

# Hardware pro počítačovou grafiku

Pavel Strachota

FJFI ČVUT v Praze

13. října 2023

# Obsah

- 1 Klasická zobrazovací zařízení
- 2 Moderní zobrazovací zařízení
- 3 Grafické adaptéry

# Obsah

- 1 Klasická zobrazovací zařízení
- 2 Moderní zobrazovací zařízení
- 3 Grafické adaptéry

# Obrazovky CRT 1/2

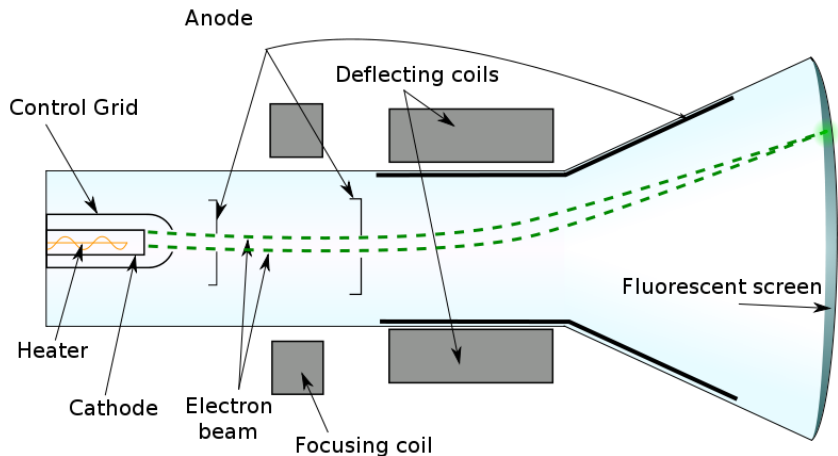
- **CRT** - Cathode Ray Tube

## Princip černobílé televize:

- 1 svazek elektronů z elektronového děla (katoda - vysoké napětí cca 15-20kV)
  - 2 je zaostřen (konvergenční mechanismus)
  - 3 svazek je odkloněn magnetickým polem (cívky) na správné místo na stínítku
  - 4 dopad elektronů na luminofor („fosfor“) na stínítku vyvolá emisi světla s exponenciálním útlumem
    - intenzita elektronového svazku (počet elektronů) ~ intenzita světla
- elektronový paprsek vykresluje obraz po řádcích
    - **(vertikální) obnovovací frekvence (*refresh rate*)** - počet snímků (=průchodů paprsku celým obrazem) za sekundu
    - **horizontální obnovovací frekvence** - počet vykreslených řádků za sekundu

# Obrazovky CRT 2/2

## Schéma černobílé televize

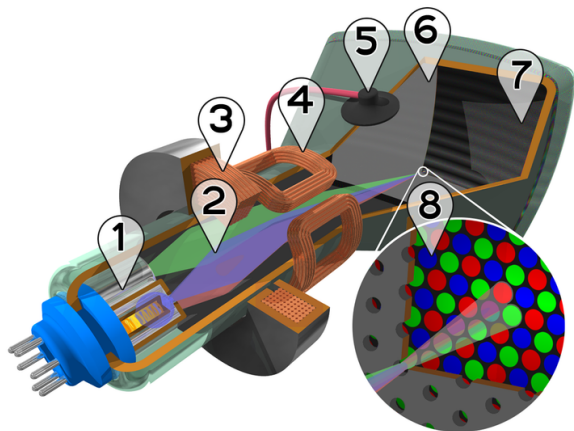


## Barevná CRT 1/2

- 3 druhy luminoforů uspořádané do mřížky:  
**R**ed, **G**reen, **B**lue
- 3 elektronová děla pro 3 barevné složky, prostorově různě umístěná
- struktura stínítka (maska) zaručuje, že elektrony z jednotlivých elektronových děl dopadnou jen na odpovídající luminofory (díky geometrické konfiguraci el. děl, samotné masky a rozmístění luminoforů)
- dojem barvy v rámci jednoho *pixelu* (elementu obrazu) vznikne složením příspěvků R,G,B jednotlivých *subpixelů*

# Barevná CRT 2/2

## Schéma barevné televize



- 1 elektronová děla
- 2 el. svazky
- 3 zaměřovací cívka
- 4 vychylovací cívky
- 5 napojení anody
- 6 stínítko (maska)
- 7 fosforová vrstva
- 8 detail fosforové vrstvy

## Typy stínítka 1/2

Stínítko - **stínící maska** (*shadow mask*) - kovový plát obsahující drobné dírkky

- trojúhelníkové uspořádání R,G,B fosforů
- trojúhelníková konfigurace elektronových děl
- zkreslení obrazu i barev kvůli zahřívání a tepelné roztažnosti masky („*doming*“)
  - invarová maska (*invar* - slitina Fe, Ni s extrémně nízkým koef. tepelné roztažnosti)

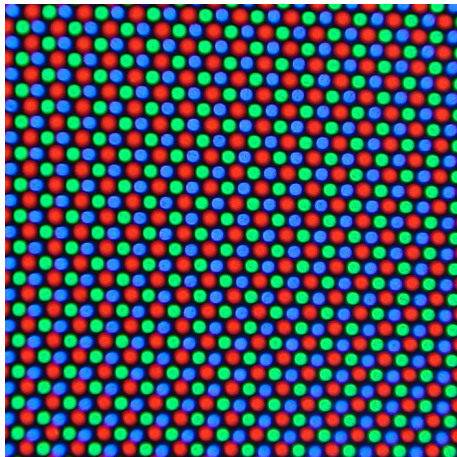




## Typy stínítka 1/2

Stínítka - **stínící maska** (*shadow mask*) - kovový plát obsahující drobné dírkky

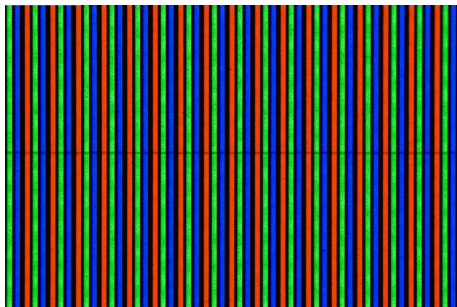
- trojúhelníkové uspořádání R,G,B fosforů
- trojúhelníková konfigurace elektronových děl
- zkreslení obrazu i barev kvůli zahřívání a tepelné roztažnosti masky („*doming*“)
  - invarová maska (*invar* - slitina Fe, Ni s extrémně nízkým koef. tepelné roztažnosti)



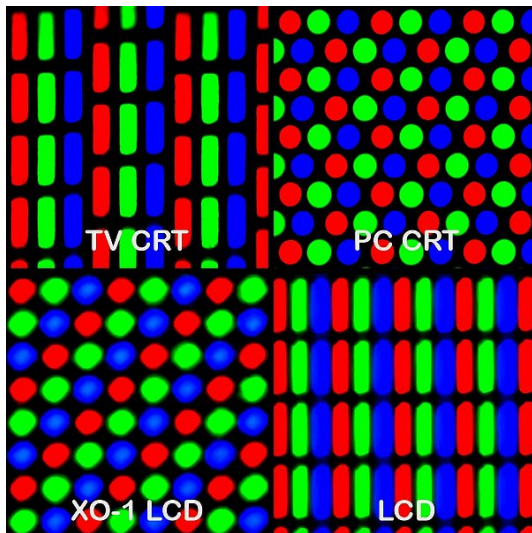
## Typy stínítka 2/2

**Štěrbínová maska** (*aperture grille*) - maska z tenkých vertikálních drátků

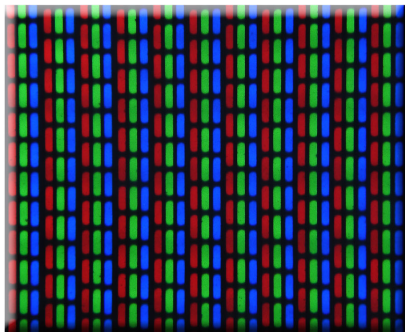
- **Sony Trinitron** cca od r. 1970
- vertikální uspořádání R,G,B fosforů
- elektronová děla vedle sebe
- menší mezery mezi pixely (hlavně vertikálně)
- plošší obrazovka
- zahřívání nezkrusuje obraz (roztažnost jen vertikálně)
- mechanická rezonance drátků  $\implies$  stabilizační horizontální drátky
- magnetizace masky  $\implies$  prohnutí drátků, zkreslení barev
- **degauss** = demagnetizace cívkou s oscilujícím mag. polem s útlumem



## Typické použití masek



# Rozlišení televize a CRT monitoru

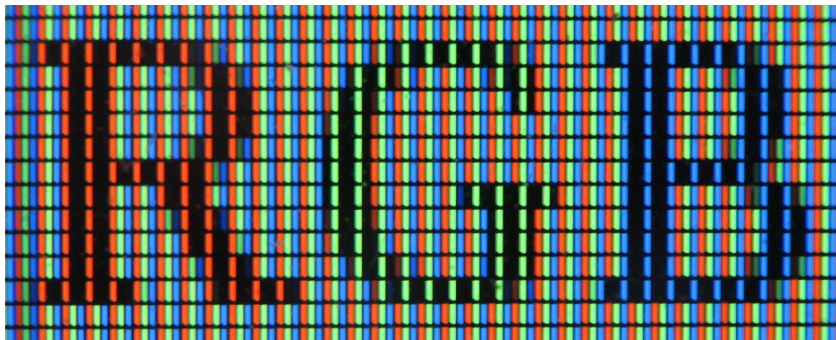


21" TV CRT Display



17" PC CRT Display

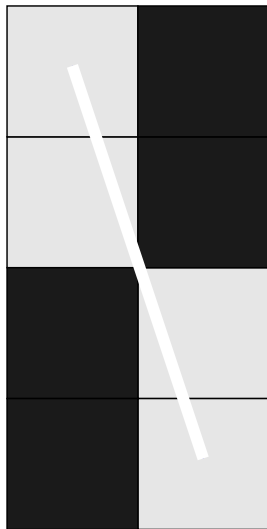
# Skládání barev na monitoru





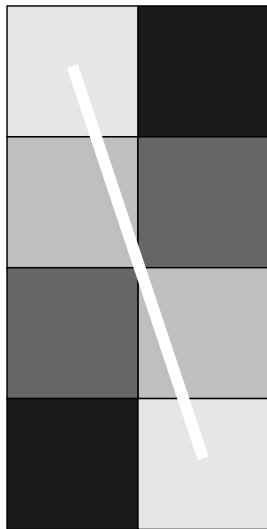
## Subpixel rendering

- využití znalosti o geometrickém uspořádání subpixelů pro zlepšení kvality obrazu (především **antialiasing**)
- V závislosti na barevných přechodech se vhodně upravují **barvy pixelů** na hranici přechodu, a tím i **intenzity subpixelů**.



## Subpixel rendering

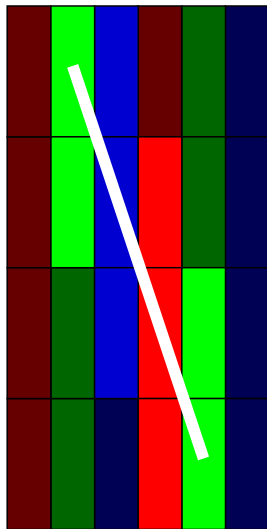
- využití znalosti o geometrickém uspořádání subpixelů pro zlepšení kvality obrazu (především **antialiasing**)
- V závislosti na barevných přechodech se vhodně upravují **barvy pixelů** na hranici přechodu, a tím i **intenzity subpixelů**.





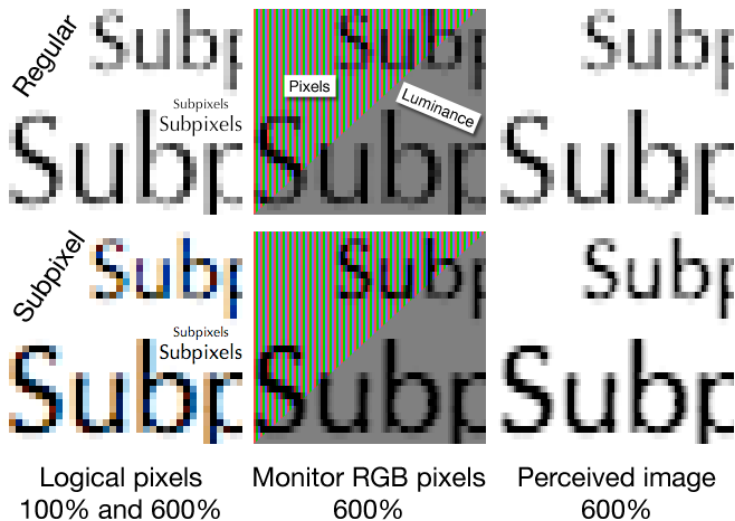
# Subpixel rendering

- využití znalosti o geometrickém uspořádání subpixelů pro zlepšení kvality obrazu (především **antialiasing**)
- V závislosti na barevných přechodech se vhodně upravují **barvy pixelů** na hranici přechodu, a tím i **intenzity subpixelů**.



# Subpixel rendering

Ukázka



## LCD obrazovky 1/7

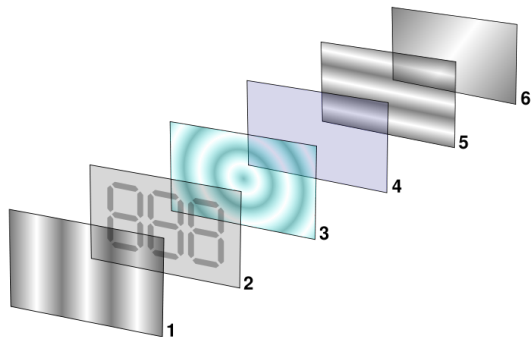
- **LCD** - *liquid crystal display* (displej z tekutých krystalů)

### Konstrukce klasického nepodsvíceného LCD

- TN (*Twisted Nematic*) - vrstva tekutých krystalů mezi dvěma polarizačními filtry s navzájem kolmou rovinou polarizace, pod nimi reflexní vrstva
- bez el. proudu mezi elektrodami jsou molekuly tekutých krystalů v šroubovici (šroubovice vychází z navzájem kolmých elektrod)
- elektrody z ITO = indium tin oxide (směs  $\text{In}_2\text{O}_3$  a  $\text{SnO}_2$ ) - průhledný vodič
- šroubovice stáčí rovinu polarizace světla, které prošlo prvním filtrem  $\implies$  světlo projde i kolmým filtrem, odrazí se od reflexní vrstvy a vrátí se zpět stejným způsobem
- po aplikaci el. proudu se šroubovice narovná, rovina polarizace světla se nestočí a přes kolmý pol. filtr. neprojde  $\implies$  černá

# LCD obrazovky 2/7

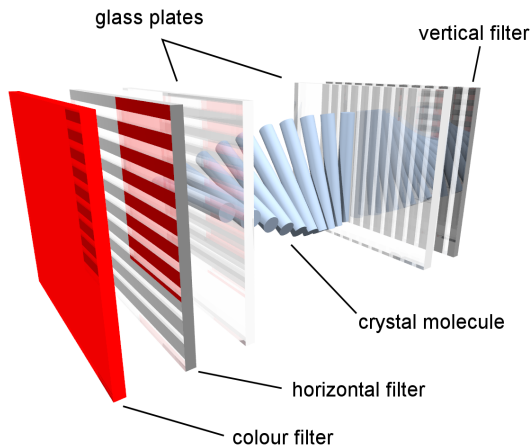
## Schéma TN LCD



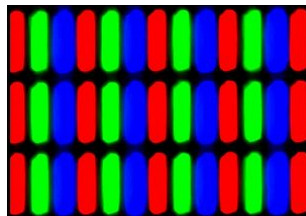
- 1 vertikální polar. filtr
- 2 sklo s ITO elektrodami ve tvaru segmentů displeje
- 3 vrstva tek. krystalů
- 4 sklo s ITO filmem (2. el. pól společný pro všechny segmenty)
- 5 horizont. polar. filtr
- 6 reflexní vrstva

# LCD obrazovky 3/7

## Konstrukce barevného LCD



- barevné filtry
- dělení na subpixely jako u CRT



# LCD obrazovky 4/7

## Typy barevných LCD

- pasivní **CSTN** (*Color Super Twisted Nematic*)
  - $m + n$  kontrolních vodičů pro rozlišení  $m \times n$ , pixely udržují setrvačností svůj stav do další obnovy
  - napětí na pixelu: [potenciál výběru řádku] – [potenciál zapnutí nebo vypnutí (video signál)] ve sloupcích
  - dostatečné napětí na změnu stavu pixelu je jen ve vybraném řádku
  - nízký kontrast, dlouhá odezva
- vylepšení: **DSTN** (*Dual Scan*)
  - displej rozdělen na poloviny, obě obnovovány současně
- v levných mobilech, levných notebookách z 90. let
- příliš pomalé pro sledování videa a hraní her - „duchy“

# LCD obrazovky 5/7

## Typy barevných LCD

- **TFT** (Thin Film Transistor) LCD
  - opět  $m + n$  kontrolních vodičů
  - tranzistor součástí každého pixelu
  - každý pixel funguje jako malý kondenzátor, tranzistor udržuje napětí na pixelu až do další obnovy (zamezí částečné aktivaci/deaktivaci pixelu vlivem nižšího přiváděného napětí ve chvíli, kdy se obnovují ostatní pixely)  $\implies$  **aktivní** displej
  - tranzistory vyrobeny nanášením polovodičové (křemíkové) vrstvy na skleněnou desku
- vlastnosti TFT
  - vyšší kontrast, rychlejší odezva (tekuté krystaly mají menší setrvačnost)
  - různé typy (**TN**, **S-IPS**, **S-PVA**, MVA, ASV, ...) - snaha zlepšit odezvu, barvy, pozorovací úhel
  - výskyt vadných (sub)pixelů (stuck = ON, dead = OFF)

# LCD obrazovky 6/7

## Barevná hloubka

- vstupní signál: **8** bitů na barevný kanál  $\implies$  16.7 miliónů barev
- řízení jasu subpixelů: většinou jen **6** bitů na subpixel
  - umožňuje rychlejší odezvu panelu
  - jen 262 tisíc barev
- simulace více barev
  - dithering: sousední pixely které mají mít stejnou barvu střídají skutečné odstíny  $\rightarrow$  viz přednáška [Barvy](#)
  - **FRC** (Frame Rate Control): pixel mění odstín s každým novým snímkem (= „časový dithering“)
  - Advanced FRC (AFRC)



# LCD obrazovky 7/7

Velmi vysoké rozlišení

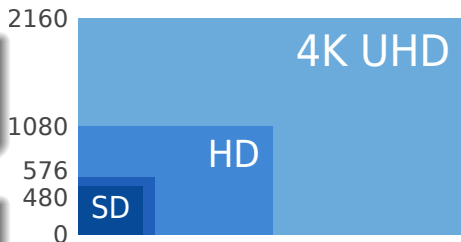
- *Ultra High Definition (UHD, 4K)* 3840 × 2160
- 8K 7680 × 4320
- u dražších tenké tranzistory z **IGZO** (Indium gallium zinc oxide)
- technologie TN nebo IPS
- barevná hloubka až 10 bitů

## Ceny za 4K v roce 2014

- TN monitory od  $\approx$  13 tis. Kč
- televize od  $\approx$  20 tis. Kč

## Ceny za UHD v r. 2020

- TN monitory od  $\approx$  6 tis. Kč
- 8K monitory  $\approx$  100 tis. Kč
- TV od 7 do 250 tis. Kč



# LCD obrazovky 7/7

Velmi vysoké rozlišení

- *Ultra High Definition (UHD, 4K)* 3840 × 2160
- 8K 7680 × 4320
- u dražších tenké tranzistory z **IGZO** (Indium gallium zinc oxide)
- technologie TN nebo IPS
- barevná hloubka až 10 bitů

## Ceny za 4K v roce 2014

- TN monitory od ≈ 13 tis. Kč
- televize od ≈ 20 tis. Kč

## Ceny za UHD v r. 2020

- TN monitory od ≈ 6 tis. Kč
- 8K monitory ≈ 100 tis. Kč
- TV od 7 do 250 tis. Kč

## Levné 4K monitory:

- jen TN oproti IPS Full HD za stejnou cenu
- **dříve max. 30Hz** (s HDMI 1.x)
- dnes již OK (HDMI 2.0, DisplayPort 1.3)

## Plazmové obrazovky 1/2

- směs vzácných plynů (**He**, **Ne**, **Xe**) uzavřené v malých buňkách
  - působením elektrického potenciálu se mění v plazmu (ionizují)
  - ionty vyzařují UV záření při dopadu na elektrody
  - UV záření excituje luminofory
- struktura R,G,B analogická jako u LCD, CRT
- tenký displej i ve velkých rozměrech (úhlopříčka až 150", 3.5m)
- vysoký kontrast, živé barvy, vysoký jas
- náchylnost k „vypálení“ (*burn-in*) statického obrazu do obrazovky, stejně jako CRT
- postupně ztrácí jas (50% jasu po 100 000 hodinách provozu)

# Plazmové obrazovky 2/2

## Burn-in efekt

Departures - All Airlines						
DESTINATION	AIRLINE	FLIGHT	TERMINAL	GATE	TIME	REMARKS
B. J. Cabo	American Airlines	1225	D	D31	9:50a	On Time
Sacramento	American Airlines	1103	C	C33	9:35a	Now 9:50a
Salt Lake	Delta Air Lines	1096	E	E13	8:30a	On Time
Salt Lake	American Airlines	689	C	C8	9:25a	Now 9:47a
San Angelo	American Airlines	3417	B	B24	8:55a	On Time
San Angelo	American Airlines	3485	B	B24	10:20a	On Time
San Antonio	American Airlines	1839	A	A34	8:15a	Now 8:40a
San Antonio	Alaska Airlines	1978	D	D17	9:25a	On Time
San Diego	American Airlines	1789	D	D16	8:40a	On Time
San Francisco	United Airlines	1247	B	B29	7:25a	On Time
San Francisco	Qantas Airways	3062	A	A23	7:55a	On Time
San Francisco	JAL	5865	C	C11	8:55a	On Time
San Francisco	Alaska Airlines	1801	A	A24	10:00a	On Time
San Francisco	American Airlines	9235	D	D20	10:00a	On Time
San Jose CA	TAM	8352	D	D29	7:55a	On Time
San Jose CA	American Airlines	1347	A	A20	10:00a	On Time
San Jose, CR	American Airlines	2163	D	D38	10:15a	On Time
San Juan	American Airlines	2058	D	D29	8:35a	On Time
Seattle	Delta Air Lines	9011	E	E10	7:00a	On Time
Seattle	American Airlines	1157	A	A10	7:10a	On Time

Saturday APR 14, 2007 6:41 AM

DFW

# Obsah

- 1 Klasická zobrazovací zařízení
- 2 Moderní zobrazovací zařízení
- 3 Grafické adaptéry

# LCD obrazovky s LED podsvícením 1/2

Obchodní název: LED TV

- klasické podsvícení: fluorescenční lampa se studenou katodou (CCFL - *Cold Cathode Fluorescent Lamp*)

## podsvícení bílými LED

- plné nebo jen okrajové podsvícení (nutné světlovody)
- lokální ztmavení (*local dimming*)  
⇒ nižší spotřeba, vyšší kontrast

[Spustit video](#)

## podsvícení maticí RGB LED (QLED)

- užší spektrum: RGB filtry propustí jen co mají
    - oproti klasickému bílému podsvícení, kde RGB filtrem LCD displeje částečně projdou i sousední frekvence
  - selektivní lokální ztmavení R,G,B
- ⇒ vyšší kontrast, širší rozsah barev (gamut)

→ viz přednáška [Barvy](#)

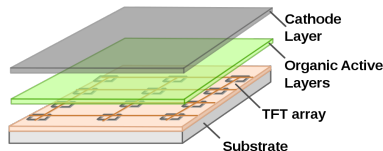
# LCD obrazovky s LED podsvícením 2/2

## Blikání

- řízení jasu monitoru - málokdy skutečným stmíváním diod
- **PWM** (Pulse-Width Modulation)
  - pulzy trvají odpovídající zlomek periody
  - levnější součástky
  - energeticky efektivnější
  - při nízkém jasu lze pozorovat *strobe-efekt*

# AMOLED displeje 1/3

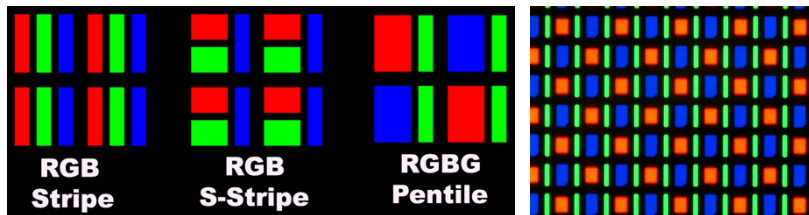
- **OLED** = Organic LED
  - organická sloučenina mezi vrstvami elektrod, funguje jako LED
  - nepotřebuje podsvícení
  - pasivní OLED displeje - levné, ale max.  $128 \times 96$  pixelů
- **AMOLED** = Active Matrix OLED
  - tok proudu do jednotlivých pixelů řízen pomocí TFT s integrovanými kondenzátory





## AMOLED displeje 2/3

- **2012:** 1280 × 800 HD Super AMOLED Plus, RGB Stripe matrix (Galaxy Tab 7.7)
- **2015:** 2560 × 1440 Quad HD Super AMOLED, 5.1", **577dpi** PenTile Matrix RG BG (Galaxy S6 / S6 **Edge**)
- **2019:**
  - 2688 × 1242 OLED Super Retina XDR 6,5" , **458dpi** (Apple iPhone 11 Pro)
  - 2560 × 1440 QHD+ Dynamic AMOLED, 6.1", **550dpi** (Galaxy S10)



# AMOLED displeje 3/3

## Vlastnosti

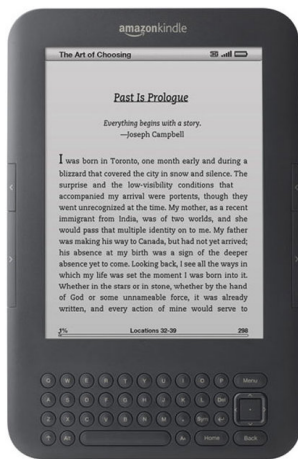
- Výhody (oproti LCD)
  - jednodušší  $\implies$  tenčí
  - nízká doba odezvy, vysoký kontrast, dobré pozorovací úhly
  - širší rozsah barev (gamut)
  - spotřeba závislá na skladbě obrazu, obvykle nižší než LCD (ale až 3× vyšší pro zcela bílou plochu)
- Nevýhody
  - omezená životnost OLED, navíc každé barvy jinak  
modrá: 50% intenzity jen za 14 000 h (5 let  $\times$  8h denně)
  - náchylnost k *burn-in* efektu (a.k.a. *image sticking*)

## Rady pro tvorbu GUI

- černé pozadí (spotřeba)
- v různých režimech nedávat stejné ikony (např. stav baterie) na stejná místa (burn in)
- pro světlý text nepoužívat zcela bílou, stačí 80% (burn in)

## Elektronický inkoust 1/2

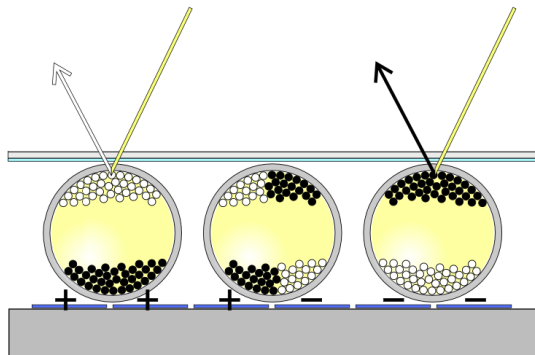
- bistabilní displej zachová obraz po odpojení zdroje energie
- čitelný bez podsvětlení
- **výhoda:** mizivá spotřeba, nepotřebuje stálý zdroj
- **nevýhody**
  - pomalé překreslování
  - někdy potřeba více cyklů (prokliknutí), aby zcela zmizel předchozí obraz
- **použití**
  - čtečky E-knih
  - displeje stavu zařízení (zaplnění USB flash disku)
  - displeje hodinek, low-cost telefonů (Motorola F3)
  - cenové etikety v obchodech



# Elektronický inkoust 2/2

## Elektroforézní displej

- mikrokapsle s nabitými částicemi barviva v kapalině
- v el. poli se částičky pohybují (elektroforéza)
- lze vyšší rozlišení elektrod než kapslí



## Dotykové obrazovky 1/2

- snímají pozici kontaktu prstu či pera (*stylus*) s obrazovkou  
⇒ ovládání zejména přenosných zařízení
  - smartphone, tablet PC, navigace, přehrávač médií, ...
- lze kombinovat s různými typy displejů
- **rezistivní** dotykové ovládání
  - 2 vrstvy vodičů oddělené nevodivou vrstvou
  - kontakt v místě stisku
  - lze ovládat čímkoliv (prstem, tužkou atd.)
  - (alespoň donedávna) levnější
- **kapacitní** dotykové ovládání
  - izolant (sklo) pokryté průhledným vodičem (např. ITO), resp. sítí vodičů
  - přiblížením prstu (funguje jako *virtuální uzemnění*) se změní *kapacitance* displeje, kterou lze (nepřímo) měřit
  - pohodlnější a přesnější ovládání (není potřeba tlačit)
  - nefunguje přes rukavice, nutný speciální vodivý stylus

## Dotykové obrazovky 2/2

- **infračervené** dotykové ovládání
  - síť infračervených diod a senzorů na okraji  
⇒ displej trochu zapuštěn
  - lze ovládat čímkoliv (perem, prstem, prstem v rukavici)
  - např. displeje samoobslužných pokladen, Amazon Kindle Touch
- další: optické snímání kamerami, rozpoznávání akustických pulzů, ...
- dotykové ovládání + **čidlo blízkosti** (*proximity sensor*)
  - displej mobilu zhasne, když je přiložen k uchu
  - *infotainment systémy* v automobilech: ovládací prvky se na displeji objeví jen při přiblížení ruky (VW Golf VII, Škoda Octavia III)

## 3D zobrazovací metody 1/3

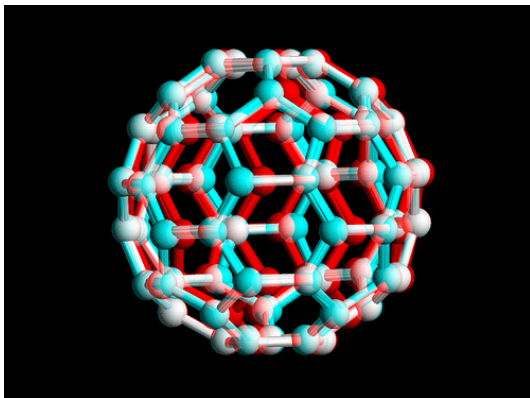
- **3D kinoprojekce** - dva (různě) lineárně polarizované obrazy promítané najednou + polarizační brýle
- **3D televize - „moderní“, ale dnes mrtvá technologie**
  - **aktivní** - střídavé promítání snímků pro pravé a levé oko + aktivní brýle synchronizované zatemňující levé a pravé oko
    - efektivně 1/2 obnovovací frekvence
    - brýle **drahé**, v ceně TV např. jen 1 brýle, někdy ani to ne
  - **pasivní** (komerčně v r. 2012) - sousední řádky obrazu různě polarizovány + polarizační brýle
    - efektivně 1/2 vert. rozlišení (Full HD 540p místo 1080p)
    - brýle **levné**  $\implies$  šlo snadno dokoupit
  - každý televizor (resp. značka) měl vlastní typ brýlí
  - 3D filmy se nedělají - **3D efekty nenahradí kvalitní scénář... :-)**
- **3D hry** - podpora monitorů i grafických akceleračních karet

# 3D zobrazovací metody 2/3

## Jednoduché 3D i na starém monitoru

### • Anaglyf

- obrazy pro pravé a levé oko přes sebe, v jiných barvách (červená, modrá)
- jednoduché barevné brýle



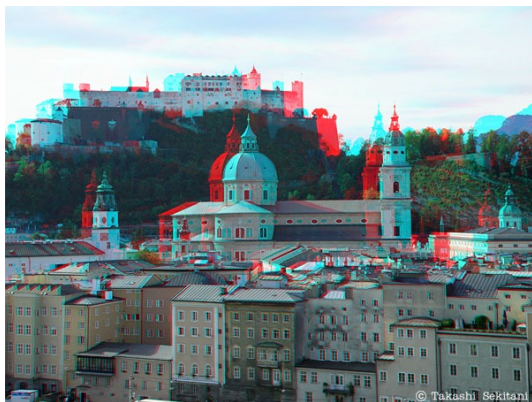


# 3D zobrazovací metody 2/3

## Jednoduché 3D i na starém monitoru

### ● Anaglyf

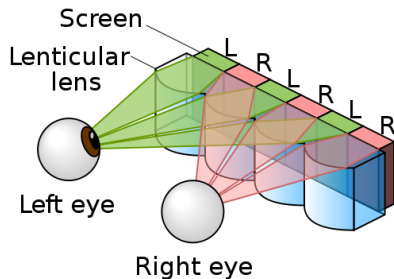
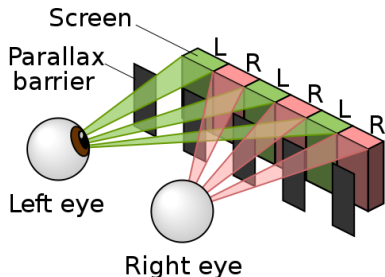
- obrazy pro pravé a levé oko přes sebe, v jiných barvách (červená, modrá)
- jednoduché barevné brýle



## 3D zobrazovací metody 3/3

3D i bez brýlí (autostereoskopie, glasses-less, glasses-free)

- nutno zajistit, aby levé a pravé oko vidělo každé svůj obraz
- bez brýlí lze, pokud je pozice diváka vůči displeji pevně daná (*Nintendo 3DS* držené v ruce)
- **2010:** Toshiba - první pokus: jen 1 divák, uprostřed a blízko
- **2013:** *Ultra-D* od Stream TV Networks ... prý funguje, detaily neznámé



# Obsah

- 1 Klasická zobrazovací zařízení
- 2 Moderní zobrazovací zařízení
- 3 Grafické adaptéry**

## Funkce grafického adaptéru

- generátor signálu pro monitor - **RAMDAC** (Random Access Memory Digital-to-Analog Converter)
- **framebuffer** - paměť snímku
  - RAM - co do ní zapíšeme, to se objeví na obrazovce
  - složitá logická struktura pro různé režimy (rozlišení) i na prastarých kartách EGA, VGA
- akcelerace 2D grafiky - funkce pro kreslení 2D primitiv (čára, obdélník, ...), vyplňování oblastí, scrollování, škálování fontů, interpolace, ...
- akcelerace 3D grafiky - přenos komplexních výpočtů pro zobrazení 3D scény z CPU na procesor grafické karty
- akcelerace dekódování videa
- obecné mnohovláknové výpočty - DirectCompute (DirectX 11), CUDA, OpenGL
- výpočty fyzikálního modelu - NVIDIA PhysX (dříve Physics Processing Unit, dnes CUDA)

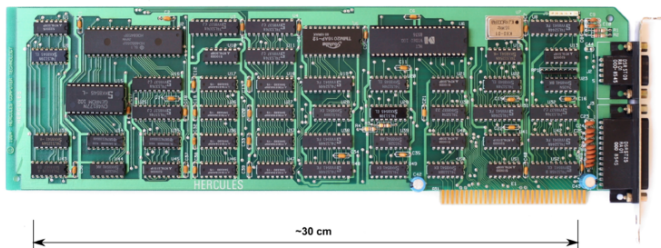
## Z historie grafických karet 1/4

- **CGA** (Color Graphics Adapter) - 1981
- 640×200 pixelů, 4 bity na pixel v paletě 16 barev, ale jen 4 barvy současně
- 80×25 znaků v textovém režimu (16 barev současně, font 8×8)



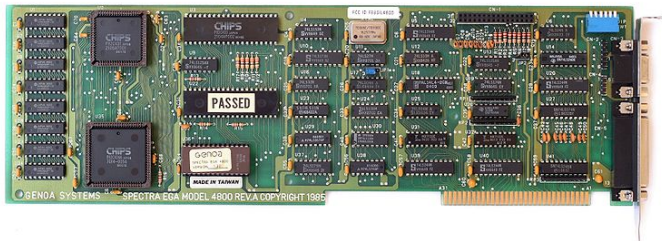
## Z historie grafických karet 2/4

- Hercules Graphics Card - 1984
- $720 \times 348$  pixelů, 1 bit na pixel
- $80 \times 25$  znaků v textovém režimu (bílá, šedá, černá, font  $9 \times 14$ )



## Z historie grafických karet 3/4

- **EGA** (Enhanced Graphics Adapter) - 1984
- 64 KB video RAM, rozšiřitelné na 256kB
- 640 × 350 pixelů, 4 bity na pixel, 16 barev z palety 64 barev



## Z historie grafických karet 4/4

- **VGA** (Video Graphics Array) - 1987 (IBM PS/2)
- 256KB video RAM
- $640 \times 480 \times 16$  barev,  $320 \times 200 \times 256$  barev



Trident TVGA9000  
SVGA adapter,  
 $2 \times 256\text{KB}$  VRAM



## Datové sběrnice pro grafické adaptéry

- **ISA** (Industry Standard Architecture) - 1981, 8/16 bit, 8MHz, 5.3MB/s
- **VESA local bus** - navíc k ISA, 1992, 32bit, 33MHz, 133MB/s. DMA kanály. Závislá na datové sběrnici i486. První 2D akcelerátory (DirectDraw).
- **PCI 1.0** (Peripheral Component Interconnect) - 1993, 32bit, 33MHz, 133MB/s.
- **AGP** (Accelerated Graphics Port) - 1997, 266MB/s (AGP 1x) - 4266MB/s (AGP 8x)
- **PCIe** (PCI Express) - 2004, sériová sběrnice, šířka 1-32 linek, až 8GT/s (v3.0, 2010)  $\implies \approx 16\text{GB/s}$  na PCIe x16

## Přenos signálu do monitoru 1/2

- **VGA** interface - analogový signál, oddělený signál pro R, G, B, horizontální a vertikální synchronizaci (retrace)
- **DVI** (Digital Visual Interface) - vhodný pro LCD (odpadne převod na analog. signál a zpět). Konektor umožňuje i přenos analogového signálu (DVI-A), redukci na VGA.



## Přenos signálu do monitoru 2/2

- **HDMI** (High Definition Multimedia Interface), 2002, nyní v. 2.1 (listopad 2017)
  - rozlišení až **8K** 7680 × 4320@60 Hz (24bit/px, 1.2GHz, 42.6 Gbit/s)
  - obsahuje audio a audio return kanál
  - 100MBit HDMI Ethernet Channel (HEC)
  - 3D over HDMI (několik formátů)
  - zejména pro spojení přehrávače a HDTV
- **DisplayPort** - 2006, nyní v. 2.0 (bude v 2020?)
  - univerzální: přes kabel i uvnitř notebooku
  - podpora USB-C konektoru
  - až 40 Gbit/s, 3×**10K** @ 60Hz
  - podpora signálu DVI a HDMI (1-kanálový)
  - výrobci neplatí licenční poplatky (*royalty-free*)



## Vývoj 3D grafických akceleratorů 1/2

- paralelní implementace operací pro zobrazování 3D vektorových dat, zpravidla v pohyblivé řádové čárce
- začátky: S3 Virge, ATI Rage - funkce pro 3D operace na čipu, ovšem čip byl kompatibilní s dřívějšími 2D akcelerátory  $\implies$  pomalá paměťová sběrnice atd.  $\implies$  slepá vývojová větev
- 3dfx Voodoo - speciální přídatná karta pro 3D operace, nutná standardní grafická karta
- další karty měly už 3D akceleraci integrovanou: NVIDIA RIVA TNT, Matrox Millennium G400, S3 Savage - sběrnice AGP
- funkce GPU na konci 90.let: rasterizace polygonů, 3D mapování textur, mip-mapping, trilineární filtrování, mlha, z-buffer (viz [POGR2](#))

## Vývoj 3D grafických akceleratorů 2/2

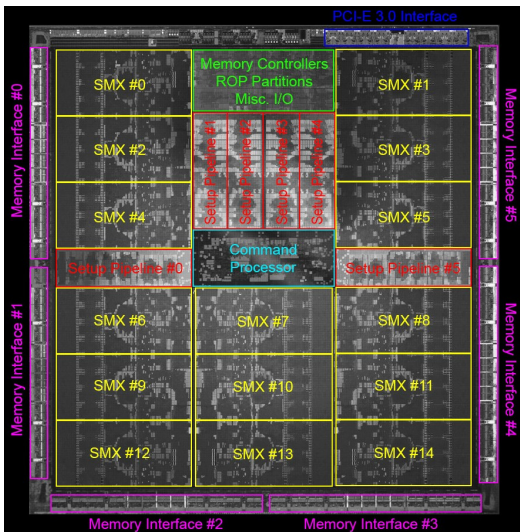
- rok 1999 - NVIDIA GeForce 256, ATI Radeon: poprvé hardwarové „Transform (,Clipping) & Lighting“ (**T&L**)
  - transformace 3D popisu scény do 2D (maticové transformace objektů - projekce, rotace, translace atd.)
  - *clipping* - oříznutí objektů pohledovým objemem, resp. hranicemi průmětny (obrazovky)
  - *lighting* - výpočet osvětlovacího modelu  $\implies$  nasvětlení scény
- rok 2001 - NVIDIA GeForce 3 - první GPU obsahující **shadery** (pixel shader, vertex shader, geometry shader)
  - velmi univerzální programovatelná jádra (výpočetní jednotky), schopná provádět komplexní výpočty nad pixely, resp. vrcholy polygonální sítě před jejich zobrazením (komplexní osvětlovací model, bump mapping atp.)
  - kanály (pipelines) pro paralelní zpracování skupiny pixelů

# Architektura moderního GPU

- **GPU** - Graphical Processing Unit
- na rozdíl od CPU, kde většinu tranzistorů zabírá cache, je v GPU většina tranzistorů věnována výpočetním jednotkám (jádrům)
- přístup do paměti je extrémně rychlý, cache není třeba
- jednotlivá jádra jsou velmi obecně programovatelné procesory (shadery) zpracovávající skupinu vláken
- efektivní paralelní výpočet a přístup do paměti vyžaduje striktní omezení
  - jaké operace mohou současně provádět různá jádra,
  - kam do paměti mohou přistupovat jednotlivá jádra atd..
- **SLI** (Scalable Link Interface) - propojení dvou grafických karet pro paralelní zpracování

# Schéma GPU

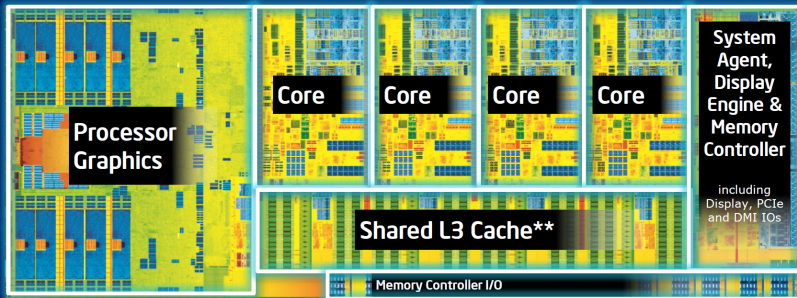
Čip NVIDIA Kepler GK110 vs. Intel Core-i7 (Haswell)



# Schéma GPU

Čip NVIDIA Kepler GK110 vs. Intel Core-i7 (Haswell)

## 4th Generation Intel® Core™ Processor Die Map *22nm Tri-Gate 3-D Transistors*



Quad core die shown above

Transistor count: 1.4 Billion

Die size: 177mm<sup>2</sup>

\*\* Cache is shared across all 4 cores and processor graphics



# Programování GPU

- dvě standardní API (rozhraní pro programování aplikací):
  - **OpenGL** - vznik již 1992, obsahuje kompletní sadu funkcí pro manipulace s 3D daty a jejich rendering
    - multiplatformní: Windows, Unix (Linux, FreeBSD, Solaris, Mac OS X atd.)
  - **DirectX** - sada API od Microsoftu (od Windows 95), její část Direct3D je alternativou k OpenGL (na Windows převažuje)
- tato API využívají dostupné funkce grafických akceleratorů, zbytek funkcí řeší softwarové knihovny

aplikace	
API: <b>OpenGL</b>	<b>DirectX</b>
ovladač	
hardware	

# Výpočty na GPU

- **GPGPU** - General Purpose Computation on Graphics Processing Units
  - NVIDIA: **CUDA** - (Compute Unified Device Architecture) - architektura umožňující programovat GPU v jazyce velmi blízkém jazyku C
  - AMD: **ATI Stream** - GPGPU na čipech ATI (AMD)
- **OpenCL** - (Open Computing Language) - platforma pro programování heterogenních platform složených z CPU i GPU, podporován čipy NVIDIA i AMD (překlad závislý na HW platformě)
- **DirectCompute** - podpora GPGPU v DirectX 11
- **NVIDIA Tesla** - karta obsahující GPU určená výhradně pro paralelní počítání - i když (už zase) má video výstup
- začátek masivnější podpory počítání v dvojnásobné přesnosti, dosud pouze **float**, resp. i **double**, ale mnohem pomaleji

## Moderní grafické akcelerátory

### 2009: NVIDIA GeForce GTX 295 SLI

- 480 jader (2x 240 ve dvou GPU), 1792MB GDDR3 RAM, 896-bit sběrnice, 223.8 GB/s, CUDA, PhysX, SLI, DirectX 10.



## Moderní grafické akcelerátory

### 2010: NVIDIA GeForce GTX 480

- 480 jader (architektura Fermi GF100), 1536MB GDDR5 RAM, 384-bit sběrnice, 177.4 GB/s, CUDA, PhysX, DirectX 11.



## Moderní grafické akcelerátory

### 2010: ATI Radeon HD 5970

- 2x GPU RV870, 1600 stream procesorů, 2048MB GDDR5 RAM, 512-bit (2x 256) sběrnice, 128 GB/s, ATI Stream, DirectX 11.



## Moderní grafické akcelerátory

### 2011: NVIDIA GeForce GTX 590

- 2x 512 jader (architektura Fermi GF110), 3072 MB GDDR5 RAM, 768-bit (2x 384) sběrnice, 2x163.9 GB/s, 2.5 TFLOPs, 365 W, CUDA, OpenCL 1.1, PhysX, DirectX 11, OpenGL 4.1.



## Moderní grafické akcelerátory

### 2011: AMD Radeon HD 6990

- 2x GPU RV870, 3072 stream procesorů, 4096MB GDDR5 RAM, 512-bit (2x 256) sběrnice, 2x160 GB/s, 5 TFLOPs, 375 W, OpenCL 1.1, DirectX 11, OpenGL 4.1.



## Moderní grafické akcelerátory

### 2012: NVIDIA GeForce GTX 690

- 3072 (2x 1536) CUDA jader (architektura Kepler GK104), 3072 MB GDDR5 RAM, 512-bit (2x 256) sběrnice, 384 GB/s, 4.6 TFLOPs, 300 W, PhysX, DirectX 11, OpenGL 4.2.





## Moderní grafické akcelerátory

### 2012: AMD Radeon HD 7970

- 28mm Tahiti GPU, 2048 stream procesorů, 3072MB GDDR5 RAM, 384-bit sběrnice, 264 GB/s, 3.79 TFLOPs, 250 W, OpenCL 1.2, DirectX 11, OpenGL 4.2.



## Moderní grafické akcelerátory

### 2013: NVIDIA GeForce GTX Titan

- 2688 CUDA jader (architektura Kepler GK110), 6144 MB GDDR5 RAM, 384-bit sběrnice, 288.4 GB/s, 4.5 TFLOPs, 250 W, 3D Vision, PhysX, DirectX 11, OpenGL 4.3.



## Moderní grafické akcelerátory

### 2013: AMD Radeon HD 7990

- 4096 stream procesorů, 6144MB GDDR5 RAM, 384-bit sběrnice, 288 GB/s, 8.2 TFLOPs single precision, OpenCL 1.2, DirectX 11.1, OpenGL 4.2.



## Moderní grafické akcelerátory

### 2014: AMD Radeon R9 295 X2 Hydra Edition

- 5632 stream procesorů (architektura Hawaii XT x2), 8192MB GDDR5 RAM, 2×512-bit sběrnice, 2×320 GB/s, 11.5 TFLOPs single precision, OpenCL 1.2, DirectX 11.2, OpenGL 4.3.



# Moderní grafické akcelerátory

## 2014: NVIDIA GeForce GTX 980

- 2688 CUDA jader (architektura Maxwell), 4096 MB GDDR5 RAM, 256-bit sběrnice, 224.4 GB/s, 4.6 TFLOPs, **165 W !!**, 3D Vision, PhysX, DirectX 12, OpenGL 4.4, GPU Boost 2.0



## Moderní grafické akcelerátory

### 2015: AMD Radeon R9 Fury X

- 4096 stream procesorů (architektura Fiji), 4096MB GDDR5 RAM, 4096-bit sběrnice, 512 GB/s, 8.6 TFLOPs single precision, OpenCL 2.1, DirectX 12, OpenGL 4.5.



## Moderní grafické akcelerátory

### 2015: NVIDIA GeForce GTX 980 Ti

- 2816 CUDA jader (GM200-310 GPU), 6144 MB GDDR5 RAM, 384-bit sběrnice, 336.5 GB/s, 5.6 TFLOPs, 250 W, 3D Vision, PhysX, DirectX 12, OpenGL 4.5, GPU Boost 2.0



## Moderní grafické akcelerátory

### 2016: AMD Radeon Pro Duo

- 8192 stream procesorů (128 výpočetních jednotek), 8192 MB GDDR5 RAM, 2×4096-bit sběrnice, 1024 GB/s, 16.38 TFLOPs single precision, OpenCL 2.0, DirectX 12, OpenGL 4.5., **350 W**.





## Moderní grafické akcelerátory

### 2016: NVIDIA Titan X (2016 edition)

- 3584 CUDA jader (Pascal GP 102, 16nm) 1,5 GHz, 12 GB GDDR5X RAM, 384-bit sběrnice, 480 GB/s, 11 TFLOPs, **250 W**, + všechny známé technologie



## Moderní grafické akcelerátory

### 2018: AMD Radeon RX Vega

- 4096 stream procesorů (64 výpočetních jednotek), 8192 MB HBM2 RAM, 2048-bit sběrnice, 483.8 GB/s, 12.5 TFLOPs single precision, OpenCL 2.0, DirectX 12, OpenGL 4.6., **295 W**.



## Moderní grafické akcelerátory

### 2018: NVIDIA GeForce RTX 2080 Ti

- 4352 CUDA jader (Turing TU 102, 12nm) 1,5 GHz, 11 GB GDDR6 RAM, 352-bit sběrnice, 616 GB/s, 13.4 TFLOPs, **250 W**, + všechny známé technologie
- **Real Time Ray Tracing (RT cores)**
- *Deep Learning Super Sampling (DLSS)* ... upscaling z nižšího na plné rozlišení



## Moderní grafické akcelerátory

### 2020: NVIDIA GeForce RTX 3090

- 10496 CUDA jader (Ampere GA102, 8nm) 1,7 GHz, RayTracing a Tensor jádra, 24 GB GDDR6X RAM, 384-bit sběrnice, 936 GB/s, 35.6 TFLOPs, **350 W**, ..., **DLSS 2.0**, PCIe 4.0x16



## Moderní grafické akcelerátory

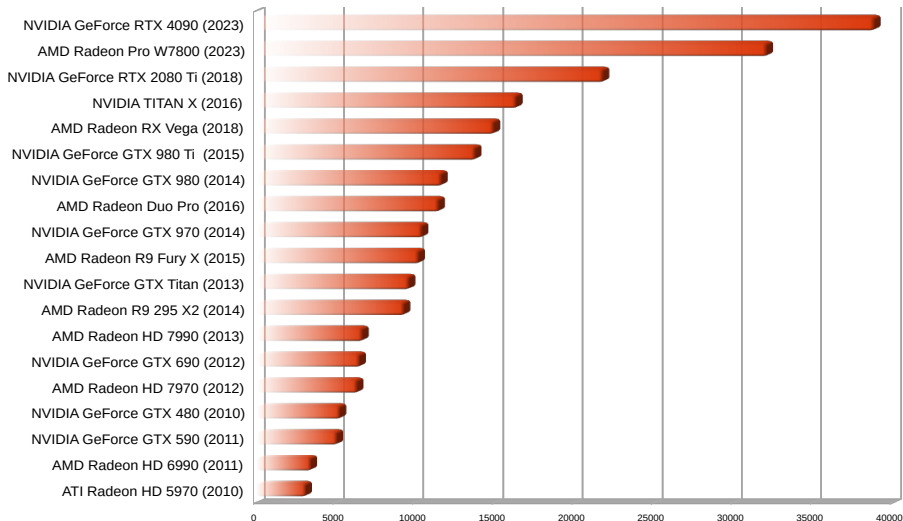
### 2021: AMD Radeon RX 6900 XT

- 5120 stream procesorů (80 jednotek Navi21, 7nm), 16GB GDDR6 RAM, 256-bit sběrnice, 512 GB/s, 20.5 TFLOPs single precision, 80 **Ray Accelerators** (RA's), **300 W**.



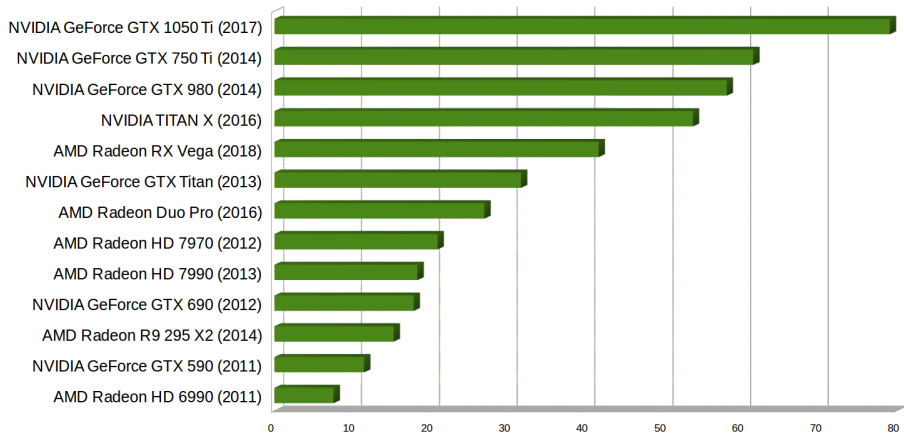
# Srovnání výkonu grafických akcelérátorů

PassMark G3DMark score 2023



# Srovnání efektivního využití elektrické energie

Power efficiency: PassMark G3DMark 2017 score per Watt



# Akcelerátory pro GPGPU

## 2011: NVIDIA TESLA C2070

- 448 jader, 6GB GDDR5, 384-bit sběrnice, 144GB/s, 238W, 515 GFLOPs v double precision, CUDA.





## Akcelerátory pro GPGPU

### 2012: NVIDIA TESLA K10

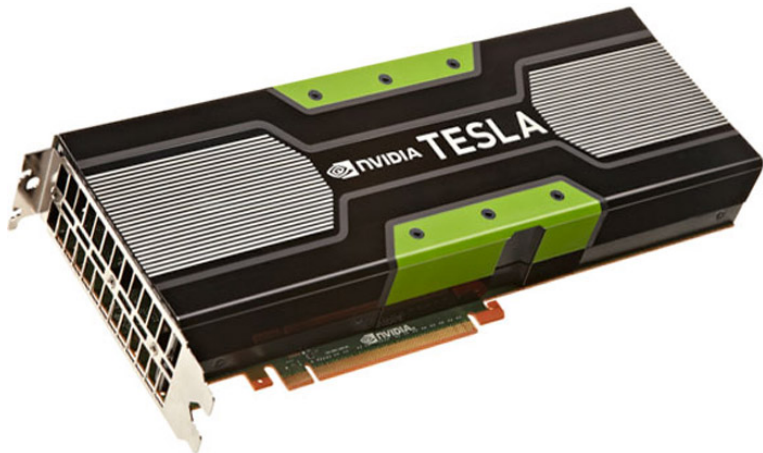
- 3072 (2x 1536) CUDA jader (architektura Kepler GK104), 8GB GDDR5, 4.6 TFLOPs v single precision, jen 190 GFLOPs v double precision.



## Akcelerátory pro GPGPU

### 2013: NVIDIA TESLA K20X

- 2688 CUDA jader (architektura Kepler GK110), 6GB GDDR5, 250 GB/s, 3.95 TFLOPs v single precision, 1.31 TFLOPs v double precision.



## Akcelerátory pro GPGPU

### 2014: NVIDIA TESLA K40

- 2880 CUDA jader (architektura Kepler GK110B), 12GB GDDR5, 280 GB/s, 4.29 TFLOPs v single precision, 1.43 TFLOPs v double precision.



## Akcelerátory pro GPGPU

### 2015: NVIDIA TESLA K80

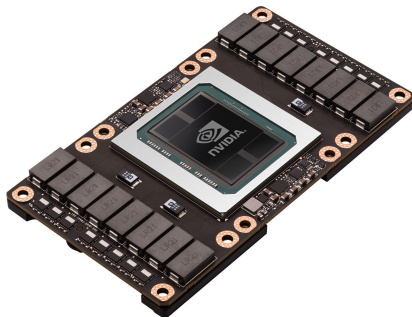
- 4992 CUDA jader (2x GPU Kepler GK210), 24GB GDDR5, 2x240 GB/s, 8.74 TFLOPs v single precision, 2.91 TFLOPs v double precision.



# Akcelerátory pro GPGPU a AI (deep learning)

## 2016: NVIDIA TESLA P100

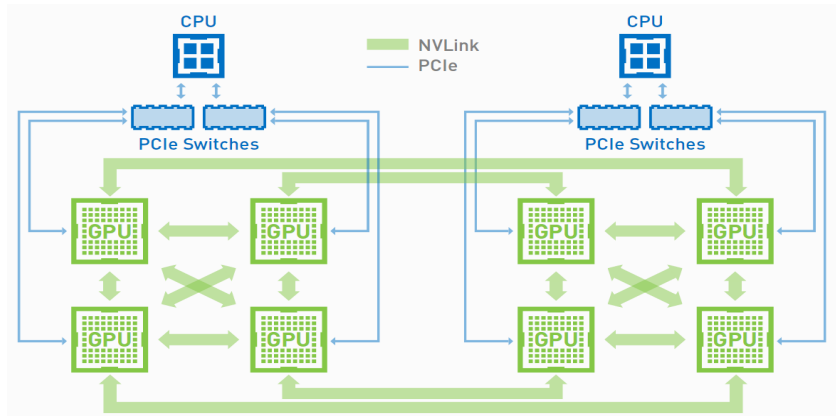
- 3840 CUDA jader (1x GPU Pascal),  
16GB HBM2, 720 GB/s, 10.6 TFLOPs single p., 5.3  
TFLOPs double p., 21.2 TFLOPs deep learning *half prec.*



# Akcelerátory pro GPGPU a AI (deep learning)

## 2016: NVIDIA TESLA P100

- 3840 CUDA jader (1x GPU Pascal),  
16GB HBM2, 720 GB/s, 10.6 TFLOPs single p., 5.3  
TFLOPs double p., 21.2 TFLOPs deep learning *half prec.*



# Akcelerátory pro GPGPU a AI (deep learning)

## 2017-2018: NVIDIA TESLA V100

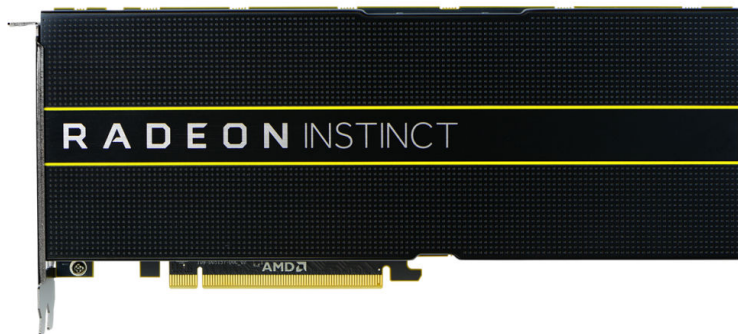
- 3840 CUDA cores, 640 *tensor cores* (1x GPU Volta), 32GB HBM2, 900 GB/s, 15.7 TFLOPs single p., 7.8 TFLOPs double p., **125 TFLOPs deep learning half prec.**, 300 W



## Akcelerátory pro GPGPU a AI (deep learning)

### AMD RADEON INSTINCT M25 (2018), MI50 (2020)

- 3840 stream procesorů (60 výpočetních jednotek) (1x GPU Vega, 7nm), 16GB HBM2, 1 TB/s, 13.3 TFLOPs single p., **6.6 TFLOPs double p.**, 26.5 TFLOPs deep learning *half prec.*, 300 W
- PCIe 4 x16 , AMD Infinity Fabric 184 GB/s

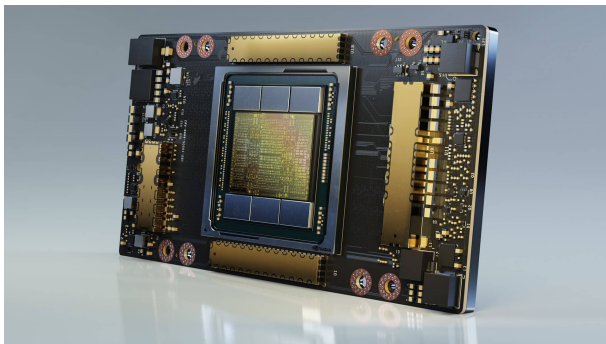




# Akcelerátory pro GPGPU a AI (deep learning)

## 2021: NVIDIA A100 Tensor Core GPU

- 1x GPU Ampere, 80GB HBM2, 1,6 TB/s, 19.5 TFLOPs single p., 9.7 TFLOPs double p., **312 TFLOPs deep learning *half prec.***, 400 W
- PCIe 4 x16 , NVLink v3 600 GB/s (SXM4)



# Akcelerátory pro GPGPU a AI (deep learning)

## 2021: AMD INSTINCT M100

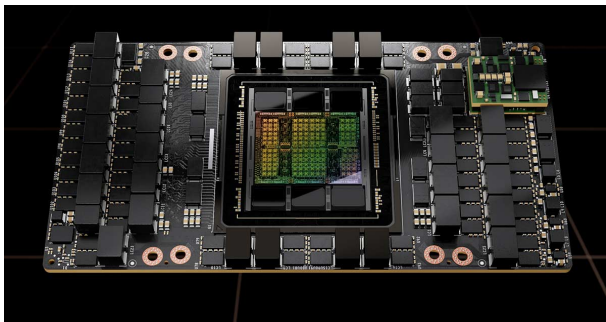
- 7680 stream procesorů (120 výpočetních jednotek)  
32GB HBM2, 1.2 TB/s,  
23.1 TFLOPs single p., **11.5 TFLOPs double p.**,  
184.5 TFLOPs deep learning *half prec.*, 300 W
- PCIe 4 x16 , AMD Infinity Fabric 276 GB/s



# Akcelerátory pro GPGPU a AI (deep learning)

## 2022: NVIDIA H100 a H100 CNX Accelerator

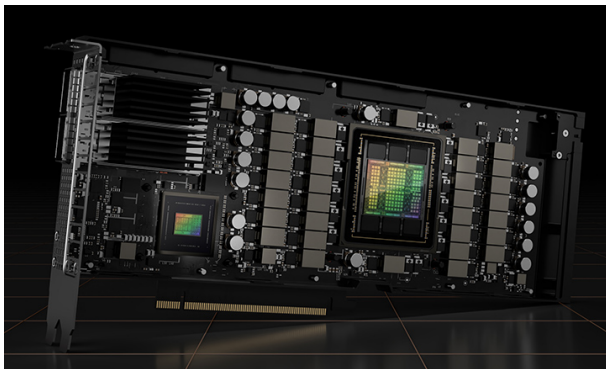
- 1x GPU Ampere, 80GB HBM2, 2,0 TB/s, 67 TFLOPs single p., 34 TFLOPs double p.,  
≈**1 PFLOP** deep learning *half prec.*, 400 W
- NVLink Switch (až 256 čipů)



# Akcelerátory pro GPGPU a AI (deep learning)

## 2022: NVIDIA H100 a H100 CNX Accelerator

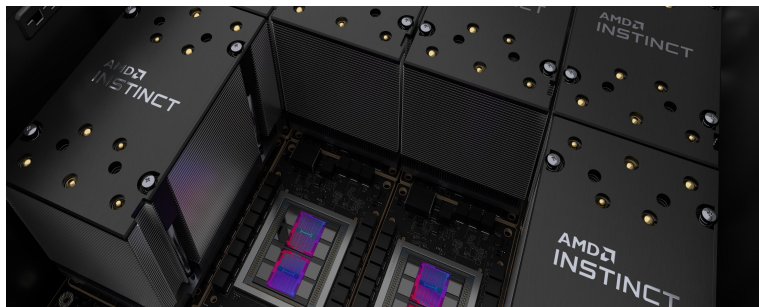
- **Converged Accelerator:** Unified Network & Compute
- 1x 400Gb/s, 2x 200Gb/s ports, Ethernet or InfiniBand



# Akcelerátory pro GPGPU a AI (deep learning)

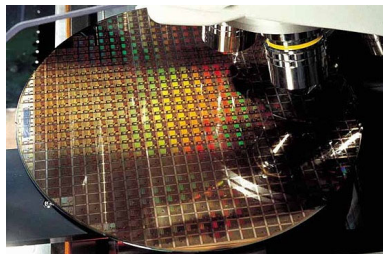
## 2022: AMD INSTINCT M250x

- 128 GB HBM2, 3.2 TB/s, 220 výpočetních jednotek
- 47.9 TFLOPs **single & double** precision,  
383 TFLOPs deep learning *half prec.*, 500 W
- AMD Infinity Fabric 800 GB/s (8 čipů na desce)
- použit v prvním exascale superpočítači (Frontier)
- AMD ROCm (analogie CUDA): prog. jazyk HIP/OpenCL



## Výroba GPU AMD (ATI) a NVIDIA

- vyvíjejí technologie, ale nemají továrny na masivní výrobu čipů (tzv. *fabless semiconductor company*)
- obě si čipy nechávají vyrábět u TSMC (Taiwan Semiconductor Manufacturing Company) - tzv. *semiconductor foundry*
  - **podzim 2010**
    - TSMC vyráběla 40nm technologií, výroba nebyla vždy bez problémů
    - výroba čipů NVIDIA GF100 (GTX 480) pouze 20% výnosnost (každý pátý se povedl), dříve jen cca 10%.
    - výroba ATI HD 5970 měla cca 60% výnosnost.



# Odkazy



<http://www.nvidia.com>



<http://www.amd.com>



<http://www.opengl.org>



<http://gpgpu.org/>



<http://www.ddworld.cz>



<http://www.wikipedia.org>



[http://www.samsung.com/uk/business/b2b/pdfs/case\\_studies/LED\\_BLU\\_White\\_Paper.pdf](http://www.samsung.com/uk/business/b2b/pdfs/case_studies/LED_BLU_White_Paper.pdf)



[www.4dsystems.com.au](http://www.4dsystems.com.au)