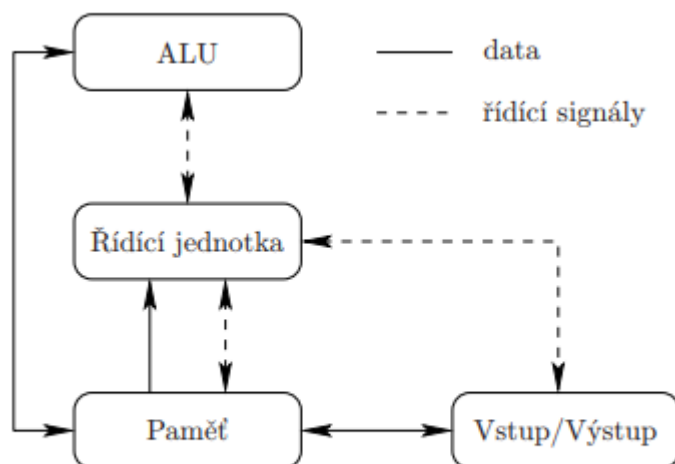


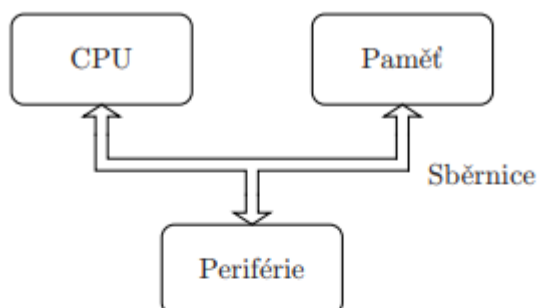
Nakreslete schéma a popište architekturu počítače dle von Neumanna a Harvardskou.

Popište princip fungování počítače jak je popsal von Neumanna. Jakou architekturu využívají dnešní běžné kancelářské počítače. Co je v těchto počítačích proti původní architektuře vylepšeno a jak ?

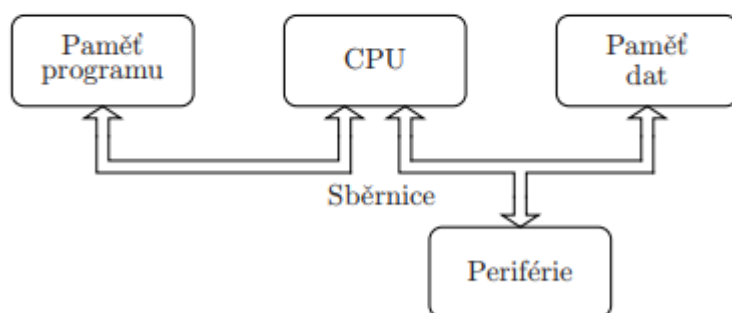


Obrázek 1: Základní schéma počítače podle von Neumanna

Zjednodušené schéma:



Obrázek 2: Počítač podle von Neumanna



Obrázek 3: Harvardská architektura počítače

von Neumann

– Výhody:

- * rozdělení paměti pro kód a data určuje programátor,
- * řídicí jednotka procesoru přistupuje do paměti pro data i pro instrukce jednotným způsobem,
- * jedna sběrnice - jednodušší výroba.

– Nevýhody:

- * společné uložení dat a kódu může mít při chybě za následek přepsání vlastního programu,
- * jediná sběrnice tvoří bottleneck.

harvardská koncepce

– Výhody:

- * oddělení paměti dat a programu přináší výhody:
 - program nemůže přepsat sám sebe,
 - paměti mohou být vyrobeny odlišnými technologiemi,
 - každá paměť může mít jinou velikost nejmenší adresovací jednotky,
 - dvě sběrnice umožňují jednoduchý paralelismus, kdy lze přistupovat pro instrukce i data současně.

– Nevýhody:

- * dvě sběrnice kladou vyšší nároky na vývoj řídicí jednotky procesoru a zvyšují i náklady na výrobu výsledného počítače,
- * nevyužitou část paměti dat nelze použít pro program a obráceně

Harvardská se sice od Neumannovy příliš nelišila, ale odstraňovala některé její nedostatky. V podstatě jde pouze o oddělení paměti pro data a program. Základním nedostatkem obou koncepcí je sekvenční vykonávání instrukcí, které sice umožňuje snadnou implementaci systému, ale nepovoluje dnes tolik potřebné paralelní zpracování. Paralelizmy se musí simulovat až na úrovni operačního systému. Bottleneck systému je také ve sběrnících, které nedovolují přistupovat současně do více míst paměti současně a navíc dovolují v daném okamžiku přenos dat jen jedním směrem.

1. **Nakreslete a popište hierarchické uspořádání paměti v počítači a vysvětlete co je důvodem tohoto uspořádání charakterizujte vlastnosti jednotlivých pamětí. Jak je v počítači (např. s procesorem Intel) implementováno virtuální paměť, proč je potřeba a k čemu slouží**

Od nejrychlejších a nejdražších po pomalé a levné (velikost/cena):

- Registry: v CPU, nejrychlejší paměti, rychlost dle taktu CPU, přímo na procesoru maximálně desítky (stovky bytů), hodnoty proměnných, SRAM
- L1 cache: v CPU, desítky kB, pro instrukce, rychlost dle taktu sběrnice (cca 1 ns), SRAM
- L2 cache: v CPU, 1 -10 MB pro data, ns až jednotky ns, SRAM
- L3 cache: dříve na MB, dnes v CPU; o něco větší a pomalejší L2; sdílena mezi všemi jádry, ns až jednotky ns, SRAM
- DRAM: desítky GiB, průměrná přístupová doba v jednotkách ns
- SSD: stovky GiB - jednotky TiB paměti, rychlost čtení/zápisu ve stovkách/tisících MiB/s, stovky ns
- HDD: TiB paměti, přístupová doba 1 -12 ms
- CD, DVD, pásky: nejpomalejší a nejlevnější paměti

Statické vs Dynamické:

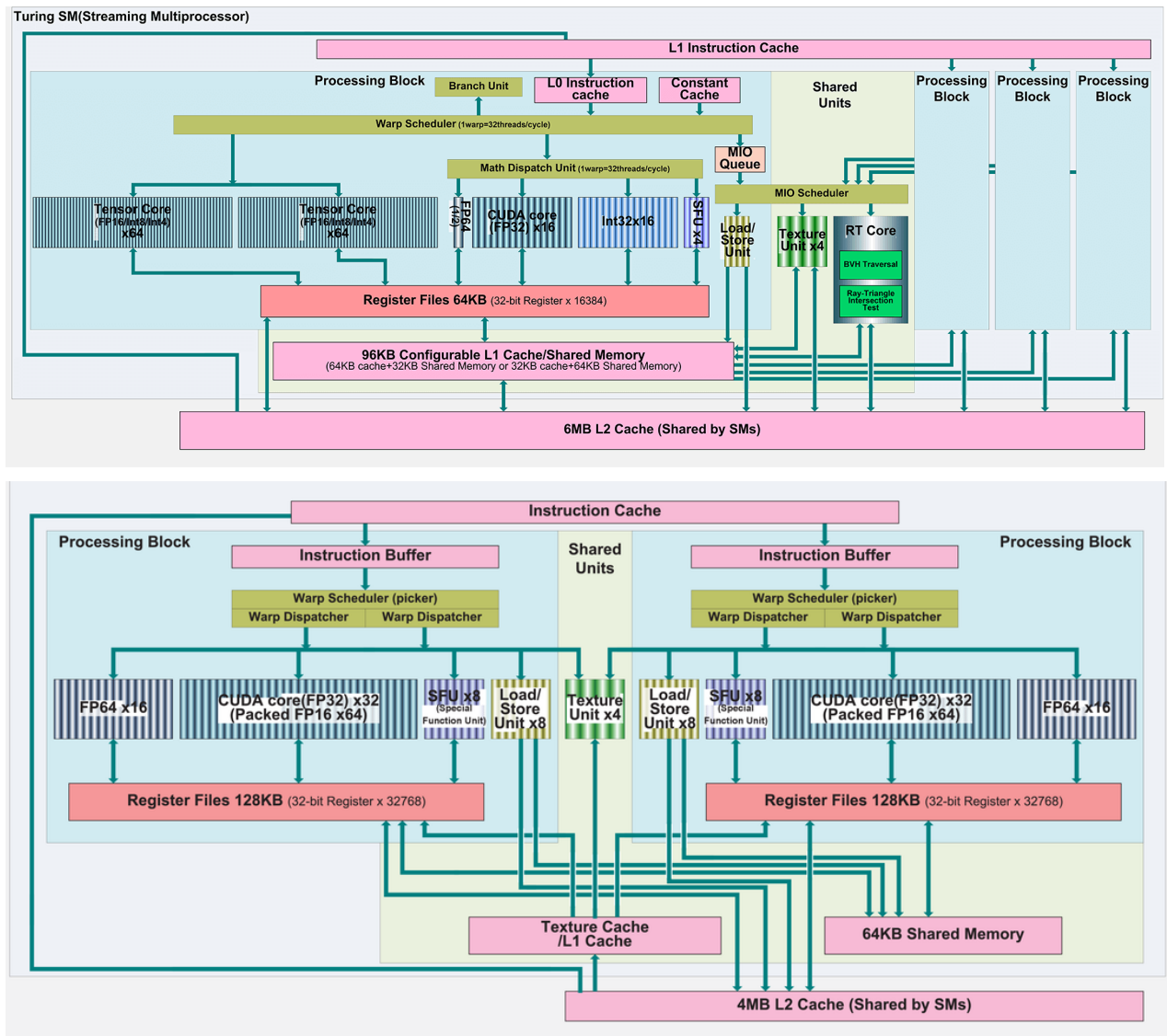
- Statické paměti fungují na klopných obvodech, každá buňka má několik tranzistorů a drží si stav, dokud jsou napájeny.
- Dynamické paměti fungují na kondenzátorech, každá buňka má jeden kondenzátor (nabitý = log 1; vybitý = log 0), který se musí obnovovat (refreshovat).

Virtuální paměť

V současných běžných operačních systémech je virtuální paměť implementována pomocí stránkování paměti spolu se stránkováním na disk, které rozšiřuje operační paměť o prostor na pevném disku

2. **Jaké jsou základní obecné konstrukční vlastnosti monolitických počítačů? Nakreslete a popište jeden vybraný monolitický počítač s RISC jádrem podrobně.**
3. **Nakreslete a pošle konstrukci pevného disku a jeho fungování. Znázorněte a popište. Jakými fyzikálními principy se na disk informace zapisují a jak se čtou. Co je to kolmý a podélný zápis**

4. Nakreslete a popište některou z posledních architekturu Nvidia GPU (Volta,Pascal,Maxwell). Popište organizaci vláken technologie CUDA, jejich identifikaci a celkový postup výpočtu.



Pascal (řada Tesla):

- SM se skládá z 128 CUDA jader
- Podpora HBM2

Cuda

- Alokace paměti na **device**
- Přesun dat z **host** na **device**
- Zpracování a provedení výpočtu **device**
- Přesun dat z **device** na **host**

- 1) Nakreslit von Neuman a Harvard. Principy počítače podle von Neumana. Nakreslit a popsat DMA.

von Neuman a Harvard je popsáný nahoře

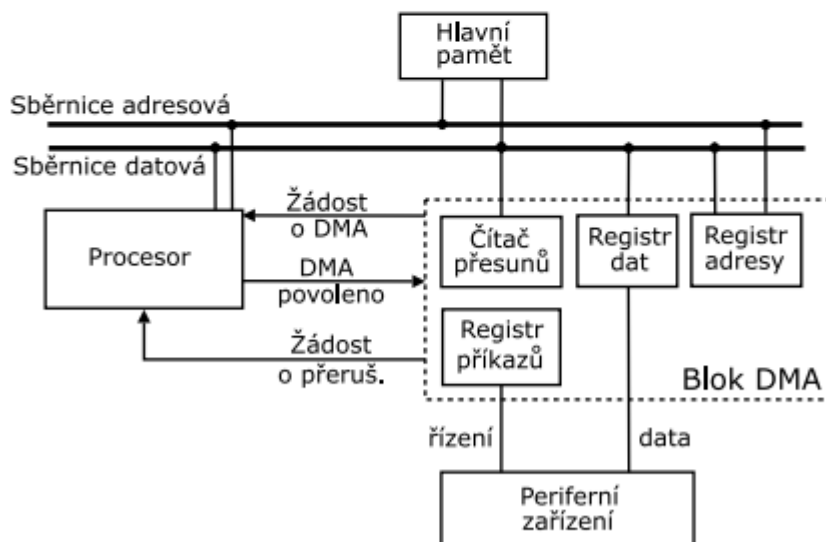
V počítači existují události, na které nelze reagovat jinak než přerušením. Princip operace spočívá v prostém přesunu informací z periferního zařízení do hlavní paměti nebo naopak. To lze realizovat výhodněji formou přímého přístupu do paměti označované zkratkou DMA.

A právě tomu zabraňuje existující další blok, který je schopen generovat adresu a určovat okamžiky přesunu dat po datové sběrnici. Tento blok se označuje jako blok DMA, někdy též označovaný kanál DMA.

- registr dat – obsahuje slovo, jež má být přesunuto z periferního zařízení do hlavní paměti nebo naopak,
- registr adresy – slouží pro uchování adresy hlavní paměti, což je adresa, na kterou bude zapsáno slovo nebo ze které bude dané slovo přečteno,
- čítač přesunů – požadovaný počet slov, které mají být ještě přesunuty v rámci jednoho spojení mezi periferním zařízením a hlavní pamětí.

Blok DMA může pracovat ve dvou režimech:

- přesouvání dat mezi periferním zařízením a hlavní pamětí jednotlivě,
- přesouvání dat v blocích.



Obrázek 6: Struktura bloku pro přímý přístup do paměti (DMA)

2) Principy RISC. Vysvětlit díky čemu jsou rychlejší. Popsat jeden procesor ARM nebo MIPS

mikroprogramový řadič může být nahrazen rychlejším obvodovým řadičem,

- používat zřetěžené zpracování instrukcí,
- celkový počet instrukcí a způsobů adresování je malý,
- data jsou z hlavní paměti vybírána a následně ukládána výhradně jen pomocí dvou instrukcí LOAD a STORE,
- instrukce mají pevnou délku a jednotný formát, který vymezuje význam jednotlivých bitů,
- je použit vyšší počet registrů
- složitost se z technického vybavení přesouvá částečně do optimalizačního kompilátoru

Mezi nevýhody RICS architektury patří třeba nutný nárůst délky programů, tvořených omezeným počtem instrukcí a také díky jednotného délce všech instrukcí.

Zpomalení, které by z toho mělo nutně plynout, se ale v praxi nepotvrdilo.

Všechny uvedené vlastnosti tvoří dobře promyšlený a provázaný celek. Např.

navýšení počtu registrů souvisí s omezením komunikace s pamětí na dvě instrukce LOAD a STORE. Pokud ostatní instrukce nemohou používat paměťové operandy, je třeba mít v procesoru uloženo více dat. Další souvislost je patrná mezi zřetěženým zpracováním a formátem instrukcí. Jednotná délka instrukcí dovoluje rychlejší výběr instrukcí z paměti a tím zajišťuje lepší plnění fronty instrukcí (viz další kapitola).

Jednotný formát pak zjednodušuje dekódování instrukcí

ARM (32bit, Mobily (malý odběr), RISC, 25 registrů, 44 instrukcí, FIQ, IRQ, 3 stupně zřetězení, 4 režimy - 2bit PSW)

(Advanced RISC Machine, původně Acorn RISC Machine)

- plně 32-bitové, nejpoužívanější procesory pro mobilní zařízení
- procesory s nízkou spotřebou používané hlavně v mobilních zařízeních, řídí se filozofií RISC
- přístup do paměti pouze pomocí instrukcí LOAD/STORE (ostatní instrukce šahají do registrů)
- 25 částečně překrytých 32 bitových registrů, (dvě banky registrů, lze mezi nimi přepínat)
- možnost podmíněného vykonání instrukcí
- jednoduchý a výkonný instrukční soubor s pevně danou šířkou instrukcí (44 instrukcí, 32 bitů)
- dva typy přerušení (Např. jiný hardware potřebuje něco od procesoru)
 - FIQ (fast interrupt request), který je pro neodkladné události
 - IRQ (interrupt request), který se používá pro události nevyžadující extrémně krátkou dobu odezvy

- 3 stupňové zřetězení instrukcí
- 4 režimy (uživatelský, supervizor, IRQ a FIQ)
- aktuální režim je určen dvěma bity programového slova (PSW - program status word)

MIPS (32/64bitový, ~~XBOX Set-top Box Mobil Xbox, Set-top box, mobil,~~ Non-interlocked pipeline, RISC, 32 registrů, 32 bitová instrukce)

(microprocessor without interlocked pipeline stages)

- používá non-interlocked pipeline → programátor/překladač se musí starat o to, aby paralelně probíhající zřetězené instrukce nevyužívaly stejných prostředků počítače
- jsou navrženy podle RISC filozofie
- přístup do paměti pouze pomocí instrukcí LOAD/STORE (ostatní instrukce sahají do registrů)
- obsahují 32 registrů, každá instrukce má délku 32 bitů
- používají se např. v herních konzolách, set-top boxech, mobilech a tiskárnách
- nejprve byl navržen kompilátor, až poté implementován obvod, s důrazem na rychlost

3) Nakreslit a popsat periferie monolitických počítačů

- obsahují procesor, paměť a vstupní/výstupní periferie integrované v jednom pouzdře
- obvykle nesou znaky RISC, ale ve zjednodušené podobě, často se pro ně používá Harvardská architektura (lze použít různé paměti pro data a program - ten je často ROM - E(E)PROM nebo Flash)

4) pevný disk, nakreslit a popsat fyzikální princip. Jak se zapisují a čtou data? Co je to kolmé a podélné zapisování

[How do hard drives work? - Kanawat Senanan](#)

Z toho to pochopí každý

5) Co je to CUDA? Jak probíhá výpočet? Popište organizaci vláken. Jaké rozšíření C/C++ si technologie vyžádala

Je unifikovaná architektura pro (obecné) výpočty na grafických kartách

- funguje pouze na kartách od Nvidie
- poskytuje programovací rozhraní pro často používané jazyky - C/C++, Fortran, Python
- optimalizováno na matematicky náročné výpočty (zpracování obrazu, problémy definované pomocí mřížek), lze použít větvení, ale to degraduje výkonnost programu

Typická operace (CPU->GPU, vlákna na GPU, výpočty GPU, GPU->CPU)

- 1) Kopírování dat z CPU na GPU
- 2) Spuštění vláken na GPU
- 3) Provedení výpočtu na GPU
- 4) Zkopírování výsledku z GPU na CPU

Vlákna (nezávislé, bloky, gridy, 3D dimenze, opakované spuštění)

- vlákna musí na sobě být nezávislé, nelze zaručit pořadí jejich vykonávání
- vlákna jsou organizovány do mřížky (bloku), bloky jsou dále také organizovány do větší mřížky (grid), lze rozšířit i do třetí dimenze
- skupiny vláken (warpy) jsou opakovaně spouštěny schedulerem a přepínají se mezi sebou (masivní paralelismus)

6) Popsat hierarchii paměti, proč ji používáme atd., plus dynamické detailně

7) Monolitický počítač s jádrem RISC jeden, nakreslit popsat.

- obsahují procesor, paměť a vstupní/výstupní periferie integrované v jednom pouzdře
- obvykle nesou znaky RISC, ale ve zjednodušené podobě, často se pro ně používá Harvardská architektura (lze použít různé paměti pro data a program - ten je často ROM - E(E)PROM nebo Flash)

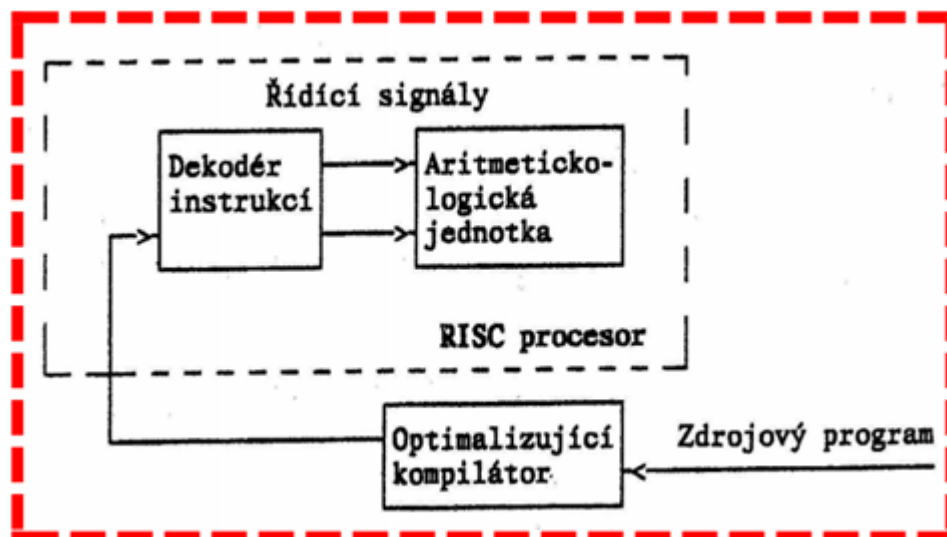
8) Nakreslit jak funguje LCD displej, základní dělení a kde se používá.

Princip činnosti

- 1) První polarizační filtr propustí světlo pouze v jednom směru.
- 2) polarizované světlo prochází strukturou tekutých krystalů, které světlo otočí o 90°
- 3) světlo projde druhým polarizačním filtrem (pootočený o 90° oproti prvnímu)

- ve výchozím stavu tedy displej svítí (propouští světlo)
- elektrickým proudem můžeme kontrolovat orientaci krystalů, což následně způsobí, že část světla bude polarizována odlišně a neprojde druhým filtrem (proud určuje intenzitu světla)
- vrstva krystalů je rozdělena na malé buňky stejné velikosti tvořící jednotlivé pixely monitoru
- je nutné displej podsvětlovat bílým světlem (elektroluminiscenční výbojky, nověji LEDky)

9) Popsat RISC a jak jeho vlastnosti zvětšují výkon a popsat ARM/MIPS



Vlastnosti, zásady a výhody platformy RISC

1) Tato architektura se vyznačuje hlavně redukovanou instrukční sadou, která obsahuje pouze ty nejzákladnější instrukce (složitější operace se řeší jejich kombinací v kódu programu), v každém strojovém cyklu je dokončena právě jedna instrukce

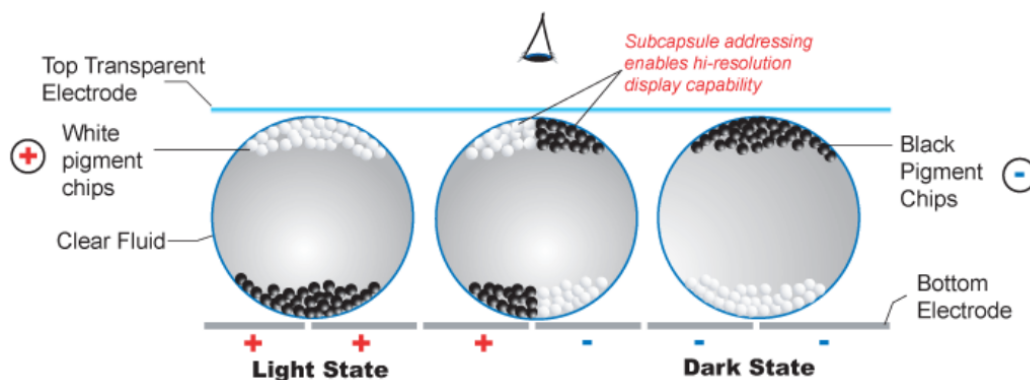
- 2) Instrukce mají pevnou délku a jednotný formát (spolu s jejich redukováním počtem dovoluje lehčí konstrukci procesoru, lze použít obvodový řadič místo mikroprogramového)
- 3) Použití rychlé dočasné paměti - velký počet registrů a použití cache paměti
- 4) Přístup do paměti provádí pouze specializované instrukce (LOAD - načíst / STORE - uložit)
- 5) Používá se zřetězení instrukcí (pipeline) pro zrychlení procesoru
- 6) Složité instrukce nejsou implementovány procesorem, o optimalizaci se stará překladač

Nevýhody: jelikož se používají pouze základní instrukce, očekávalo se, že se délka programů několikanásobně zvýší, toto se však v praxi nepotvrdilo

Zástupce: ARM, MIPS

10) Historie procesoru Intel x86

11) E-Ink - Princip



Skládá se z tzv. kapslí (jakoby pixel). Obsahují světlé (kladně nabitě) a tmavé (záporně nabitě) částice a transparentní chemicky stabilní olej. Změna stavu kapslí využívá elektroforetického principu (pohyb nabitých molekul v elektrickém poli, funkce jasná z obrázku). Vysoká hustota oleje pak drží částice na místě.

Barevné E-Ink fungují:

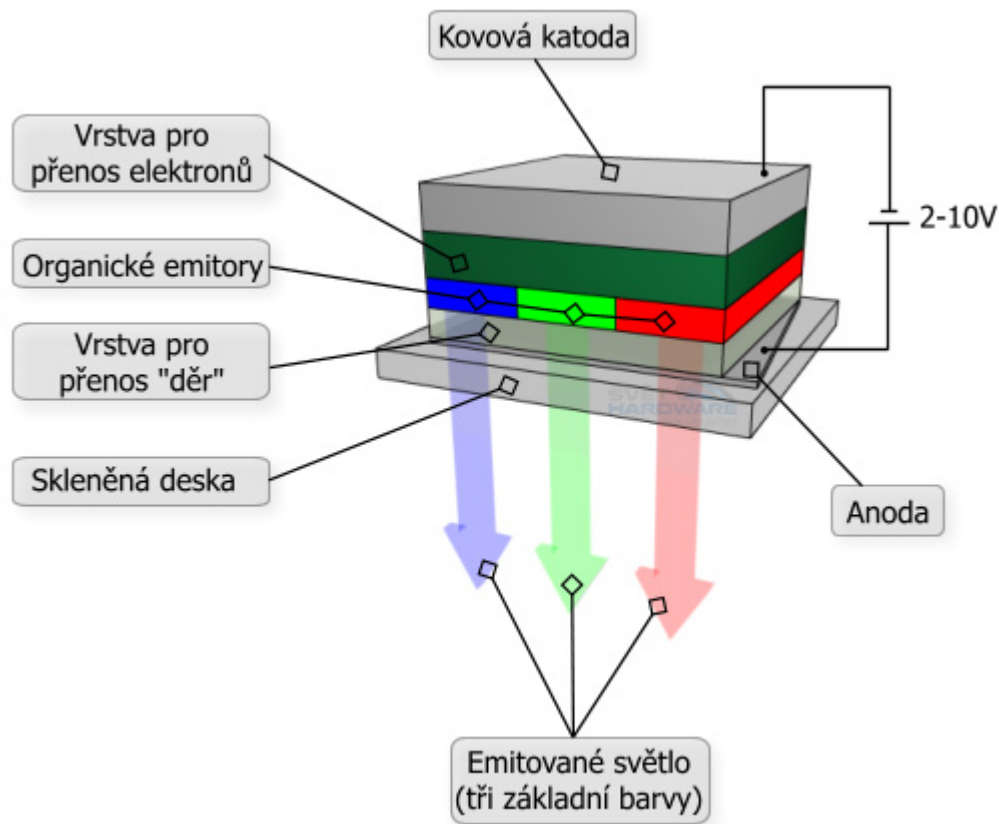
Buď podobně jako OLED/LCD, kdy jeden CFA (Colorized Filter Array) je tvořen 4 B/W subpixely, na kterých jsou barevné filtry.

A nebo používají více různých barevných částic v jedné kapsli (krom černé), z nichž každá barva jinak reaguje na elektromagnetické síly. Různými náboji jsou pak barvy řazeny a tvoří tak odstíny barev.

12) Pentium Pro - vykonávání instrukcí

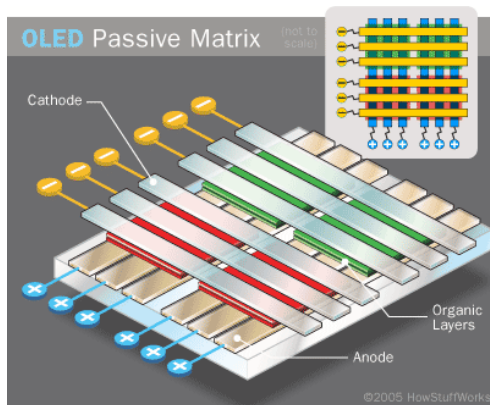
13) I2C sběrnice

14) OLED, princip aktivní, pasivní, využití, výhody a nevýhody.



- hlavním prvkem displeje je organická dioda emitující světlo
- v buňce je kovová katoda, vrstva pro přenos elektronů, organická vrstva emitující světlo, vrstva pro přenos děr a průhledná anoda
- po přivedení napětí na obě elektrody se začnou hromadit elektrony na straně anody (vrstva pro elektrony), díry (kladné částice) se nahromadí u katody (vrstva pro přenos děr)
- v organické vrstvě začne docházet ke srážkám mezi elektrony a dírami
- srážky vyvolají jejich vzájemnou eliminaci (rekombinace), přičemž dojde k vyzáření fotonu

Pasivní: Obsahuje mřížku z na sebe kolmých elektrod a anod. Místa, kde se kříží tvoří pixely, které emitují světlo. Výroba je jednodušší, ale kvůli externím ovládacím obvodům mají vyšší spotřebu než aktivní (ovšem stále nižší než LCD).

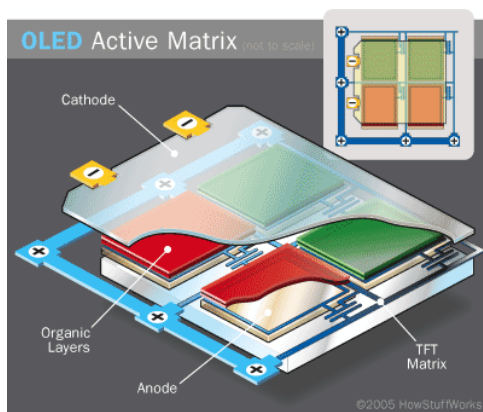


Aktivní: Každá buňka (subpixel) má dva tranzistory. Jeden řídí vybíjení a nabíjení a druhý je jako stabilizátor kvůli zajištění konstantní velikosti proudu.

Je tvořen celými vrstvami (místo mřížek anod a katod).

TFT mřížka > Anoda (vrstva) > Organická vrstva > Katoda (vrstva)

Samo TFT je obvod, který určuje jaké pixely se rozsvítí a jaké ne. Proto má menší spotřebu.



Výhody:

- nepotřebují podsvícení, vysoký kontrast, velmi tenké, nízká spotřeba, dobrý pozorovací úhel, minimální zpoždění, snadná výroba, možnost použití v pružných displejích

Nevýhody:

- životnost, náchylnost na zničení ~~vedou~~ vlhkostí

15) Periferie monolitu**Vstupní a výstupní brány**

Jako vstup a výstup monolitů je paralelní brána – port. Obvykle 4 nebo 8 jednobitové vývody, kde lze současně zapisovat i číst logické hodnoty. Lze nastavit chování vývodů jako vstupní, nebo výstupní. K mnoha účelům.

Sériové rozhraní**Čítače a časovače**

Čítač vnějších událostí: nastavení zda bude inkrementovat při náběžné, nebo sestupné hraně vnějšího signálu.

Časovač: Není inkrementován vnějším signálem, ale hodinovým signálem.

A/D převodníky

Převod vnějšího analogového signálu na digitální

Komparační A/D převodník: Porovnávání měřené veličiny (napětí) s referenční hodnotou

A/D převodníky s D/A převodem

Integrační A/D převodník

Převodníky s časovacím RC článkem: využívá se nabíjení/vybíjení kondenzátoru a měří se čas pomocí časovače

PWM (Pulse Width Modulation)

Šířková modulace pulzů. Pro převod z šířkově upravených pulsů na analog se používá RC článek.

Obvod reálného času RTC (Real Time Clock)

Jde o řízení v reálném čase. Speciální periferie

Speciální periferie**16) RISC a přínos (zřetězení)****17) Popište a nakreslete funkčnost procesoru I7 (i5 nebo i3)**