

Grafické adaptéry, zobrazovací jednotky, GPU

Petr Jedlička, Cyril Klimeš, Tomáš Pitner

Předmluva

HTML · Slides · PDF | Přehled předmětu

◎ Cíle přednášky

- Poznat účel, principy, historii a především současné použití a možnosti zobrazovacího subsystému v počítačích a podobných zařízeních, například herních konzolách
- Pochopit, kam jde leading-edge vývoj směrem k integrovaným GPU a pamětem

🔑 **Grafické adaptéry** jsou obvody nebo i celé moduly generující **obraz**, který se pak promítá na monitoru, televizoru, displeji… Bez nich bychom nic neviděli nebo se vše muselo tisknout.

Co potřebujeme zobrazovat?

Záleží na tom, z čeho a kam:

- K ovládání vestavěného zařízení s osmibitovým kontrolérem sloužícího ke snímání hodnot, jejich zobrazení, ukládání a odesílání obvykle vystačí s jednoduchým *displejem* (řádkovým, segmentovým či bitmapovým) nebo dokonce jen *kontrolkami* (LED).
- Tam, kde už zadáváme vstup ve formě čísel a textů a chceme to i vidět, obvykle nastupují větší (LCD) *displeje* nebo *monitory* jako samostatná zařízení.
- Specifickým zobrazovacím zařízením je *e-papír* pro aplikace, kde se obraz nemění často (čtečky knih, ale i displeje průmyslových zařízení, popisky na zboží apod.); dnes už existuje i v barevné podobě.

Pravěk počítačového zobrazování

První generace počítačů (cca 1940–1956)

výstup na **papír** tiskem, **kontrolky** svítí/nesvítí

Druhá generace (cca 1956–1963)

tiskové terminály umožňující komunikaci pomocí psaného textu (ale ne na obrazovce), počátky **znakových CRT** displejů

Třetí generace (cca 1964–1971)

alfa-numerické CRT terminály, video-terminály DEC VT05

První generace: kontrolky



Druhá generace: tiskové terminály

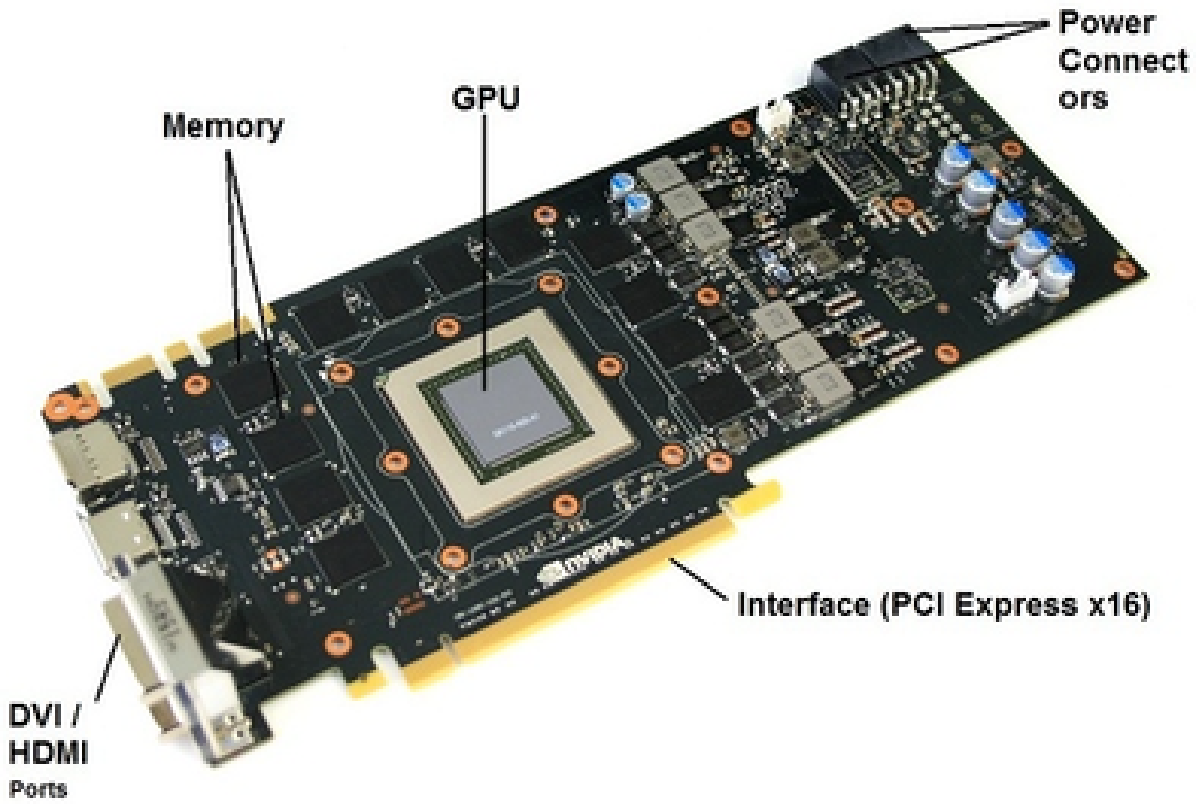


Třetí generace: CRT terminály



Grafické adaptéry

- Počítaje nejpozději třetí generací počítačů jsou tudíž potřeba speciální obvody (zařízení), jejichž účelem je produkovat obraz, který se přenese určitým rozhraním (kabelem) a zobrazí na výstupním zobrazovacím zařízení (monitoru, displeji).
- Taková zařízení starající se o generování obrazu se označují jako *grafické adaptéry* nebo později *Graphics Processing Unit* (GPU).



Princip tvorby obrazu v adaptéru

- Obraz se většinou vykresluje dle dat ve *videopaměti* (*video-RAM*).
- Videopaměť v nejjednodušším provedení obsahuje *bitový obraz* (bitmapu, matici, rastr) o rozměrech `width * height` obrazových bodů (*pixelů*), třeba 320*240 pixelů, dnes ale mnohem více, třeba Full HD (1920*1080) nebo 4K, 8K …
- Grafický adaptér generuje obrazové signály dle obsahu *rastru* neboli *framebufferu*.
- Tato činnost se musí dít kvůli periodickému obnovování obrazu (např. 50x za vteřinu) nepřetržitě a v zásadě nezávisle na CPU: je proto realizována vyhrazeným grafickým adaptérem.

Barva

- V ideálním případě (který dříve rozhodně vždy neplatil) může být každý pixel zobrazen jinou barvou než sousední, nezávisle na nich.
- Počet možných barev jednoho pixelu označujeme jako barevná hloubka.
- Jeden pixel u *monochromatického* ("černobílého") zobrazování bez dalších vylepšení znamená jeden bit.
- Osmibitové i první 16bitové osobní počítače měly kromě monochromu barevná zobrazování se *16 barvami*, tedy co pixel, to 4 bity (2^4 = 16 barev).

RGB skládání barev

Na soudobých displejích již od doby CRT obrazovek přes plazmové a LCD displeje až po OLED je barva pixelu tvořena *složením intenzit barevných složek* **Red Green Blue** (RGB, aditivní míchání barev).

Barevná kódování na dnešních displejích

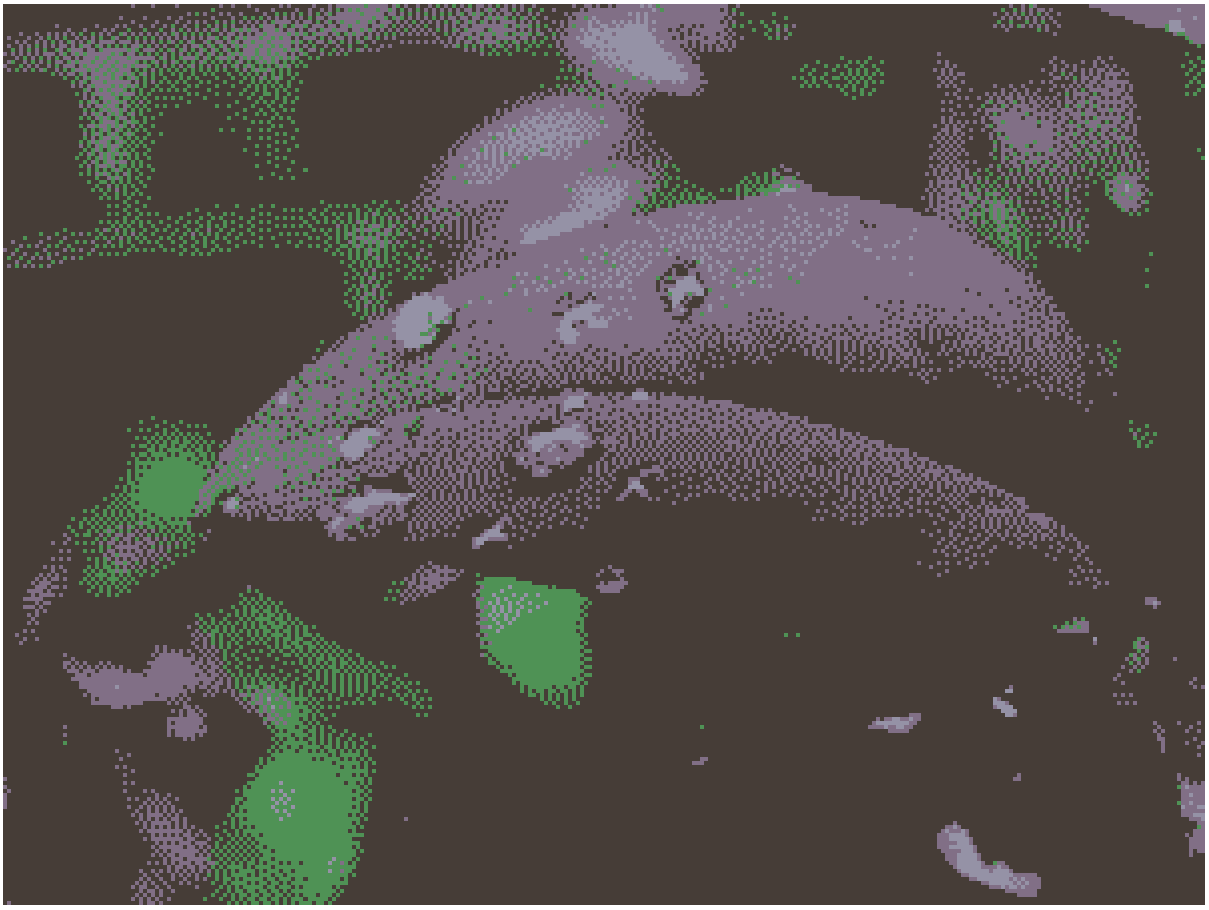
True Color	24bitová barva (2^{24} = 16 777 216 barev) = 8+8+8 bitů každá složka RGB).
True Color + alpha	totéž + 8 bitový alpha-channel určující průhlednost daného pixelu v některých obrazových formátech jako je PNG.
Deep Color	48bitová barva (2^{48} = 281 474 976 710 656 = 281,5 biliónů barev), 16 bitů každá složka RGB.



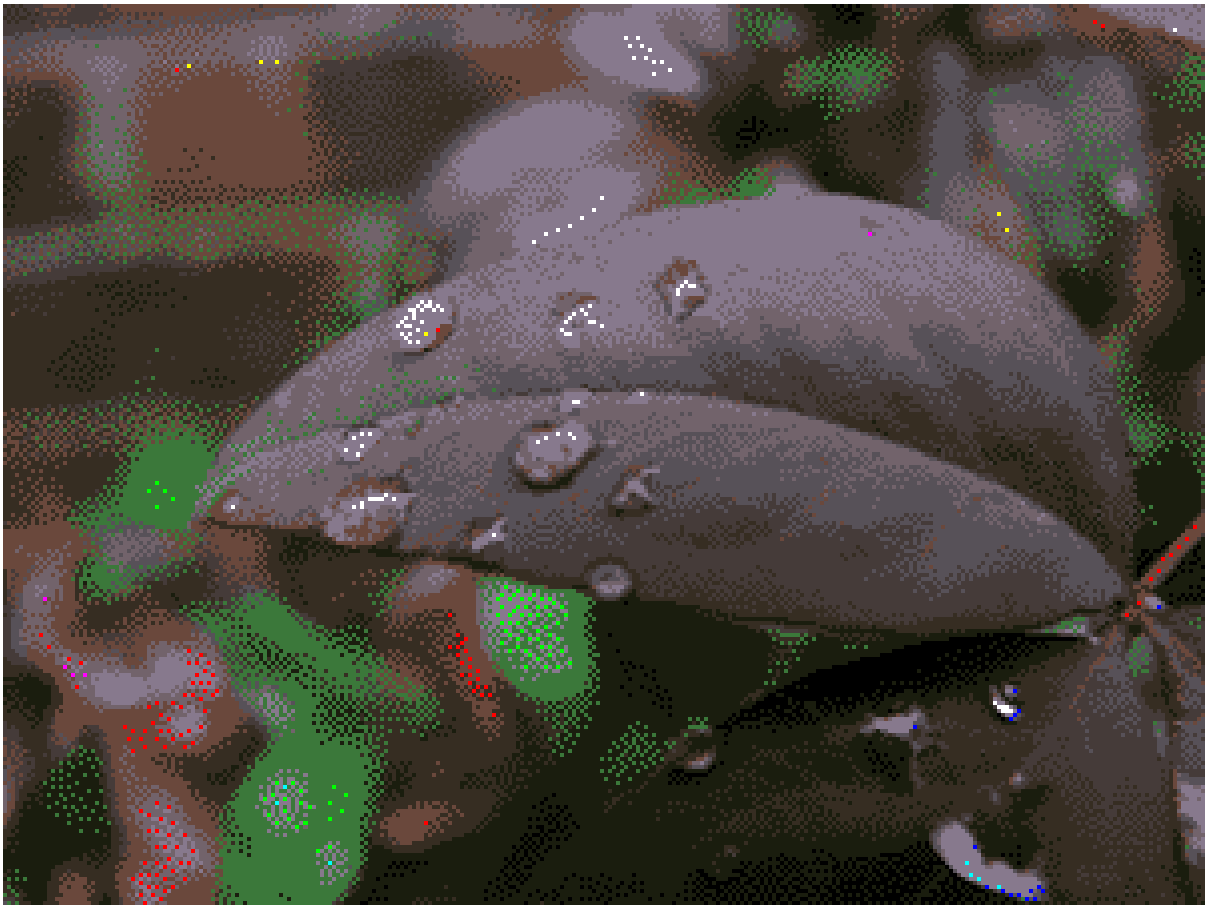
Existují i další varianty kódování barev, například při 16 bitech složky RGB kódovány 5+5+5+1.

Srovnání barevné hloubky (1-2 bit)





Srovnání barevné hloubky (4-8 bit)





- ukázka VGA256

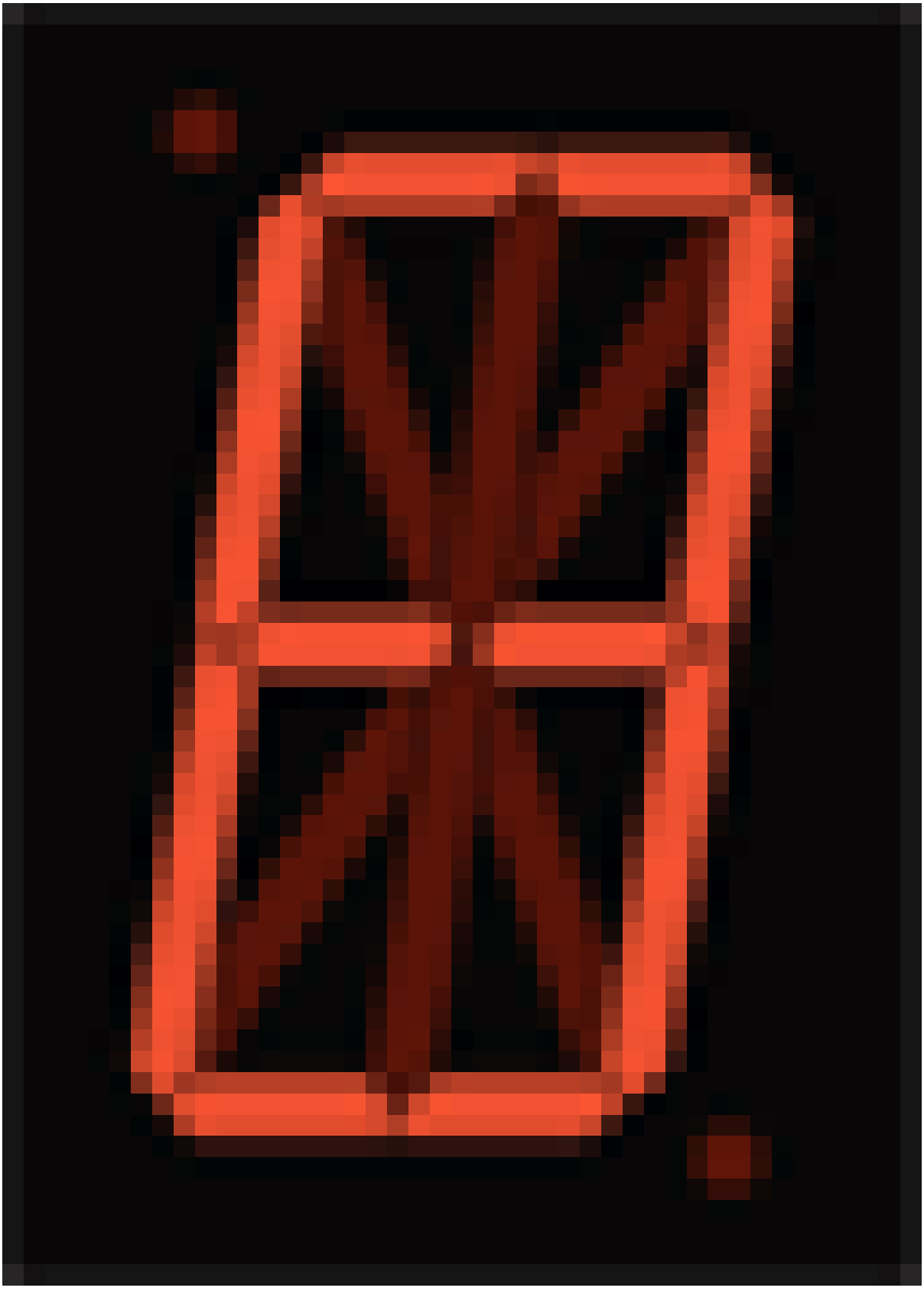
Srovnání barevné hloubky (24 bit)



Displeje

- Z anglického "display" = *ukázat se*
- Jsou koncová zobrazovací zařízení, na nichž finálně vidíme obraz generovaný grafickým adaptérem/GPU.
- V pravěku počítačů stačily jednotlivé žárovky (svítí = 1, nesvítí = 0).
- Pak *segmentové displeje*: uměly ukázat číslice, obvykle navíc tečku, někdy mezi čísly dvojtečku, případně některá písmena a znaky.

Segmentové displeje



- Šestnáctisegmentový, to už je fraječina. Zde na bázi červených LED.



- 7-segmentové displeje s tečkou; rovněž na bázi LED, nemusí být červené.

CRT displeje

- Klasické monitory pracovaly po desetiletí na bázi vakuové obrazovky (CRT, Cathode Ray Tube, katodová trubice).


IBM

Dnešní monitory

- Zobrazení na dnešních displejích/monitorech pracuje obvykle na polovodičových principech a zahrnuje:

Panel	samotná zobrazovací plocha, většinou matice pixelů alespoň 1920 * 1080 (Full HD)
Podsvícení	u LCD panelů nutné; většinou na bázi LED, *miniLED* nebo *kvantových teček*
Řídicí elektronika	dekódování vstupního signálu, převod na povelování zobrazovací matice

Samotné zobrazení

LCD	Liquid Crystal Display Tekuté krystaly mění polarizaci světla podle napětí a tím propouštějí nebo blokují světlo z podsvícení.
OLED	Organic LED Každý pixel sám svítí (není třeba podsvícení), čímž se dosahuje velmi věrné podání barev vč. černé, vysoký kontrast, ale panel se časem "vysvěcuje", slábne, a je drahý.
Mini-LED	využívají tzv. *lokálního stmívání* (local dimming), kdy je plocha rozdělena na zóny a ne každá svítí stejným jasem: tudíž lepší zobrazení černé za nižší cenu než OLED.
Micro-LED	miniaturní LED diody, mnohem větší počty pixelů; podobná kvalita, ale na rozdíl od OLED dlouhá životnost díky použití GaN (nitrid galia) namísto organických polovodičů.

Kvalita monitorů

Rozlišení	Počet pixelů (např. Full HD, QHD, 4K)
Obnovovací frekvence	Počet snímků za sekundu (60 Hz, 144 Hz, 240 Hz…)
Odezva	Doba přepnutí pixelu (např. 1 ms) — důležité pro hry
Proměnná obnovovací frekvence	VRR (Variable Refresh Rate) přizpůsobuje frekvenci snímkování GPU — technologie FreeSync (AMD) a G-Sync (NVIDIA) zabraňuje trhání obrazu

Kvalita monitorů (2)

Jas a kontrast	Svítivost a poměr světlé/tmavé — důležité pro čitelnost a HDR
Barevný gamut	Rozsah barev (např. sRGB, AdobeRGB, DCI-P3)

HDR	Vysoký dynamický rozsah — lepší barvy a světla ve scénách
Úhly pohledu	Jak moc se mění barvy při pohledu ze strany

Typy dle užití

Kancelářská práce	prohlížení obrázků, videí, čtení textu; zpracování text, tabulky, částečně grafika
Programátor	čtení a tvorba textu, složité rozložení na obrazovce, dlouhé "civění" na jedno místo
Grafika	velmi přesné zobrazení barev, vysoký kontrast, HDR, vysoké rozlišení
Hry	rychlá odezva, vysoká obnovovací frekvence

Monitory vs televizory

Televizory	• účel: sledování zdálky, na velké ploše a pohyblivých obrazů vč. kvalitní reprodukce zvuku • vstupy digitální (obvykle HDMI), DVB-T, DVB-C (kabel), DVB-S (satelit), • evt. ještě analogový TV signál, analogový videosignál, SCART
Monitory	• účel: sledování soustředěně detailů zblízka, • stačí horší nebo žádný zvuk • vícero digitálních vstupů (HDMI, DisplayPort, DVI)

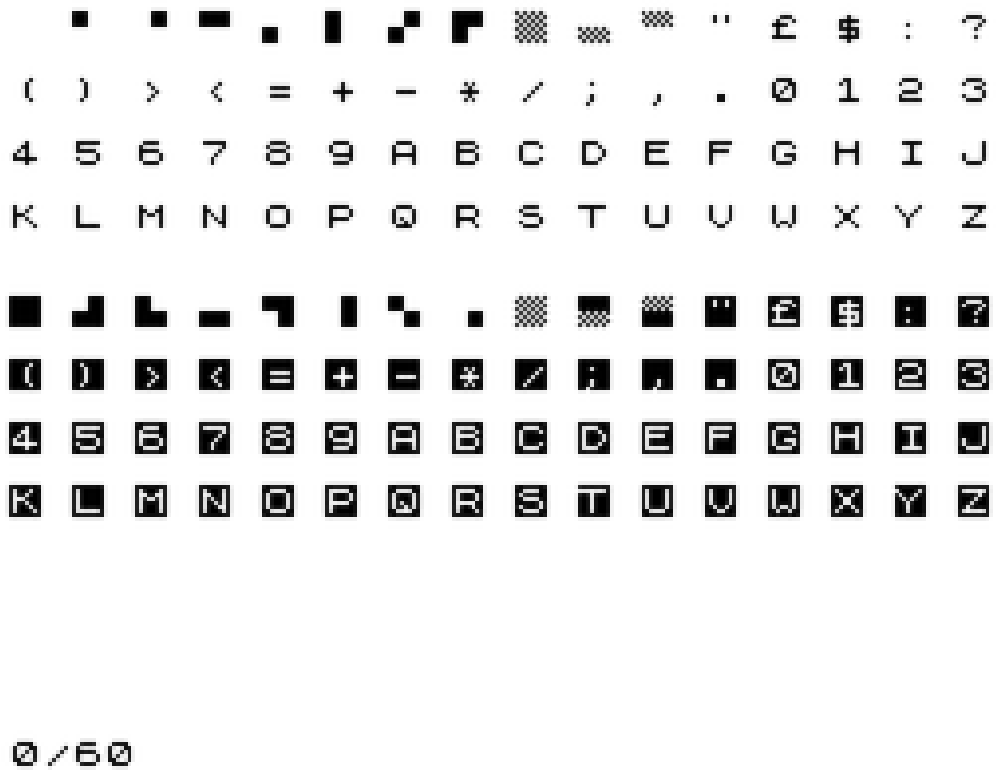
Znaková zobrazení a pseudografika

- Jenže v pravěku počítačů byla i ta *bitmapová paměť drahá* a *procesory pomalé*:
- používala se namísto toho bajtová *matice znaků* po třeba 32 znacích na 24 řádcích a obraz se generoval z bitového obrazu daného znaku uloženého zvlášť.
- Znak byl tedy definován maticí bitů typicky 8*8, co bit, to černo-bílý pixel.

Příklad osmibit ZX81

- Hezký příklad je osmibitový ZX81, kde na každém políčku v matici 32 sloupců na 22 řádcích obrazu mohl být buďto číslo, písmeno z velké anglické abecedy nebo znak *pseudografiky*: různé puntíky, obdélníčky, části úseček, rohy,... pomocí nichž se dala "kreslit" okna, rámečky, atd.
- Každé políčko obrazovky tedy zabralo ve videopaměti pouhý jeden bajt, plus musela existovat pevná či modifikovatelná paměť na pixelové matice jednotlivých znaků pro vykreslování.

Znaková sada zobrazitelná na ZX81



- Sada zahrnovala nejen samotné tvary znaků, ale i provedení téhož tvaru v inverzní podobě
- vše jen *monochrom* (černobílé).
- Navíc ZX81 *nepoužíval ASCII* znakovou sadu, neměl například malá písmena.

Výhody a důvody znakové grafiky

- Obecnou výhodou znakové grafiky byla extrémní paměťová úspornost a tím rychlost práce.
- S obrazem tehdy manipuloval hlavní procesor (CPU), ten navíc neuměl operace s více daty najednou (SIMD), vše se dělo po bajtech.
- Uložení celé obrazovky bylo například na 40*24 = 960 bajtech.
- Přesuny a vykreslování objektů po obrazovce spočívaly v umístění nebo přesunu pár desítek nebo stovek bajtů, což CPU zvládla.

Znaky nebo plný rastr?

- Dnešní grafické adaptéry umějí jak znakový režim — typicky při startu počítače --, tak plně rastrový (grafický) režim.
- Disponují většinou více framebuffery, což dovoluje rychlé přepínání mezi více předpřipravenými obrazy.
- Framebuffery jsou většinou dnes součástí jádra grafického adaptéru spolu se zobrazovacími obvody generujícími výstupní obrazový signál.

Rastr a rozlišení

- **Rastr** (bitmapa, matice) je plný grafický obraz: hodnota barvy prostě pro každý *pixel* rastru.
- Klíčovým parametrem rastru je **rozlišení** — počet pixelů `šířka x výška`.
- S rozlišením roste nejen požadavek na *videopaměť*, ale i na *přenosovou kapacitu* propojení GPU → displej.

Rozlišení v praxi

- Základní rozlišení dnešních displejů a monitorů (poměr stran 16:9):
  - **Full HD** (FHD): 1920 × 1080
  - **QHD**: 2560 × 1440
  - **4K Ultra HD** (UHD): 3840 × 2160
  - **8K**: 7680 × 4320
- Nároky na vyšší počet pixelů dnes jdou ruku v ruce se *zobrazovacími rozhraními* (HDMI 2.2, DisplayPort), jež dokážou přenést i 8K a vyšší rozlišení.



Kolik paměti potřebuje framebuffer?

- Velikost videopaměti pro jeden snímek: `šířka x výška x počet bajtů na pixel`.
- Full HD (1920 × 1080) v True Color (24 bity = 3 bajty na pixel):
  - `1920 x 1080 x 3 = 6 220 800 B` cca **6,2 MB** na snímek.
- Při obnově 60× za sekundu je potřeba z videopaměti přečíst a odeslat cca **373 MB/s**.
- Reálná grafická paměť se ale počítá na jednotky až desítky GB: framebuffer tvoří jen zlomek

toho, co zabírají **textury**, **Z-buffer** a datové struktury 3D scény.

Anatomie grafického adaptéru

- Co musí obsahovat "grafika", aby fungovala?
- **Graphics processing unit**
- **Videopaměť**
- **Paměť BIOSu**

Graphics Processing Unit

- **GPU**, jednotka starající se o výpočetní úkoly s
- tvorbou grafických objektů ve videopaměti (framebufferu),
- jejich přesun, stínování, kopírování;
- jde de facto o počítač se srovnatelnou nebo i vyšší výpočetní silou než CPU, ale na jiné úlohy.

Videopaměť

- Videopaměť je určena pro uložení rastru k 1:1 zobrazení na obrazovce (*framebuffer*),
- ale také pro *Z-buffer*, který spravuje hloubkové souřadnice ve 3D grafice,
- dále *textury*, *vertexové buffery* a přeložené (zkompilované) *shaderové programy*.
- Buďto je vyhrazená na grafickém adaptéru, nebo součástí.

Paměť BIOSu

- Paměť s inicializačním systémovým kódem (firmware), který se spustí při startu adaptéru,
- je schopen adaptér nastavit,
- přečíst HW konfiguraci,
- nabízí základní obslužné funkce *ovladačům*, které jsou již řádným systémovým softwarem daného OS a počítače.

Dedikované a integrované GPU

- Dnešní "grafika" existuje ve dvou podobách: jako **samostatný (dedikovaný) adaptér** nebo jako **integrované GPU** (iGPU) přímo v procesoru či SoC.

Dedikované (samostatné) GPU

- Vlastní karta na sběrnici (dnes PCI Express), s vlastní *videopamětí* a vlastním chlazením.
- Nejvyšší výkon; spotřeba nejvýkonnějších modelů přesahuje 300 W.
- Hlavní výrobci čipů: **NVIDIA** (GeForce RTX, architektura Blackwell), **AMD** (Radeon RX, architektura RDNA), **Intel** (Arc, architektura Xe).

Integrované GPU (iGPU)

- GPU je součást procesoru (APU) nebo systému na čipu, sdílí s CPU operační paměť.
- Postačí pro kancelářskou práci, video i hry v nižších detailech, a to s minimální spotřebou.
- Např.: **AMD Radeon Graphics** (RDNA), **Intel Arc Graphics** (Xe) v procesorech, **GPU Apple Silicon**.
- Ke GPU se dnes v procesorech přidává i **NPU** (Neural Processing Unit) pro akceleraci umělé inteligence.

Unified memory

- U mobilních SoC (Apple Silicon) a herních konzolí sdílí CPU a GPU **jeden jednotný adresní prostor** — tzv. *unified memory*.
- Není nutné kopírovat data mezi pamětí CPU a GPU; paměť se využívá efektivněji a GPU má k dispozici celou kapacitu systému.
- Např. konzole PlayStation 5 a Xbox Series X|S používají APU AMD se sdílenou grafikou a pamětí GDDR6.

Paměti grafických adaptérů

- Videopaměti se dnes používají typy optimalizované pro velké datové toky:

GDDR6, GDDR6X	standard u herních a pracovních GPU (NVIDIA RTX, AMD Radeon RX)
GDDR7	nová generace, ještě rychlejší, poprvé u NVIDIA GeForce RTX řady 50
HBM (High Bandwidth Memory)	vertikálně skládané čipy s extrémní propustností, pro výpočetní akcelerátory a AI (např. NVIDIA H100/B200, AMD Instinct)
LPDDR / sdílená RAM	u integrovaných GPU a unified memory (Apple Silicon, konzole)

GPU dnes není jen grafika

- GPU je od počátku 21. století i univerzálním **paralelním výpočetním akcelerátorem** (GPGPU — General-Purpose computing on GPU).
- Programování masivně paralelních výpočtů: **CUDA** (NVIDIA), **ROCm** (AMD), **OpenCL**, **Vulkan** compute.
- Oblasti: simulace, vědecké výpočty, kryptoměny, ale hlavně **umělá inteligence** a strojové učení.
- Specializované jednotky:
  - **Tensor jádra** (NVIDIA) a **AI akcelerátory** (AMD, Intel) pro maticové operace neuronových sítí;
  - **RT jádra** pro výpočet *ray tracingu* v reálném čase (realistické osvětlení a odrazy ve hrách).

AI upscaling a generování snímků

- Moderní GPU *umějí i softwarově "vyrobit" obraz*: vykreslí scénu v nižším rozlišení a neuronová síť ji v reálném čase dopočítá na vyšší.
- **DLSS** (NVIDIA, dnes verze 4.x s Multi Frame Generation), **FSR** (AMD), **XeSS** (Intel).
- Výsledek: vyšší snímkovou frekvenci při zachování dobré kvality obrazu.
- GPU dnes rovněž zvládají hardwarové kódování videí (AV1, HEVC) — u NVIDIA např. jednotky NVENC.

Současný trh grafických adaptérů (2026)

- **NVIDIA**: GeForce RTX řady 50, architektura Blackwell, paměti GDDR7, ray tracing 4. generace, tensor jádra 5. generace, DLSS 4.
- **AMD**: Radeon RX řady 9000, architektura RDNA 4 se specializovanými AI akcelerátory po vzoru NVIDIA, FSR 4 založené na neuronových sítích.
- **Intel**: Arc řady B (Battlemage), zaměření na cenově dostupné modely (např. Arc B580) a na integrovanou grafiku Xe.
- Směr vývoje: stále výkonnější **integrované GPU a unified memory**, důraz na **AI výkony** (vyjadřované v TOPS) a **energetickou efektivitu**.

Na další kapitolu: Periferie počítačů

HTML · Slides · PDF | Přehled předmětu