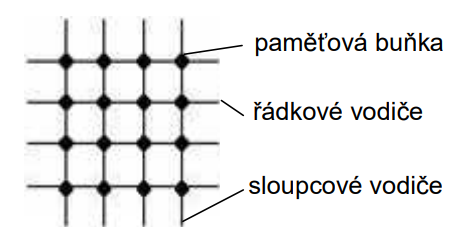
**Úvod**

Polovodičové paměti se skládají z paměťových buněk.

Paměťová buňka je realizována pomocí integrované součástky nebo obvodu, umožňující trvale nebo dočasně vyvolat dva logické stavy 1 nebo 0. Každá základní paměťová buňka má kapacitu 1 bit.

Podle toho, čím je paměťová buňka tvořena se mění vlastnosti polovodičové paměti.

**Struktura**



Paměťové buňky jsou na polovodičovém čipu uspořádány maticově (tvoří jakousi mřížku).

Poloha každé paměťové buňky je určena řádkovým a sloupcovým vodičem.

O nalezení (adresování) příslušné buňky v paměti se stará paměťový řadič, jehož úkolem je také řídit proces čtení, zápisu dat a zabezpečení přenosu dat.

**Dělení podle možností operací**

1. **RWM (Read / Write Memory)** – umožňují libovolné čtení i zápis dat.

Jedná se o paměti energeticky závislé (k udržení informace vyžaduje neustálé napájení).

Paměti s přímým přístupem se označují jako paměti RAM.

Celkově se tato skupina dá označit **RAM (Random Access Memory)** místo RWM

1. **ROM (Read Only Memory)** - jsou určeny především pro čtení zapsané informace.

Jedná se o paměti energeticky nezávislé (dokáže uchovat uloženou informaci bez napájení).

**Paměti RAM (RWM)**

Paměti typu RAM lze rozdělit podle obvodu, který tvoří paměťovou buňku (reprezentace logických hodnot 0 a 1)

**Statická paměť RAM (SRAM - Static Random Access Memory)**

Tato buňka je velice rychlá, vyžaduje menší proud než paměť dynamická, ale fyzicky zabírá na polovodičovém čipu paměti poměrně velký prostor. Jedná se tedy o paměti malé kapacity.

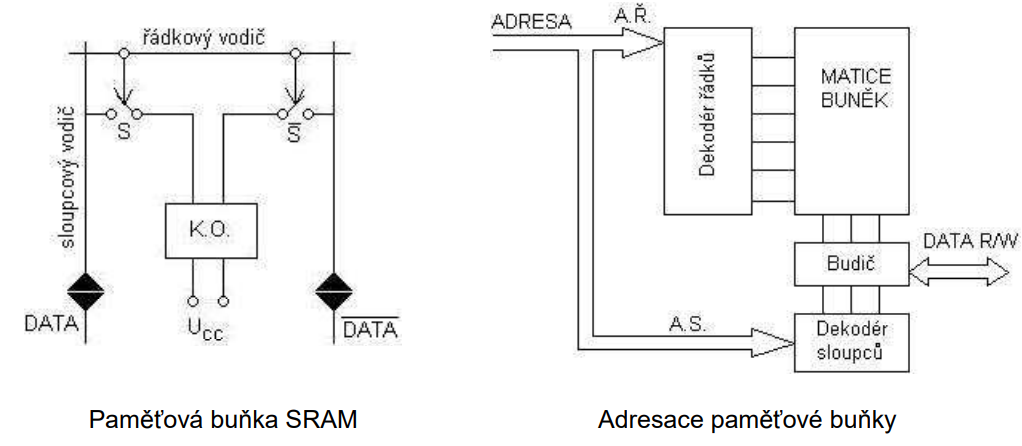
Propustnost 300 GB/s

Zpoždění 1 – 2 ns

Využití: registry, vyrovnávací paměť CACHE.

Paměťová buňka je tvořena bistabilním klopným obvodem (BKO).

Stav klopného obvodu reprezentuje logickou hodnotu 0 nebo 1.



Jedna paměťová buňka obsahuje minimálně 4 tranzistory (2 tvoří samotný BKO, zbývající řídí proces čtení/zápis).

**Fungování:**

Po výběru příslušné paměťové buňky se řádkovým vodičem sepnou spínače, čímž dojde k připojení paměťové buňky na datové (sloupcové) vodiče.

V této chvíli můžeme z buňky číst nebo do ní zapisovat (resp. změnit stav BKO). Při čtení se zkoumá jak skutečná, tak inverzní hodnota (slouží ke kontrole správnosti čtení).

**Dynamická paměť RAM (DRAM - Dynamic Random Access Memory)**

Potřebuje obnovovací cykly, aby se obsah paměti nezměnil.

Využití: operační paměť, videopaměť grafické karty

Propustnost 25GB/s

Zpoždění 50 ns

Paměťová buňka je tvořena kondenzátorem (ve skutečnosti se využívá kapacity PN přechodu) a tranzistorem typu MOSFET, kterým se řídí nabíjení a vybíjení paměťového kondenzátoru.

Logické hodnoty 0 a 1 odpovídají vybitému a nabitému kondenzátoru.

**DRAM Refresh**

Protože je kapacita paměťové buňky reálná a velmi malá, dochází k rychlému samovolnému vybíjení (ztrátě informace).

Aby ke ztrátě informace nedošlo, provádí se periodická obnova dat (refresh).

Refresh (obnovování informace) se provádí po celých řádcích.

V okamžiku obnovování informace nelze provádět operace čtení/zápis.

**Destruktivní paměť při čtení (důvod proč je pomalejší než SRAM)**

Při čtení je na adresový vodič přivedena hodnota log. 1, která způsobí otevření tranzistoru. Pokud byl kondenzátor nabitý, zapsaná hodnota přejde na datový vodič. Tímto čtením však dojde k vybití kondenzátoru a tedy ztrátě uložené informace.

Přečtenou hodnotu je nutné opět do paměti zapsat. Z důvodu periodické obnovy informace a obnovy informace po jejím přečtení jsou paměti DRAM pomalejší při čtení/zápisu než paměti SRAM.

Tato paměť se rozděluje na: SDRAM, DDR, DDR2, DDR3, DDR4, RDRAM

**Synchronous Dynamic RAM (SDRAM) (stará)**

Během jednoho hodinového impulsu provede 1 operaci přenosu dat

Paměti SDRAM jsou schopny pracovat s kmitočtem 66, 100, 133 MHz

Datová propustnost byla 533 až 1066 MB/s, napájecí napětí 3,3 V, kapacita do 512 MB.

**DDR (Double Data Rate)**

DDR paměť přenáší data jak během náběžné, tak sestupné hrany hodinového impulsu.

To znamená, že během jednoho impulsu se tedy provedou 2 operace

Pracovní kmitočet 200–600 MHz, propustnost 1,6–4,8 GB/s, napájecí napětí 2,5 V

Kapacita jednoho paměťového modulu se pohybuje od 64 MB až 2 GB

**DDR2**

Pracují stejně jako DDR.

Rozdíl je v použití nižšího napájecího napětí (1,8 V). Díky nižšímu napětí je spotřeba pamětí DDR2 nižší a je možné je taktovat na vyšší pracovní frekvence.

Kapacita jednoho paměťového modulu se pohybuje od 256 MB až 4 GB

Pracovní kmitočet pamětí se pohybuje v rozmezí 400–1066 MHz s datovou propustností až 8,5 GB/s

Technologie QPB (Quad Pumped Bus)

Efektivní hodnoty kmitočtu, umožní zvýšit rychlost o čtyřnásobek základního kmitočtu paměťové sběrnice

Příklad:

Paměť DDR2-800, pracuje vnitřně na kmitočtu 200 MHz, díky technologii QPB je její efektivní kmitočet 4× vyšší, tedy 800 MHz

**DDR3**

Napájecí napětí bylo sníženo na 1,5 V (nižší spotřeba modulů + nižší vyzářené teplo).

Pracovní kmitočet pamětí se pohybuje v rozmezí 800–2133 MHz s datovou propustností až 17 GB/s.

Kapacita jednoho paměťového modulu se pohybuje od 512 MB až 8 GB

**DDR4**

Využívá napájecí napětí 1,2 V (nižší spotřeba modulů + nižší vyzářené teplo).

Pracovní kmitočty pamětí se pohybují v rozmezí 2133–2400 MHz, standardně jsou řízeny 4-kanálovým paměťovým řadičem.

Kapacita jednoho paměťového modulu se pohybuje v rozmezí 8 GB až 128 GB.

**RDRAM (Rambus DRAM)**

Paměti vyvinuté firmou Rambus a Intel, určené pro procesory Pentium 4 s podporou technologie QPB.

Data se přenášela 2x za hodinový cyklus podobně jako paměti DDR.

Velká datová propustnost byla dosažena velkým taktovacím kmitočtem paměti.

Vysoká cena modulů (RIMM modul), nutnost osadit 2 nebo 4 moduly a nepoužité banky osadit terminátorem (CRIMM modul) způsobilo, že se tyto paměti na trhu neprosadili.

**Moduly pamětí DRAM**

Polovodičové paměti typu DRAM se využívají především pro konstrukci operační paměti.

Pouzdro DIP (starý)

V minulosti byla operační paměť v provedení klasického integrovaného obvodu – pouzdro DIP (Dual In-line Package).

DIP měl běžně 16 nebo 18 pinů.

Operační paměť měla pevnou kapacitu.

Používaný s 8086

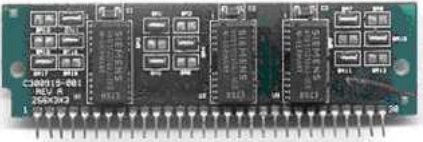


**SIPP (Single In-line Pin Package) (taky starý)**

Obě strany modulů jsou stejné.

SIMM moduly se dělí na krátké (30 pinů - šířka datové sběrnice byla 8 bitů) a dlouhé (72 pinů - šířka datové sběrnice 32 bitů).

Dual nebo Quad channel



**DIMM (Dual In-line Memory Module)**

Novější SDRAM, DDR, DDR2, DDR3, DDR4

Šířka datové sběrnice je 64 bitů

Paměťové moduly DIMM mají po stranách otvory pro zajištění v patici na základní desce.

SDRAM – 168 pinů, DDR – 184 pinů, DDR2 – 240 pinů, DDR3 – 240 pinů, DDR4 – 288 pinů

SO-DIMM

moduly jsou určeny pro notebooky a jsou osazeny stejným typem paměti jako DIMM

jen mají menší pošty pinů. Pod a včetně 100 pinů je 32bit sběrnice, nad je 64bit sběrnice

**RIMM (Rambus Inline Memory Module)**

používal se výhradně pro RDRAM

**Paměti ROM**

většinou určené pouze pro čtení zapsaného obsahu

Slouží především k uložení firmware v elektronických zařízeních, popřípadě BIOSu základní desky

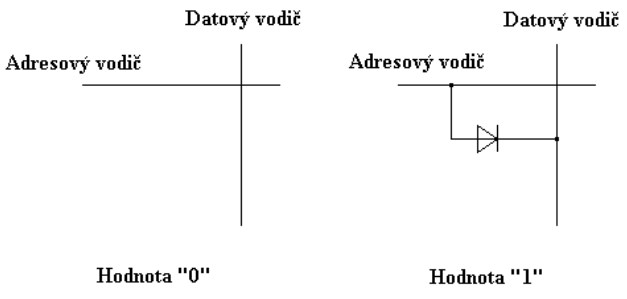
Dělí se na:

1. po naprogramování nelze změnit obsah paměti – **ROM** a **PROM**
2. paměti lze omezeně přeprogramovat (změnit obsah) – **EPROM**, **EEPROM**, **Flash ROM**

**ROM (Read Only Memory)**

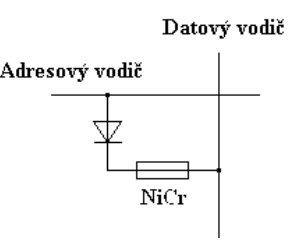
Informace je do této paměti zapsána při její výrobě a poté již není možné žádným způsobem obsah změnit.

Paměťová buňka paměti ROM může být realizována jako dvojice nespojených vodičů a vodičů propojených přes polovodičovou diodu

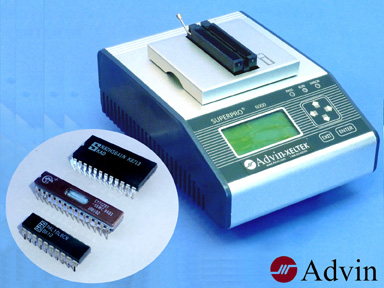


**PROM (Programmable ROM)**

Paměť, kterou si může uživatel naprogramovat sám.



Program se do paměti ukládá pomocí speciálního zařízení – programátoru (obrázek)



Při výrobě je vytvořena matice obsahující spojené adresové vodiče s datovými vodiči přes polovodičovou diodu a tavnou pojistku.

Takto vyrobená paměť obsahuje na začátku samé hodnoty 1.

Během programování (zápisu informace) se na příslušnou paměťovou buňku přivede takový proudový impuls, který způsobí přepálení pojistky (stav 0)

Tento zápis je možné provést pouze jednou.

**EPROM (Erasable PROM)**

Jsou realizovány pomocí speciálních unipolárních tranzistorů

Zapsané informace je možné vymazat působením UV záření (doba působení je asi půl hodiny)

Paměti EPROM jsou charakteristické malým okénkem v pouzdře integrovaného obvodu obsahujícího tuto paměť.

Okénko bývá přelepeno ochranným štítkem proti UV záření

**EEPROM (Electrically Erasable PROM)**

konstrukcí podobný pamětem EPROM

Oproti pamětem EPROM se mažou elektrickými impulsy, doba mazání se pohybuje v milisekundách

**Flash ROM**

Rychlejší než EPROM a EEPROM

S pamětí se dá pracovat jako s RAM, ale po odpojení napájení se informace nevymaže.

Vymazání informace se provádí elektrickou cestou, jejich přeprogramování je možné provést přímo v počítači.

Vlastnosti

* Stabilní uložení informací - uchová data i bez napájení
* Umožňují přímé čtení i zápis (stejně jako paměti typu RAM)
* Nízká úroveň napájení - velice nízký příkon vhodný pro přenosná zařízení
* Stálost - jsou schopny odolat velkým otřesům bez ztráty dat
* Rychlost - extrémně krátká vybavovací doba

Nevýhoda - Mají omezenou životnost na cca 100 000 cyklů výmaz / zápis.

Využití: externí flash disky, paměťové karty, SSD disky, BIOS

**Technologie operační paměti**

**Vícekanálový paměťový řadič**

Dual channel = 2 přenosové kanály.

Quad channel = 4 přenosové kanály.

Eight channel = 8 přenosových kanálů

(existuje i triple channel atd…)

Tuto technologi podporuje pouze DDR3 a DDR4

Místo jedné paměťové sběrnice spojující operační paměť s řadičem paměti je použiti více sběrnic. Propustnost tak teoreticky vzroste několikrát.

Pro vyšší datovou propustnost je potřeba splnit tyto podmínky:

1. Podpora čipové sady na základní desce
2. Paměťové moduly se musí osazovat v páru.

Paměťové banky na základní desce jsou z tohoto důvodu barevně odlišeny

1. Oba kanály musí být osazeny stejným typem paměťového modulu DIMM (proto se prodávají kity daných modulů)

**SPD (Serial Presence Detect)**

Pamět typu EEPROM pro ukládání informací o paměťových modulech (frekvence, napětí, výrobce..)

**EPP (Enhanced Performance Profiles)**

Vylepšené SPD (pomatuje si i přetaktování paměti)

**ECC (Error-Correcting Code)** – použítí v serverech

technologie detekce a opravy chyb, kontroluje integritu dat v paměti. Dokáže opravit 1 nebo 2 bity.

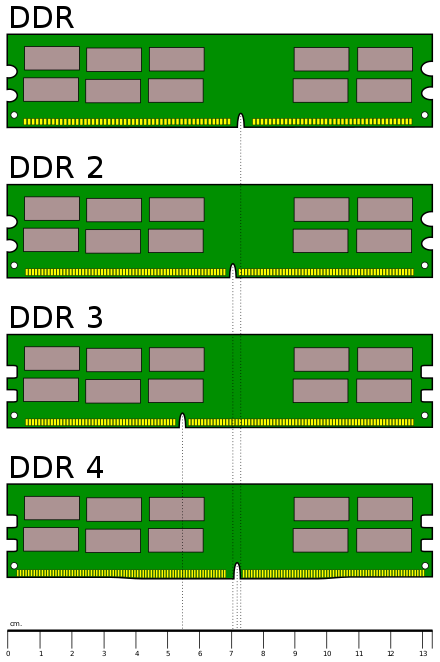
**Chipkill** - použítí v serverech

Metoda lepší než ECC, pokud detekuje více chyb než je schopen opravit (4 bity), tak vyřadí daný čip z provozu, aby zabránil vzniku dalších chyb.

**Memory Scrubbing**

proces vyhledávání jednobitových chyb a jejich případná oprava v době, kdy je paměť nejméně vytížena. Má aktivně předcházet vzniku chyb ve více bitech

Key – (klíč) jsou u cache pamětí tagy, ale označuje se i tak otvory pro správnou orientaci modulu (taky se jim říká zámky)



**Pojmy**

**Adresa**

Procesor pomocí adresy vybírá přímo požadovanou buňku paměti, kde se o její nalezení stará řadič. Adresa je najita v mřížce pomocí sloupcového a řádkového vodiče.

**WR operace**

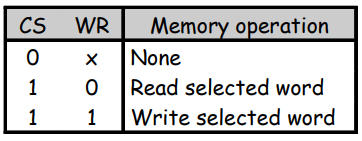
nachází se v rozhraní RAM paměti.

Vybírá mezi čtením a zapisováním do paměti.

K přečtení paměti se WR nastaví na 0

K zápisu do paměti se WR nastaví na 1

Ještě k tomu je zapotřebí CS (Chip Select), 0 = deaktivovat čip RAM, 1 = aktivovat čip RAM



**RD operace -?**

**Cache**

Úkol: překonat paměťovou bariéru mezi procesorem a pamětí

(výkon procesoru roste rychleji než výkon paměti a paměť je dost pozadu)

Nachází se v procesorech nebo pevných discích

Cache Memory je velmi rychlá paměť, která je zpravidla umístěna mezi procesorem a operační pamětí.

Ve cache paměti je uložena ta část (nějaká data) hlavní paměti, která je právě procesorem používána.

Jsou-li procesorem požadovaná data v cache paměti (hledají se podle klíče a nejdříve se prohledává L1, poté L2 a nakonec L3), tak jsou z ní přečteny a není proveden vstup do pomalejší OP (když jsou najity tak tomu se říka “hit“). Ale pokud požadovaná data nejsou v cache paměti, tak jsou přečteny z OP (jsou přeneseny do procesoru i do cache paměti, nejdříve do L3, poté L2 a nakonec L1) (tomu se říka “miss“).

= Volba správné strategie přesunu údajů mezi pamětí cache a operační pamětí má podstatný vliv na výkon systému

Procesory často používají koprocesor, který přebírá veškeré úlohy spojené s přípravou a aktualizací dat v cache paměti.

(ve starých počítačích byla L1 byla v procesoru a L2 na základní desce a L3 nebyla)

**Organizace cache**

Cache paměti bývají organizovány jako **asociativní paměti**. Asociativní paměti jsou tvořeny tabulkou (tabulkami), která obsahuje vždy sloupec, v němž jsou umístěny tzv. **tagy** (klíče), podle kterých se v asociativní paměti vyhledává. Dále jsou v tabulce umístěna data, která paměť uchovává, a popř. další informace nutné k zajištění správné funkce paměti

**Pojmy**

**Klíč** (někdy Tag) – cache je tvořena tabulkami, které obsahují sloupec klíčů podle kterého se v ní vyhledává. (Tabulce se někdy říká banka) Klíč má několik částí (část adresy, přístupové bity, statistika používání)

**Řádek** – tabulky se dělí na řádky, kde každý řádek má klíč. Na řádku jsou poté data, která paměť uchovává. V datové části jsou operandy (obvykle 4 nebo 8)

Taková banka by se dala představit jako Dictionary <Klíč,List<Operandy>>

**Příklad tabulky**

Paměť 16KB = 4 Tabulky (Banky) po 4KB (4096 B)

V tabulce 128 řádků – 4096 / 128 = 32B jeden řádek

L1 – paměti o malé kapacitě je přímo součástí procesoru a je stejně rychlá, jako procesor (nejspíše obsahuje data které jádro zrovna zpracovává)

je obvykle rozdělena do dvou částí:

1. Mezipaměť instrukcí - se zabývá informacemi o operaci, kterou musí CPU provést
2. datová mezipaměť - obsahuje data, na nichž má být operace provedena.

L2 - je ppomalejší než L1, ale je zase větší

L3 – je největší, ale také nejpomalejší.

Je pro všechna jádra společná

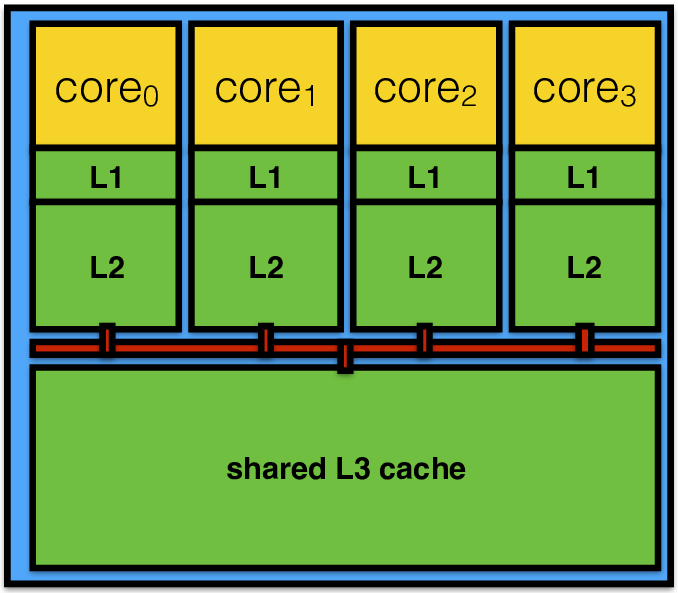
**Příklad** i7 – 10700 (8 jader)

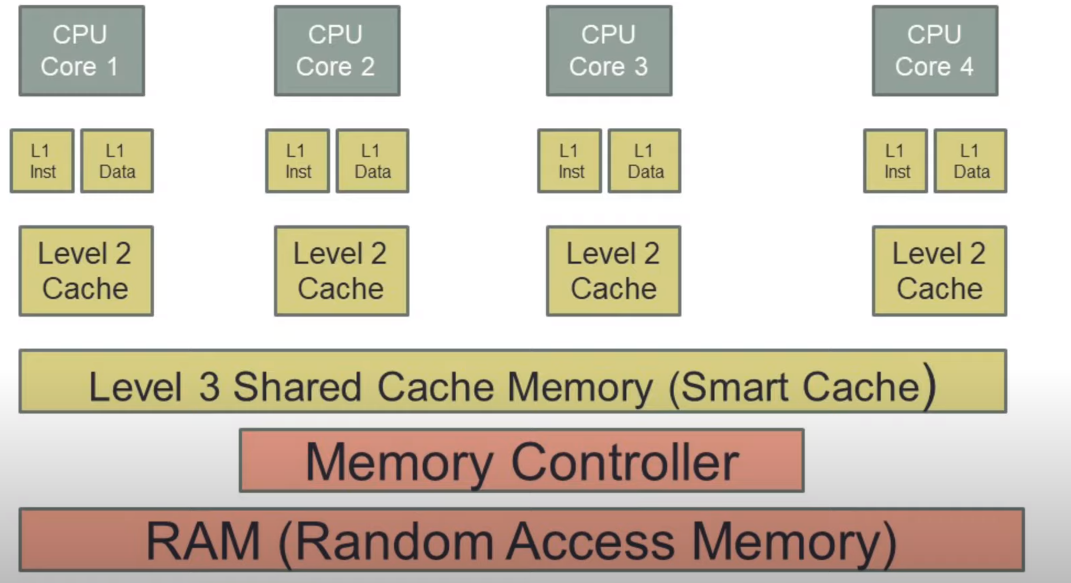
L1 – 64 KB (každé jádro) = 0,5 MB celkem

L2 – 256 KB (každé jádro) = 2 MB celkem

L3 – 16 MB (společná pro všechna jádra)

**Zajímavost:** Mezipaměť paměti L1 je obvykle stokrát rychlejší než paměť RAM, zatímco mezipaměť L2 je zhruba 25krát rychlejší





**Smart Cache**

Vynalezena Intelem

Sdílení L2 nebo L3 mezi jádry procesoru.

L2 je stává společnou

Nějaké jádro může používat více paměti, než má vyhrazeno, pokud jí nějaké jádra nepotřebují (nějakou část nebo celou)

=

Pokud nějaké jádro potřebuje data ke zpracování tak mu bude přiděleno více paměti

<https://www.youtube.com/watch?v=bE9EbQOeb_U&ab_channel=pctambayan>

<https://www.youtube.com/watch?v=GlrJTc651AM&ab_channel=TheAudiopedia>