

PROGRAMMING GPU

An introduction.

Bartosz Boczula
Intel Technology Poland

AGENDA

1. HARDWARE
2. DRIVER
3. DIRECTX
4. GPA Live Session

WHAT IS CPU?

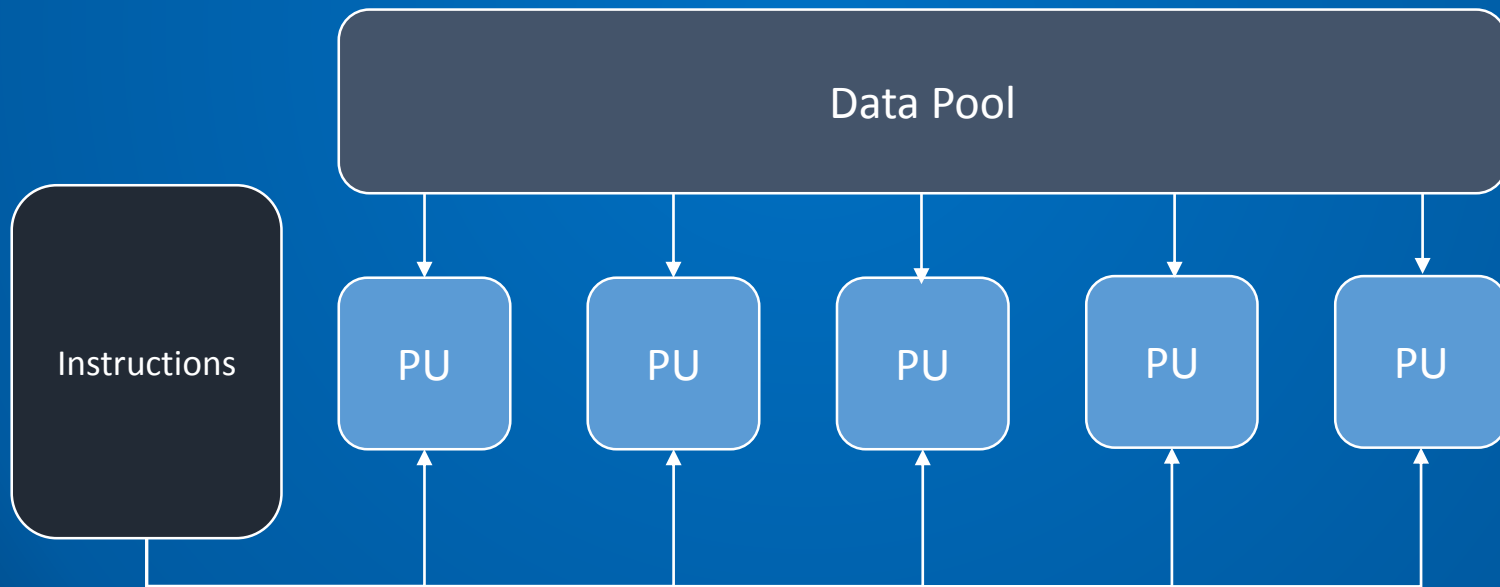
- A processing unit
- Used to executes a computer program
- Wide instruction set
- High frequency, not too much cores
- Big cache, branch prediction

WHAT IS GPU?

- A processing unit, like CPU
- Also executes programs, called shaders
- Optimized for 3D graphics computation
- Implements number of graphics primitive operations
- Either on separate card or integrated with the CPU
- Highly parallel structure
- Efficient for algorithms where the processing of large block of data is done in parallel

WHAT IS SIMD?

- Architecture class
- Multiple processing units that perform the same operation on multiple data points simultaneously



GPU VS. CPU





CPUs composed of few cores and are really good in making single threads go really fast

CPUs tend to skip all over the place during execution

GPUs are composed of hundreds of cores and are really good in making thousands of threads in aggregate to go really fast

GPUs are optimized for taking huge batches of data and performing the same simple linear operation very quickly

GPU VS. CPU

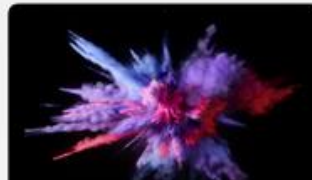
Device	CPU	GPU
 <p>ZenPhone 2</p>	?	?
 <p>PlayStation 4</p>	?	?
 <p>MacBook Pro</p>	?	?
 <p>Infinity LC-900</p>	?	?

WHAT IS SOC?



MSI X99A Tomahawk: militarna płyta pod Intel **Broadwell-E**
benchmark.pl - 1 sie 2016
X99A Tomaha
bogaty zestaw

Intel **Broadw**



Microsoft **Surface Pro 5** To Feature **Skylake** Chipset Instead Of Kaby ...
Mobile & Apps - 21 lis 2016

Rumors are rife that the Microsoft **Surface Pro 5** will be unveiled next year. Some reports suggested that the highly anticipated device will be ...

Microsoft **Surface Pro 5** News, Updates & Specs: New Tablet To Be ...
Gamenguide - 22 lis 2016

Microsoft **Surface Pro 4** Black Friday 2016: Best Deals And Bundles ...
Headlines & Global News - 25 lis 2016

Microsoft **Surface Phone** & **Surface Pro 5** may be released together ...
International Business Times AU - 24 lis 2016

Microsoft **Surface Pro 5** Vs. Microsoft **Surface Book i7**: Does Intel ...
Counsel & Heal - 21 lis 2016

Wyświetl wszystkie

Wyświetl wsz

Two Thunderbolt 3 ports

Buy >

with Touch Bar
Touch ID

al-core Intel Core i5 processor

st up to 3.3GHz

e-based SSD

BMHz memory

raphics 550

h trackpad

Four Thunderbolt 3 ports

Buy >

WHAT IS SOC?

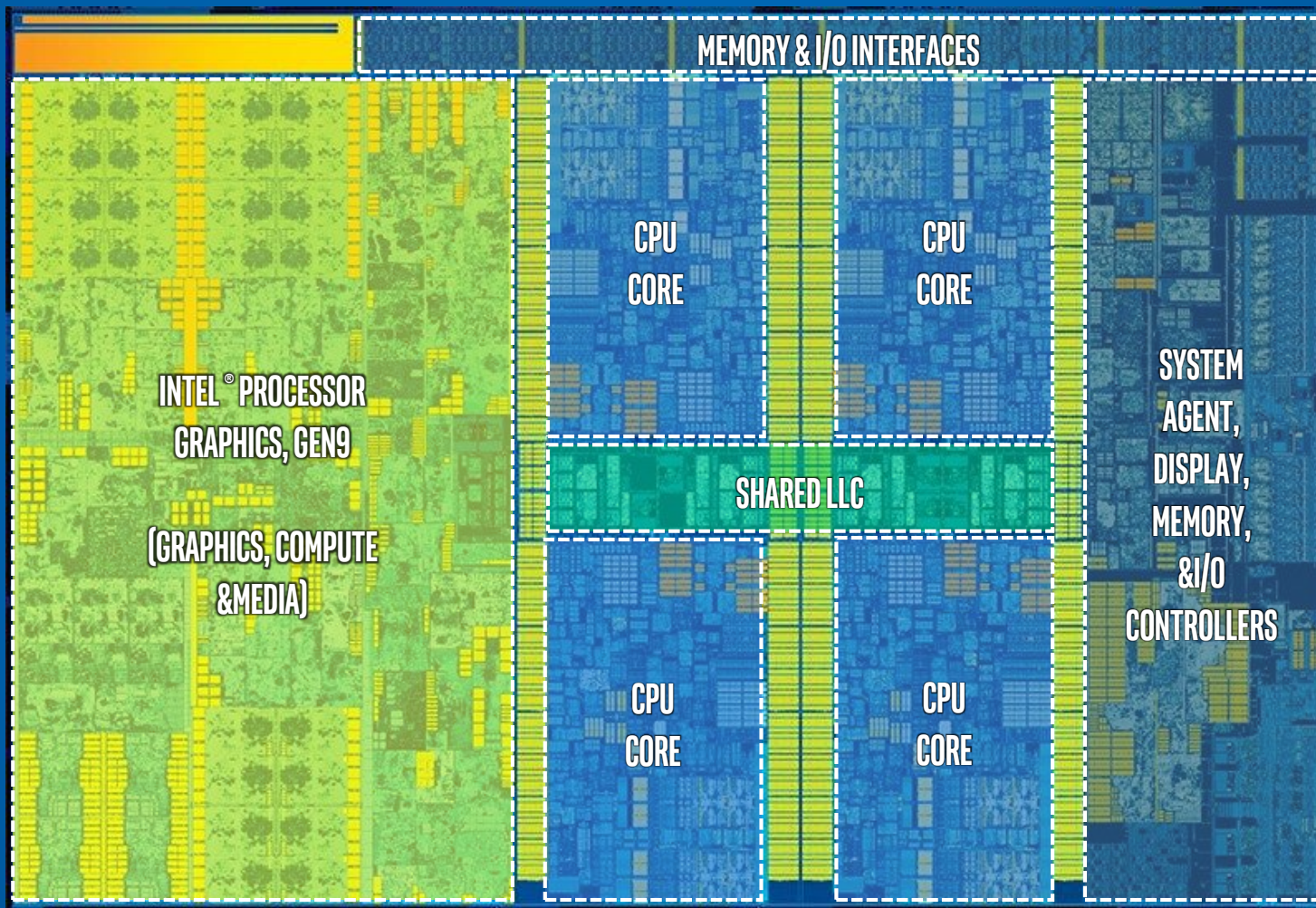


WHAT IS SOC?

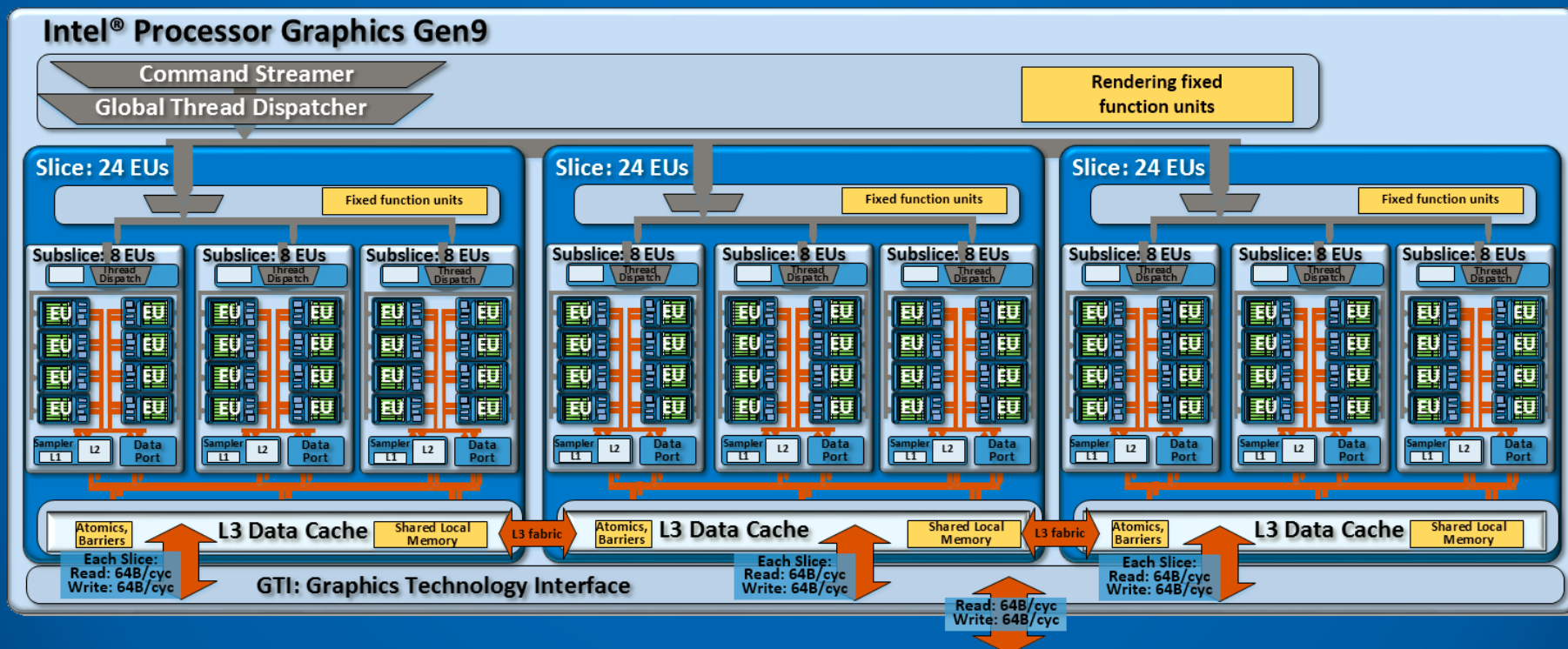
System-on-a-Chip (SoC) integrates a complete set of system components into a single chip:

- Usually contains CPU cores
- With integrated graphics, also contains GPU
- Additional components (busses, controllers, etc.)
- Caches, EDRAM, LLC

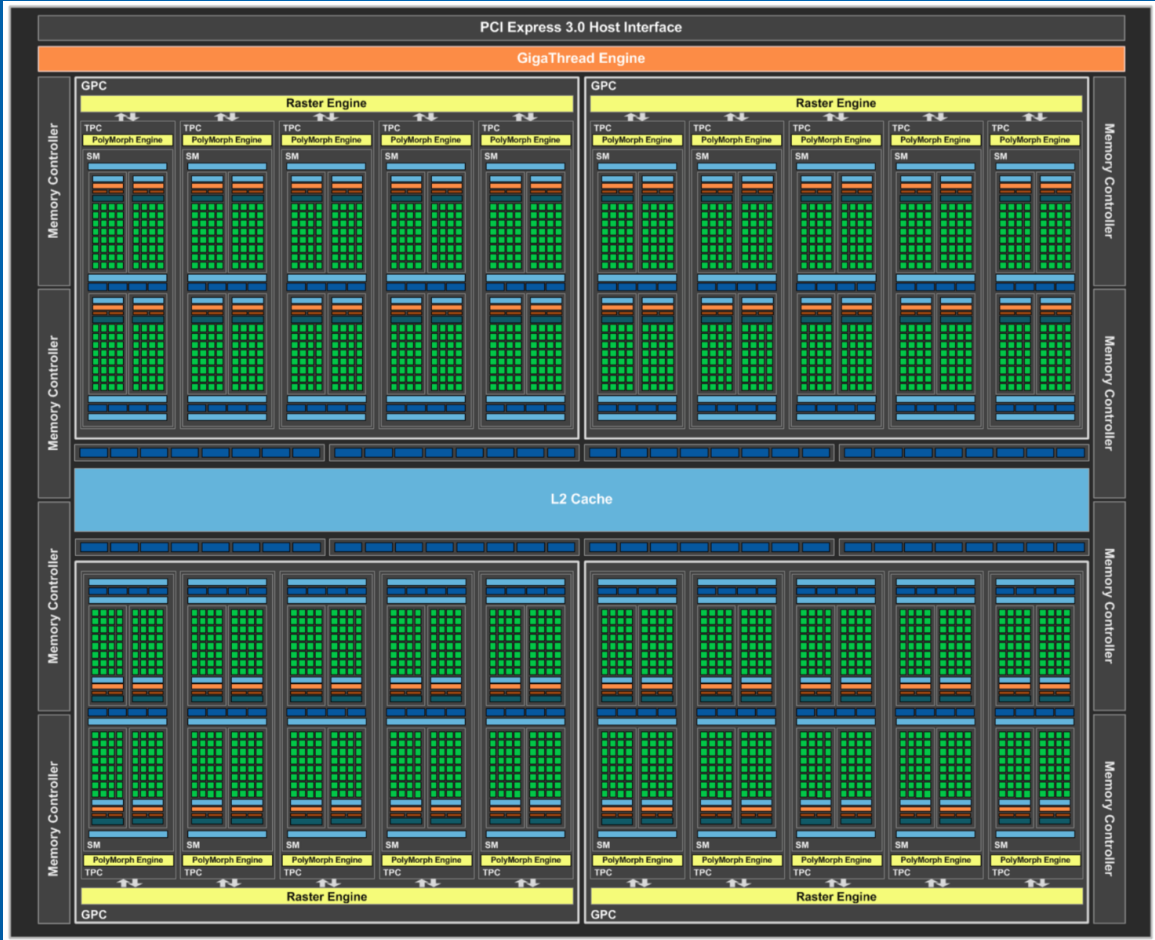
WHAT IS SOC?



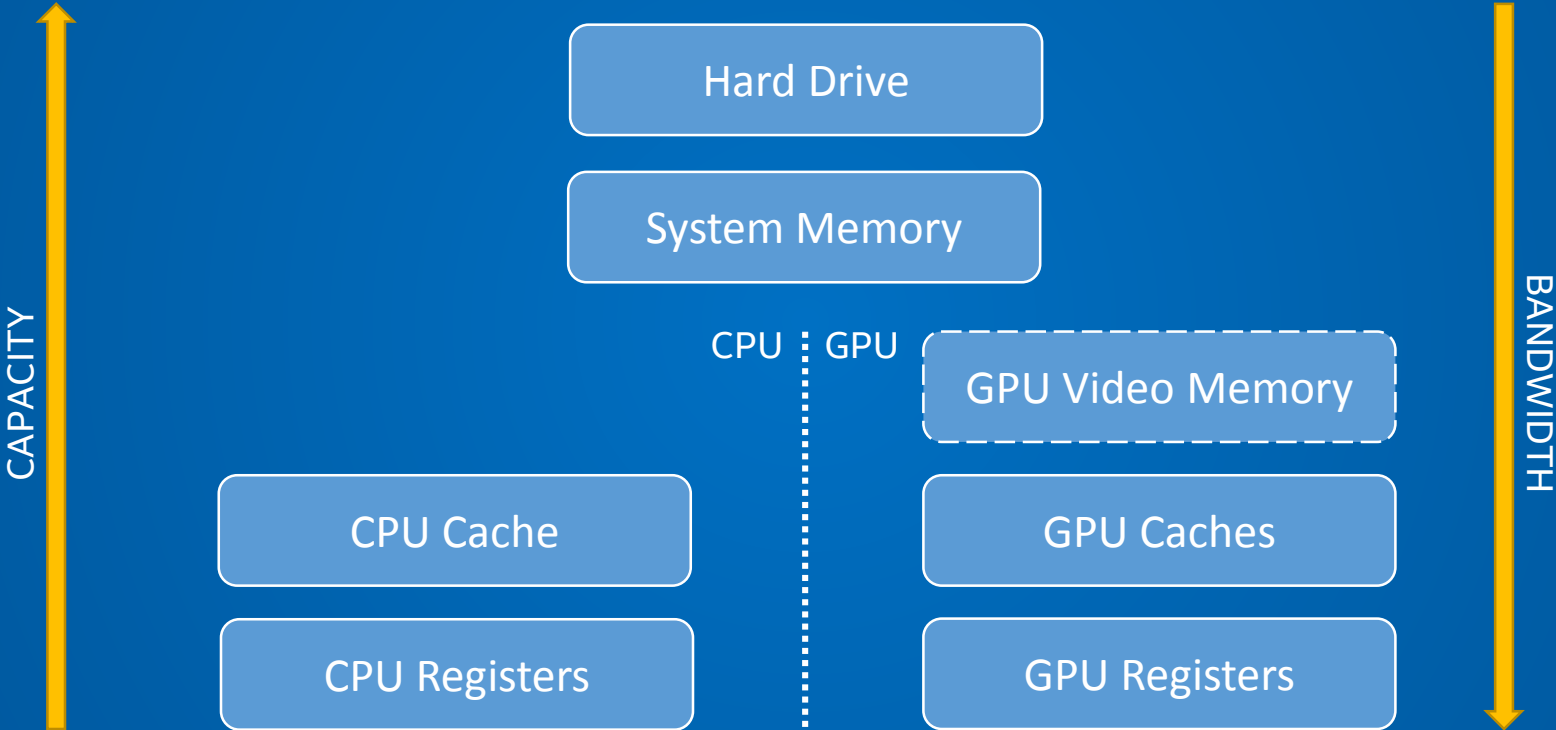
INTEL GPU ARCHITECTURE



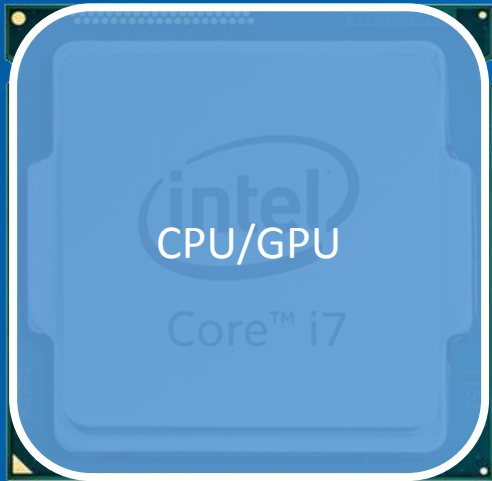
NVIDIA GPU ARCHITECTURE



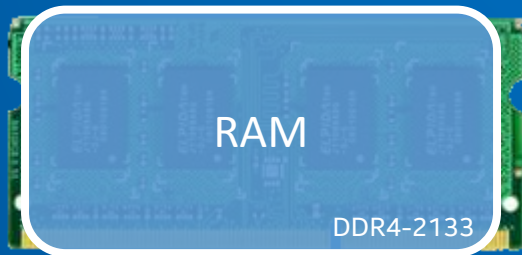
MEMORY MODEL



MEMORY MODEL (INTEGRATED)

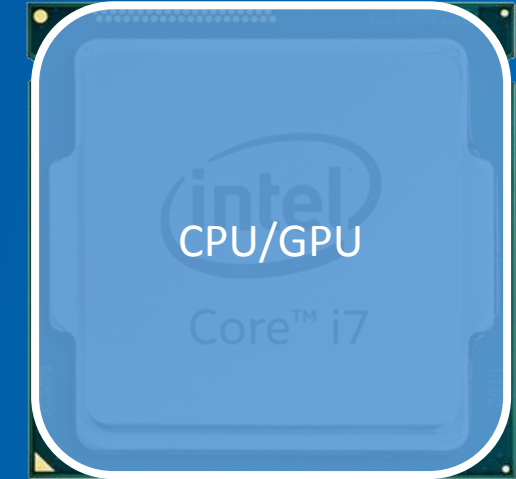


↑↓ 20.04 GB/s

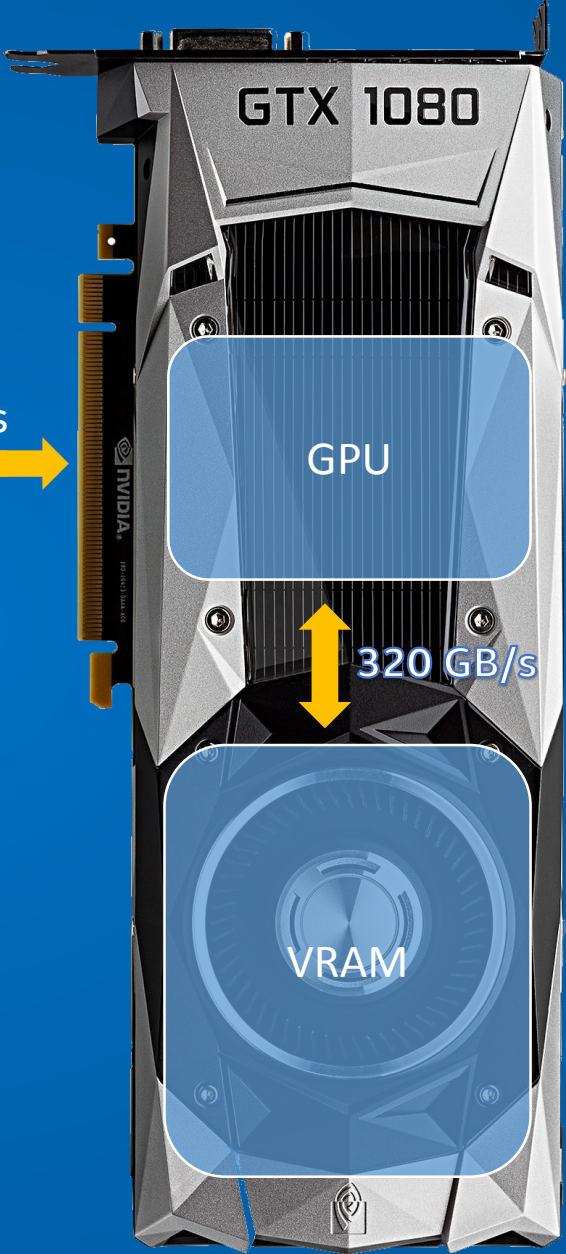


- Integrated graphics
- Unified Memory Architecture
- Zero copy data between CPU and GPU

MEMORY MODEL (DEDICATED)



32 GB/s
PCIExpress

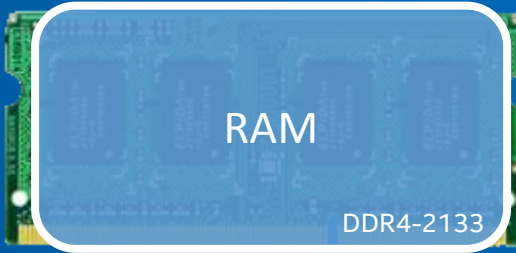


GPU

320 GB/s

VRAM

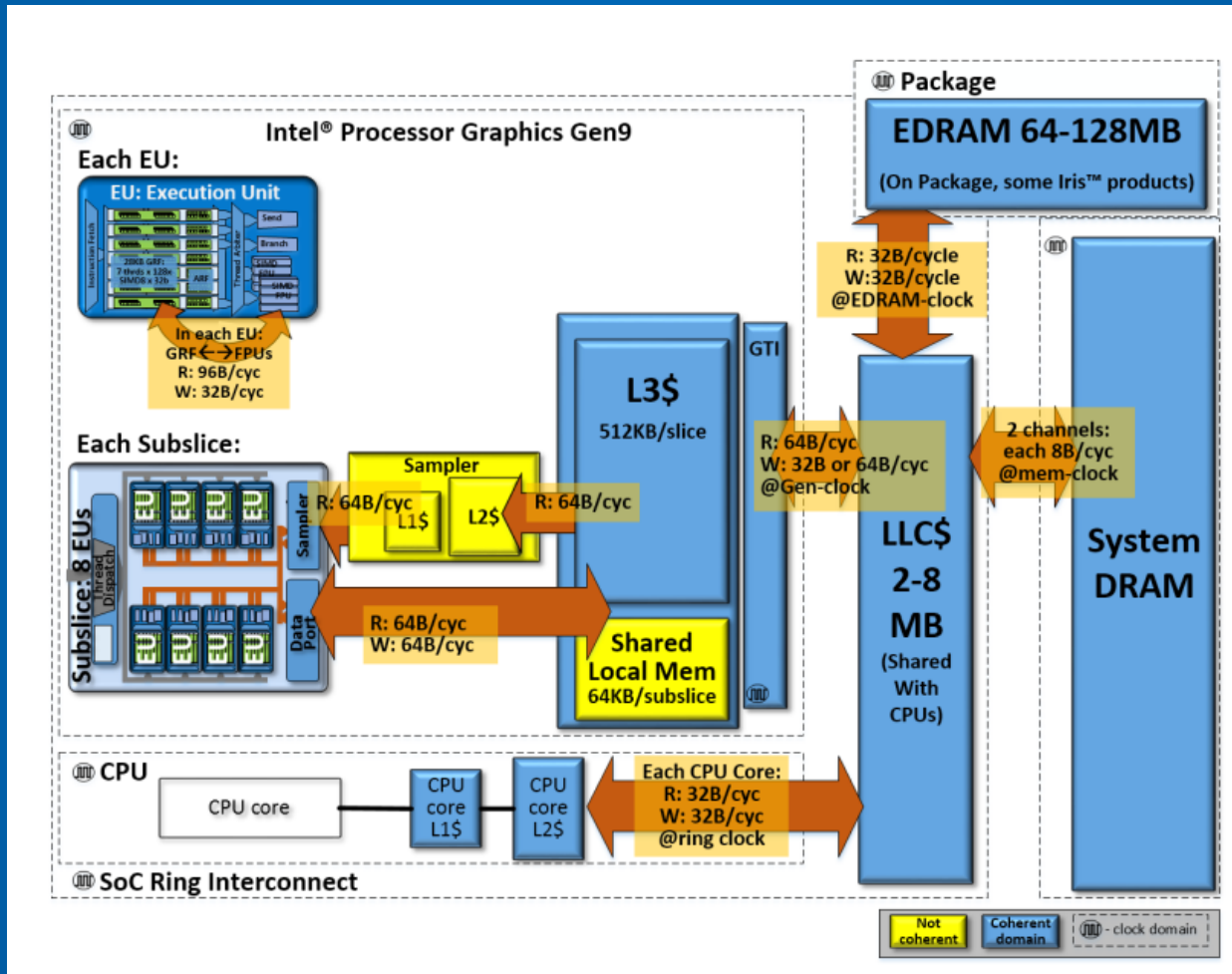
20.4 GB/s



RAM

DDR4-2133

MEMORY MODEL



TDP

- Thermal Power Design
- Represents the average power, in watts, the SOC dissipates when operating at base frequency with all cores active
- On integrated graphics, the power is shared between CPU, GPU, LLC, EDRAM, and every other component

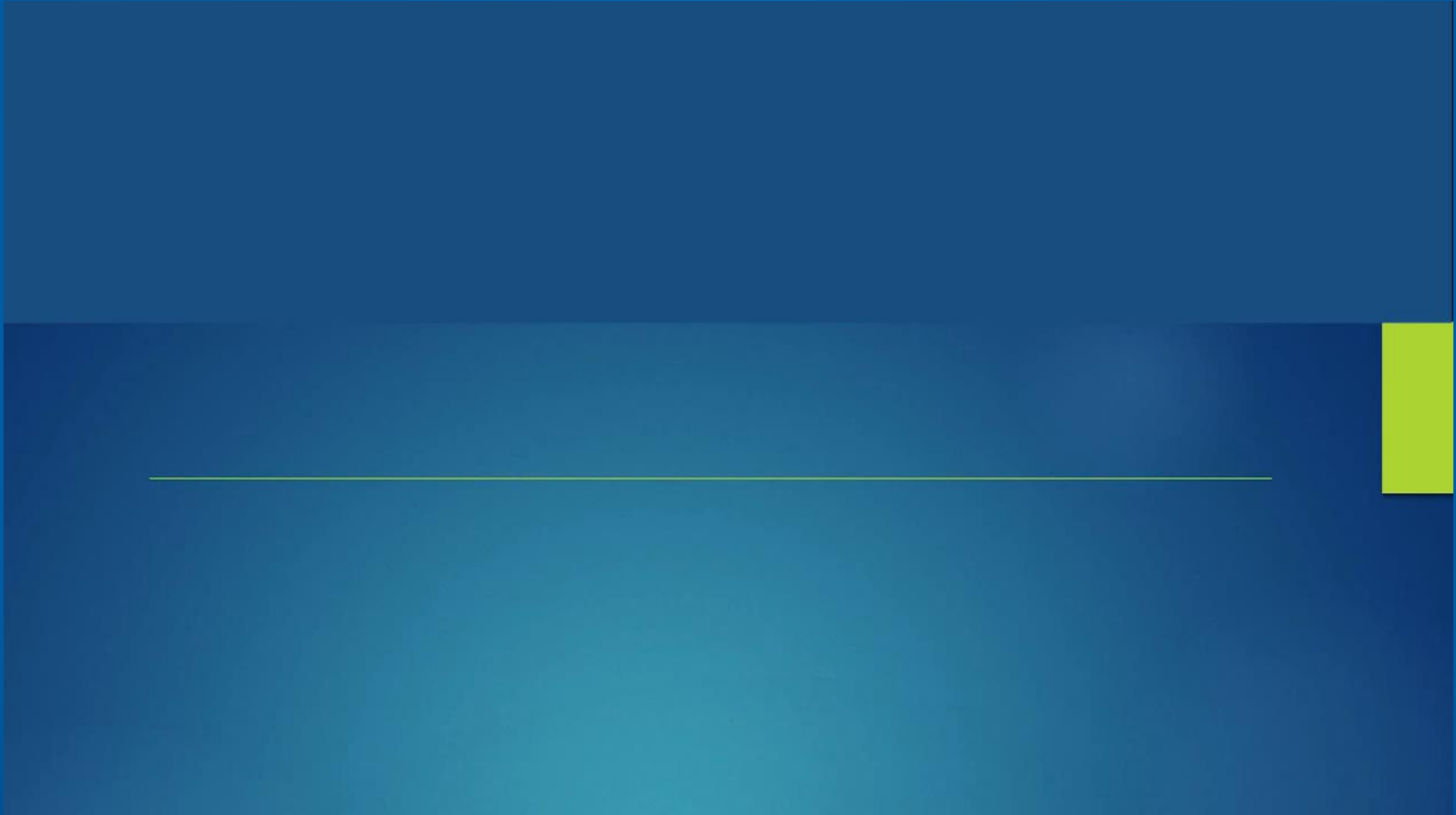
GFLOPS

- FLOPS, Floating Point Operations Per Second
- Used for measuring performance
- Calculation is complicated
- Based on frequency and GPU architecture

GFLOPS

PLATFORM	GPU	GFLOPS
	AMD Radeon GCN	1311
	Nvidia G70	240
	AMD Radeon GCN	1840
	Skylake GT4e	1152
	GP104	8228

SUMMARY



SUMMARY

- Three resources: CPU, GPU, Memory
- Always have that in mind
- Many different configuration of those resources
- Each resource have it's own advantages and disadvantages

DRIVER



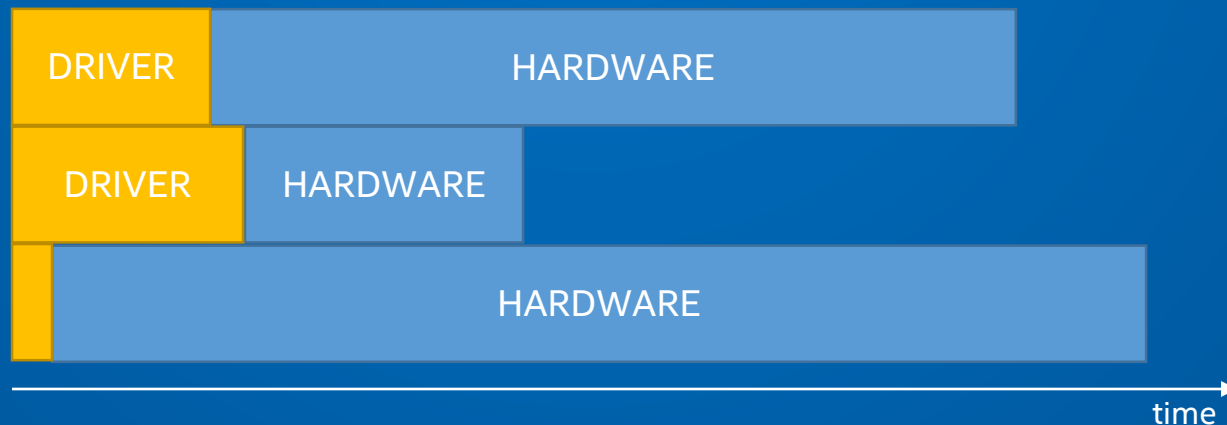
- Don't know hardware details
- Don't really want to know hardware details 😊
- Hardware vendors also supply drivers
- Between hardware and API
- You can always update software

DRIVER



WHAT DRIVER DOES?

- Not a tiny program (~600 MB!)
- Also uses and consumes HW resources – REMEMBER!
- Keeps track of the graphics resources
- Validates the data
- Controls the hardware (...but how?)
- Compiles shaders



WHAT DRIVER DOES?

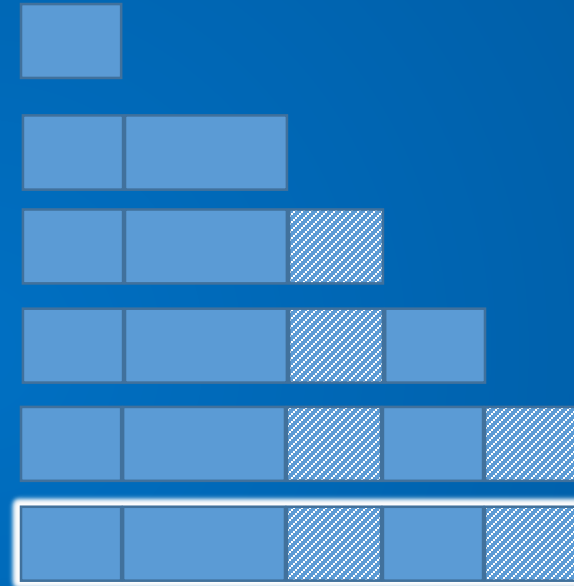
- DirectX provides function *CreateTexture2D()*
 - Application doesn't really care how this will be done
 - Application just want to have a pointer to a valid texture
 - Driver has to allocate the memory for it
 - Decides how this texture will be stored
 - Allocates memory for mip-maps
 - Copy the data from system memory (or not)
-
- DirectX provides function *ClearRenderTargetView()*
 - Driver may spawn own shaders
 - A lot to do, so also consumes resources

HOW DRIVER CONTROLS GPU?

- Driver interprets API calls and prepares a work for GPU in a form of command buffers
- Command buffers are then delivered to the GPU and enqueued in command buffer queue
- GPU is constantly reading command buffers from the queue and executes them
- Sometimes there is no work to do on the GPU, sometimes there is too much work

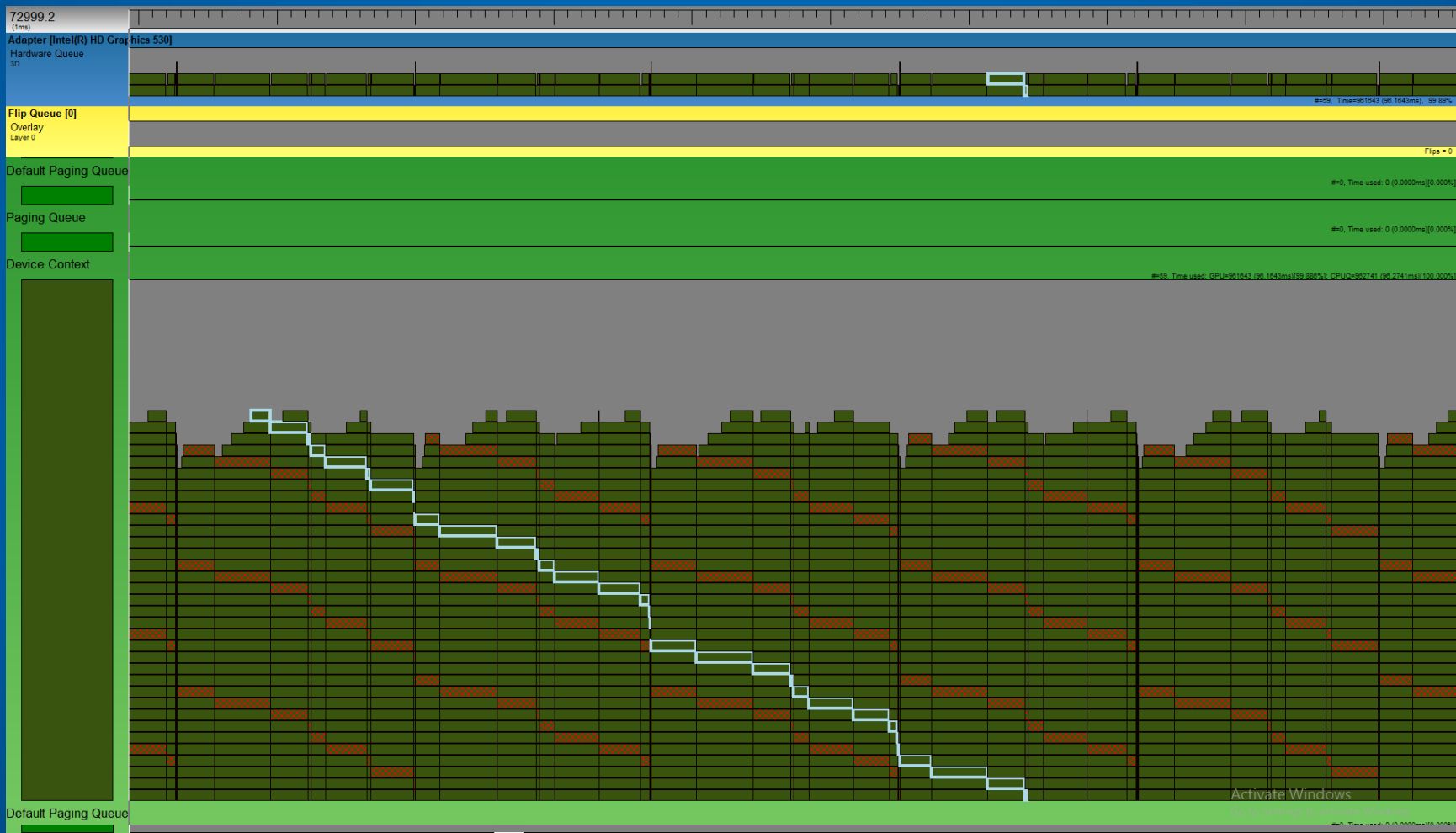
HOW DRIVER CONTROLS GPU?

- PSSetShader()
- IASetVertexBuffer()
- Draw()
- PSSetShader()
- Draw()
- Present()



(command buffer)

HOW DRIVER CONTROLS GPU?



DIRECTX

- Set of application programming interfaces
- Developed by Microsoft, only for Windows platforms
- First version introduced in 1995
- Latest version is DirectX 12, introduced in 2014

DIRECTX

GAME/APP

Microsoft®
DirectX^{API}®

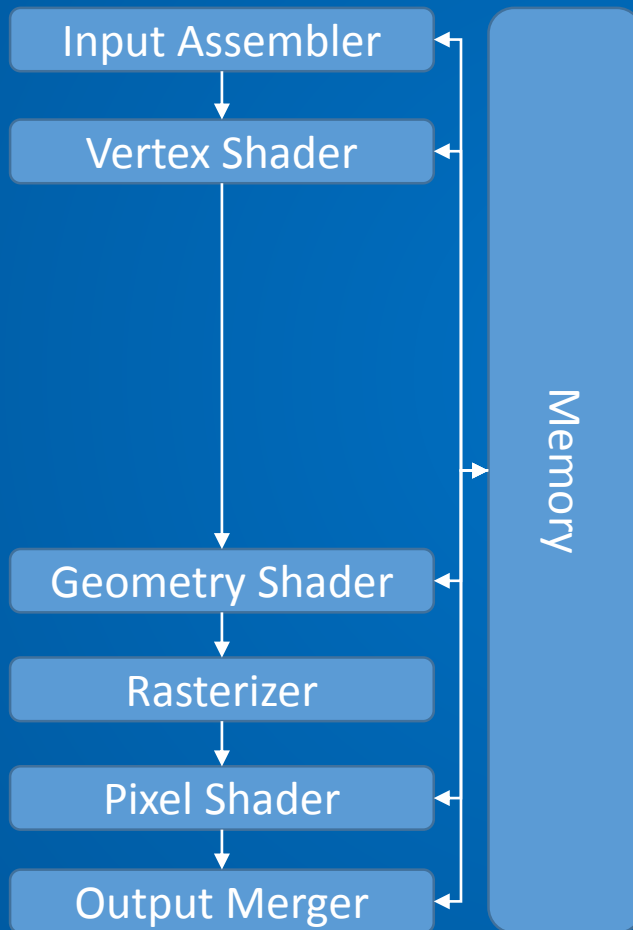
DRIVER

HARDWARE

DRAWCALL

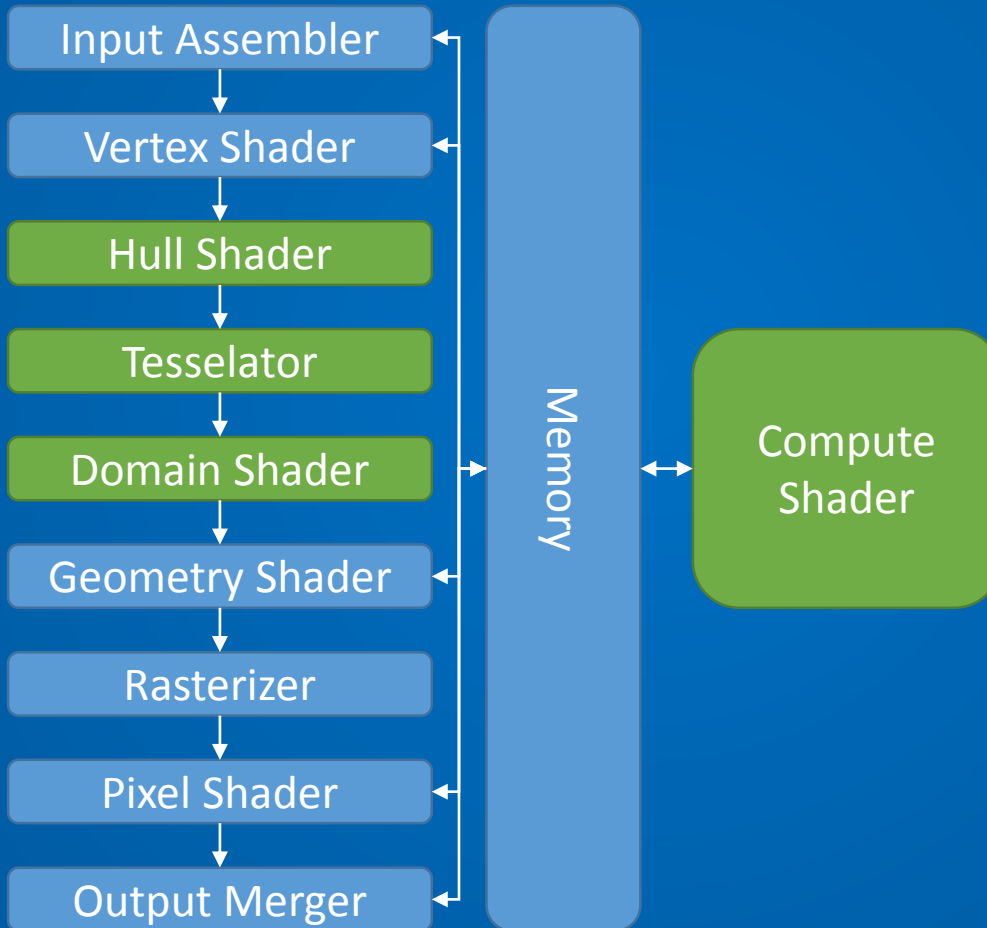
- Any API call that triggers pipeline to do some work
- Typically Draw(), DrawInstanced(), DrawIndexed()
- Also Clears and Copies

DIRECTX 10 PIPELINE



- Represents a fundamental architecture change

DIRECTX 11 PIPELINE



DIRECTX DEVICE

- Represented by ID3D11Device
- Used to create resources
 - CreateBuffer
 - CreateTexture2D
 - CreateRenderTargetView
 - CreateDepthStencilView
 - CreatePixelShader
 - CreateVertexShader

DIRECTX DEVICE CONTEXT

- Also called Device Context
- Used to set pipeline states
- Generates rendering commands
- Using resources owned by a Device
- Examples:
 - Draw
 - IASetVertexBuffer/IASetIndexBuffer
 - PSSetShader
 - OMSetRenderTarget

DIRECTX SWAP CHAIN

- SwapChain is a collection of buffers that are used for displaying frames to the user
- Each time application presents new frame, the first buffer (Front Buffer) in the Swap Chain takes the place of the displayed buffer
- We render to the last buffer (Back Buffer)
- This process is called swapping or flipping



BackBuffer

FrontBuffer



DIRECTX RESOURCES

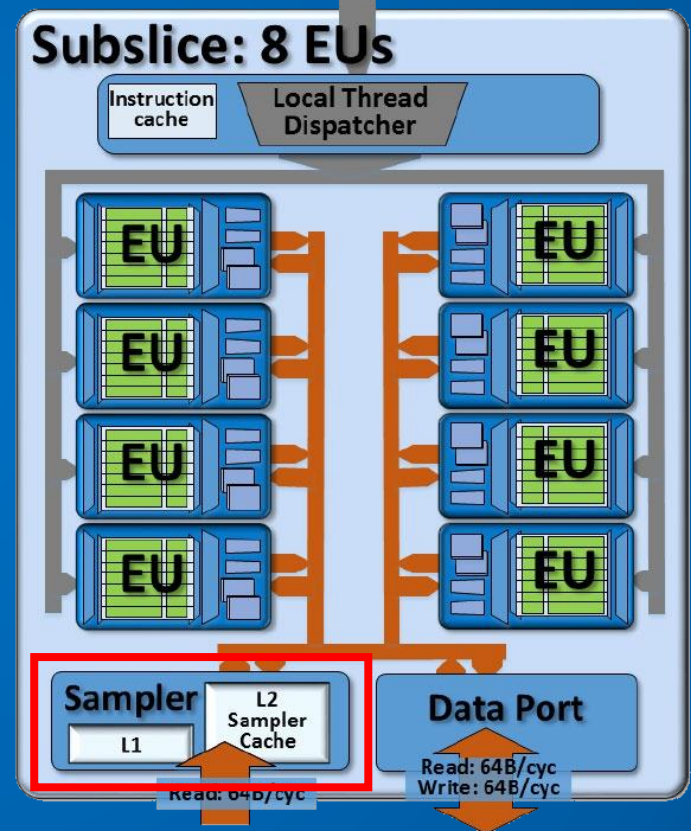
- Building blocks of the scene
- Areas in memory accessible by DirectX pipeline
- Buffers or Textures
- Read or Write access
- Accessible by CPU or GPU
- Create, Bind, Release

DIRECTX BUFFERS

- Collection of elements
- Vertex Buffer, Index Buffer, Constant Buffer
- Unstructured resource
- Cannot contain any mipmap levels
- Cannot get filtered when read
- Cannot be multisampled

DIRECTX TEXTURES

- Stores texels
- Can be filtered by samplers
- Can be read by shader units
- Hardware support - SAMPLERS
- Each texel can have 1 – 4 components



DIRECTX RENDER TARGET

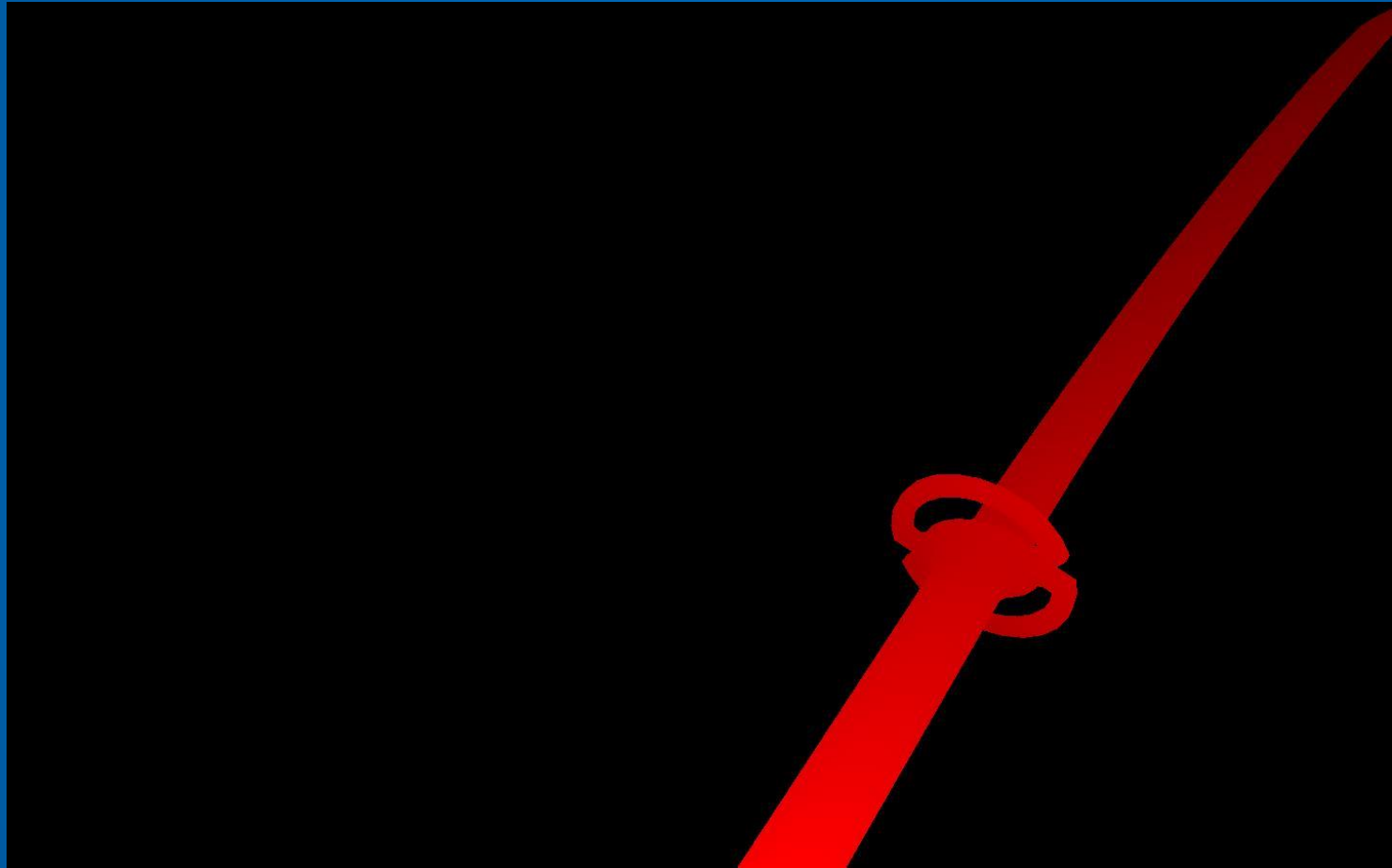
- Usually a texture, where GPU draws pixels for a scene that is being rendered
- In DX11 it is Render Target View – a cast of a resource
- Render Target doesn't necessarily need to be a back buffer

DIRECTX DEPTH BUFFER

- Contains per-pixel data for z-depth of each pixel rendered
- Strongly related to Render Target
- Each render target has to have Depth Buffer
- Depth Buffer can be independent

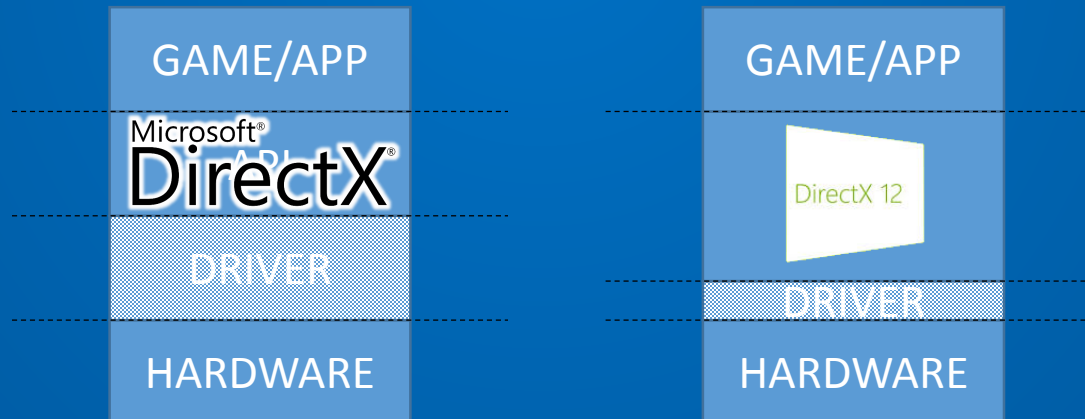


DIRECTX DEPTH BUFFER



DIRECTX 12 / VULKAN

- So called low-level API
- Reduced CPU overhead
- Application has much more work
- Driver is very thin



THANK YOU!

bartosz.boczula@intel.com