

V textech je občas nějaké "viz C1" apod. Je to část jak vidíte zde A, tak níže bude C a otázka 1. Prosím, pokud něčemu nerozumíte, používejte komentáře (označit -> pravým -> komentář), pokud chcete doplnit/opravit informaci, můžete vepsat.

## A

*1. Nakreslete zjednodušené schéma a popište základní principy činnosti počítače, jak je definoval von Neumann. Jaké jsou obecné výhody a nevýhody Neumannovy koncepce? (7)*

zbytek viz C1

*2. Popište základní konstrukční rysy (vlastnosti, řešení) procesorů RISC a důvody, proč jsou implementovány. Vysvětlete princip zřetěženého zpracování instrukcí v RISC procesorech a jeho přínos. (7)*

- tato architektura byla vymyšlena a navržena v průběhu 70. let, cílem bylo redukovat složitost výroby procesorů a dosáhnout zpracování jedné instrukce během jednoho strojového cyklu

### **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 redukovaným 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 zřetěžení 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

**Zřetěženého zpracování instrukcí:** každá instrukce je rozdělení zpracování jedné instrukce mezi

různé části procesoru a tím i dosažení možnosti zpracovávat více instrukcí najednou.

*3. Co to jsou dynamické paměti? Nakreslete a popište realizaci paměťové buňky a organizaci buněk na čipu. Charakterizujte stručně historický vývoj těchto pamětí a kde se používají. (7)*

viz C3

Používají se jako operační paměť v počítači.

FPM DRAM - není nutné vkládat znovu adresu řádku

EDO DRAM - čas mezi CAS signály je kratší

SDRAM - Synchroní s systémovou frekvencí

DDR1,2,3,4 - používá jak sestupnou tak náběžnou hranu

*4. Co to je architektura CUDA. Nakreslete zobecněný pohled na architekturu CUDA. Vysvětlete, jaký je postup výpočtu programu, využívajícího CUDA. Jaké rozšíření jazyka C/C++ si technologie CUDA vyžádala? (7)*

viz C4

Tyto dva obrázky jsou totožné: Jeden multiprocessor = SM; Procesor = Core

*5. Nakreslete a popište princip fungování technologie E-Ink, používané pro zobrazovací jednotky. Jaké má tato technologie výhody a nevýhody, kde se používá? (7)*

Energie se spotřebovává pouze při překreslení zobrazových bodů, jinak molekuly vydrží velmi dlouhou dobu zůstat na místě.

### **Výhody**

- Vysoké rozlišení
- Dobrý kontrast a pozorovací úhly
- Čitelnost na přímém slunci
- Není nutné podsvícení
- Možnost používat na pružném podkladu
- Minimální spotřeba při překreslení, nulová spotřeba při statickém zobrazení

### **Nevýhody**

- Pouze 16 odstínů šedi (existují barevné verze po 4096 barev)
- Špatné barevné rozlišení
- Zpoždění při překreslování

*6. Popište technické vlastnosti a konstrukci jednoho RISC procesoru, který znáte. Např. ARM, MIPS, nebo Itanium.. . (7)*

MIPS (microprocessor without interlocked pipeline stages)

- používá non-interlocked pipeline → programátor/překladač se musí starat o to, aby paralelně probíhající žetězené instrukce nevyužívaly stejných prostředků počítače
- jsou navrženy podle RISC filozofie

- přístup do paměti pouze pomocí instrukcí LOAD/STORE (ostatní instrukce šahají do registrů) - obsahují 32 registrů, každá instrukce má délku 32 bitů
- používají se např. v herních konzolích, set-top boxech, mobilech a tiskárnách - nejprve byl navržen kompilátor, až poté implementován obvod, s důrazem na rychlost

*7. Nakreslete a popište fungování pevného disku. Jak jsou v disku organizována data? Jakým způsobem (fyzikálním principem) se na tato média ukládají a následně vybavují informace? (7)*

[http://www.outech-havirov.cz/chmiel/files/ovt\\_epo\\_ps/et/cast2\\_03\\_pmv\\_princip.pdf](http://www.outech-havirov.cz/chmiel/files/ovt_epo_ps/et/cast2_03_pmv_princip.pdf)

### **Stopa**

- rozdělení plotny na kružnice, u moderních disků statisíce

### **Cylindr**

- stopy nad sebou tvoří pomyslný válec - cylindr

### **Sektor**

- nejmenší adresovatelná jednotka - velikost je dána formátováním (kdysi 512 B, dnes 4096 B) - číslování od 1, na začátku je hlavička s číslem sektoru, na konci je zakončení sektoru, obsahující kontrolní součet
- před a za sektorem jsou mezisektorové mezery, do kterých nelze uložit data

*8. Nakreslete a popište princip multiplexního řízení zobrazovací jednotky Kdy, kde a proč je potřeba. (6)*

*Stačí zjednodušeně nakreslit, že do znaku vede jeden adresní vodič a x datových.*

- multiplexované displeje nejsou ovládány naráz, ale po částech (řádky, znaky, pixely) - data nejsou zasílána naráz, ale za sebou (např. řádek, sloupec, znak) - používá se časový multiplex, na displeji se zobrazuje naráz pouze část dat, ale vzhledem k rychlému blikání (alespoň 30 - 50 za sekundu) se to lidskému oku jeví jako souvislý obraz

Vodičů: adresa znaku + číslo segmentu. (Narozdíl od statického kde je  $n*m$  vodičů.)

**B**

B1 = C1

*2. Popište základní konstrukční rysy (vlastnosti, řešení) procesorů RISC a důvody, proč jsou implementovány. Co to je predikce skoku, kdy a jak se v procesoru používá, způsoby implementace. (7)*

viz A2

Vzniká problém při větvení v kódu, pokud dojde ke skoku, musí se pipeline vymazat a načíst odznovu - zpomalení. Používá se proto např. predikce skoku, která se snaží předpovědět, jestli se skok provede nebo ne (buď staticky - predikce je součástí instrukce anebo dynamicky - např. 2bitová predikce favorizuje jednu akci až po dvou stejných operacích po sobě). Popřípadě v superskalární architektuře lze provádět obě dvě možnosti (kód při skoku i kód, pokud se skok neprovede) současně, a neplatná možnost se poté zahodí. Skoky lze také odkládat, pokud za nimi následují na skoku nezávislé instrukce.

*3. Co to jsou statické paměti? Nakreslete a popište realizaci pamětové bunky a organizaci buněk na čipu. Kde a v jakých kapacitách se tyto paměti používají, jak jsou rychlé. Je možné je použít jako hlavní paměť počítače a proč? (7)*

Jsou extrémně rychlé paměti, kapacity jsou velmi malé. Najdeme je v registrech, Lx cache (x= číslo). Můžeme je použít jako hlavní paměť, ale bude to sakra drahé, takže raději ne.

*4. Nakreslete a popište architekturu NVidia Fermi. Vysvětlete, jaký je postup výpočtu programu, využívajícího CUDA. Jaké rozšíření jazyka C/C++ si technologie CUDA vyžádala? (7)*

- 3 miliardy tranzistorů, 512 jader, 64-bitová adresní sběrnice
- s CPU komunikuje přes host interface (8 GiB/s)
- 6x64b řadiče typu GDDR5
- optimalizace pro výpočty s dvojitou přesností (double-precision floating point) - FMA instrukce (fused multiply-add, pouze s jedním zaokrouhlením), 1 takt SP, 2 takty DP - ECC ochrana

viz C4

5. *Nakreslete a popište princip technologie OLED. Co to je AMOLED a PMOLED, stručně popište rozdíl. Jaké má tato technologie výhody a nevýhody, kde se používá? (7)*

Jsou organizovány do matice (matrix).

Pasivní: dvě na sebe kolmé elektrody, procházející celou šířkou a výškou displeje.

Aktivní: každá bunka má tranzistor.

### **Výhody**

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

### **Nevýhody**

- prakticky žádné, body se rády vypalují

*6. Charakterizujte obecné konstrukční vlastnosti monolitických počítačů. Stručně popište architekturu procesoru ATmega32, používaného ve cvičení. (7)*

- 40 pinů (dva napájecí, dva pro oscilátor, jeden pro RESET, tři pro napájení a poskytování referenčního napětí analogově-digitálnímu převodníku a 32 pro libovolné I/O operace)
- piny umí fungovat v analogovém módu
- 3 vnitřní timery
- aproximační ADC (analog-to-digital converter)

- USART (universal synchronous-asynchronous receiver/transmitter), SPI, two wire interface - 32 KiB Flash programovatelné paměti, 1 KiB EEPROM, 2 KiB interní SRAM - frekvence 1 až 16 MHz
- lze programovat pomocí SPI, paralelně anebo pomocí JTAGu (Joint Test Action Group, slouží k testování plošných spojů a interních obvodů procesoru) - lze nastavit fuse bity

*7. Charakterizujte Flynnovu klasifikaci paralelních systémů. Kam v této klasifikaci patří SMP systémy. Jaké statické a dynamické sítě a přepínače se v paralelních systémech používají pro přístup procesorů do globální paměti? (7)*

Flynnovou klasifikací lze třídit paralelní systémy podle počtu toků instrukcí a dat.

**SMP** - symetrický multiprocessing, všechny procesory jsou stejně výkonné -

MIMD

**Sítě:**

Statické: cesta, kružnice, 2x kružnice, bin. strom, krychle

Dynamické: křížové, promíchání s výměnou, sběrnice

*8. Vysvětlete pojem PWM (Pulse Wide Modulation) - pulsní šířková modulace, nebo li také střída. Jak lze PWM použít pro řízení jasu diody žárovky? (6)*

Jedná se řízení v poměru vypnutí a zapnutí.

60% času je 1 => rozsvíceno na 60% jasu

C

*1. Nakreslete zjednodušené schéma harvardské architektury a popište základní principy činnosti počítače, jak je definoval von Neumann. Jaké jsou obecné výhody a nevýhody harvardské architektury? (7)*

von Neumann:

1. Počítač se skládá z paměti, řídicí jednotky, aritmetické jednotky, vstup a výstupní jednotky
2. Struktura pc je nezávislá na typu řešené ulohy, počítač se programuje podle obsahu v paměti
3. Následující krok je závislý na kroku předchozím
4. Instrukce a operandy jsou v téže paměti
5. Paměť je rozdělena do buněk stejné velikosti, jejich pořadová čísla se využívají jako adresy
6. Program je tvořen posloupností instrukcí, ty se vykonávají jednotlivě v pořadí, v jakém jsou zapsány do paměti
7. Změna pořadí provádění instrukcí se provede instrukcí podmíněného/nepodmíněného skoku
8. Pro reprezentaci instrukcí, čísel, adres a znaků se používá dvojková soustava

Hardvardské arch.:

1. Výhody: oddělená paměť dat a programu má výhody:
  - a. program nemůže přepsat sám sebe
  - b. paměti mohou být vyrobeny odlišně
  - c. každá paměť může mít odlišnou velikost
  - d. dvě sběrnice umožňují paralelismus
2. Nevýhody:
  - a. dvě sběrnice kladou vyšší nároky na řídicí jednotku + náklady
  - b. paměť nelze použít zároveň pro program a data

*2. Nakreslete a popište architekturu procesoru P-Pro, nebo jeho libovolného následníka - Pentium II/III/M, Core, Core 2. Popište podrobněji jakým způsobem jsou v procesoru dekódovány instrukce a v jakém pořadí prováděny. (7)*



**Bus interface unit:** stará se o komunikaci mezi procesorem a systémovou sběrnici

**L2:** Uchovává informace mezi výpočetní jednotkou a hlavní pamětí, pro rychlou správu dat. (Z dob Pentii: Write-Back Data zůstávají v L2 a nesynchronizují se okamžitě, a až jsou odstraněny flushnou do hlavní paměti.)

**L1: Instrukční:** uchovává instrukce, **Datová** - uchovává data pro vykonávání

instrukcí **Fetch/Decode Unit:** dekóduje CISC instrukci na RISC (mikrooperace).

**Instruction pool:** místo, kde jsou uloženy mikrooperace (až 40), i ty hotové.

**Dispatch/Execute Unit:** jednotka vybírá instrukce (i mimo pořadí - out of order, díky tomu se provádění může optimalizovat/maximalizovat) -> po vykonání jsou uloženy do banky instrukcí (Instruction pool).

**Retire Unit:** hotové instrukce jsou vybírány a data uloženy z L1 do L2.

*3. Co to jsou dynamické paměti? Nakreslete a popište realizaci paměťové buňky a organizaci buněk na čipu. Popište a nakreslete, jak je v počítači realizována virtuální paměť. (7)*

Dynamické paměti fungují na kondenzátorech a tranzistorech (báze je připojena na adresový vodič).  
**Virtuální pamět:**

Způsob přidělování paměti procesům tak, aby byla lineární a zdánlivě využívala celý adresní prostor.

*4. Co to je architektura CUDA. Nakreslete a vysvětlete organizaci vláken pro realizaci výpočtu. Vysvětlete, jaký je postup výpočtu programu, využívajícího CUDA. Jaké rozšíření jazyka C/C++ si technologie CUDA vyžádala? (7)*

Je programové i hardwarové rozšíření od Nvidia, které dává možnost jednoduchého využívání výkonu GPU.

**CUDA jádro:** obsahuje jednu FPU a jednu ALU jednotku (integer unit)

#### **Roz šíření C/C++**

kvantifikátory funkcí a proměnných: `__global__`, `__device__`, `__host__`

spouštění kernelů: `fn<<<pocet_bloku,pocet_vlaken>>>(params)`

vestavěné datové typy (není přímo rozšíření): `intx`, `charx`, ... (kde x je číslo)

#### **Postup výpočtů**

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

*5. Nakreslete a popište princip technologie LCD. Co to je aktivní a pasivní displej, stručně popište rozdíl. Jaké má tato technologie výhody a nevýhody, kde se používá? (7)*

#### **Pasivní displeje**

- obsahují mřížku vodičů, body se nacházejí na průsečících mřížky
- při vyšším počtu bodů narůstá potřebné napětí → rozostřený obraz, velká doba odezvy (3 FPS) - nevhodné pro hry, filmy, televizi atd.
- používá se v zařízeních s malým displejem (hodinky)

#### **Aktivní displeje (TFT - thin film transistor)**

- každý průsečík v matici obsahuje tranzistor nebo diodu, která řídí činnost daného bodu, je možné velmi přesně a rychle ovládat svítivost každého bodu
- šířka subpixelů je zhruba 0.2 - 0.3 mm

#### **Výhody**

- kvalita obrazu, životnost, spotřeba energie, odrazivost a oslnivost, bez emisí

#### **Nevýhody**

- citlivost na teplotu, pevné rozlišení, vadné pixely, doba odezvy

*6. Nakreslete a popište typické základní periférie monolitických počítačů. (7)*

**V/V brány** - nejjednodušší způsob komunikace - paralelní brána (port, realizován registrem - 4 nebo 8 jednobitových vývodů, u kterých lze nastavit, jestli slouží k zápisu nebo čtení), vstupy mohou používat pull-down nebo pull-up rezistory pro udržení logické 0 nebo 1 při nezapojeném vstupu

**Čítač/časovač** - registry, které jsou inkrementovány (čítač vnějšími událostmi, časovač vnitřním signálem procesoru), při přetečení vyvolají přerušení

#### **A/D převodníky**

Komparační - porovnává napětí s referenční hodnotou, rychlé, ale jejich potřebný počet narůstá velmi rychle, pokud chceme vyšší rozlišení

D/A převod - obsahuje komparátor

Sledovací - který zkouší porovnat napětí se vstupem, pokud neodpovídá, tak jej inkrementuje nebo dekrementuje (nehodí se pro skokově se měnící vstupy)

Aproximační - používá metodu půlení intervalu (maximálně  $n$  kroků,  $n$  je počet bitů rozlišení převodníku)

Integrační - používá metodu dvojité integrace (potřebuje externí kondenzátor) S RC článkem - doba nabití/vybití kondenzátoru závisí na vstupním odporu

## D/A převodníky

PWM (pulse-width modulation) - šířková modulace pulzů

- výstupní napětí se mění pomocí měnění poměru doby logické 1 a 0 na výstupu - převod šířkově modulovaných pulzů na analogovou veličinu zajišťuje RC článek (musí platit  $R \cdot C \gg T$ )

Paralelní převodníky

- přímý převod číselné hodnoty na proud (používá se odporová síť R-2R - pouze odpory hodnoty R a dvojnásobku R)

## Obvody RTC

- potřebují záložní zdroj (baterii), aby mohly "počítání" času provádět i při odpojení napájení - je nutno zajistit atomicitu čtení času (hodnoty se na pozadí mění v čase), řeší se vhodným programovým řízením anebo pomocným registrem (do toho se uloží čas před čtením a během čtení se nemůže změnit)

*7. Charakterizujte základní dělení paralelních systémů podle organizace paměti. Kam v této klasifikaci patří SMP systémy. Jaké statické a dynamické sítě a přepínače se v paralelních systémech používají pro přístup procesorů do globální paměti? (7)*

## Paměť

Sdílený prostor - procesy mají přístup k jedné paměti

Distribuovaná paměť - každý proces má vlastní adresní paměť

Paměť	Ne (sdílené)	Ano (sdílené)
Ne (distribuované)	klasické 1x CPU	UMA, SMP (desítky CPU)
Ano (distribuované)	Clustery, multipočítače	NUMA

**SMP** - symetrický multiprocessing, všechny procesory jsou stejně výkonné

## Sítě:

Statické: cesta, kružnice, 2x kružnice, bin. strom, krychle

Dynamické: křížové, promíchání s výměnou, sběrnice

*8. Vlákna umožňují programátorovi lépe využít které části počítače? Jak musí programátor implementovat svůj program, aby co nejlépe využil výpočetní výkon počítače? (6)*

Lépe využít výpočetní jednotky, popřípadě přístupy ke zdrojům.

Non-blocking, uměl spávně a efektivně komunikovat mezi vlákny.