Graphics Hardware

The hardware associated with the generation and rendering of graphical images on the display device

Graphics Cards

The hardware components that jointly collaborate to generate graphical display on the display device

Graphics cards can be categorized depending on whether the memory and processing unit used is a part of the motherboard or they are separate

Integrated Graphics card

It is a part of the motherboard and the memory (RAM) used is shared with the central processing unit (CPU) of the system

Graphical rending in these types of systems are slower

External/Dedicated Graphics card

They come with separate cards with their own memory (RAM) and processors

Graphical rendering is much fasters

Motherboard contains **CPU with fewer cores,** whereas a Graphics card is add-in board with powerful **GPU that consists of hundreds of cores** for parallel processing of data

Graphics Processing Unit (GPU)

It is a specifically designed processor to accelerate computer graphics workloads

It is coupled with high speed RAM to store pixel data that are accessed by DAC (Digital to Analogue Converter) to translate the image into analogue signal to render image on the screen

The performance of a GPU is measured in terms of frame rate (FPS)

A human eye perceives smooth animation if it is shown 24 frames per second but GPUs process game data at the rate of more than 60 frames per second to render smooth transitions.

Accelerates rendering of 3D graphics for

- interesting visual effects
- realistic scenes
- advanced lighting
- shadowing techniques.

Now also used to accelerate additional workloads in

- high performance computing (HPC)

- deep learning

GPU Types

- **Integrated**

  Does not come on its own separate card but instead **embedded alongside CPU**

- **Discrete**

  Distinct chip that is mounted on its **own circuit board**, typically attached to a PCI Express slot.

Features:

CPU On-Board Facilities

3D graphics, is a **calculation intensive** environment.

The calculations are usually simple and can be performed in integer math

Graphics coprocessors often rely on the main CPU for performing this basic arithmetic.

Special purpose graphics processing units (GPUs)

Hundreds of processors for **parallelly executing graphics algorithms** more quickly

Workload sufficiency

Processes **thousands of floating point operations per pixel**

Strong market demand

Examples:

NVDIA GeForce 9800 GTX

GTX can render 340 million small triangles per second, almost 11 billion pixels per second

576 billion floating point operations per second

Chipset supports memory transfer at up to 25.6 GB/s

NVDIA GeForce 9800 GTX has 16 cores each with eight float point ALUs

GeForce GTX 1080

NVIDIA CUDA Cores: 2560

Boost Clock: 1733 MHz

Memory Size: 8 GB

GeForce RTX 3090Ti

NVIDIA CUDA Cores: 10752

Boost Clock: 1.86 GHz

Memory Size: 24 GB

Graphics Software

Software or programs associated with the creation and manipulation of 2D and 3D graphical images

Special purpose packages are for non-programmers who can generate images without having to worry about graphics procedures like Photoshop, Blender etc

General programming graphics packages are for programmers who make use of library of graphics functions in programming languages like C, C++, JAVA, Python for generating graphics thru API calls like Vulkan, DirectX, OpenGL etc

Graphics drivers are software written for operating systems for communicating with the hardware device responsible for generating graphics. Like Intel graphics driver for Windows for communicating with Intel graphics card

Working Mechanism of GPU

GPUs work by using a method called parallel processing, where multiple processors handle separate parts of a single task.

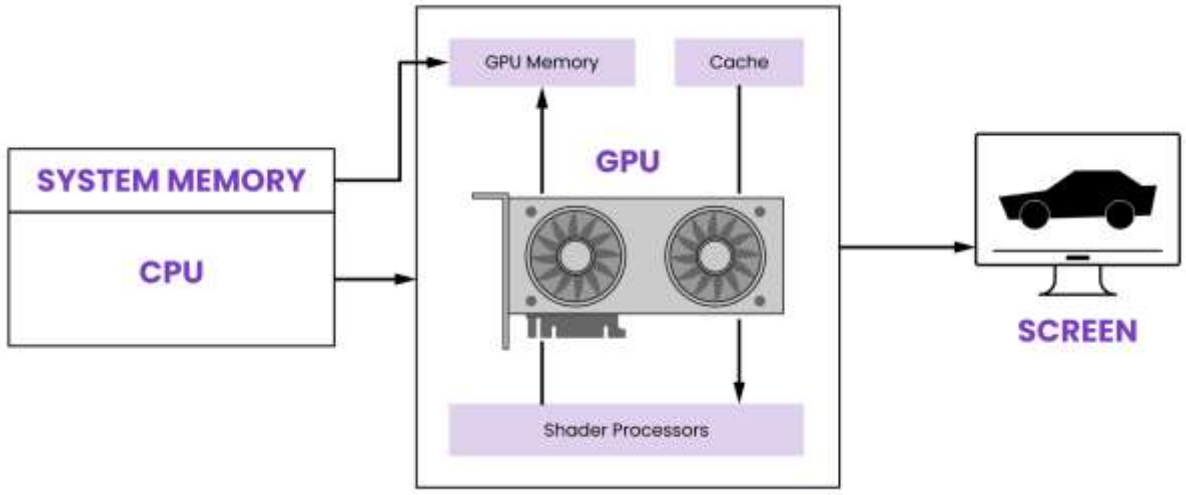
A GPU will also have its own RAM to store the data it is processing.

This RAM is designed specifically to hold the large amounts of information coming into the GPU for highly intensive graphics use cases.

For graphics applications, the CPU sends instructions to the GPU for drawing the graphics content on screen.

The GPU executes the instructions in parallel and at high speeds to display the content on the device

-- a process known as the graphics or rendering pipeline.



Usage of GPU

Accelerating the rendering of real-time 2D and 3D graphics applications.

Video editing and video content creation.

Video game graphics.

Accelerating ML applications such as image recognition and facial detection and recognition.

Training deep learning neural networks.

mine bitcoin and other cryptocurrencies such as Ethereum.

Cloud GPU

A cloud GPU is suitable for companies that require heavy computing power or need to work with machine learning or 3D visualizations.

A cloud GPU is a cloud-based GPU service or virtual GPU that removes the need to deploy a GPU or associated hardware and software on a local device.

Benefits

freeing up local resources,

saving time and cost, and

greater scalability.

Cloud service providers (Google, Amazon, Azure) provide cloud GPUs.

GPUs work by using a method called parallel processing, where multiple processors handle separate parts of a single task.

A GPU will also have its own RAM to store the data it is processing.

This RAM is designed specifically to hold the large amounts of information coming into the GPU for highly intensive graphics use cases.

For graphics applications, the CPU sends instructions to the GPU for drawing the graphics content on screen.

The GPU executes the instructions in parallel and at high speeds to display the content on the device

-- a process known as the graphics or rendering pipeline.

GPU use cases: What GPUs are used for today

Accelerating the rendering of real-time 2D and 3D graphics applications.

Video editing and video content creation.

Video game graphics.

Accelerating ML applications such as image recognition and facial detection and recognition.

Training deep learning neural networks.

mine bitcoin and other cryptocurrencies such as Ethereum.

cloud GPU?

A cloud GPU is suitable for companies that require heavy computing power or need to work with machine learning or 3D visualizations.

A cloud GPU is a cloud-based GPU service or virtual GPU that removes the need to deploy a GPU or associated hardware and software on a local device.

Benefits

freeing up local resources,

saving time and cost, and

greater scalability.

Cloud service providers (Google, Amazon, Azure) provide cloud GPUs.

Google Cloud offers high-performance GPUs for many applications. A range of GPU types are available that are suitable for a variety of workloads, budgets and performance requirements.

GPU vs. CPU

A GPU might be found integrated with a CPU on the same electronic circuit -- on a graphics card or in the motherboard of a PC or server. GPUs and CPUs are fairly similar in construction. However, CPUs are used to respond to and process the basic instructions that drive a computer, while GPUs are designed specifically to quickly render high-resolution images and video.

Essentially, CPUs are responsible for interpreting most of a computer's commands, while GPUs perform more complex mathematical and geometric calculations to focus on graphics rendering and other applications that require intensive calculations.

Both processors are available with different numbers of cores and transistors. The core can be thought of as the processor within the processor. Each core can process its own tasks, or threads. A CPU uses fewer cores and performs tasks sequentially. A GPU, in contrast, might have hundreds or thousands of cores, which allow for parallel processing and lightning-fast graphics output.

A single-core CPU usually does not have the capability for parallel processing, but multicore processors can perform calculations in parallel by combining more than one CPU onto the same chip. GPUs can also contain more transistors than a CPU.

In addition, a CPU has a higher clock speed, meaning it can perform an individual calculation faster than a GPU, so it is often better equipped to handle basic computing tasks