Recent Advances and Future Trends in Graphics Hardware

Michael Doggett
Architect
November 23, 2005

Overview

- XBOX360 GPU : Xenos
  - Rendering performance
  - GPU architecture
  - Unified shader
  - Memory Export
  - Texture/Vertex Fetch
  - HDR rendering
  - Displaced subdivision surfaces
- RADEON X1800
- Graphics Hardware
  - Graphics APIs
  - GPU Research

ATI - Driving the Visual Experience Everywhere

- Products from cell phones to super computers

Gaming

Gaming Console

Integrated

Embedded Display

Notebook

Digital TV

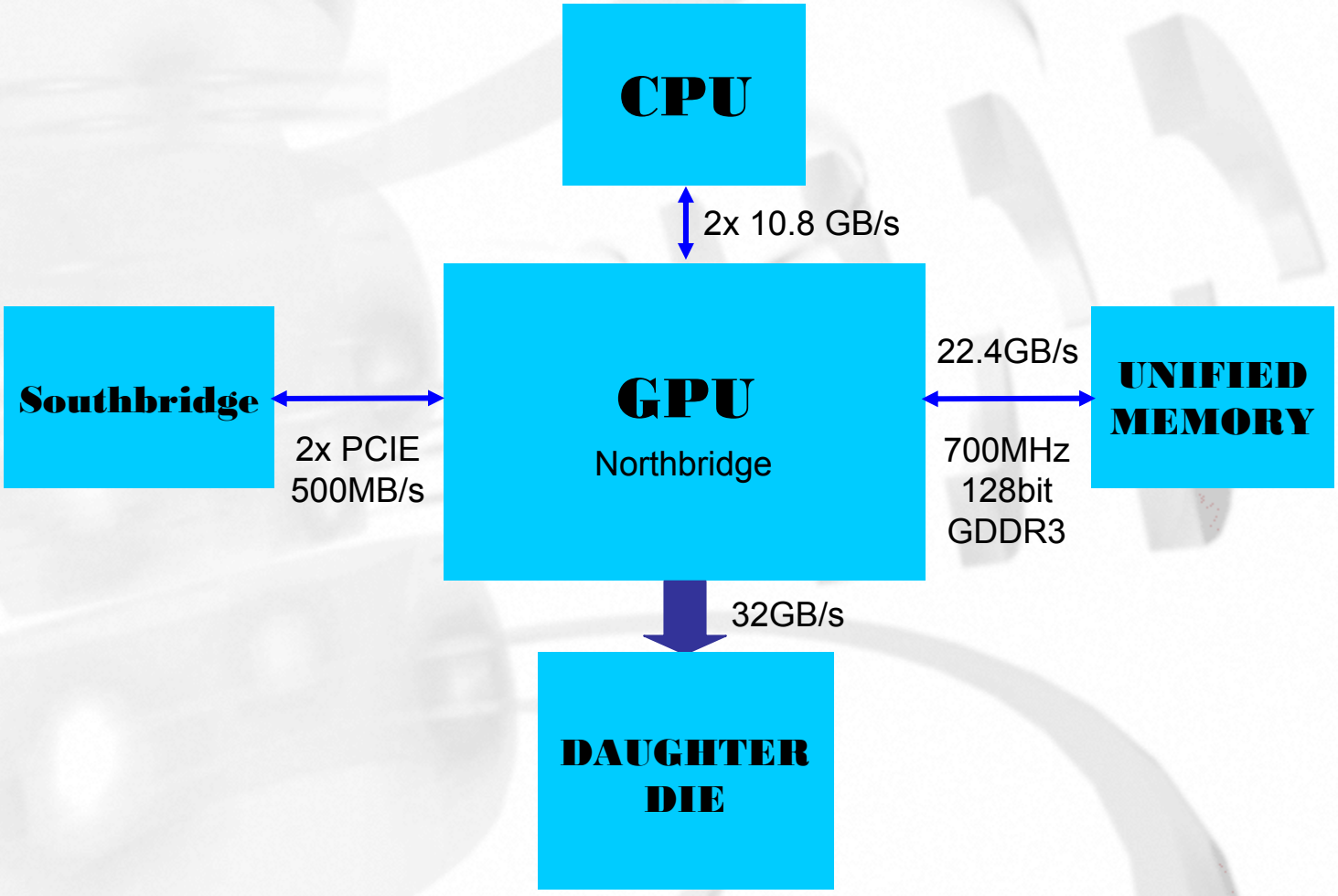
Color Phone Display

Multimedia

Workstation

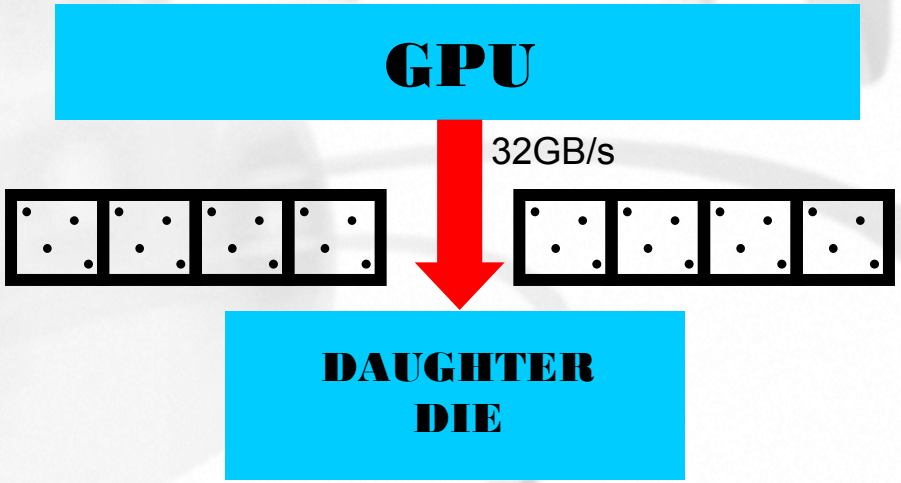
Multi Monitor Display

System architecture



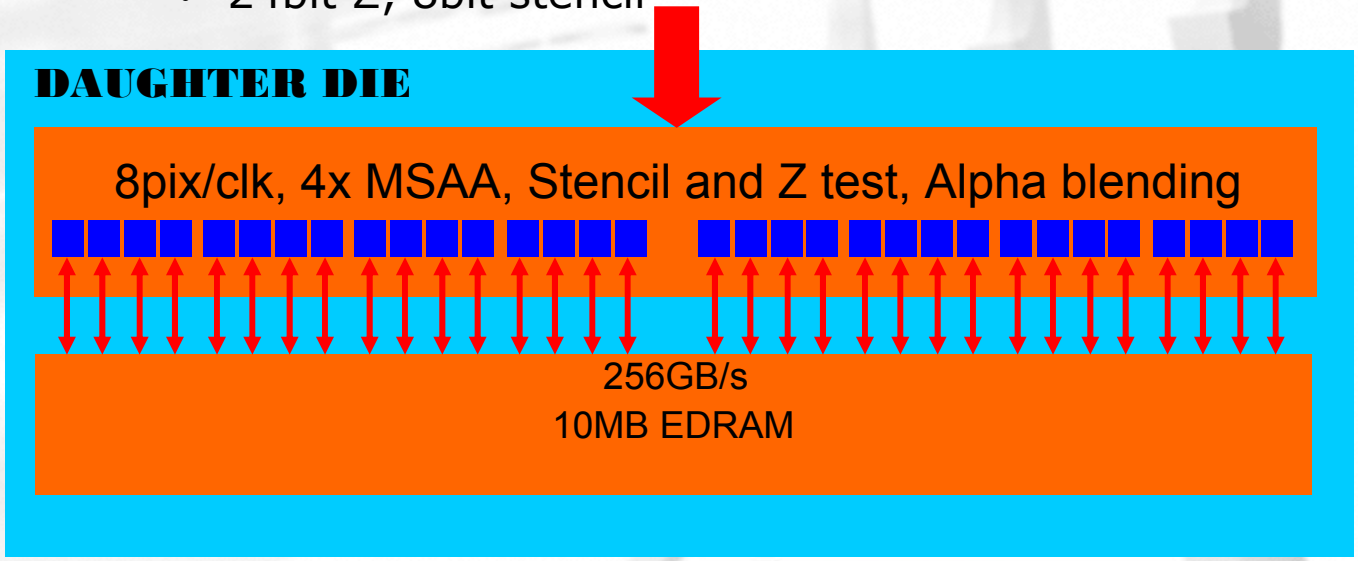
Rendering performance

- GPU to Daughter Die interface
  - 8 pixels/clk
    - 32BPP color
    - 4 samples Z - Lossless compression
  - 16 pixels/clk – Double Z
    - 4 samples Z - Lossless compression

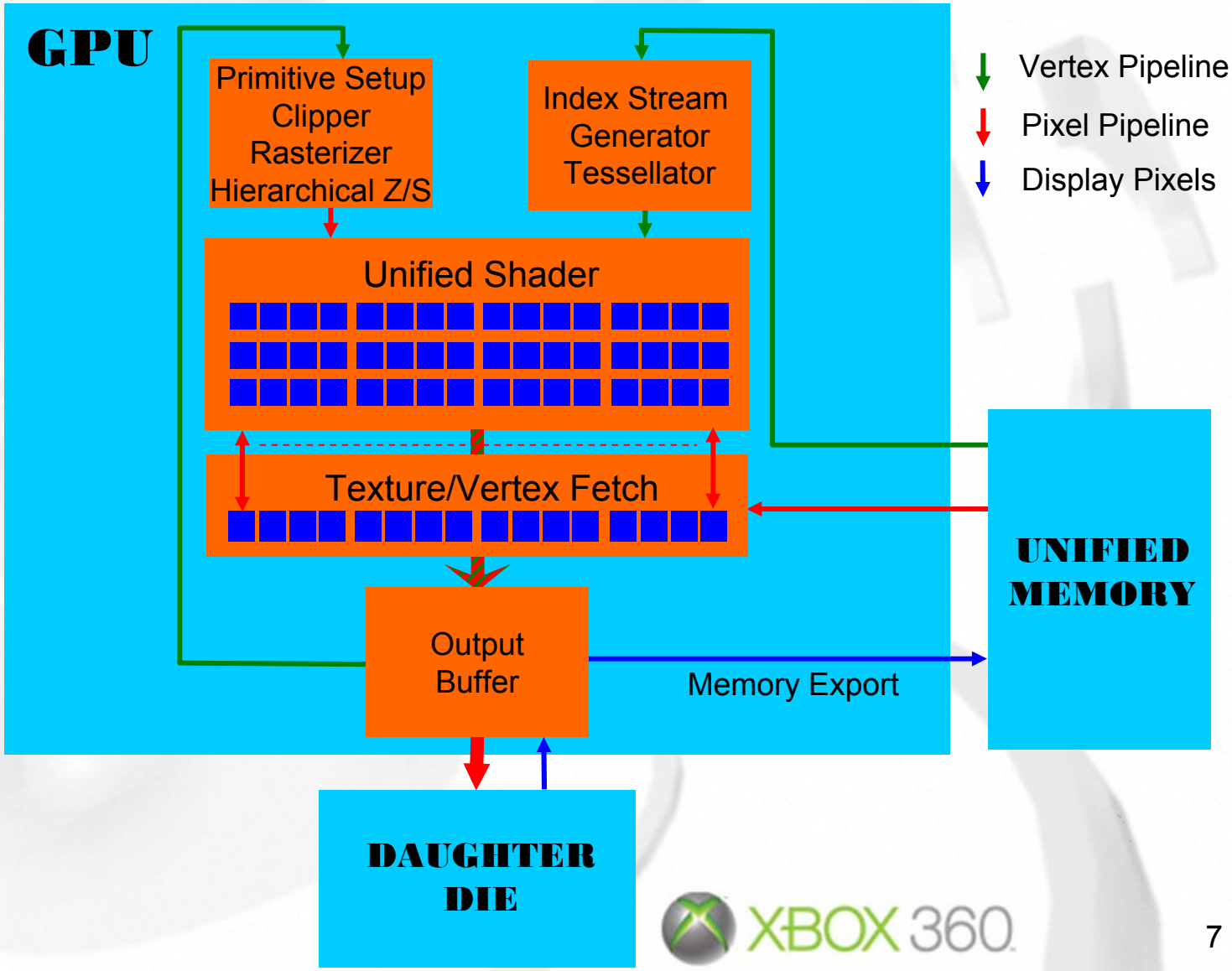


Rendering performance

- Alpha and Z logic to EDRAM interface
  - 256GB/s
  - Color and Z - 32 samples
    - 32bit color, 24bit Z, 8bit stencil
  - Double Z - 64 samples
    - 24bit Z, 8bit stencil

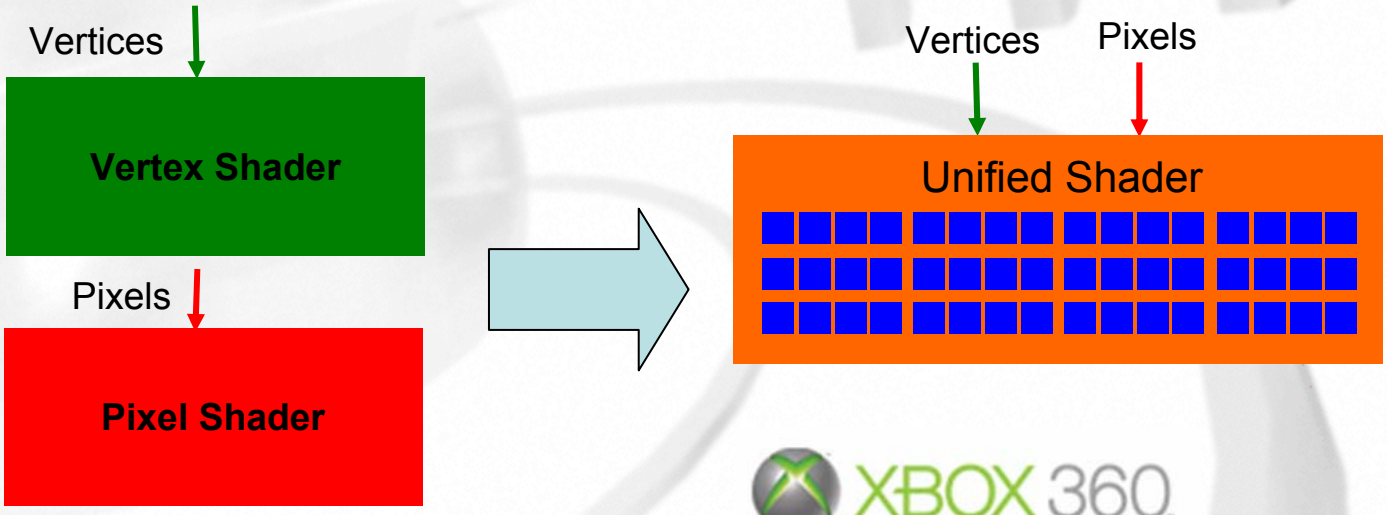


GPU architecture



Unified Shader

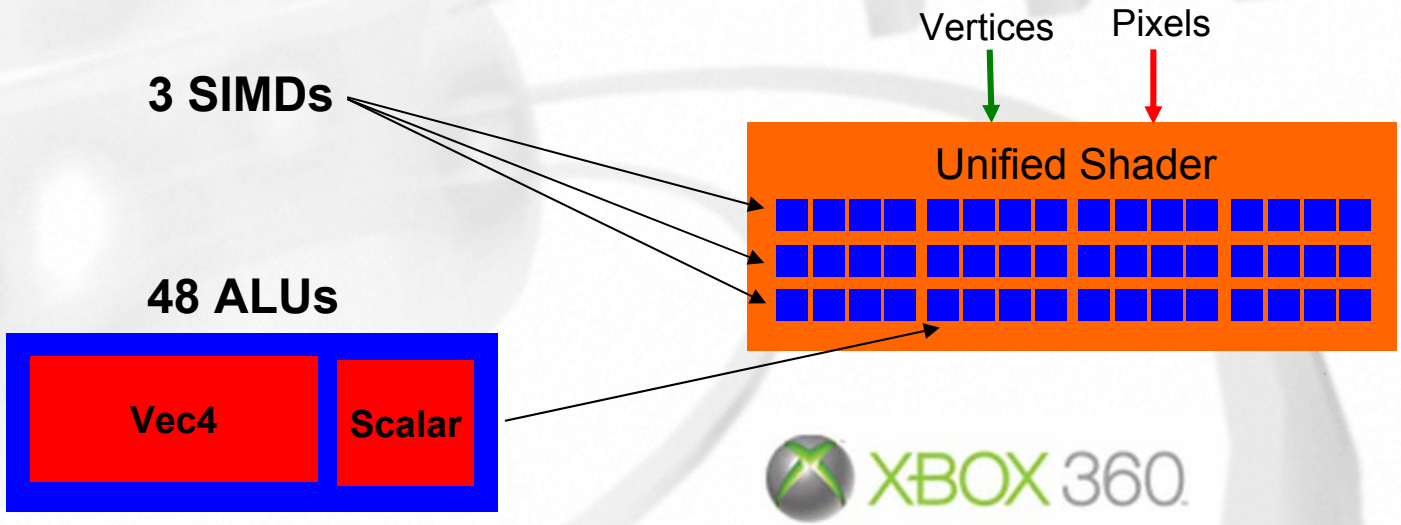
- A revolutionary step in Graphics Hardware
- One hardware design that performs both Vertex and Pixel shaders
- Vertex processing power



XBOX 360

Unified Shader

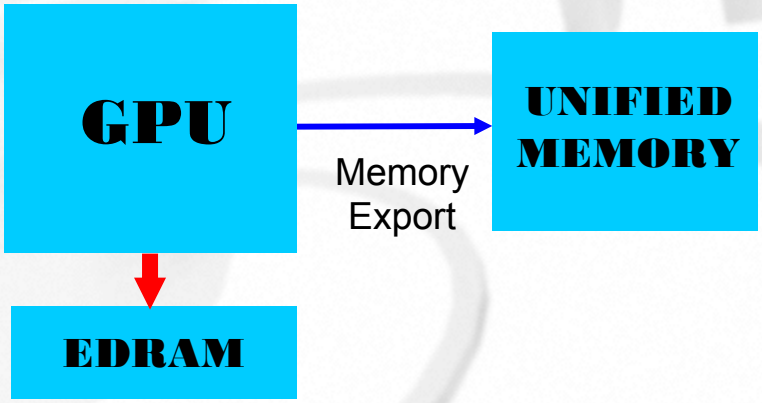
- GPU based vertex and pixel load balancing
  - Better vertex and pixel resource usage
- Union of features
  - E.g. Control flow, indexable constant, …
- DX9 Shader Model 3.0+



XBOX 360

Memory Export

- Shader output to a computed address
- Virtualize shader resources - multipass
- Shader debug
- Randomly update data structures from Vertex or Pixel Shader
- Scatter write

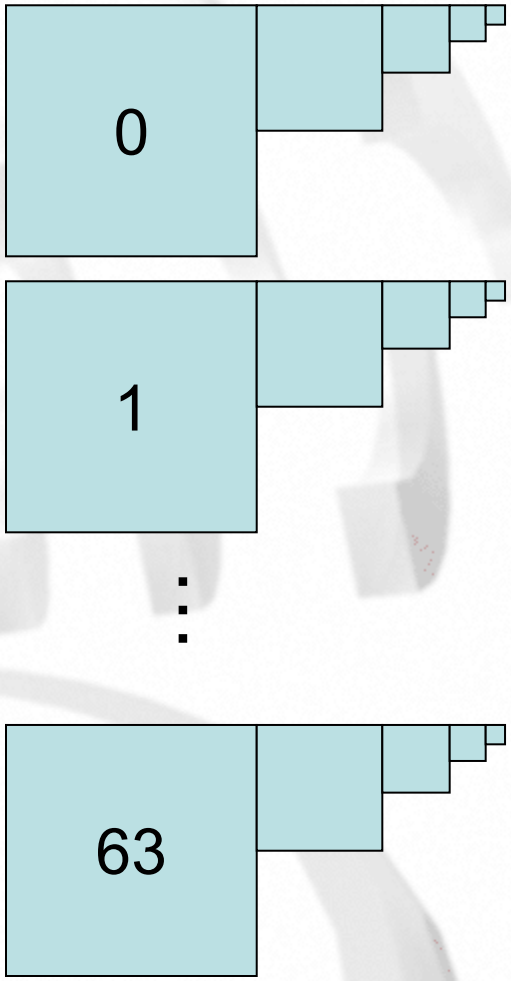


Texture/Vertex Fetch

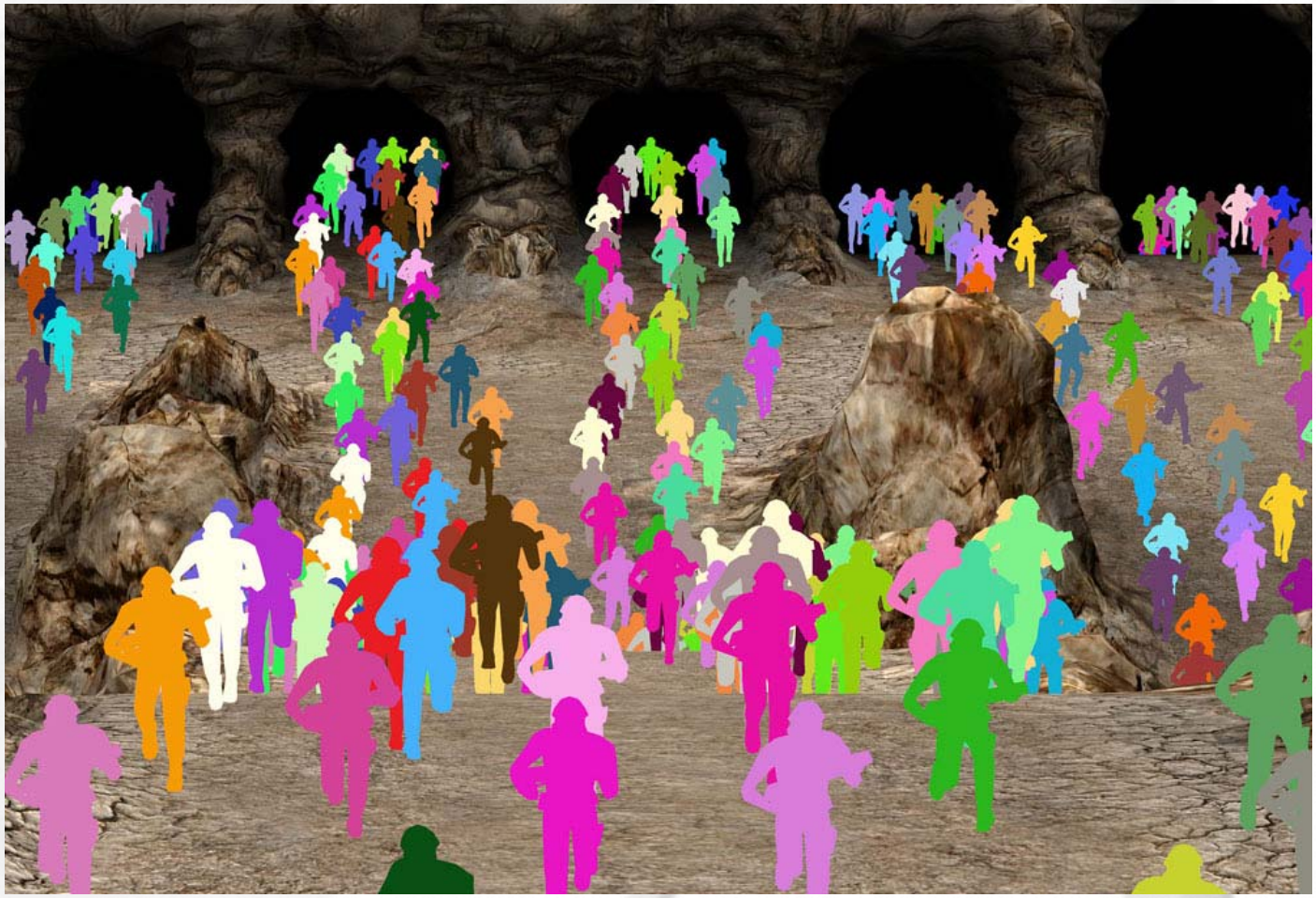
- Shader fetch can be either:
  - Texture fetch (16 units)
    - LOD computation
    - Linear, Bi-linear, Tri-linear Filtering
    - Uses cache optimized for 2D, 3D texture data with varying pixel sizes
    - Unified texture cache
  - Vertex fetch (16 units)
    - Uses cache optimized for vertex-style data

Texture Arrays

- Generalization of 6 faced cube maps to 64 faces
- Each face is a 2D mip mapped surface
- *Not* volume texture
- Applications
  - Animation frames
  - Varying skins for instanced characters / objects
  - Character shadow texture flipbook animations



Texture array application :
Unique seeds for instanced shading

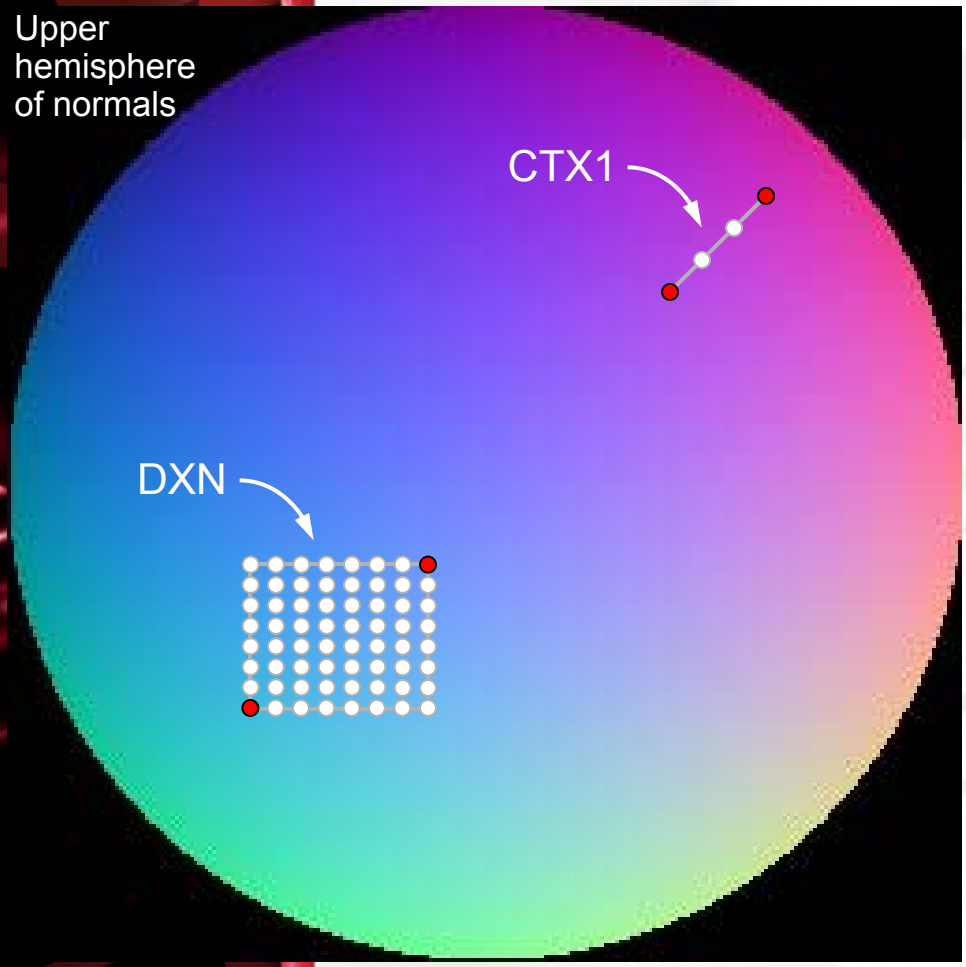


Texture array application : Hundreds of instanced characters

Texture compression

- All of the old DXT formats
  - DXT1, DXT2/3, DXT4/5
- Several new formats (variations on above formats)
  - DXT3A
    - 4 bit scalar replicated into four channels in shader
  - DXT3A as 1111
    - 1 bit per channel pixel
  - DXT5A
    - 3bit selection between 2 8bit endpoints
  - DXN
    - 3Dc normal compression,
    - 2-channel version of DXT5A
  - CTX1
    - 2bit selection between 2 8.8bit endpoints

CTX1 vs. DXN



- Both have 8.8 anchor points
- CTX1
  - *x* and *y* share 2-bit interpolation value
  - 4 representable normals per 4×4 block of texels
  - 4 bits per texel
- DXN
  - Independent 3-bit interpolation values for *x* and *y*
  - 64 representable normals per 4×4 block of texels
  - 8 bits per texel

High Dynamic Range Rendering

- Special compact HDR render target format:
  - Just 32 bits: 7e3 7e3 7e3 2
  - Compatible with multisample antialiasing
  - R, G and B are unsigned floating point numbers
    - 7 bits of mantissa
    - 3 bits of exponent
    - Range of 0..16
  - 2 bits of alpha channel
- 16-bit fixed point at half speed
  - With full blending

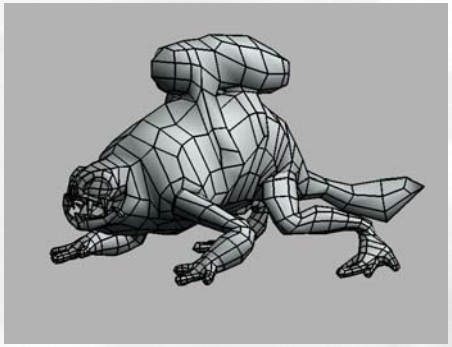
Displaced subdivision surfaces

- Prototype algorithm
- Vineet Goel, ATI research Orlando

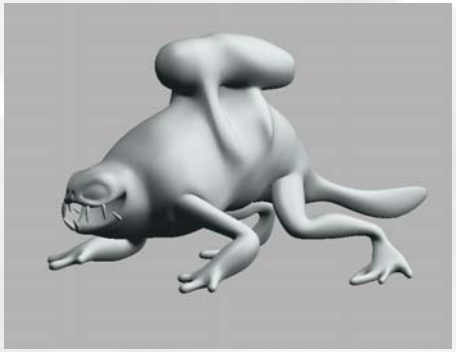
Displaced subdivision surface algorithm

- Tessellator:
  - Generates 64 vertices for each patch that are fed into the VS.
- Vertex Shader:
  - Reads in one-ring, computes Stam's method using precomputed table lookup
  - Adds Displacement map
- Pixel Shader
  - Adds bump mapping and surface color

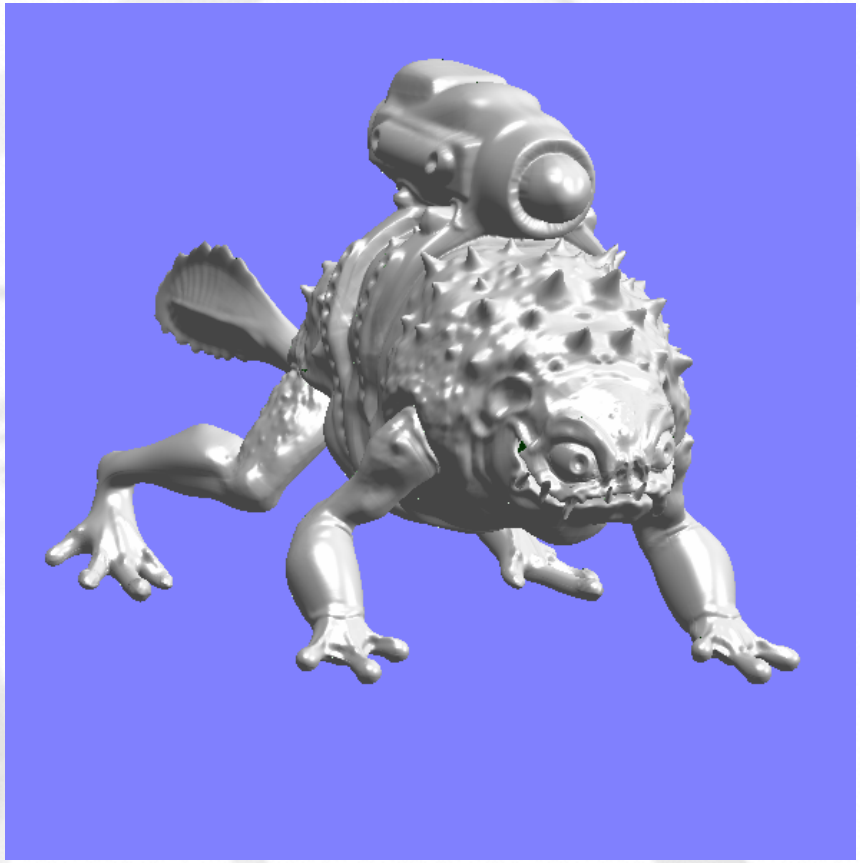
Displaced subdivision surfaces



Base mesh
- Used by Tessellator to generate vertices

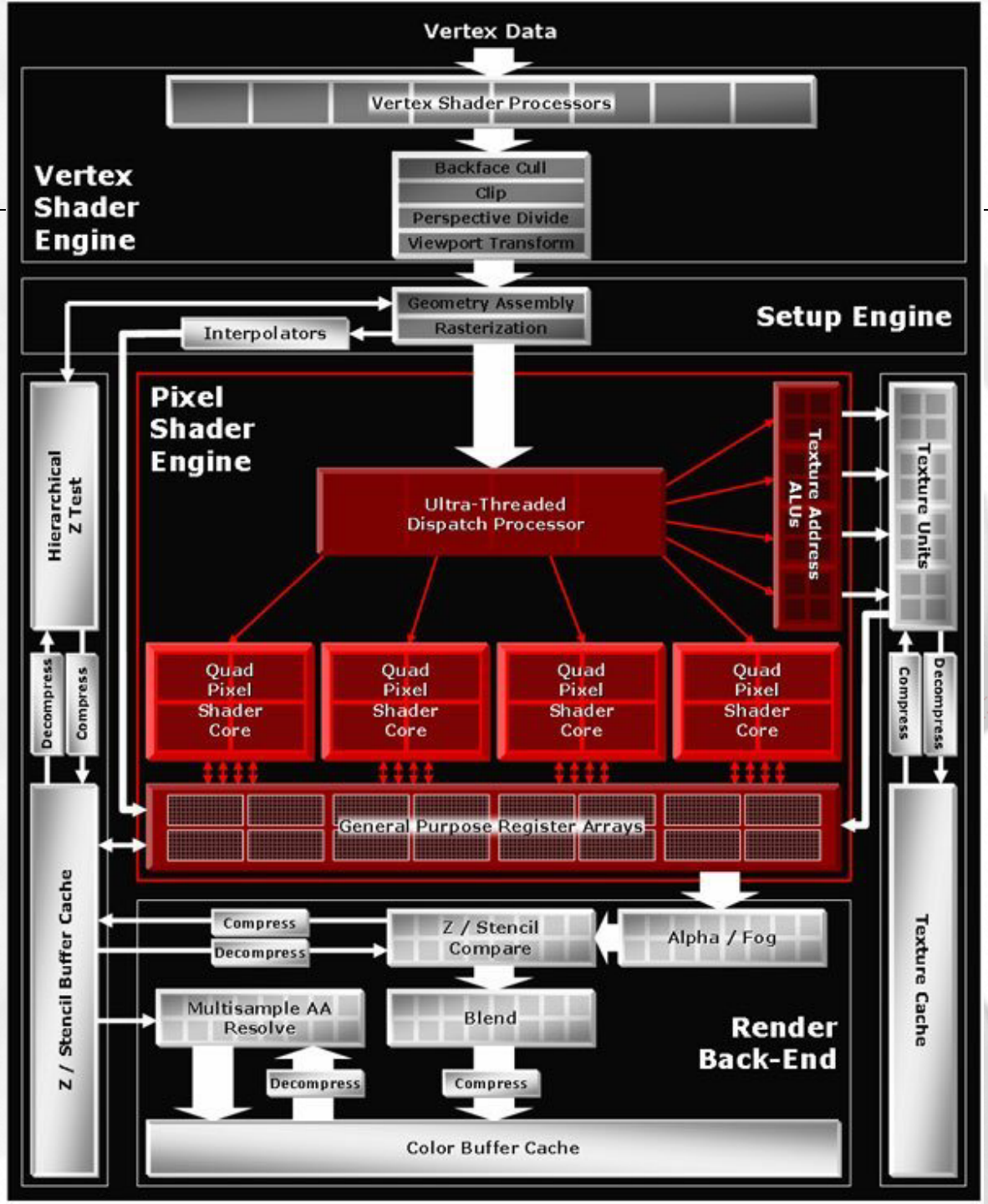


Subdivision surface



Displaced subdivision surface

RADEON X1800


Vertex Data
Vertex Shader Processors
Vertex Shader Engine
Backface Cull
Clip
Perspective Divide
Viewport Transform
Geometry Assembly
Rasterization
Interpolators
Setup Engine
Pixel Shader Engine
Hierarchical Z Test
Ultra-Threaded Dispatch Processor
Texture Address ALUs
Texture Units
Quad Pixel Shader Core
Quad Pixel Shader Core
Quad Pixel Shader Core
Quad Pixel Shader Core
Decompress
Compress
Compress
Decompress
General Purpose Register Arrays
Z / Stencil Buffer Cache
Compress
Decompress
Z / Stencil Compare
Alpha / Fog
Multisample AA Resolve
Blend
Render Back-End
Decompress
Compress
Color Buffer Cache
Texture Cache

RADEON X1800

- DirectX 9.0 Shader Model 3.0
- Pixel Shader
  - 32bit IEEE Single Precision float
  - 4 quad based, threaded units
  - 512 threads of 16 pixels allow efficient dynamic branching

Graphics Hardware

Graphics APIs

- Windows Vista
  - Virtual Memory
  - Improved state change efficiency
- Vista DirectX9
  - ClearType
- DirectX10
  - Unified shader programming model
  - Geometry Shader
    - Access to entire triangle and adjacent vertices
    - Output to array of render targets, cube maps
  - Stream output from Geometry Shader

GPU research

- Improved multi-GPU performance and antialiasing
  - CrossFire
  - Multi-chip, multi-core
- Higher Order Surfaces
  - Subdivision surfaces, NURBS
- General Purpose GPU (GPGPU)
  - Driving improved 32bit float performance
  - Xenos shader scatter write

GPU research - Shadows

- Performance enhancements for stencil shadows and shadow buffers
- Hierarchical rendering
  - User low resolution shadow map to find areas that require detail rasterization
    - Chan, Durand EGSR04
  - Stencil shadow volumes using 8x8 pixel tiles
    - Aila, Akenine-Möller GH04
- Improving Z, stencil performance

GPU research – Ray Tracing

- Accelerating ray tracing on GPUs
  - Uniform grid
    - Purcell et. al. SIG02
  - KD-Tree, performance improvement greater than hardware improvement
    - Foley et. al. GH05
  - EarlyZ Volume rendering ray casting
    - J. Krüger, R. Westermann VIS03
  - Build and update acceleration structures
    - Xenos shader scatter write
- FPGA based ray tracing design
    - RPU, Woop et. al. SIG05
- GPU evolution affected by ray tracing

GPU research – Anti-Aliasing

- Multisample efficient, effective solution
- Custom filtering of multisampled images
- Use Multisample sample mask to control transparency
  - Alpha to coverage
  - Foliage, Chicken wire
- Order independent transparency

RADEON X1800 Demos

- The Assassin
- Parthenon
- Toyshop

Questions ?