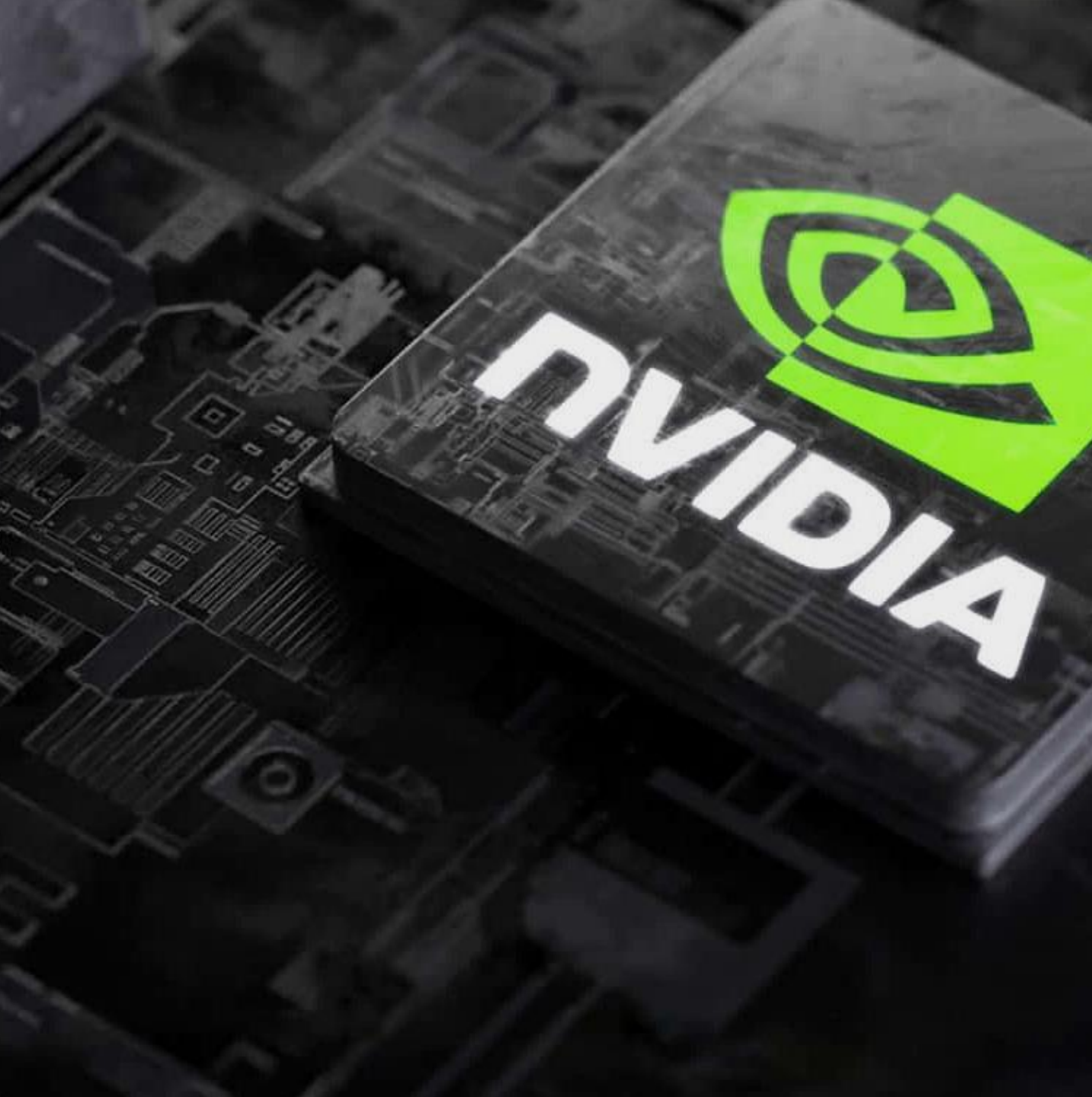


Zobrazenie informácie v počítači

GPU nVidia

Univerzita sv. Cyrila a Metoda
Samuel Csomor
II. ročník

A close-up, angled photograph of a black NVIDIA graphics card. The card is positioned diagonally across the frame. The NVIDIA logo, a green stylized eye, is prominently displayed in the upper right corner of the card's surface. Below the logo, the word "NVIDIA" is printed in large, white, sans-serif capital letters. The background shows the intricate circuitry of the printed circuit board (PCB) in shades of grey and black.

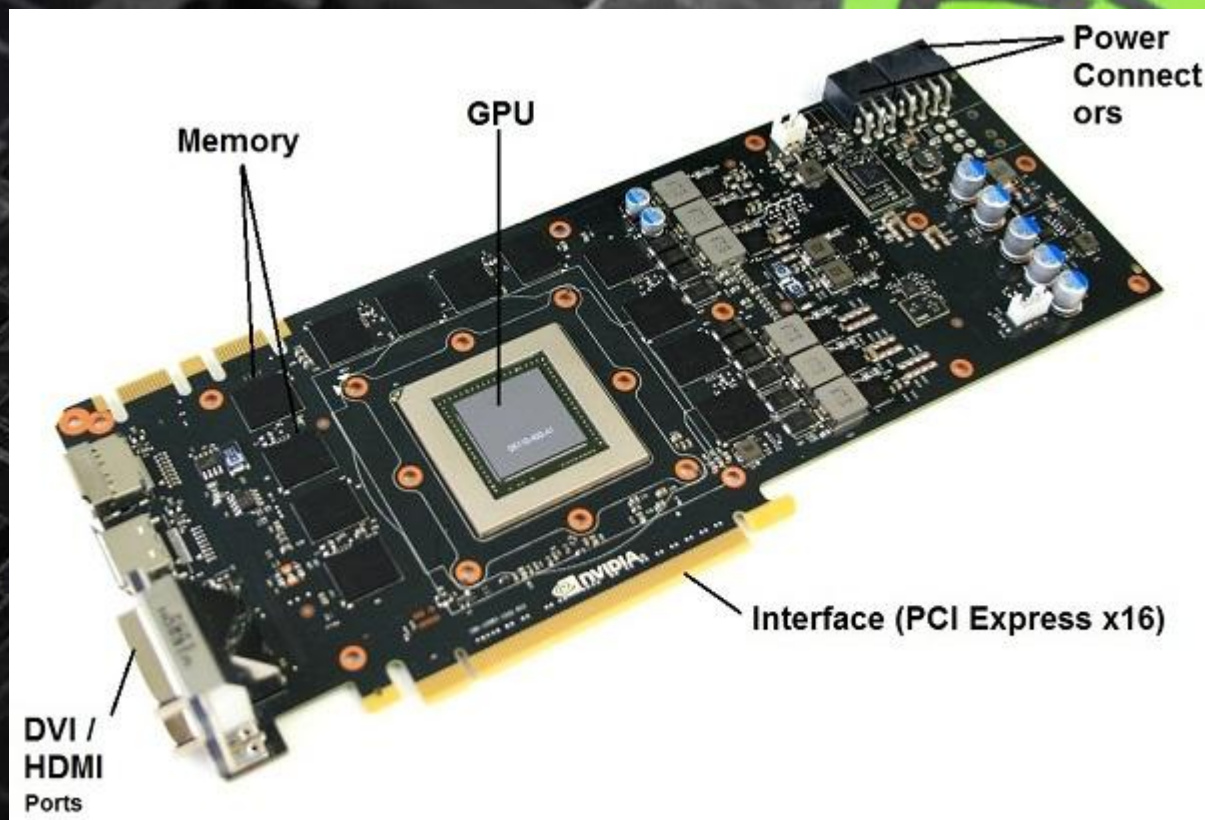
Grafická karta

Grafická karta je hardverová súčasť počítača

Spracúvava grafické dáta a zobrazuje ich na monitore

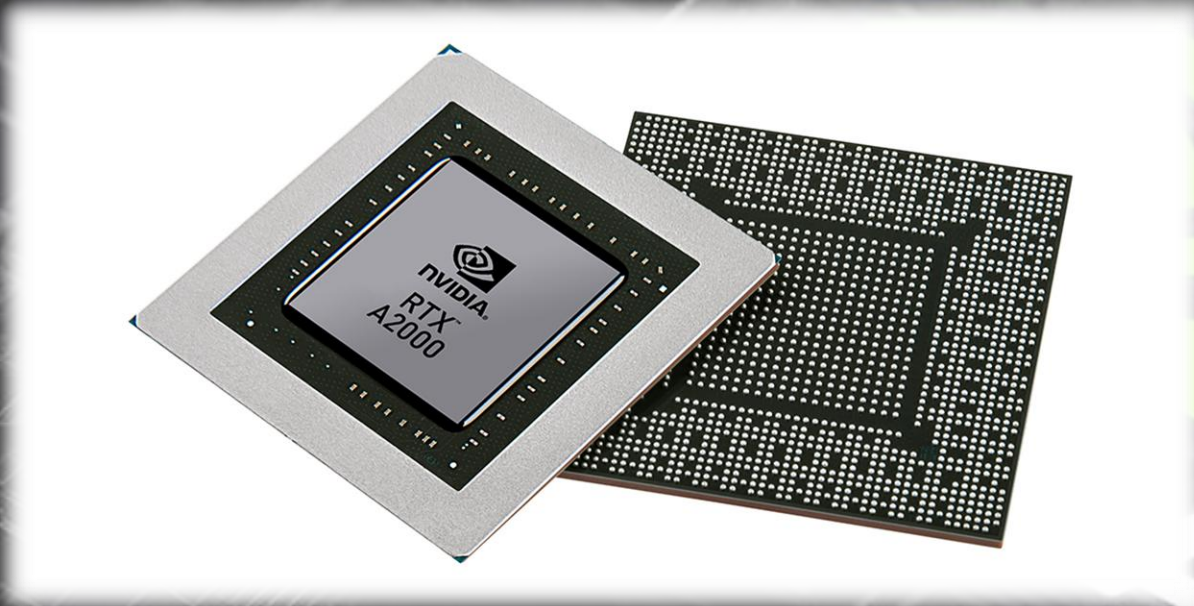
Vykonáva náročné výpočty – generovanie obrazov v hre, renderovanie, 3D modelovanie, vedecké výpočty, zobrazenie operačného systému

Hlavné komponenty



GPU(processor)
VRAM(pamäť)
Chladienie
Výstupné porty
Napájanie

Všetko je uložené na PCB(Printed Circuit Board)



GPU

GPU je špecializovaný čip navrhnutý na paralelné spracovanie veľkého množstva úloh. Moderné GPU obsahujú tisíce jadier, ktoré umožňujú efektívne vykonávať mnohé operácie súčasne.

Rozdiel medzi GPU a CPU

Vlastnosti

	GPU	CPU
Počet jadier	Niekoľko tisíc	4 až 32
Spracovanie	Paralelné	Sériové
Rýchlosť jadier	Vysoká	Nižšia
Flexibilita	univerzálne použitie	Špecializovaná (grafika)

CPU sú flexibilnejšie, pretože dokážu spracovať veľa rôznych programov a inštrukcií.

GPU sú oveľa menej flexibilné a spracúvajú len jednoduché inštrukcie, ako je základná aritmetika.

V podstate, ak potrebujeme spracovať veľké množstvo dát, GPU to zvládne rýchlejšie, a keď máme podstatne menej dát, CPU bude rýchlejšie.



Výstupné porty

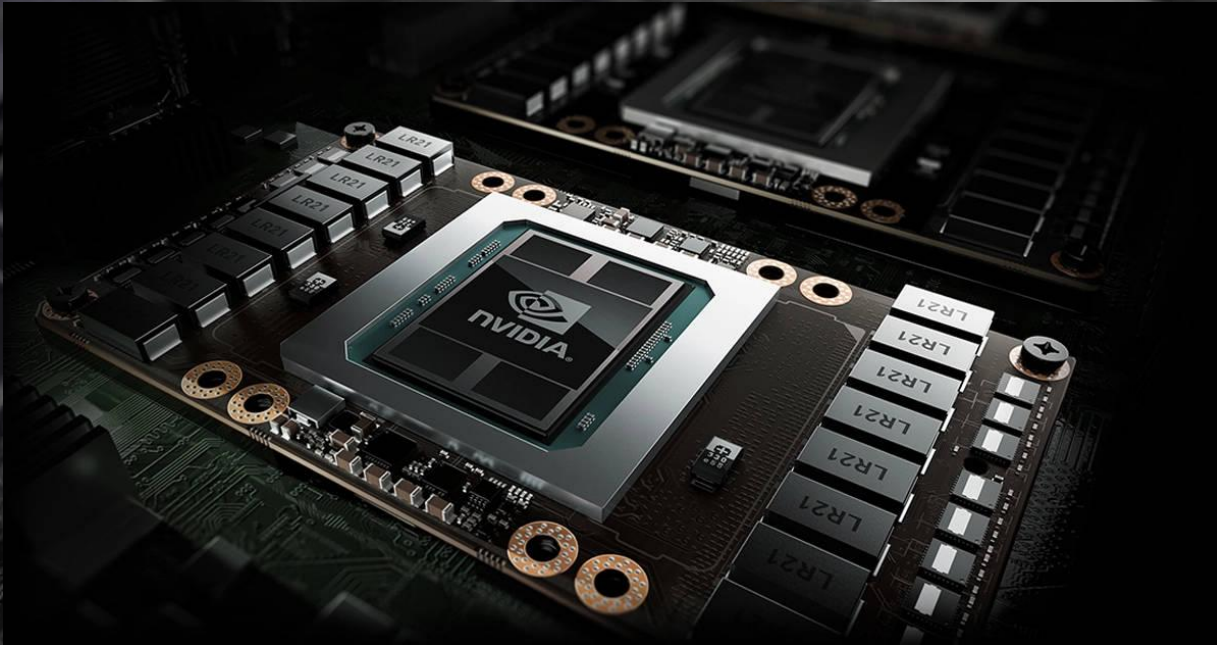
HDMI, DisplayPort alebo DVI umožňujú pripojenie monitora. Moderné karty podporujú aj viacero monitorov naraz.



Napájanie

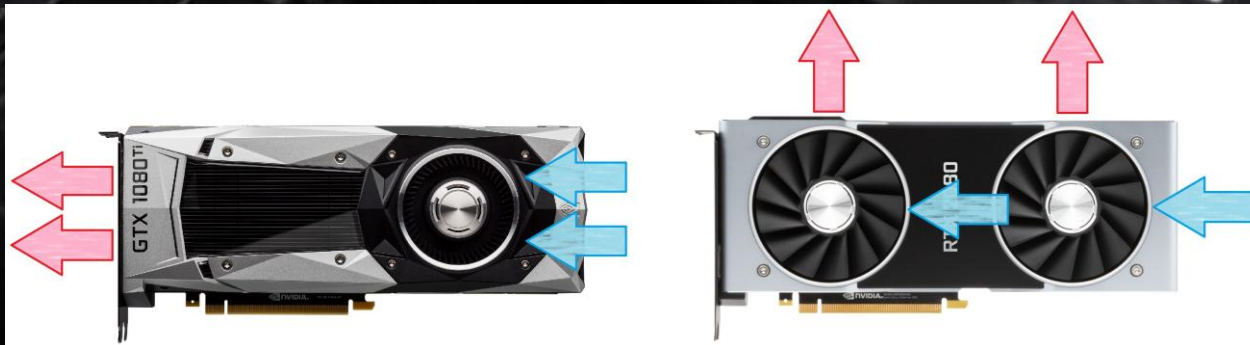
Grafické karty spotrebujú značné množstvo energie, pričom niektoré modely vyžadujú až 450 W.

Karty sú napájané cez PCIe slot a dodatočné napájacie konektory (6-pin, 8-pin)



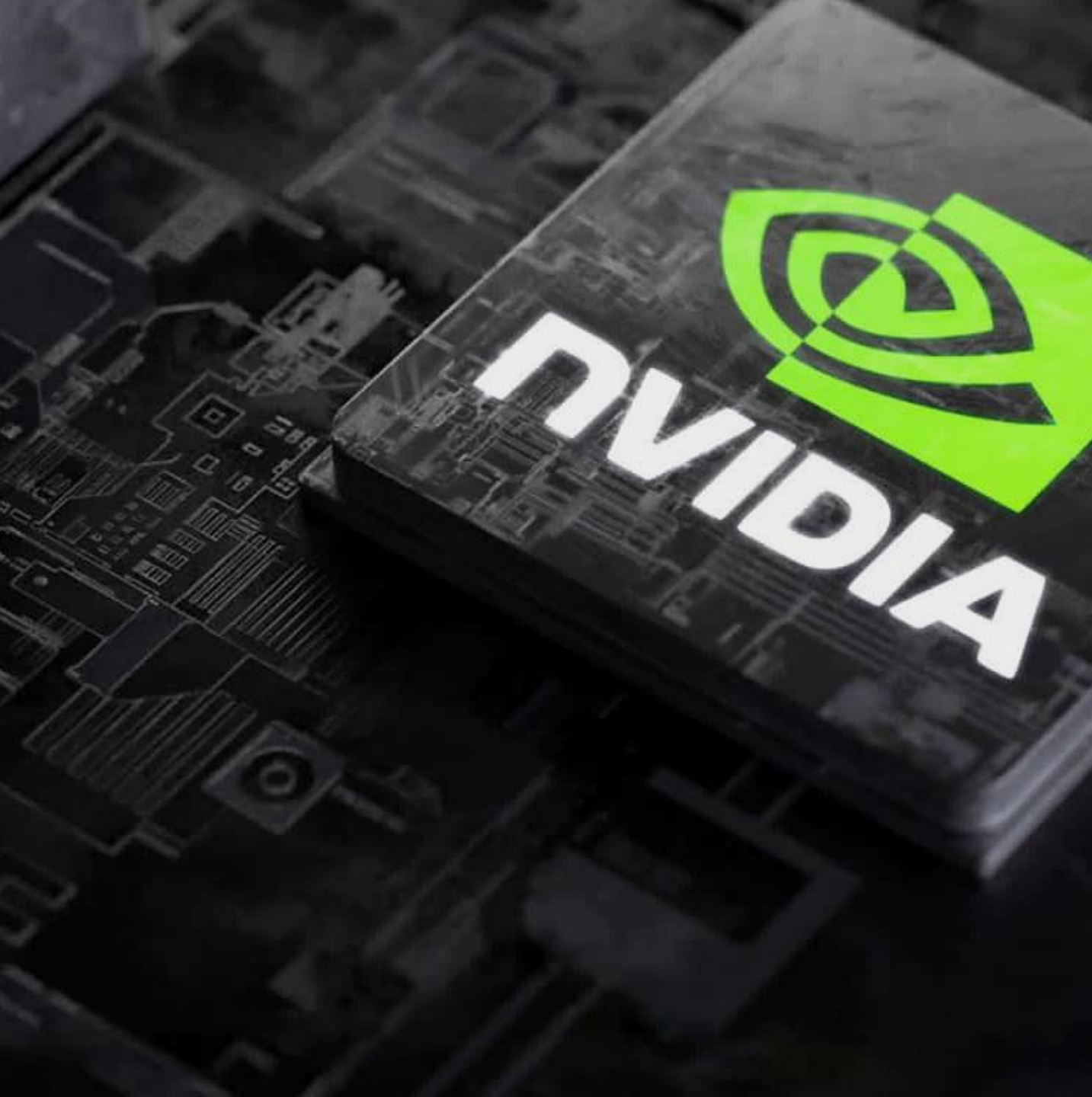
VRAM

V súčasnosti 8 až 24 GB napr.
Typu typu GDDR6 alebo HBM2
Dostatok pamäte je kľúčový
pre plynulé spracovanie
náročných grafických úloh.



Chladenie

GPU generujú veľké množstvo
tepla
sú vybavené ventilátormi,
chladičmi, vodným chladením

A close-up, high-angle photograph of a black NVIDIA GPU chip. The chip is rectangular with a complex circuit pattern visible on its surface. The NVIDIA logo, a green stylized eye, is prominently displayed in the upper right corner of the chip. Below the logo, the word "NVIDIA" is printed in large, white, sans-serif capital letters. The background is dark and out of focus, showing other components of a circuit board.

Dôležité medzníky v historií nVidií

1993 – Založenie nVidia (Jensen Huang, Chris Malachowsky, Curtis Priem)

1999 – Prvá **GPU** na svete:
GeForce 256

2006 – Zavedenie CUDA (Compute Unified Device Architecture)

2018 – Architektúra Turing: ray tracing v reálnom čase

Všetky Architektúri nVidie



Maxwell Architecture (2014) **Blackwell Architecture (March 2024)**
Kepler Architecture (2012) **Hopper Architecture (March 2022)**
Fermi Architecture (2010) **Ada Lovelace Architecture (September 2022)**
Tesla Architecture (2006) **Ampere Architecture (2020)**
Curie Architecture (2004) **Turing Architecture (2018)**
Rankine (2003) **Volta Architecture (2017)**
Kelvin (2001) **Pascal Architecture (2016)**
Celsius (1999)

GeForce 256 – "Prvá GPU na svete"

11. October 1999

17 miliónov tranzistorov

32 MB SDR VRAM

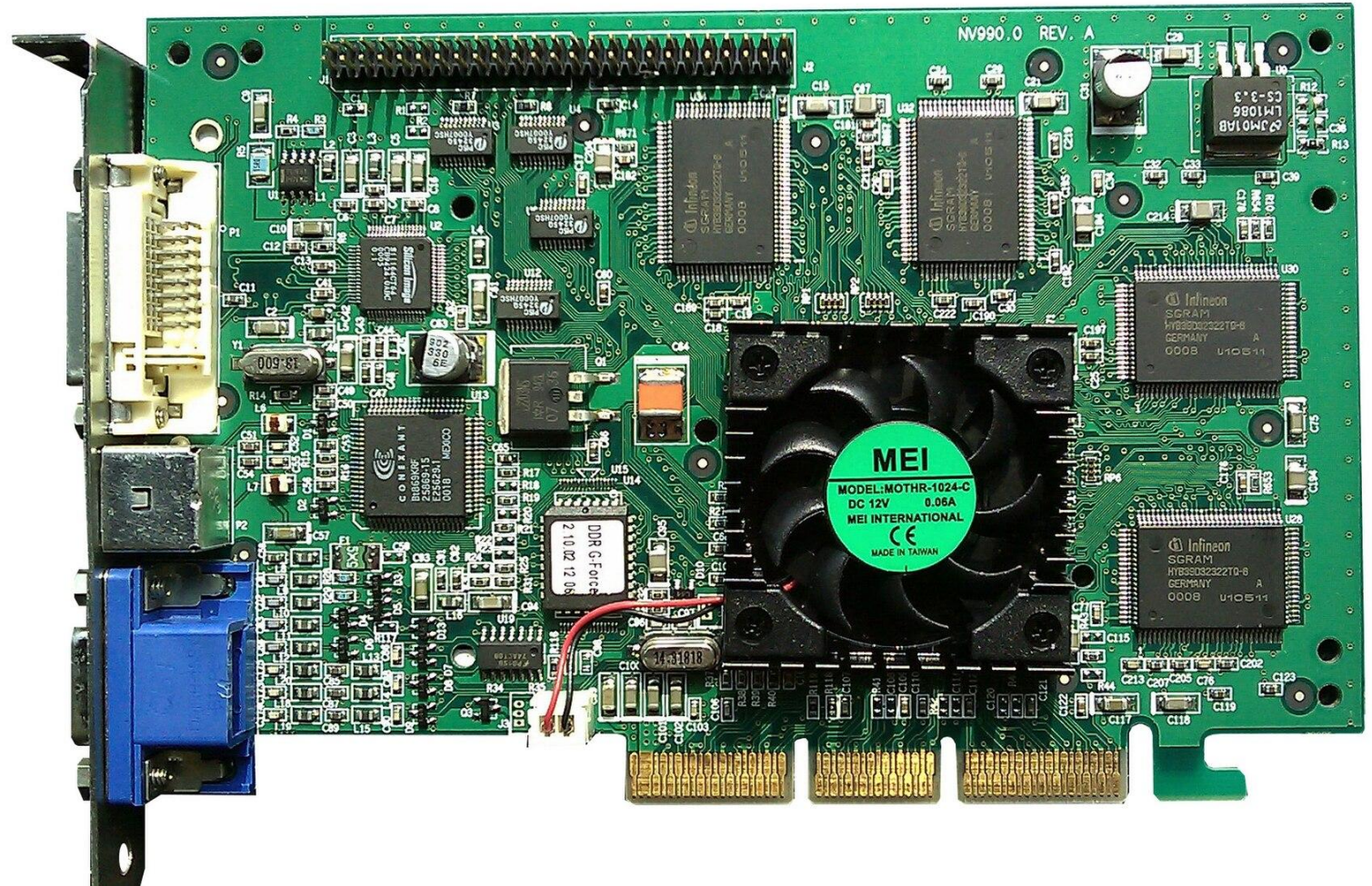
Architektúra
Celsius

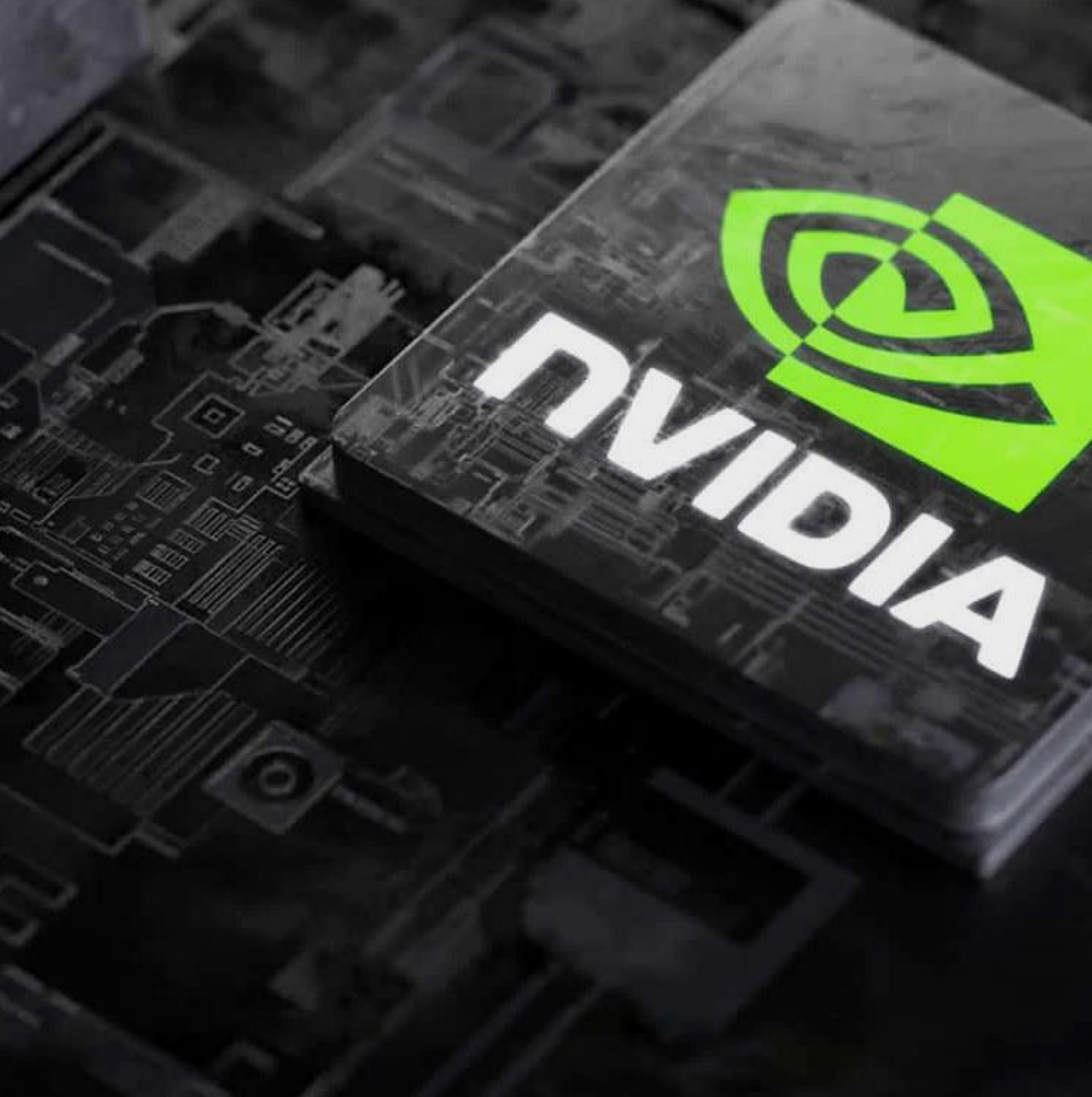
Die Size
139 mm²

Process size
220 nm

Cena
249 USD

GPU clock 120 MHz



A close-up, high-angle photograph of a black NVIDIA GPU chip mounted on a circuit board. The chip features the NVIDIA logo, a green stylized eye, and the word "NVIDIA" in white. The background is a dark, textured surface of the circuit board.

Architektúra Tesla(2006)

Priniesla unifikované shadery

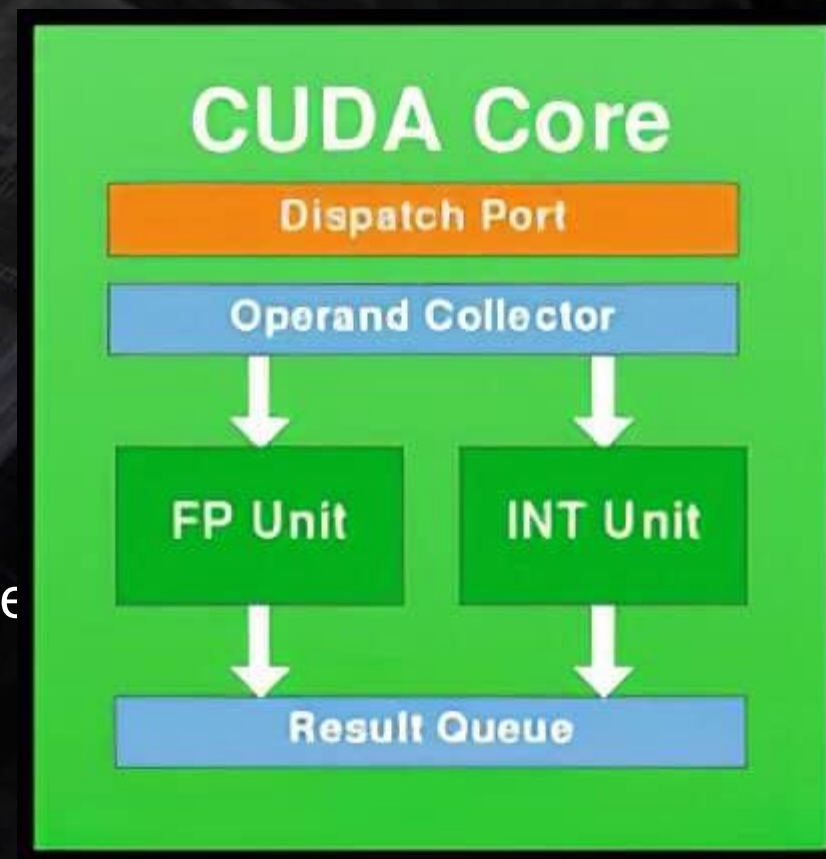
CUDA (Compute Unified Device Architecture) umožňuje používať GPU na všeobecné výpočty
napr. Vedecké simulácie, AI,
bioinformatiku

CUDA cores

CUDA jadrá sú **fyzické hardvérové výpočtové jednotky** integrované v grafických procesoroch (GPU) spoločnosti NVIDIA.

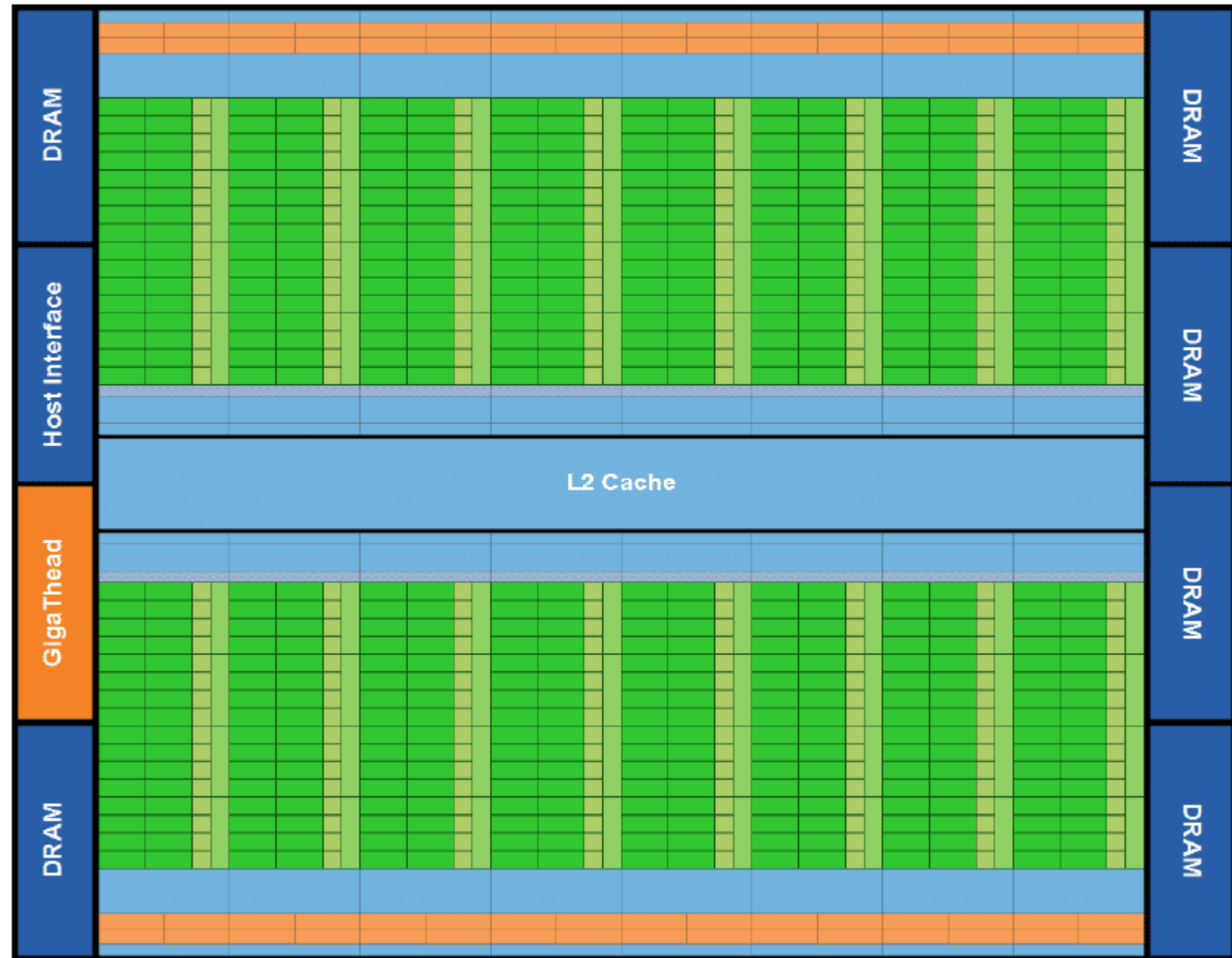
Predstavujú **základný stavebný prvok** pre vykonávanie paralelných výpočtov na GPU.

Každé CUDA jadro je schopné vykonávať jednoduchú operáciu (napríklad sčítanie, násobenie) na jednom vlákne. Keďže GPU obsahujú **veľké množstvo CUDA jadier**, dokážu súčasne spracovávať tisíce až milióny operácií.



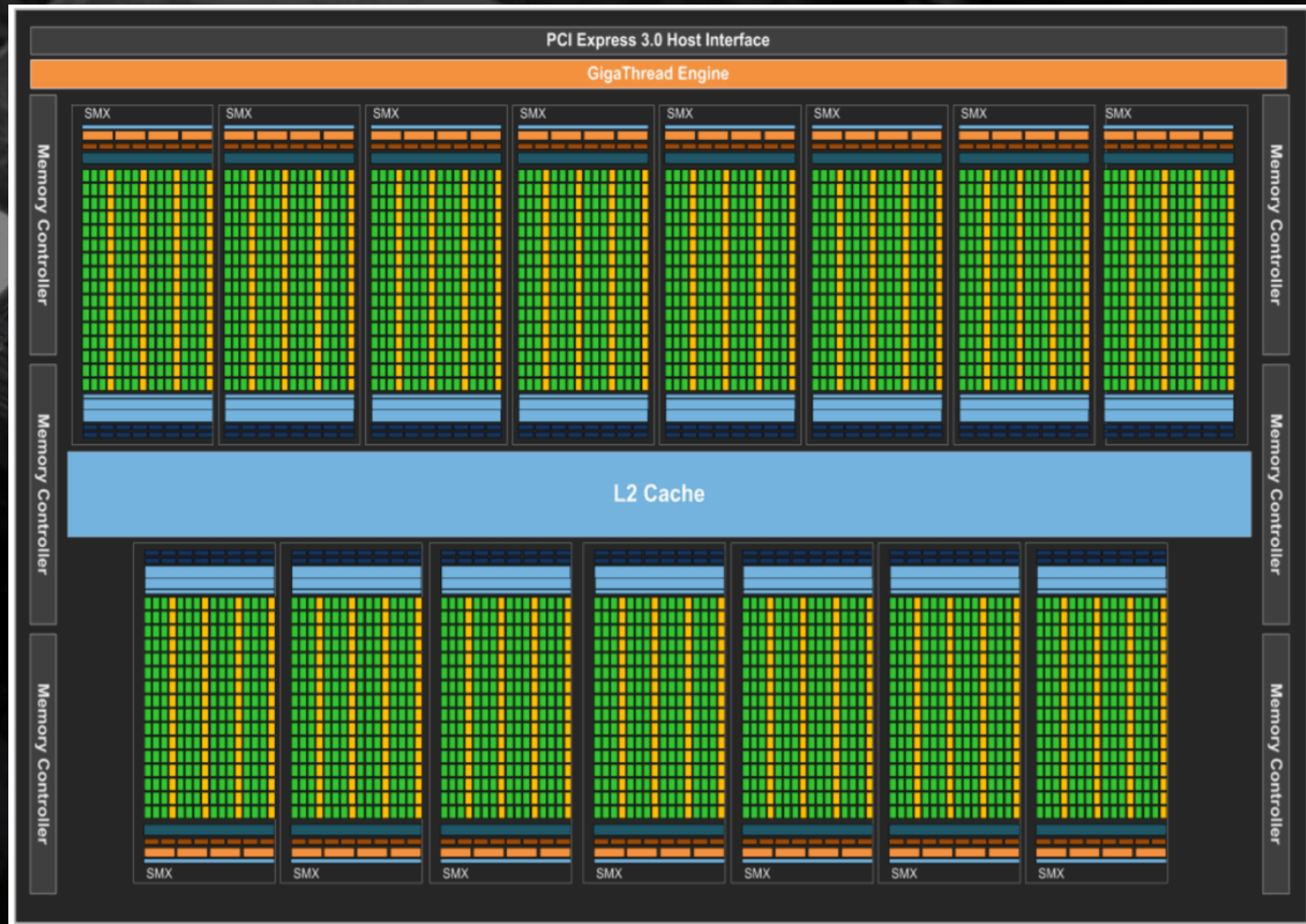
CUDA cores

Prvá GPU z Fermiho architektúri(2010) mala až do 512 CUDA cores uložených v 12 Streaming Multiprocesoroch(1 SM = 32 CUDA cores)



CUDA cores

Keplerova
architektúra(2012)
strojnásobila CUDA cores
(až 1536)
Založená na 28 nm
výrobnej technológii



CUDA cores

Každou ďalšou generáciou nVidia pridávala viac a viac CUDA jadier.

Maxwell (2014) a Pascal (2016) mali až 128 CUDA jadier na jeden SM (Streaming Multiprocessor).



GeForce 8800 GTX – CUDA

8. November 2006

681 miliónov tranzistorov

758 MB GDDR3 VRAM

Architektúra

Tesla

Die Size

428 mm²

Process size

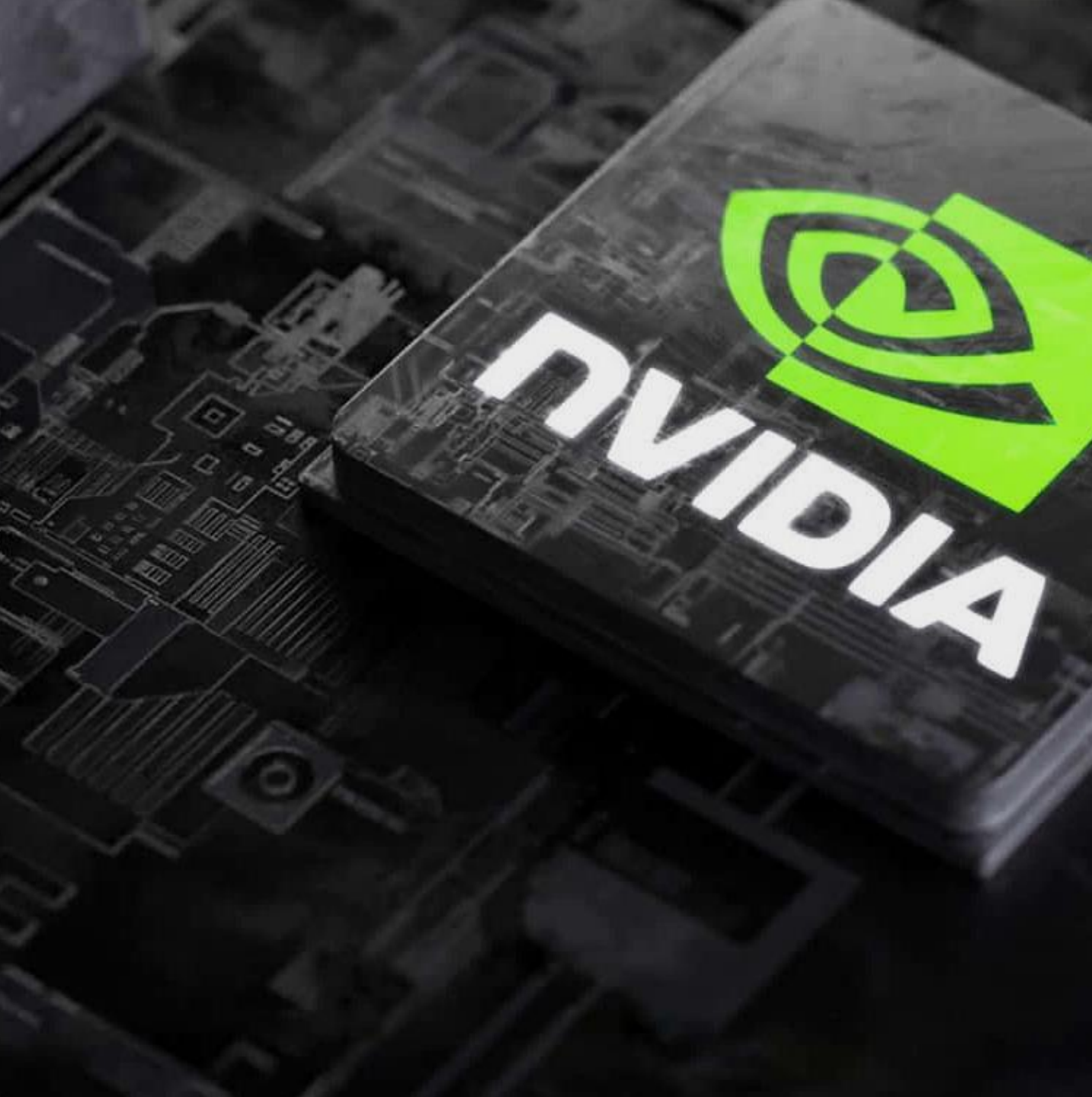
90 nm

Cena

599 USD

GPU clock 576 MHz



A close-up, angled view of a black NVIDIA GPU chip mounted on a circuit board. The chip features the NVIDIA logo, which consists of a green stylized eye icon above the word "NVIDIA" in white capital letters. The background shows the intricate patterns of the circuit board.

Architektúra Turing(2018)

Ray Tracing (RTX) jadrá špeciálne
na výpočet odrazov a lomu svetla,
osvetlenie

Prelom v renderovaní

Tensor jadrá – deep learning

GeForce RTX 2080 – Ray Tracing, Tensor jadrá

20. September 2018

13.6 biliónov tranzistorov

8 GB GDDR6 VRAM

Architektúra
Turing

Die Size
545 mm²

Process size
12 nm

Cena
699 USD

GPU clock 1710 MHz



Architektúra Touring

Počet CUDA jadier bol znížený zo 128 na 64 a boli pridané Tensor jadrá a Ray Tracing jadrá.

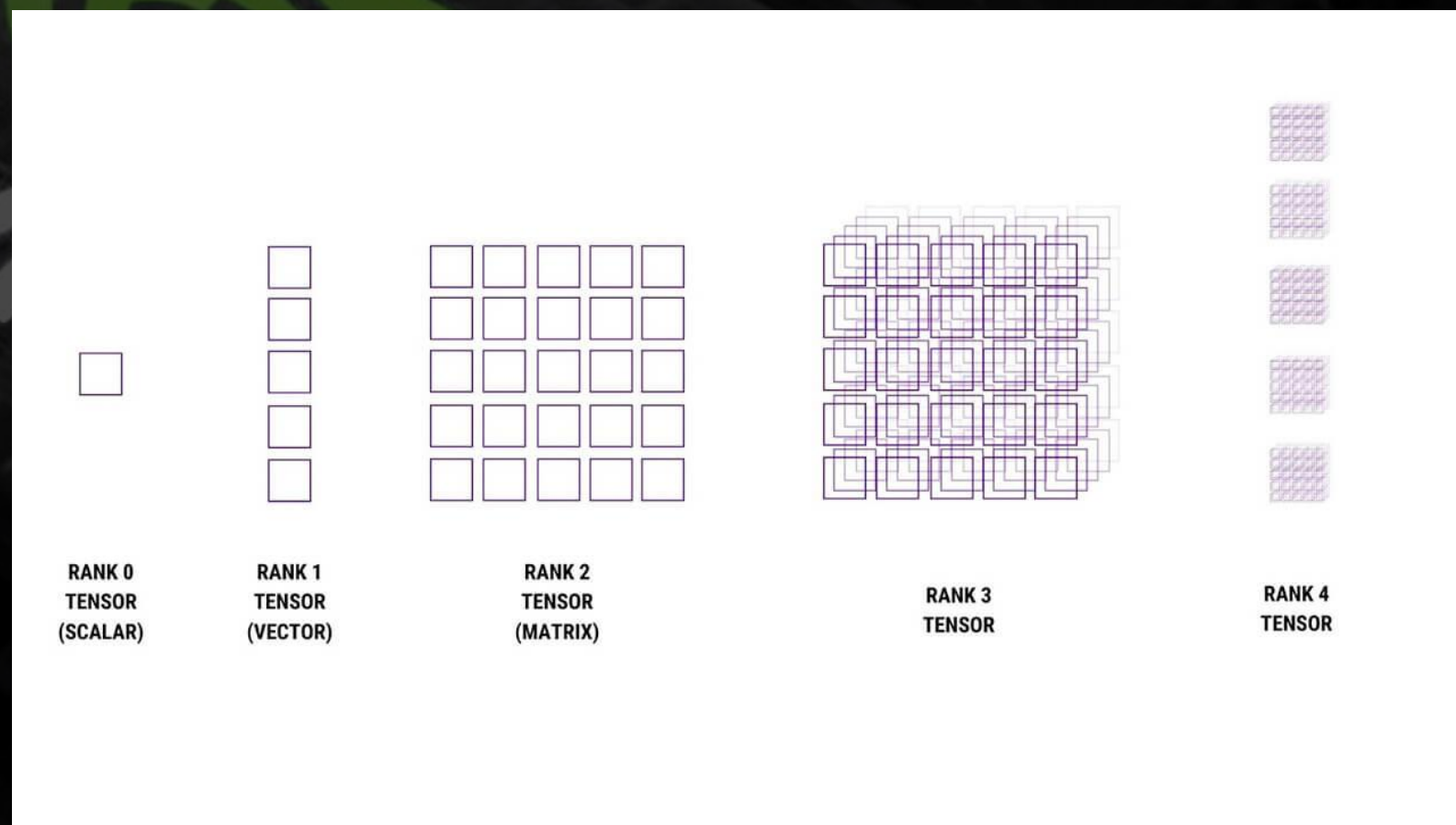
Použitá výrobná technológia bola 12 nm.

Integer a floating-point operácie boli od seba oddelené.



Architektúra Touring - Tensor

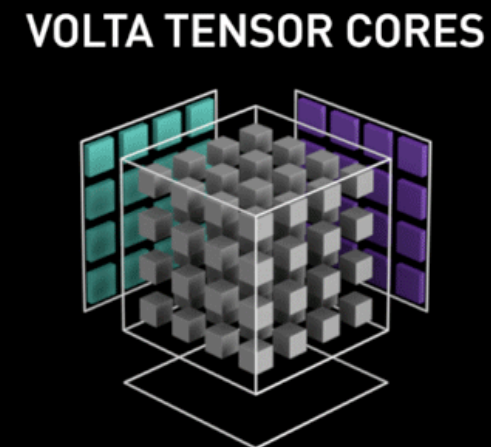
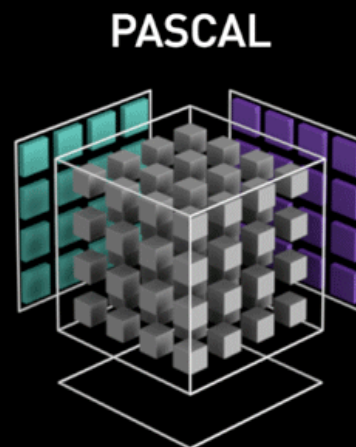
Tensor je dátový typ, ktorý dokáže reprezentovať všetky typy dát. Môžeme si ho predstaviť ako kontajner, v ktorom je uložená viacrozmerná množina údajov



Architektúra Touring – Tensor cores

Tensor Cores vykonávajú veľmi rýchle násobenie a sčítanie týchto tensorov, hlavne v operáciách typu maticové násobenie, čo je kľúčové napríklad pri geometrických transformáciách alebo pri tréňovaní neurónových sietí.

Prvé Tensor cores mala architektúra Volta(2017)



TURING TENSOR CORE

114 TFLOPS FP16

228 TOPS INT8

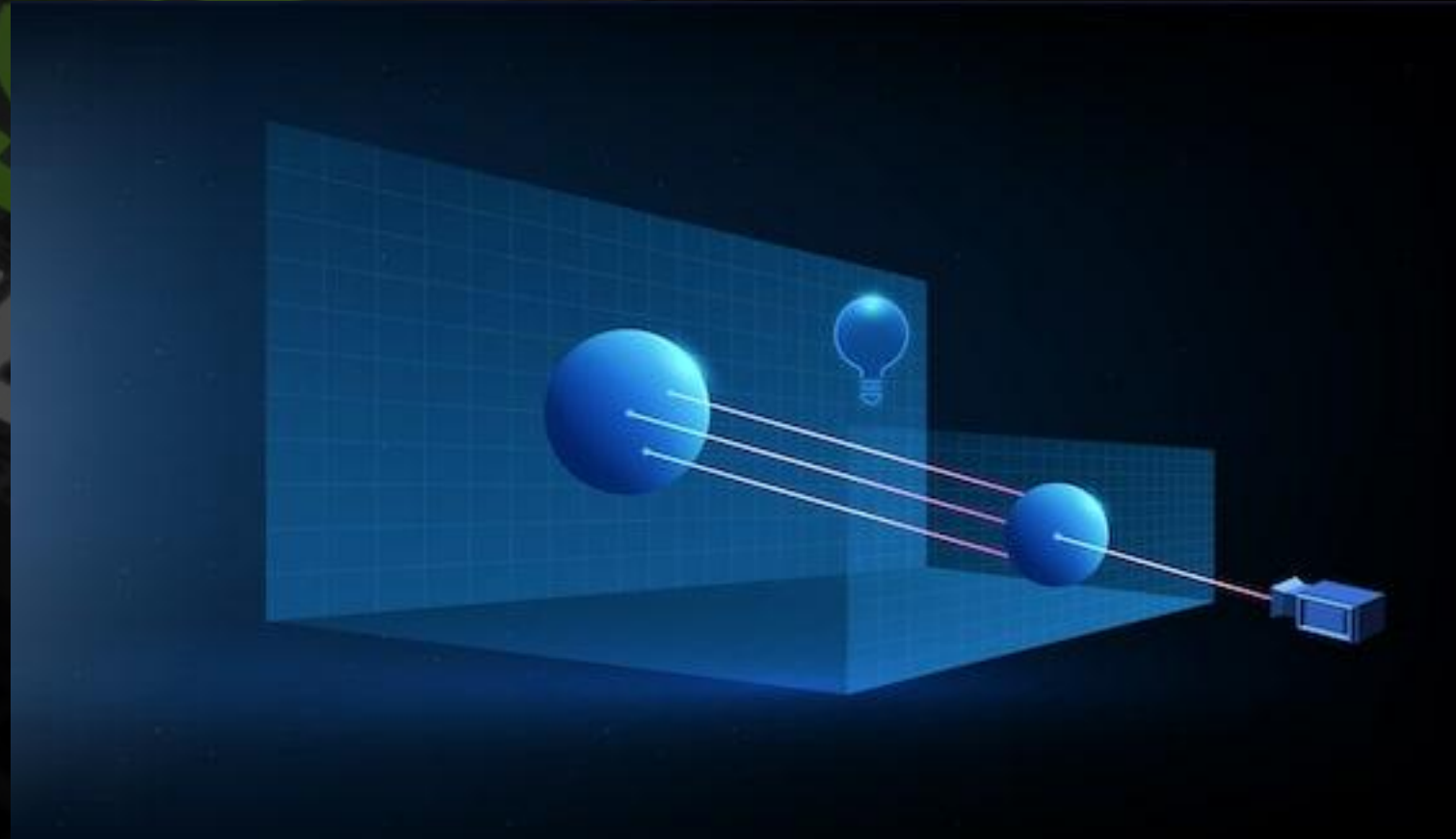
455 TOPS INT4

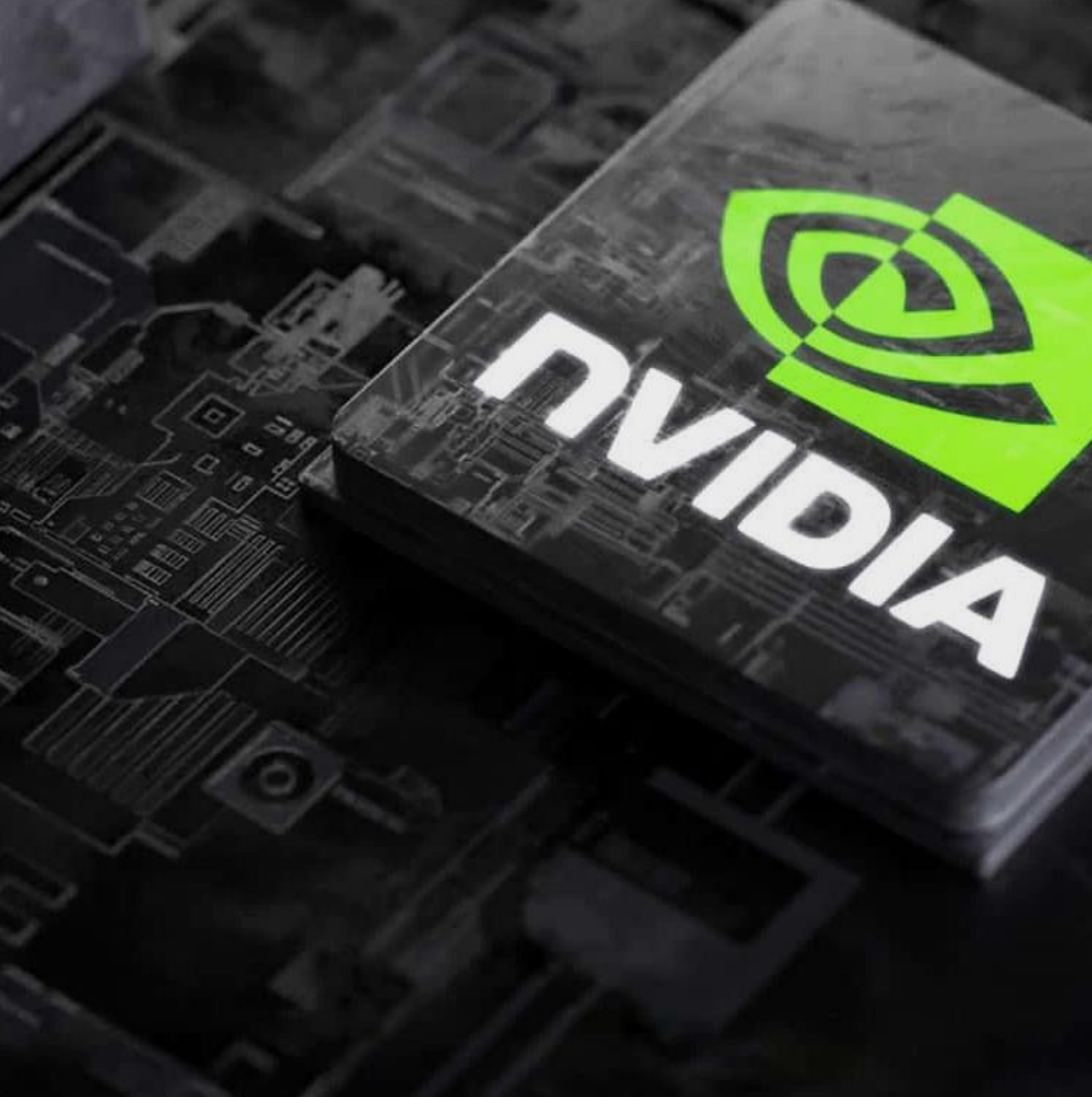


Architektúra Touring – Ray Tracing cores

Ray Tracing je výpočtovo náročná technológia, vyžaduje simuláciu dráhy svetelných lúčov, ktoré sa odrážajú od objektov v scéne a interagujú s rôznymi materiálmi.

RT Cores urýchľujú túto simuláciu tým, že vykonávajú špecifické výpočty pre určovanie priesečníkov lúčov so scénou (ray-triangle intersection)



A close-up, angled view of a black NVIDIA GPU chip. The chip's surface is covered in intricate circuitry. The NVIDIA logo, a green stylized eye, is prominently displayed in the upper right, and the word "NVIDIA" in white capital letters is printed below it.

Architektúra Blackwell 2.0(2025)

DLSS 4

pomocou umelej inteligencie
zvyšuje výkon hier generovaním
dodatočných snímok a zlepšuje
kvalitu obrazu pri nižšom výkone.
detailnejšie ray tracing efekty a
vyššie FPS aj pri náročných
scénach.

3 AI framy za 1 skutočný

Tensor jadrá 5. generácie

GeForce RTX 5090 – DLLS 4

30. Január 2025

92,2 biliónov tranzistorov

32 GB GDDR7 VRAM

Architektúra
Blackwell 2.0

Die Size
750 mm²

Process size
5 nm

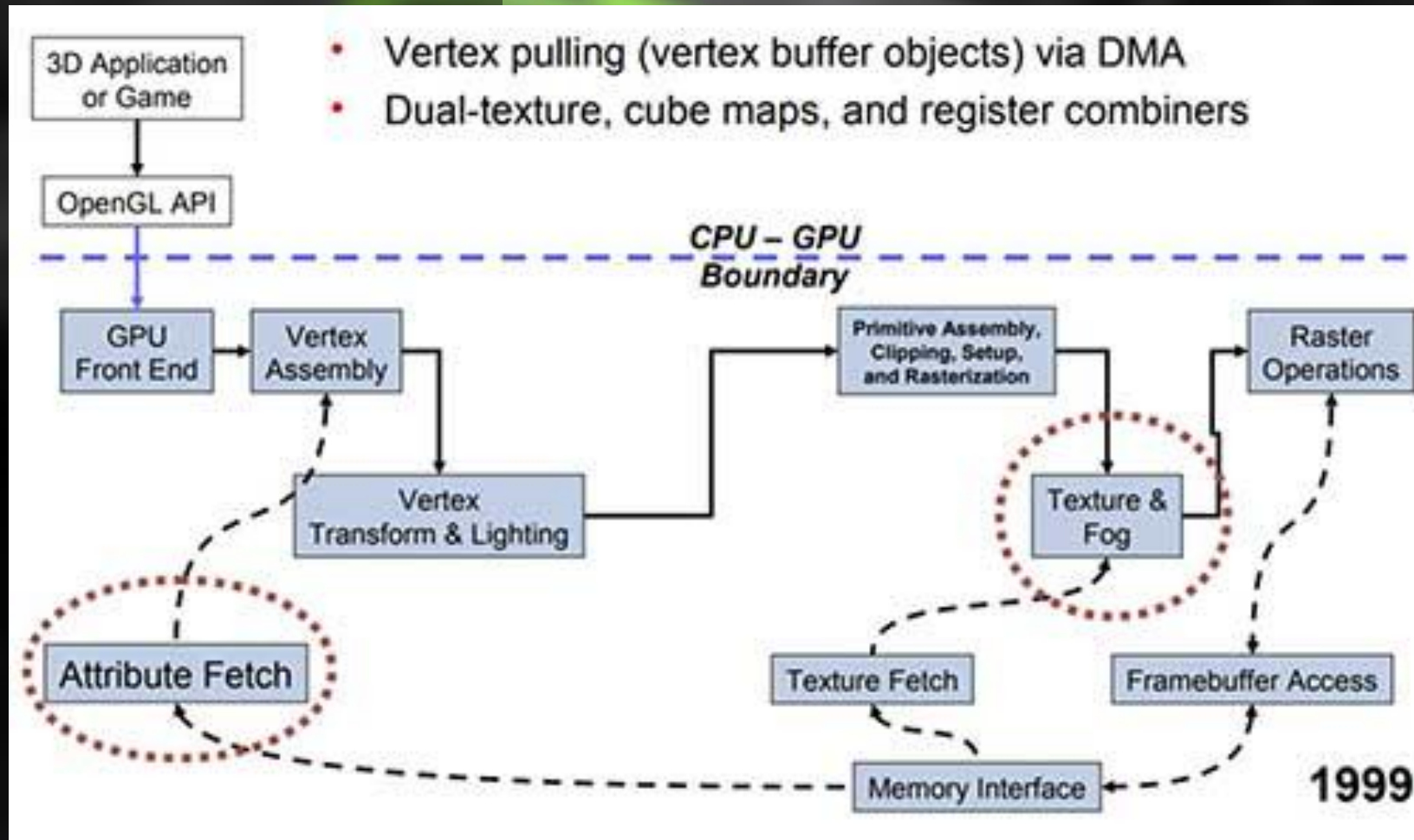
Cena
1999 USD

GPU clock 2017 MHz

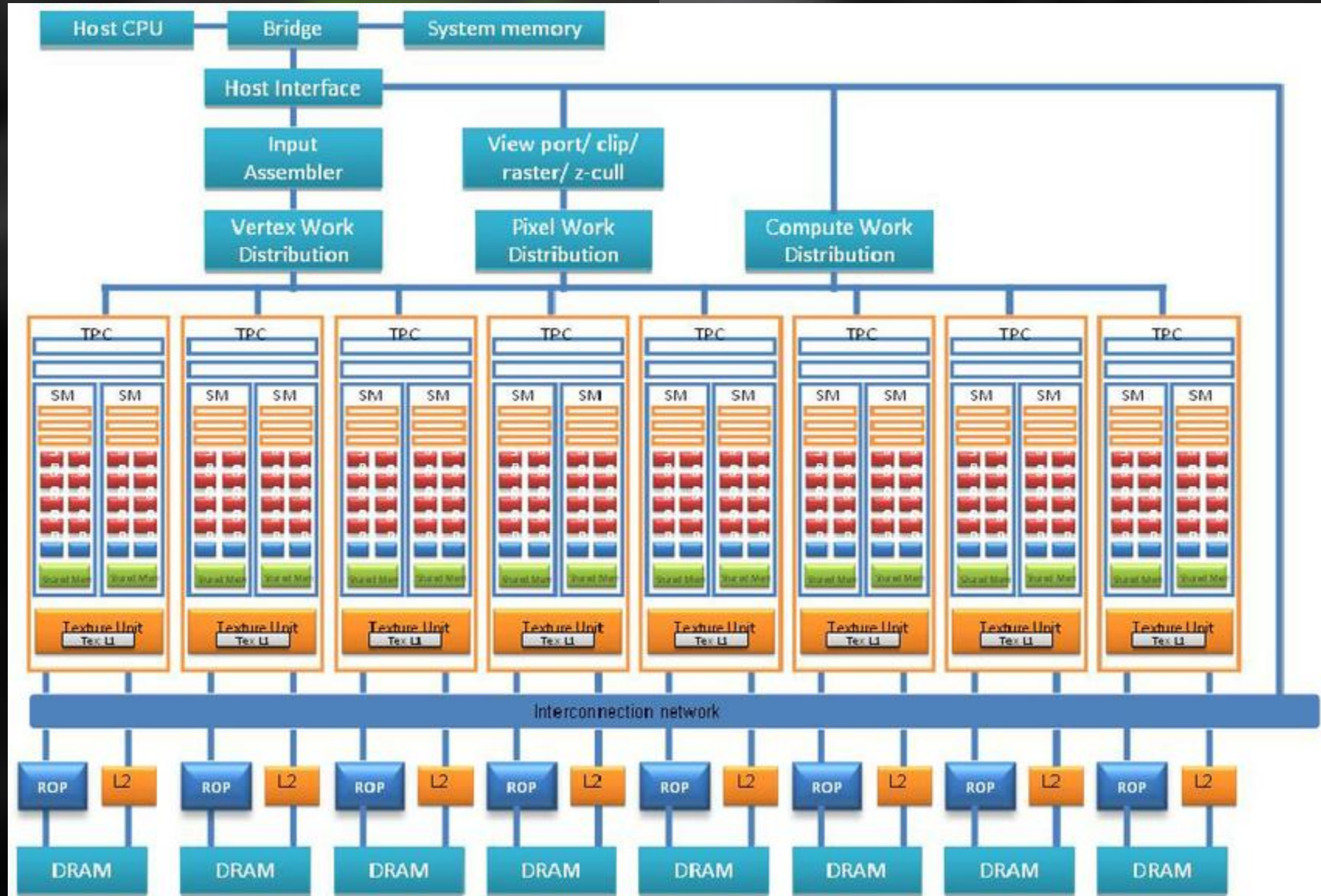


TECHPOWERUP

GeForce 256

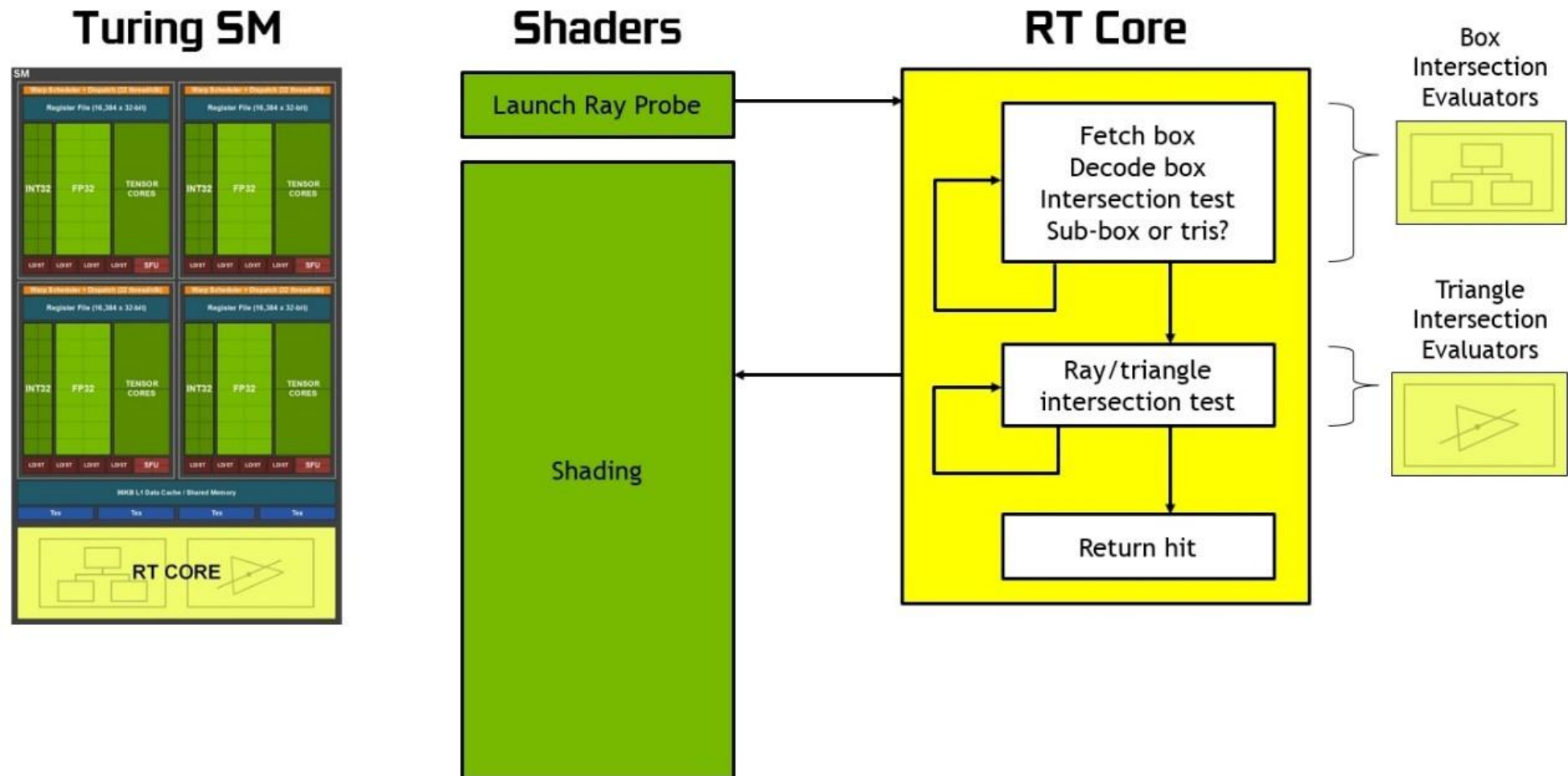


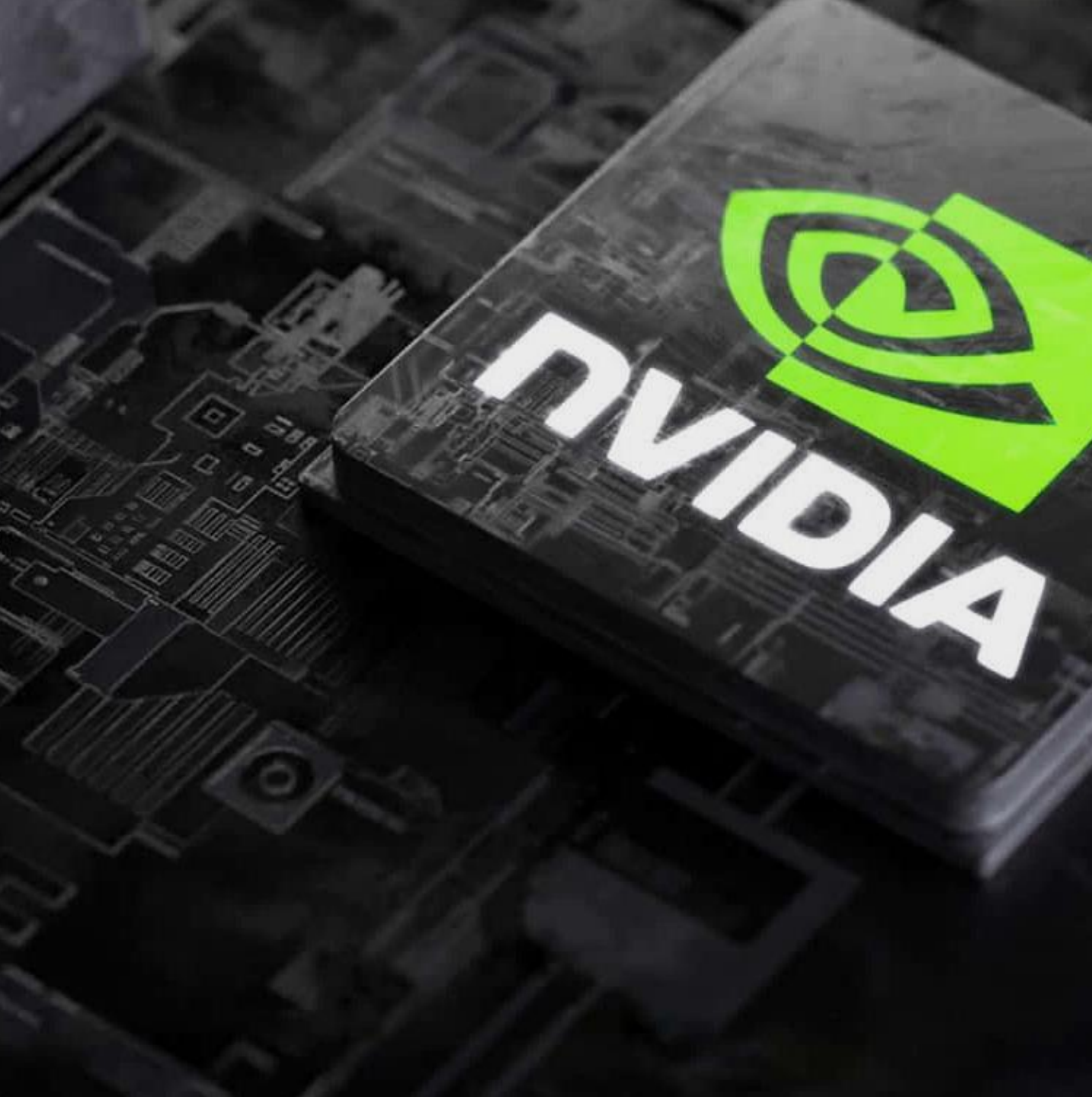
GeForce 8800 GTX



Turing Ray Tracing s RT jadrami

Hardware Acceleration Replaces Software Emulation



A close-up, angled view of a black NVIDIA graphics card. The card is resting on a dark, textured surface that appears to be a circuit board. The NVIDIA logo, a green stylized eye, is prominently displayed in the upper right corner of the card. Below the logo, the word "NVIDIA" is printed in large, white, sans-serif capital letters. The background is dark and out of focus, showing various electronic components and traces of the circuit board.

Vízie do Budúcnosti

Nová architektúra – Rubin
Môžeme očakávať čoraz viac
prispôsobenie AI – lepšie Tensor
jadrá

Cloudové hranie – GeForce NOW
podpory pre virtuálnu a rozšírenú
realitu.

Diagnostika, správa, BIOS, servis

Diagnostické nástroje:

GPU-Z, HWMonitor, MSI Afterburner, nVidia-smi (CLI nástroj)

Nastavenia v BIOSe:

PCIe sloty, integrovaná vs. dedikovaná grafika, Re-Size BAR support

Firmware:

GPU BIOS – dá sa aktualizovať cez NVFlash

Testovanie stability a výkonu:

3DMark, FurMark, Unigine Heaven

Možné poškodenia a servis:

Prehrievanie, zlý kontakt v PCIe slote, poškodené VRAM moduly

Riešenia: repaste, reflow, výmena chladienia alebo GPU

Zdroje

GRAFICKEKARTY.SK. *Čo je grafická karta a prečo je dôležitá pri hrách a práci?* [online]. Dostupné na:
<https://grafickekarty.sk/navody/co-je-graficka-karta/>

BRANCH EDUCATION. *How do SSDs Work? How does your Smartphone store data?* [online]. YouTube, 2024. Dostupné na:
<https://www.youtube.com/watch?v=h9Z4oGN89MU>

NIKTEK. *The Evolution of Nvidia GeForce Graphics (1999–2022)* [online]. YouTube, 2022. Dostupné na:
<https://www.youtube.com/watch?v=as-aVVm9JZI>

TECHPOWERUP. *NVIDIA GeForce 256 DDR Specs* [online]. Dostupné na: <https://www.techpowerup.com/gpu-specs/geforce-256-ddr.c734>

Zdroje

WIKIPÉDIA. Grafická karta. [online] Dostupné na:
https://sk.wikipedia.org/wiki/Grafick%C3%A1_karta

KIRK, David B. a NVIDIA. *NVIDIA CUDA Software and GPU Parallel Computing Architecture*. [online] January 2007. Dostupné na:
https://www.researchgate.net/publication/39432029_NVIDIA_CUDA_Software_and_GPU_Parallel_Computing_Architecture.

ALZA.CZ. Jak funguje grafická karta? [online] 4. 5. 2020. Dostupné na: <https://www.alza.cz/jak-funguje-graficka-karta>.

TECHCENTURION. **NVIDIA CUDA Cores Explained** [online].
Dostupné na: <https://www.techcenturion.com/nvidia-cuda-cores/>.
[Citované 28. 4. 2025].

Ďakujem za pozornosť