

The Graphics Pipeline

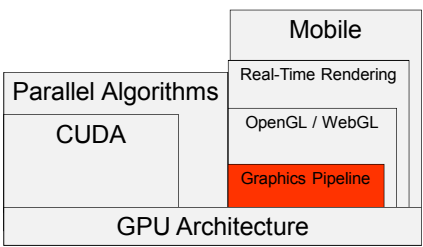
Patrick Cozzi
University of Pennsylvania
CIS 565 - Fall 2012

Announcements

- 10/29 – Eric Lengyel Game Engine Architecture Guest Lecture
- 10/30 – Class in SIG lab
 - Instead of class on 10/31
- Build vs. buy

2

Course Contents



GPU Architecture

Parallel Algorithms

CUDA

Mobile

Real-Time Rendering

OpenGL / WebGL

Graphics Pipeline

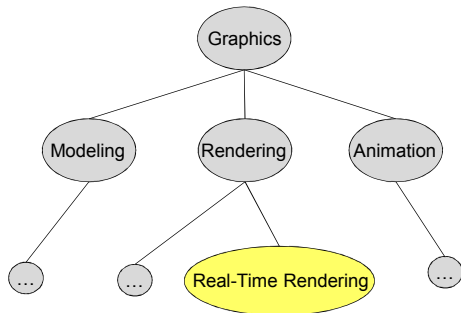
3

Agenda

- Brief Graphics Review
- Graphics Pipeline
- Mapping the Graphics Pipeline to Hardware

4

Graphics Review



5

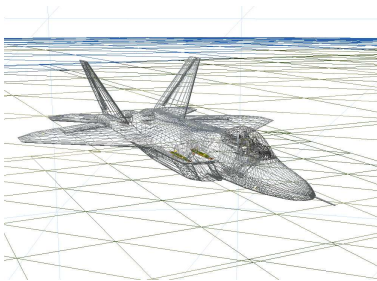
Graphics Review: Modeling

■ Modeling

- Polygons vs. triangles
 - How do we store a triangle mesh?
- Implicit Surfaces
- ...

6

Triangles



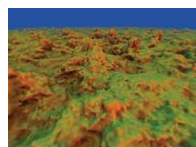
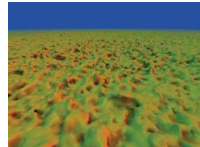
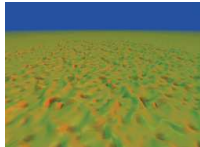
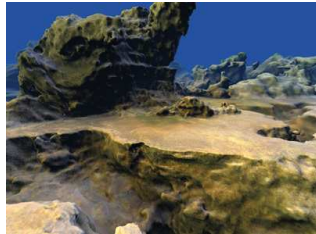
7

Triangles



8

Implicit Surfaces



Images from http://to.developer.nvidia.com/GPUGems3/opugems3_ch01.htm

9

Graphics Review: Rendering



Image credit: Henrik Wann Jensen

Model of a scene:

3D surface geometry (e.g., triangle mesh)
surface materials
lights
camera

Image

How does each triangle contribute to each pixel in the image?

Kayvon Fatahallian CMU 15-869, Fall 2011

Image from http://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15869-f11/www/lectures/01_intro.pdf

10

Graphics Review: Rendering

- Rendering
 - Goal: Assign color to pixels
- Two Parts
 - Visible surfaces
 - What is in front of what for a given view
 - Shading
 - Simulate the interaction of material and light to produce a pixel color

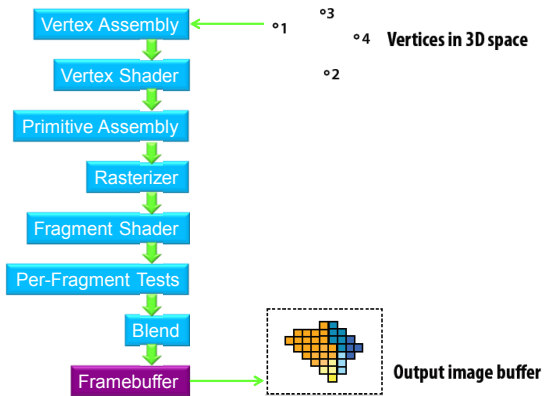
11

Graphics Review: Animation

- Move the camera and/or agents, and re-render the scene
 - In less than 16.6 ms (60 fps)

12

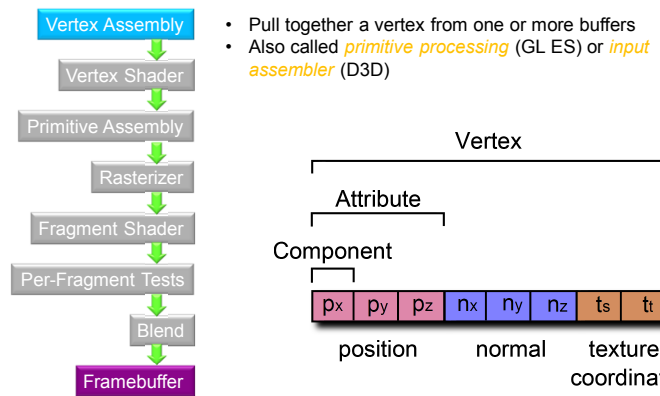
Graphics Pipeline Walkthrough



13

Images from http://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15869-f11/www/lectures/01_intro.pdf

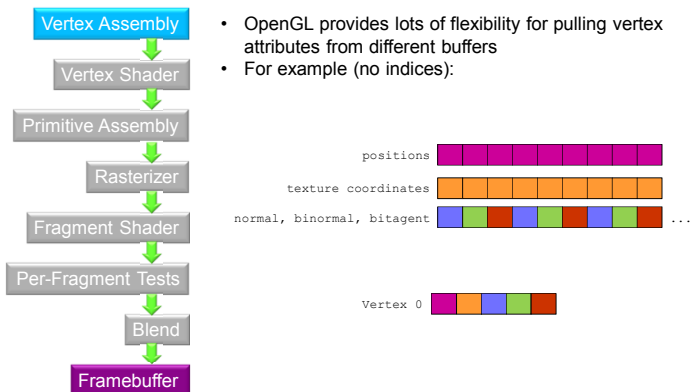
Vertex Assembly



14

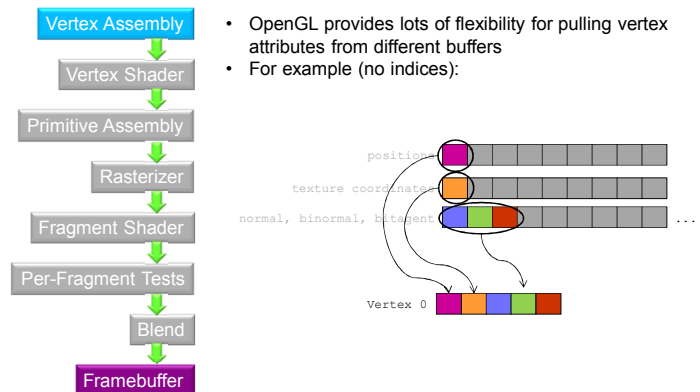
Image from <http://www.virtualbook.com>

Vertex Assembly



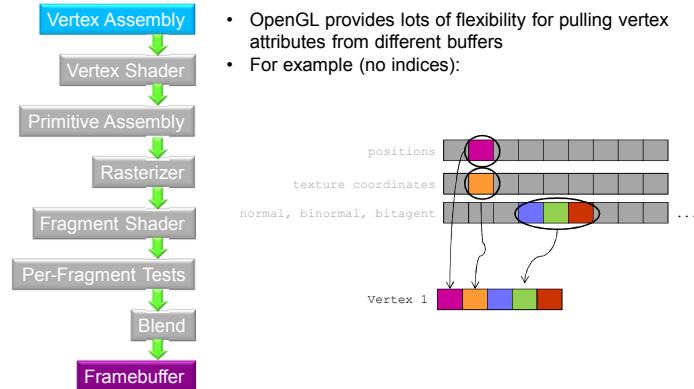
15

Vertex Assembly



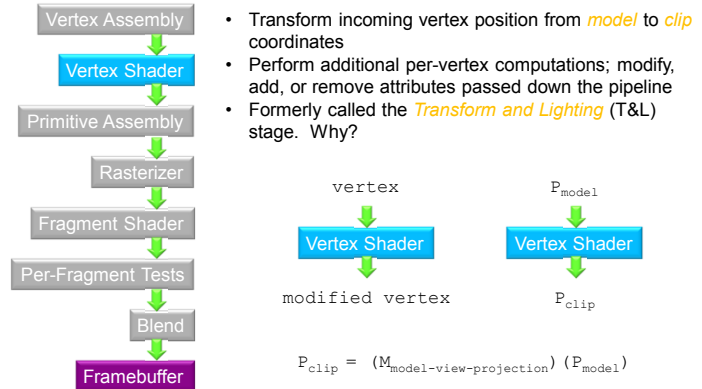
16

Vertex Assembly



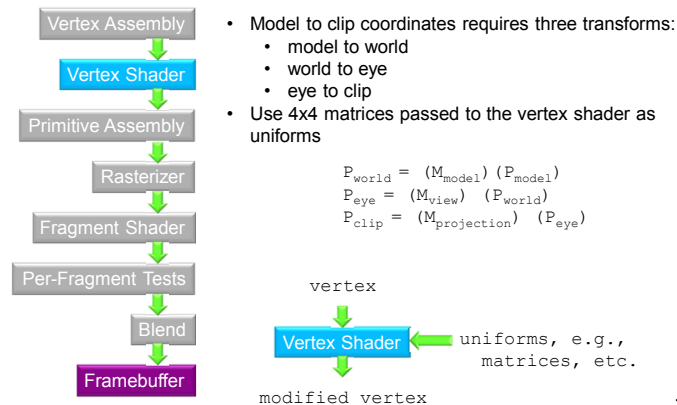
17

Vertex Shader



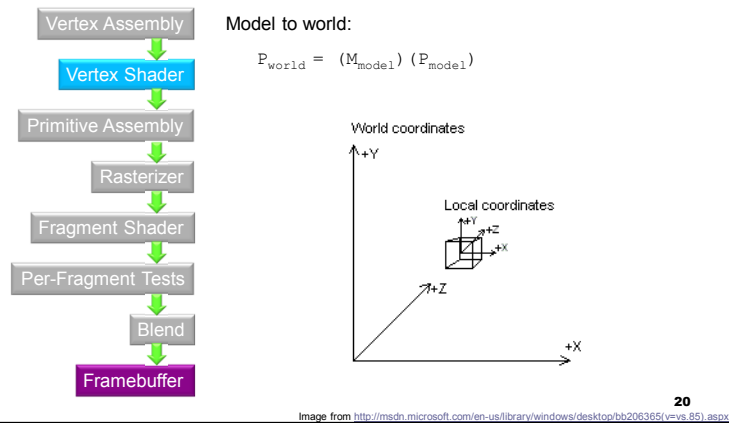
18

Vertex Shader



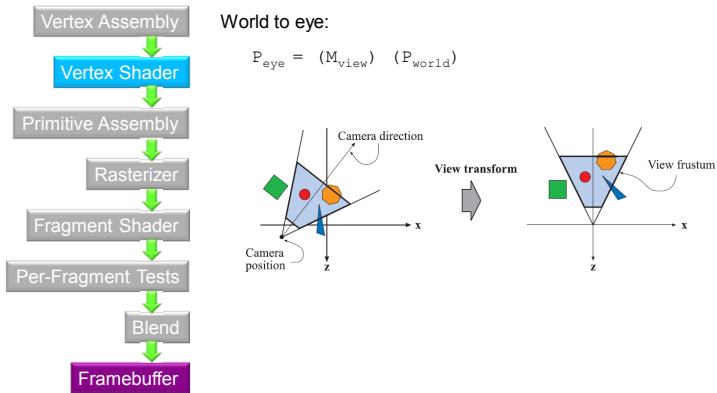
19

Vertex Shader



20

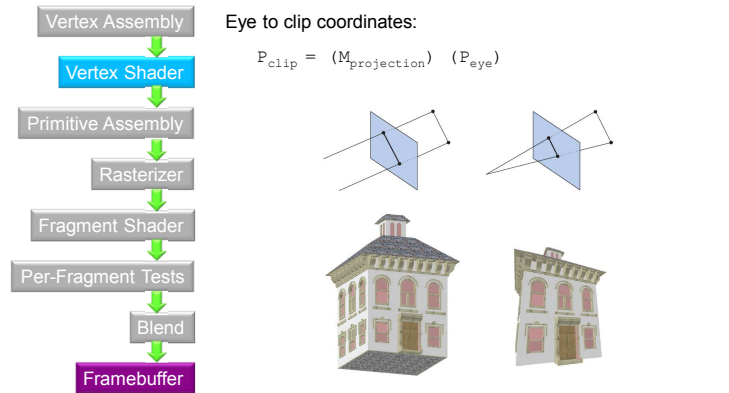
Vertex Shader



21

Image from <http://www.realtimerendering.com/>

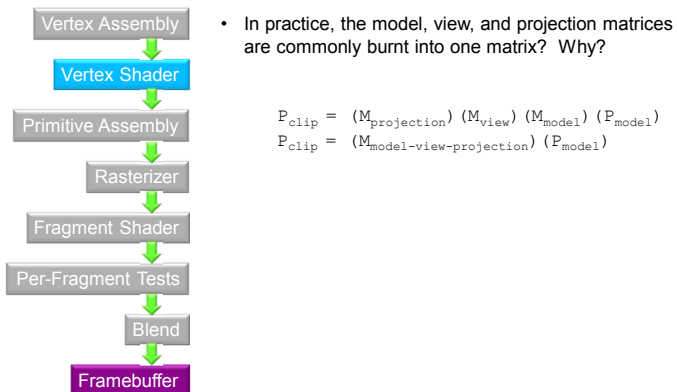
Vertex Shader



22

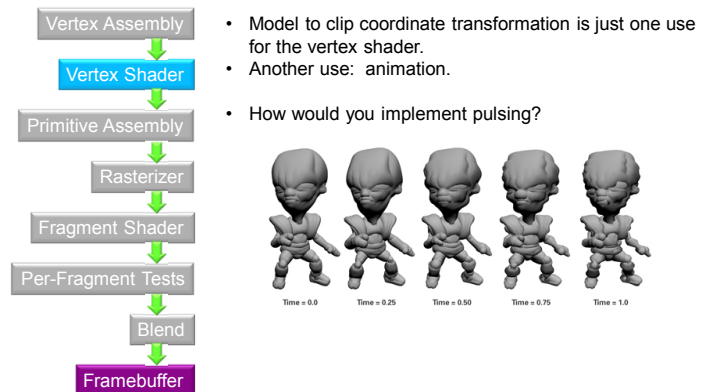
Image from <http://www.realtimerendering.com/>

Vertex Shader



23

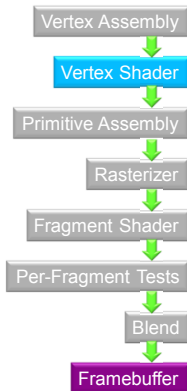
Vertex Shader



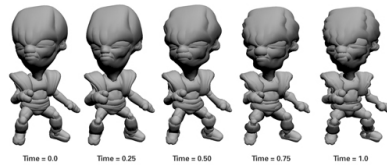
24

Image from http://http.developer.nvidia.com/CgTutorial/cg_tutorial_chapter06.html

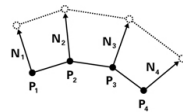
Vertex Shader



- How would you implement pulsing?



- Displace position along surface normal over time

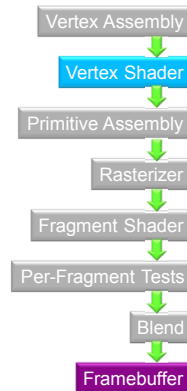


- How do we compute the displacement?

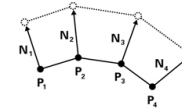
25

Image from http://http.developer.nvidia.com/CgTutorial/cg_tutorial_chapter06.html

Vertex Shader



- How do we compute the displacement?



- Consider:

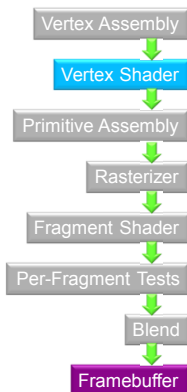
```
float displacement =
    0.5 * (sin(u_time) + 1.0);
```

- What are the shortcomings?

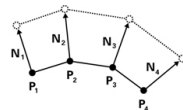
26

Image from http://http.developer.nvidia.com/CgTutorial/cg_tutorial_chapter06.html

Vertex Shader



- How do we compute the displacement?



- Consider:

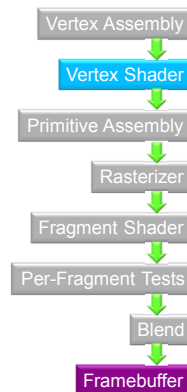
```
float displacement =
    u_scaleFactor * 0.5 *
    (sin(u_frequency * u_time)
    + 1.0);
```

- What are the other shortcomings?

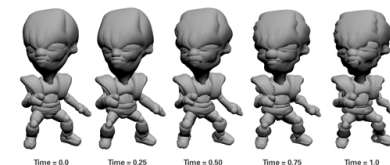
27

Image from http://http.developer.nvidia.com/CgTutorial/cg_tutorial_chapter06.html

Vertex Shader



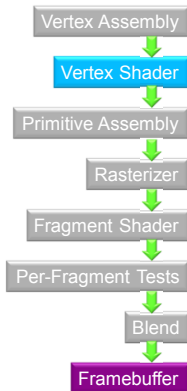
- How do we get the varying bulge?



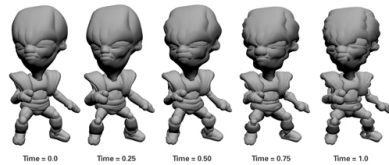
28

Image from http://http.developer.nvidia.com/CgTutorial/cg_tutorial_chapter06.html

Vertex Shader



- How do we get the varying bulge?



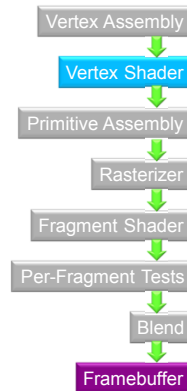
- Consider

```
float displacement =
    u_scaleFactor * 0.5 *
    (sin(position.y * u_frequency *
    u_time) + 1.0);
```

29

Image from http://http.developer.nvidia.com/CaTutorials/cg_tutorial_chapter08.html

Vertex Shader

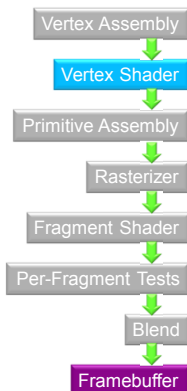


- What varies per-vertex and what does not?

```
float displacement =
    u_scaleFactor * 0.5 *
    (sin(position.y *
    u_frequency * u_time) +
    1.0);
```

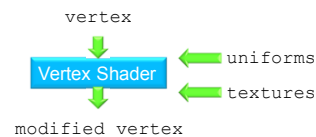
30

Vertex Shader



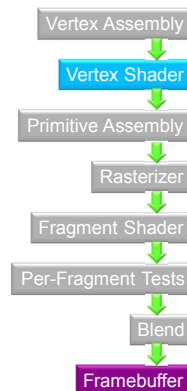
- On all modern GPUs, vertex shaders can read from textures as well as uniform variables.

- What is this useful for?

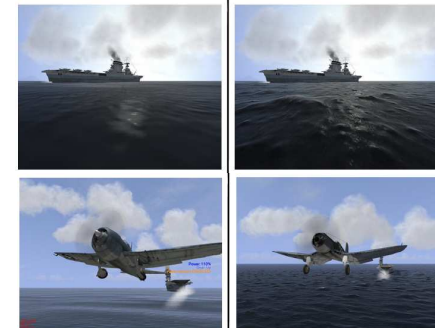


31

Vertex Shader



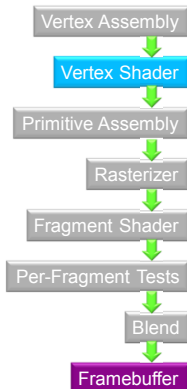
- Example: Textures can provide height maps for displacement mapping



32

Images from <http://developer.nvidia.com/content/vertex-texture-fetch>

Vertex Shader



- Technology preview: vertex shaders are becoming available to CSS on the web as *CSS shaders*

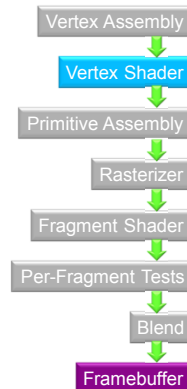
Demo

<http://www.adobe.com/devnet/html5/articles/css-shaders.html>

33

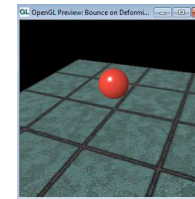
More info on CSS shaders: <https://d3cs.w3.org/hy/EXTF/aw-file/tip/custom/index.html>

Vertex Shader



- RenderMonkey Demos

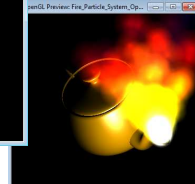
Bounce



Morph



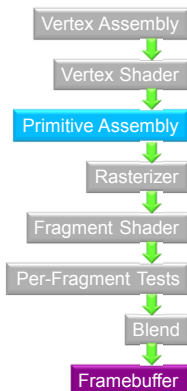
Particle System



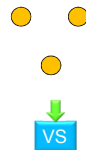
34

RenderMonkey: <http://developer.amd.com/archive/oc/rendermonkey/pages/default.aspx>

Primitive Assembly

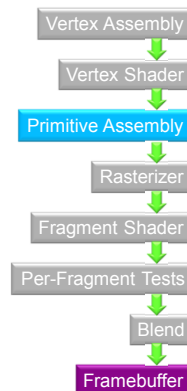


- A vertex shader processes one vertex. *Primitive assembly* groups vertices forming one primitive, e.g., a triangle, line, etc.



35

Primitive Assembly

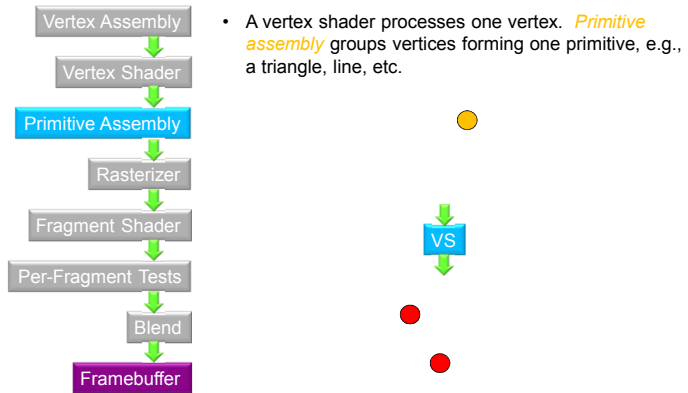


- A vertex shader processes one vertex. *Primitive assembly* groups vertices forming one primitive, e.g., a triangle, line, etc.



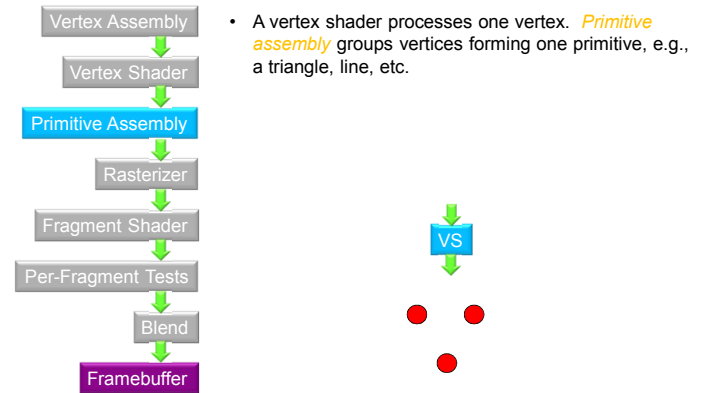
36

Primitive Assembly



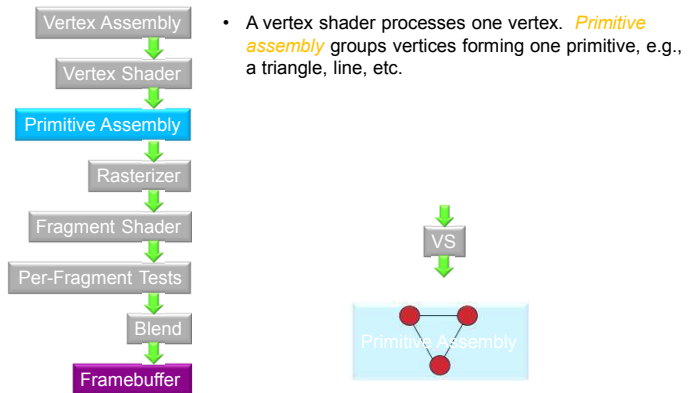
37

Primitive Assembly



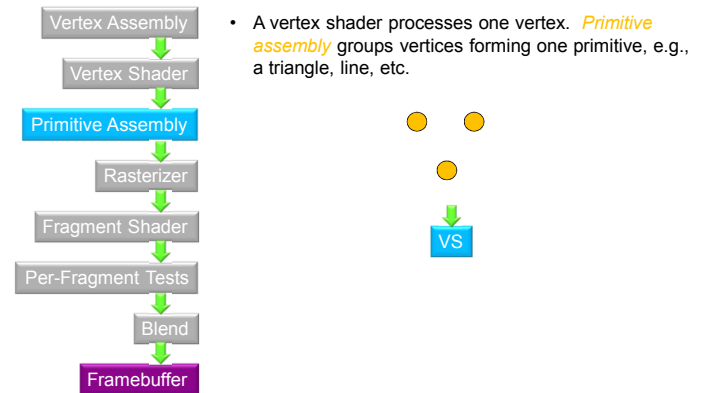
38

Primitive Assembly



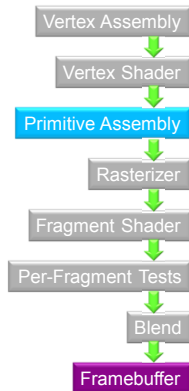
39

Primitive Assembly

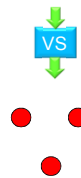


40

Primitive Assembly

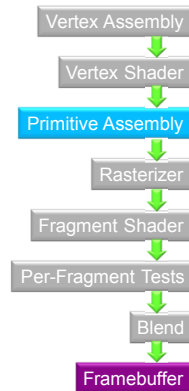


- A vertex shader processes one vertex. *Primitive assembly* groups vertices forming one primitive, e.g., a triangle, line, etc.

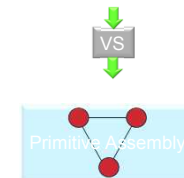


41

Primitive Assembly

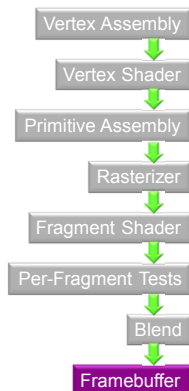


- A vertex shader processes one vertex. *Primitive assembly* groups vertices forming one primitive, e.g., a triangle, line, etc.



42

Perspective Division and Viewport Transform



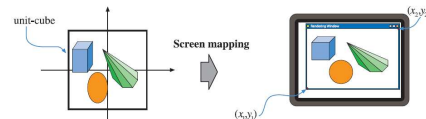
- There are a series of stages between primitive assembly and rasterization.

- Perspective division*

$$P_{ndc} = (P_{clip}) \cdot xyz / (P_{clip}) \cdot w$$

- Viewport transform*

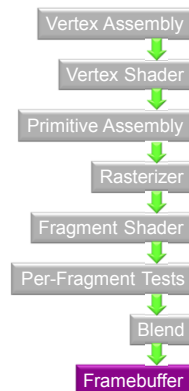
$$P_{window} = (M_{viewport-transform}) (P_{ndc})$$



43

Image from <http://www.realtimerendering.com/>

Clipping



- There are a series of stages between primitive assembly and rasterization.

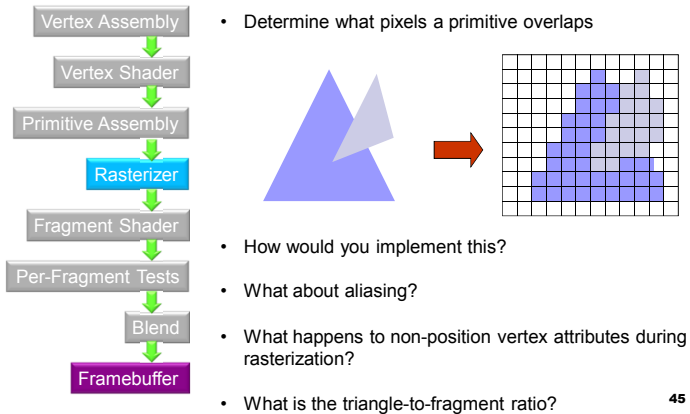
- Clipping*



44

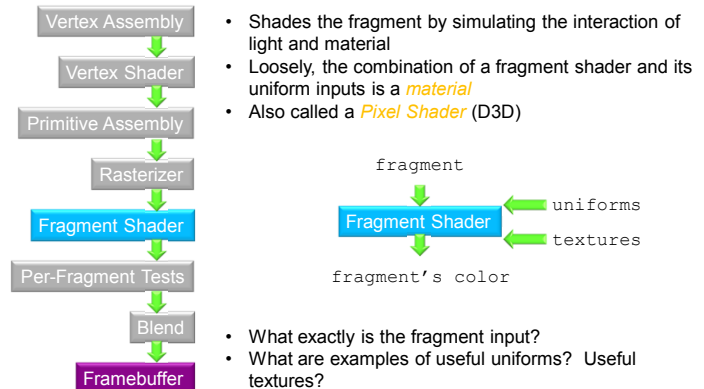
Image from <http://www.realtimerendering.com/>

Rasterization



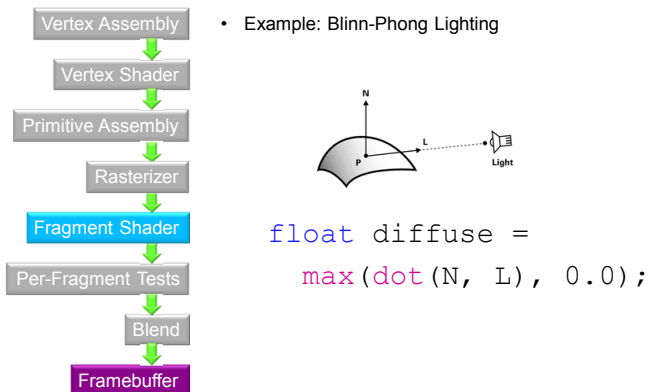
45

Fragment Shader



46

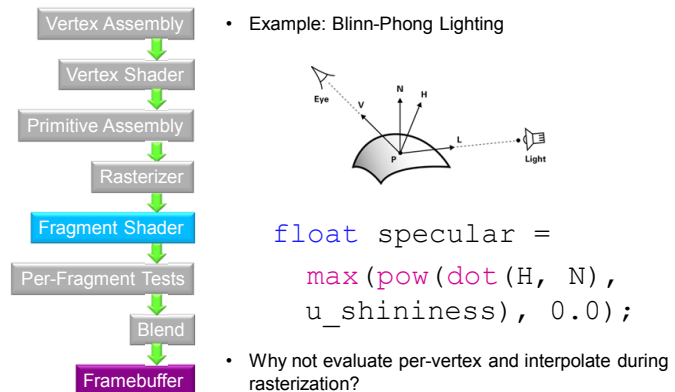
Fragment Shader



47

Image from http://http.developer.nvidia.com/CgTutorial/cg_tutorial_chapter05.html

Fragment Shader



48

Image from http://http.developer.nvidia.com/CgTutorial/cg_tutorial_chapter05.html

Fragment Shader

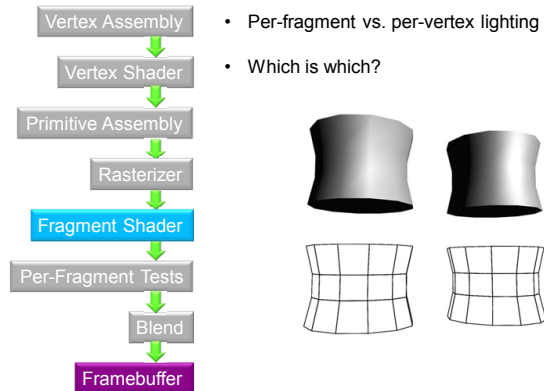


Image from http://http.developer.nvidia.com/CgTutorial/cg_tutorial_chapter05.html

49

Fragment Shader

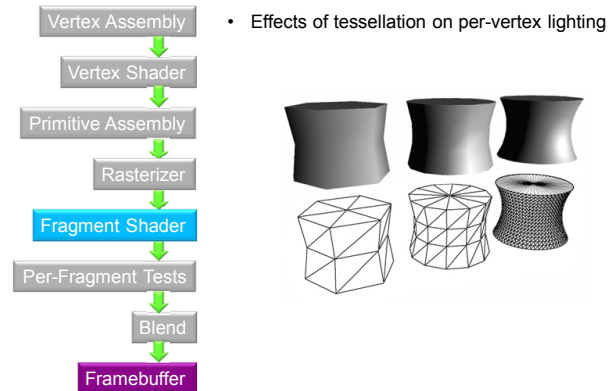


Image from http://http.developer.nvidia.com/CgTutorial/cg_tutorial_chapter05.html

50

Fragment Shader

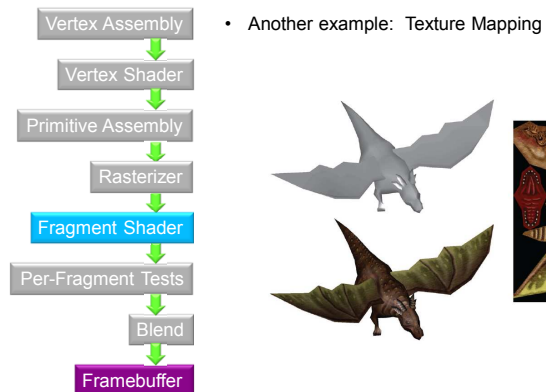


Image from <http://www.realtimerendering.com/>

51

Fragment Shader

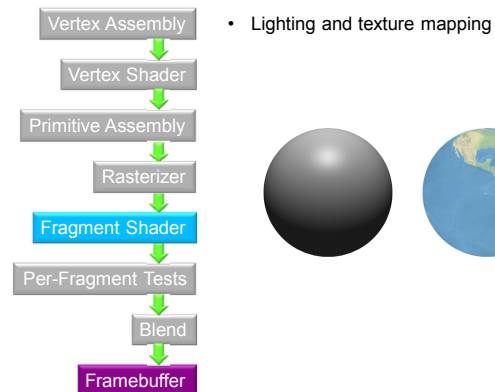
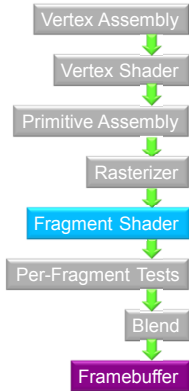


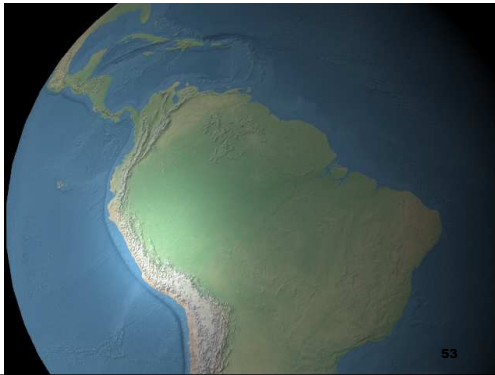
Image from <http://www.virtualglobebook.com/>

52

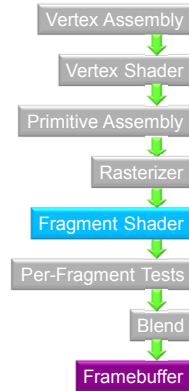
Fragment Shader



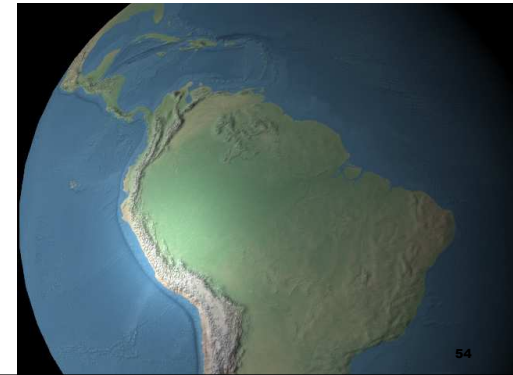
- Another example: Bump mapping



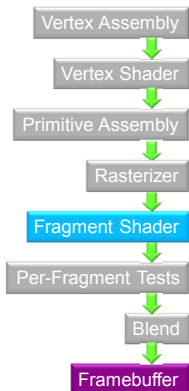
Fragment Shader



- Another example: Bump mapping



Fragment Shader



- Fragment shaders can be computationally intense

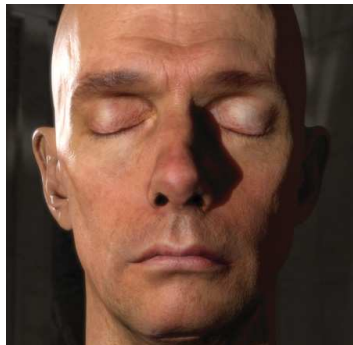
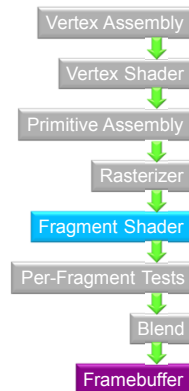


Image from http://http.developer.nvidia.com/GPUGems3/gpugems3_ch14.html

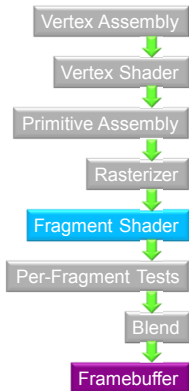
Fragment Shader



- A fragment shader can output color, but what else would be useful?

56

Fragment Shader

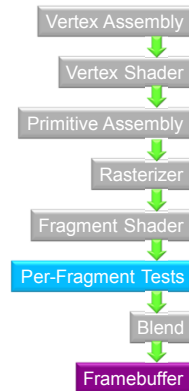


- A fragment shader can output color, but what else would be useful?

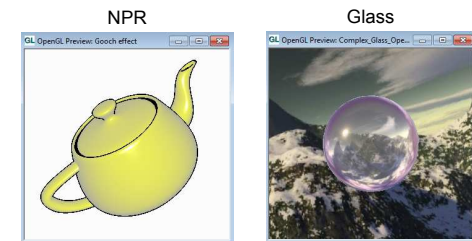
- Discard the fragment. Why?
- Depth. Why?
- Multiple colors. Why?

57

Fragment Shader



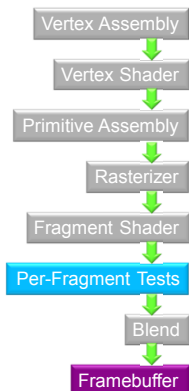
- RenderMonkey Demos



58

RenderMonkey: <http://developer.amd.com/archive/oc/rendermonkey/pages/default.aspx>

Per-Fragment Tests

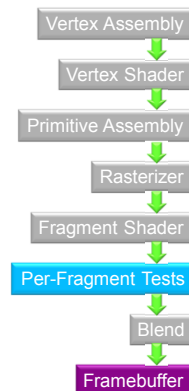


- A fragment must go through a series of tests to make to the framebuffer

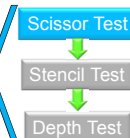
- What tests are useful?

59

Scissor Test

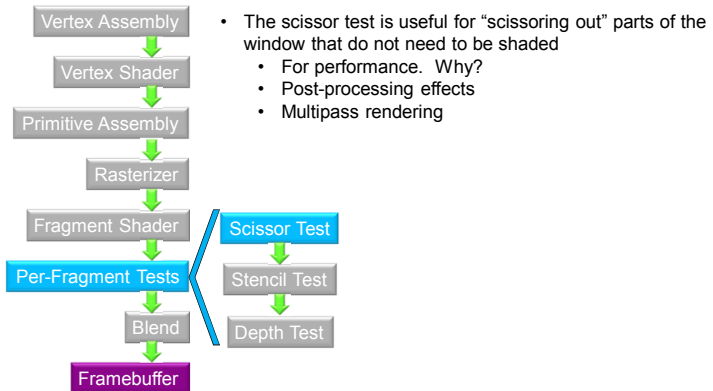


- Discard a fragment if it is within a rectangle defined in window coordinates
 - Why is this useful?
 - Does this need to happen after fragment shading?



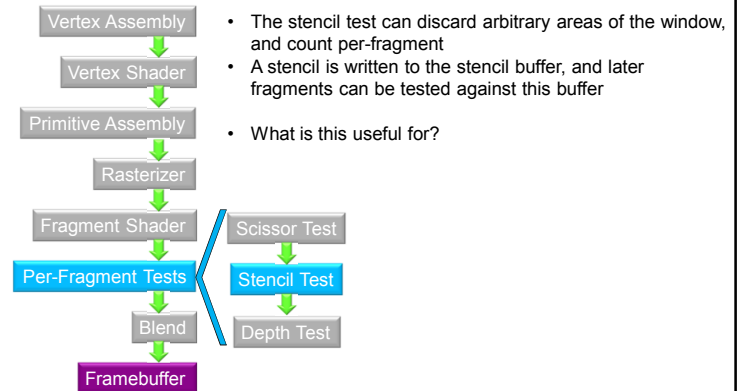
60

Scissor Test



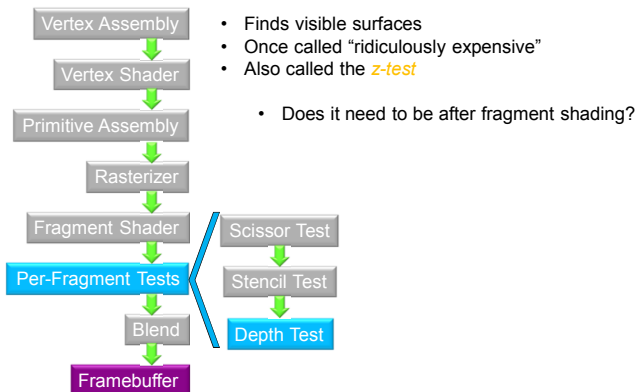
61

Stencil Test



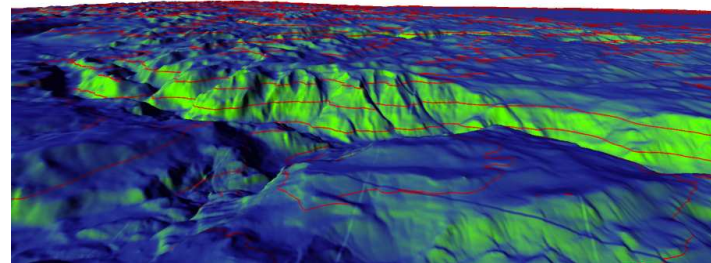
62

Depth Test



63

Depth Test



64

Image from <http://www.virtualglobebook.com/>

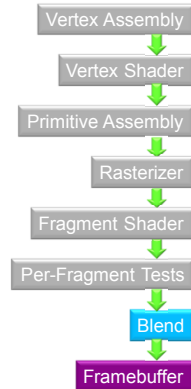
Depth Test



65

Image from <http://www.virtualjobbook.com/>

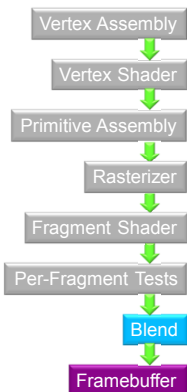
Blending



- Combine fragment color with framebuffer color
 - Can weight each color
 - Can use different operations: +, -, etc.
- Why is this useful?

66

Blending



- Example: Translucency

- Additive Blending

$$C_{dest} = (C_{source}.rgb) (C_{source}.a) + (C_{dest}.rgb);$$

- Alpha Blending

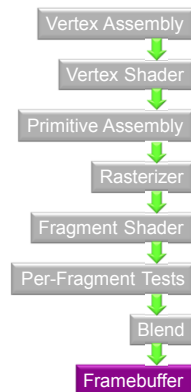
$$C_{dest} = (C_{source}.rgb) (C_{source}.a) + (C_{dest}.rgb) (1 - C_{source}.a);$$



67

Image from http://http.developer.nvidia.com/GPU/Gems/gpugems_ch06.html

Graphics Pipeline Walkthrough



- After all that, write to the framebuffer!



Output image buffer

68

Images from http://www.cs.cmu.edu/afs/cs.cmu.edu/academic/class/15869-f11/www/lectures/01_intro.pdf

Evolution of the Programmable Graphics Pipeline

- Pre GPU
- Fixed function GPU
- Programmable GPU
- Unified Shader Processors

69

Early 90s – Pre GPU



Wolfenstein 3D, 1992

Doom I, 1993

- Interactive software rendering (no GPUs yet)
- NOTE: SGI was building interactive rendering supercomputers, but this was beginning of interactive 3D graphics on PC

70

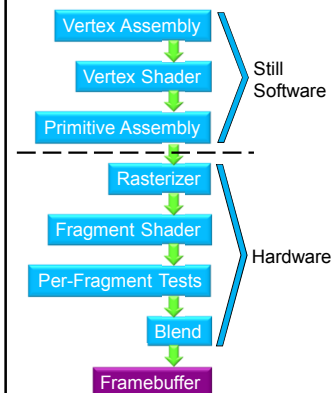
Slide from <http://s09.idav.ucdavis.edu/talks/01-BPS-SIGGRAPH09-mhouston.pdf>

Why GPUs?

- Exploit Parallelism
 - Pipeline parallel
 - Data-parallel
 - CPU and GPU executing in parallel
- Hardware: texture filtering, rasterization, MAD, sqrt, etc.

71

3dfx Voodoo (1996)



In hardware:

- Fixed-function rasterization, texture mapping, depth testing, etc.
- 4 - 6 MB memory
- PCI bus
- \$299



72

Image from <http://www.thedodgegarage.com/3dfx/v1.htm>

Aside: Mario Kart 64

- High fragment load / low vertex load



Image from http://www.gamespot.com/users/my_shoe/

73

Aside: Mario Kart Wii

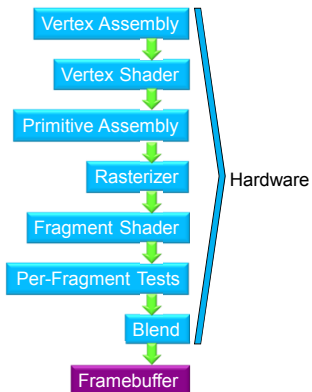
- High fragment load / low vertex load?



Image from <http://wii.ign.com/dor/objects/949580/mario-kart-wii/images/>

74

NVIDIA GeForce 256 (1999)



In hardware:

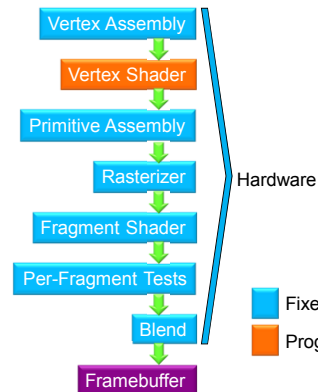
- Fixed-function vertex shading (T&L)
- Multi-texturing: bump maps, light maps, etc.
- 10 million polygons per second
- Direct3D 7
- AGP bus



Image from http://en.wikipedia.org/wiki/File:VisionTek_GeForce_256.jpg

75

NVIDIA GeForce 3 (2001)

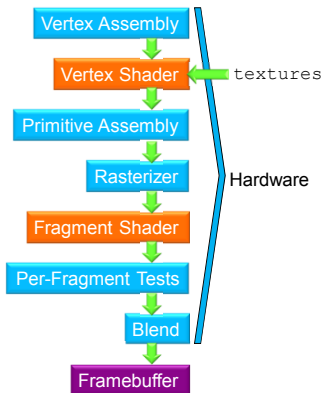


- Optionally bypass fixed-function T&L with a programmable vertex shader
- Optionally bypass fixed-function fragment shading with a programmable fragment shader
- Many programming limits
- Direct3D 8
- Pentium IV – 20 stages
- GeForce 3 – 600-800 stages

- Fixed-function stage
- Programmable stage

76

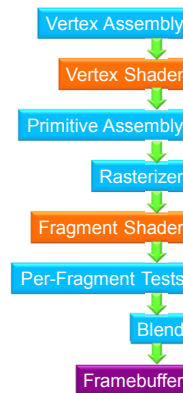
NVIDIA GeForce 6 (2004)



- Much better programmable fragment shaders
- Vertex shader can read textures
- Dynamic branches
- Multiple render targets
- PCIe bus
- OpenGL 2 / Direct3D 9

77

NVIDIA GeForce 6 (2004)



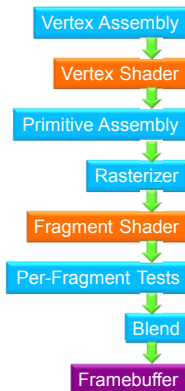
- Vertex shader can read textures
- Dynamic branches
- Multiple render targets
- PCIe bus
- OpenGL 2 / Direct3D 9



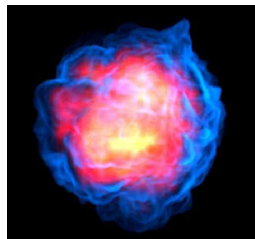
78

Image from http://download.nvidia.com/developer/presentations/2004/GPU_Jackpot/Shader_Model_3.pdf

NVIDIA GeForce 6 (2004)



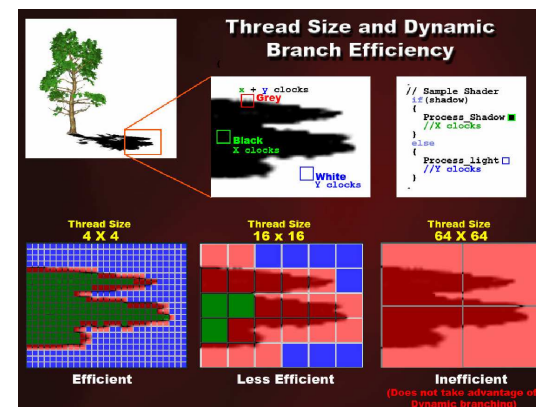
- Vertex shader can read textures
- Dynamic branches
- Multiple render targets
- PCIe bus
- OpenGL 2 / Direct3D 9



79

Image from http://download.nvidia.com/developer/presentations/2004/GPU_Jackpot/Shader_Model_3.pdf

Dynamic Branches



80

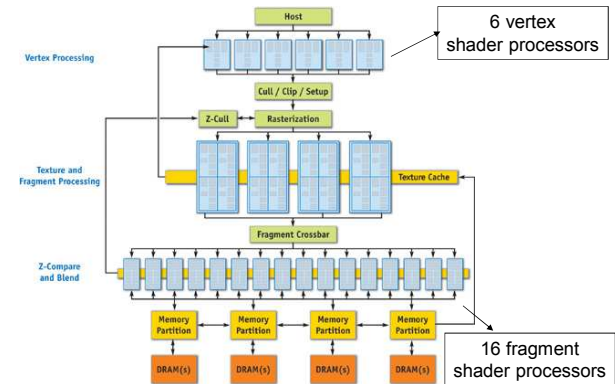
Image from http://developer.amd.com/media/gpu_assets/03_Clever_Shader_Tricks.pdf

Dynamic Branches

- For best performance, fragment shader dynamic branches should be coherent in screen-space
- How does this relate to warp partitioning in CUDA?

81

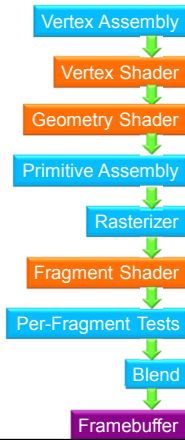
NVIDIA GeForce 6 (2004)



82

Image from http://http.developer.nvidia.com/GPUGems2/gpugems2_chapter30.htm

NVIDIA GeForce 8 (2006)



- Ground-up GPU redesign
- Geometry Shaders
- Transform-feedback
- OpenGL 3 / Direct3D 10
- Unified shader processors
- Support for GPU Compute

83

Geometry Shaders

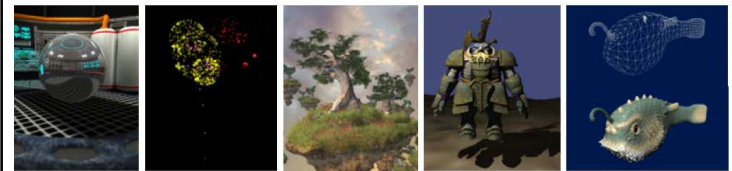
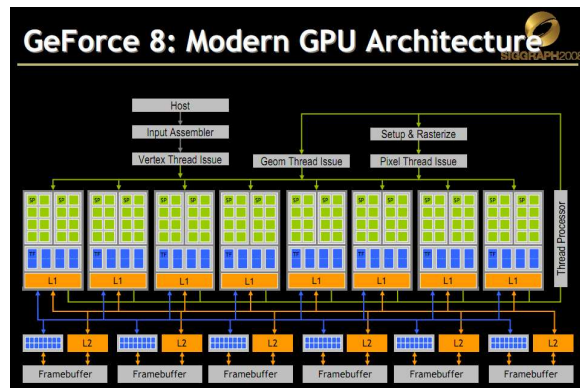


Figure 5: From left — render to cube map, particle system, instancing, shadow volume, displacement mapping.

84

Image from David Blythe : http://download.microsoft.com/download/f2/d/f2d5ee2c-b7ba-4cd0-9686-b6508b5479a1/direct3d10_web.pdf

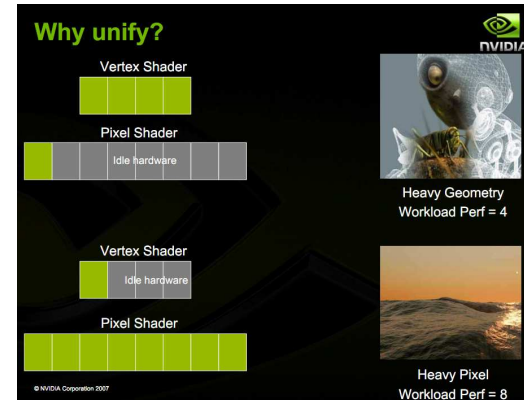
NVIDIA G80 Architecture



85

Slide from [http://s08.idav.ucdavis.edu/~lehke-nvidia-gpu-architecture.pdf](http://s08.idav.ucdavis.edu/~lehke/nvidia-gpu-architecture.pdf)

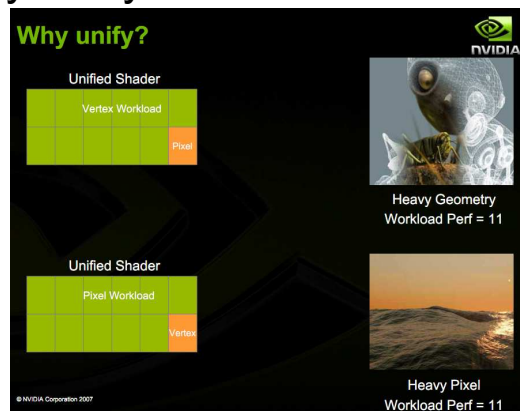
Why Unify Shader Processors?



86

Slide from <http://s08.idav.ucdavis.edu/~lehke-nvidia-gpu-architecture.pdf>

Why Unify Shader Processors?



87

Slide from <http://s08.idav.ucdavis.edu/~lehke-nvidia-gpu-architecture.pdf>

Terminology

| Shader Model | Direct3D | OpenGL | Video card Example |
|--------------|----------|--------|--|
| 3 | 9 | 2.x | NVIDIA GeForce 6800 ATI Radeon X800 |
| 4 | 10.x | 3.x | NVIDIA GeForce 8800 ATI Radeon HD 2900 |
| 5 | 11.x | 4.x | NVIDIA GeForce GTX 480 ATI Radeon HD 5870 |

88

Shader Capabilities

| | SM 2.0/2.X | SM 3.0 | SM 4.0 |
|---------------------------|-----------------------|----------------|--------------------|
| Introduced | DX 9.0, 2002 | DX 9.0c, 2004 | DX 10, 2007 |
| VS Instruction Slots | 256 | $\geq 512^a$ | 4096 |
| VS Max. Steps Executed | 65536 | 65536 | ∞ |
| PS Instruction Slots | $\geq 96^b$ | $\geq 512^a$ | $\geq 65536^a$ |
| PS Max. Steps Executed | $\geq 96^b$ | 65536 | ∞ |
| Temp. Registers | $\geq 12^a$ | 32 | 4096 |
| VS Constant Registers | $\geq 256^a$ | $\geq 256^a$ | 14×4096^c |
| PS Constant Registers | 32 | 224 | 14×4096^c |
| Flow Control, Predication | Optional ^d | Yes | Yes |
| VS Textures | None | 4 ^e | 128×512^f |
| PS Textures | 16 | 16 | 128×512^f |
| Integer Support | No | No | Yes |
| VS Input Registers | 16 | 16 | 16 |
| Interpolator Registers | 8 ^g | 10 | $16/32^h$ |
| PS Output Registers | 4 | 4 | 8 |

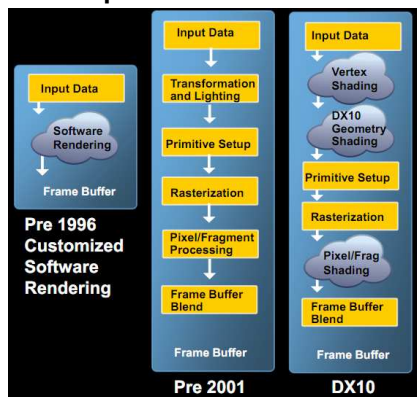
89
Table courtesy of A K Peters, Ltd. <http://www.realtimerendering.com/>

Shader Capabilities

| | SM 2.0/2.X | SM 3.0 | SM 4.0 |
|---------------------------|-----------------------|----------------|--------------------|
| Introduced | DX 9.0, 2002 | DX 9.0c, 2004 | DX 10, 2007 |
| VS Instruction Slots | 256 | $\geq 512^a$ | 4096 |
| VS Max. Steps Executed | 65536 | 65536 | ∞ |
| PS Instruction Slots | $\geq 96^b$ | $\geq 512^a$ | $\geq 65536^a$ |
| PS Max. Steps Executed | $\geq 96^b$ | 65536 | ∞ |
| Temp. Registers | $\geq 12^a$ | 32 | 4096 |
| VS Constant Registers | $\geq 256^a$ | $\geq 256^a$ | 14×4096^c |
| PS Constant Registers | 32 | 224 | 14×4096^c |
| Flow Control, Predication | Optional ^d | Yes | Yes |
| VS Textures | None | 4 ^e | 128×512^f |
| PS Textures | 16 | 16 | 128×512^f |
| Integer Support | No | No | Yes |
| VS Input Registers | 16 | 16 | 16 |
| Interpolator Registers | 8 ^g | 10 | $16/32^h$ |
| PS Output Registers | 4 | 4 | 8 |

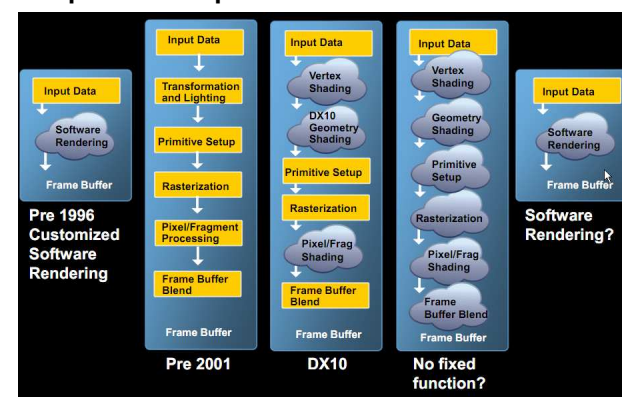
90
Table courtesy of A K Peters, Ltd. <http://www.realtimerendering.com/>

Evolution of the Programmable Graphics Pipeline



91
Slide from Mike Houston: <http://s09.idav.ucdavis.edu/talks/01-BPS-SIGGRAPH09-mhouston.pdf>

Evolution of the Programmable Graphics Pipeline



92
Slide from Mike Houston: <http://s09.idav.ucdavis.edu/talks/01-BPS-SIGGRAPH09-mhouston.pdf>