GPU

GPU stands for **Graphics Processing Unit**. It is a specialized electronic circuit designed to accelerate the creation of images for output to a display device.

^ **CPU** = General-purpose processor that handles most tasks on a computer. It can execute a few complex tasks at a time.

IN SIMPLER TERMS : The teacher, who knows a lot and manages all kinds of tasks.

^ **GPU** = Highly parallelized processor designed for performing thousands of smaller, similar tasks simultaneously, making it ideal for graphics rendering and large-scale computations.

WE CAN ALSO UNDERSTAND LIKE : A group of super-fast artists and builders, helping create beautiful, detailed worlds in video games and movies!

Architecture and Functionality

GPU consists of **hundreds to thousands of small cores** designed for parallel processing.

**1.Shader Cores (CUDA Cores / Stream Processors):** The main computational units that perform calculations for graphics and other tasks(  rendering graphics).

**2.Memory Interface:** GPUs have high-bandwidth memory (like GDDR6 or HBM2) for fast data transfer.

**3.Cache Hierarchy:** Similar to CPUs, GPUs have L1 and L2 caches to speed up memory access.

**4.Rasterization Unit:** Converts vector graphics (polygons) into pixels on the screen.

**5.Ray Tracing Cores (Modern GPUs):** Dedicated hardware for real-time ray tracing, which simulates light and shadows for more realistic images.

How GPUs Process Graphics?

1. **Vertex Processing:** Calculates the position of vertices in 3D space.
2. **Rasterization:** Converts 3D shapes into 2D pixels.
3. **Fragment Shading:** Applies color, texture, and lighting to each pixel.
4. **Output:** The final image is displayed on the screen.

**History and Evolution of GPUs**

**1.Early Days (1970s – 1990s):**

* Early computers had **graphics controllers** for simple 2D rendering.
* **1981:** IBM introduced the **Monochrome Display Adapter (MDA)**, one of the first graphics cards.
* **1987:** IBM released the **VGA (Video Graphics Array)**, which set the standard for early PC graphics.

**2.Rise of the Modern GPU (1990s – 2000s):**

* **1995:** NVIDIA entered the market and introduced **NV1**, its first graphics card.
* **1999:** NVIDIA launched the **GeForce 256**, the first GPU marketed as a "GPU," capable of hardware-based transformation and lighting (T&L). This marked the start of the modern GPU era.
* **2000s:** GPUs evolved rapidly to support 3D rendering, real-time shading, and high-definition video playback.

**3.The Era of GPGPU (2000s – Present):**

* GPUs began to be used for **general-purpose computing (GPGPU)**, performing tasks outside of graphics rendering—such as scientific simulations, deep learning, and data analysis.
* **CUDA (Compute Unified Device Architecture)** by NVIDIA and **OpenCL (Open Computing Language)** were key innovations that enabled developers to harness GPU power for non-graphics tasks.

**Types of GPUs**

**1. Integrated GPUs**

* Built into the same chip as the CPU.
* Found in laptops and budget PCs.
* Less powerful but more energy-efficient.
* Example: **Intel Iris Xe, AMD Radeon Vega**.

**2. Discrete GPUs**

* Separate hardware installed on a graphics card.
* Offers significantly higher performance.
* Common in gaming PCs, workstations, and high-performance computing.
* Example: **NVIDIA GeForce RTX, AMD Radeon RX**.

**3. External GPUs (eGPUs)**

* External graphics units that connect to a laptop via Thunderbolt.
* Provide desktop-level performance to laptops for gaming and rendering.

**Applications of GPUs**

**1. Gaming**

GPUs are critical for rendering complex 3D environments and enabling real-time physics, lighting, and textures. The introduction of **ray tracing** technology has revolutionized gaming graphics, bringing cinematic-quality visuals to real-time games.

**2. Artificial Intelligence (AI) and Machine Learning (ML)**

* GPUs are highly efficient at processing large amounts of data, making them essential for AI and deep learning.
* They accelerate neural network training and inference. NVIDIA’s **Tensor Cores** are specially designed for AI tasks.

**3. Scientific Research**

* Used in weather prediction, molecular dynamics, astrophysics, and genomics.
* GPU-powered supercomputers perform calculations much faster than traditional CPU-based systems.

**4. Video Production and Rendering**

* GPUs accelerate video editing, 3D animation, and rendering in software like **Adobe Premiere Pro, Blender**, and **Maya**.
* Real-time rendering allows for faster previews and final renders.

**5. Cryptocurrency Mining**

GPUs are used for mining cryptocurrencies like **Ethereum** due to their ability to perform repetitive calculations quickly.

**6. Medical Imaging and Diagnosis**

GPUs help in advanced imaging techniques such as MRI, CT scans, and simulations for surgery planning.

**Major GPU Manufacturers**

1. **NVIDIA:** Leader in gaming, AI, and professional GPUs. Known for innovations like CUDA, RTX, and DLSS.
2. **AMD:** Known for high-performance gaming GPUs and APUs (Accelerated Processing Units). Offers competitive pricing.
3. **Intel:** Recently entered the discrete GPU market with the **Intel Arc series**.

**Future Trends in GPUs**

1. **AI and Deep Learning:** Future GPUs will have even more specialized AI hardware.
2. **Quantum Computing Integration:** GPUs may be used alongside quantum computers for hybrid computational tasks.
3. **Energy Efficiency:** Companies are focusing on improving performance-per-watt to reduce power consumption.
4. **Advanced Ray Tracing:** More realistic visuals in games and movies.
5. **Cloud Gaming and Virtualization:** GPUs will play a key role in cloud gaming services like NVIDIA GeForce NOW and Google Stadia.