



Untersuchung und Vorstellung moderner Grafikchiparchitekturen

Hauptseminar Technische Informatik, 4. Juni 2008

Niko Joram

Gliederung

1. Einleitung
2. Grafikverarbeitung
 - klassische Grafikpipeline
 - Unified Shader Architektur
3. Ausgewählte Architekturen
 - nVidia GeForce Serie 9
 - ATI Radeon Serie HD3000
 - Vergleich
4. Zusammenfassung und Ausblick

Quellen

1 Einleitung

- heutige GPUs reichen bezüglich Rechenleistung an CPUs heran
- Entwicklung der Integrationsdichte übertrifft Mooresches Gesetz

Intel Core 2: 45 nm Prozess, je Kern 400 Millionen Transistoren

ATI Radeon HD3800: 55 nm Prozess, 660 Millionen Transistoren

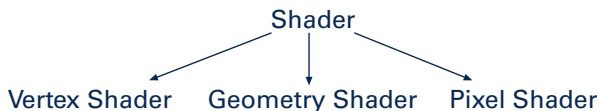
- hohes Maß an Parallelisierung (bei 3D-Berechnungen gut möglich)
- mittlerweile nicht nur für Grafik nutzbar (General-Purpose-GPU)

↪ Untersuchung der Architekturen lohnenswert

2 Grafikverarbeitung

klassische Grafikpipeline

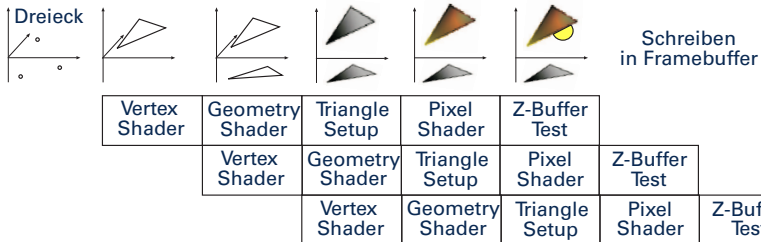
- wichtigste Verarbeitungseinheit: Shader
- Shader entsprechen spezialisierten Floating Point Units



- **Vertex Shader:** Geometriemanipulation
- **Geometry Shader:** Erzeugung zusätzlicher Geometrie
- **Pixel Shader:** Bildpunktmanipulation
- je Grafikchip gibt es mehrere Einheiten jeder Sorte

2 Grafikverarbeitung

klassische Grafikpipeline



- je Taktzyklus wird im Normalfall ein Dreieck fertiggestellt
- Erweiterung auf superskalare Pipeline

2 Grafikverarbeitung

Unified Shader Architektur

Probleme der klassischen Grafikpipeline:

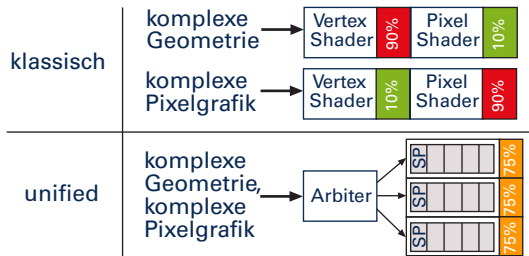
- feste Anzahl von Geometrie- u. Pixeleinheiten
- klassische Grafikpipeline zu unflexibel für moderne 3D-Grafiken
- relativ komplexe Pipelinestufen \rightsquigarrow geringe maximale Taktfrequenz

Lösung: Unified Shader Architektur

- Nutzung extrem vieler einfacher FPU's
- Zusammenfassung zu parallel arbeitenden Rechenwerken
- dynamische Funktionszuweisung durch Programmierung

2 Grafikverarbeitung

Unified Shader Architektur



- Nutzung der Shader als Streamprozessoren
- superskalare Architektur, Vektorrechner
- gleichmäßige Lastverteilung auf die Rechenwerke durch Arbiter

2 Grafikverarbeitung

Unified Shader Architektur

Probleme der Unified Shader Architektur

Hardware-Sicht:

- Arbitrer muß ständig für gleichmäßige Auslastung sorgen
- Arbitrierungsalgorithmus/-hardware komplex

Software-Sicht:

- Algorithmen für Funktionszuweisung der Shader
- eigentliche Programmierung der Shader

3 Ausgewählte Architekturen

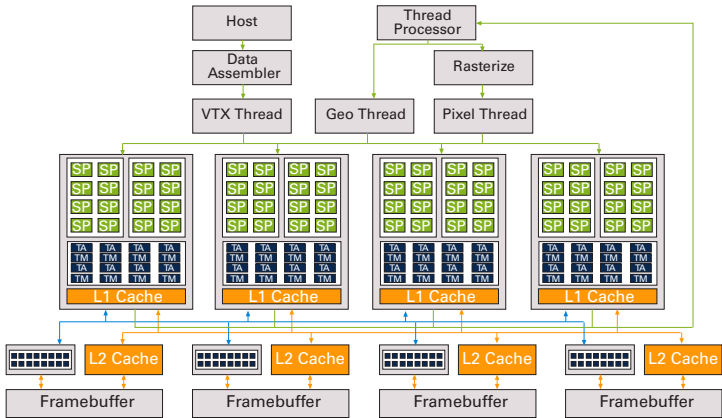
nVidia GeForce Serie 9

Eckdaten:

- 65 nm Prozess, 505 Millionen Transistoren
- Unified Shader Architektur
- 650 MHz Kerntakt (1,6 GHz Takt der ALUs)
- 64 ALUs mit 32 Bit Genauigkeit nach IEEE 754
- Controller für externe Schnittstellen integriert (RAMDAC, DVI, HDMI)
- skalierbar mit SLI (scaleable link interface)

3 Ausgewählte Architekturen

nVidia GeForce Serie 9



3 Ausgewählte Architekturen

nVidia GeForce Serie 9

Thread Processor

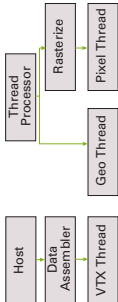
- zuständig für Verteilung des Datenstroms
- Funktionszuweisung für einzelne Shader
- übernimmt Aufgaben eines Arbiters

Vertex-, Geometry-, Pixel Thread Issue

- Vorverarbeitung für jeweiligen Datentyp
- Datenweiterleitung zu freiem Shader-Block

Data Assembler

- Auswahl von Geometrie und anderen Daten
- Zusammenfassung zu Standard-Datenstrukturen



3 Ausgewählte Architekturen

nVidia GeForce Serie 9

Ausführungseinheit

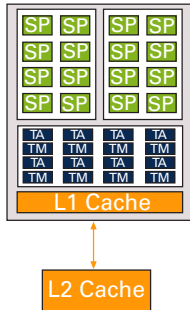
- jede kann andere Aufgabe übernehmen
- Anzahl problemlos skalierbar
- Caches halten beliebige benötigte Datenstrukturen
- Datenfluß selbst ist *streaming* (nicht *cached*)

Stream Processor (SP)

- jeder SP kann je Takt Multiply-Add und Multiply

Textur-Verarbeitung

- selbständiges Prefetching (TA)
- selbständige Filterung (TM)
- SP werden dafür nicht benötigt

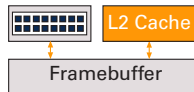


3 Ausgewählte Architekturen

nVidia GeForce Serie 9

Raster Operation Processor

- Anzahl skalierbar
 - jeder besitzt mehrere Anti-Aliasing-Einheiten
 - Kompressionsalgorithmen für anfallende Daten
 - Speichercontroller, Schreibpuffer, L2 Cache
- ↪ keine Konflikte untereinander



Funktionalität

- Anti-Aliasing (bis zu 12 Pixel/Takt)
- Z-Test (32 Bit Z-Buffer)
- Farbverwaltung
- Schreiben der fertigen Pixel in Framebuffer

3 Ausgewählte Architekturen

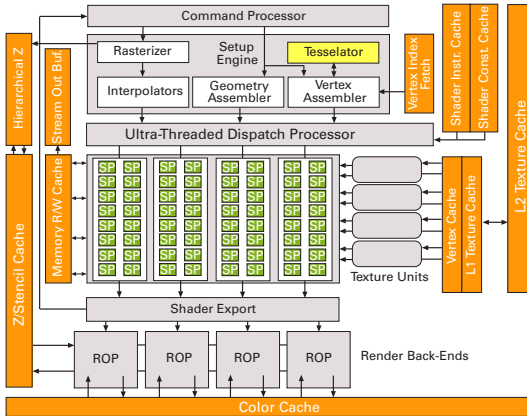
ATI Radeon Serie HD3000

Eckdaten:

- 55 nm Prozess, 660 Millionen Transistoren
- Unified Shader Architektur
- 775 MHz Kerntakt
- 64 ALUs mit 32 Bit Genauigkeit nach IEEE 754
- Skalierbarkeit mit CrossFire (mehrere GPUs im Verbund)

3 Ausgewählte Architekturen

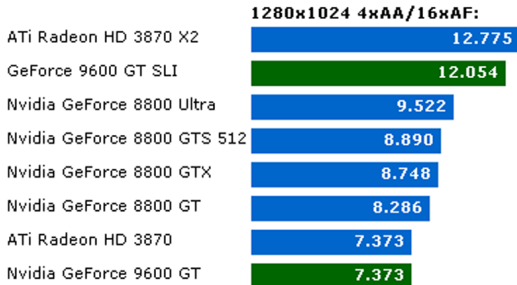
ATI Radeon Serie HD3000



3 Ausgewählte Architekturen

Vergleich

Benchmark mit 3DMark06



4 Zusammenfassung und Ausblick

- moderne GPUs sind CPUs bezüglich Parallelverarbeitung überlegen
- **Unified Shader Architektur**: skalierbarer Verbund vieler Recheneinheiten
- jede Recheneinheit kann frei programmiert werden

↪ Nutzung als schneller, paralleler Rechner nicht nur für Grafik

CUDA (Compute Unified Device Architecture) von nVidia

- API und Compiler
- enthält Bibliotheken mit Algorithmen (z.B. FFT, DCT, ...)
- spezieller Modus der Hardware (keine Zerlegung in Pixel)

Quellen

Beschreibung der Architektur des nVidia G80, G94

<http://www.digit-life.com/articles2/video/g80-part1.html>

<http://www.digit-life.com/articles3/video/g94-part1.html>

Beschreibung der Architektur des Radeon R600, R670

<http://www.digit-life.com/articles2/video/r600-part1.html>

http://www.computerbase.de/artikel/hardware/grafikkarten/2007/test_ati_radeon_hd_3850_rv670/2

allgemeines zu GPUs

<http://de.wikipedia.org/wiki/Grafikprozessor>

http://en.wikipedia.org/wiki/Graphics_processing_unit

Unified Shader Architektur

<http://de.wikipedia.org/wiki/Shader>

<http://www.computerbase.de/bild/article/650/6>

Streamprozessor

<http://de.wikipedia.org/wiki/Streamprozessor>