

Grafický Subsystem AMD

Michal Rýzek

Počítačové Architektúry · Univerzita Cyrila a Metoda · 2024/2025

AMD

Advanced Micro Devices

OBSAH PREZENTÁCIE

01

História AMD GPU

Vznik, vývoj a kľúčové míľniky

03

Stavba GPU čipu

CU, ROP, TMU, cache hierarchia

05

Pamäťový subsystém

GDDR6, HBM, Infinity Cache

07

Výkonnostné štatistiky

Benchmarky, TDP, špičkové produkty

02

Architektúry AMD GPU

GCN, RDNA, CDNA generácie

04

Renderovací pipeline

Rasterizácia, ray tracing, compute

06

Softvérový stack

ROCm, DirectX, OpenGL, Vulkan

08

Budúcnosť & AI

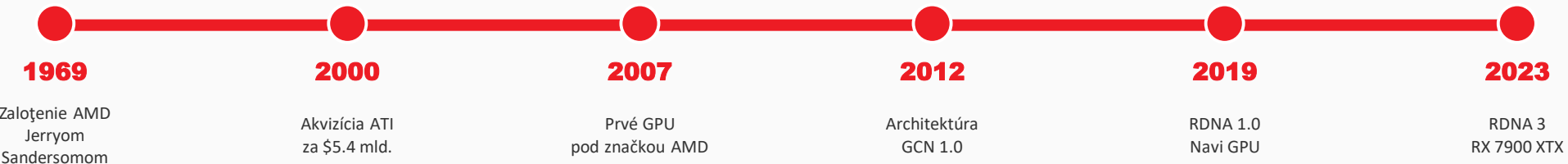
XDNA, MI300, budúce smerovanie

HISTÓRIA AMD

Zrod technologického giganta

Kľúčový moment: Akvizícia ATI Technologies v roku 2006

AMD sa akvizíciou ATI stala jedným z mála výrobcov, ktorý ovláda celý zásobovací reťazec od CPU cez GPU až po integrované riešenia. Táto strategická akvizícia v hodnote 5,4 miliardy dolárov transformovala AMD z výrobcu procesorov na full-stack poskytovateľa výpočtových platforiem.



1969

Rok založenia

\$5.4B

Akvizícia ATI

~26K

Zamestnancov

#2

GPU výrobca

ATI → AMD: EVOLÚCIA GRAFICKEJ DIVÍZIE

ATI Mach 64 (1994)

Prvé 64-bit grafické čipy, podpora VESA a PCI

Rage 128 (1998)

Podpora DVD prehrávania, 2D/3D akcelerácia

Radeon R100 (2000)

Prvý Radeon, podpora DirectX 7, T&L hardvér

Radeon 9700 Pro (2002)

Prvý DX9 čip, programmable shadery, lídrom trhu

CrossFire (2005)

Technológia multi-GPU, odpoveď na SLI

AMD Radeon HD 2900 (2007)

Prvá GPU pod značkou AMD po akvizícii ATI

ARCHITEKTÚRY AMD GPU – PREHĽAD GENERÁCIÍ

GCN 1.0–5.0

2012–2019

Proces: 28nm–14nm

Compute-first design, GPGPU, OpenCL

RDNA 1.0

2019

Proces: 7nm

Navi, 1.25× perf/W vs GCN, PCIe 4.0

RDNA 2.0

2020

Proces: 7nm TSMC

Ray Tracing HW, Infinity Cache, SAM

RDNA 3.0

2022

Proces: 5nm+6nm chiplet

AI akcelerácia, AV1, 54% perf/W rast

CDNA 1/2/3

2020–2023

Proces: 7nm–5nm

HPC/AI, HBM2e, MI200/MI300 Instinct

GCN ARCHITEKTÚRA

Graphics Core Next – Revolúcia 2012

GCN (Graphics Core Next) predstavoval zásadný architektonický zlom. AMD prešla od VLIW (Very Long Instruction Word) dizajnu k skalárnej/vektorovej architektúre s unified shader modelom.

Compute Unit (CU)

64 shader processors, 4 SIMD16 jednotky, 1 skalar procesor, 4 texturovanie jednotky, 64KB LDS

Unified Memory

Zdieľaná adresovateľná pamäť pre CPU+GPU, základ pre APU platformy

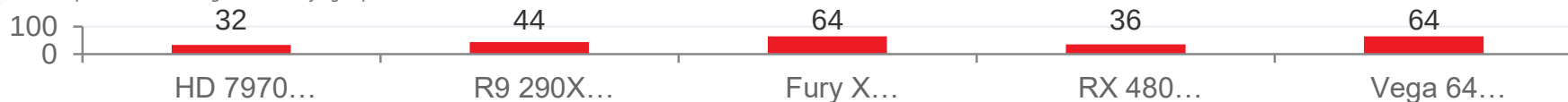
Asynchronous Compute

Simultánne vykonávanie grafických a compute workloadov na rovnakom HW

GPGPU podpora

Natívna OpenCL/HSA podpora, základ pre HPC aplikácie a ťažbu kryptomien

Počet Compute Units v GCN generáciách flagship modelov



RDNA ARCHITEKTÚRA

Radeon DNA – nová éra výkonnosti

GCN vs. RDNA – Kľúčové rozdiely

Vlastnosť	GCN	RDNA 1	RDNA 2
SIMD šírka	64 lanes (SIMD64)	32 lanes (SIMD32)	32 lanes + RT
Front End	1 per 4 CU	1 per CU	Vylepšený
Cache L1	16KB/CU	32KB/CU (2×)	32KB/CU
Cache L2	512KB–2MB	2MB	4MB
Infinity Cache	N/A	N/A	128MB–256MB
Ray Tracing	Nie	Nie	Áno (HW)
Perf/Watt	Základ	+25%	+54%

25%

Výkon/Watt rast
RDNA1 vs GCN

54%

Výkon/Watt rast
RDNA2 vs GCN

7nm

Výrobný proces
RDNA1 (TSMC)

HW RT

Ray Tracing
od RDNA 2.0

RDNA 3 – CHIPLET DIZAJN

Prvá GPU na svete s multi-chiplet architektúrou

Radeon RX 7900 XTX (Navi 31) – Chiplet GPU

MCD
6nm TSMC
2x Infinity Cache
32MB GDDR6

MCD
6nm TSMC
2x Infinity Cache
32MB GDDR6

GCD
(Graphics Compute Die)
5nm TSMC
12 534M tranzistorov
48 CU

MCD
6nm TSMC
2x Infinity Cache
32MB GDDR6

MCD
6nm TSMC
2x Infinity Cache
32MB GDDR6

Výroba

GCD: 5nm TSMC
MCD: 6nm TSMC
6× MCD chipletov

Tranzistory

GCD: 12.5 mld.
MCD (×6): ~33 mld.
Spolu: ~57.7 mld.

Infinity Cache

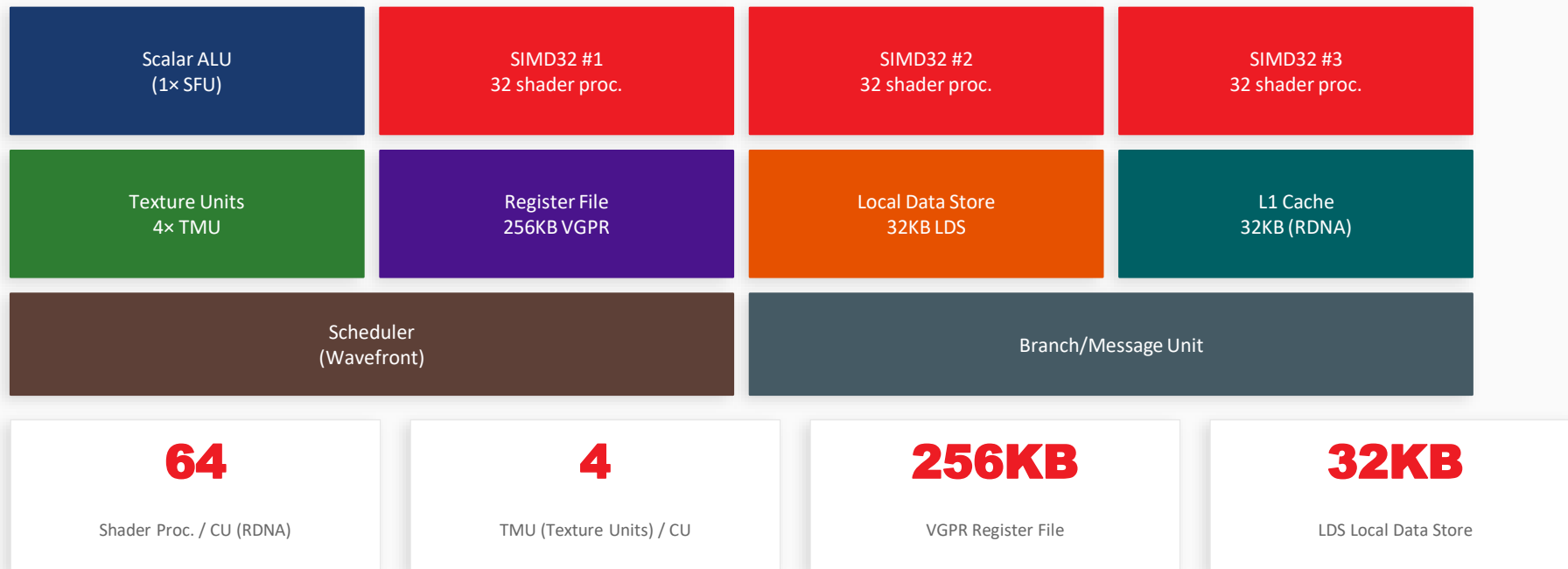
96MB L3 cache
(6×16MB MCD)
Replacement for HBM

Infinity Fabric prepojenie

Výhody chiplet dizajnu: Vyšší yield, nižšia cena, miešanie výrobných procesov pre optimal PPA

COMPUTE UNIT (CU) – ZÁKLADNÁ STAVEBNÁ JEDNOTKA

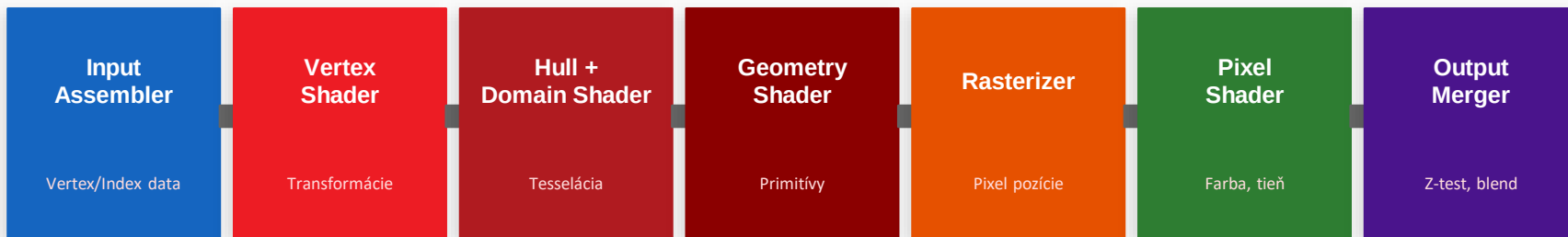
Compute Unit je základná výpočtová bunka AMD GPU. Každý CU obsahuje všetky potrebné komponenty pre spracovanie grafických aj výpočtových úloh.



RDNA: 2 CU = 1 WGP (Work Group Processor) — zdieľajú L1 cache a Shader Array riadenie

RENDEROVACÍ PIPELINE

Cesta od geometrie po pixel na obrazovke



Vertex Processing

Vertex Shader transformuje 3D súradnice pomocou matíc (model → view → projection space). Rieši osvetlenie, animácie kostí (skeletal animation) a vertex displacement.

Rasterizácia

Premena vektorových primitívov (trojuholníkov) na diskkrétne pixely/fragmenty. AMD implementuje hierarchický rasterizér pre efektívne culling mimo viditeľnej oblasti.

Pixel/Fragment Processing

Najvýpočtovo náročná fáza. Pixel Shader počíta finálnu farbu každého fragmentu – PBR tienenie, tieňové mapy, ambient occlusion, post-processing efekty.

Celý pipeline je programovateľný – Vertex, Hull, Domain, Geometry a Pixel Shadery sú HLSL/GLSL programy bežiacie na CU

RAY TRACING

Fyzikálne korektné zobrazenie svetla

Hardware Ray Tracing v AMD RDNA 2+

Ray Tracing simuluje fyzikálne správanie svetelných lúčov. Každý pixel 'vystreľuje' lúče do scény, ktoré sa odrážajú, lámu a absorbujú na povrchoch. AMD RDNA 2 zaviedla dedikované Ray Accelerator jednotky v každom CU.

BVH Štruktúra

Bounding Volume Hierarchy – hierarchický strom pre rýchle pretínanie lúčov s geometriou. AMD Ray Accelerator traversuje BVH hardvérovo.

Ray/Triangle Test

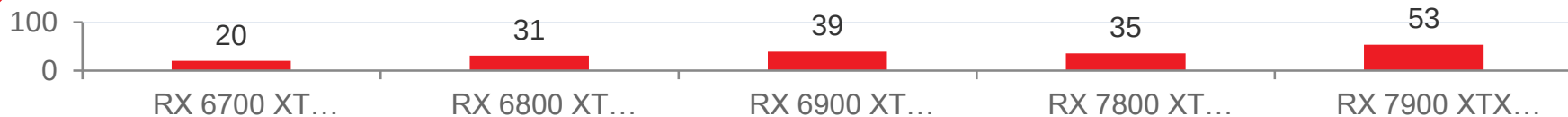
Möller-Trumbore algoritmus pre výpočet priesečníku lúča s trojuholníkom. Až 4× rýchlejší v HW než v shaderoch.

Denoising (AMD RIS)

Radeon Image Sharpening a FSR 2 používajú temporal denoising pre plynulé výsledky aj pri 1/4 vzorkovaní.

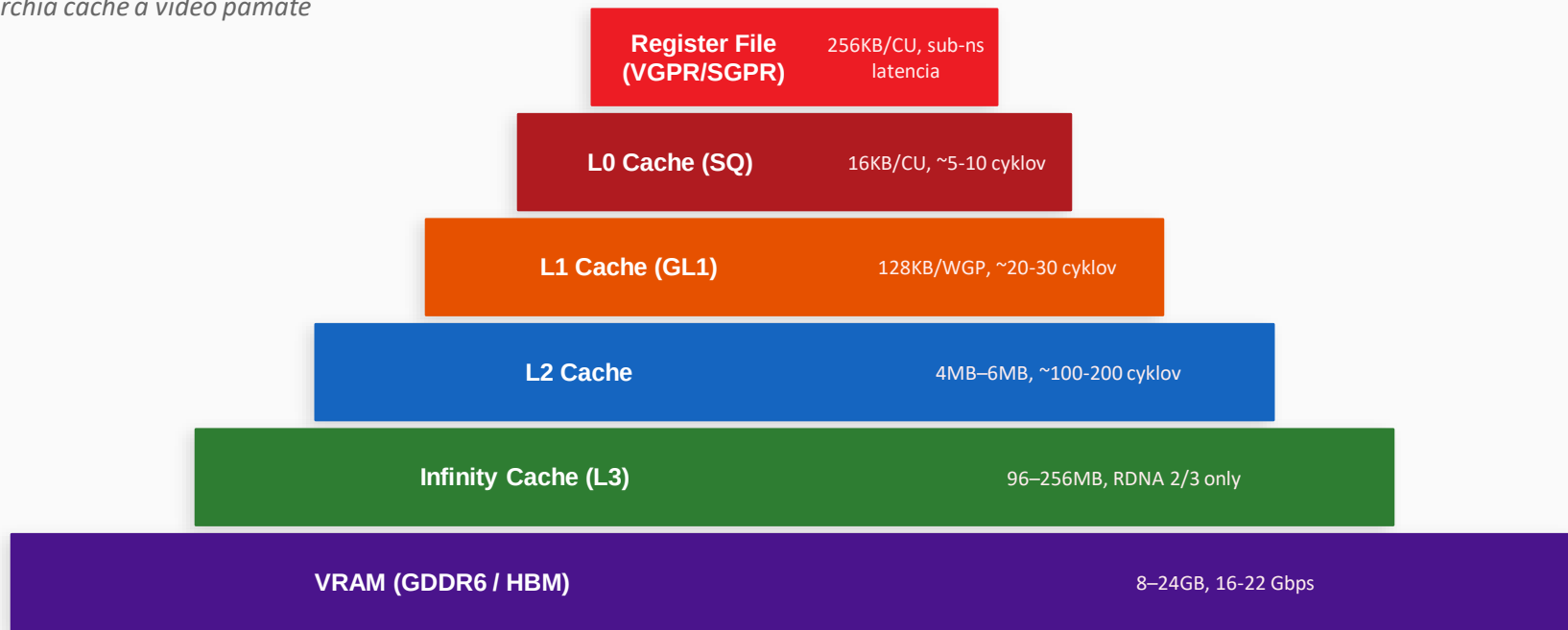
Hybrid Rendering

Kombinácia rasterizácie + RT len tam kde je to viditeľné (odlesky, tieň, AO) minimalizuje výkonnostnú záťaž.



PAMÄŤOVÝ SUBSYSTÉM

Hierarchia cache a video pamäte



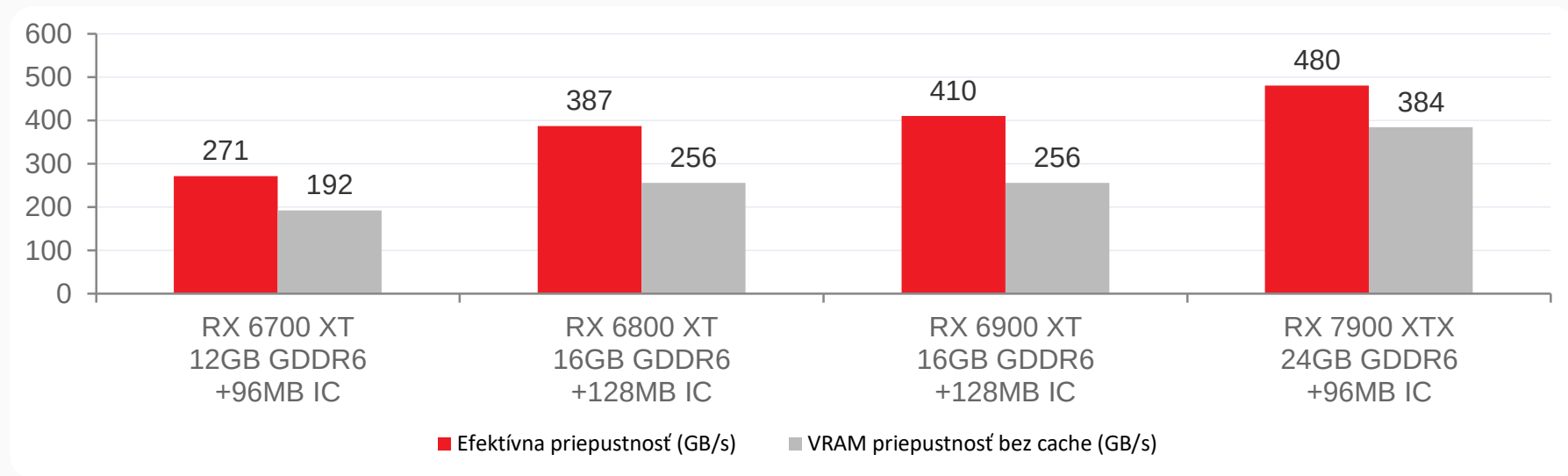
↑ Rýchlosť, Kapacita ↓

↓ Latencia, Cena/bit ↑

INFINITY CACHE

AMD inovatívne riešenie pamäťovej priepustnosti

Infinity Cache je veľkokapacitná L3 cache implementovaná na GCD/MCD chiplete v RDNA 2 a 3. Nahrádza HBM pamäť pre mainstream tier pri zachovaní vysokej efektívnej šírky pásma.



96MB

IC v RX 7900 XTX
(6×16MB MCD)

3.5×

Efektívna BW boost
vs samotná VRAM

192-bit

GDDR6 zbernica
RDNA 3 flagship

HIGH BANDWIDTH MEMORY (HBM)

Pre HPC, AI a profesionálne GPU

HBM je 3D stackovaná pamäť priamo na GPU package. Používa sa v AMD Instinct (MI) radoch pre datacentrá a HPC. Poskytuje dramaticky vyššiu priepustnosť než GDDR pri nižšej spotrebe energie.

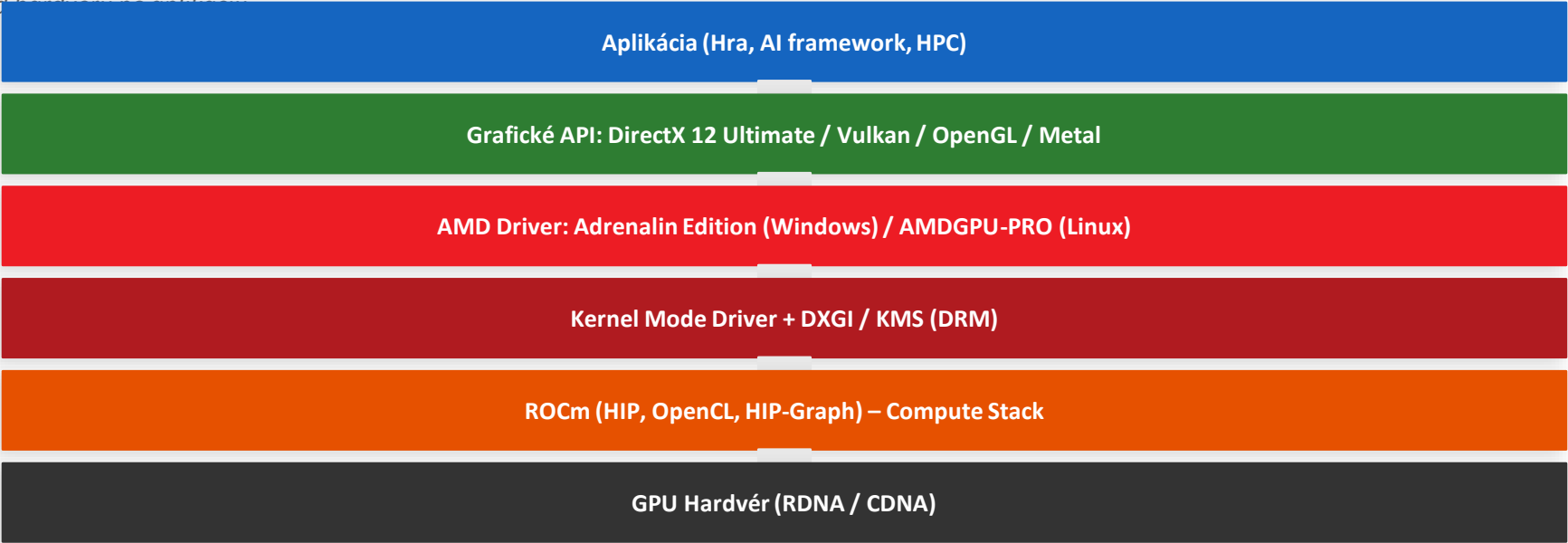
Parameter	GDDR6	HBM2e	HBM3
Zbernica	256-bit / 512-bit	4096-bit / stack	8192-bit / stack
Max. BW	~512 GB/s (512b)	~1.6 TB/s (4 stack)	~3.2 TB/s (4 stack)
Energia	~1.35V, vyššia	~1.2V, nižšia	~1.1V, najnižšia
Kapacita	8–24 GB (GPU)	16–80 GB / stack	24–192 GB / stack
Integ. forma	Samostatné čipy	CoWoS / 2.5D interposer	CoWoS / 2.5D
Cena	Nízka	Veľmi vysoká	Extrémna
Použitie	Gaming GPU	MI200, profesionálne	MI300, HPC

AMD MI300X: 192GB HBM3 • 5.2 TB/s efektívna BW • 8× HBM3 stack • 153 miliárd tranzistorov

Zdroj: AMD Instinct MI300X Product Brief 2023 | HBM Consortium Specification

SOFTVÉROVÝ STACK AMD

Od aplikácie po hardvér



ROCm (Radeon Open Compute)

ROCm je open-source platforma pre GPU výpočty. Obsahuje HIP (CUDA kompatibilita), OpenCL, ROCBlas, MIOpen (AI), a podpora PyTorch/TensorFlow. Kľúčový pre AMD vstup do AI tréningovej infraštruktúry.

Grafické API

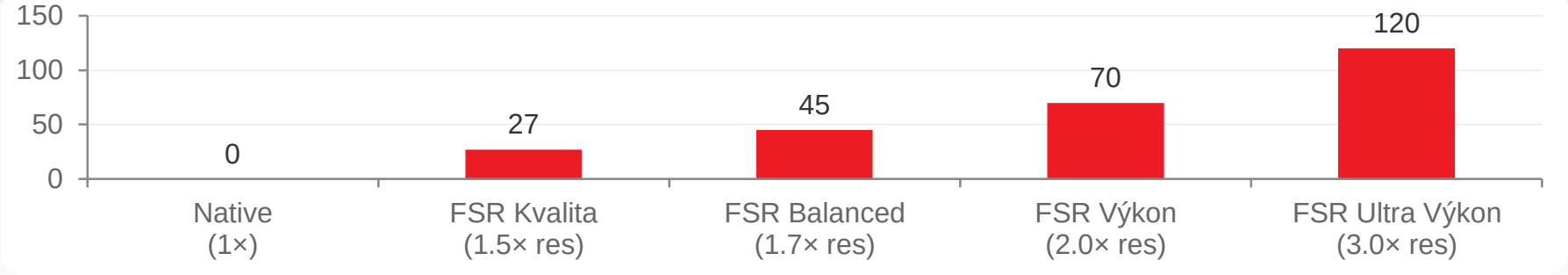
DirectX 12 Ultimate	DXR RT, Mesh Shadery, VRS
Vulkan 1.3	Nízkoúrov. API, multi-thread
OpenGL 4.6	Legacy, wide compat.
HIP/ROCm	GPGPU, AI, HPC

FIDELITYFX SUPER RESOLUTION (FSR)

AMD odpoveď na DLSS – otvorená upscaling technológia

FSR je open-source upscaling technológia AMD. Na rozdiel od NVIDIA DLSS nevyžaduje dedikovaný HW (Tensor Core) – funguje na akejkolvek GPU vrátane Intel a NVIDIA. FSR 2+ používa temporal upscaling pre kvalitu porovnateľnú s DLSS 2.

Verzia	Rok	Metóda	Kľúčová vlastnosť
FSR 1.0	2021	Spatial (EASU+RCAS)	Bez AI, akákoľvek GPU, open-source
FSR 2.0	2022	Temporal (TAA-based)	Ghost-free, motion vektory, sub-native
FSR 3.0	2023	FSR2 + Frame Gen	AMD Fluid Motion Frames, 2× FPS
FSR 3.1	2024	Vylepšený Frame Gen	AI-enhanced, Native AA mód



Priemerný FPS nárast podľa FSR 2 módu (Cyberpunk 2077, RX 6800 XT, 4K)

DISPLAY SUBSYSTEM

DisplayPort 2.1 (RDNA 3)

Max. priepustnosť: 80 Gbps (UHBR20)
Podpora: 8K@165Hz alebo 4K@480Hz
Podpora DSC (Display Stream Compression)

HDMI 2.1 (RDNA 2/3)

Max. priepustnosť: 48 Gbps
Podpora: 4K@120Hz, 8K@60Hz
Variáble Refresh Rate (VRR) natívne

AMD FreeSync Premium Pro

Adaptive Sync VRR: eliminuje tearing
LFC (Low Frame Comp.): min 2× LFC
HDR podpora, SDR compensation

Multi-Display & Eyefinity

Eyefinity: až 6 monitorov z 1 GPU
Hardvérový compositor (DXGI)
Mono + stereo display výstupy

Výstupné porty RX 7900 XTX:

2× DisplayPort 2.1

UHBR20, 80Gbps

2× HDMI 2.1

48Gbps, 4K@120Hz

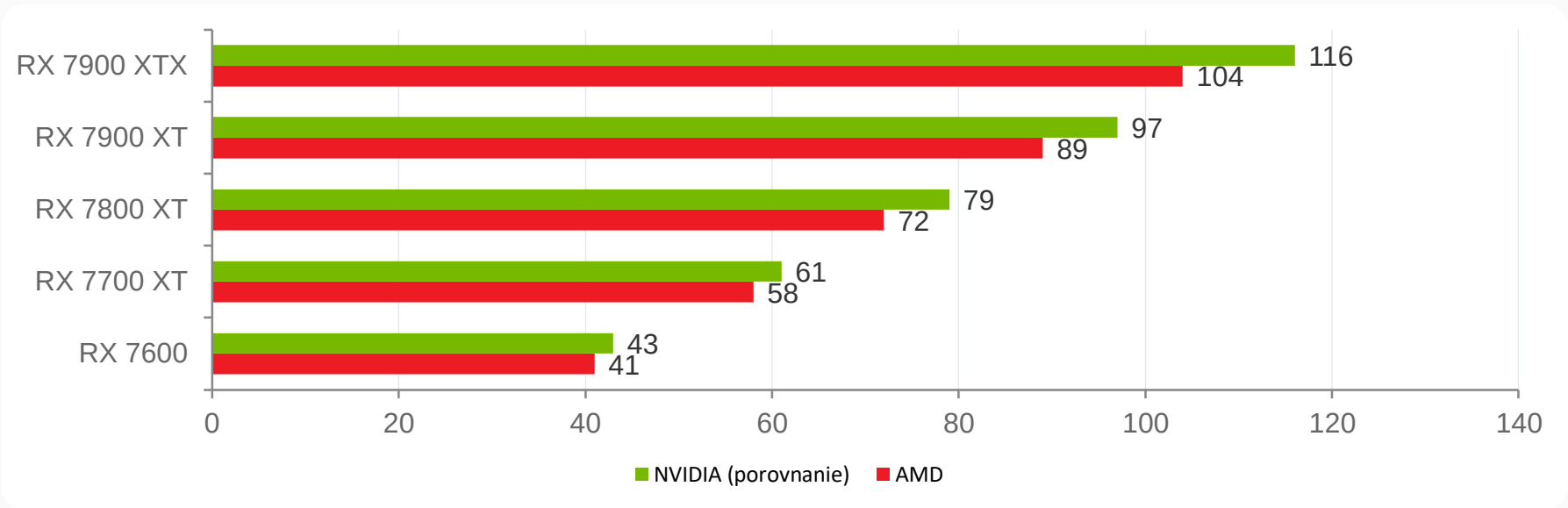
1× USB-C (DP Alt)

DP 2.1 cez USB-C

AMD FreeSync je kompatibilné s VESA Adaptive-Sync štandardom → funguje aj s NVIDIA GPU od GTX 10xx série

GAMING BENCHMARKY

Porovnanie výkonu AMD vs NVIDIA flagship GPU
Priemer FPS – 4K Ultra, DirectX 12 (Cyberpunk 2077 + Forza Horizon 5 + Spider-Man PC)



* Orientačné priemerné hodnoty z Hardware Unboxed, Digital Foundry 2023/2024. Výkon závisí od konkrétnej hry a nastavení.

104 FPS

RX 7900 XTX
4K priemer

\$999

MSRP RX 7900 XTX
launched Dec 2022

355W

TDP RX 7900 XTX
max TGP

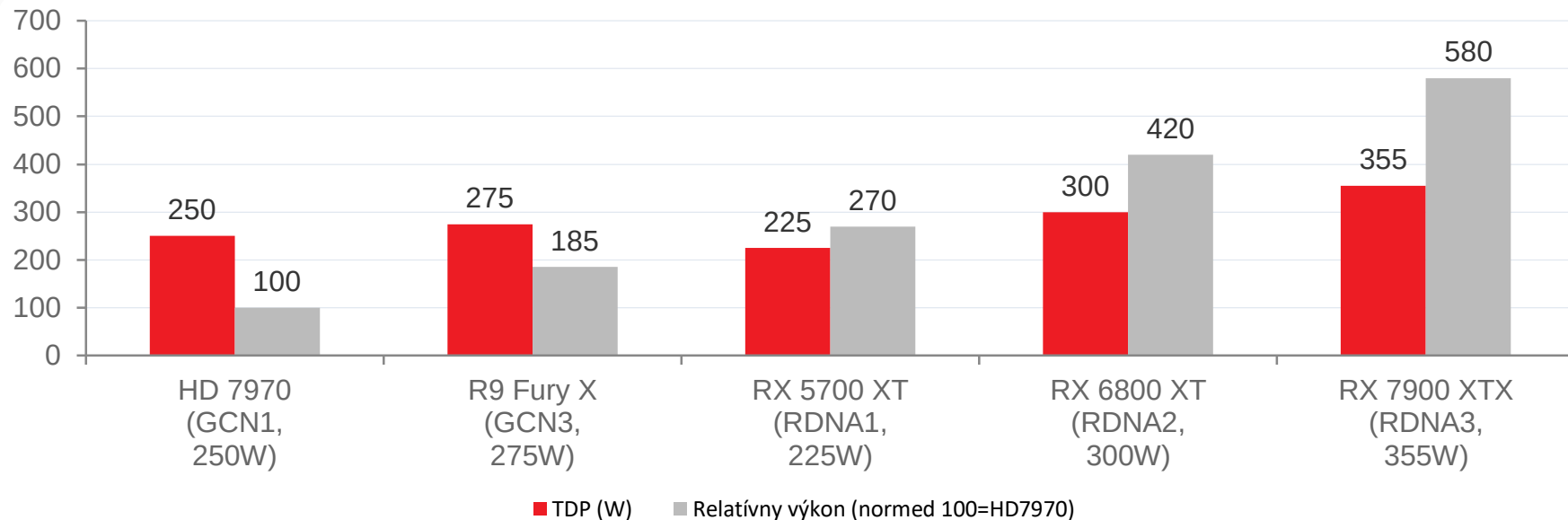
24GB

VRAM
GDDR6

SPOTREBA ENERGIE A TDP

Výkon na watt – kľúčový ukazovateľ efektivity

TDP (Thermal Design Power) udáva maximálny tepelný výkon čipu za plného zaťaženia. RDNA 3 dosiahla dramatické zlepšenie efektivity cez chiplet dizajn a 5nm proces.

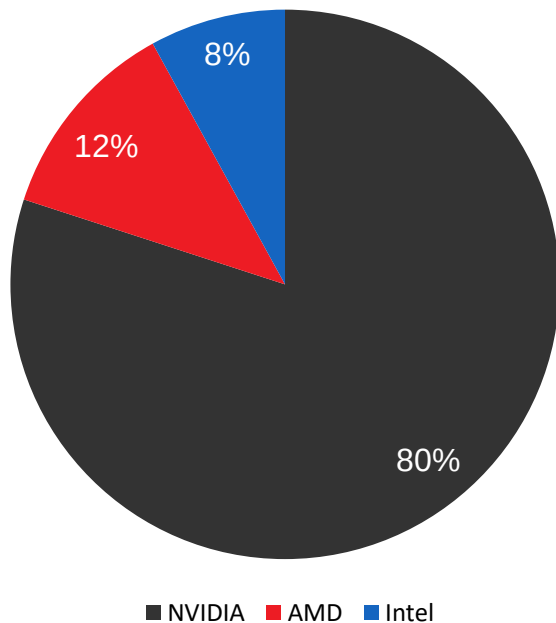


Výkon/Watt RDNA 3 vs GCN 1.0: ~5.8× viac výkonu pri len 1.4× vyššom TDP — 4.1× efektivita

RDNA 3: Smart power management – GPU dynamicky škáluje výkon podľa workloadu, nie iba podľa TDP limitu

TRHOVÝ PODIEL A KONKURENCIA

AMD vs NVIDIA vs Intel na GPU trhu
Diskrétny GPU trh (Discrete GPU, Q3 2023)



* Jon Peddie Research Q3 2023 | AMD získala cez 20% podiel v niektorých kvartáloch 2022

AMD Segmenty GPU:

Radeon RX Gaming

Mainstream a enthusiast gaming GPU
RX 7600–7900 XTX

Radeon Pro

Profesionálne workstation GPU
CAD, VFX, rendering

Instinct MI Series

HPC, AI, datacentrá
MI300X, MI200

Integrated (RDNA)

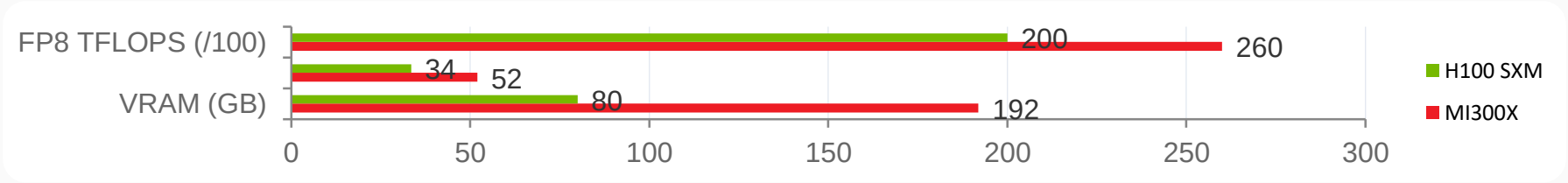
APU: Ryzen 6000/7000 iGPU
Steam Deck, Xbox/PS5

AMD INSTINCT MI300X

Najvýkonnejší AI GPU akcelerátor 2023/2024
MI300X je AMD vlajkový HPC/AI GPU akcelerátor navrhnutý priamo pre konkurenciu s NVIDIA H100. Používa 3D chiplet architektúru (UCIe) s kombináciou CDNA 3 GCD a HBM3.

Tranzistory	153 miliárd	FP16 výkon	1.3 PFLOPS
Výrobný proces	5nm TSMC (GCD) + 6nm (MCD)	BF16 výkon	1.3 PFLOPS
VRAM	192 GB HBM3	TDP	750 W
Pamäťová BW	5.2 TB/s	Prepojenie	PCIe Gen5, XGMI/Infinity Fabric
FP8 výkon (AI)	2.6 PFLOPS	Cena (2024)	~\$15 000 – \$20 000

MI300X vs NVIDIA H100 (80GB SXM):



AMD GPU V HERNÝCH KONZOLÁCH

Xbox Series X/S a PlayStation 5

AMD dominuje herné konzolové GPU – dodáva grafické subsystémy pre obe najväčšie platformy (Xbox aj PlayStation). Obe konzoly 9. generácie sú postavené na RDNA 2 architektúre.

Xbox Series X

GPU architektúra	RDNA 2 (Custom)
Compute Units	52 CU
GPU frekvencia	1825 MHz
Teraflops	12 TFLOPS
VRAM	16 GB GDDR6
Pamäťová BW	336 GB/s
CPU	8-core Zen 2 @ 3.8 GHz
Výstup	4K@120fps, 8K

PlayStation 5

GPU architektúra	RDNA 2 (Custom Sony)
Compute Units	36 CU
GPU frekvencia	2233 MHz (variable)
Teraflops	10.28 TFLOPS
VRAM	16 GB GDDR6
Pamäťová BW	448 GB/s
CPU	8-core Zen 2 @ 3.5 GHz
SSD	825 GB, 5.5 GB/s raw

XDNA – NEURAL PROCESSING UNIT (NPU)

AMD vstup do AI hardvéru pre edge zariadenia

XDNA je AMD špecializovaná architektúra pre inferenčné AI workloads. NPU (Neural Processing Unit) je integrovaný priamo do Ryzen AI čipov (Ryzen 7040+, Ryzen AI 300). Zabezpečuje efektívne lokálne AI spracovanie bez zaťaženia CPU/GPU.

AIE (AI Engine)

Array of VLIW SIMD procesorov optimalizovaných pre maticové operácie (INT8, FP16, BF16). Masívne paralelné viacjadrové spracovanie.

Pamäťová hierarchia

Lokálna SRAM pre každý AIE tile, zdieľaná L2 cache, priame DMA prenosy. Minimalizuje latenciu inference.

Ryzen AI 300 (Strix)

Až 50 TOPS NPU výkon. Integrovaný RDNA 3.5 GPU + Zen 5 CPU. Navrhnutý pre Copilot+ PC platformu Microsoftu.

AI aplikácie

Windows Studio Effects (blur, noise), LLM on-device inference, real-time preklad, automatická titulky, Adobe Firefly akcelerácia.



RDNA 4 A BUDÚCNOSŤ AMD GPU

Navi 48 – RX 9070 XT a ďalší vývoj

RDNA 4 (kódové meno Navi 4x) bola predstavená na CES 2025. Radeon RX 9070 XT a RX 9070 prinášajú dramatické vylepšenia v Ray Tracing, AI a efektívnosti oproti RDNA 3.

RDNA 4 Ray Tracing

2x výkon Ray Tracing oproti RDNA 3. Nové BVH traverzálne jednotky, 2x Ray Accelerator per CU.

AI & Compute

Nové AI akceleračné jednotky, Wave Matrix (WMMA) inštrukcie, FP8/FP4 podpora pre AI inference.

Výrobný proces

4nm TSMC (GCD) + 6nm MCD. Menší die area, vyššia frekvencia, lepší power management.

FSR 4 + AI

AI-based upscaling (FSR 4) priamo na RDNA 4 HW. Konkurencia pre DLSS 3 Frame Generation.

AMD GPU Roadmap – pohľad do budúcnosti:

2024
RDNA 3.5 (iGPU)

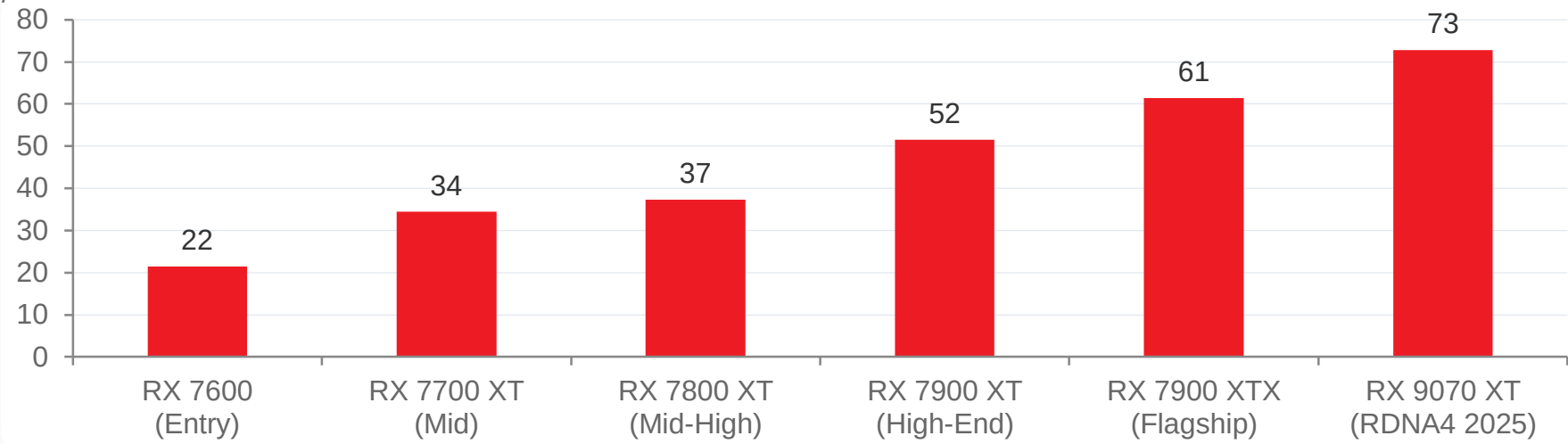
2025
RDNA 4 (RX 9070)

2025
CDNA 4 (MI350)

2026+
RDNA 5 (Navi 5x)

PRODUKTOVÉ RADY AMD RADEON 2024

Kde každá GPU odpovídá do klasifikace:



FP32 Teraflops podľa produktového radu (Radeon 7000 + preview RDNA4)

Entry (RX 7600)	Mid (RX 7700 XT)	High (RX 7900 XT)	Flagship (7900 XTX)
MSRP: \$269 VRAM: 8GB 1080p Gaming	MSRP: \$349 VRAM: 12GB 1440p Budget	MSRP: \$749 VRAM: 20GB 4K High Settings	MSRP: \$999 VRAM: 24GB 4K Ultra / Creator

TEPELNÝ MANAŽMENT A CHLADENIE

Riešenia od AMD a AIB partnerov

Moderné GPU s TDP 200–355W vyžadujú sofistikovaný tepelný dizajn. AMD navrhuje referenčný chladič (Radeon Reference Design), AIB partneri (Sapphire, XFX, PowerColor, ASRock) pridávajú vlastné riešenia.

Blower / Single Fan

Vzduch nasávajú spredu, vyfukujú vzadu cez PCI bracket. Vhodné pre tesné skrinky a servery. Vyšší hluk, nižší OC potenciál.

Open-Air (2-3 ventilátory)

Najrozšírenejšie AIB riešenie. Vzduch vyfukujú cez rebra vo vnútri skrinky. Výhodné: nízky hluk, vyšší OC. Nevýhoda: zahrievajú vnútro.

Liquid / AIO

Chladič s uzavretou vodnou slučkou (Sapphire Nitro+ Liquid). Najnižší hluk a teploty, možný OC. Drahšie, zložitejšia inštalácia.

Vapour Chamber (VC)

AMD RX 7900 XTX ref. design. Plochá tepelná trubica pre rovnomerné rozloženie tepla cez celý GPU die. Bez chladiacej kvapaliny.

83°C

Typická teplota GPU
hotspot pri zaťažení

110°C

Max. hotspot limit
RX 7900 XTX (tjunction)

95W

Typický TGP
RX 7600 (entry)

355W

Max TGP
RX 7900 XTX

AMD vs NVIDIA – TECHNOLOGICKÉ POROVNANIE

Oblasť	AMD Radeon	NVIDIA GeForce
Architektúra	RDNA 3 (chiplet, 5nm+6nm)	Ada Lovelace (monolith, 4nm)
Ray Tracing HW	Ray Accelerator / CU	RT Core (3. gen, dedikovaný)
AI upscaling	FSR 4 (open, akýkoľvek GPU)	DLSS 4 (uzavretý, len NVIDIA)
Pamäť flagship	GDDR6 + 96MB Infinity Cache	GDDR6X (RTX 4080)
Compute/GPGPU	ROCm, HIP, OpenCL	CUDA (dominantná platforma)
AI akcelérátor	XDNA NPU v Ryzen AI	Tensor Core (v GPU priamo)
Ovládače Linux	AMDGPU open-source	Proprietary / Nouveau
Frame Generation	AFMF (akákoľvek GPU)	DLSS FG (len RTX 4000+)
HPC/AI datacenter	Instinct MI300X	H100 / H200 / B100
Konzoly	Xbox + PlayStation	–

ZÁVER – ČO SME SA NAUČILI

1 AMD GPU prešla dlhou cestou od ATI (1985) cez akvizíciu (2006) po dnešné RDNA 3 a Instinct MI300X.

2 Architektúra RDNA prináša dramatické zlepšenia výkonu/watt oproti GCN – 5× efektivita za 10 rokov.

3 Compute Unit je základná jednotka GPU – obsahuje SIMD, TMU, cache, register file a scheduler.

4 Pamäťová hierarchia (Register → L0 → L1 → L2 → Infinity Cache → VRAM) je kľúčom k GPU výkonu.

5 Softvérový stack (ROCm, DirectX, Vulkan, FSR) rovnako dôležitý ako samotný HW – bez softvéru nič nefunguje.

6 AMD dominuje konzolový trh (Xbox + PS5), bojuje o AI datacenter (MI300X vs H100), rastie v NPU (XDNA).

ZDROJE A LITERATÚRA

[1] AMD Inc. (2023). RDNA 3 Architecture Whitepaper. <https://www.amd.com/en/technologies/rdna-3>

[2] AMD Inc. (2023). AMD Instinct MI300X Product Brief. <https://www.amd.com/en/products/accelerators/instinct/mi300/mi300x>

[3] Jon Peddie Research (2023). GPU Market Share Q3 2023 Report.

[4] AnandTech – Ian Cutress: Inside RDNA 3 Architecture (2022)

[5] Hardware Unboxed (2023/2024). RX 7900 XTX Review & Benchmarks. YouTube.

[6] Digital Foundry (2023). AMD FSR 2 vs DLSS 3 Deep Dive Analysis.

[7] JEDEC Solid State Technology Association (2023). HBM3 Specification JESD238A.

[8] Khronos Group (2023). Vulkan 1.3 Specification. <https://registry.khronos.org/vulkan/>

[9] Microsoft (2023). DirectX 12 Ultimate Feature Guide. Microsoft Learn.

[10] Sony Interactive Entertainment (2020). PS5 System Architecture – Mark Cerny GDC Talk.

[11] AMD ROCm Documentation (2024). <https://rocm.docs.amd.com/>

[12] TechPowerUp GPU Database (2024). <https://www.techpowerup.com/gpu-specs/>

AMD

Grafický
Subsystem

Ďakujem za pozornosť!

Otázky & Diskusia

Michal Rýzek

Počítačové Architektúry · Univerzita Cyrila a Metoda · 2024/2025

RDNA 3

MI300X

XDNA NPU

FSR 4

CDNA 3