

Architektura počítačů

**Studijní materiál pro předmět
Architektury počítačů a paralelních systémů**

Ing. Petr Olivka, Ph.D.
katedra informatiky FEI VŠB-TU Ostrava
email: petr.olivka@vsb.cz

Ostrava, 2021

1 Architektura počítačů

Pojem architektura je převzat z jiného oboru lidské činnosti, než počítače. Neurčuje jednoznačné definice, schémata či principy. Hovoří o tom, že počítač se skládá z menších částí a teprve jejich vzájemné propojení a soulad vytváří funkční celek. Většina konstrukčních hledisek je neměřitelných a vychází ze zkušeností, technického citu, fantazie a tvůrčí invence autorů.

Architekturu je potřeba moderně chápat jako pohled na podstatné vlastnosti počítačů. Ty je možno rozdělit do čtyř základních kategorií:

- **struktura, uspořádání:** popis jednotlivých funkčních částí a jejich propojení,
- **součinnost, interakce:** popis řízení dynamické komunikace mezi funkčními bloky,
- **realizace, provedení:** popisuje vnitřní strukturu jednotlivých funkčních bloků,
- **funkcionalita, činnost:** výsledné chování počítače jako celku.

Lidé z různých oborů se budou na počítač dívat samozřejmě s preferováním různých hledisek. Pro většinu uživatelů však bude asi nejdůležitější poslední výše jmenované hledisko.

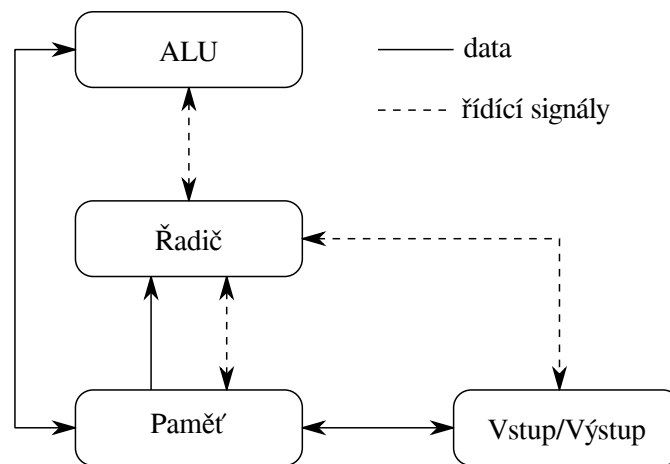
Pokud je ale cílem pochopit principy činnosti počítače, je zapotřebí se zabývat všemi uvedenými hledisky.

1.1 Základní koncepce

Americký matematik maďarského původu John von Neumann definoval v roce 1945 základní koncepci počítače EDVAC (Electronic Discrete Variable Automatic Computer). Byl to výsledek výzkumu týmu odborníků Univerzity státu Pensylvánie ve Filadelfii, kteří tak položili základ koncepce počítače řízeného obsahem paměti. Od té doby se objevilo několik různých modifikací i odlišných modelů, ale v podstatě počítače, tak jak je známe a používáme v dnešní době, vychází z tohoto původního modelu.

Před samotným sestavením počítače si von Neumann stanovil určitá kritéria a principy, které musí počítač splňovat. Ty se dají ve stručnosti shrnout do následujících bodů:

1. Struktura počítače je nezávislá na typu řešené úlohy, počítač se programuje obsahem paměti.
2. Strojové instrukce a operandy (data) jsou v téže paměti. Pro instrukce i pro data je tedy možno přistupovat jednotným způsobem.



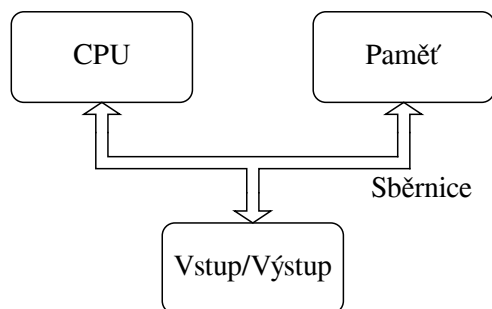
Obrázek 1: Původní schéma počítače podle von Neumanna

3. Paměť je rozdělena do buněk stejné velikosti, jejich pořadová čísla se využívají jako adresy.
4. Následující krok počítače je závislý na kroku předchozím.
5. Program je tvořen posloupností instrukcí, ty se vykonávají sekvenčně, tedy jednotlivě a v pořadí v jakém jsou zapsány do paměti.
6. Změna pořadí provádění instrukcí se provede instrukcí podmíněného či nepodmíněného skoku.
7. Pro reprezentaci instrukcí, čísel, adres, znaků atd., se používá dvojková číselná soustava.
8. Počítač se skládá z řídicí a aritmeticko-logické jednotky (dnes společně označované jako CPU), paměti a vstupně-výstupní jednotky.

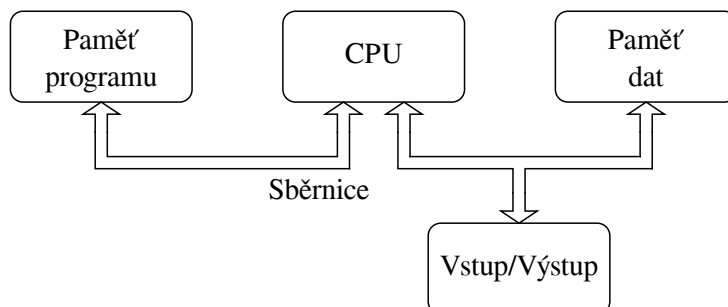
Původní schéma počítače, jak jej navrhl von Neumann, je na obrázku 1. Na tomto schématu je ještě vidět samostatně řadič a ALU. Ve schématu je také vidět role řadiče, který řídicími signály ovládá jednotlivé části počítače. Data směřují vždy z paměti nebo do paměti.

Bylo však již zmíněno, že v dnešní době je řídicí obvod počítače integrován společně s ALU do jednoho obvodu, který je označován jako CPU – centrální procesorová jednotka. Tohoto označení se užívá přibližně od šedesátých let minulého století. Dále se také již v počátcích vývoje počítačů sdružily datové a řídicí signály do společné skupiny, označované jako sběrnice.

Tyto změny však nijak nemění základní principy fungování počítače. Je však možno díky těmto úpravám znázornit architekturu počítače dle von Neumanna jednodušeji a přehledněji, jak je vidět na obrázku 2.



Obrázek 2: Počítač podle von Neumanna



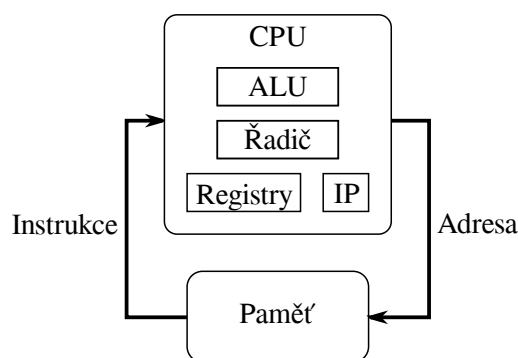
Obrázek 3: Harvardská architektura počítače

Počítač dle von Neumanna, zobrazený na obrázku 2 se tedy z dnešního pohledu skládá z procesoru, paměti a periférií a tyto části jsou propojeny sběrnici. Pokud by některá z těchto částí chyběla, nemůžeme výsledek považovat za funkční počítač.

1.2 Harvardská architektura

Několik let po von Neumannovi, přišel vývojový tým odborníků z Harvardské univerzity s vlastní koncepcí počítače, která se sice od Neumannovy příliš nelišila, ale částečně řešila některé její nedostatky. V podstatě jde pouze o oddělení paměti pro data a program. Schéma harvardské koncepce je na obrázku 3.

Základní koncepce a principy fungování jsou u harvardské i von Neumannovy architektury stejné a v dalších tématických okruzích APPS nebude potřeba se na obě architektury dívat jako na dvě odlišné.



Obrázek 4: Zjednodušený princip fungování počítače

1.3 Princip fungování počítače

Pro vysvětlení základního principu fungování počítače bude stačit vybrat z obrázků 2 a 3 pouze procesor a paměť. Dále pak je potřeba mít na paměti informace z kapitoly 1.1. Zde budou důležité zejména informace: počítač je programován obsahem paměti, instrukce z paměti se vykonávají sekvenčně a každý následující krok závisí na kroku předchozím.

Z toho je zřejmé, že procesor je sekvenční obvod a vstupem tohoto obvodu budou strojové instrukce z paměti. Pro názornost a další vysvětlení fungování je procesor i s pamětí zobrazen na obrázku 4.

Na obrázku 4 je jako součást procesoru pro připomenutí uvedena i jeho řídicí a výpočetní část. Procesor, jako sekvenční obvod, využívá samozřejmě paměť. Dosud byla paměť uváděna jako samostatná část počítače. Ve skutečnosti je velmi malá část paměti počítače umístěna i přímo v procesoru a to ve formě registrů.

Jedním z těchto registrů je IP (Instruction Pointer), ukazatel do paměti na instrukci, která má být provedena, či je právě prováděna. U některých procesorů bývá tento registr označován jako PC (Program Counter), jeho význam je však stejný jako IP.

Má-li procesor vykonat strojovou instrukci, jeho řídicí část využije aktuální hodnotu registru IP a prostřednictvím sběrnice si vyžádá (přečte) z paměti instrukci na adrese IP. Na obrázku 4 je toto čtení instrukce z paměti znázorněno šipkami.

Poté co procesor získá z paměti strojovou instrukci, tak ji provede. Následně zvýší IP o délku právě provedené instrukce a celý cyklus čtení a provedení instrukce se opakuje.

Pokud právě prováděná instrukce potřebuje data z paměti, tak si je procesor vyžádá stejně, jako si načel instrukci samotnou. Dle obrázku 4 však z paměti do procesoru nesměřuje instrukce, ale data. Adresa těchto

dat je obvykle součástí prováděné strojové instrukce.

Uložení výsledku řeší procesor podle dokončené strojové instrukce. Výsledky se mohou ukládat do registrů a v tom případě zůstávají v procesoru. Do paměti se výsledky strojové instrukce uloží opět prostřednictvím sběrnice.

1.4 Základních vlastnosti architektur počítačů

Pro popis základních vlastností a pro porovnání obou koncepcí počítačů je možno vycházet ze zjednodušených schémat na obrázku 2 a na obrázku 3.

Základním nedostatkem obou koncepcí je sekvenční vykonávání instrukcí, které sice umožňuje snadnou implementaci systému, ale nepovoluje dnes tolik potřebné paralelní zpracování. Viz kapitola 1.3, jeden IP procesoru neumožňuje rozpracovat více sekvencí instrukcí. Paralelizmy se proto musí simulovat až na úrovni operačního systému.

Úzké místo systému je také ve sběrnicích, které nedovolují přistupovat současně do více míst paměti současně, např. pro instrukci i pro data, a navíc sběrnice dovolují v daném okamžiku přenos dat jen jedním směrem.

Porovnání vlastností obou koncepcí jednotlivě:

- 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 úzké místo.

- 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 souběžně přistupovat pro instrukce i data.

- 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.

1.5 Kapacita paměti

Paměť počítače, jak bylo uvedeno v kapitole 1.1, se skládá z buněk stejné velikosti. Velikost těchto buněk je různým v počítačích, zejména mikropočítačích, mnohdy odlišná.

Naprosto nejrozšířenější velikostí paměťové buňky je však buňka velikosti jednoho bajtu (byte, značí se B), který obsahuje 8 bitů (značí se b). Pojem bajt vznikl v šedesátých letech minulého století ve firmě IBM. V technické praxi je také užíván alternativní pojem pro bajt, a to oktét.

Velikost paměťové buňky jeden bajt se stala také základní jednotkou pro vyjádření kapacity paměti. Jednotka bajt a její násobky jsou standardizovány mezinárodní normou IEC 60027-2. Tato norma je zařazena i mezi české technické normy jako ČSN IEC 60027-2. Cílem této normy je zavedení předpon a jejich zkratk pro násobky mocniny 2, které jsou blízké násobkům SI jednotek, jako je kilo (10^3), mega (10^6), giga (10^9), atd.

Platí totiž přibližný vztah mezi mocninami čísla 10 a 2:

$$2^{10} \cong 10^3, \text{ tedy } 1024 \cong 1000.$$

Kapacita paměti se vyjadřuje v bajtech s předponami kibi (2^{10}), mibi (2^{20}), gibí (2^{30}), atd.

- byte [B] - 8 bitů [b]
- $1KiB = 2^{10}B = 1024B$
- $1MiB = 2^{20}B = 1048576B$
- $1GiB = 2^{30}B = 1073741824B$
- $1TiB = 2^{40}B = 1099511627776B$

Jak je patrné, rozdíly mezi mocninami čísla 10 a čísla 2 postupně narůstají se zvyšující se kapacitou. V případě údaje MB a MiB je tato odchylka skoro 5%, u kapacity udávané v GB a GiB už ale přesahuje 7%.

Ne vždy se v praxi jednotky kapacity používají správně. Důvodem může být někdy neznalost, jindy i záměr. Je proto potřeba si na jednotky a jejich násobky dávat v praxi pozor. Nespolehat se zejména na marketingové informace, ale ověřovat si údaje v technické dokumentaci.

2 Historie vývoje počítačů - generace počítačů

Vývoj počítačů byl v minulém století po jistou dobu klasifikován do několika generací. Dnes se tato klasifikace již nepoužívá. Přesto je v praxi možno narážet na informace, že nějaký princip či nějaká technologie vznikla v době určité počítačové generace. Proto zde velmi krátký přehled jednotlivých generací počítačů.

Vývoj počítačů by členěn postupně do několika generací v závislosti na použité technologii výroby jednotlivých komponent počítačů.

Generace	1	2	3	3.5	4
Rok	1951	1957	1964	1971	1981
Prvky	elektronky	tranzistory	SSI	MSI	LSI
Paměť	buben	ferity	ferity	MSI	LSI
Kapacita paměti	1kB	10kB	1MB	1MB	10MB
Řízení periférií	CPU syn.	CPU as.	kanály	adaptéry	perif. proc

Tabulka 1: Generace počítačů

Z tabulky 1 je patrný vývoj v oblasti komunikace s perifériemi, který zásadním způsobem ovlivňuje průchodnost celého systému. Je zde také vidět postupný nárůst objemu operační paměti a využívání vyššího stupně integrace pro výrobu jednotlivých prvků počítačů.

3 Kontrolní otázky

1. Jaké jsou základní principy fungování počítače.
2. Jaké má výhody a nevýhody architektura počítače dle von Neumanna?
3. Přinesla harvardská architektura nějaká vylepšení proti von Neumannově?
4. Jaká je podpora paralelismu u obou architektur počítačů?
5. Je lepší mít oddělené paměti pro data i program? Proč ano a proč ne?
6. Může fungovat počítač bez paměti či bez periférií?
7. K čemu se v počítači využívá dvojková soustava?
8. Zvyšují sběrnice výkon počítače?
9. Je možné, aby procesoru prováděl instrukce jinak, než sekvenčně?
10. Jak je v počítači organizovaná paměť?