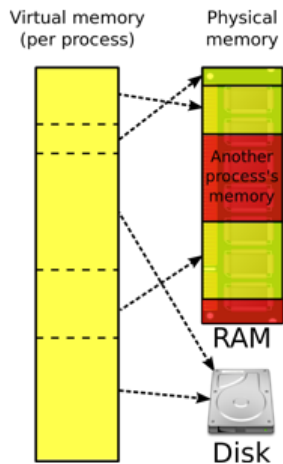


M13 Paměť

#technicke_vybaveni_pocitacu

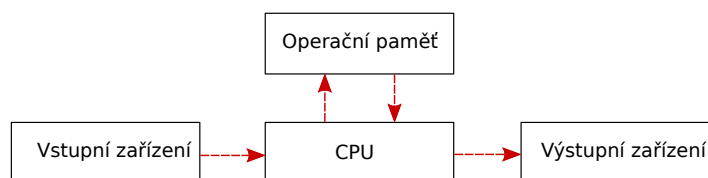
- fyzická zařízení určená k ukládání programů nebo dat pro okamžitou nebo trvalou potřebu
- rozdělení fyzických zařízení
 - vnitřní - RAM
 - vnější - ukládání programů a dat
- vnitřní paměť uložená v vnější se nazývá "virtuální" paměť (*nebo swap*)



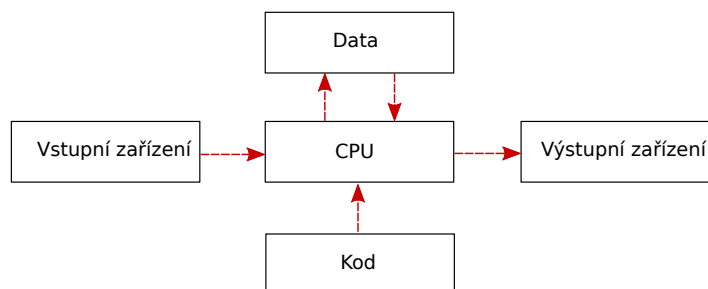
- energetické rozdělení
 - nezávislé (*nevolatilní*) - Flash (vnější i vnitřní), ROM/PROM/EPROM/EEPROM (pro uložení firmware)
 - závislé (*volatilní*) - vnitřní paměť DRAM a vyrovnávací (SRAM)
- paměťové médium popisuje vnější paměť jako je např.: magnetopáska, optický disk (CD/DVD)

Operační paměť

- slouží k ukládání dat po dobu běhu programu
- přístup k op je rychlejší než přístup k vnější paměti
- procesor pomocí adresy vybírá požadovanou buňku
- paměť je spojena s procesorem pomocí rychlé sběrnice; mezi op a procesor se vkládá ještě [cache](#)
- dnes realizována jako polovodičová paměť typu RAM; ztrácí informace při odpojení napájení; obsah paměti je třeba občas obnovovat čtením všech řádků
- je spravován operačním systémem
- uchovává kód programů (kód procesů a jejich mezivýsledky), základní datové struktury kernelu, atd.
- Fyzický adresový prostor (*FAP*) paměti je souvislý prostor paměťových buněk určité velikosti (1, 2, 4 nebo 8 bytů); buňky jsou lineárně adresovány adresami pevné délky; velikost buňky je dána délkou adresy (adresa n bytů; buněk 2^n); celý FAP nemusí být vyplněn; některé bloky se mohou objevit vícekrát
- správa paměti
 - přidělení paměť. regionu na požádání procesu
 - uvolnění regionu na požádání procesu
 - udržení informace o obsazení adresového prostoru
 - zabránění přístupu procesu k paměti mimo jeho přidělený region
 - podpora střídaného běhu více procesů
- architektury
 - [Von Neumannova](#)



- Harvardská



- metody správy
 - monolitická
 - FAP je rozdělen na dva bloky
 - jeden provádí rutiny kernelu a jeho datové struktury - "Kernel memory"
 - druhý je přiřazen na požadání aplikací - "Application memory"
 - je-li paměť volná, je přidělena procesu celá bez ohledu na požadovanou velikost (nesmí přesáhnout velikost bloku); v obsazené paměti je požadavek zamítnut
 - statické bloky
 - paměť je rozdělena do několika bloků o pevné velikosti, které lze samostatně alokovat
 - maximální počet procesů je omezen počtem bloků; proces může přesahovat jeden blok
 - velikost bloku se liší podle využití
 - ochranu zajišťuje limitní registr procesoru - v registru je uložena hodnota aktuálního paměťového regionu; hodnota lok. adresy se porovnává s hodnotou registru; pokud je hodnota větší je vyvolána výjimka (*proces se pokouší zapsat mimo region*)
 - dynamická
 - paměť je rozdělena na bloky jejichž velikost se dynamicky upravuje dle požadavků procesů; před alokací prvního regionu tvoří paměť aplikačního prostoru jeden blok
 - po uvolnění bloků je nutné provádět scelování volných bloků
 - obsazení paměti je realizováno na počátku každého bloku jakousi hlavičkou
 - paměť je chráněna limitním registrem

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

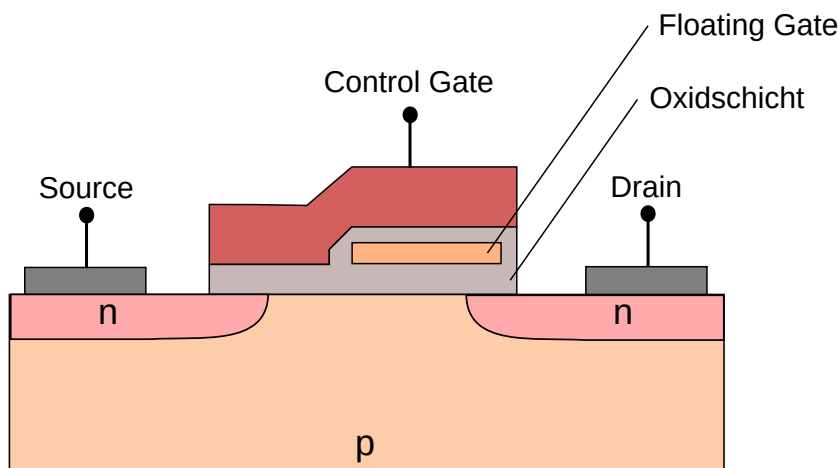
- proto OS vyžadují před vypnutím proces **shutdown** který korektně ukončí procesy systému a uloží obsah diskového cache do souborů na disku
- před odpojením je důležité odmountovat vyměnitelná média jinak může dojít k poškození souborového systému
- moderní systém se snaží problém eliminovat zapomocí 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

Paměť flash

- je energicky nezávislá a elektricky zapisovatelná
- asynchronní a nedá se taktovat
- organizována po blocích (1 blok = [?] bytů); každý blok lze programovat samostatně
- používá se jako paměť typu ROM např. pro uložení firmware
- lze ji znovu naprogramovat bez nutnosti vyjmutí čipu
- využívá se v přenosném datovém médiu (např.: SD karta, USB Flash disk, SSD disk)
- má omezenou přepisovatelnost
- princip ukládání
 - data jsou ukládána v unipolárních tranzistorech (1 tranzistor = 1 bit (*SLC*) / 3+ bitů (*MLC*)); SLC nabízí větší rychlost a stabilitu, MLC naopak větší kapacitu a menší cenu
 - tranzistor obsahuje dvě hradla - ovládací (*CG*) a plovoucí (*FG*) izolované vrstvou oxidu; všechny elektrony na FG přivedené jsou zde „uvězněny“, tím je informace uchována
 - když jsou na FG elektrony, částečně ruší el. pole přicházející z CG, což modifikuje prahové napětí U_t buňky
 - buňka je aktivována přivedením určitého elektrického napětí na CG, což ovlivňuje elektrický proud tranzistorem; tento proud proudí, nebo neproudí, což závisí na úrovni U_t buňky, která je závislá na množství elektronů na hradle FG
 - přítomnost nebo nepřítomnost elektrického proudu je interpretována jako log. **1** nebo **0**



EEPROM

- elektricky mazatelnou energeticky nezávislou paměť typu ROM-RAM
- typická životnost je 200 000 zápisů (ATmega16) (více než flash); životnost dat je 20 let
- nevýhodou je vyšší složitost paměťové buňky → nižší hustota → vyšší cena
- využívá se jako úložiště dat, která se mění častěji než je životnost paměti flash (např. nastavení hlasitosti u TV)
- používá tranzistory vyrobené technologií MNOS; na řídicí elektrodě je nanášena vrstva nitridu křemíku a pod ní je umístěna tenká vrstva oxidu křemičitého
- buňka paměti pracuje na principu ukládání elektrického náboje na přechod těchto dvou vrstev

Paměťová buňka

statická

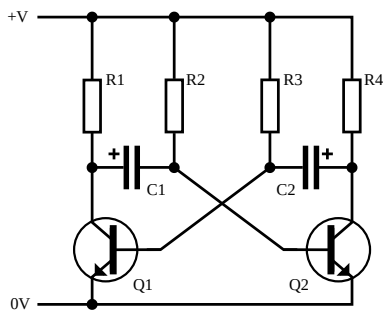
- uchovává informaci v sobě uloženou po celou dobu, kdy jsou připojeny ke zdroji elektrického napájení
- má rychlý přístupový čas, což znamená, že data lze číst nebo zapisovat rychle
- používá bistabilní klopný obvod k uložení informace

dynamická

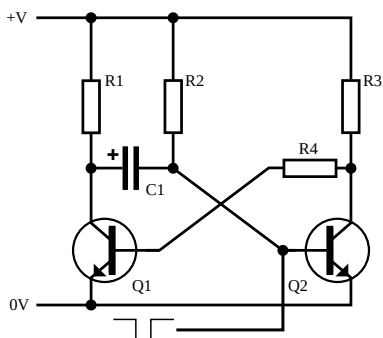
- uchovává informaci v sobě uloženou i po odpojení zdroje elektrického napájení
- vyžaduje periodickou obnovu informace aby se zabránilo ztrátě dat
- používá malého náboje uloženého v kondenzátoru k uložení informace
- levnější na výrobu

Klopné obvody

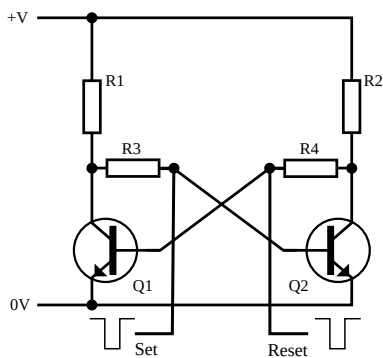
- elektronický obvod, který přechází mezi několika stavy
- ke změně mezi stavy dochází skokově
- skládají z hradel
- lze je použít jako např.: paměťové prvky, impulzní generátory, časovače nebo oscilátory
- astabilní
 - nemají žádný stabilní stav
 - obvody neustále oscilují mezi jedním a druhým stavem podle nastavené časové konstanty



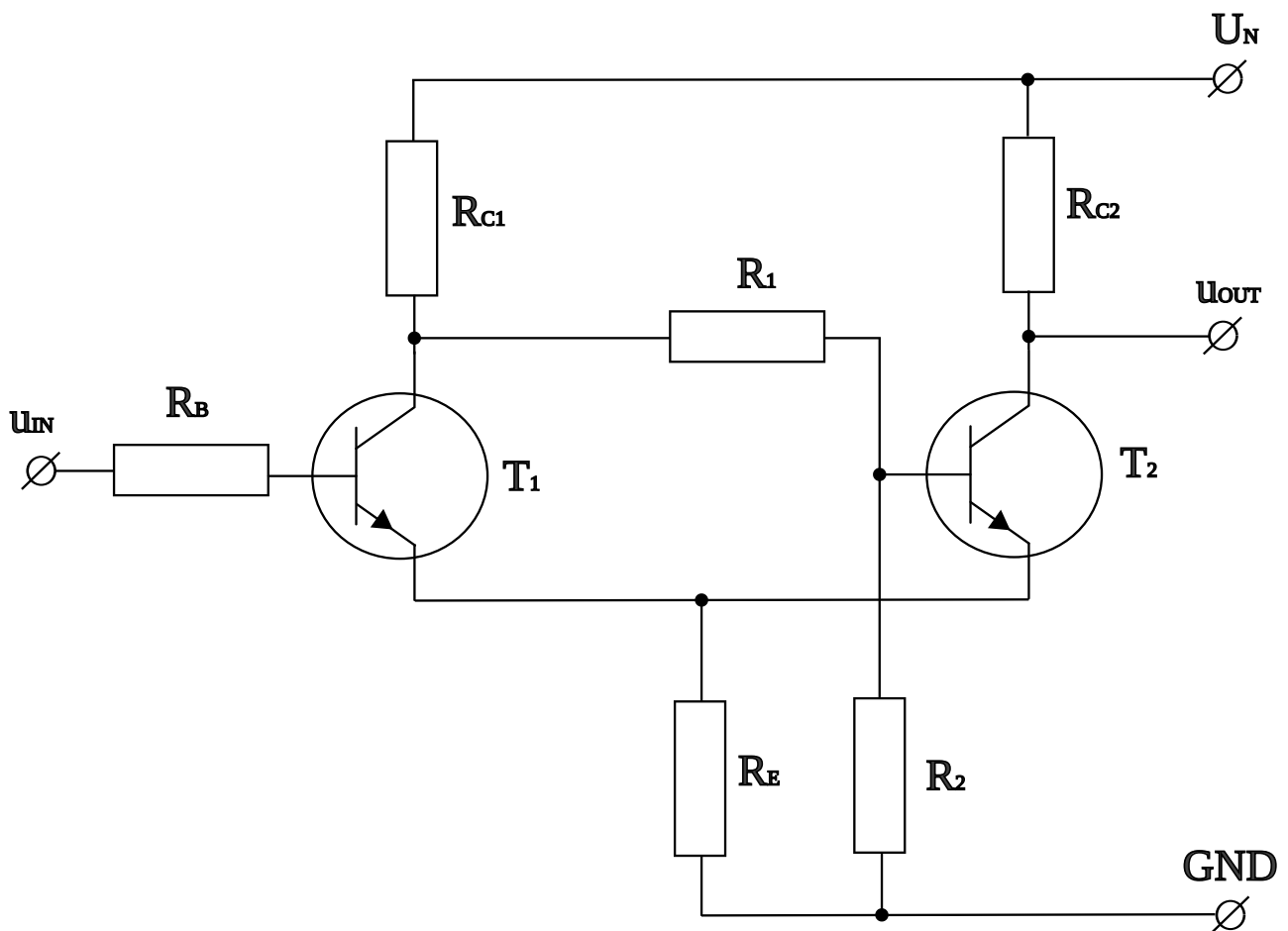
- monostabilní
 - jeden stabilní stav
 - sám se po určité době přepne zpět do stabilního stavu



- bistabilní
 - dva stabilní stavy
 - mezi stavy lze přepínat pomocí signálů přivedených na vstupy
 - obvody se používají jako paměťové prvky



- Schmittův
 - slouží k úpravě tvaru impulzů
 - základní vlastností je hystereze
 - výstup je závislý nejen na hodnotě vstupu, ale i na jeho původním stavu
 - hystereze zde zabraňuje vzniku zákmitů výstupního signálu v okolí střední úrovně spínání



Registry

- úložiště pro informace o velikosti 1 slova (ATmega16 je to 8 bitů); může být odvozena od velikosti sběrnice
- využívá se pro uložení mezivýpočtů a adres
- procesor má několik málo desítek registrů různých typů
 - uživatelsky-přístupné registry
 - datové registry - pro uložení číselných hodnot, jako jsou **integer** a float
 - adresové registry
 - uchovávají adresy
 - používané instrukcemi, které přistupují do paměti nepřímo
 - mohou obsahovat jak adresu tak data

- mohou ukládat číselné hodnoty (použity jako indexové registry (jako offset z některých adres))
 - stack register - udává relativní adresu poslední přidané položky na vrchol zásobníku
- podmínkové - obsahují pravdivostní hodnotu často využívanou k zjištění, zda některá instrukce měla nebo neměla být vykonána
- konstantní - pouze pro čtení takové hodnoty, jako je 0, 1, nebo π
- mohou být paměťově mapované - přístupné stejným způsobem, jako data v operační paměti
- výhodou je jejich rychlost
- řídicí registry mění činnost procesoru
- LIFO (last in, first out) nebo FIFO (first in, first out)
- dělí se podle počtu bitů (8-bitový, 32-bitový)
- dnes implementovány jako soubory registrů
- registr příznaků
 - není interpretován jako číslo; každý z jeho bitů má zvláštní význam
 - provedení některých instrukcí může měnit hodnotu některých příznaků; ovlivňuje chování některých instrukcí (např. podmíněné skoky)
 - je možnost přímo číst
 - velikost a významy bitů závisí na architektuře
 - dělení
 - aritmeticko-logické příznaky - **carry**, **null**, **negative** (dvojkový doplněk), **overflow** a **parity** (počet jedničkových bitů v binární reprezentaci výsledku poslední operace je sudý)
 - řídicí příznaky - **step** a **interrupt**

Latence

- doba kterou trvá procesoru získat data z paměti; interval mezi požadavkem a doručením
- 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)

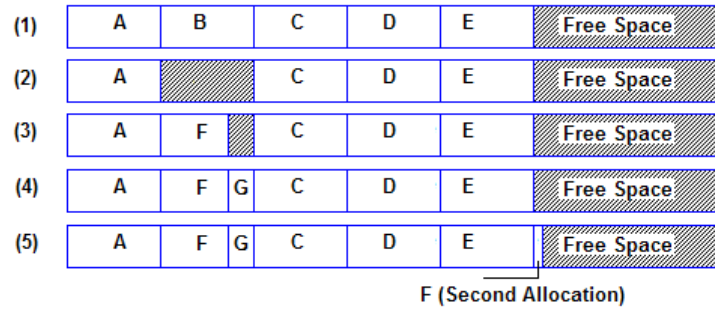
Synchronní a asynchronní

- asynchronní
 - vyznačují se tím, že na svém čipu většinou obsahují pouze minimum podpůrných obvodů
 - obvody pracují maximální možnou rychlostí bez přítomnosti jakýchkoli hodinových signálů
 - o dodržení doby přečtení/zápisu se musí starat paměťový subsystém, v jednodušších případech přímo mikroprocesor
 - mezi vystavením adresy na adresovou sběrnici a přečtením dat z paměti uběhne doba 50 či 60 ns; aby mikroprocesor tuto dobu dodržel, musí provádět jinou činnost, v nejhorším případě pouze nečinně několik taktů čekat
 - dnes se prakticky neobjevují
- synchronní
 - u synchronních pamětí jsou použity hodinové signály; podpůrné obvody pracují na základě stavového automatu
 - rozšiřuje se protokol použitý pro přenosy dat do a z paměti
 - paměťové čipy většinou dokáží automaticky provádět obnovu dat
 - nejsou automaticky rychlejší než asynchronní
 - neuvádí se žádné přístupové časy

Souborové systémy

- datová struktura kterou používá operační systém k řízení ukládání a získávání dat
- určuje kde jednotlivá data končí a kde začínají
- některé systémy umožňují přístup pomocí síťového protokolu (např. Samba)

- jak jsou soubory postupně upravovány, mazány a přidávány vzniká fragmentace → defragmentace je proces spojování



fragmentů souborů k sobě