CS 152
Computer Architecture and Engineering

Lecture 22 – Graphics Processors

2006-11-14

John Lazzaro
(www.cs.berkeley.edu/~lazzaro)

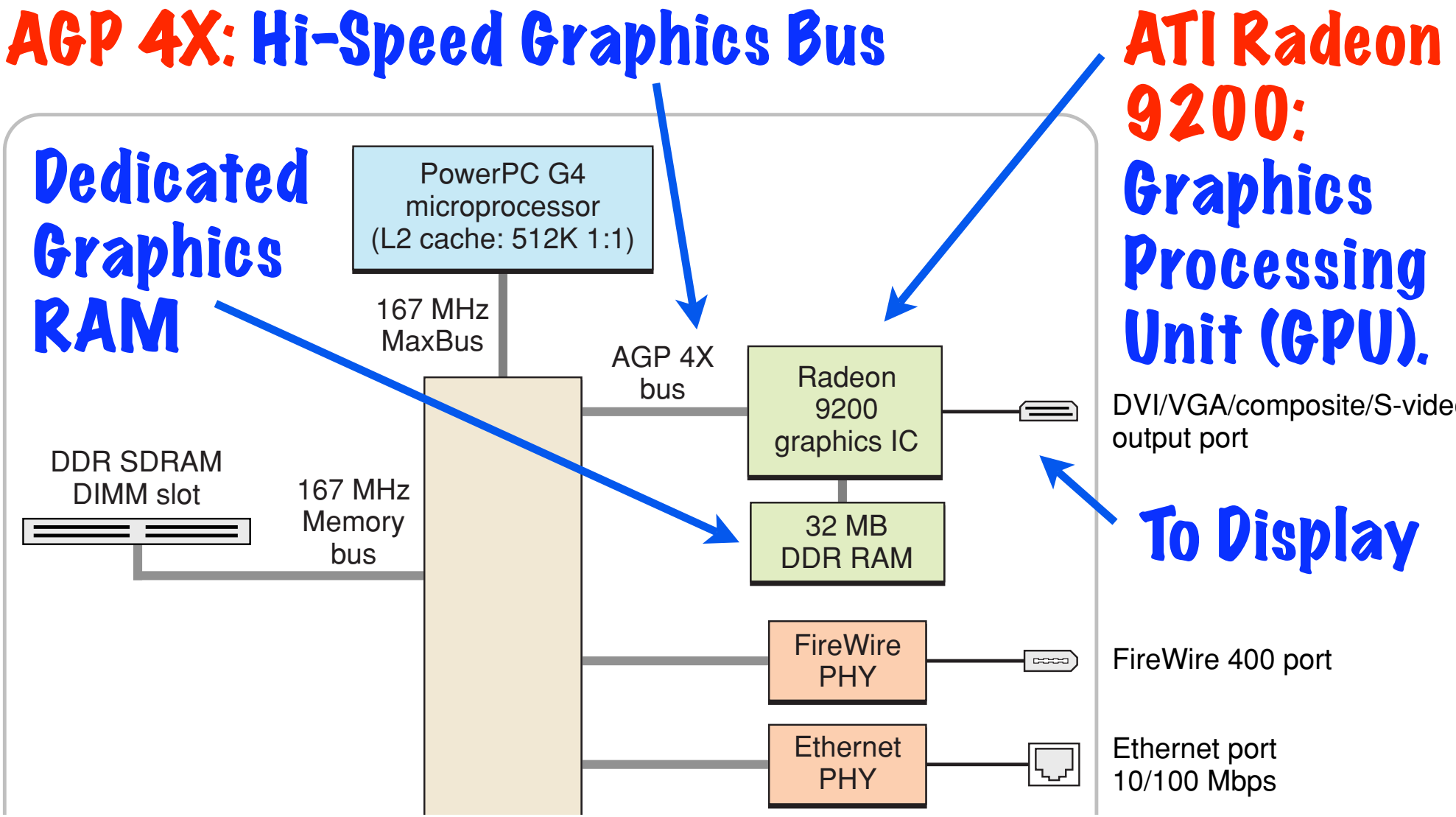
TAs: Udam Saini and Jue Sun

www-inst.eecs.berkeley.edu/~cs152/

Today: Graphics Processors

- **Computer Graphics.** A brief introduction to "the pipeline".
- **Stream Processing.** Casting the graphics pipeline into hardware.
- **Unified Pipelines.** GeForce 8800, the new architecture from Nvidia.

Recall: Mac Mini G4 System Diagram

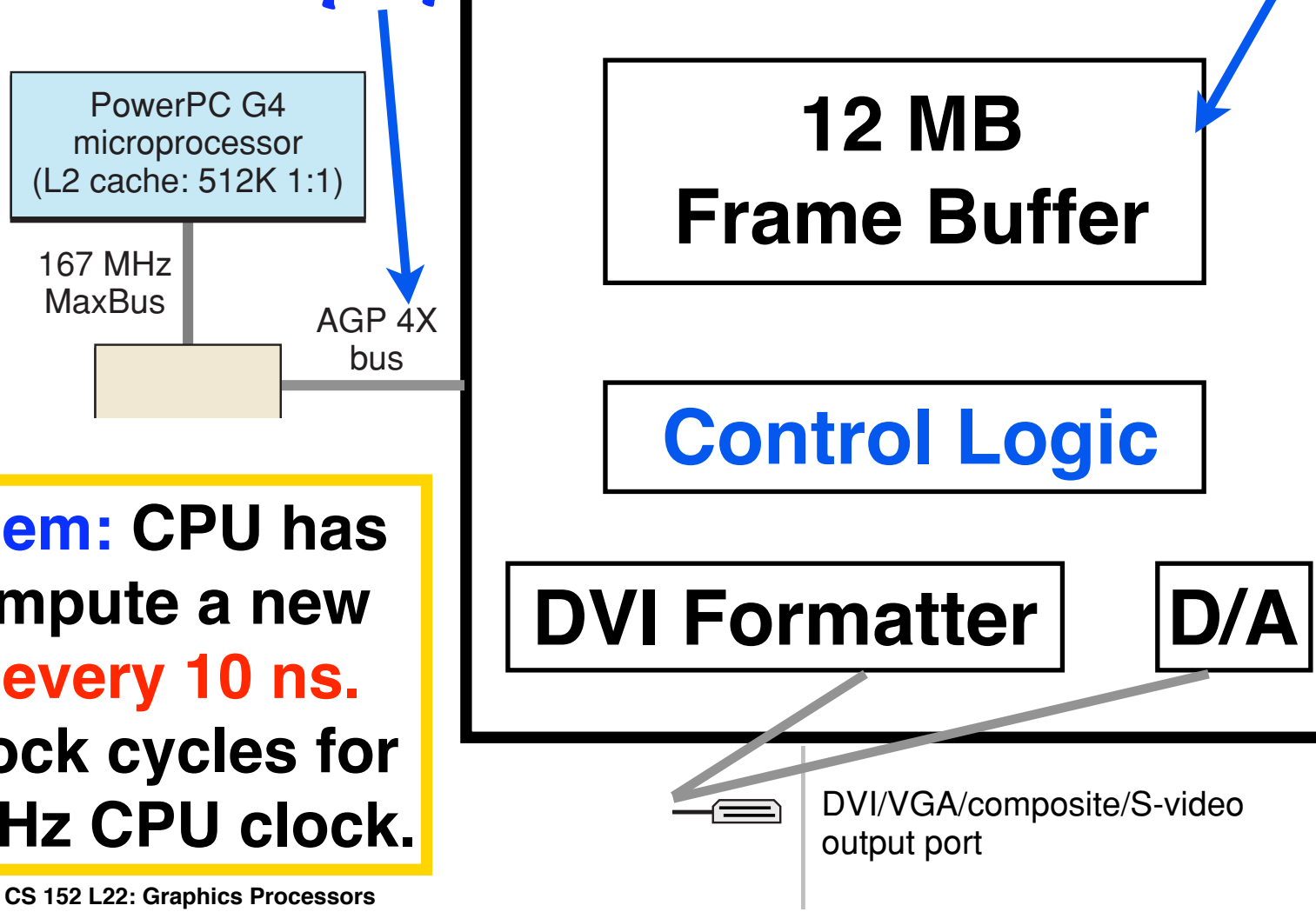


Average selling price (ASP) for GPUs: $30


2560
1600
About 12 MB/frame (24-bit pixels)
24 frames/sec: 300 MB/second

A "dumb" graphics card ...

AGP 4X: 1.1 GB/s. Can handle 24 f/s (300 MB/s) for a 2560x1600 display.



Double Buffering:

CPU writes "next frame" in one buffer.

Control logic sends "this frame" out of other buffer to display.

Problem: CPU has to compute a new pixel every 10 ns. 10 clock cycles for a 1 GHz CPU clock.

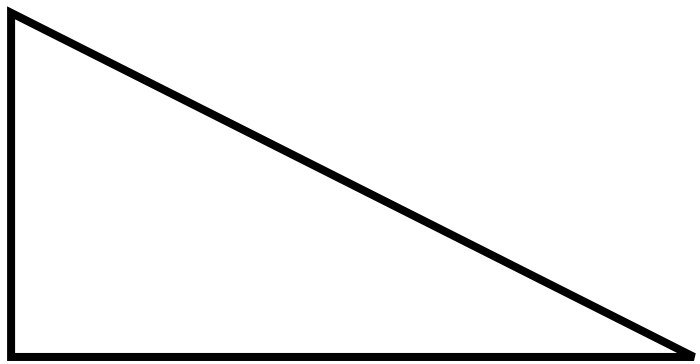
Q. What kind of graphics are we accelerating?
A. In 2006, interactive entertainment (3-D games). In the 1990s, 2-D acceleration (fast windowing systems, games like Pac-Man).

Graphics Acceleration

Q. In a multi-core world, why should we use a special processor for graphics?
A. Programmers generally use a certain coding style for graphics. We can design a processor to fit the style.

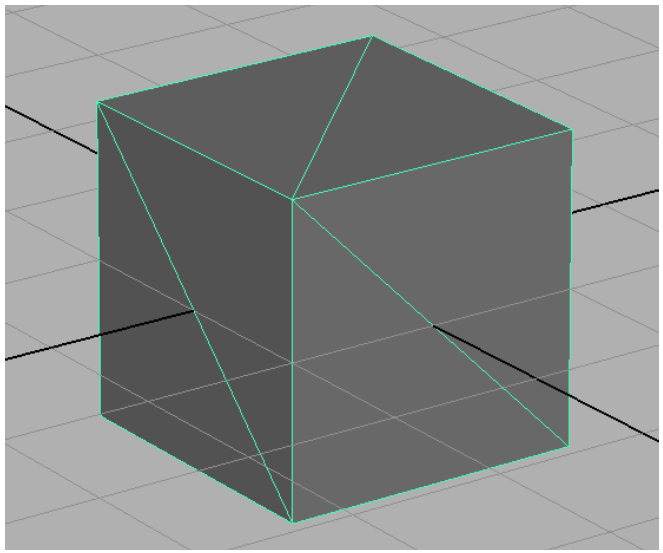
Next: An intro to 3-D graphics.

The Triangle ...



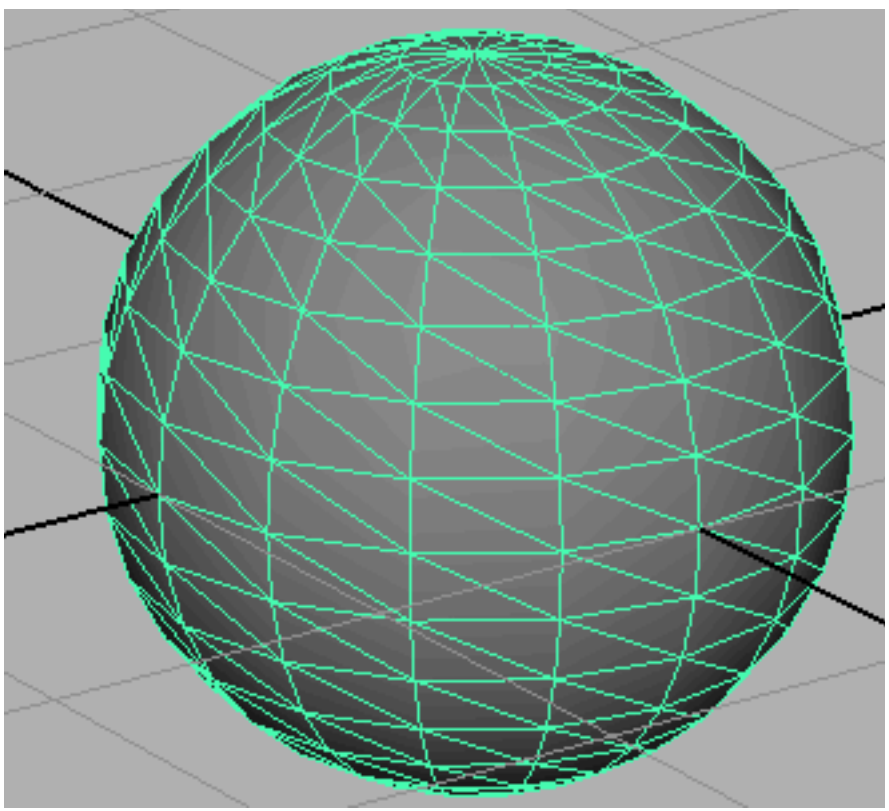
Simplest closed shape that may be defined by straight edges.

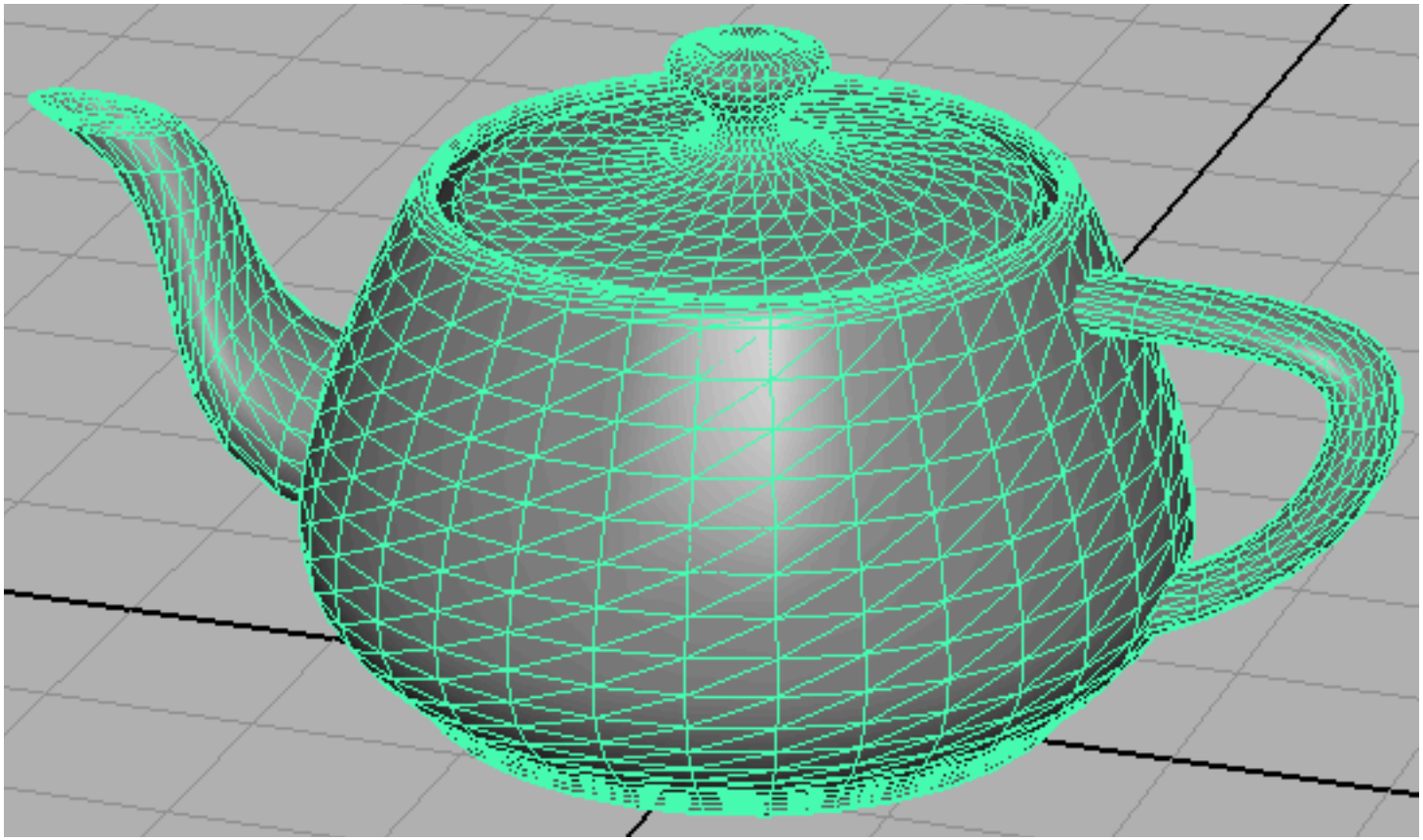
With enough triangles, you can make anything.



A cube whose faces are made up of triangles. This is a 3-D model of a cube -- model includes faces we can't see in this view.

A sphere whose faces are made up of triangles. With enough triangles, the curvature of the sphere can be made arbitrarily smooth.

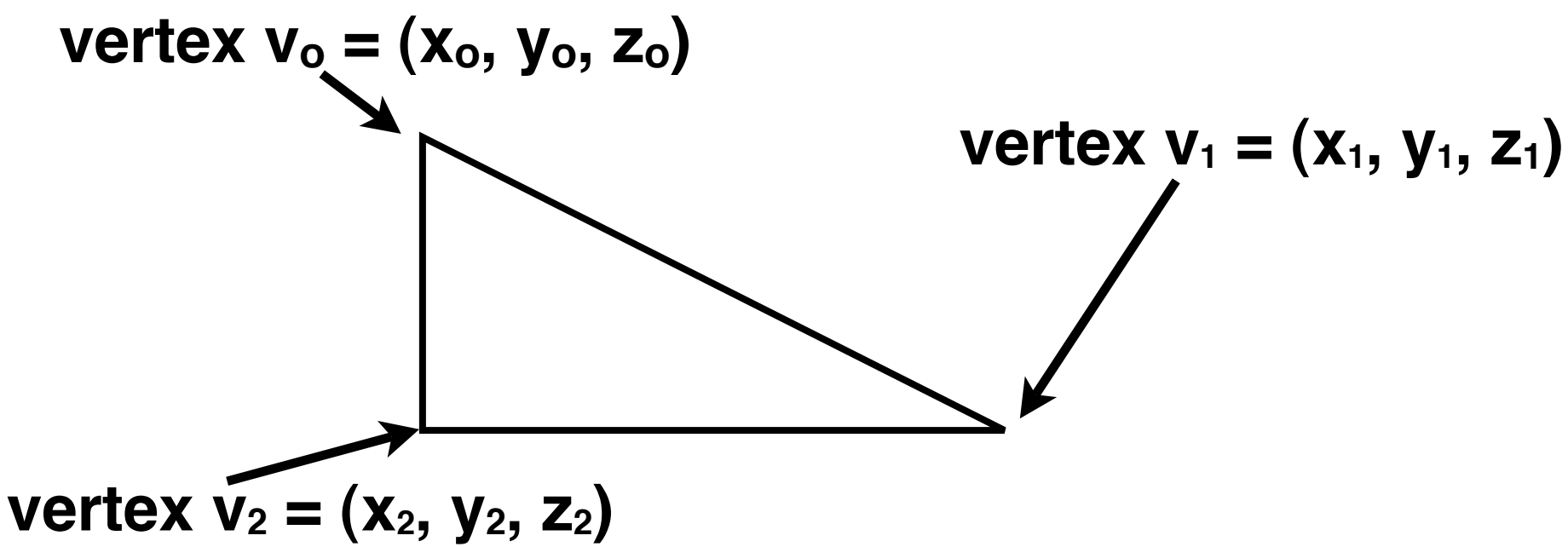




A teapot (famous object in computer graphics history). A "wire-frame" of triangles can capture the 3-D shape of complex, man-made objects.

Triangle defined by 3 vertices

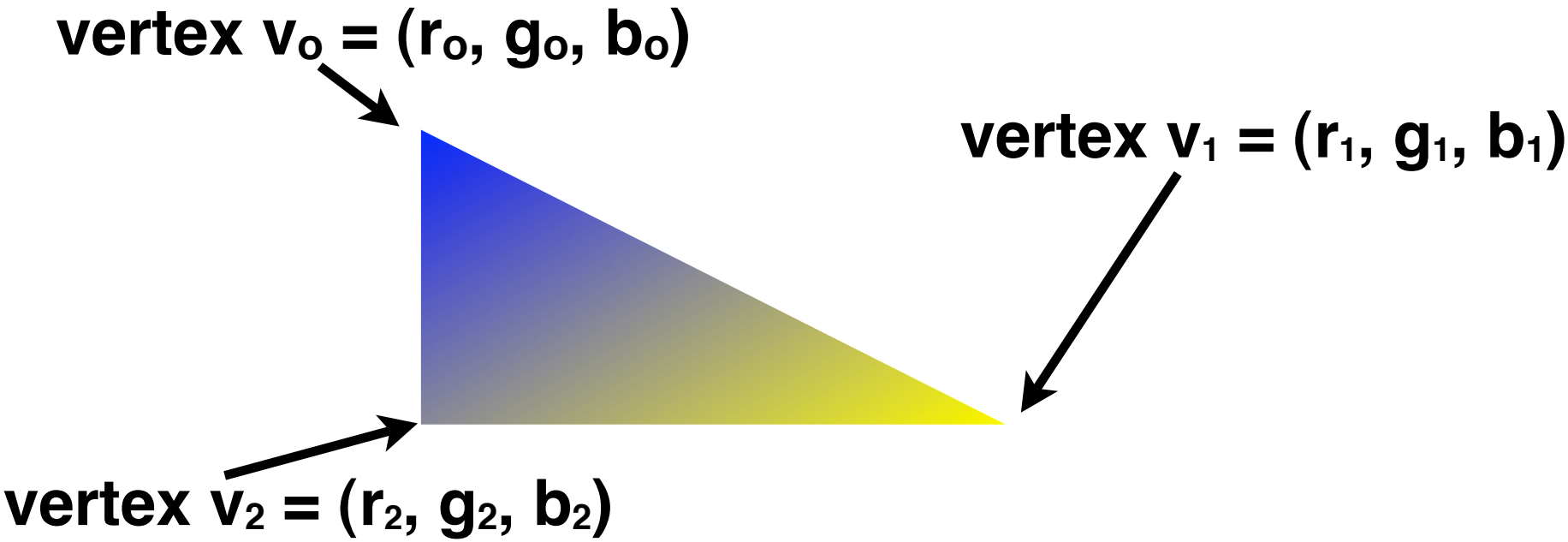
By transforming (v' = f(v)) all vertices in a 3-D object (like the teapot), you can move it in the 3-D world, change it's size, rotate it, etc.



If a teapot has 10,000 triangles, need to transform 30,000 vertices to move it in a 3-D scene ... per frame!

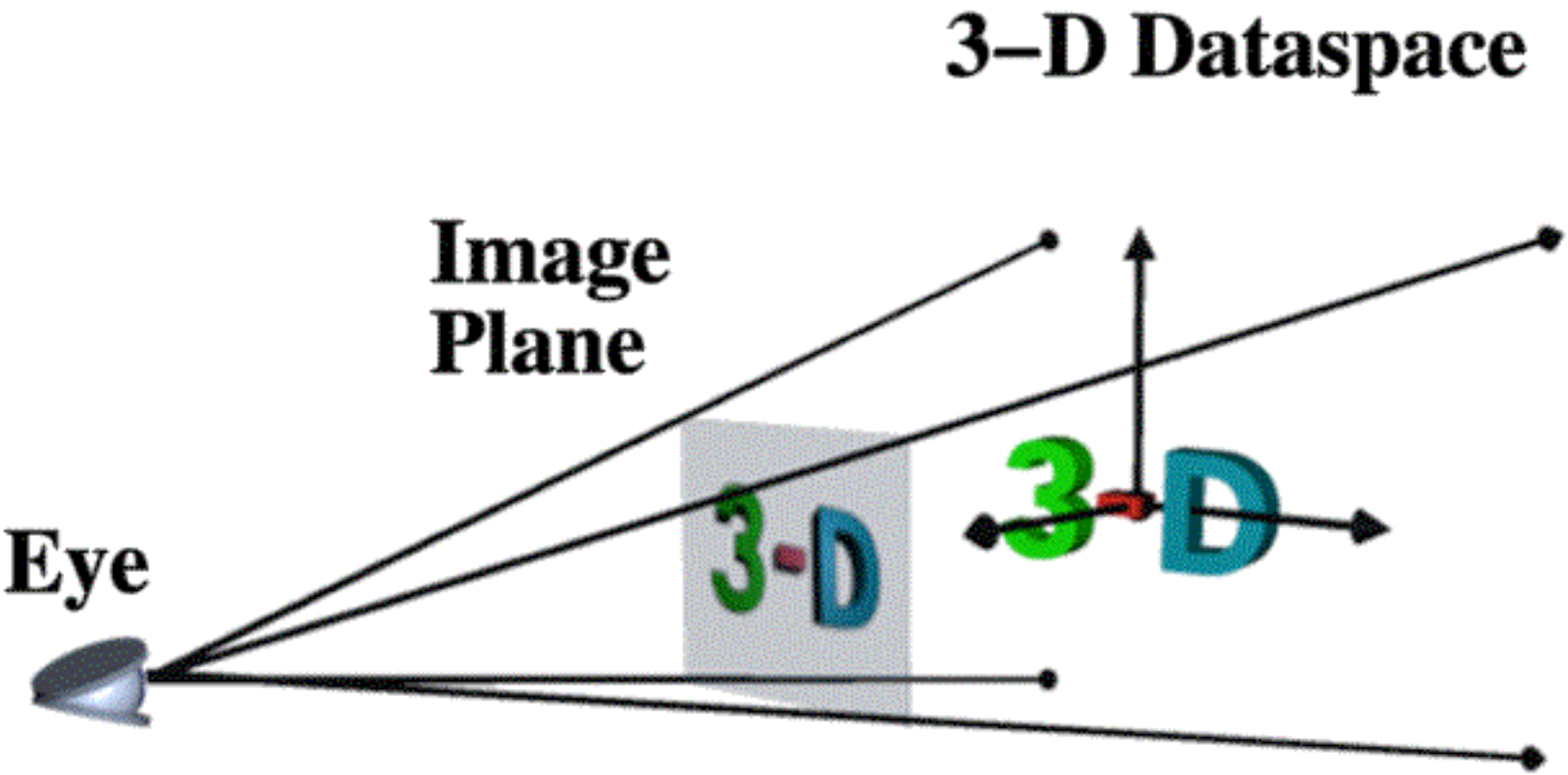
Vertex can have color, lighting info ...

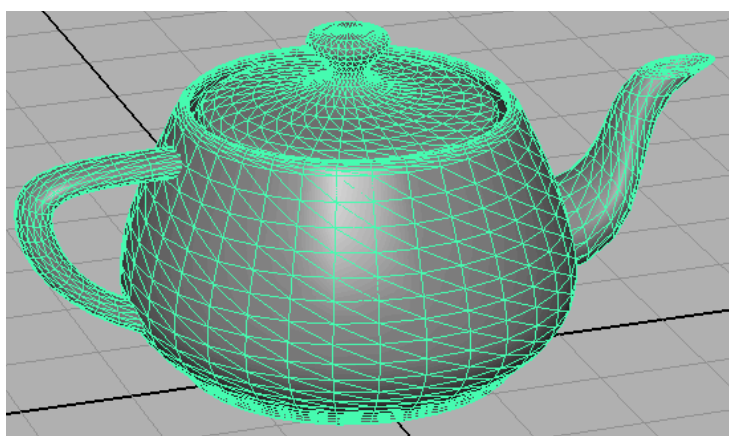
If vertices colors are different, this means that a smooth gradient of color washes across triangle.



More realistic graphics models include light sources in the scene. Per-vertex information can carry information about how light hits the vertex.

We see a 2-D window into the 3-D world

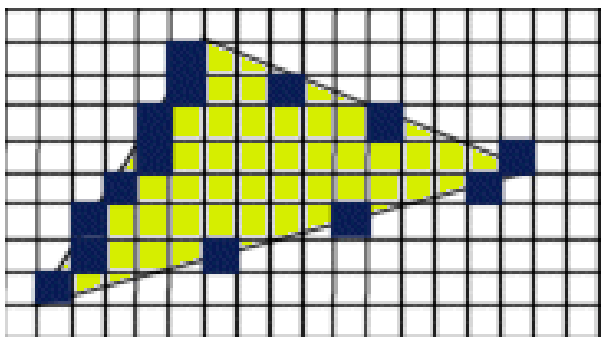




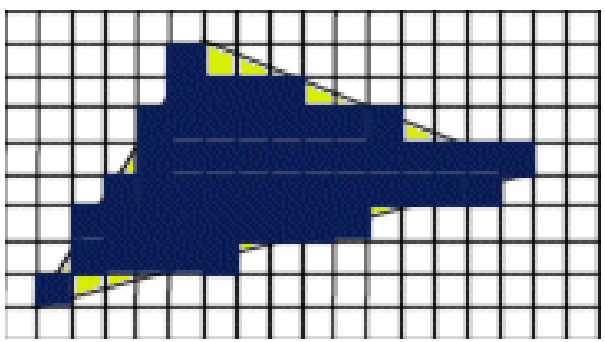
From 3-d triangles to screen pixels

First, project each 3-D triangle that might "face" the "eye" onto the image plane.

Then, create "pixel fragments" on the boundary of the image plane triangle



Then, create "pixel fragments" to fill in the triangle (rasterization).



Why "pixel fragments"? A screen pixel color might depend on many triangles (example: a glass teapot).

Process each fragment to "shade" it.

Algorithmic approach: Per-pixel computational model of metal and how light reflects off of it. Move teapot and what reflects off it changes.

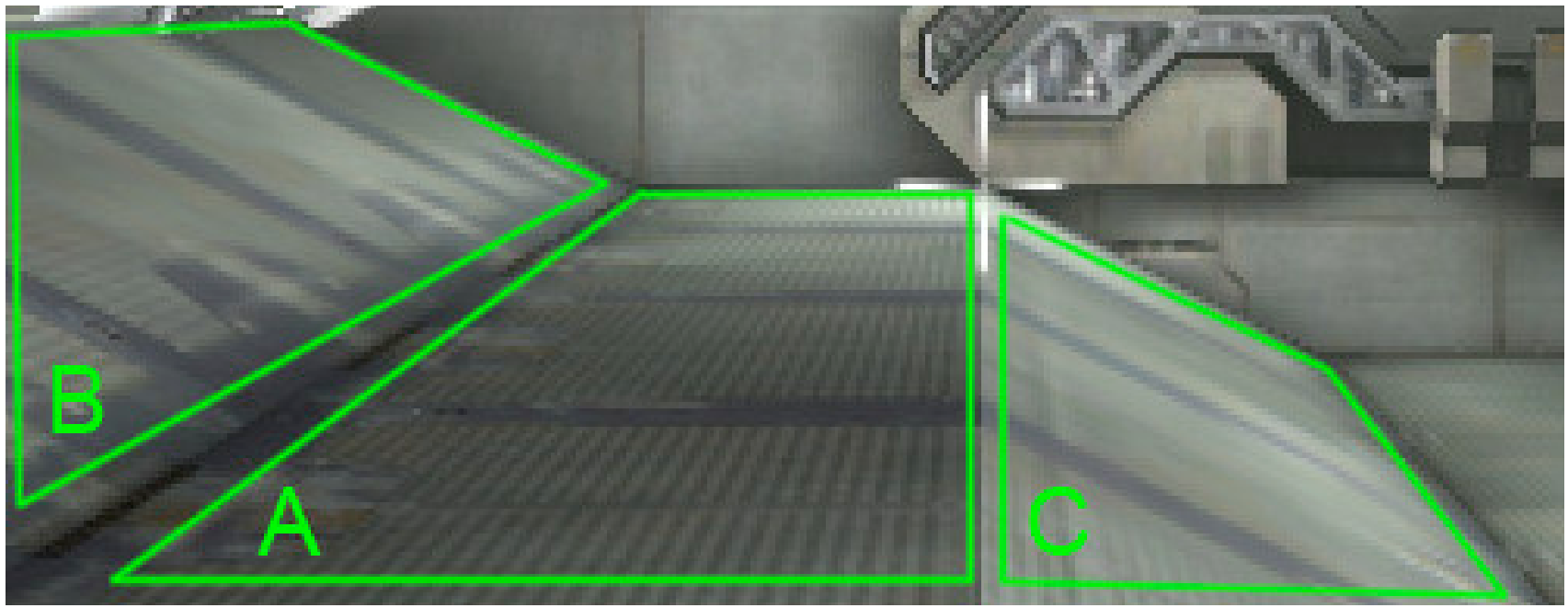
Process each fragment to "shade" it.

Artistic approach: Artist paints surface of teapot in Photoshop. We "map" this "texture" onto each pixel fragment during shading.

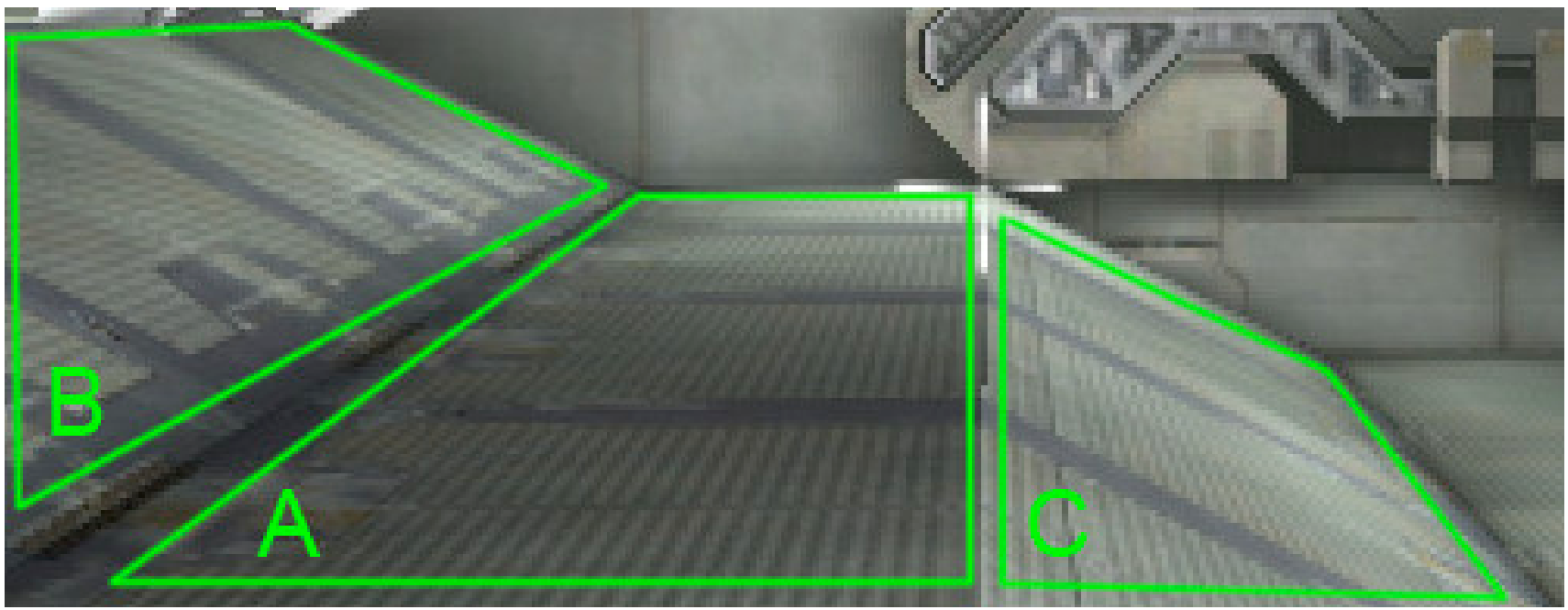
Final step: Output Merge. Assemble pixel fragments to make final 2-d image pixels.

Applying texture maps: Quality matters

"Good" algorithm. B and C look blurry.



"Better" algorithm. B and C are detailed.



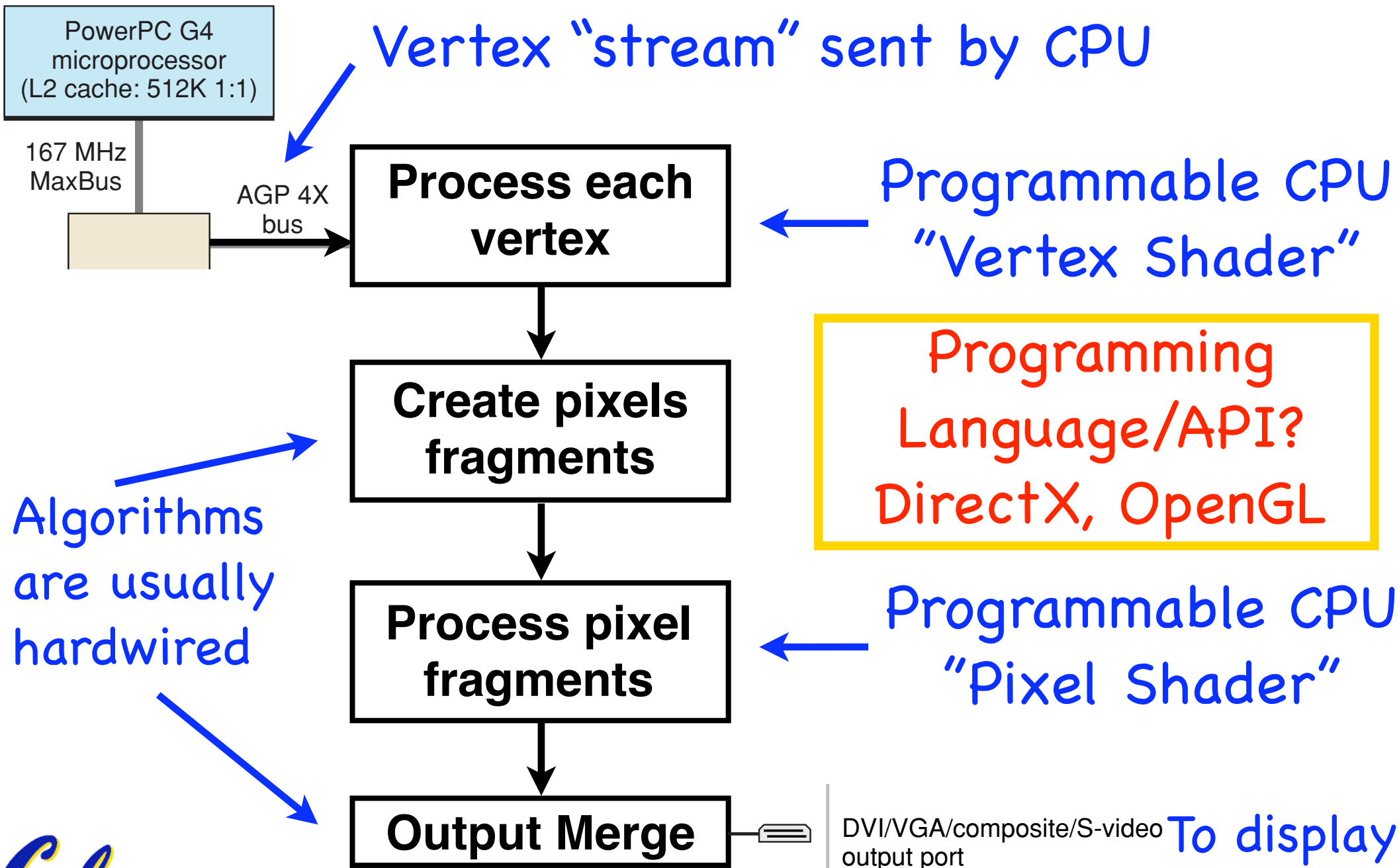
Putting it All Together ...

Luxo, Jr: Short movie made by Pixar, shown at SIGGRAPH in 1986.

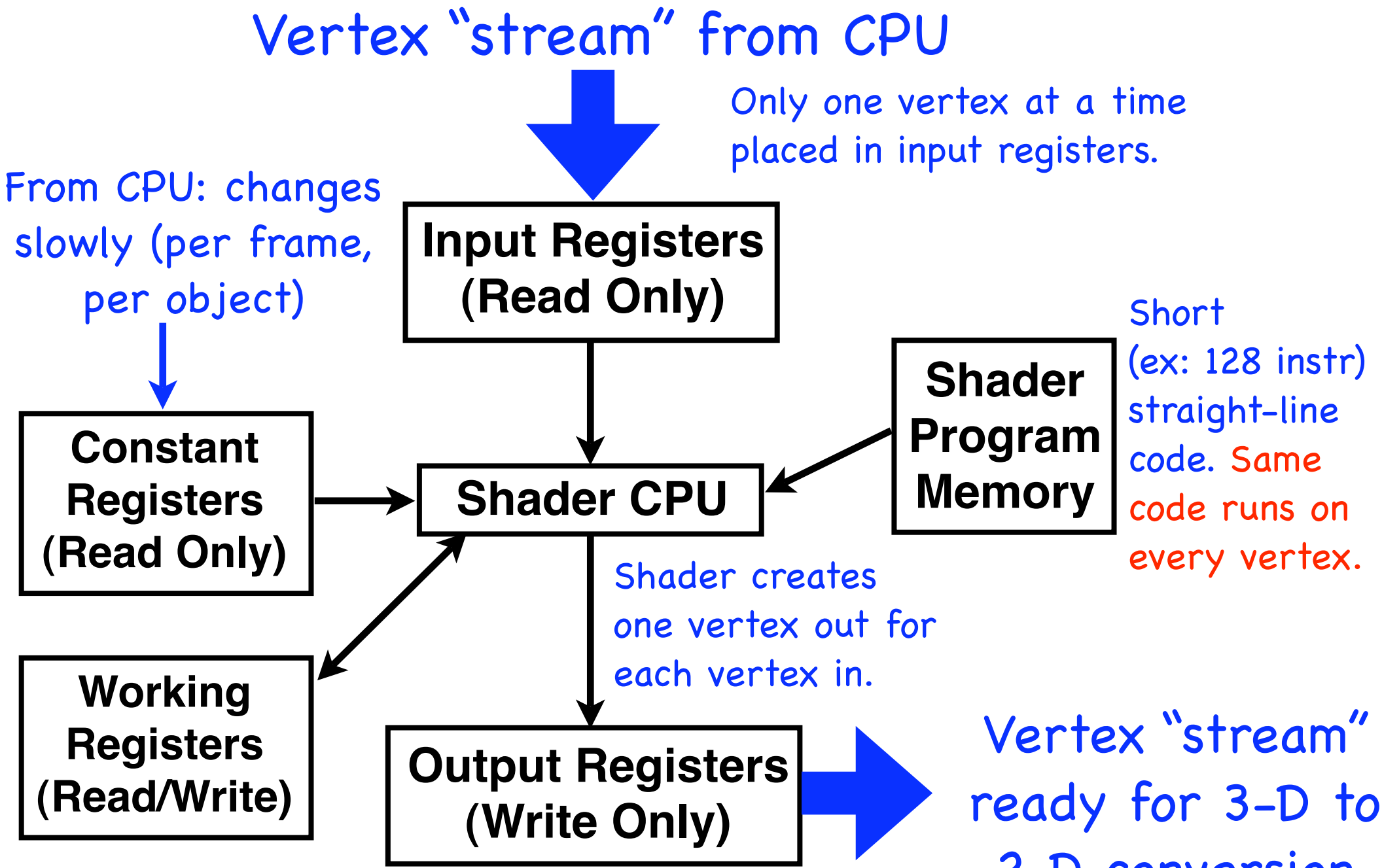
First Academy Award given to a computer graphics movie.


© PIXAR

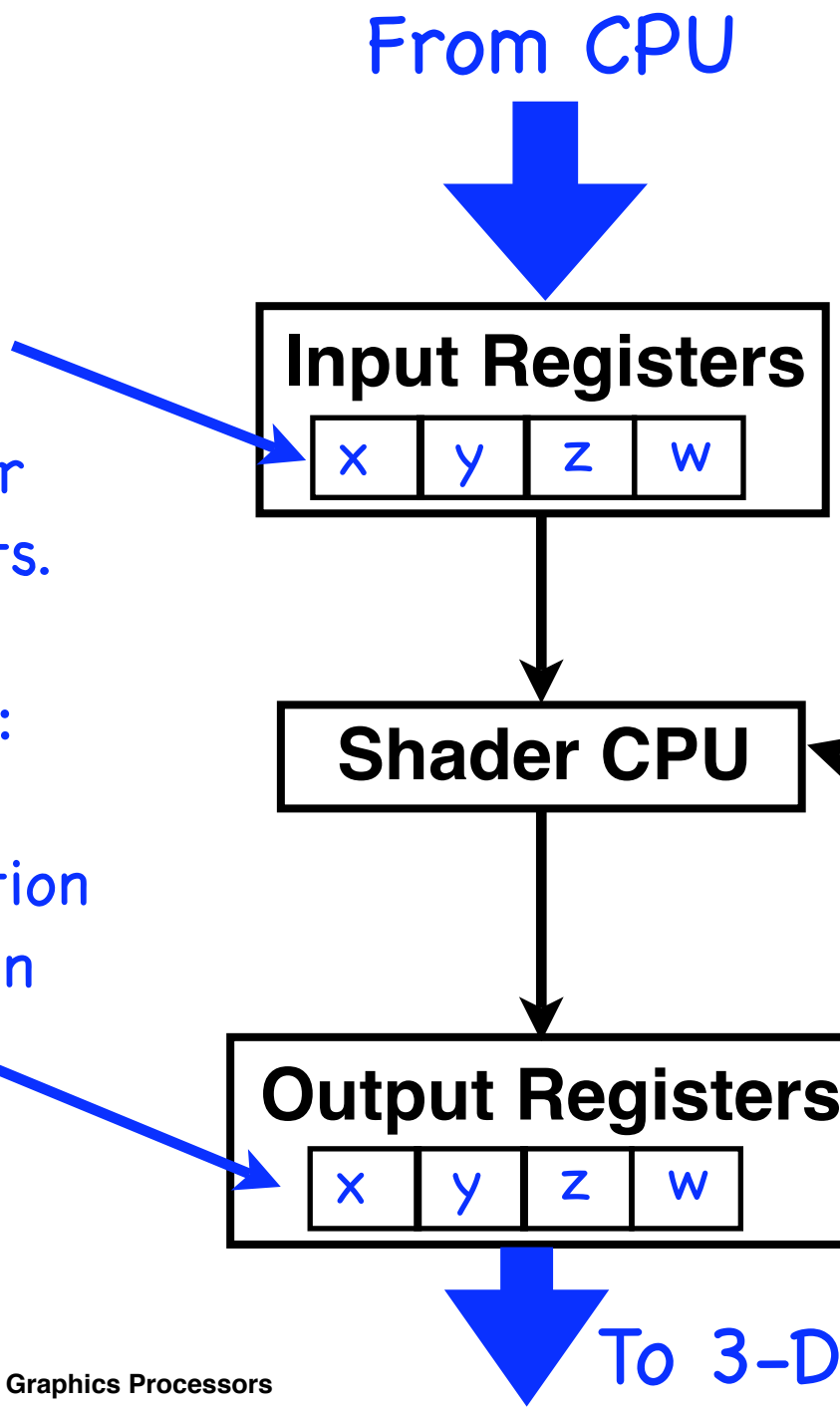
The graphics pipeline in hardware (2004)



Vertex Shader: A "stream processor"



Optimized instructions and data formats



128-bit registers, holding four 32-bit floats.

Typical use: (x,y,z,w) representation of a point in 3-D space.

Typical instruction:

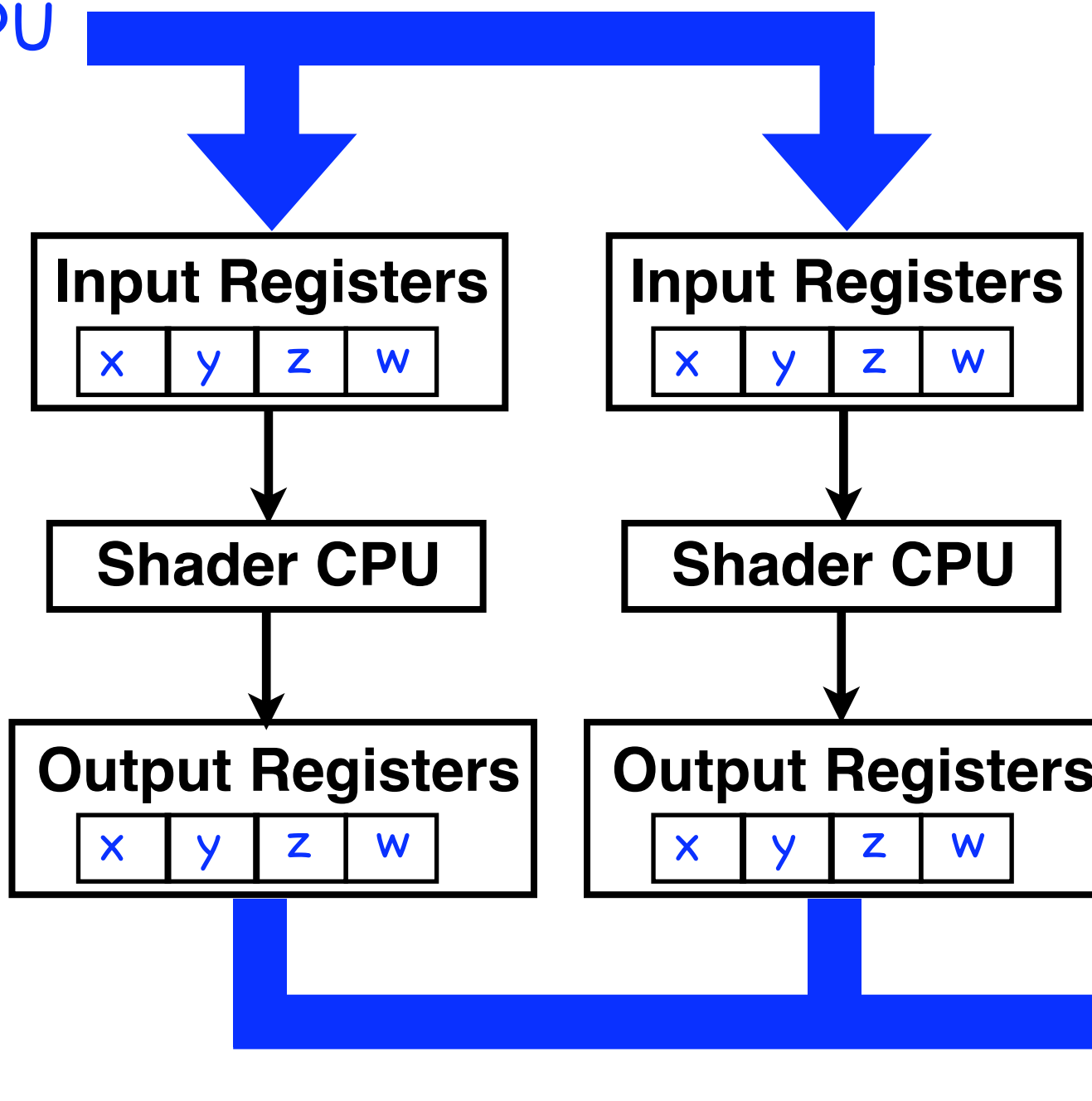
rsq dest src

dest.{x,y,z,w} = 1.0/sqrt(abs(src.w)).

If src.w=0, dest ∞.

The 1/sqrt() function is often used in graphics.

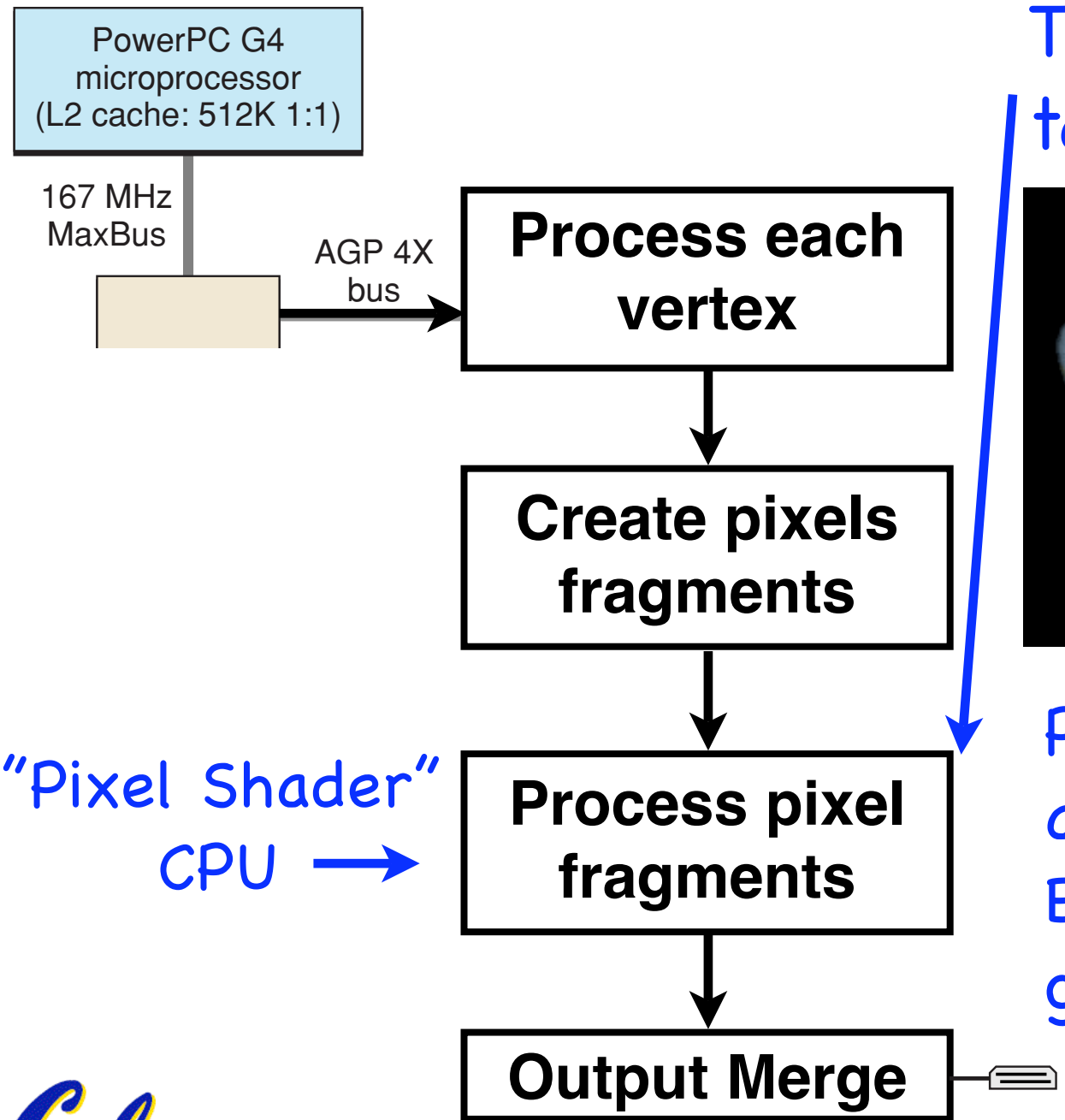
Easy to parallelize: Vertices independent



Caveat: Care might be needed when merging streams.

Why? 3-D to 2-D may expect triangle vertices in order in the stream.

Pixel shader specializations ...

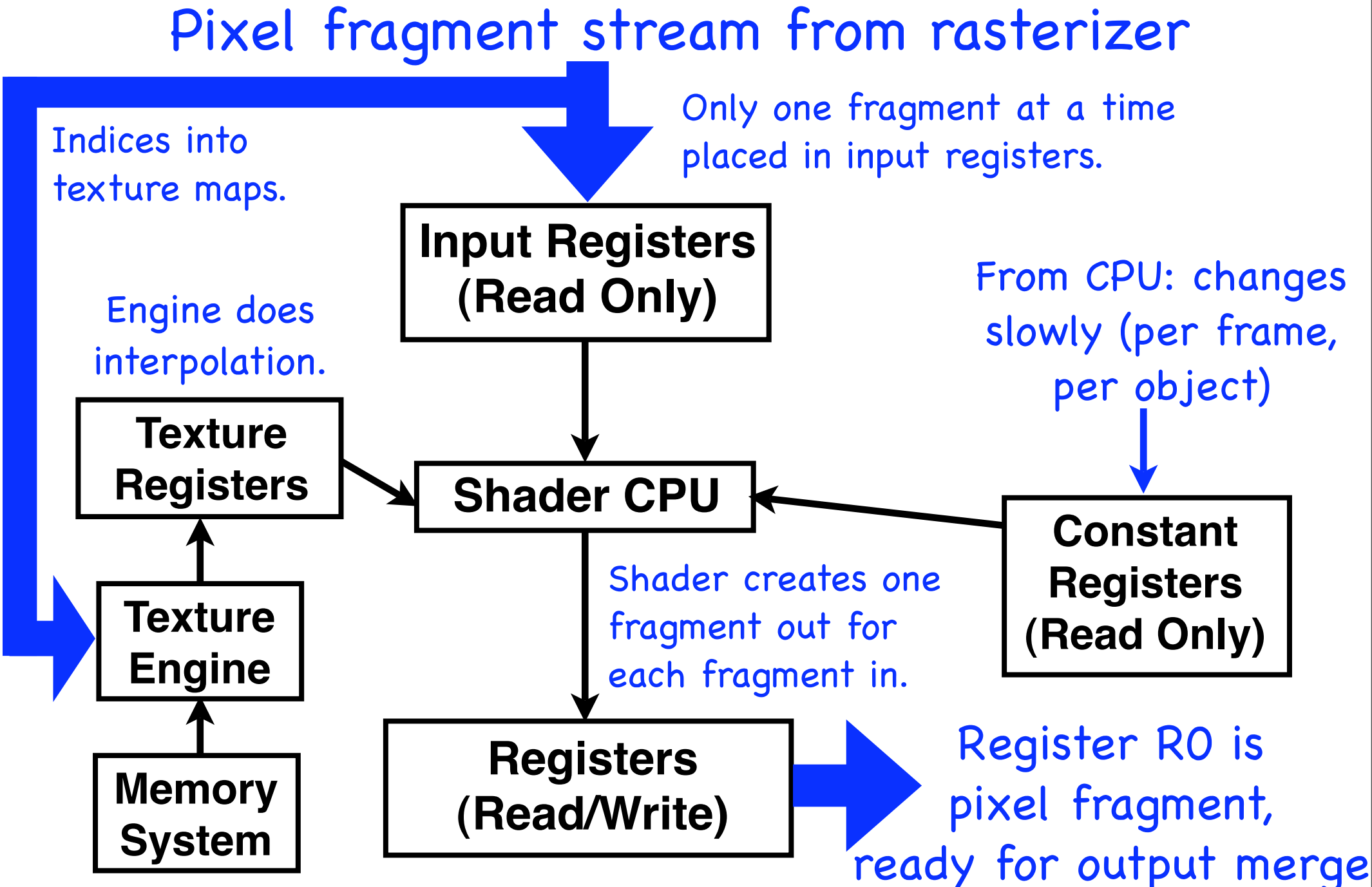


"Pixel Shader" CPU →

Texture maps (look-up tables) play a key role.

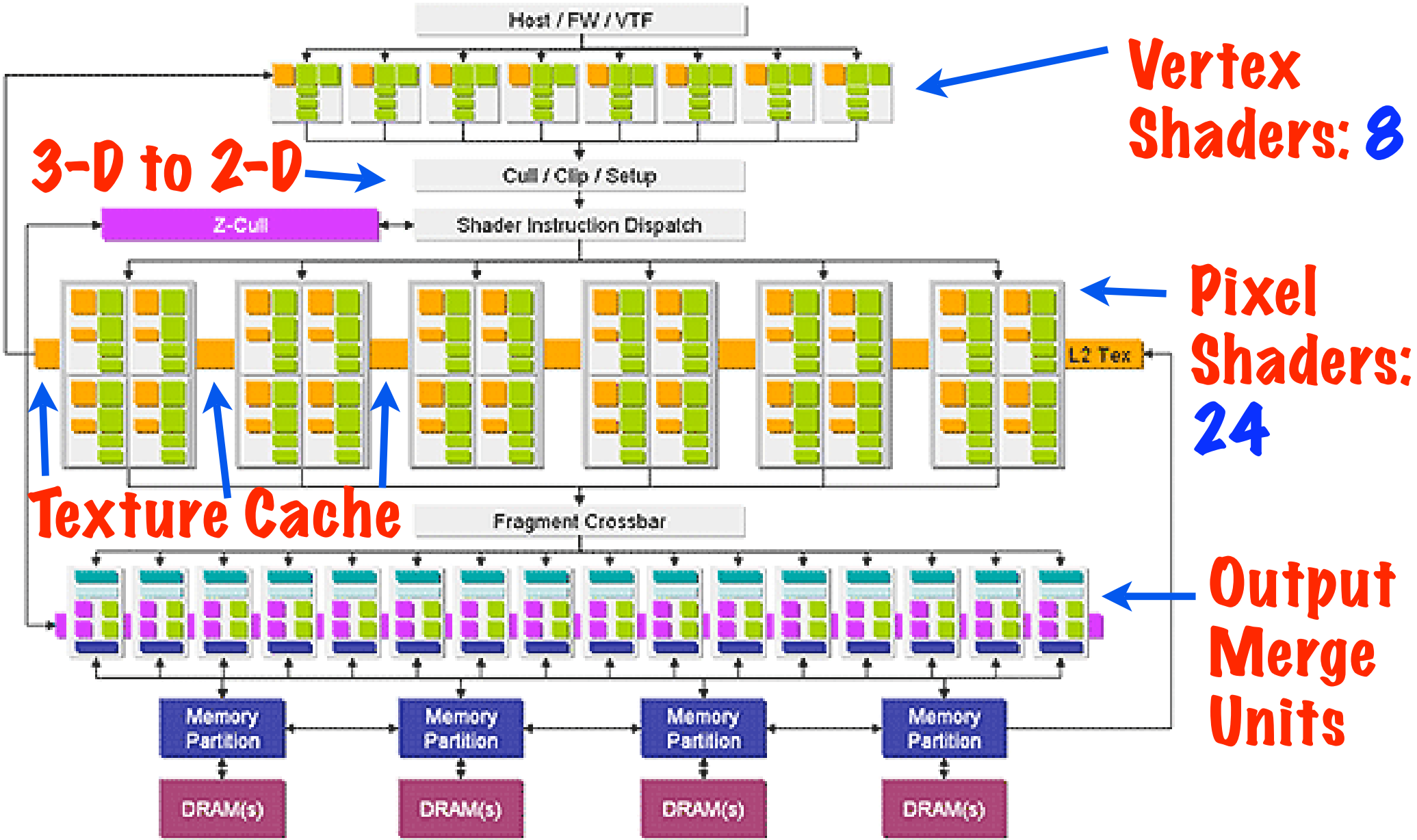
Pixel shader needs fast access to the map of Europe on teapot (via graphics card RAM).

Pixel Shader: Stream processor + Memory



Recent Design: Nvidia GeForce 7900

278 Million Transistors, 650 MHz clock, 90 nm process



Basic idea: Replace specialized logic (vertex shader, pixel shader, hardwired algorithms) with many copies of one unified CPU design.

Unified Architectures

Consequence: You no longer "see" the graphics pipeline when you look at the architecture block diagram.

Designed for: DirectX 10 (Microsoft Vista), and new non-graphics markets for GPUs.

DirectX 10 (Vista): Towards Shader Unity

Earlier APIs: Pixel and Vertex CPUs very different …

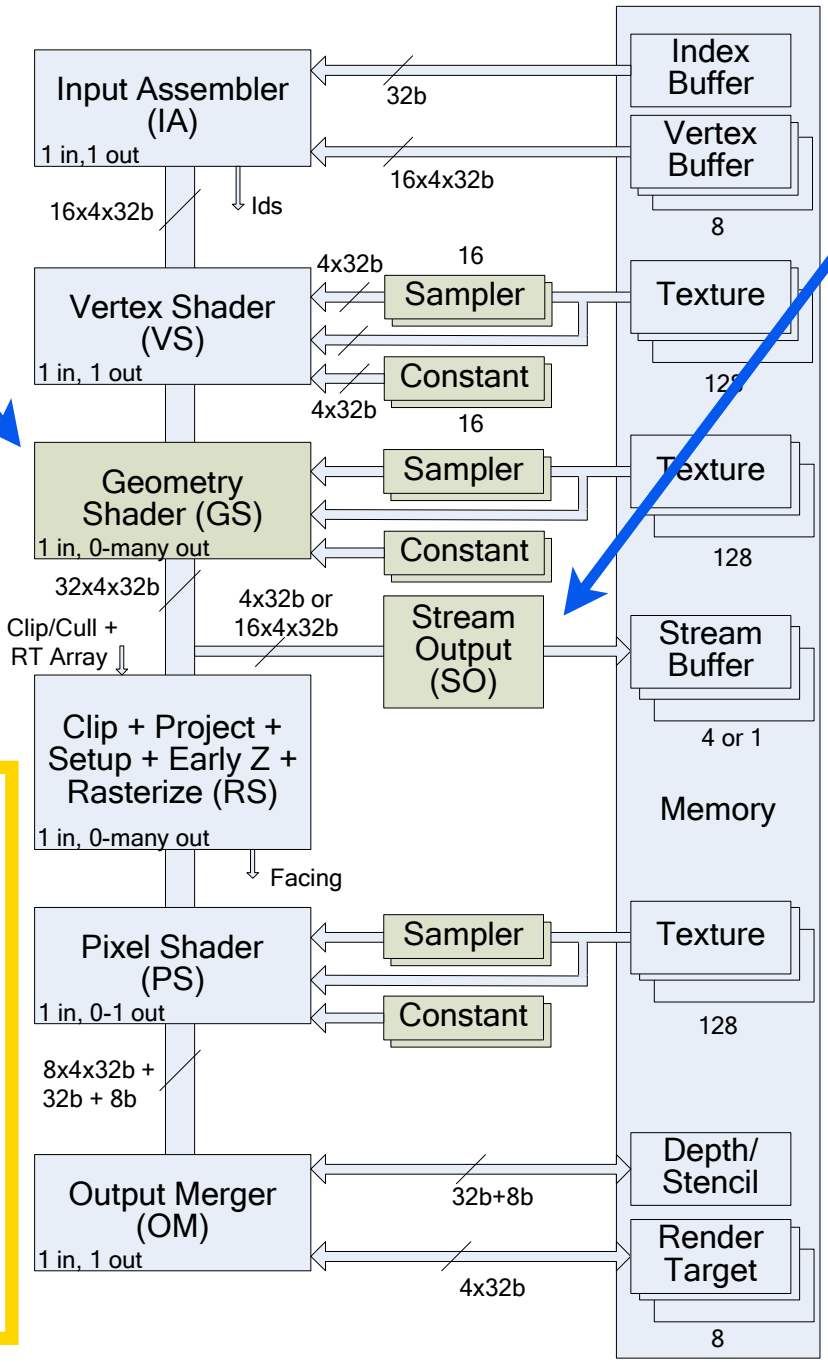
Feature	1.1 2001	2.0 2002	3.0 2004†	4.0 2006
instruction slots	128	256	≥512	≥64K
	4+8‡	32+64‡	≥512	
constant registers	≥96	≥256	≥256	16x4096
	8	32	224	
tmp registers	12	12	32	4096
	2	12	32	
input registers	16	16	16	16
	4+2§	8+2§	10	32
render targets	1	4	4	8
samplers	8	16	16	16
textures			4	128
	8	16	16	
2D tex size			2Kx2K	8Kx8K
integer ops				✓
load op				✓
sample offsets				✓
transcendental ops	✓	✓	✓	✓
		✓	✓	
derivative op			✓	✓
flow control		static	stat/dyn	dynamic
			stat/dyn	

Table 1: Shader model feature comparison summary.

DirectX 10: Many specs are identical for Pixel and Vertex CPUs

DirectX 10 : New Pipeline Features ...

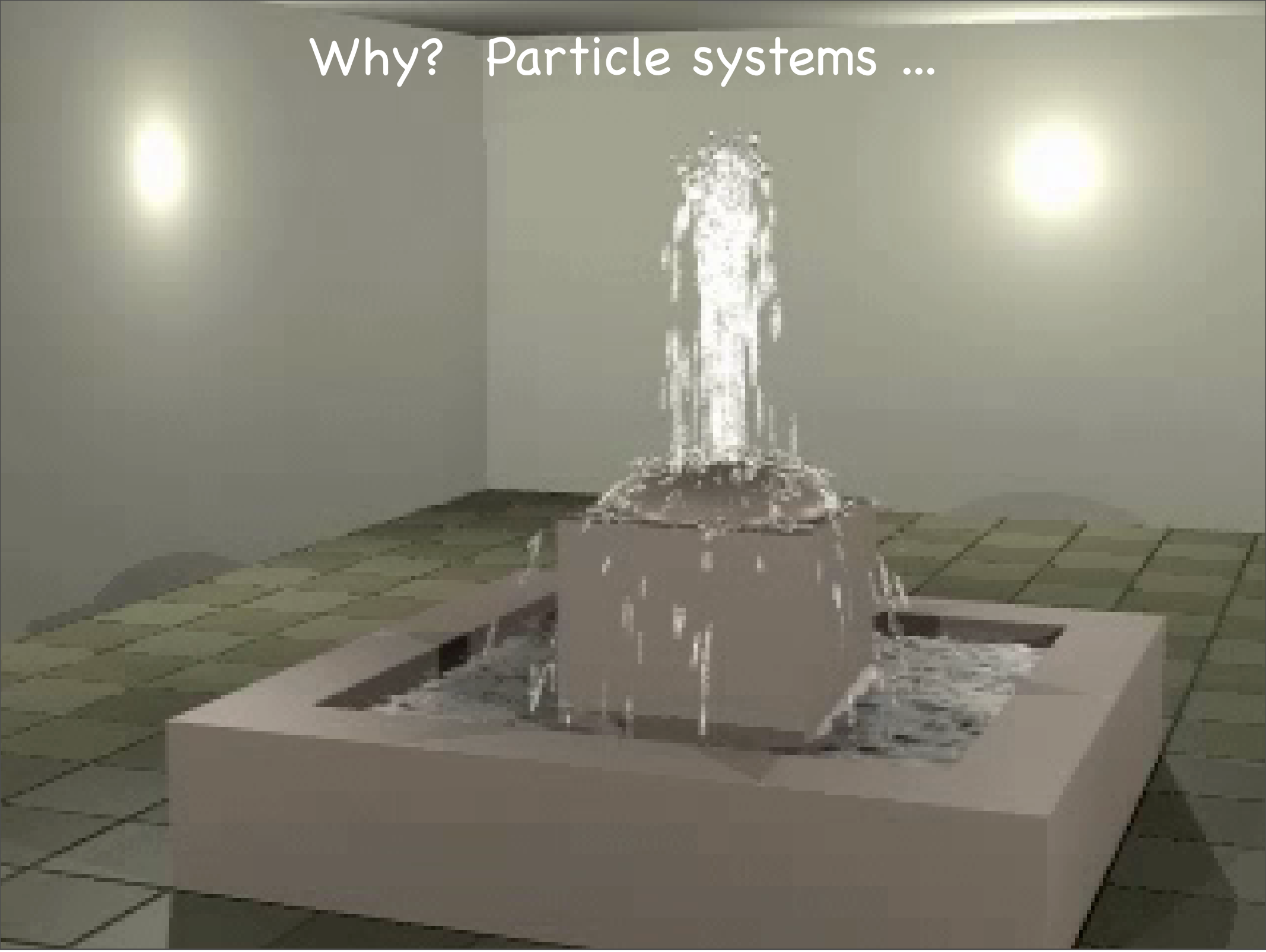
Geometry Shader: Lets a shader program create new triangles.



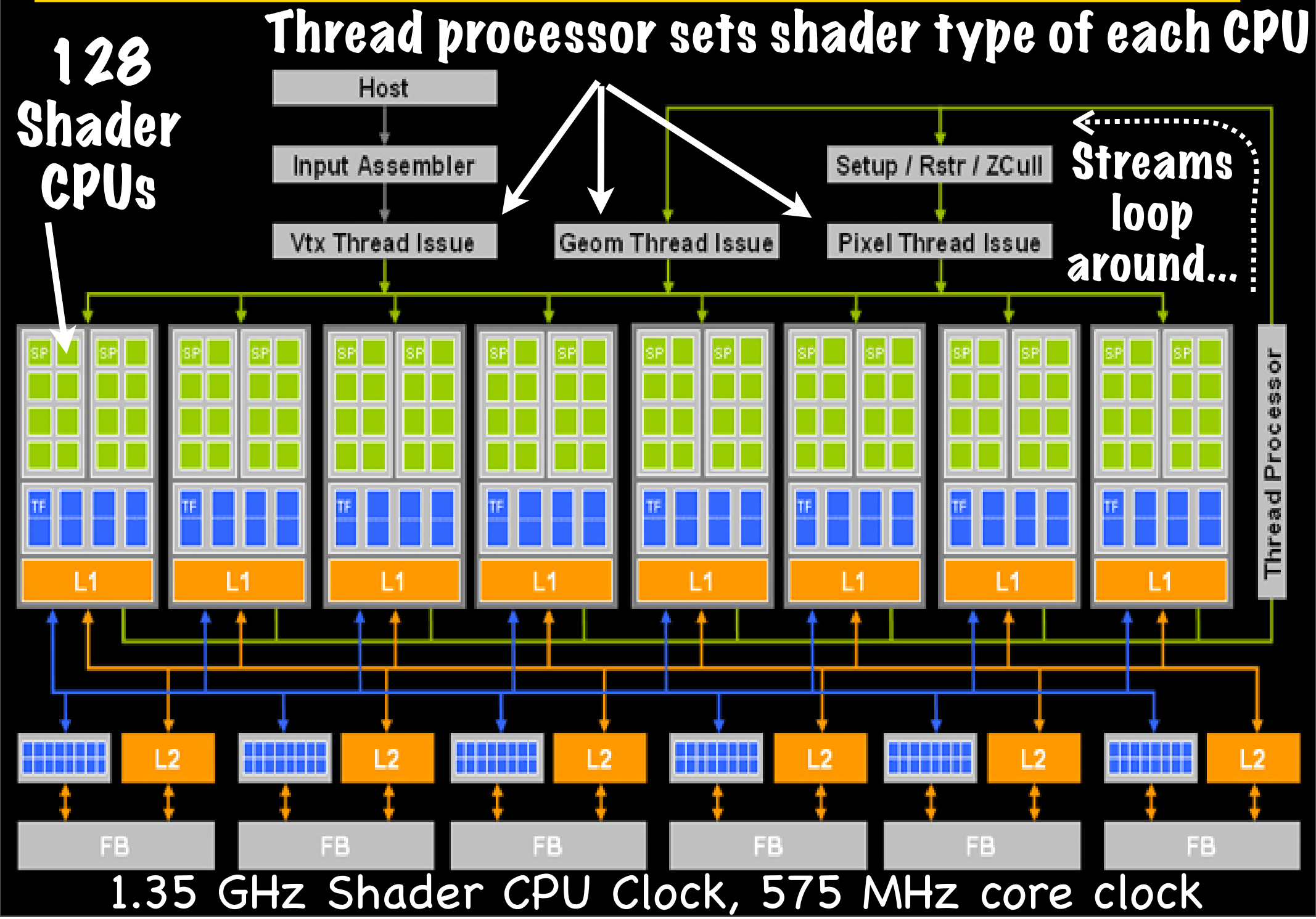
Stream Output: Lets vertex stream recirculate through shaders many times ... (and also, back to CPU)

Also: Shader CPUs are more like RISC machines in many ways.

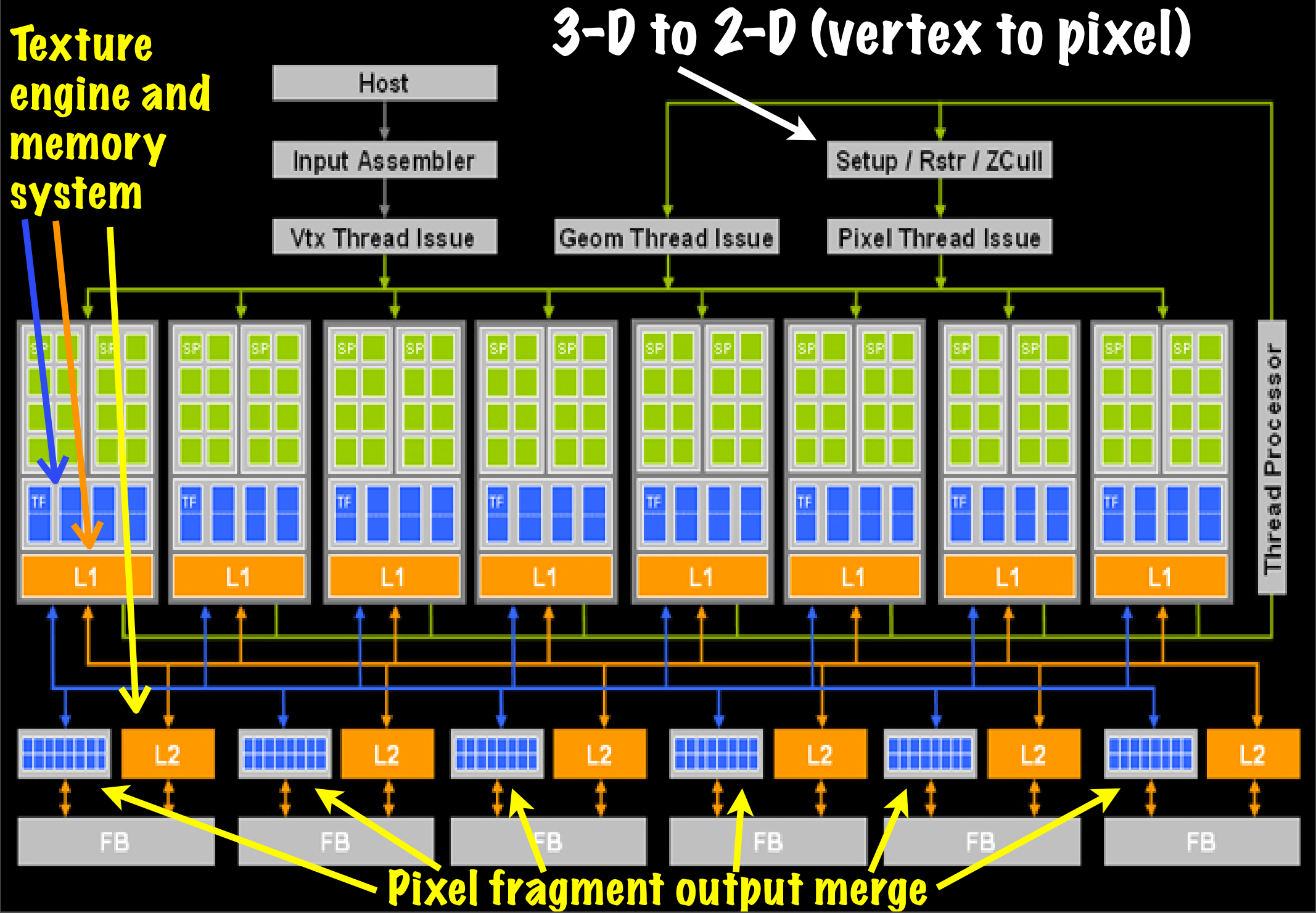
Why? Particle systems ...



NVidia 8800: Unified GPU, Announced last week



Graphics-centric functionality …

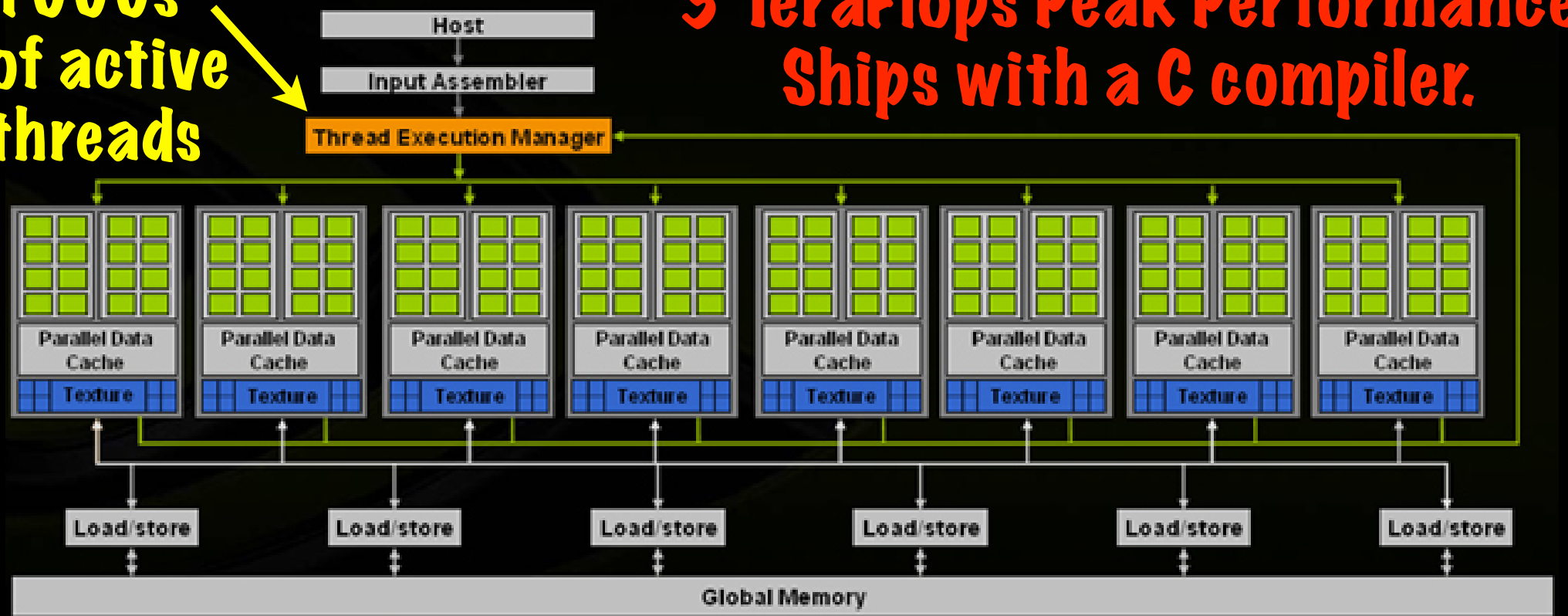


Can be reconfigured with graphics logic hidden …

128 scalar 1.35 GHz processors: Integer ALU, dual-issue single-precision IEEE floats.

1000s of active threads

3 TeraFlops Peak Performance
Ships with a C compiler.



Texture system set up to look like a conventional memory system (768MB GDDR3, 86 GB/s)

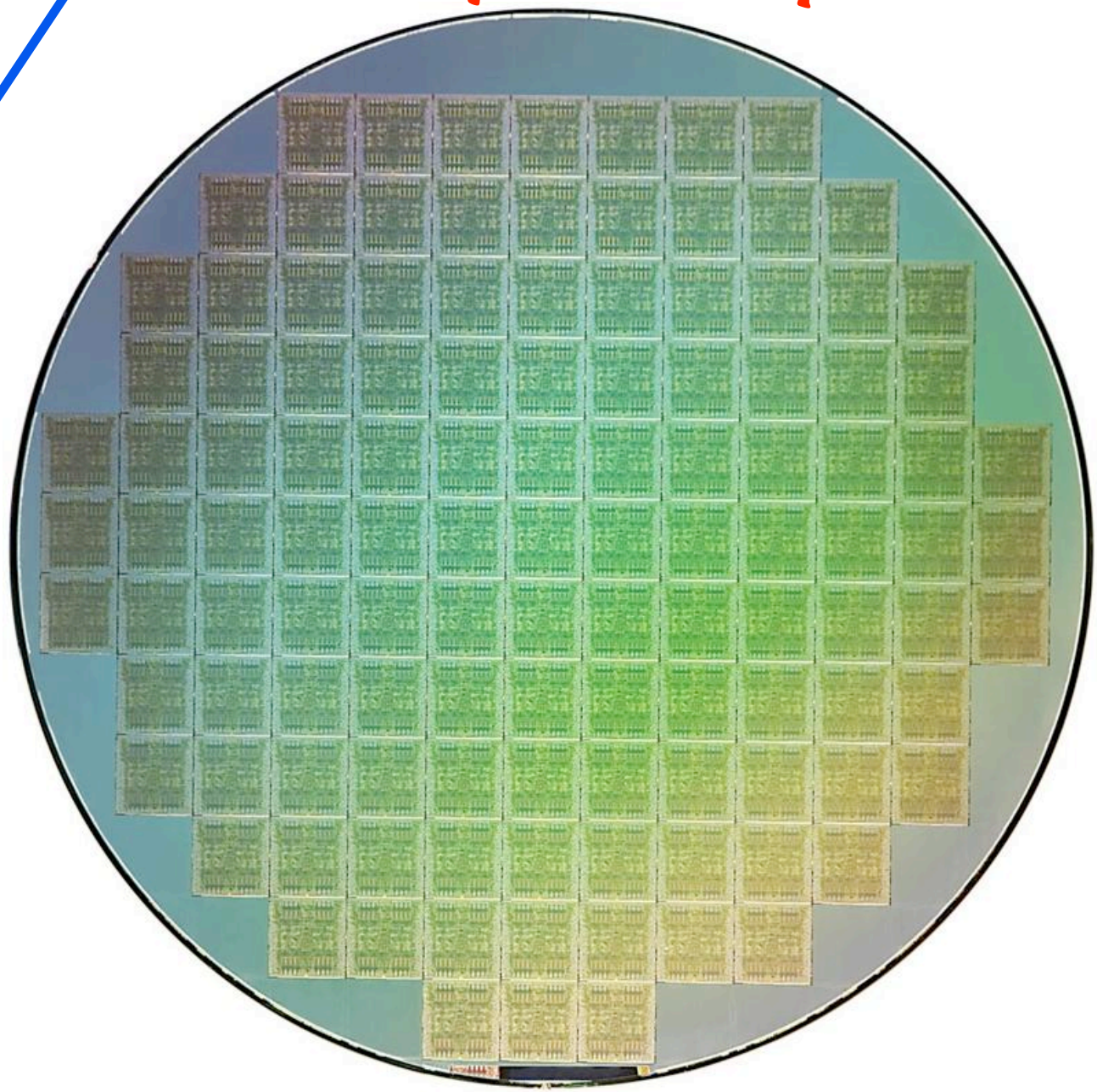
Chip Facts

90nm process

681M Transistors

80 die/wafer (pre-testing)

A big die. Many chips will not work (low yield). Low profits.



Design Facts

4 year design cycle

$400 Million design budget

600 person-years: 10 people at start, 300 at peak

Some products are "loss-leaders"

Breakthrough product creates "free" publicity you can't buy.


Nvidia's Powerful Chip Moves Closer to 'Reality' - WSJ.com

file:///Users/lazzaro/Desktop/wsj-nvidia.webarchive

WSJ.com THE WALL STREET JOURNAL. ONLINE

Article Search

Advanced SearchAdvanced Search

Free Dow Jones Sites | Free Dow Jones Sites | As of Thursday, November 9, 2006 | Online Journ

Home
News
Technology
Markets
Personal Journal
Weekend & Leisure
Opinion
TODAY'S NEWSPAPER
MY ONLINE JOURNAL

Nvidia's Powerful Chip Moves Closer to 'Reality'

Graphics Product Captures Moving Shoulder Blades, Adrianne's Authentic Pout

By DON CLARK

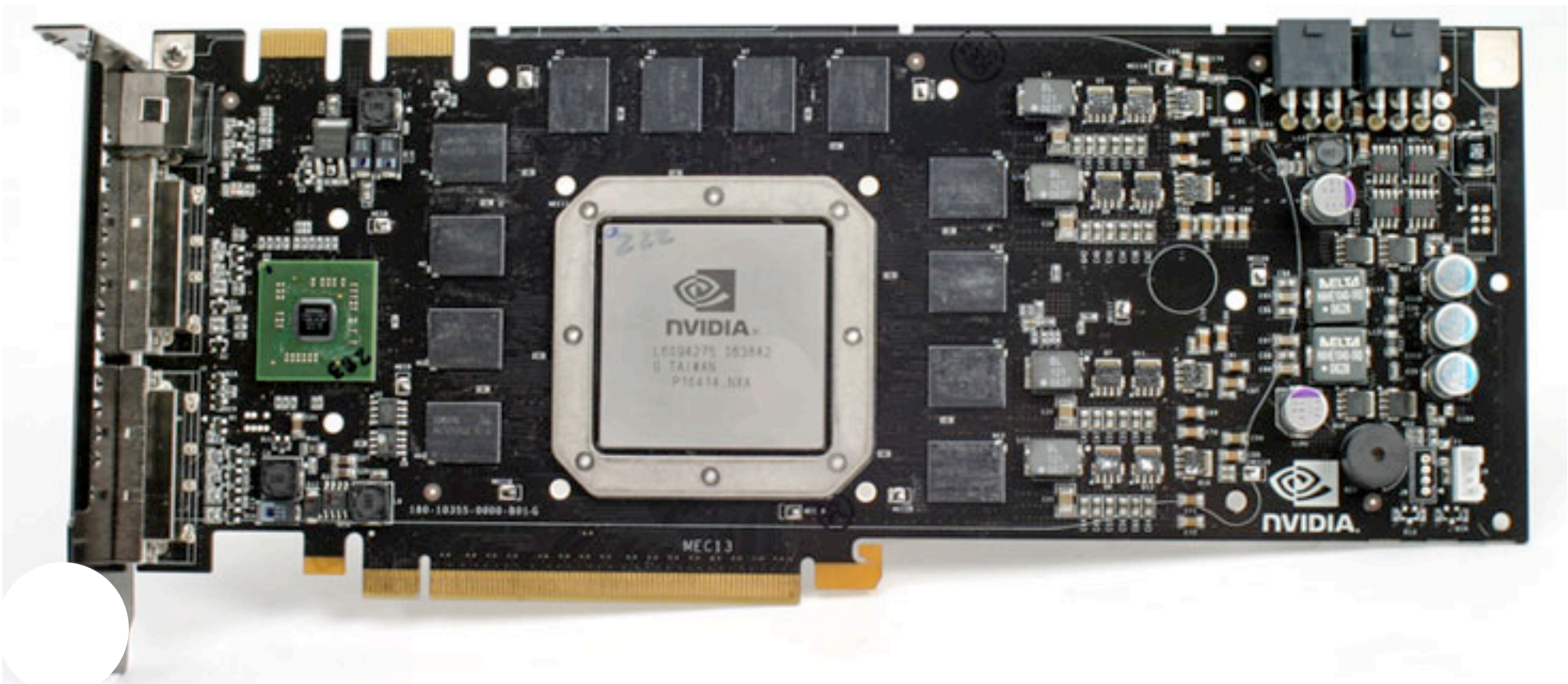
November 9, 2006; Page B3

(1) When chip is "shrunk" to 65nm fab process, die will be smaller, yields will improve, profits will rise.

(2) Simpler versions of the design will be made to create an entire product family, some very profitable. "We tape out a chip a month", NVidia CEO quote.

GeForce 8800 GTX Card: $599 List Price

PCI-Express 16X Card - 2 Aux Power Plugs!



185 Watts Thermal Design Point (TDP) --
TDP is a "real-world" maximum power spec.

Dustbuster-style fan to move 185 Watts


39
SW
W

Face was "scanned" to create a vertex model. 8800 GTX was used to do skin, eye, lips and hair rendering.

History and Graphics Processors

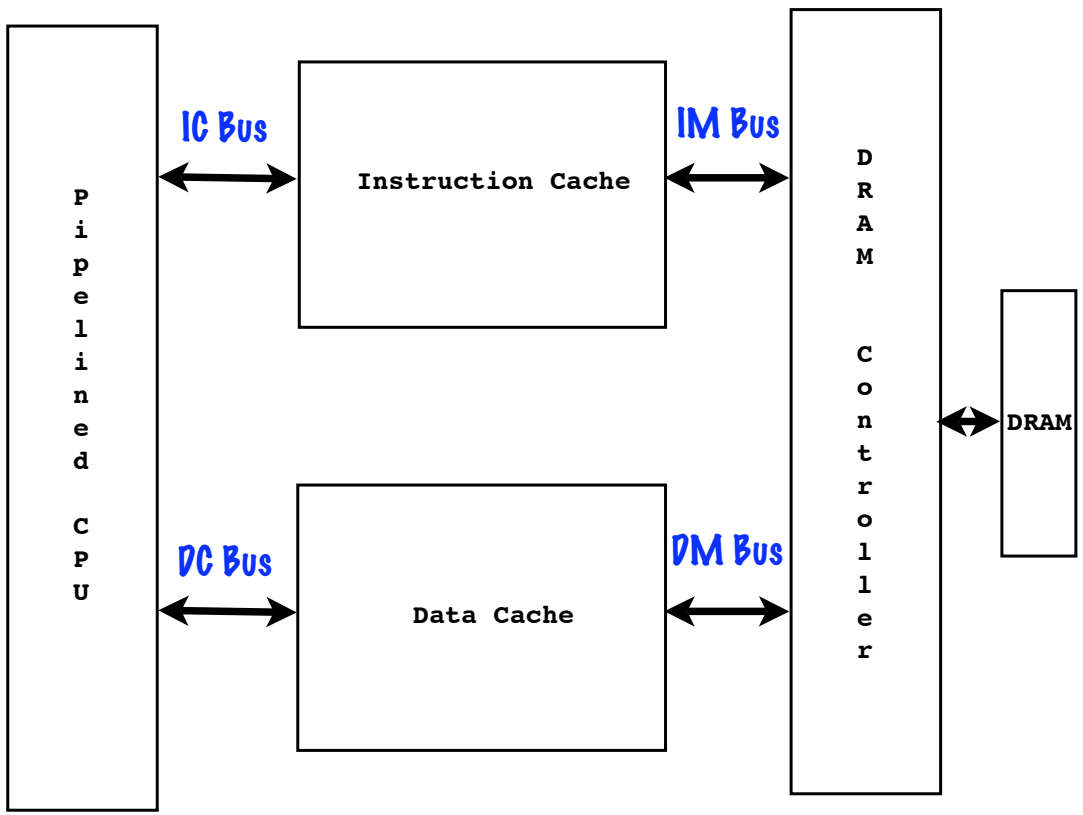
- **Create standard model from common practice:** Wire-frame geometry, triangle rasterization, pixel shading.
- **Put model in hardware:** Block diagram of chip matches computer graphics math.
- **Evolve to be programmable:** At some point, it becomes hard to see the math in the block diagram.

"Wheel of reincarnation" -- Hardwired graphics hardware evolves to look like general-purpose CPU. EECS visitor Ivan Sutherland co-wrote a paper on this topic in 1968!

Reminder: Final Checkoff this Friday!

F 11/17				Final Project: Final Checkoff, 12-2PM or 3-5PM, 125 Cory

Final report due following Monday, 11:59 PM



TAs will provide "secret" MIPS machine code tests.

Bonus points if these tests run by end of section. If not, TAs give you test code to use over weekend

Mid-term, group talks after Thanksgiving