BOMBARDINO CROCODILO GAMING

mridul noorain parth

WELCOME TO:

APPLICATION PROCESSORS

Graphics, Gaming and Networking

A NEW ERA IN PROCESSING

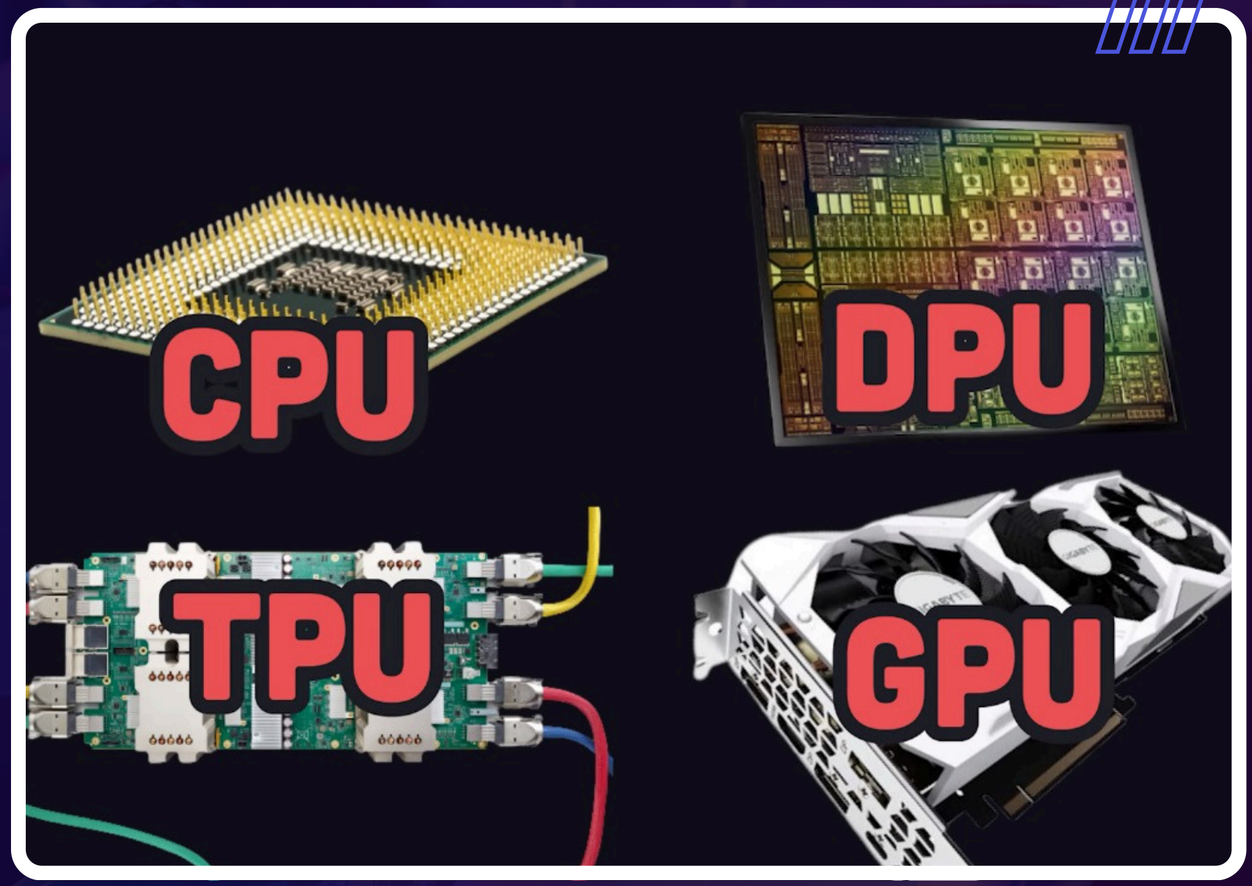
PROF. SHALABH GUPTA

m n p

PROCESSING UNITS:

1) CPU - CENTRAL PROCESSING UNIT
2) GPU - GRAPHICS PROCESSING UNIT
3) DPU - DATA PROCESSING UNIT
4) NPU - NETWORK PROCESSING UNIT
5) QPU - QUANTUM PROCESSING UNIT
6) TPU - TENSOR PROCESSING UNIT

and many more ...

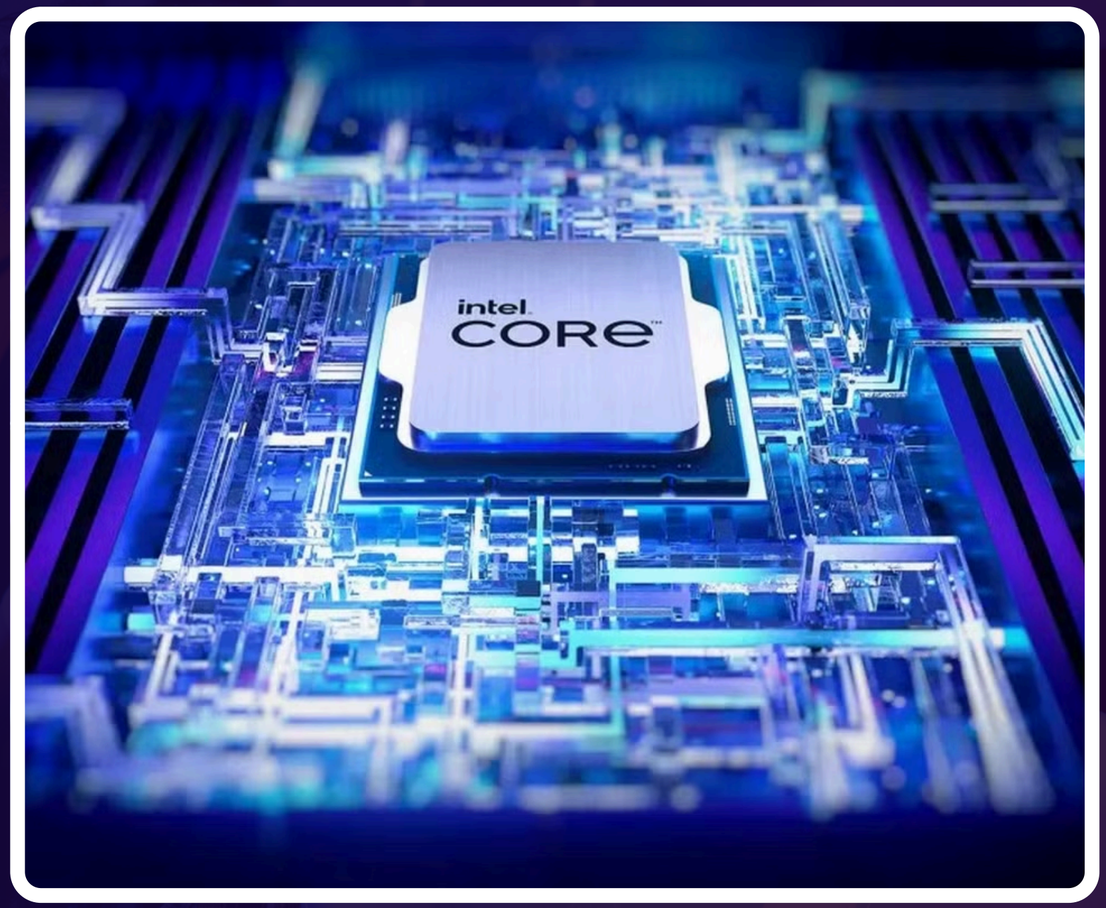


m n p

CPU

What is CPU?

A CPU (Central Processing Unit) is the primary component of a computer that carries out instructions from programs. It processes data, performs calculations, and manages the flow of information through the system.



Key components of CPU:

1. ALU
2. CU
3. Registers
4. Cache

How a CPU works?

1. **Fetch** – Get the next instruction from memory (RAM).
2. Decode – Understand what the instruction means.
3. Execute – Perform the operation (math, move data, etc.).

GPU

m n p

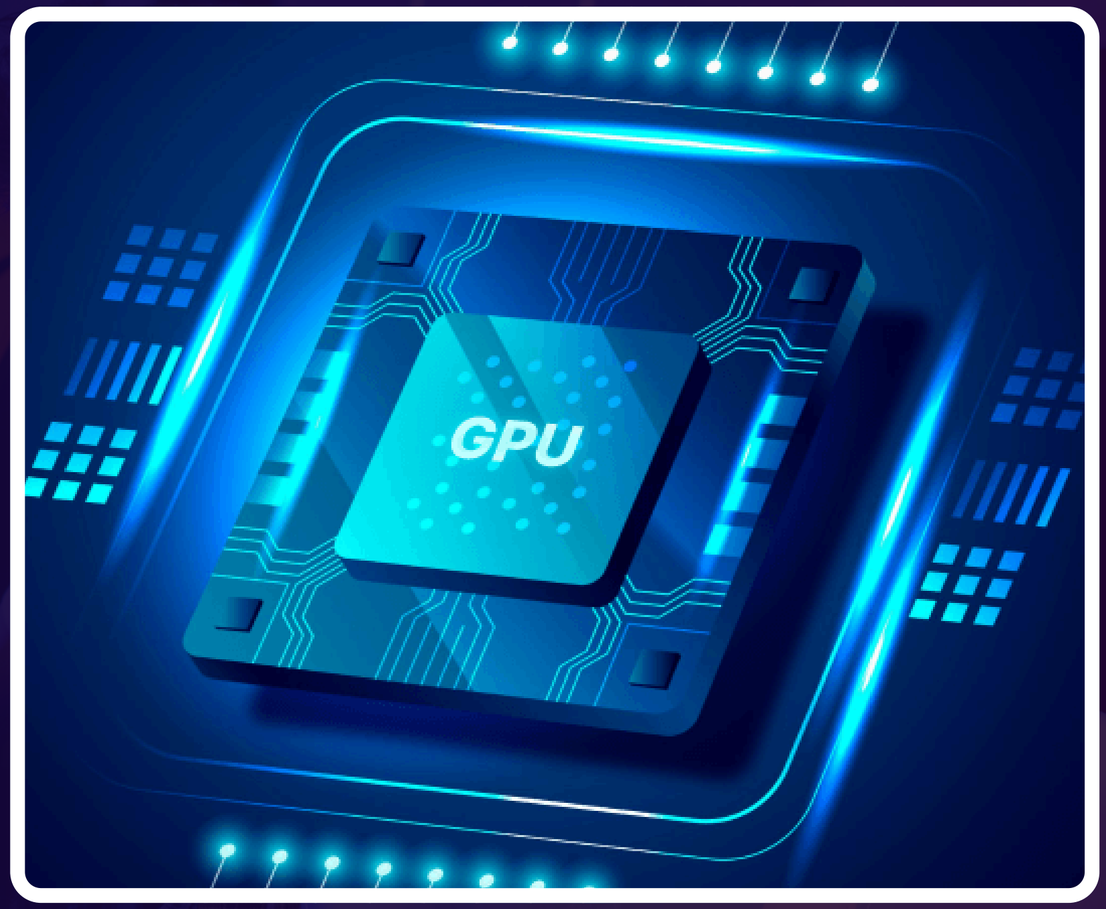
What is GPU?

A GPU (Graphics Processing Unit) is a specialized processor designed to handle graphics rendering, parallel computations, and visual output tasks.

How a GPU works?

1. Gets the instruction from CPU.
2. Breakdown of bigger task into smaller tasks which are performed parallely(SIMD:Single Instruction Multiple Data).
3. Graphics Pipeline:
   - **Vertex Processing:**Transforms 3D coordinates into 2D screen positions.
   - **Rasterization:** Converts shapes into pixels.
   - **Fragment/Pixel Shading**: Calculates the final color of each pixel using textures, lights, shadows, etc.
4. Reads/Writes to VRAM(Video RAM).

VRAM is high bandwidth and optimized for fast data transfer between memory and cores.



CPU VS GPU

Parameter	CPU	GPU
Function	Generalized component that deals with the main processing functions of a computer.	Specilized component that is great for parallel computing.
Processing	Runs Processes Serially.	Runs Processes Parallely.
Design	Fewer but more powerful cores.	More cores(less powerful than CPU cores)
Emphasis	Low Latency	High throughput
Use Case	General Purpose Computing Devices	High-Performance Computing Devices

m n p

PROCESSORS FOR GAMING

HOW ARE THEY DIFFERENT?

- High Clock Frequencies.
- Higher number of Cores and Threads.
- Larger Cache Sizes (L1 L2 & L3)
- Higher TDP
- Overclocking support

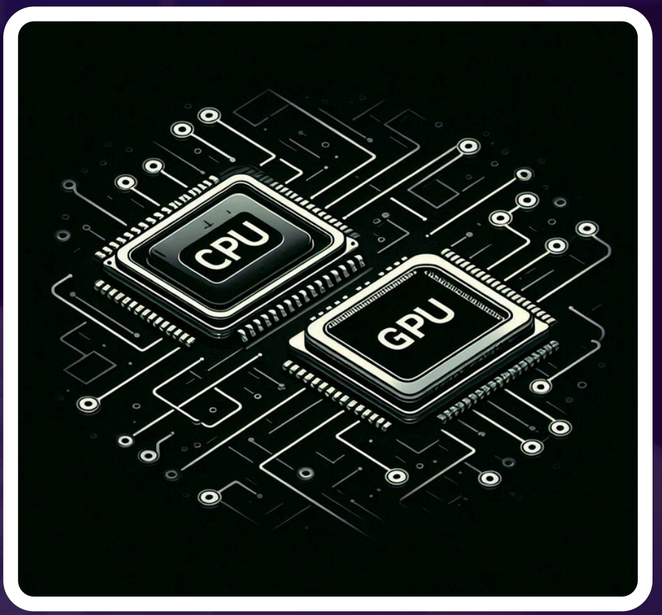
Feature	Gaming CPU	General Logic CPU/Processor
Clock Speed (GHz)	Higher clock speeds (4–6 GHz) for faster single-threaded performance	Often balanced for power efficiency and multitasking
Core & Thread Count	Medium-high core counts (6–16 cores), but with strong single-core performance	May have fewer cores (4–8) for desktop, or very many (32–128) for servers
Cache Size	Large L3 cache to quickly access game data and reduce memory bottlenecks	Cache size optimized based on task type — may not prioritize large caches

m n p

HOW DO THEY WORK

CPU 🤝 GPU

- Distributes Game Tasks Across Multiple Cores:
  - The CPU divides tasks like game logic, physics, AI etc using multi-threading, so they can run simultaneously.
- Uses Multi-Level Cache for Fast Data Access:
  - The CPU uses L1, L2, and L3 caches to store frequently used game data — like player positions, object coordinates, AI states, and recent physics calculations for quick access.
- Sends Draw Calls to the GPU for Rendering:
  - After handling game logic and updating scene data, the CPU sends draw calls (rendering instructions) to the GPU, telling it what to display on screen, including positions, textures, effects, and animations.

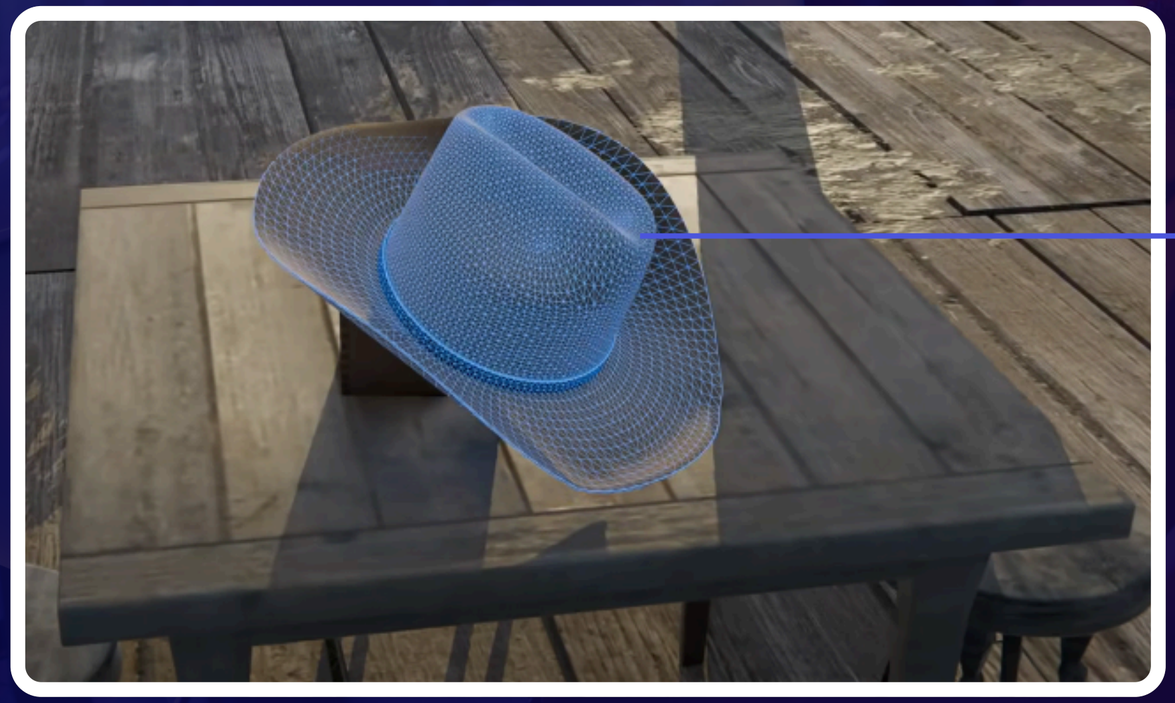


AN EXAMPLE

m n p

AN EXAMPLE

m n p

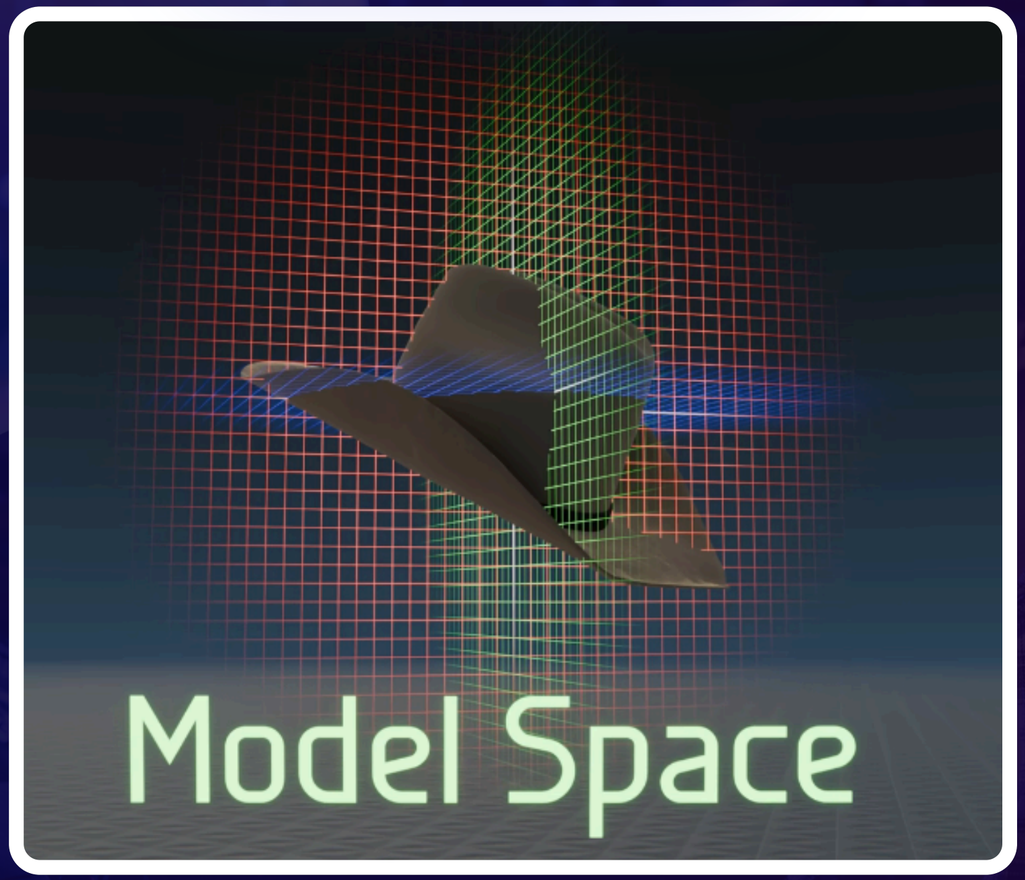


28,000 TRIANGLES

14,000 VERTICES

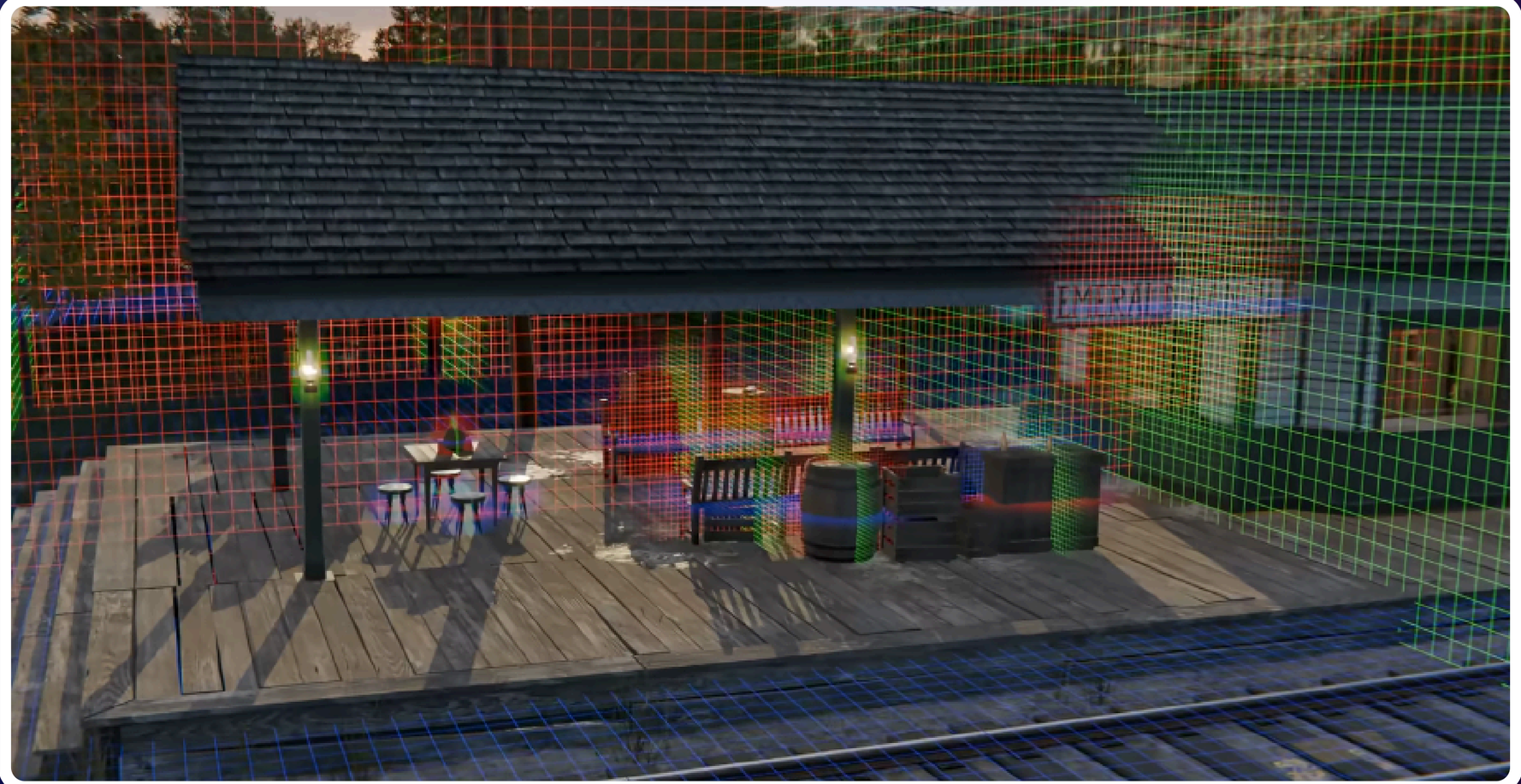
AN EXAMPLE

m n p



AN EXAMPLE

m n p



m n p

THE NETWORK PROCESSING UNIT

A Network Processing Unit (NPU) is a specialized processor designed specifically for network packet processing. Unlike general-purpose CPUs, NPUs are multi-core processors optimized for high-throughput, low-latency tasks. This makes them ideal for applications such as routers, switches, firewalls, and intrusion detection systems, where rapid packet processing is crucial.

Key Applications:

Deep packet inspection (DPI)

Firewall filtering

Encryption/decryption (IPsec, SSL offload)

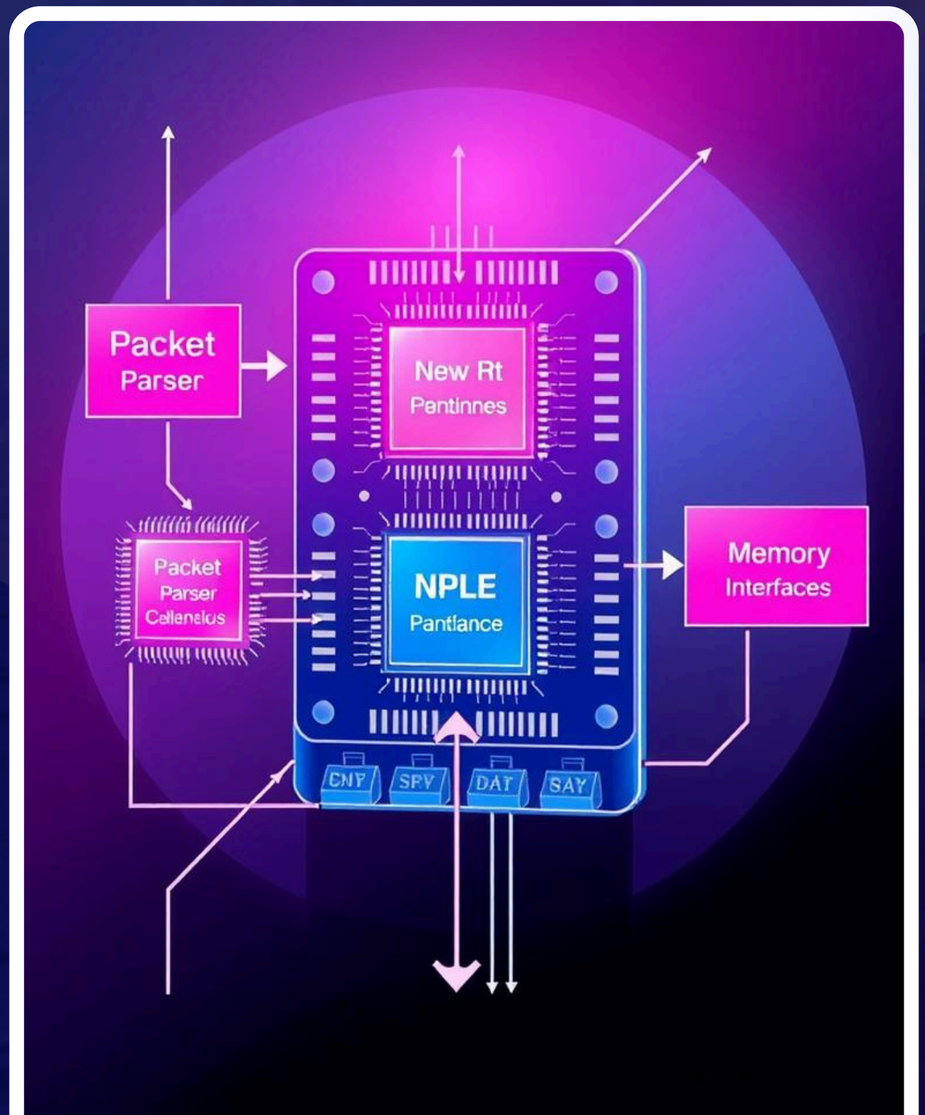
QoS (Quality of Service) enforcement

Load balancing

m n p

NPU INTERNAL ARCHITECTURE

The internal architecture of an NPU is designed for parallel processing. It typically features a multi-core architecture optimized for handling network tasks simultaneously. Key components include a packet parser, classification engine, forwarding engine, and traffic manager.



MULTI-CORE ARCHITECTURE

Optimized for parallel processing of network tasks.

PACKET PARSER

Analyzes incoming packets to extract relevant information.

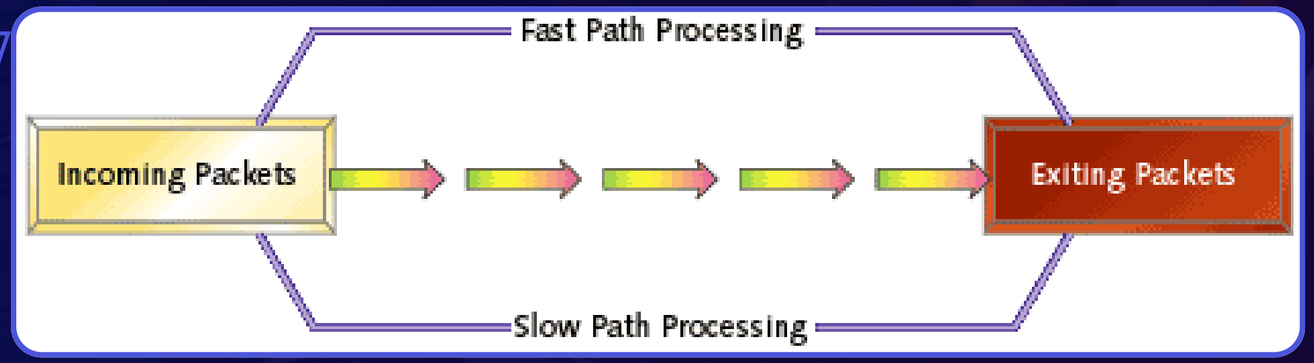
CLASSIFICATION ENGINE

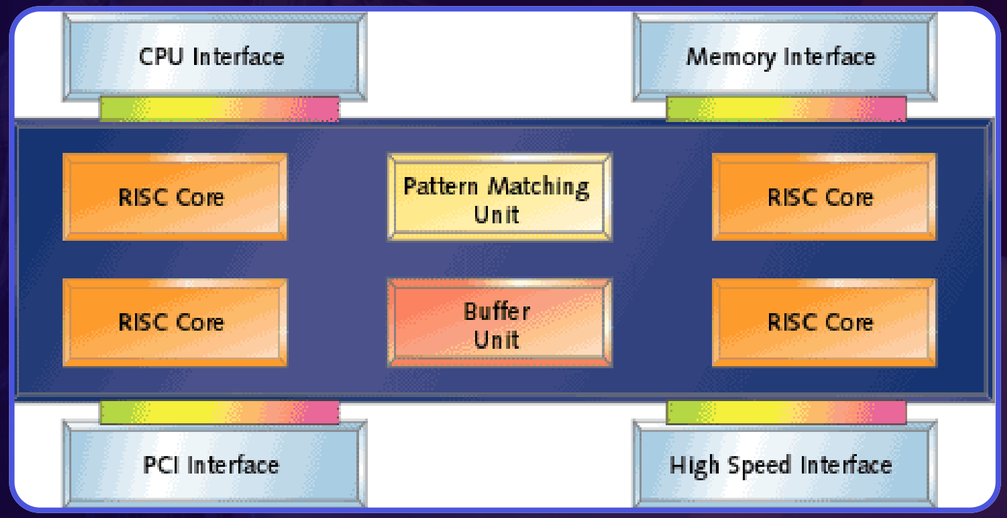
Categorizes packets based on predefined rules and policies.

m n p

NPU OPERATION: A DEEP DIVE

The operation of an NPU involves a systematic flow of packets through various stages. Packets enter the NPU (ingress), undergo parsing to extract relevant information, are classified based on predefined rules, have policies enforced, and finally exit the NPU (egress).





KEY ATTRIBUTES

m n p

PHYSICAL SIZE	Chip-integrated to standard Rack-sized Network Processing Units
Processing Size	High throughput: Keeps up with fast links (10Gbps, 100Gbps, 10 million PPS, Deterministic Performance: Predictable even under load
Architectural Edge	Lookup engines: Performs fast searches (e.g., routing tables) Flow tables / TCAMs: For matching rules (e.g., firewalls, access control) Memory system: Fast-access memory (SRAM, DRAM) for headers, states
Programmable	Broadcom SDKs (for sophisticated vendors.) P4 (a language to describe packet processing) DPDK (Data Plane Development Kit) (used with CPUs but sometimes extends to programmable NPUs)
Use Case	Enterprise and carrier-grade routers, Data center switches, Security appliances (VPNs, firewalls), and 5G infrastructure
Vendors	Broadcom (Trident, Tomahawk, Jericho series) Intel (Tofino via Barefoot Networks, uses P4) Cisco (in-house NPUs like UADP)

m n p

THANK YOU