



# Radeon™ HD 2900

Michael Doggett

August 5, 2007

# Overview

- Starting Point
- Requirements
- Top level
- Pipeline Blocks from 'top to bottom'
  - Command Processor
  - Shader Setup Engine
  - Ultra Threaded Dispatch Processor
  - Shader Core
  - Texture
  - Render Backend
  - Memory Controller
- Conclusion

# Starting point

- Combine the best of existing technology
  - R5xx series
    - Heavily threaded shader cores
      - Hides latency of memory fetch
    - Vec4+1 Vertex and Vec3+1 Pixel shaders
    - Ringbus memory subsystem
  - XBOX 360 GPU
    - Unified shader architecture
      - Vertex and Pixel
      - Vec4+1
      - Stream Out
    - Unified L1 texture cache
    - Introduced Tessellator

# Requirements

- DirectX10 compatible
- Support new driver model
  - Vista driver model
- Scalable family
  - “Number” of shader cores, texture units, render back-ends.
  - Shader scalable in number of pipes, SIMDs.
  - Target specific cost, feature set and performance levels for each part

# Top Level

Red - Compute

Yellow - Cache

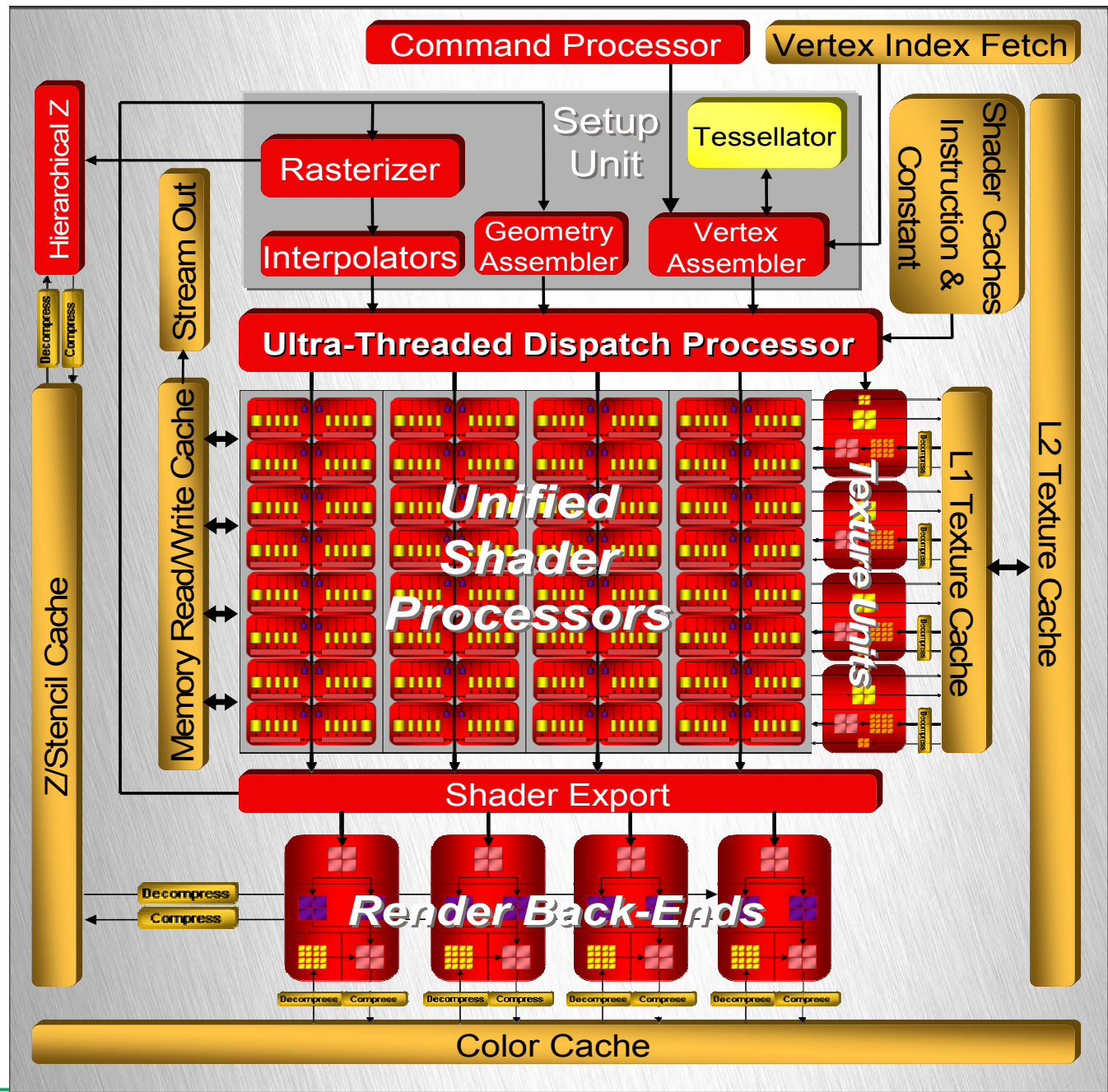
Unified shader

Shader R/W

Instr./Const. cache

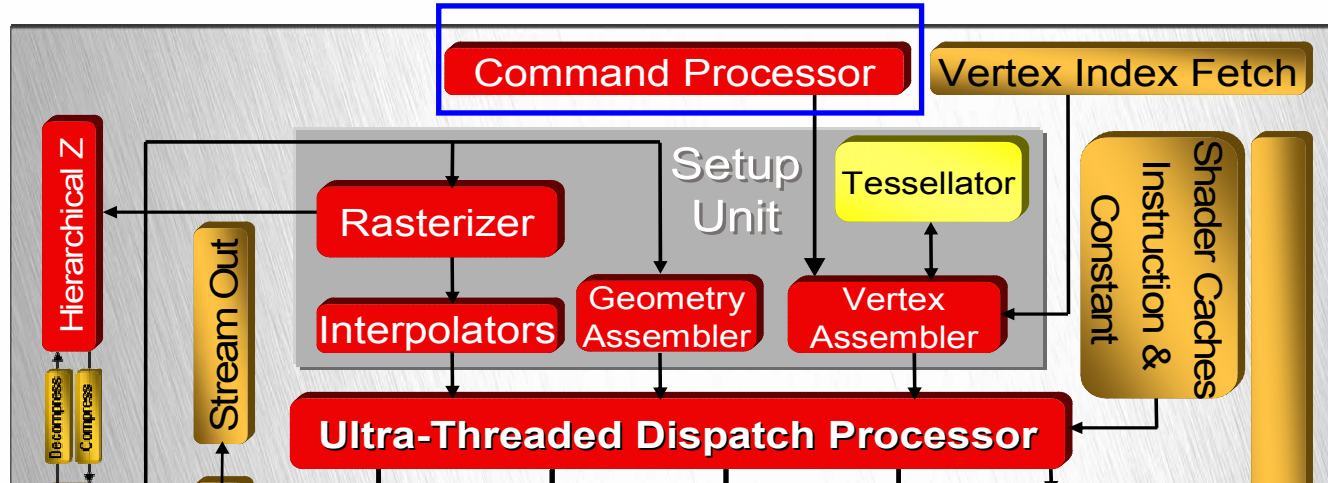
Unified texture cache

Compression

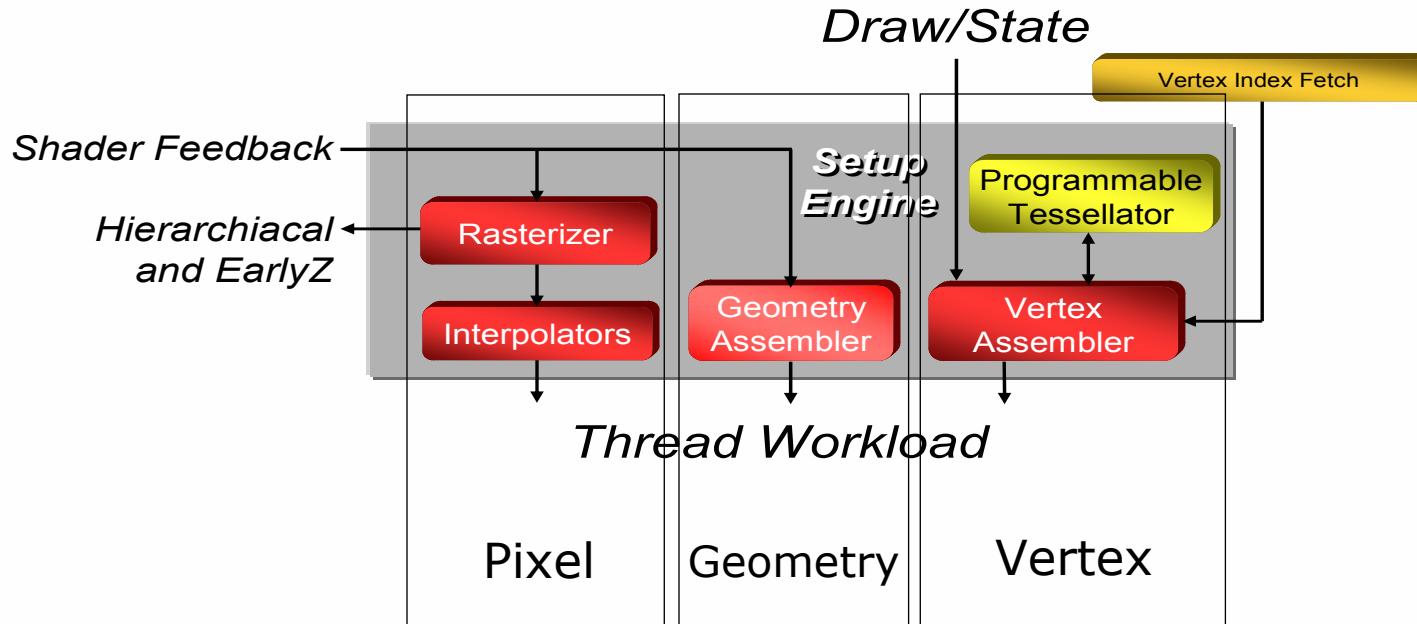


# Command Processor

- GPU interface with host
- A custom RISC based Micro-Coded engine
- First class memory client with Read/Write access
- State management

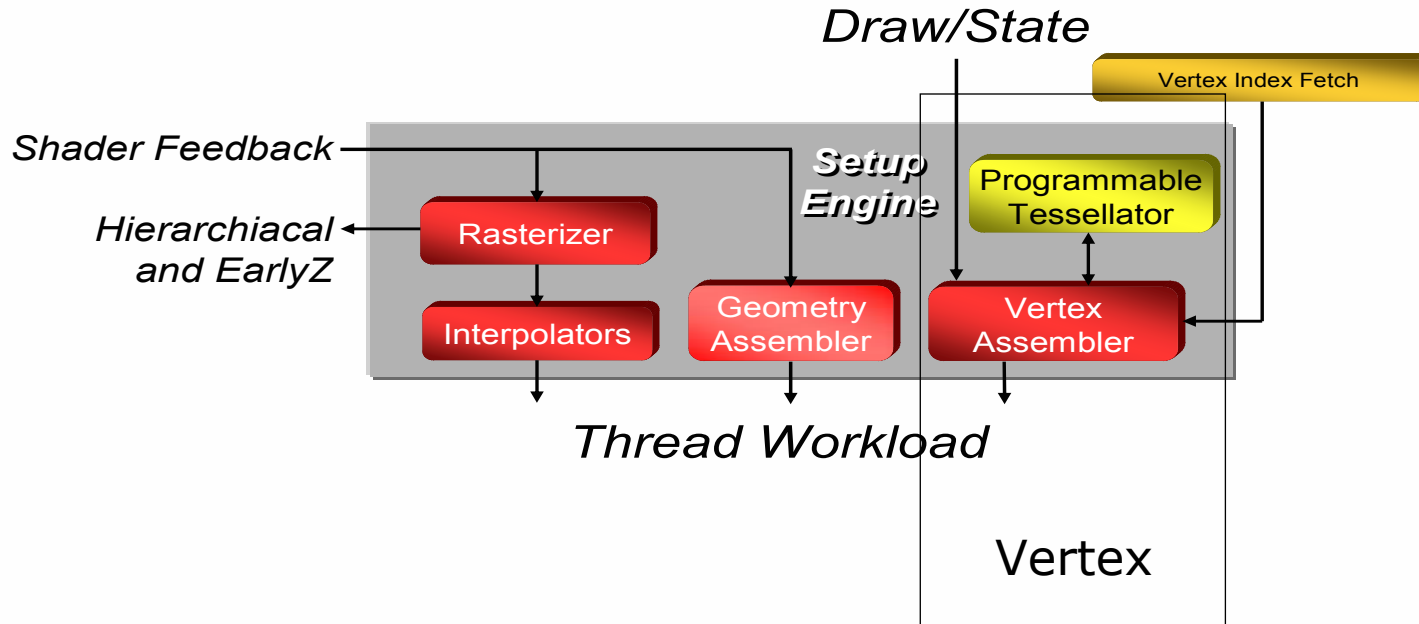


# Shader 'Setup Engine'



- 3 groups of blocks feeding 3 data streams
  - Each group feeding 16 elements (Vertices/Geometry/Pixels)/cycle

# Shader 'Setup Engine'

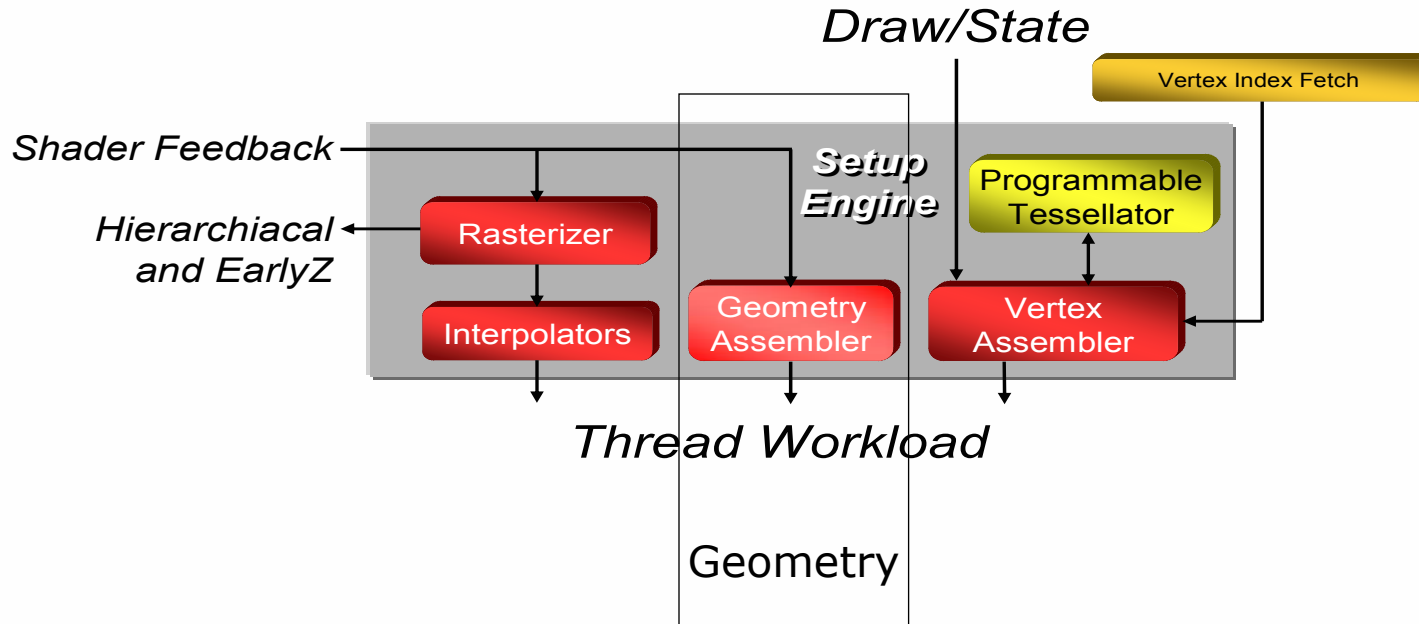


- Vertex blocks

- Primitive Tessellation
- Inputs - Index & instancing
- Sends vertex addresses to shader core

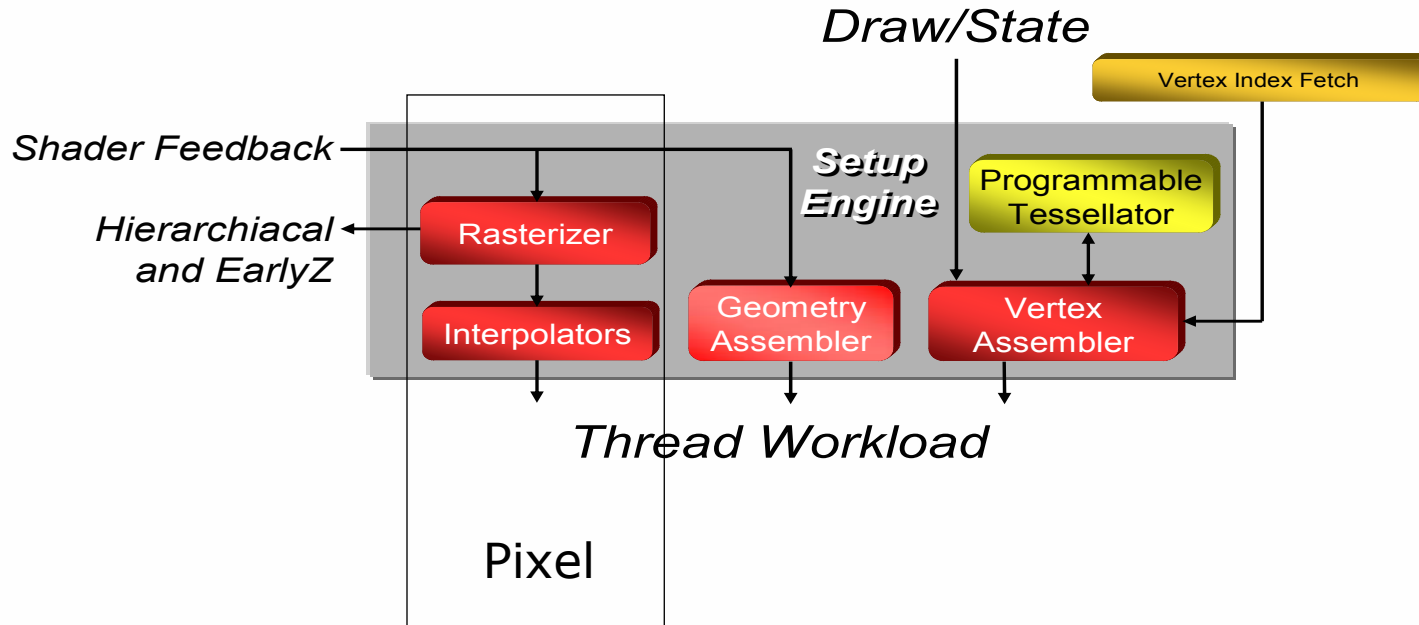


# Shader 'Setup Engine'



- Geometry blocks
  - Uses on/off chip staging
  - Sends processed vertex addresses, near neighbor addresses and topological information to shader core

# Shader 'Setup Engine'

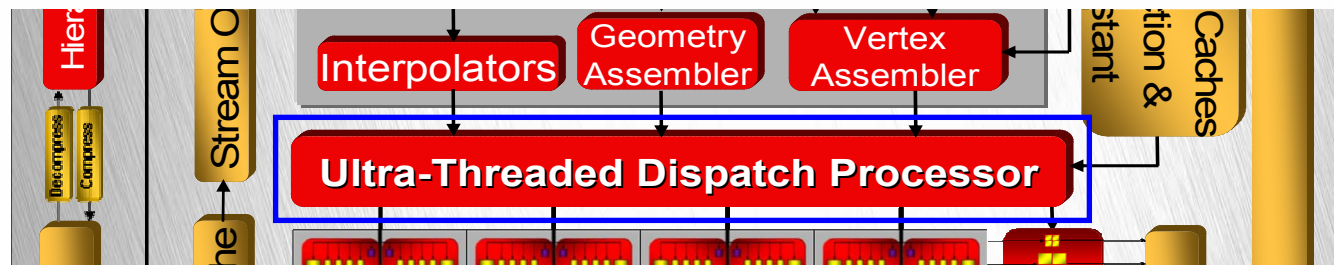


- Pixel blocks

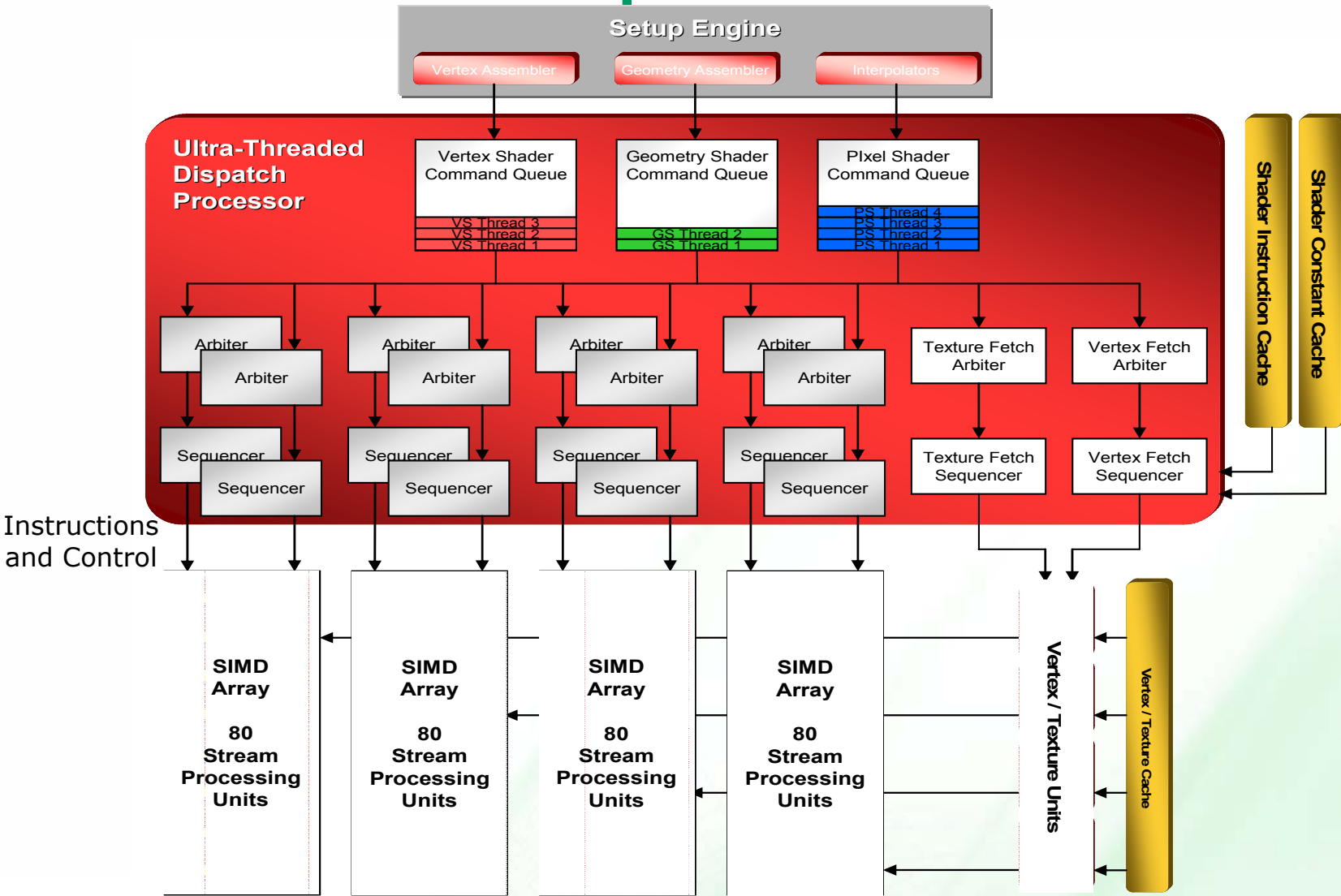
- Triangle setup, Rasterization and Interpolation
- Interfaces to depth to perform HiZ/Early Z checks

# Ultra-Threaded Dispatch Processor

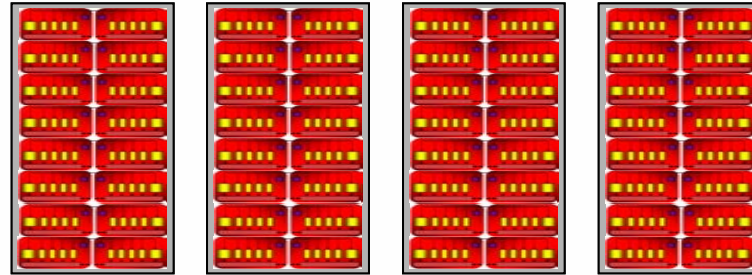
- Main control for the shader core
  - All workloads have threads of 64 elements
  - 100's of threads in flight
  - Threads are put to sleep when they request a slow responding resource
- Arbitration policy
  - Age/Need/Availability
  - When in doubt favor pixels
  - Programmable



# Ultra-Threaded Dispatch Processor



# Shader Core



- 4 parallel SIMD units
- Each unit receives independent ALU instruction
- Very Long Instruction Word (VLIW)
- ALU Instruction (1 to 7 64-bit words)
  - 5 scalar ops – 64 bits for src/dst/cntrls/op
  - 2 additional for literal constant(s)

# Stream Processing Units

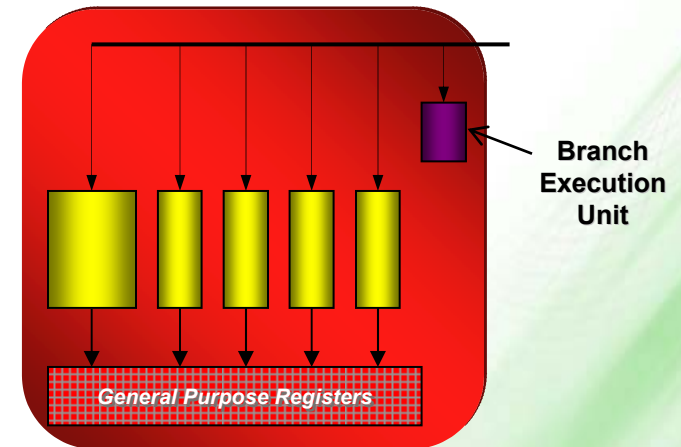
## 5 Scalar Units

- Each scalar unit does FP Multiply-Add (MAD) and integer operations
- One also handles transcendental instructions
- IEEE 32-bit floating point precision

## Branch Execution Unit

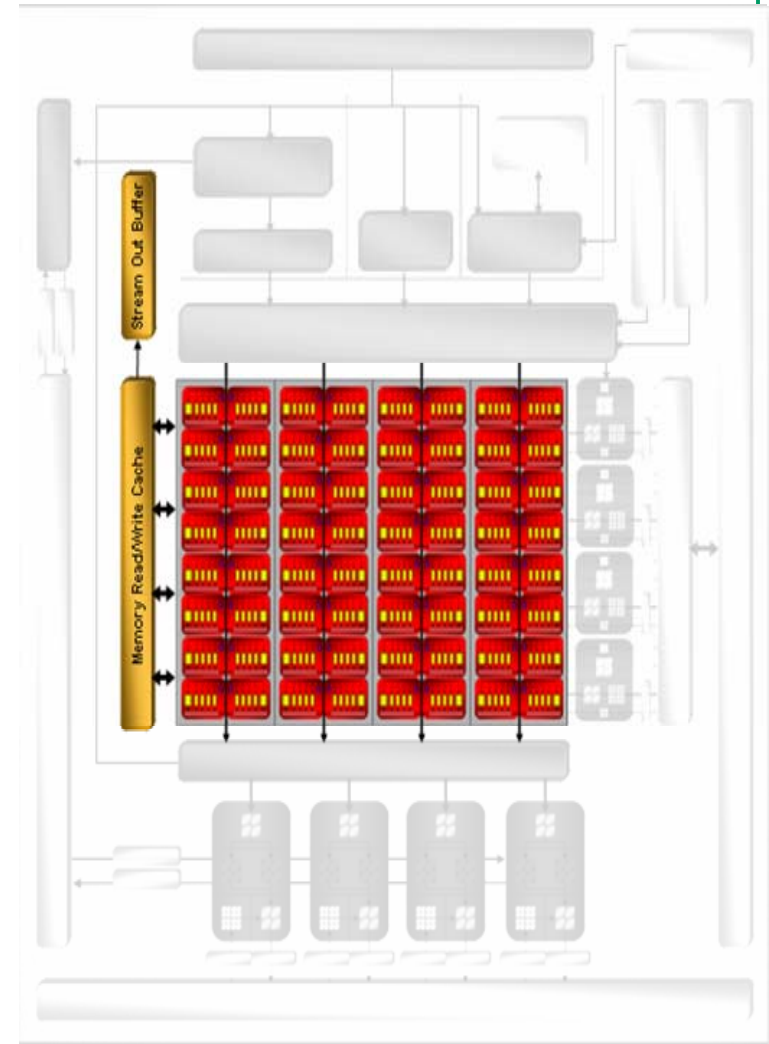
- Flow control instructions

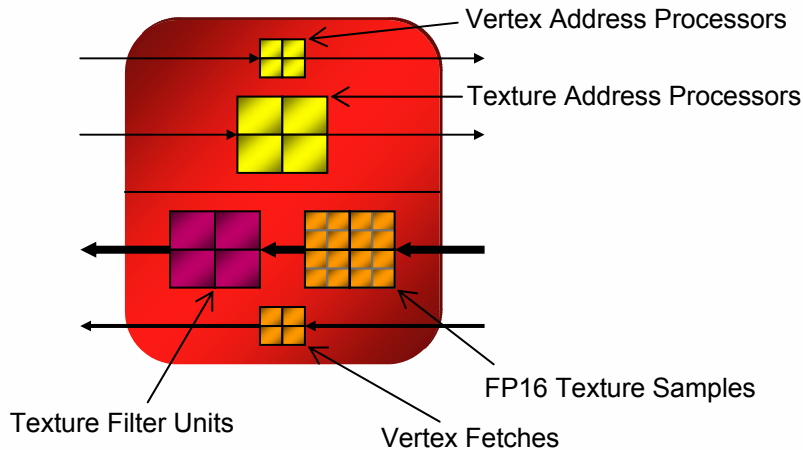
**Up to 6 operations co-issued**



# Memory Read/Write Cache

- Virtualizes register space
  - Allows overflow to graphics memory
  - Can be read from or written to by any SIMD (texture & vertex caches are read-only)
  - 8KB Fully associative cache, write combining
- Stream Out
  - Allows shader output to bypass render back-ends and color buffer
  - Outputs sequential stream of data instead of bitmaps
- Used for Inter-thread communication



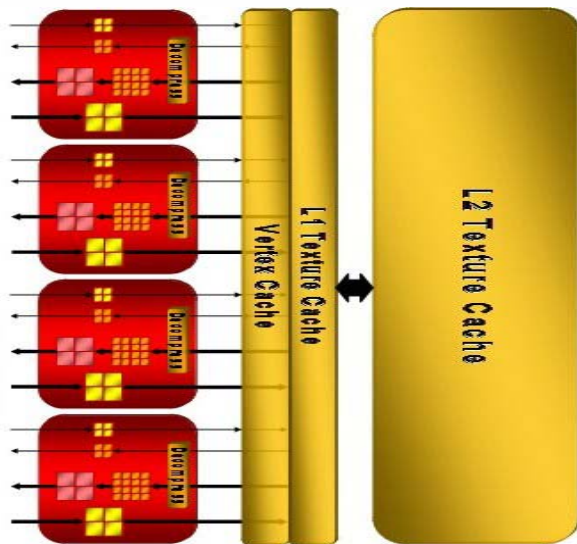


## Fetch Units

- 8 Fetch Address Processors each
  - 4 filtered and 4 unfiltered
- 20 Texture Samplers each
  - Can fetch a single data value per clock
- 4 filtered texels (with BW)
  - Bilinear filter one 64-bit FP color value per clock, 128b FP per 2 clocks for each pixel

## Fetch Caches

- Unified caches across all SIMDs
- Vertex/Unfiltered cache
  - 4kb L1, 32Kb L2
- Texture cache
  - 32KB L1, 256KB L2
  - Texture





# Render Back-Ends

Double rate depth/stencil test

- 32 pixels per clock for HD 2900
- New HiStencil

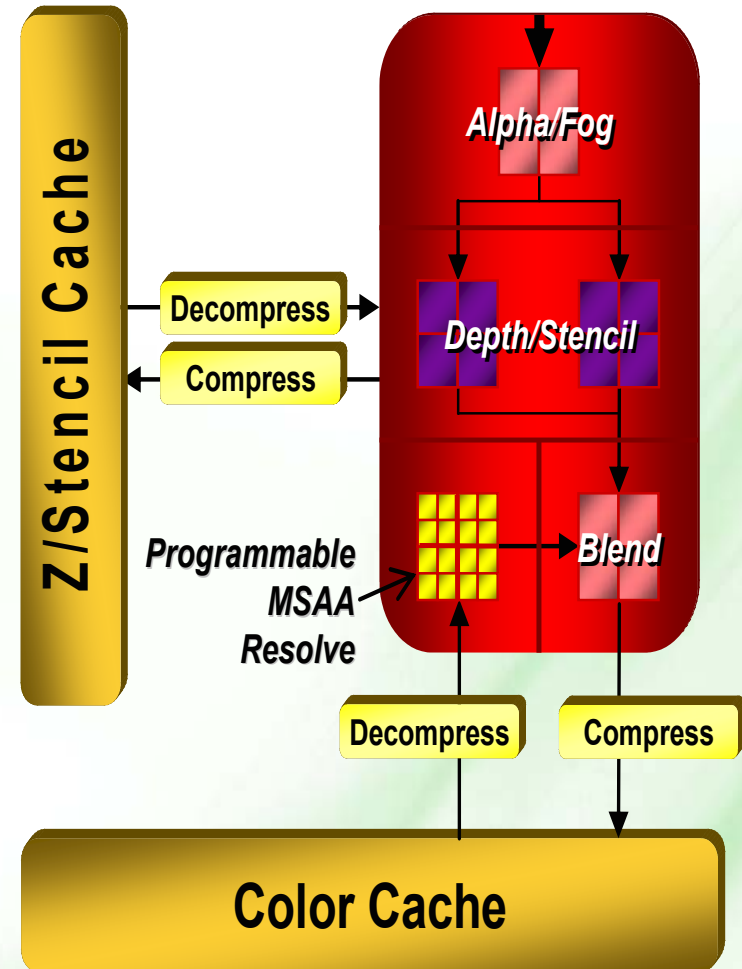
Programmable MSAA resolve

- Allows Custom AA Filters

New blend-able DX10 surface formats

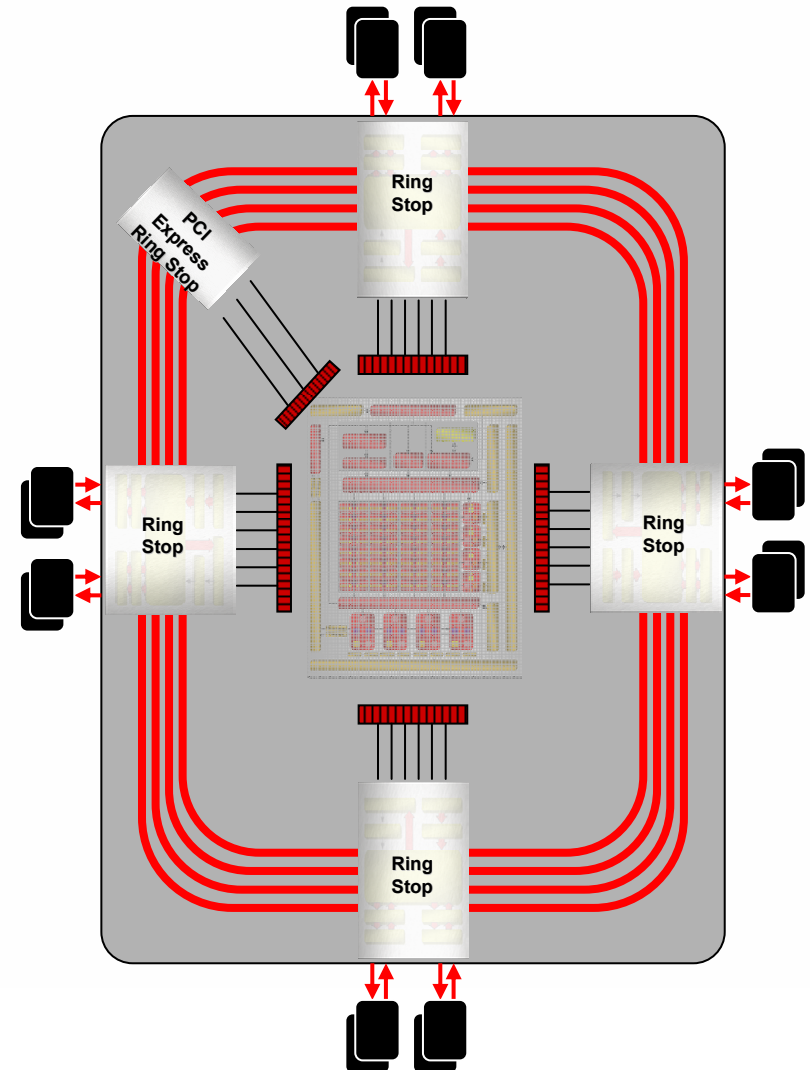
- 128-bit and 11:11:10 floating point format

Up to 8 Multiple Render Targets with MSAA support



# Memory Interface and Controller

- 512-bit Interface
  - Compact, stacked I/O pad design
  - More bandwidth with existing memory technology
  - Improved cost:bandwidth ratio
  - 8 x 64 bit memory channels
- Double ringbus
  - 512 bit read and write



# Radeon HD 2000 Series

Radeon	2900	2600	2400
Stream Processors	320	120	40
SIMDs	4	3	2
Pipelines	16	8	4
Texture Units	16	8	4
Render Backends	16	4	4
L2 texture cache (KB)	256	128	0
Technology (nm)	80	65	65
Area (mm <sup>2</sup> )	420	153	82
Transistors (Millions)	720	390	180
Memory Bandwidth	512	128	64

## Where next ?

- Move fixed functions blocks to shader
  - Improve programmability, reduce area, improve reuse, maintain/target performance
- Enhancements for GPGPU
  - Improved precision and compliance
  - New APIs, new functions
- New technologies such as 65, 55, 45, 32...
- Graphics and gaming keeps on evolving
  - DX-next is already being discussed
  - We are well into next generation and next-next generations
  - ...

# Radeon HD 2900

- Unified shader
  - Vertex, Geometry and Pixel
  - Multiple SIMD
  - 5-way scalar
- Shader cached memory read/write
- Geometry shader on/off chip storage
- 512 bit stacked I/O Memory Interface
- Full DX10 functionality

## Questions and Demo

- See more about the tessellator in Course **28. Advanced Real-Time Rendering in 3D Graphics and Games**, Natalya Tatarchuk, Wednesday
- See more about CTM in **GPGPU course**, Justin Hensley, Tuesday
  
- Thanks to Eric Demers and Mike Mantor