Real Time Graphics

Dieter Schmalstieg

Introduction

Partially based on material from Michael Wimmer, Vienna University of Technology

Grand Theft Auto V

- 250M$ development cost
- 1000M$ sales in 3 days

Playstation 4: Destiny

- September 2014
- Online shooter
- 500M$ sales on first day

Real-time Rendering: What is it all about?

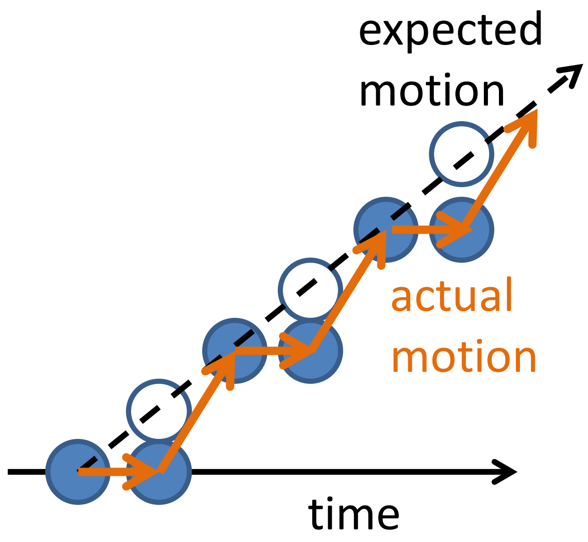

artificial intelligence?
bump maps
levels of detail
texturing
IGN.COM

But, most importantly…

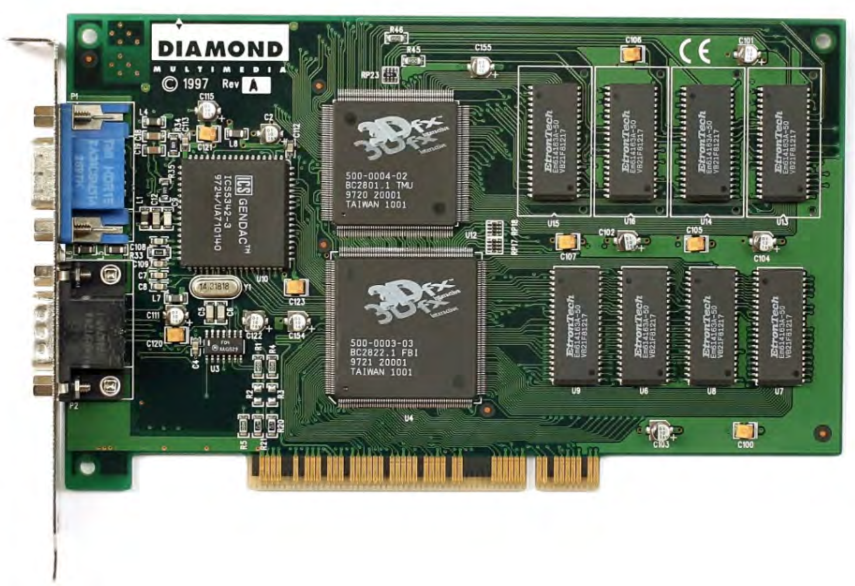
… all of this at 60 frames per second!

Why 60Hz?

- Actually, more might be needed
- Aliasing with monitor refresh rate!
- Jerky or blurry motion perceived
- Head mounted displays (Oculus Rift) need even higher refresh rate



Hardware Development



Diamond Monster 3D 3dfx Voodoo 1 (1997)

NVIDIA GeForce GTX 980 (2014)

Evolution of Real-time Graphics

- Some important phases:
  - Early research
  - Flight simulation
  - SGI workstations
  - PC
- Hardware generations:
  - Different development track for SGI/PC
  - Defined by feature set, but:
  - Any feature can be implemented in hardware
    - Early SGI: hardware geometry, no texturing
    - Early PC: hardware texturing, no geometry

Categories of Advancement

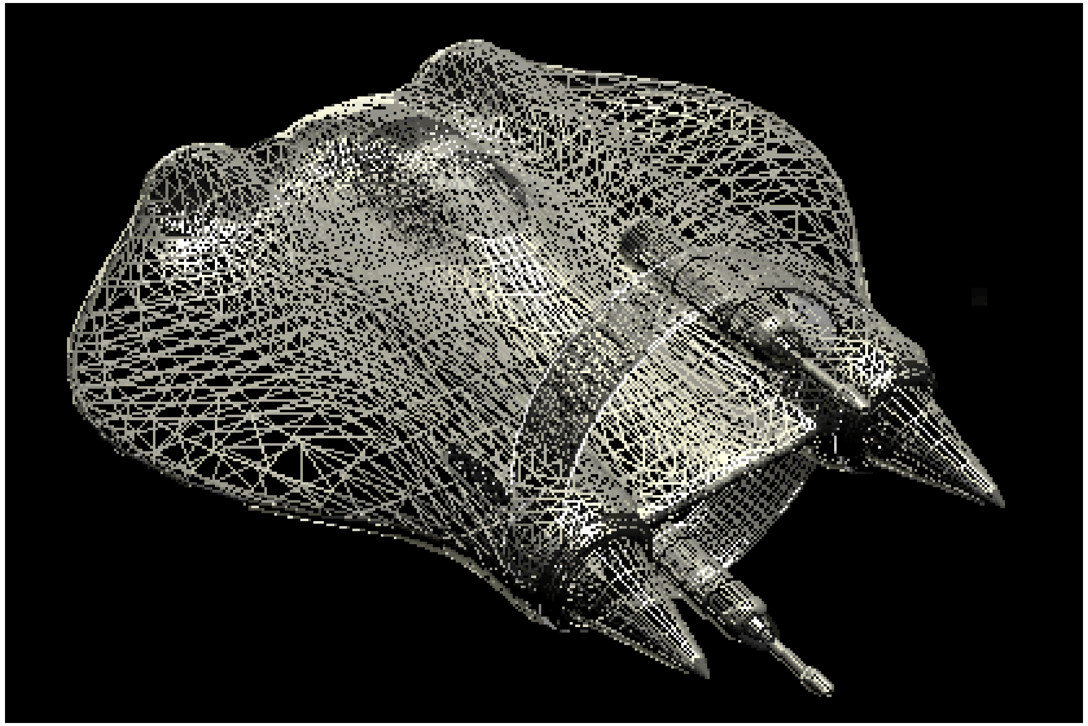
- Performance
  - Triangles/second
  - Fragments/second
- Features
  - Hidden surface elimination
  - Texture mapping
  - Programmable shading
- Quality
  - Numeric precision: int 8/10/16 bit, float 16/24/32 bit
  - Supersampling/antialiasing

Relationship Hardware/Software

- In software, *performance* is inversely proportional to algorithm *complexity*
- Hardware is different: performance is either
  - invariant to complexity (texture-filtering), or
  - falls off catastrophically (software fallbacks).
  - If a feature is “free”, rendering without it may be even slower (e.g. mipmapping).
- Pipelining leads to “free” features
  - Geometric transformations, multi-texturing

First Generation – Wireframe

- Prior to 1987
- Vertex: transform, clip, project
- Fragment: color interpolation
- Frame buffer: overwrite



Second Generation – Shaded Solids

- 1987–1992
- Vertex: lighting calculations
- Fragment: depth interpolation, triangles
- Frame buffer: depth buffer, blending

Third Generation – Texture Mapping

- 1992–2000
- Vertex: texture coordinate transformation
- Fragment: texture coordinate interpolation
  texture evaluation and filtering

Fourth Generation – Programmability

- Programmable shading
  - Replaced fixed function pipeline
- DirectX8/9
  - Vertex shading
  - Fragment shading
- DirectX10/11
  - Unified Shader Model
  - Geometry Shading
  - Tesselation

DirectX11/12

- For Windows Vista and Windows 7
- Shader stage: Tessellation
  - Input: low-detail mesh
  - Output : high-detail mesh
- DirectCompute
  - GPGPU

CryEngine3

Authoring without Tessellation

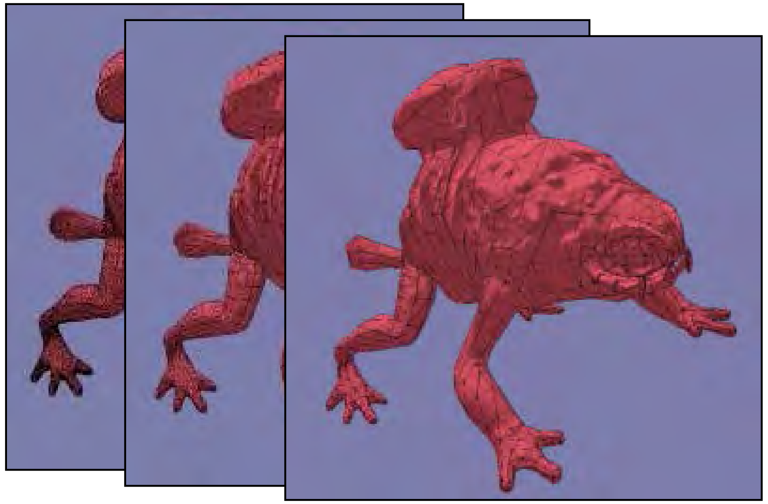
Sub-D Modeling

Animation

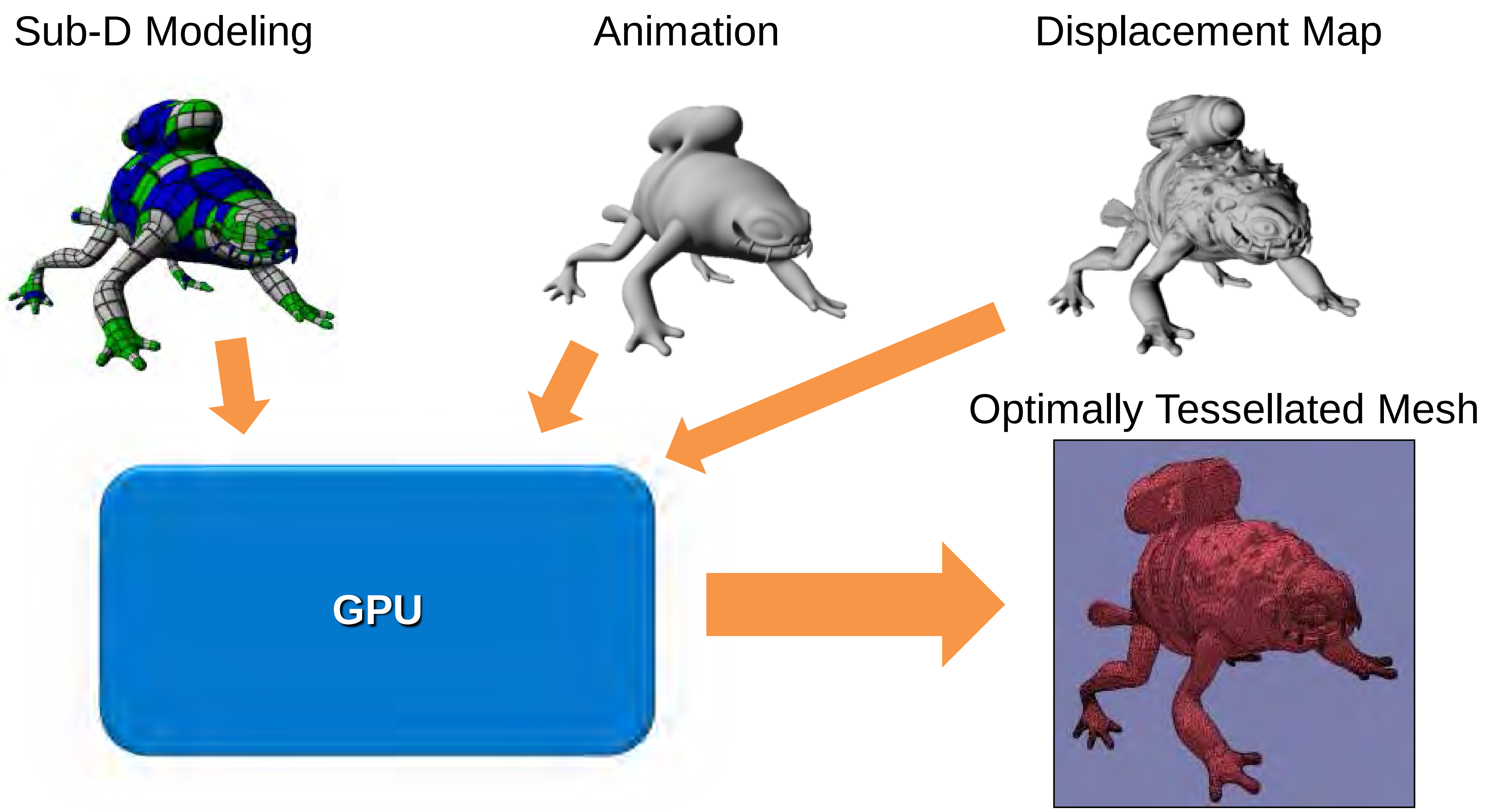
Displacement Map

Polygon Mesh

Generate LODs

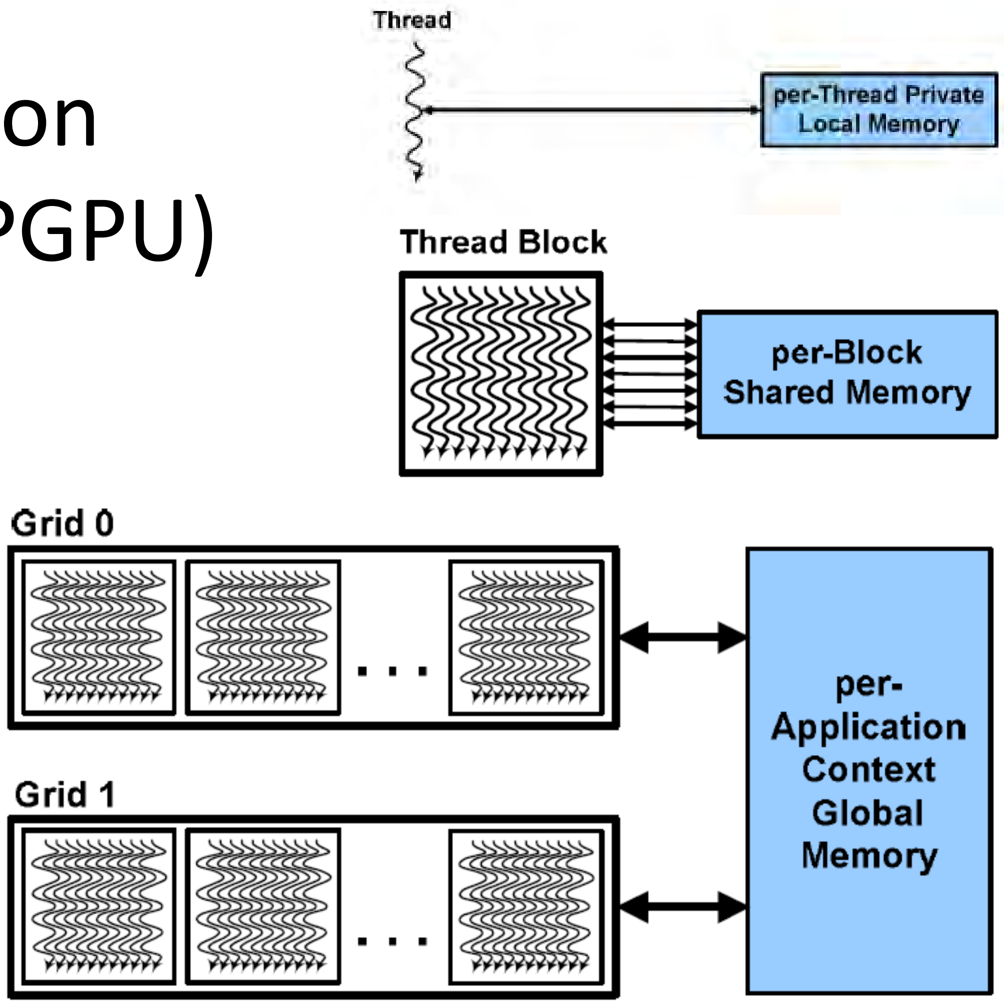


Authoring with Tessellation



Fifth Generation

- General-purpose computing on graphics processing units (GPGPU)
  - Enabled by unified shader architecture
  - Massive parallelization on graphics card



Other GPGPU Applications

- Massive parallel visual computing
- Ray tracing, radiosity, photon mapping
- Physics simulation (e.g. NVIDIA's PhysX)
- Scientific computation
- New approaches
  - Intel Xeon Phi: 64 CPU cores + vector instructions
  - Intel, AMD: Combined CPU+GPU, shared memory

NVIDIA GeForce GTX980 Ti

GEFORCE GTX 1080

GPU Engine Specs:

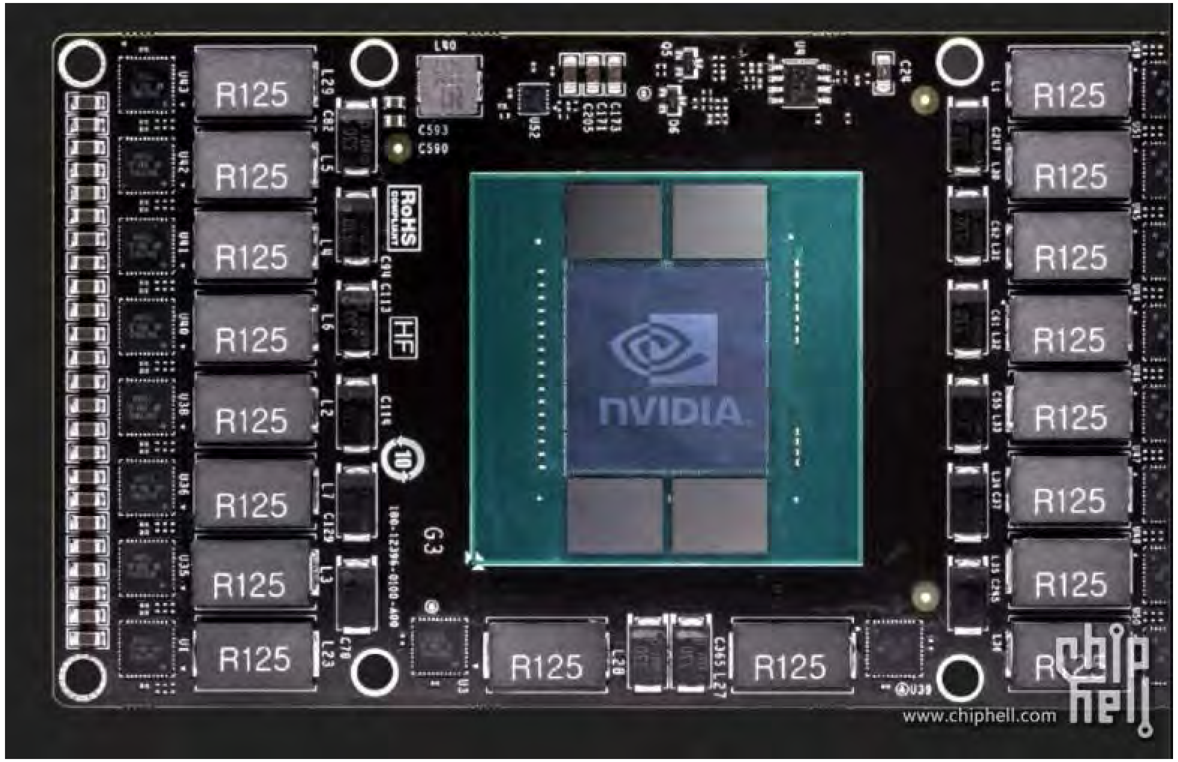
NVIDIA CUDA® Cores	2560
Base Clock (MHz)	1607
Boost Clock (MHz)	1733

Memory Specs:

Memory Speed	10 Gbps
Standard Memory Config	8 GB GDDR5X
Memory Interface Width	256-bit
Memory Bandwidth (GB/sec)	320

Technology Support:

Simultaneous Multi-Projection	Yes
VR Ready	Yes
NVIDIA Ansel	Yes
NVIDIA SLI® Ready	Yes - SLI HB Bridge Supported
NVIDIA G-SYNC™-Ready	Yes
NVIDIA GameStream™-Ready	Yes
NVIDIA GPU Boost™	3.0
Microsoft DirectX	12 API with feature level 12_1
Vulkan API	Yes
OpenGL	4.5
Bus Support	PCIe 3.0
OS Certification	Windows 7-10, Linux, FreeBSDx86



GK 110

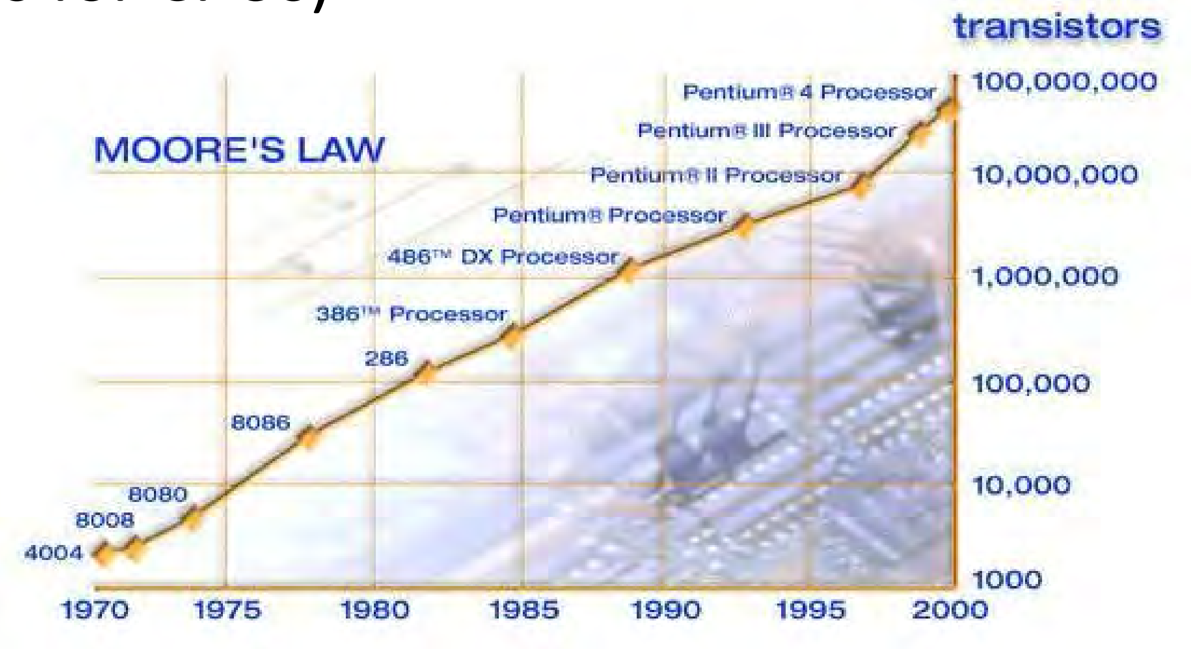
GPGPU Progress

GTX960

	GeForce GTX 980	GeForce GTX 970	GeForce GTX 780	GeForce GTX 780 Ti
GPU	GM204	GM204	GK110	GK110
Fertigung	28 nm	28 nm	28 nm	28 nm
Transistoren	5,2 Mrd.	5,2 Mrd.	7,1 Mrd.	7,1 Mrd.
Shader-Rechenkerne	2048	1664	2304	2880
Rechengruppen	16 SMM	13 SMM	12 SMX	15 SMX
Textureinheiten	128	104	192	240
Texeldurchsatz	144,1 GTex/s	109,2 GTex/s	165,7 GTex/s	210 GTex/s
Rasterendstufen	64	64	48	48
GPU-/Turbo-Takt	1126 / 1216 MHz	1050 / 1178 MHz	863 / 902 MHz	875 / 928 MHz
Rechenleistung (SP)	4,61 TFlops	3,49 TFlops	3,97 TFlops	5,04 TFlops
Rechenleistung (DP)	0,192 TFlops	0,145 TFlops	0,165 Tflops	0,21 Tflops
Speicher	4 GByte GDDR5	4 GByte GDDR5	3 GByte GDDR5	3 GByte GDDR5
Speicher-Takt (R/W)	3506 MHz	3506 MHz	3004 MHz	3500 MHz
Speicher-Anbindung	256 Bit	256 Bit	384 Bit	384 Bit
Datentransferrate	224 GByte/s	224 GByte/s	288,3 GByte/s	336 GByte/s
Stromanschlüsse	2 × 6-pin	2 × 6-pin	1×6-pin, 1×8-pin	1×6-pin, 1×8-pin
Formfaktor	Dual-Slot	Dual-Slot	Dual-Slot	Dual-Slot
Display-Anschlüsse	3 × DP 1.2, HDMI 2.0, DL-DVI	3 × DP 1.2, HDMI 2.0, DL-DVI	2 × DL-DVI, HDMI, DP	2 × DL-DVI, HDMI, DP
Mehrschirmbetrieb	4	4	3+1	3+1
Temperatur-Grenzwert	95 °C	95 °C	95 °C	95 °C
TDP	165 Watt	145 Watt	250 Watt	250 Watt
Direct3D Feature Level	11_2	11_2	11_0	11_0
Preis ab	540 Euro	330 Euro	350 Euro	429 Euro

Moore's Law

- Gordon Moore, 1965
- Exponential growth in number of transistors
- Doubles every 18 months (holds for CPUs)
  → yearly growth: 1.6



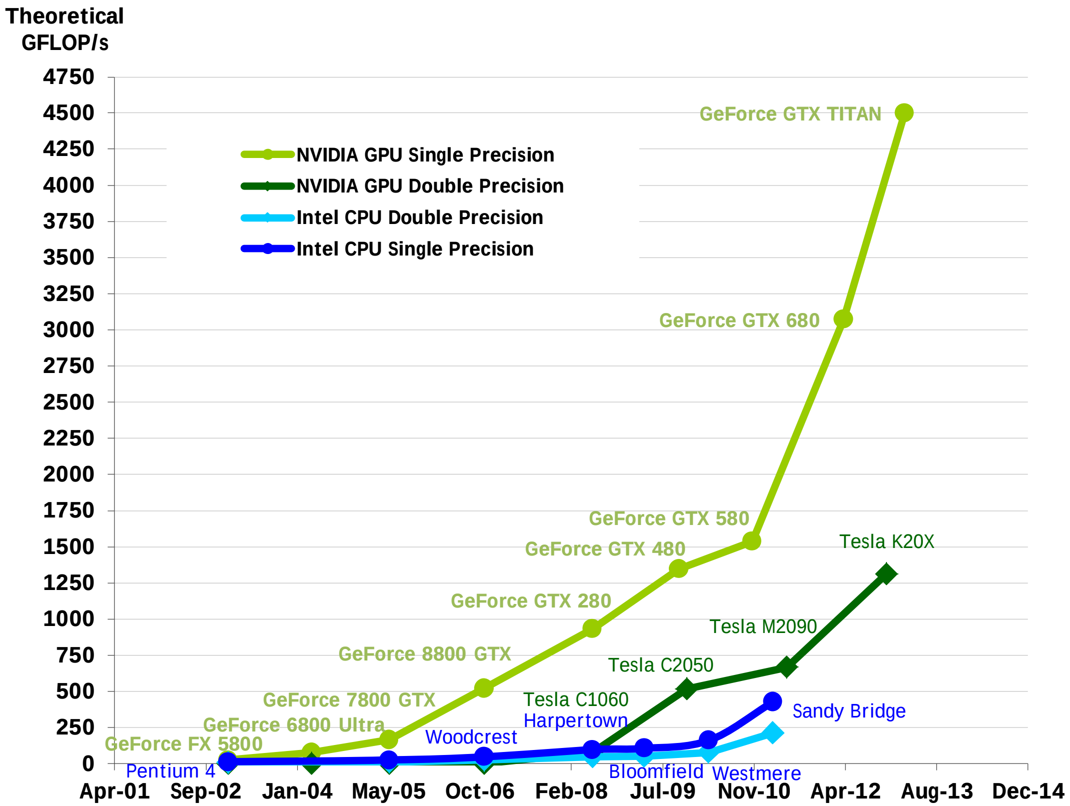
Development (SGI Workstations)

Gen	Year	Product	Fill rate (no Z)	Yr rate	Tri rate (no Z)	Yr rate
1st	1984	Iris 2000	100K (46M)	-	0.8K (10K)	-
2nd	1988	GTX	40M (80M)	1.2	135K	1.9
3rd	1992	RealityEngine	380M	1.5	2M	2.0
		(Cost: $1,000,000, size of a domestic fridge!)				
3rd	1996	InfiniteReality	1000M	1.3	12M	1.6
				1.3		1.8

Yearly tri rate growth above Moore's law!

NVIDIA and Moore’s Law

- GPU computing power increases much faster compared to CPU



Synthetic Benchmarks

- Peak numbers are not achievable in practical situations
- Usually less than peak performance:
  - State changes, pipeline stalls
  - CPU/drivers issues
  - Non-optimal geometry arrangement
  - Bus bandwidth for geometry
  - Cache inefficiencies
  - Nontrivial shading/texturing
- Influenced by:
  - z/texture compression
  - (no) antialiasing

Application Benchmarks

- More important than peak numbers:
  - Shader instructions
  - Feature set
  - Quality (antialiasing, supersampling)
- Simple textured triangles not a representative measure!
- Todays Benchmarks are frame rates in games

Where are we headed?

- Note: development driven by games!
- Toy Story (1995) took an average of 7 hours per frame to render (max. ~90 hours)
- Alvy Ray Smith (MS Graphics Research Fellow & Pixar tech guy) would like 80M polys per frame
  – That’s 4.8 billion polys per sec at 60 Hz
- “Shrek 1” characters: ~800K polys each
- And that’s all Renderman rendered...

Lecture Topics

- Graphics Pipeline, OpenGL, Vulkan
- Texturing and Shading
- Special effects (2D), shading effects (3D)
- Rendering acceleration (LOD, visibility)
- Global illumination