The ABCs of GPU

1. Why is GPU so popular?
2. What is GPU
3. History of GPU
4. Linux Graphic
5 . Cutting Edge Techs

Never heard of GPU?

- You may not hear GPU in 2018, but you definity heard of below:

  Bitcoin

Never heard of GPU?

- You may not hear GPU in 2018, but you definity heard of below:

  Bitcoin

  AI/Big-Data/Deep-learning

Never heard of GPU?

You may not hear GPU in 2018, but you definitely heard of below:

Bitcoin

AI/Big-Data/Deep-learning

Player Unknown's Battlegrounds

Never heard of GPU?

One thing in common:

A very good GPU cards!

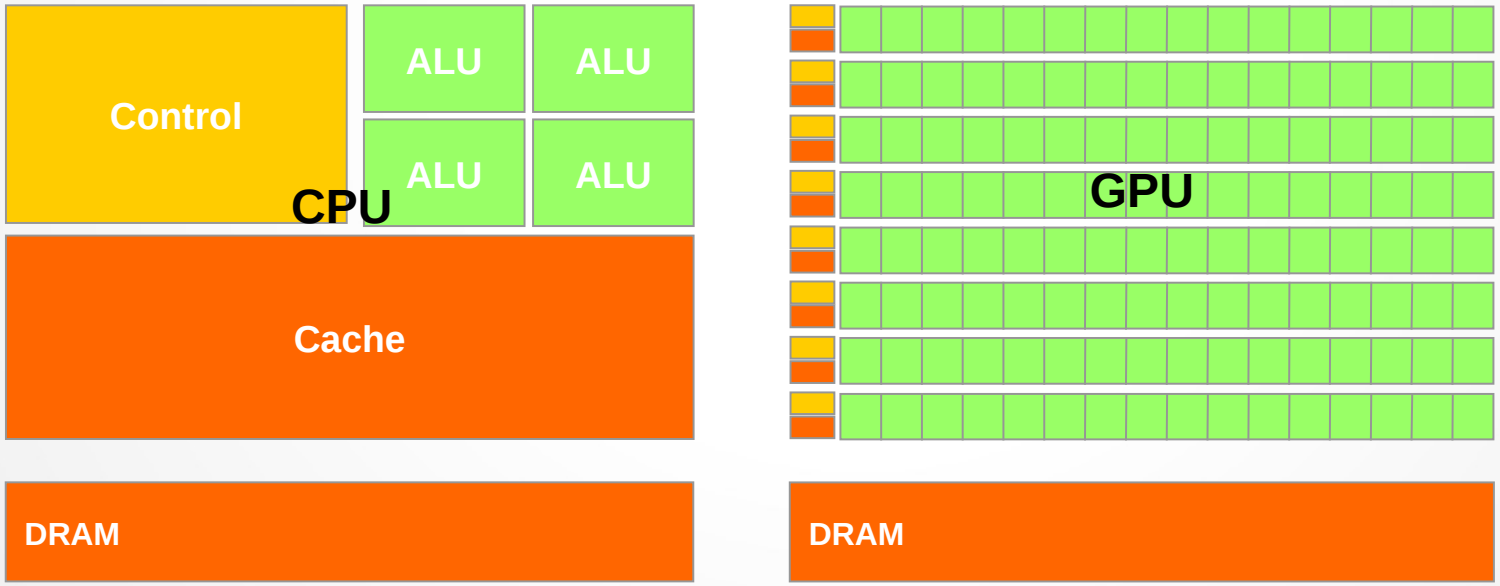
Dell Price: $19,508.16 Amazon Entry level price: $8399 HP Price: $18,782.75

What is GPU

GPU: graphics processing unit
a RISC specialized processor

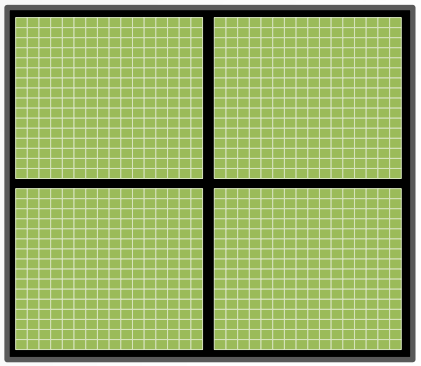
What does it do?
1. offloads 3D graphics rendering from microprocessor
2. General Purpose: no graphic outpu

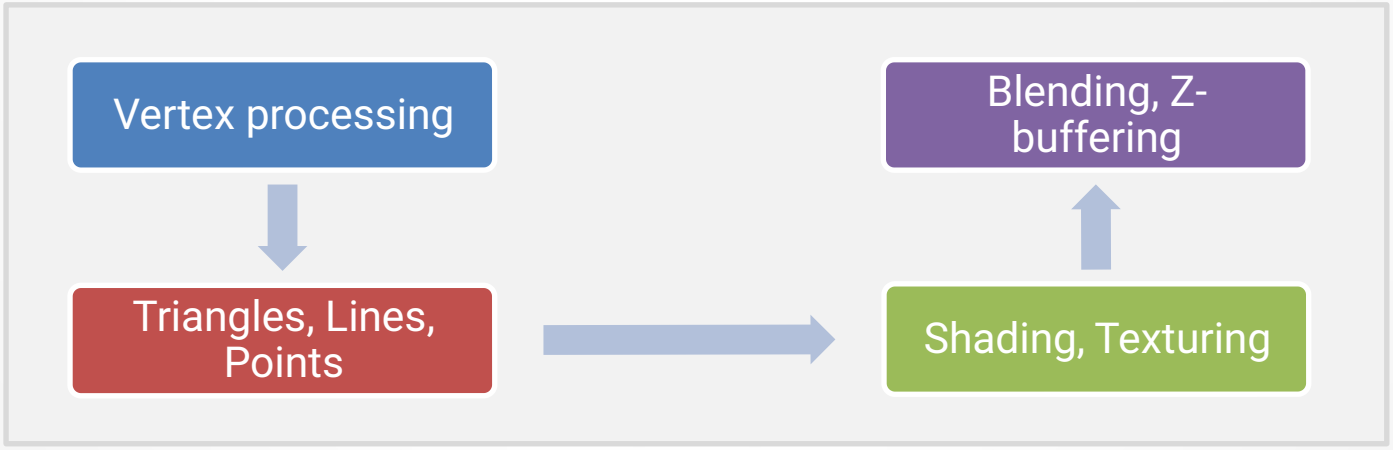
Difference with CPU?



What GPU does

Traditional GPU workflow





GPU inside

SM 84

64 CUDA SM

8 Tensor

CUDA cores: 5376

Tensor cores: 672

GPU Evolution

Console

VGA-2D
Trident
S3
Matrox

GPU Evolution

VGA-3D:

3dfx
3dlabs
PowerVR
Nvidia GForce
ATI Radeon

GPU:
Nvidia
AMD
Intel

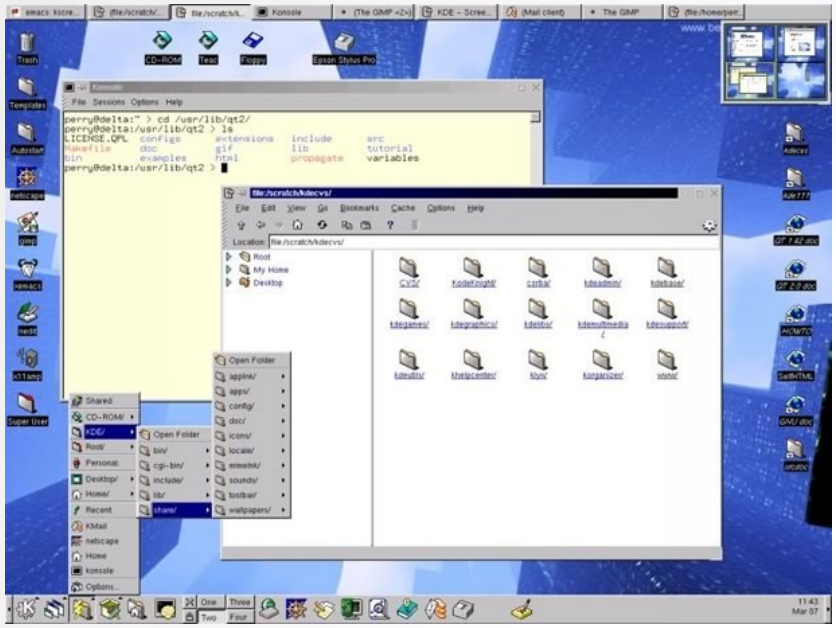
GPU Evolution

- 1980’s – No GPU. PC used VGA controller
- 1990’s – Add more function into VGA controller
- 1997 – 3D acceleration functions:
  Hardware for triangle setup and rasterization
  Texture mapping
  Shading
- 2000 – A single chip graphics processor (beginning of GPU term)
- 2005 – Massively parallel programmable processors
- 2007 – CUDA (Compute Unified Device Architecture)

Graphic in Operating System

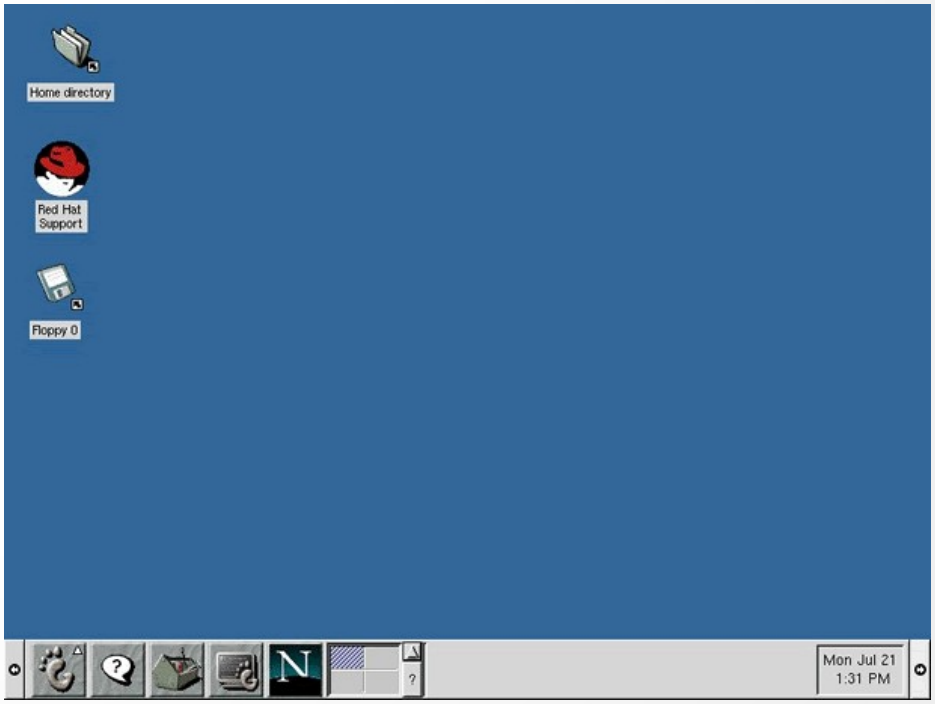
Graphic in Operating System

Graphic in Linux



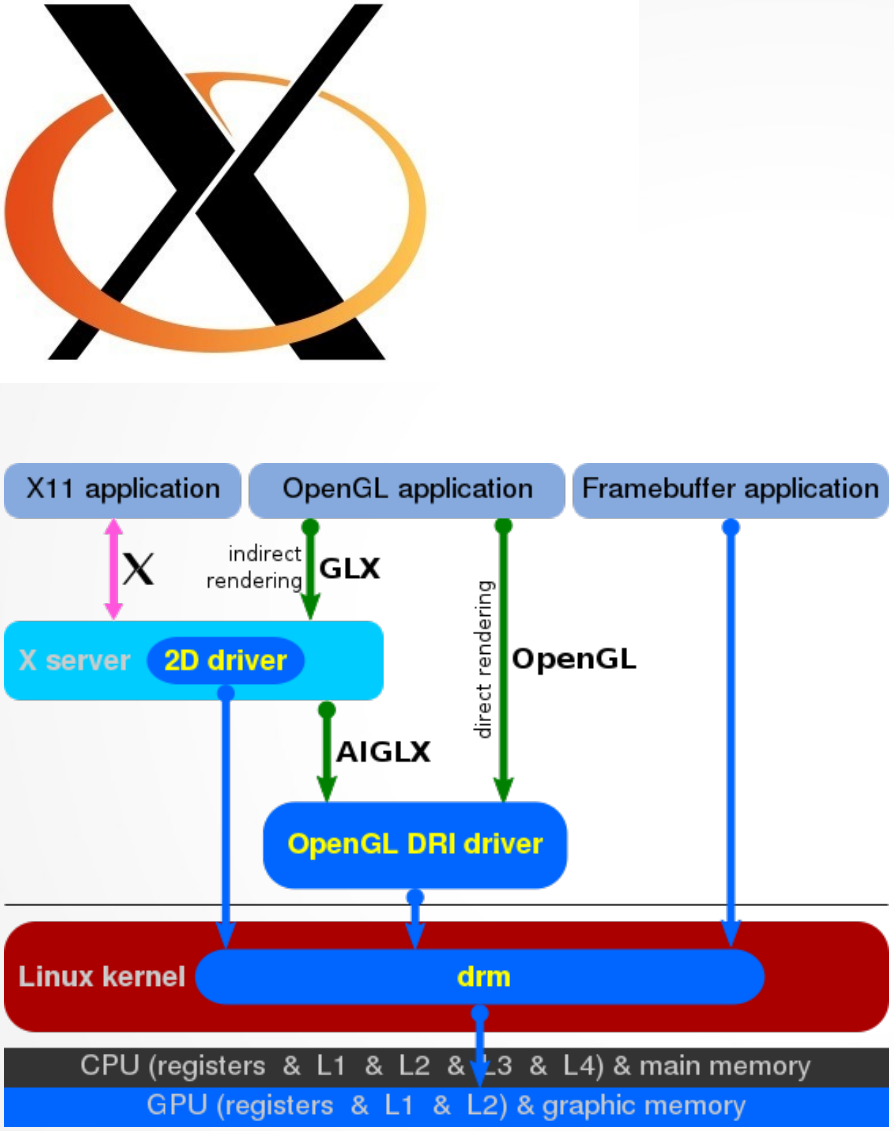
Linux 1.0 (released in 1994)

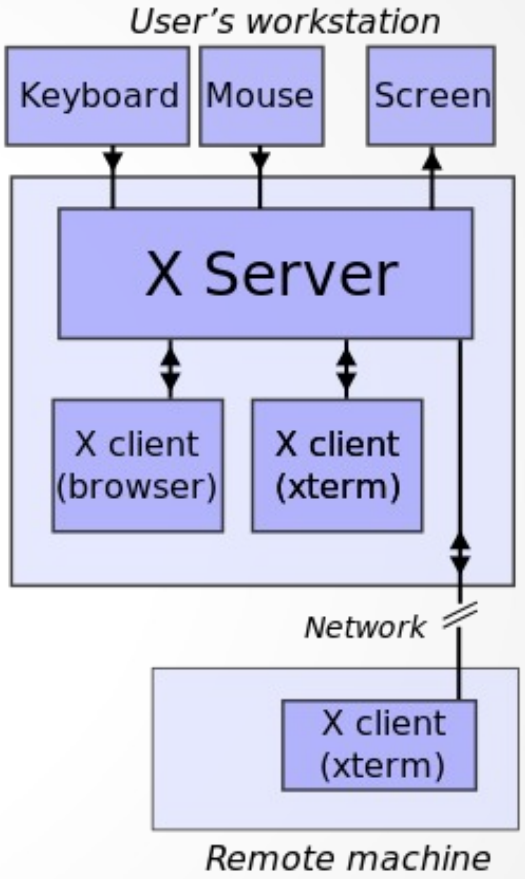
KDE 1.0 (released in 1998)



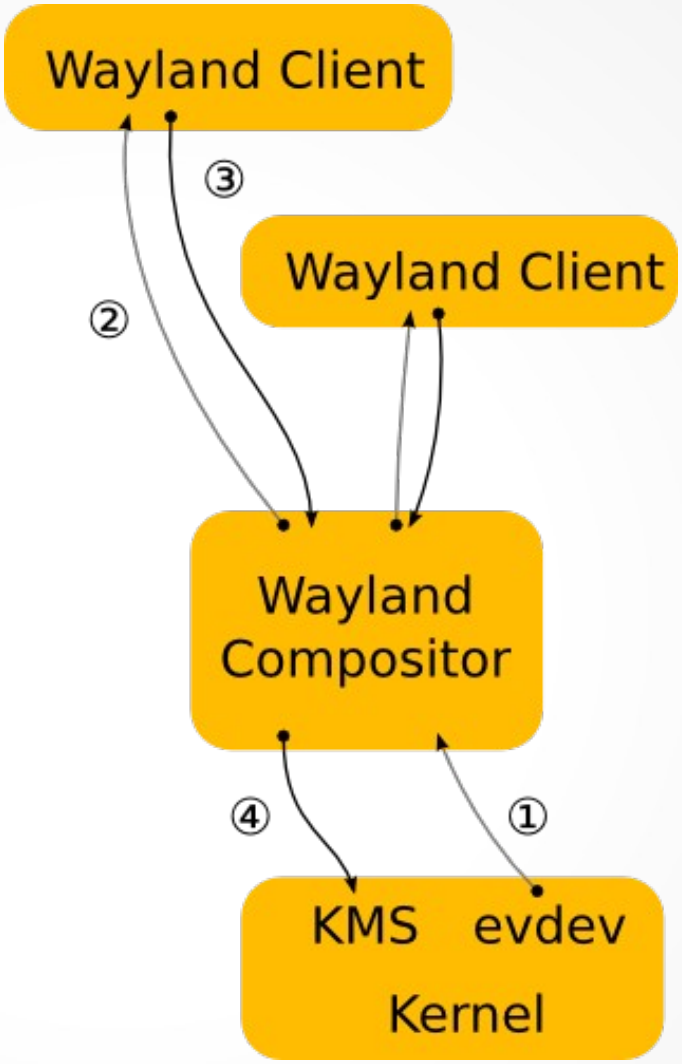
GNOME 1.0 (released in 1999)

Xorg in Linux





Wayland in Linux

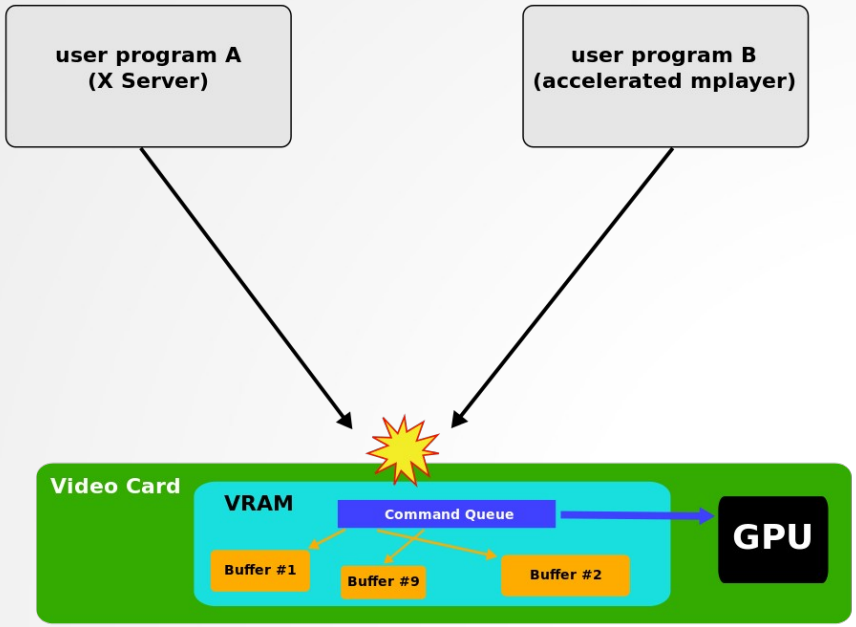


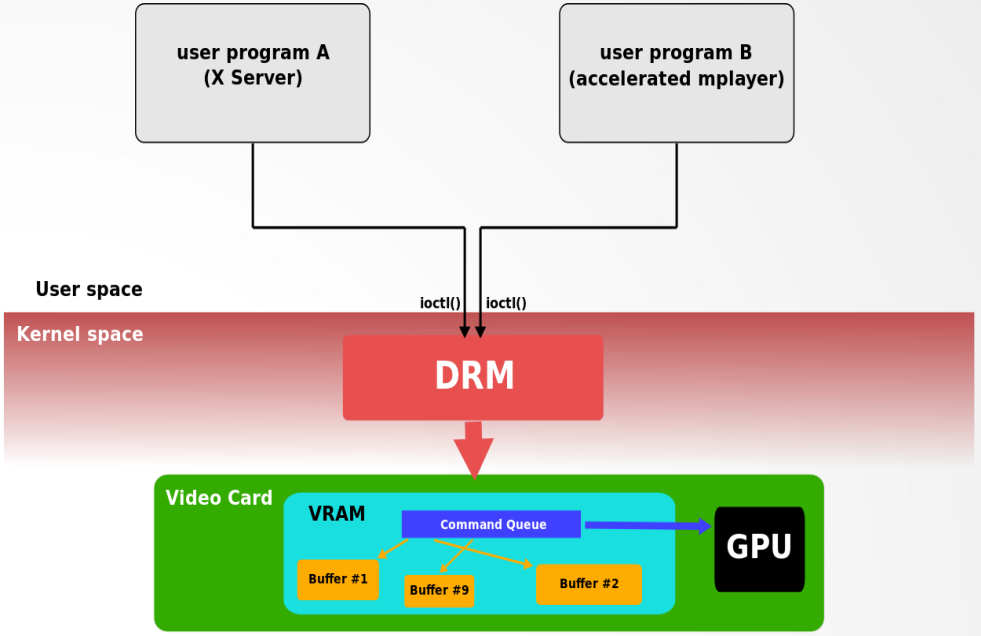
Development in Linux

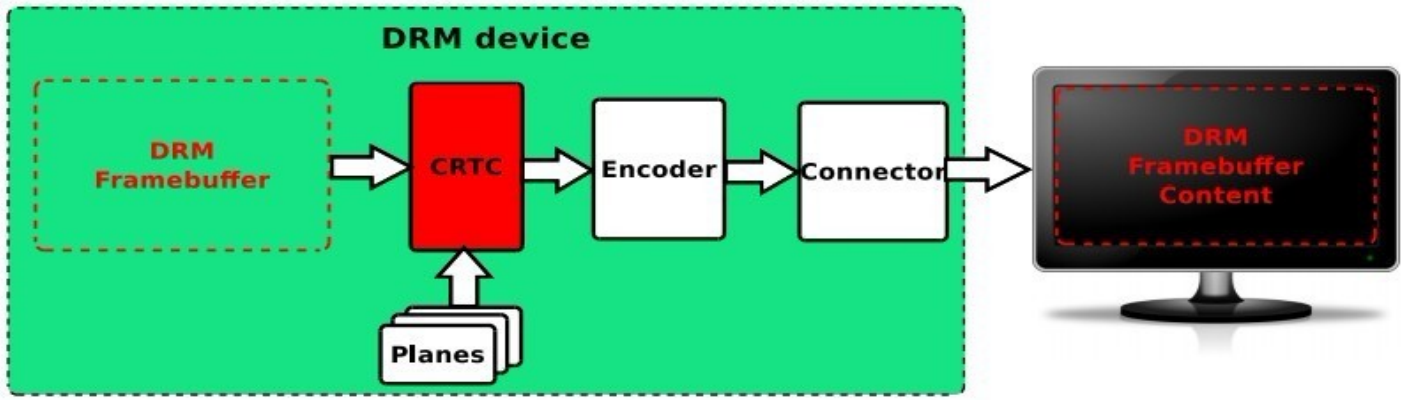
1. Userspace application
Xorg, performance is really bad, blank

2. DRM/ DRI , Direct Render Manager/Infrastructure
It is like frontend of GPU driver, will provide GPU memory management, TTM/GEM
KMS: kernel module setting (4.2)

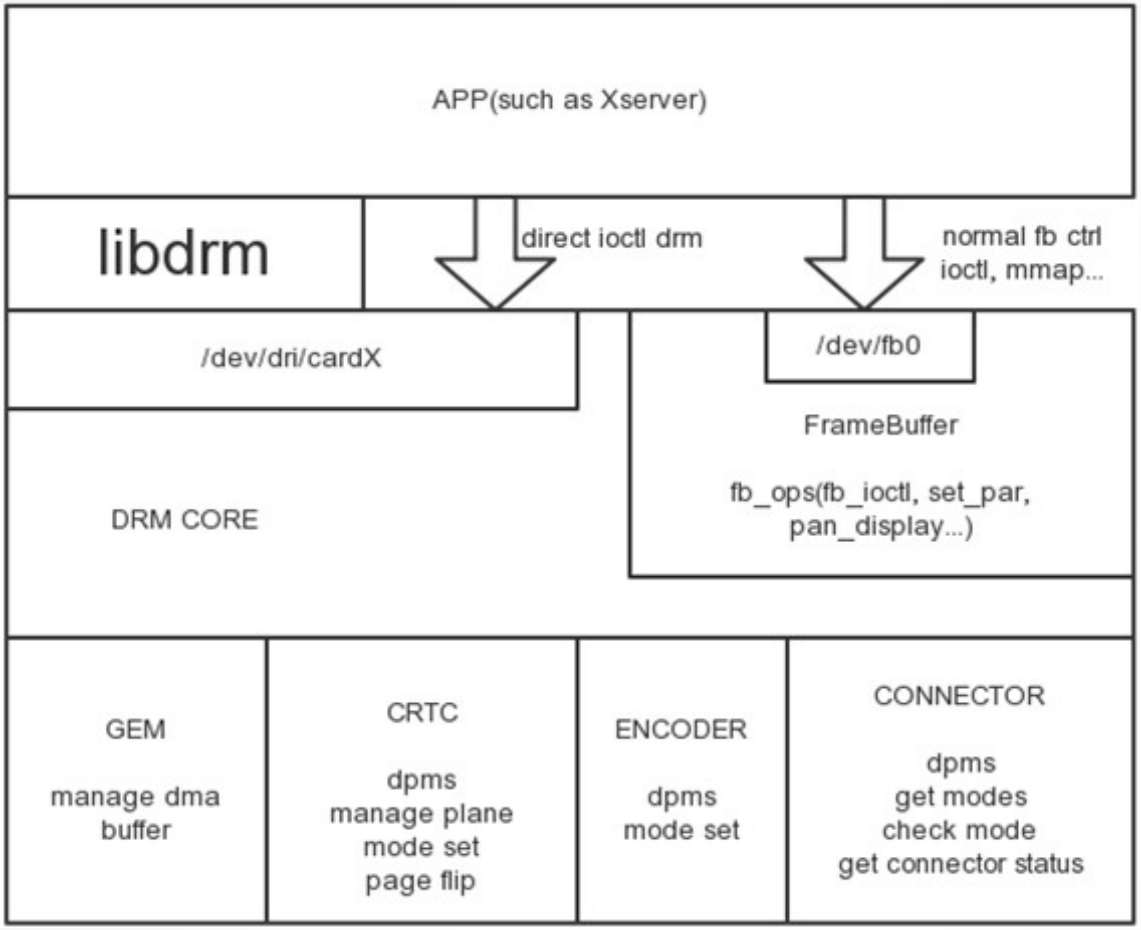
Development in Linux







DRM architecture



Development in Linux

1. Userspace application
Xorg, performance is really bad, blank

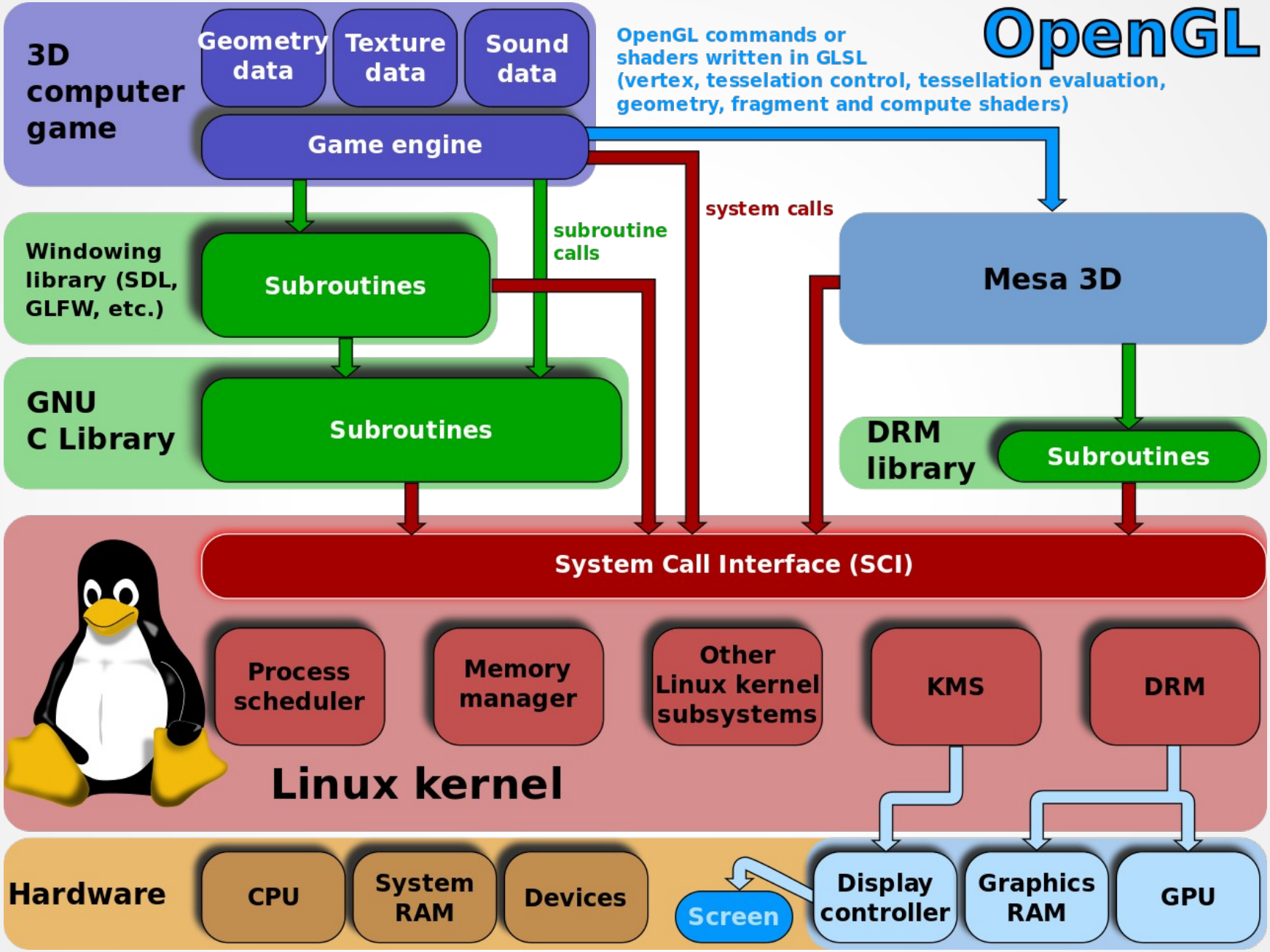
2. DRM/ DRI ， Direct Render Manager/Infrastructure
It is like frontend of GPU driver, will provide GPU memory management, TTM/GEM
KMS: kernel module setting (4.2)

3. Hardware Acceleration

The idea is driving will provide some APIs for application and let applications access hardware rather than cpu directly

Mesa(openGL)

Development in Linux



Cutting Edge Techs

General compute:

heterogeneous architecture
cpu+gpu
cpu+fpga
cpu+tpu

OpenCL
CUDA

Cut-Edge Techs

GPU virtualization

GPU is so expensive, and there are so many cores on it, how could I use it efficiently?

GPU virtualization is Your Answer!

Intel:	MDEV	software implement
AMD:	SRIOV	hardware implement
Nvidia:	MDEV	software implement

GPU virtualization

Physical GPU shared among multiple virtual machines

Great performance and suitable for different workload

Full API compatibility – GPU vendor driver inside the virtual machine

Full device visibility to the hypervisor/host -

allows for device-specific features such as dynamically monitoring and tuning performance, detailed error reporting, etc.

Thanks!

Question?

Reference

https://zhuanlan.zhihu.com/p/30797693
https://www.nvidia.com/en-us/data-center/tesla-v100/
https://www.anandtech.com/show/11367/nvidia-volta-unveiled-gv100-gpu-and-tesla-v100-accelerator-announced
http://www.178linux.com/2233
https://en.wikipedia.org/wiki/Direct_Rendering_Manager
https://wayland.freedesktop.org/architecture.html
https://www.x.org/wiki/
https://blog.csdn.net/prike/article/details/79063867