaný bajt)
4. ke každému klíči je přiřazen jeden bit označující platnost dat (validitu)
5. má-li se cache naplnit obsahem fyzické adresy, musí se otestovat, zda je některý z klíčů na určeném řádku neplatný – je-li neplatný, použije se pro zápis
6. je-li platný, musí se algoritmus pseudo-LRU (Last Recently Used) – účelem LRU je najít nejdéle nepoužívanou hodnotu, který je z cache přesunuta do operační paměti, čímž je v cachi získáno místo pro novou položku



USB (Universal Serial Hub)

Externí sběrnice počítače. Nahrazuje dříve používané způsoby připojení (sériový port, paralelní port, PS2, gameport). Zvládá spolehlivou i nespolehlivou komunikaci. Podporuje izochronní přenosy².

Umožňuje k počítači připojovat nejrůznější zařízení, pracuje na principu master-slave. V rámci USB sítě může být až 127 zařízení, podporováno je až 5 úrovní zařízení. Je optimalizované pro různé typy přenosů – malé/velké bloky dat, spolehlivá/nespolehlivá komunikace, proudový přenos v reálném čase, ... Rychlosti jsou od několika Mbps až po několik Gbps (dle verze).

Podpora Plug and Play³, Hot Swap⁴. Podpora více současných operací na několika zařízeních. Levná implementace, žádné licenční poplatky. Podpora Power Management (automatické *uspávání* momentálně nepracujících periferií).

Ve vedení jsou použity i dva vodiče s napájecím napětím 5 V – lze připojit periferii, jejichž odběr nepřekročí 100 mA. Datový tok řídí procesor.

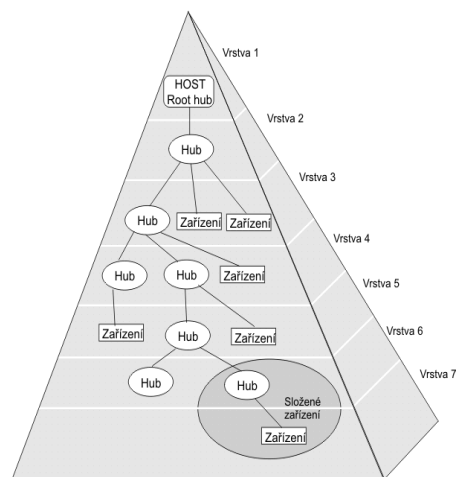
Princip

Používá *vrstvenou hvězdicovou topologii*. Na vrcholku pyramidy stojí hostitel/host/root hub/kořen považovaný za nejvyšší úroveň (většinou vestavěný do základní desky počítače), který řídí podřízená zařízení. Od hostitele vede jediný kabel, který se může větvit k dalším periferiím. Větvení probíhá buď v některé z periferií, nebo jsou ve vedení zařazeny *huby*, v nichž se kabel větví.

Rozhraní mezi USB systémem a počítačem se označuje jako „hostitelský řadič“. S ohledem na zpoždění, která mohou v hubech a kabelech nastat, celá norma udává, že maximálně můžeme mít 7 úrovní včetně kořenové vrstvy. Mezi kořenovým řadičem a koncovým zařízením může být max. 5 hubů. Hostitel je v systému jediný – typické pro PC – ten má roli *master* – řídí datové přenosy v celém systému, rozpoznávání zařízení, hubu, řídí konfiguraci apod.

Každý hub je složen z řadiče, opakovače a převaděče. Hub detekuje připojení, odpojení, může distribuovat a/nebo koncentrovat datový tok, může povolovat nebo zakazovat přenosy na jednotlivé porty, detekuje rychlost (low speed, high speed, ...), umí tyto rychlosti překládat (různá USB mohou pracovat v různých rychlostních režimech).

Koncové zařízení pak přijímá nebo vysílá data a řídicí informace. Může jít o periferii připojenou do hubu nebo složené zařízení (obsahující vestavěný hub). Složené zařízení se pro systém tváří jako hub s jedním nebo více neodpojitelnými koncovými zařízeními. Zařízení mohou být napájena ze sběrnice nebo mohou mít vlastní zdroj.



² **izochronní přenos** = proudový přenos v reálném čase

³ **Plug and Play** – automatická konfigurace připojeného zařízení

⁴ **Hot Swap** – detekce připojení a odpojení za běhu počítače/systému

Přenosové rychlosti

USB není omezeno nejslabším článkem – kde je High Speed probíhá komunikace High Speed, kde je zařízení pomalejší, tam proběhne komunikace v pomalém režimu. Směr přenosu ke koncovému zařízení = *downstream*, opačně *upstream*.

USB 1.1 – Low Speed – 1,5 Mb/s (Full Speed – 12 Mb/s), pro pomalá zařízení (mohou se použít i nestíněné kabely o max. délce 3 m); klávesnice, myši, herní zařízení; half duplex

USB 2.0 – High Speed – 480 Mb/s, stejný kabel, pro videozařízení a zařízení pro ukládání dat (disky USB); half duplex

USB 3.0 a 3.1 – Super Speed – 5 Gb/s a 10 Gb/s; díky Power Managementu nižší spotřeba energie; odlišný konektor, zpětně kompatibilní s USB 2.0

Konektory

A – plochý; vždy pro upstream; nacházejí se na PC na hubu

B – čtvercový; pro downstream; mini-B – zmenšený pro malá zařízení (fotoaparát); micro-B – zejména na mobilních telefonech

C – oboustranný

USB do verze 3 měla 4 vodiče (napájení, data, data, zem), od verze 3 jich má 9 (napájení, napájení, 4× Super Speed, 2× zpětná kompatibilita, zem). Kabel může mít až 5 metrů.

