

# Grafické karty

Struktura, základní parametry,  
akcelerace, vytváření 3D scény, použití  
pro negrafické výpočty

- Grafická karta je komponenta počítače
- Stará se o grafický výstup na monitor, či jinou zobrazovací jednotku
- Využívá se jak ke grafickým, tak negrafickým výpočtům (využití shaderů k obecným výpočtům matematik, fyziky; neuronové sítě)

## Struktura

### Integrované

- Používá operační paměť počítače
- Levnější
- Méně výkonné
- Nejdříve integrované na základní desce (v severním můstku), později přímo v procesoru

### Dedikované

- Vlastní paměť
- Dražší
- Výkonnější

### GPU

- Grafický čip
- Buď na dedikované grafické kartě nebo v integrovan viz. Výše
- Zajišťuje výpočty a vykreslování dat, které jsou uloženy v operační paměti
- Největší výrobci jsou nVidia, AMD, Intel

### Paměť

- Uložení dat/informací potřebných pro činnost grafického čipu

### Výstupy

#### VGA

- (Video Graphics Array)
- Nejstarší
- Analogový přenos obrazu
- Velmi zastaralý a dnes již moc nepoužívaný, maximálně u monitoru, který ještě zpracovává signál analogově (CRT)
- U moderních monitorů GPU nejprve musí digitální signál převést na analog, který musí monitor poté znovu zdigitalizovat

#### DVI

- Digital Video Interface
- Digitální přenos obrazu
- 2 typy:
  - DVI-D
    - Čistě digitální
  - DVI-I
    - Jak digitální, tak analogový signál

- Zvláštnost je DVI-A
  - Čistě analogový

## HDMI

- High-Definition Multimedia Interface
- Digitální přenos zvuku i obrazu
- Nejčastější konektor dnes
- Společně s DP přenášejí rámce tzn. Mohou přenášet i ethernet

## DP

- Display Port
- Nejnovější
- Hodně podobný HDMI

## Sběrnice

- ISA, AGP, VL-Bus
  - Velmi zastaralé a dnes již nepoužívané
  - PCI, PCI-Express (x1, x4, x8, x16)

## Základní parametry

- Velikost paměti (GB)
- Typ paměti (rychlost)
  - GDDR4
  - GDDR5
  - GDDR6
  - Liší se v propustnosti (GB/s) a v efektivní frekvenci paměti (MHz)
- Šířka paměťové sběrnice
  - Paměťová sběrnice je spojení mezi GPU a pamětí grafické karty
  - Její šířka udávaná v bitech říká, kolik dat za jeden takt může mezi pamětí a GPU projít
  - Nejrozšířenější velikost je v současnosti 256bit
- Počet stream procesorů
  - Cca odpovídá hrubému výkonu grafické karty
- Verze grafických technologií
  - Verze DirectX (API, které obsahuje definici funkcí, které má karta hardwarově podporovat)
  - OpenGL
- Podpora CrossFire, SLI
  - Propojení dvou grafických karet
- Rozlišení v textovém režimu
  - Počet znaků, které je možné v textovém režimu zobrazit na jednom řádku, a počet řádků, které je možné umístit na obrazovku (typické hodnoty 40/80x25)
- Rozlišení v grafickém režimu (jemnost obrazu)
  - Počet pixelů, které je možné v horizontálním a vertikálním směru zobrazit
- Opakovací frekvence

- Kolikrát je grafická karta schopná poslat do displeje obraz

## Akcelerace

- Jedná se o řadu obvodů uvnitř GPU, které jsou konkrétně zaměřeny na určité výpočty
- Umožňují rychlejší vykreslování grafických objektů, a tudíž i rychlejší zpracování obrazu
- Příklady pokročilejších akcelerací:

## Geometry shader

- Pracují s celou kostrou objektu a jako jediné umí měnit počet polygonů a tvarovat objekt
- Můžete tedy v reálném čase pozorovat růst stromů a další činnosti
- Toto se dosud muselo řešit několika různými objekty, které se postupně objevovaly a mizely, což bylo značně náročné na výkon

## Vertex shader

- Vytváří z polygonů trojrozměrné objekty, umísťuje je do prostoru a stará se o světelné efekty

## Pixel shader

- U trojrozměrných objektů definuje barvu a průhlednost všech obrazových bodů

## Raster Operation Pipeline

- Sloučí obrazová data a optimalizuje je pro přenesení

## TMU jednotky (texture mapping unit)

- Mapuje textury na objekty

## Vytváření 3D scény

- Renderování obrazu je prováděno na základě grafických algoritmů
- 3D obraz je vytvářen na základě počítačového modelu, charakteristik scény a okolí
- Nejprve je zapotřebí vytvořit kostru modelu pomocí drátové konstrukce
- Kostra je pokryta povrchem, kterému je přiřazen materiál a vztah k dopadajícímu světlu
- Scénu charakterizuje umístění světel a kamery, okolí pak barvy, odraz, mlhu apod....

## Proces vytváření

- Nejprve musí aplikace popsat, jak má obraz vypadat
- Tento popis je předán přes programové rozhraní API ke zpracování HW
- Ve Windows se používají 2 rozhraní API: DirectX a Open GL
- Následuje matematický popis zobrazovacích těles, který se většinou provádí převedením povrchu těla na síť trojúhelníků
- Tělesa se natočí do polohy odpovídající pohledu pozorovatele a trojúhelníky, které jsou mimo dosah viditelnosti se odstraní
- Toto způsobí značné zmenšení potřebného výpočetního výkonu
- Poté následuje vykreslení povrchu těles
- Trojúhelníková síť je vymalována a barva se upravuje podle stínu
- Filtrování obrazu odstraní chvění obrazu při rychlých změnách

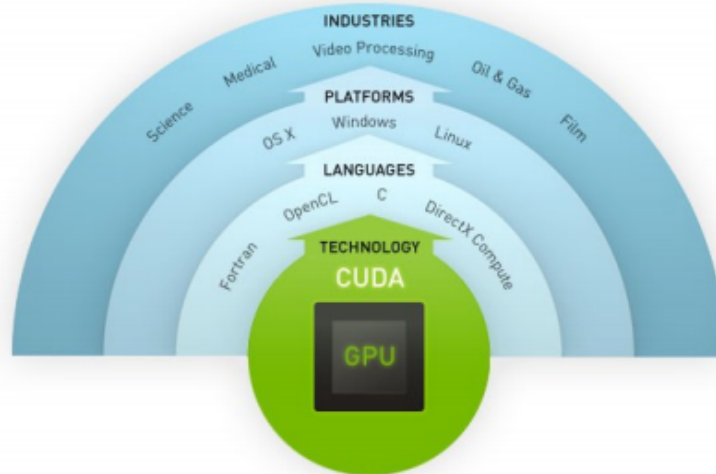
- Vyhlazení roztřepených hran odstraní rušivé přechody barev
- Nakonec se výsledný obraz pošle na zpracování pro zobrazovací zařízení
- Během zpracování obrazu pro monitor se začíná zpracovávat další snímek

## Použití pro negrafické výpočty

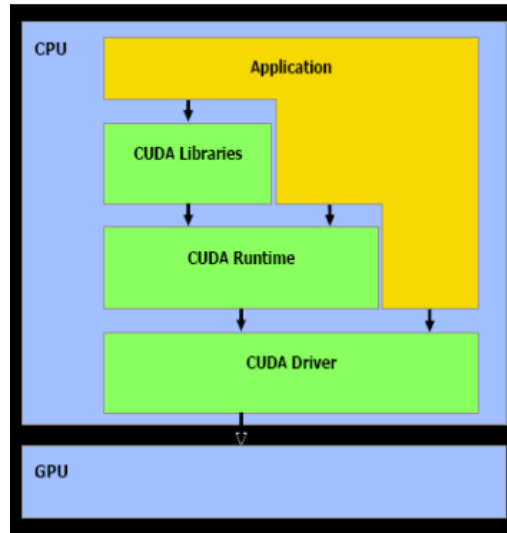
- GPU se na rozdíl CPU zaměřují ne na rychlost, ale na výkon (velké množství jednoduchých operací)
- GPU nepracují na tak vysoké frekvenci, ale díky velkému počtu jader a vyšší paměťové propustnosti zvládají až deseti násobek výkonu
- Nové CPU mají 4 jádra, přičemž GPU jich mají stovky
- GPU excelují při nasazení, které lze dobře paralelizovat – vědecké výpočty, časově náročné jednoduché operace (rendering), manipulace s obrázky

## GPGPU (General Purpose Processing on GPUs), CUDA

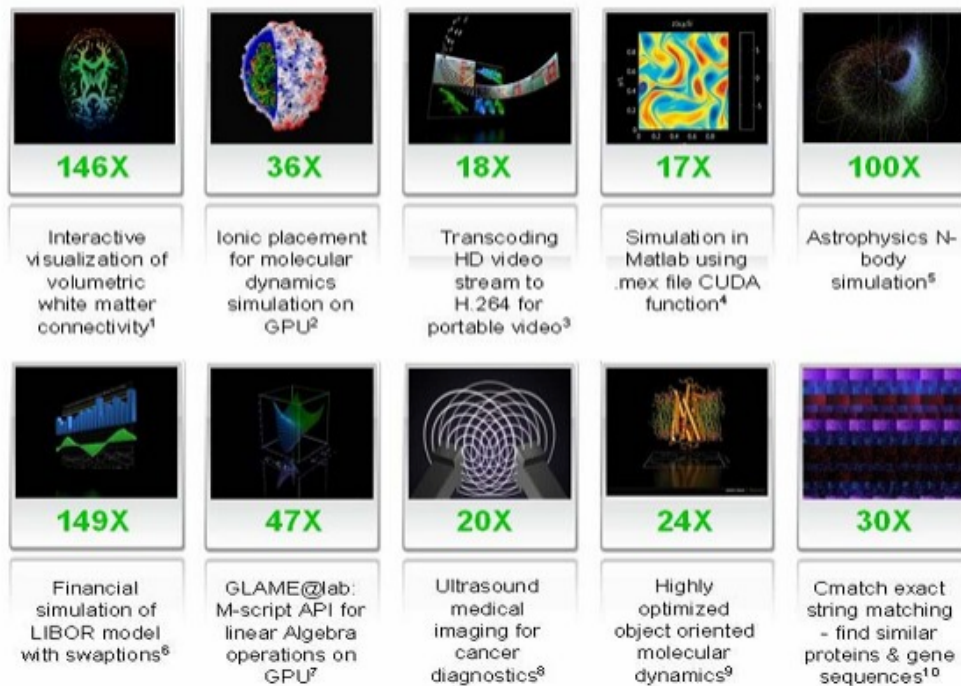
- NVIDIA vytvořila programovací prostředí CUDA (Compute Unified Device Architecture), které by mělo pomoci GPU výpočtům
  - Umožňuje programovat negrafické aplikace pro grafické procesory pomocí jazyka C
  - Nejstarší



- 
- CUDA aplikace může přistupovat k jakékoliv části API CUDA



- GPGPU lze účinně využívat hlavně v aplikacích, které lze velice dobře paralelizovat
- Také se hlavně využívá k časově náročným, ale jednoduchým operacím (rendering, zpracování velkých obrázků, práce s HD filmy)
- GPGPU lze využít například i v automobilovém průmyslu pro návrh automobilů za pomoci GPU raytracingu pod platformou CUDA
- Nebo také ve zdravotnictví například pro efektivnější nacházení prvotních stádií rakoviny



obr. 4 : Zrychlení algoritmů pomocí GPU

