**Graphics processing unit**

Processing units are employed to do calculations in a computing device. With the advent of technology concepts like 3D imagery, High Definition video streaming, Graphics, etc.. are introduced. To implement these concepts on a hardware device large and complex mathematical operations have to be performed and with greater speed.

The central processing unit, although it has high frequency cannot process the calculations of such a large scale effectively. So, a dedicated processing unit for executing larger calculations with high frequency was introduced. This processing unit was called a Graphical Processing Unit. GPU is a specialized electronic device used mainly for calculations based on computer graphics and image processing. These are either embedded on the [SoC](https://www.elprocus.com/difference-between-soc-system-on-chip-single-board-computer/" \t "_blank) along with the microprocessor or the main processor or available as the stand-alone chips with dedicated memory units.

**Computational Functions**

For calculations related to 3D computer graphics, GPU uses the transistors present in its design. The computations around the 3D graphics include geometric operations such as rotation and translation of vertices into different coordinate systems, texture mapping, and rendering polygons. Many recent GPU functions also include the functionality of CPU, oversampling and interpolation techniques for reducing aliasing.

Today there has been a tremendous increase seen in the usage of GPU with the increase in deep learning and machine learning technologies. To train a deep learning model a larger number of complex calculations have to be done. The use of GPU has made the training of machine learning models easier task.

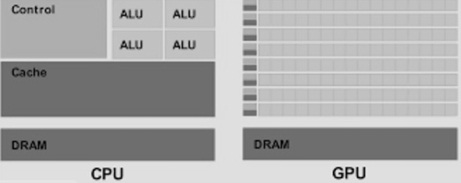
Graphical processing units are found to be 250 times faster than CPU. In GPU accelerated video decoding, the GPU performs the portions of the video decoding process and video post-processing. The commonly used API for this purpose are DxVA, VDPAU, VAAPI, XvMC, XvBA. Here DxVA is for the windows-based operating system and the remaining are for Linux based and Unix like operating systems. XvMC can only decode videos encoded with MPEG-1 and MPEG-2.

The video decoding processes that can be performed by GPU are as follows-

* Motion Compensation
* Inverse Discrete Cosine Transform
* Inverse modified discrete cosine transform.
* In-loop deblocking filter
* Intra frame prediction
* Inverse quantization
* Variable-length Decoding
* Spatial-temporal deinterlacing
* Automatic interlace source detection
* Bitstream processing
* Perfect pixel positioning

**Graphical Processing Unit Architecture**

GPU is usually used as a co-processor along with the CPU. By this, the CPU can perform general-purpose scientific and engineering computing with higher frequency. Here, the time consuming and computation-intensive part of the code is moved onto the GPU whereas the remaining code still works on the CPU. GPU does parallel processing of the code thereby boosting the performance of the system. This type of computing is known as Hybrid Computing.

Graphical Processing Unit Architecture

Unlike CPU that contains two to eight CPU cores, GPU is made up of hundreds of smaller cores. All these cores work together in parallel processing. To effectively use the functions of parallel computing architecture of GPU, application developers at NVIDIA have designed a parallel programming model called ‘CUDA’.

GPU architecture differs based on its model. The general architecture of the GPU consists of Multiple Processing Clusters. These clusters contain multiple Streaming multiprocessors. Here, each of the streaming [multiprocessors](https://www.elprocus.com/wp-content/uploads/2014/12/multi-core-processors.pdf) contains a layer of layer-1 instruction cache along with its associated cores.

**GPU Forms**

Based on their functionality and processing methods there are different forms of GPU available in the market. There are two main forms of GPUin personal computers – dedicated graphics card, integrated graphics. The dedicated graphics card is also known as Discrete GPU. Integrated graphics is also known as Unified memory architecture, shared graphics solutions.

Most of the GPU are designed considering their application like for 3D graphics processing, gaming, etc..GeForceGTX is specifically designed for gaming, Nvidia Titan is designed for cloud computing, Nvidia Quadro is designed for workstation and 3D animations, Nvidia Tesla designed for cloud workstation and artificial intelligence training, Nvidia Drive PX designed for the automated car, etc…

**Dedicated Graphics Card**

Systems with dedicated GPU are known as ‘DIS Systems”. Here the dedicated refers to the fact that these GPU chips have a dedicated [RAM](https://www.elprocus.com/ram-memory-organization-types/) used exclusively by the card. These are usually interfaced with the motherboard using expansion slots such as PCI Express or Accelerated Graphics Port. These chips are easily replaced or upgraded. Due to size and weight constraints dedicated GPU on the portable computers are interfaced through a non-standard slot.

**Integrated Graphics Processing Unit**

This type of GPU does not have a dedicated RAM unit. Instead, it uses a portion of computer memory for its operation. This GPU can be integrated onto the motherboard either as part of its chipset or built on the same die with CPU. These have lesser capacity than the dedicated graphics card but are less costlier to implement. Intel HD Graphics and AMD Accelerated processing Unit are the examples of this GPU.

**Hybrid Graphics Processing**

The functionality of this GPU lies between the dedicated graphics card and the integrated graphics card. This uses a portion of system memory and also has a small dedicated memory cache. This dedicated cache makes up for the high latency of the RAM. ATI’s hyper memory and Nvidia’s TurboCache are the commonly used Hybrid Graphics Processing Units.

**Stream Processing and General Processing GPU’s**

These are popularly called as GPGPU’s. General-purpose graphics processing unit is commonly used as the modified stream processor to perform computer kernels. Using this concept the massive computational power of the modern graphics accelerator’s shader is used as the general-purpose computing power. For massive vector operations, this method gives higher performance than a simple CPU.

**External GPU**

Similar to a large external hard drive, this graphic processing unit is also present at the outside of the computer unit. These are also externally connected to laptop computers. Laptops usually have a good amount of RAM and a sufficiently powerful CPU. Instead of a powerful graphics processor laptops are embedded with a less powerful but more energy-efficient onboard graphics chip. These are not powerful enough to perform game graphics and does not support higher graphics games. So, this External GPU is used with laptops for higher performances.

With the increasing demand for high graphics and good image resolutions, the demand for more powerful GPUs is also increasing. With the availability of powerful GPU, a lot more can be achieved in the field of high processing technologies such as machine learning and deep learning. GPU has also accelerated a tremendous boom in the gaming industry. Many high graphic games have been launched that fully utilize the power of GPU. Which type of GPU can be attached externally to laptops?

Top of Form

Bottom of Form