Massively Parallel Computing on the FUZION Chip

Ray McConnell
CTO
PixelFusion plc
Overview

- History
- Architecture
- Graphics applications
- Future
- Conclusion
Thoroughbred Technology

- Architecture has evolved over 15 years
  - ’99 FUZION1 - First Product
    • Evolutionary single/multi chip solution
  - ‘97 Pixel Flow
    • 8 bit ALU, Sub pixel composition network
  - ’91 Pixel Planes 5
    • 1 bit ALU, Multiple regions connected via crossbar
  - ’86 Pixel Planes 4
    • 1 bit ALU, full 512x512 processor instantiation
  - ‘82 Logic Enhanced Memory Patents filed
    • Assignees: UNC (University of North Carolina)

- Unique patented portfolio
  - 30+ additional patents filed(Architecture+Algorithms)
  - More in process
Architecture - System

FUZION Core

Microcode Store

Array Controller

Channel Controller

Thread Manager

Binning Unit

N=6 FUZION Blocks

FUZION BUS

AGP / PCI Core Logic

Display Unit

Pixel Select & Format

Palette & Gamma RAM

Video DAC

Audio Unit

VIOP FBI

Video Output Port

Video Input Port

VIOP Source Muxing

AGP Master

PCI Master

AUX BUS

EPU

ARC Core + Caches

Thread Data

Thread Control

Thread Instruction Feed

AGP Master

PCI Master

FUZION BUS

LOCAL MEMORY

RI[3]

Direct Rambus RMC

Direct Rambus RAC

RI[0]

Direct Rambus RMC

Direct Rambus RAC

PCI Master

Thread Instruction Feed

Thread Control
Architecture - System EPU

- Generic 32bit EPU (ARC CPU)
  - Embedded Processor Unit (Synthesised)
  - Data and instruction caches

- Tightly coupled EPU to SIMD Core
  - Low latency multitasking OS (EOS)
  - Exception control
  - Memory management
  - Host protocol management
  - High quality real time emulation of VGA
    - Uses SIMD Core
  - Spare processing cycles on EPU for application extensions
Architecture - Core

Thread Manager

Array Controller

Load/Store Controller

Channel Controller

Fuzion Block (1 of 6)

LEE Coefficients

Linear Expression Evaluator

256 Processor Elements

SEQ

Channel

Binning Unit

PE Memory

Fuzion Bus

instructions
Architecture - Blocks

Linear Expression Evaluator

Processor Element

PE Register File

PE Page Reg

PE Page Reg

Processor Element

Register File

... 256 PEs

swizzle network to next block

LEE Stage

... Block I/O Bus

Block I/O buffer

sense amps

PE Memory:
2 KBytes DRAM

... Fuzion Bus

sense amps

Block I/O Bus

Transfer Engine

... Binning Unit

Fuzion Bus Interface

... Fuzion Bus

... Binning Unit

Fuzion Bus

... Transfer Engine

... Block I/O Bus

... swizzle network to next block

... 256 PEs
Architecture - LEE

- Replaces interpolation
  - Evaluates $Ax + By + C$ directly at every PE

- Drawing triangles
  - Evaluate edge equations

- Z-buffering, Gouraud shading
  - All linear in screen space, e.g. $Z = Ax + By + C$

- Perspective-correct texturing
  - Texture coordinates = $Ax + By + C$

- LEE accelerates all rasterization operations
  - Also can be used for data distribution functions
Architecture - PE

- ALU (8 bits)
- Register File: 32 bytes
- PE Memory: 2KByte DRAM
- Block I/O Channel
- LEE Feedback
- Instructions
- LEE Result

- Connections:
  - Left PE: 16 bytes
  - Right PE: 16 bytes
  - ALU: 8 bits
Characteristics – Key points

- PE usage efficiency
  - Algorithmic

- PE data flow efficiency
  - Architectural

- I/O Latency efficiency
  - Data driven instruction issue control

- Multithreading
  - Control units interact via semaphores

- Enough PE data workspace
  - 2K Bytes DRAM
  - More complex PE
Characteristics - Silicon

- >500mm²
- 35 Watts
  - Forced Air
- 1036 Pins
  - HPBGA
- ~70M Transistors
- eDRAM process
  - Custom DRAM
- Redundancy
  - PE units + DRAM
  - Yield management
Graphics - AGP Pro Thermal Analysis
Architecture - Compute numbers

- 1.5 TeraOPS \((1.5 \times 10^{12})\) 8 bit multiplies and adds
  - LEE + PE
  - 12K bit data-path \((8 \times 1536)\)
- 600 Gigabytes/sec on chip DRAM bandwidth
  - 50k bit wide data-bus \((32 \times 1536)\)
- 1.2 Terabytes/sec PE register file data bandwidth
- 1.2 Terabytes/sec inter PE data bandwidth
- >3 GigaFlops (IEEE compatible)
  - Can be faster with custom format
- 7 GMACs \((16 \times 16)\)
- 150M 3D Transformations/sec
- 50M Tris/sec
Graphics Applications

- High quality, standard OpenGL
  - 100% compliant v1.2
- High quality, standard Direct 3D
  - 100% compliant DirectX 7
- High quality, standard 2D windows
  - 100% GDI+ compliant
  - Full 256 ROP acceleration
  - Full line & text acceleration with transparency
  - Sophisticated 2D filtering and scaling
- High quality, high resolution MPEG II
- Software is ultimate future-proofing
  - Radical extendibility
  - Completely web revision controlled
SIMD OpenGL

- 100% OpenGL v1.2
  - Region based
  - Efficient Bin sort
  - Bin memory management
  - Efficient with small triangles
  - Blending and texturing is maximally efficient

- Standard extensions
  - Fragment lighting (per pixel shade)
  - Light textures
  - Vertex co-ordinate frames and normal perturbation
  - Multi-sample AA
  - Multi texture and shadow maps
SIMD OpenGL

- Unbounded API extensions
  - Customisable pipeline for lib function mix-match
  - New primitive types (Curved surfaces)
  - User programmable shaders

- Multiple alpha blended underlay & overlay planes

- In field delivery of API enhancements
  - WWW

- Optional fast/highest-quality anti-aliasing

- Optional geometrical processing

- Scaleable software design
  - Will operate on larger/smaller SIMD arrays
Silicon Schedule

- Tapeout Expected by Oct. ’99
- First Silicon early December
  - UMC/USIC .25 µ eDRAM Fab
- First Boards Q1 00
- Full System Qualification and optimized OpenGL & Direct3D in Q2 00
Future – Graphics

- Advanced API acceleration
  - Fahrenheit low level (when available)
- Advanced Anti-Aliasing
  - Extremely high resolution coverage sampling
- Advanced 3D effects
  - Depth of field, motion blur, custom shaders, ..
- Advanced shading
  - Torrance Sparrow micro-facet modelling
- Advanced texturing
  - Image based rendering (texture with Z)
- Advanced geometry
  - Function (particles) or map driven
- Volumetric visualization
- 2D within a 3D desktop -natural next step
  - Readable text in perspective
Future - Silicon

- Larger and faster SIMD arrays
- Smaller and cheaper SIMD arrays
- PE extensions
- Multi-chip implementations
- Target Market Variations
  - Core IP licensable
  - Third party IP integration
Future - Software Development Kit

- SDK availability
  - Library of software modules
    - 2D, 3D (Gems), Imaging, Numerical etc.
  - EPU C Compiler
    - EOS extensions
  - Simulators
    - PC Farm acceleration
  - Assemblers
  - Debuggers
    - Hierarchical views of PE register/memory
  - Performance monitors
    - hardware assisted

- SIMD C++ scheduled for '00
  - Syntax extensions
  - Explicit Multi-threading
Future - Applications

- Widening application space
  - Advanced Signal, 2D Image and Video processing
  - Advanced motion detection for video compression
  - Fast fuzzy data search/convolution (DNA sequencing)
  - Machine vision
  - Artificial neural networks
  - Encryption/Decryption/Compression/Decompression

- Can make use of installed graphics hardware
  - CAD acceleration
  - PC Farm super-computing
Future price/performance

Price/Performance > 3x / Year
Conclusion

- Open functionality and re-use of software
  - Real applications, real markets
  - Performance compares with hardwired functions
  - Market scope larger than 3D graphics
  - Software will port rapidly to future generations

- Roadmap into future, beyond 0.13u technologies
  - Verification task massively simplified with replication
  - Highly regular custom silicon layout
    • Deep sub-micron design barrier solution
  - 5k man/gate/day productivity (from ground zero)
    • Higher on next generations
  - Highly parallel embedded test (Using EPU)

- Redundancy for yield management
  - Cost reduction on large die