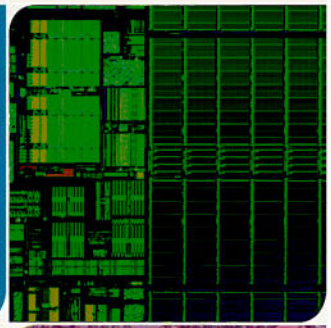
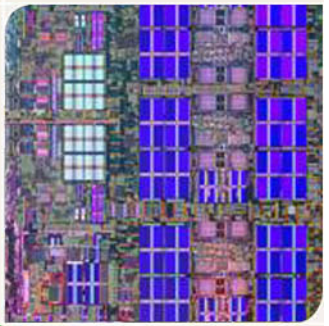
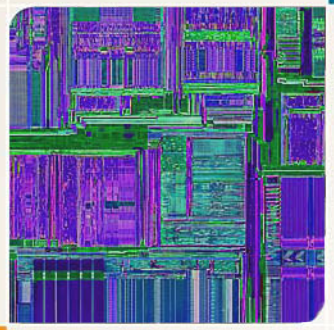
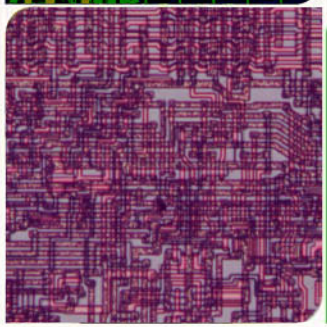
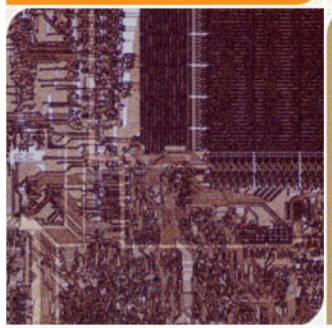





CS4803DGC Design Game Consoles

Spring 2009

Prof. Hyesoon Kim

Georgia Tech College of Computing



Amdahl's Law (1)

$$\text{Speedup} = \frac{\text{Execution Time without Enhancement}}{\text{Execution Time with Enhancement}} = \frac{\text{Execution Time}_{\text{old}}}{\text{Execution Time}_{\text{new}}}$$

What if enhancement does not enhance everything?

$$\text{Speedup} = \frac{\text{Execution Time without using Enhancement at all}}{\text{Execution Time using Enhancement when Possible}}$$

$$\text{Execution Time}_{\text{new}} =$$

$$\left(\text{Execution Time}_{\text{old}} \times (1 - \text{Fraction}_{\text{Enhanced}}) + \text{Execution Time}_{\text{old}} \times \frac{\text{Fraction}_{\text{Enhanced}}}{\text{Speedup}_{\text{Enhanced}}}\right)$$

$$\text{Overall Speedup} = \frac{1}{\left((1 - \text{Fraction}_{\text{Enhanced}}) + \frac{\text{Fraction}_{\text{Enhanced}}}{\text{Speedup}_{\text{Enhanced}}}\right)}$$



Question

- Profile data (before parallelizing application)

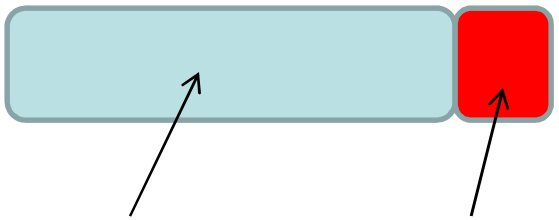
  func(A): 20% → non parallelizable section

  func(B): 80% → parallelizable section

Maximum speedup using 8-core processor?

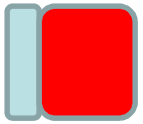
Maximum speedup using 128-core GPU?

Answer



Parallel Sequential

80 sec 20 sec

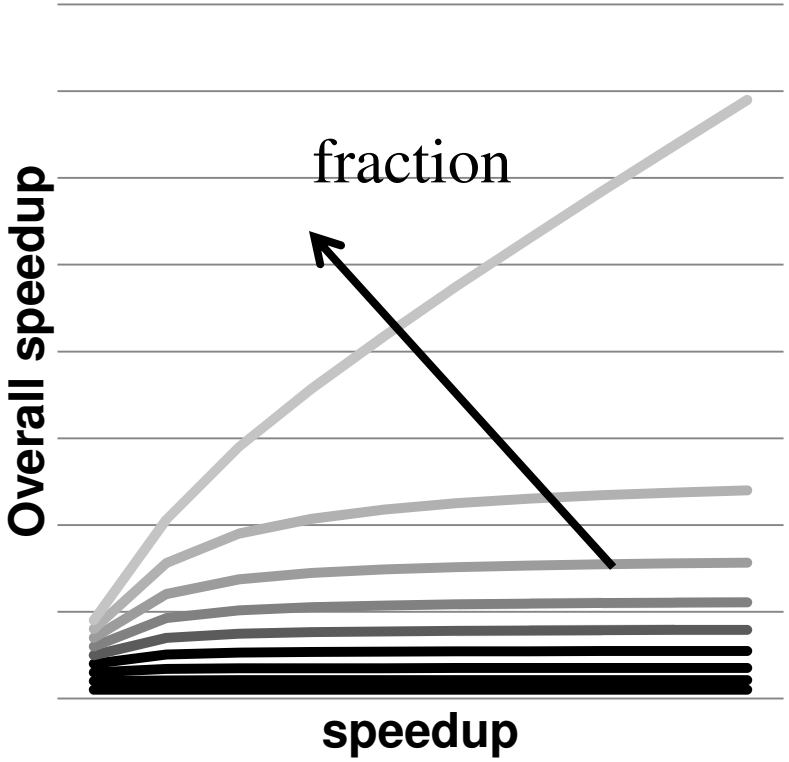


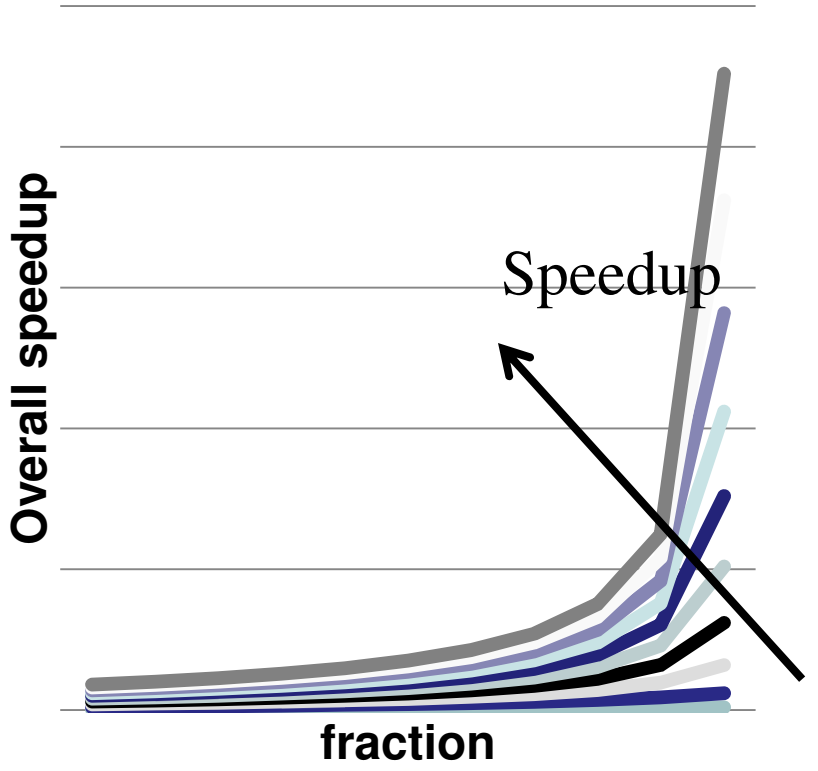
10 sec 20 sec

0.625 sec 20 sec

Amdahl's Law (2)

$$\text{Overall Speedup} = \frac{1}{\left((1 - \text{Fraction}_{\text{Enhanced}}) + \frac{\text{Fraction}_{\text{Enhanced}}}{\text{Speedup}_{\text{Enhanced}}}\right)}$$







Heterogeneous Architecture

- Fast core: Sequential portion
- simple cores: parallel portion

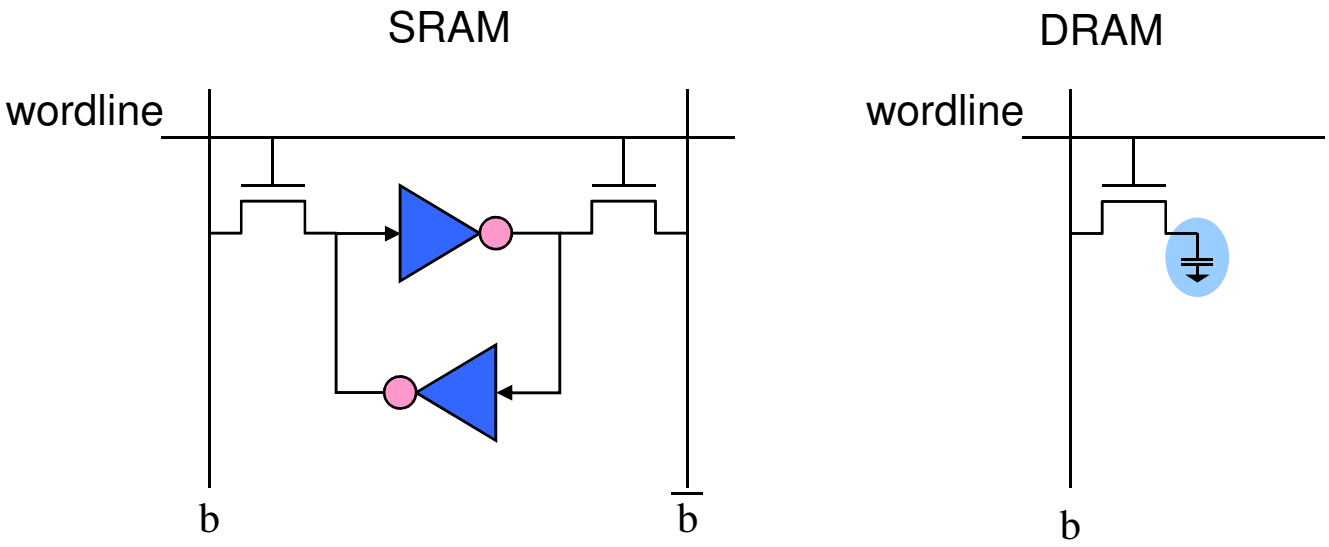


DRAM MEMORY CONTROLLERS

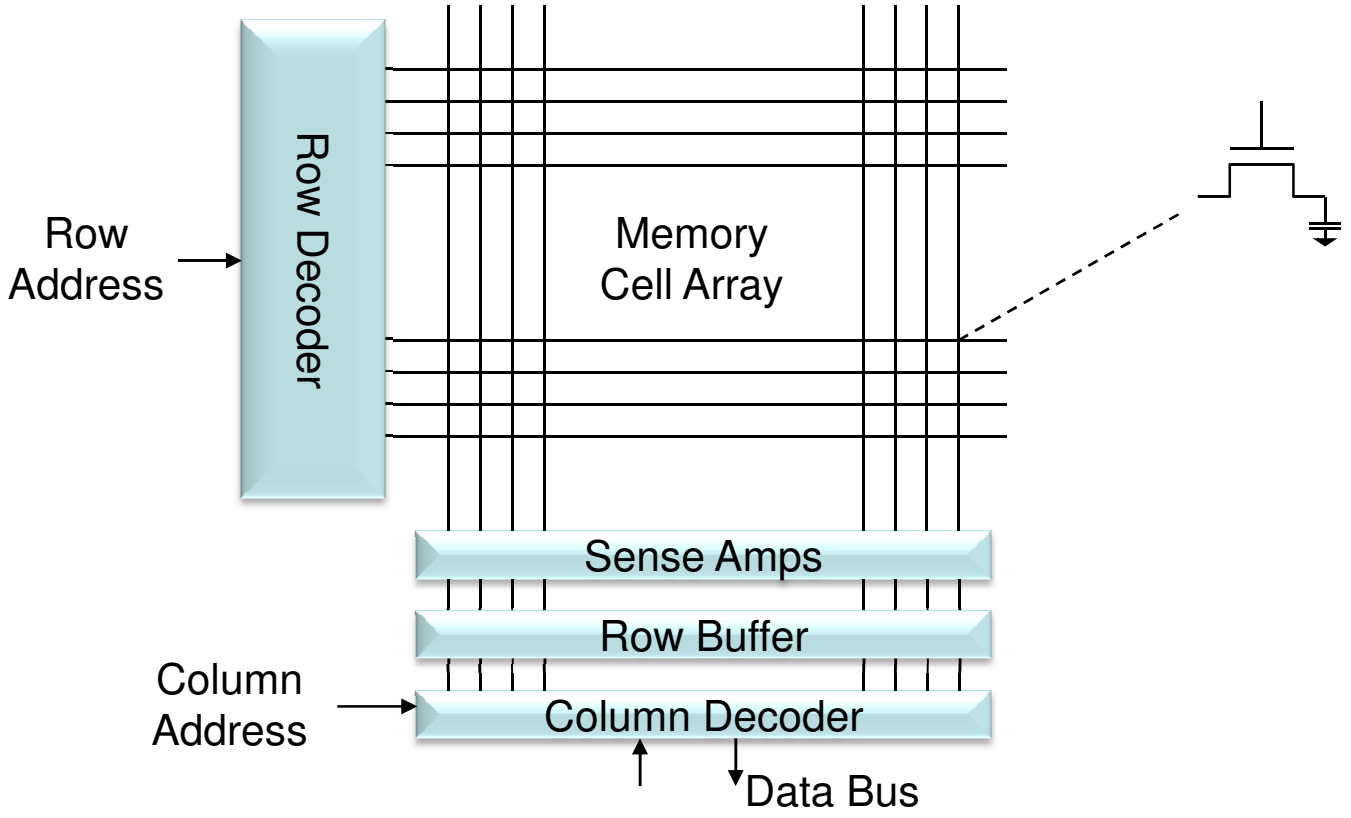
Georgia Tech College of Computing



REVIEW: Hardware Structures



REVIEW: DRAM Chip Organization

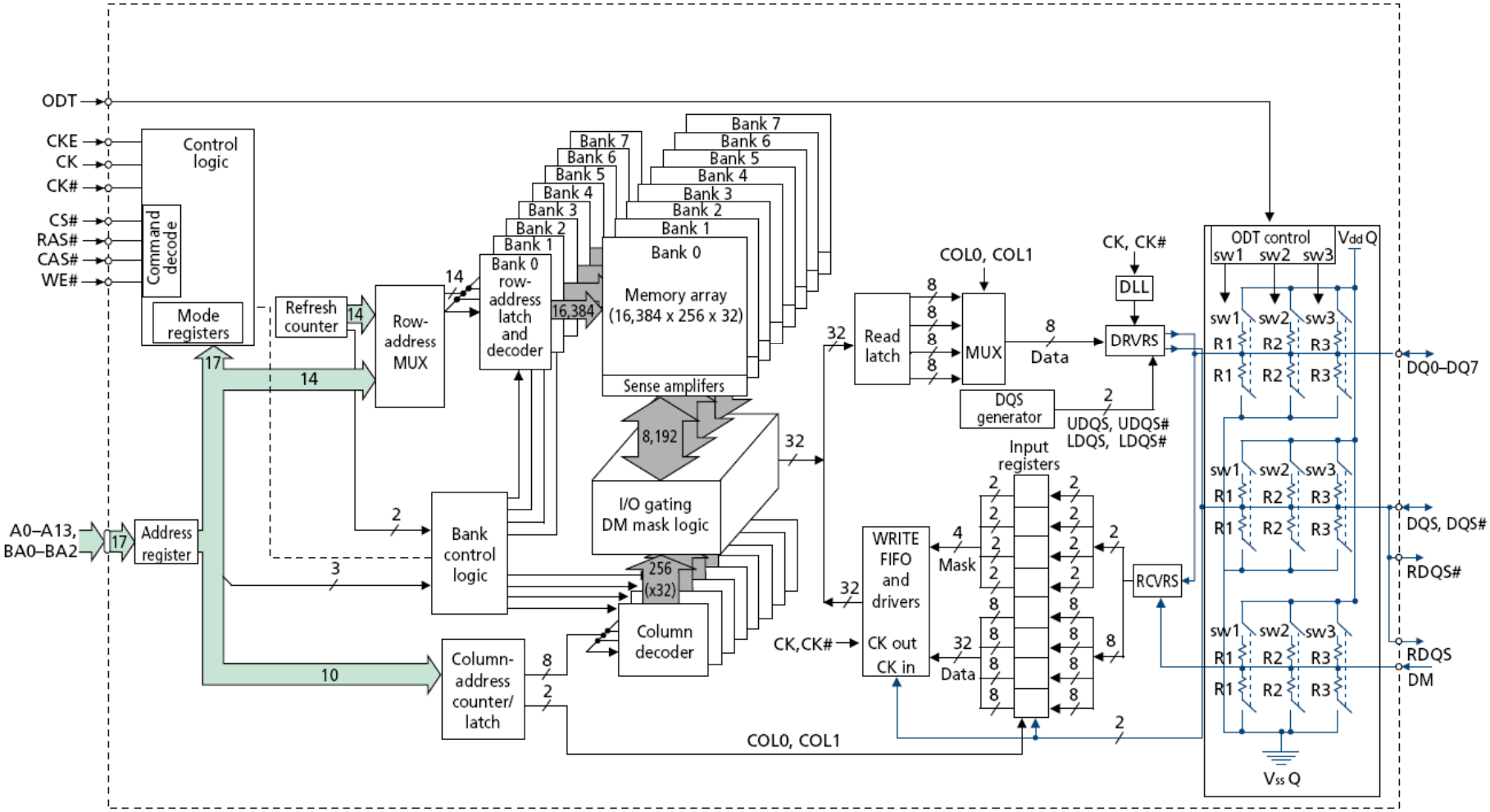




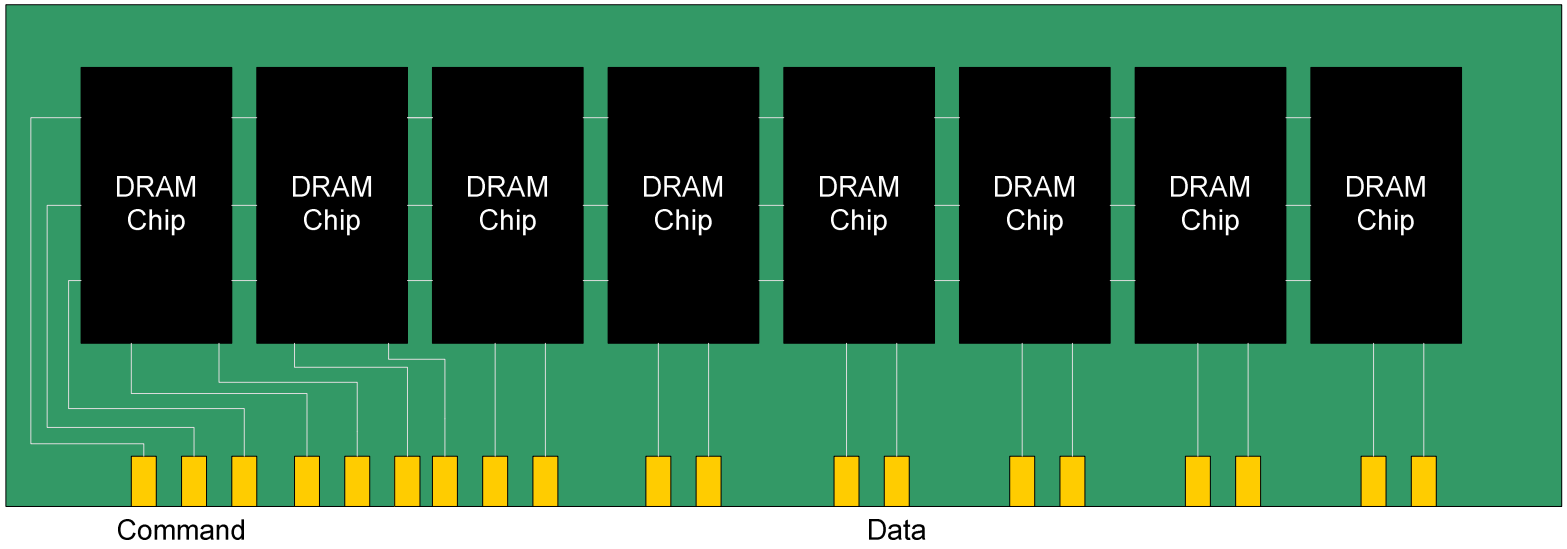
Duty of Memory Controllers

- Memory requesting scheduling
  - Ordering memory requests
  - FCFS, FRFCFS etc.
- Send commands
  - Read, write, CS
- DRAM refresh counter control

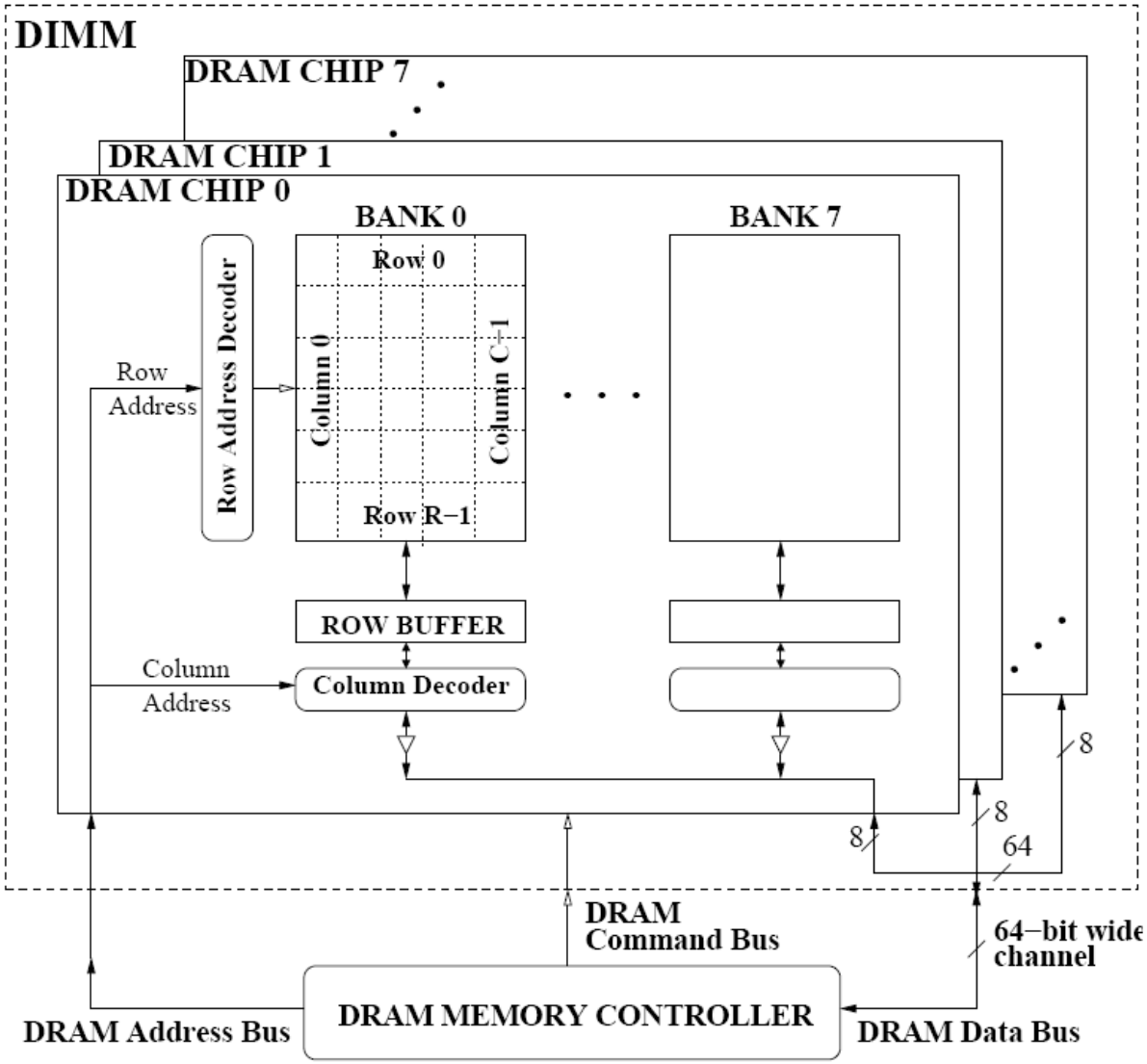
128M x 8-bit DRAM Chip



A 64-bit Wide DIMM

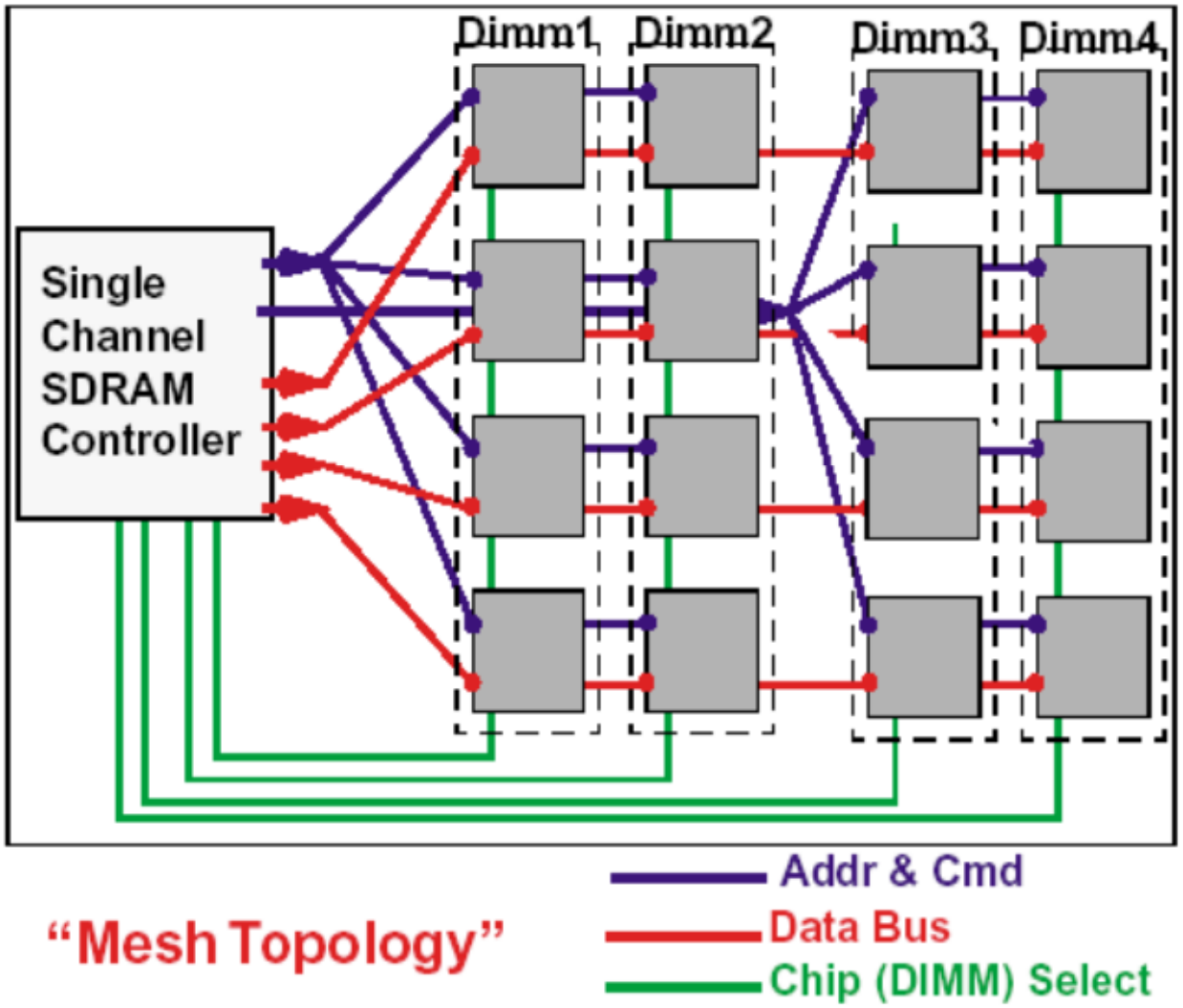


A 64-bit Wide DIMM



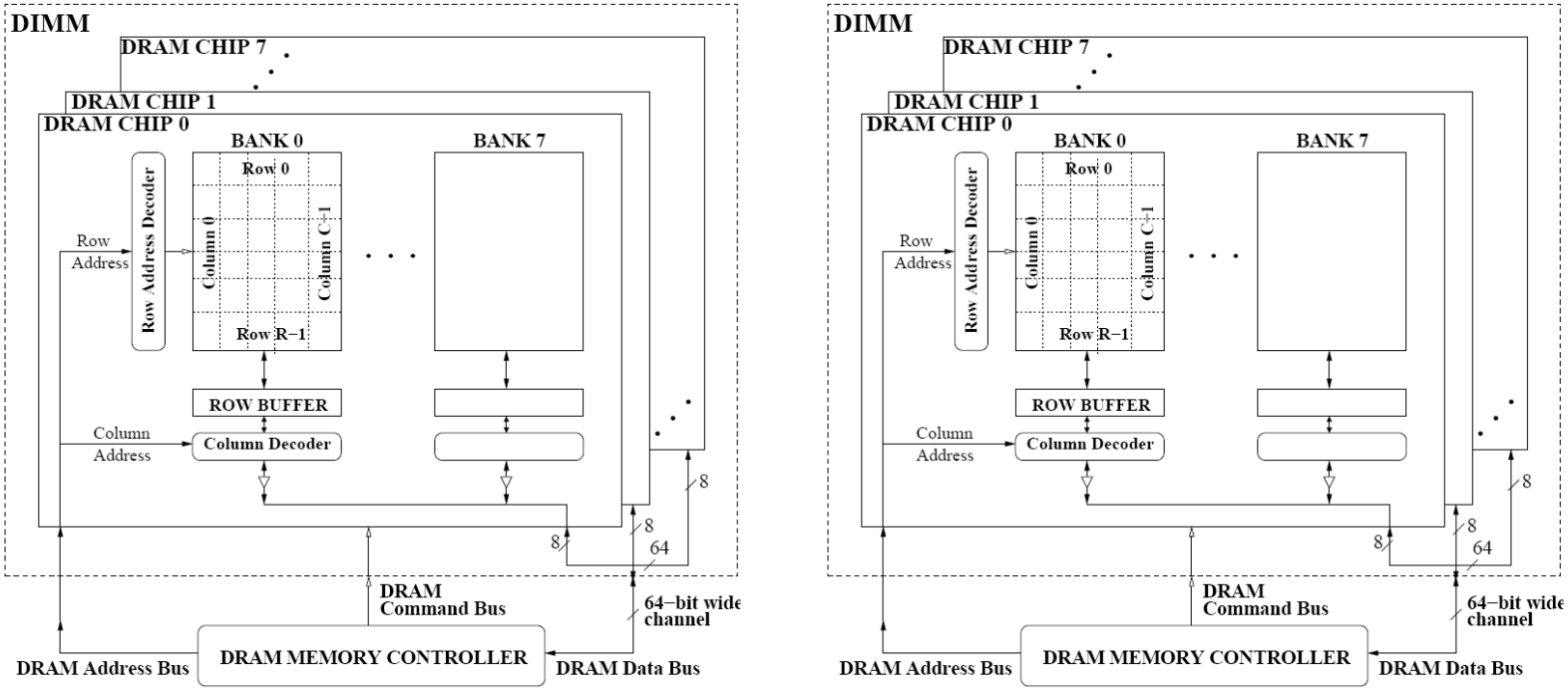
- Advantages:
  - Acts like a high-capacity DRAM chip with a wide interface
  - Flexibility: memory controller does not need to deal with individual chips
- Disadvantages:
  - Granularity: Accesses cannot be smaller than the interface width

Multiple DIMMs



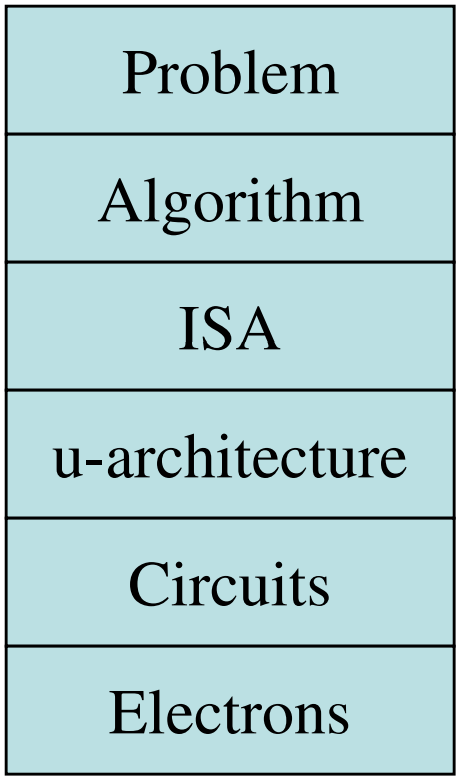
- Advantages:
  - Enables even higher capacity
- Disadvantages:
  - Interconnect complexity and energy consumption can be high

DRAM Channels



- 2 Independent Channels: 2 Memory Controllers (Above)
- 2 Dependent/Lockstep Channels: 1 Memory Controller with wide interface (Not Shown above)

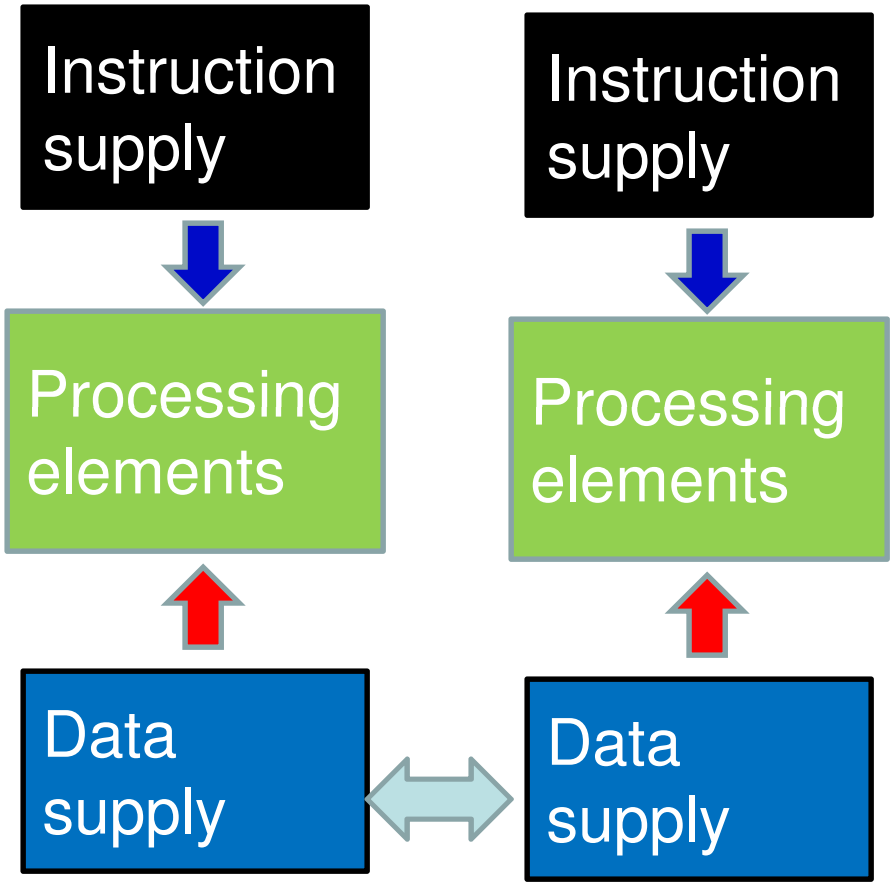
Computer Architecture Research



- Architecture
  - Core design
  - Performance evaluations
  - Simulation
- Architecture + VLSI/circuit
- Compiler + architecture
- O/S + architecture



Processor Design





More Detail

- Instruction supply
  - Branch predictors
  - Instruction prefetch, ooo fetch
- Data supply
  - Data prefetch
    - Speculative execution
  - Cache compression, replacement policy, increase bandwidth
  - Memory scheduler/controller, performance, fairness



More Detail

- Core design
  - Increase ILP: scheduling, register renaming, cluster, load store queue design, memory disambiguation, data-flow graph
  - Increase MLP: speculative execution, run-ahead execution,
  - Increase TLP: SMT, transactional memory
  - High performance, energy efficiency
- Interconnect
  - Memory, cache connections



More Detail

- New technologies
  - 3D technology
  - Bio/nano
  - Quantum computing
  - New memory technology (PCM)
- Reliability
- Performance debugging
  - Deterministic exeuction
- Security
- Programmability



REVIEW



Requirements for GC

- Time constrain
- Lots of Data
- Heavy use of graphics
- Both Integer/floating point operations are important
- Floating point → low precision
- Stream applications
- Embedded systems
- Various I/O devices
- No comparability issues (no reason to support legacy code)
- All the platform is stable:
- Platform optimizations

Xbox 360 System Block Diagram

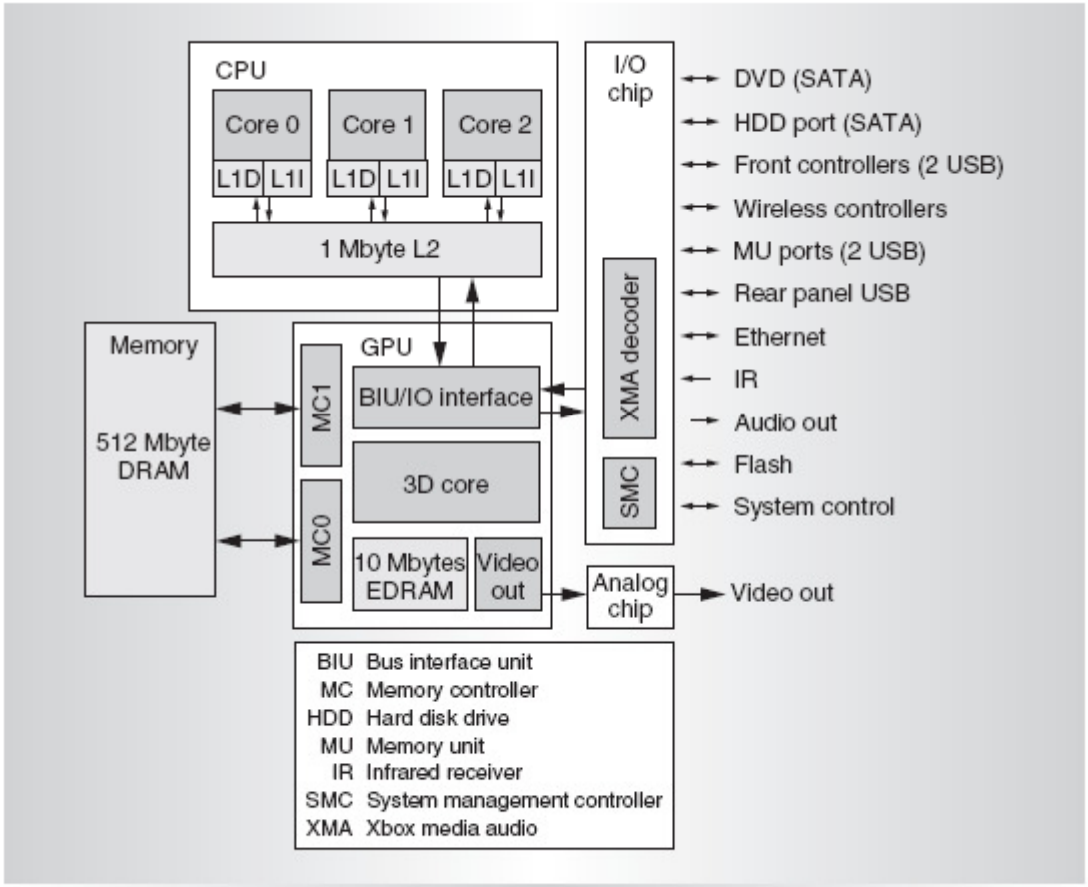


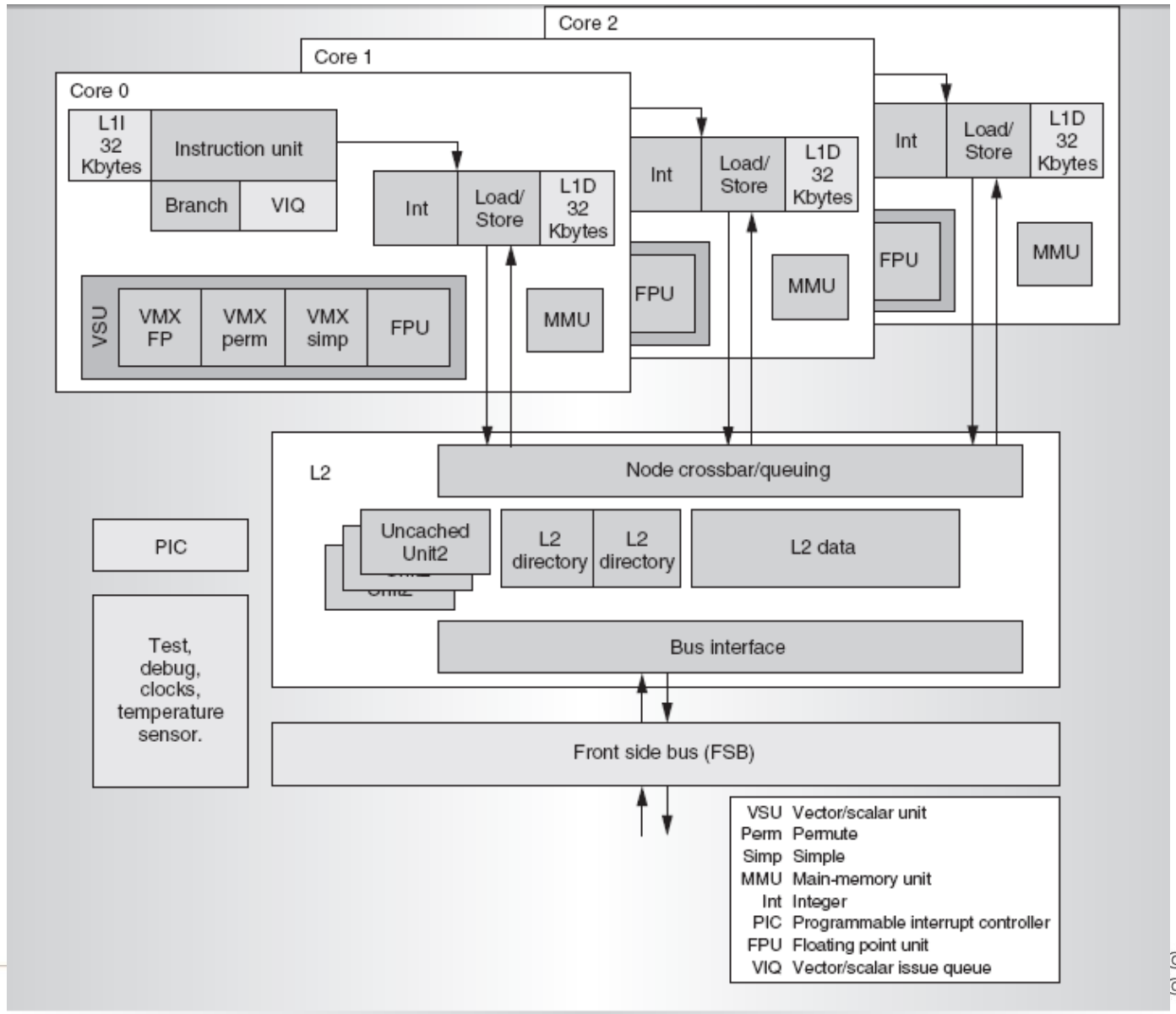
Figure 2. Xbox 360 system block diagram.



Xbox 360 Architecture

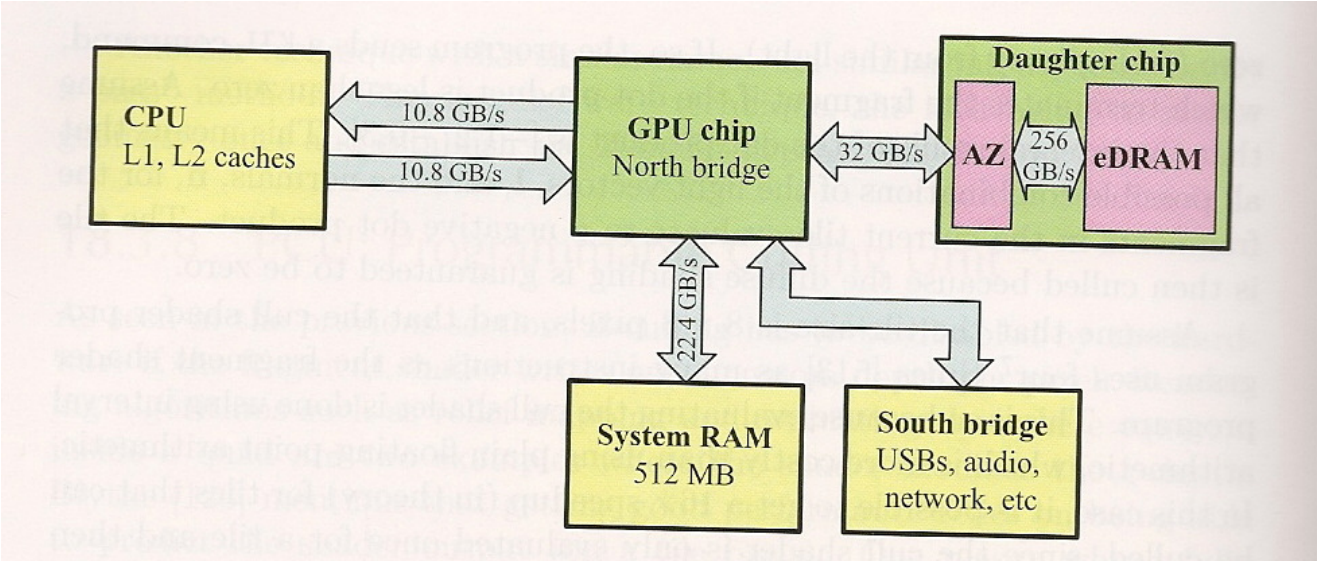
- 3 CPU cores
  - 4-way SIMD vector units
  - 8-way 1MB L2 cache (3.2 GHz)
- 48 unified shaders
- 3D graphics units
- 512-Mbyte DRAM main memory
- FSB (Front-side bus): 5.4 Gbps/pin/s
- 10.8 Gbyte/s read and write

Xbox 360 CPU Block Diagram



Xbox 360 Memory Architecture

- embedded DRAM(eDRAM): frame buffers,
- 10MB
- Daughter chip: AZ: all alpha and Depth testing.



Xbox 360 Graphics Processors

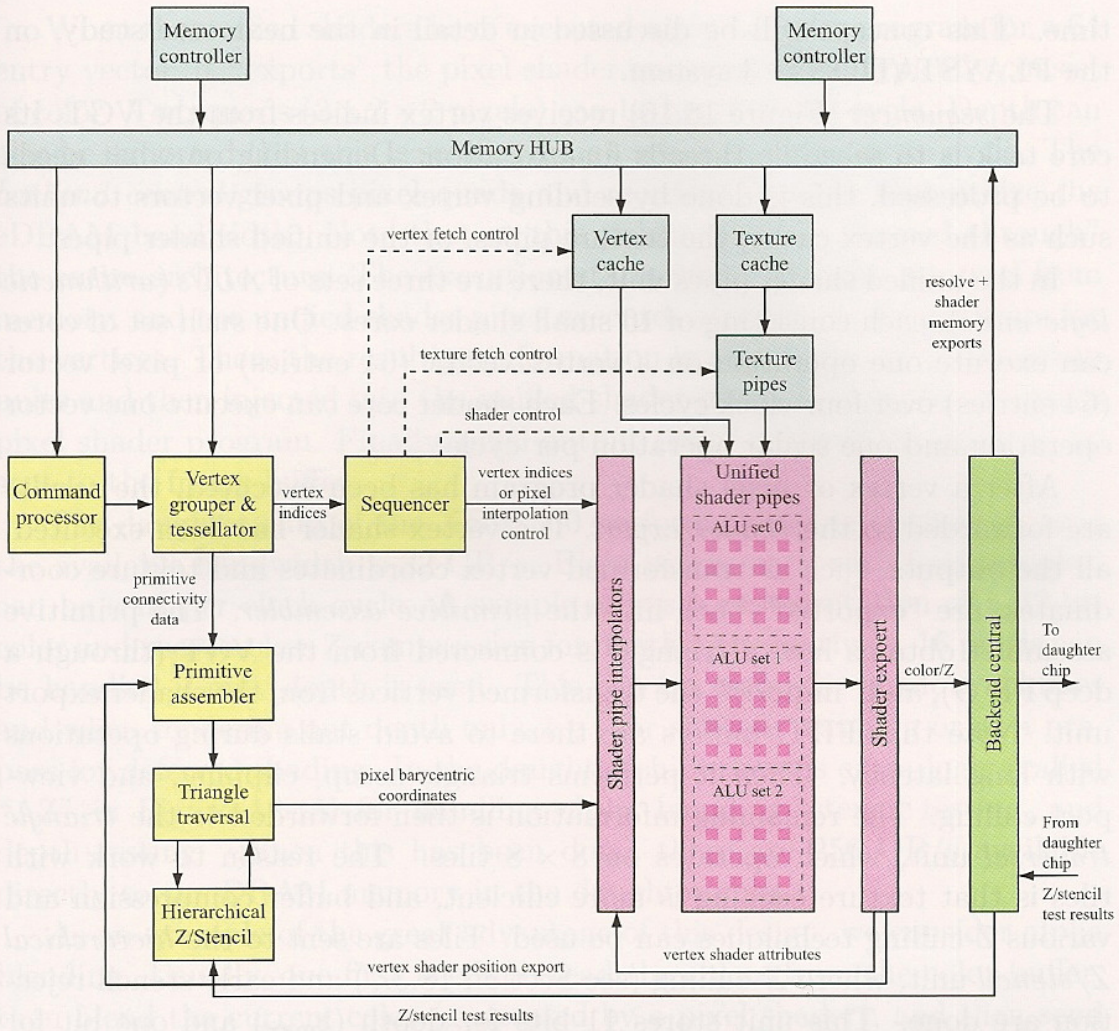


Figure 18.16. Block diagram of the Xbox 360 graphics processor.

Unified shader
Command processor:
reads commands from memory

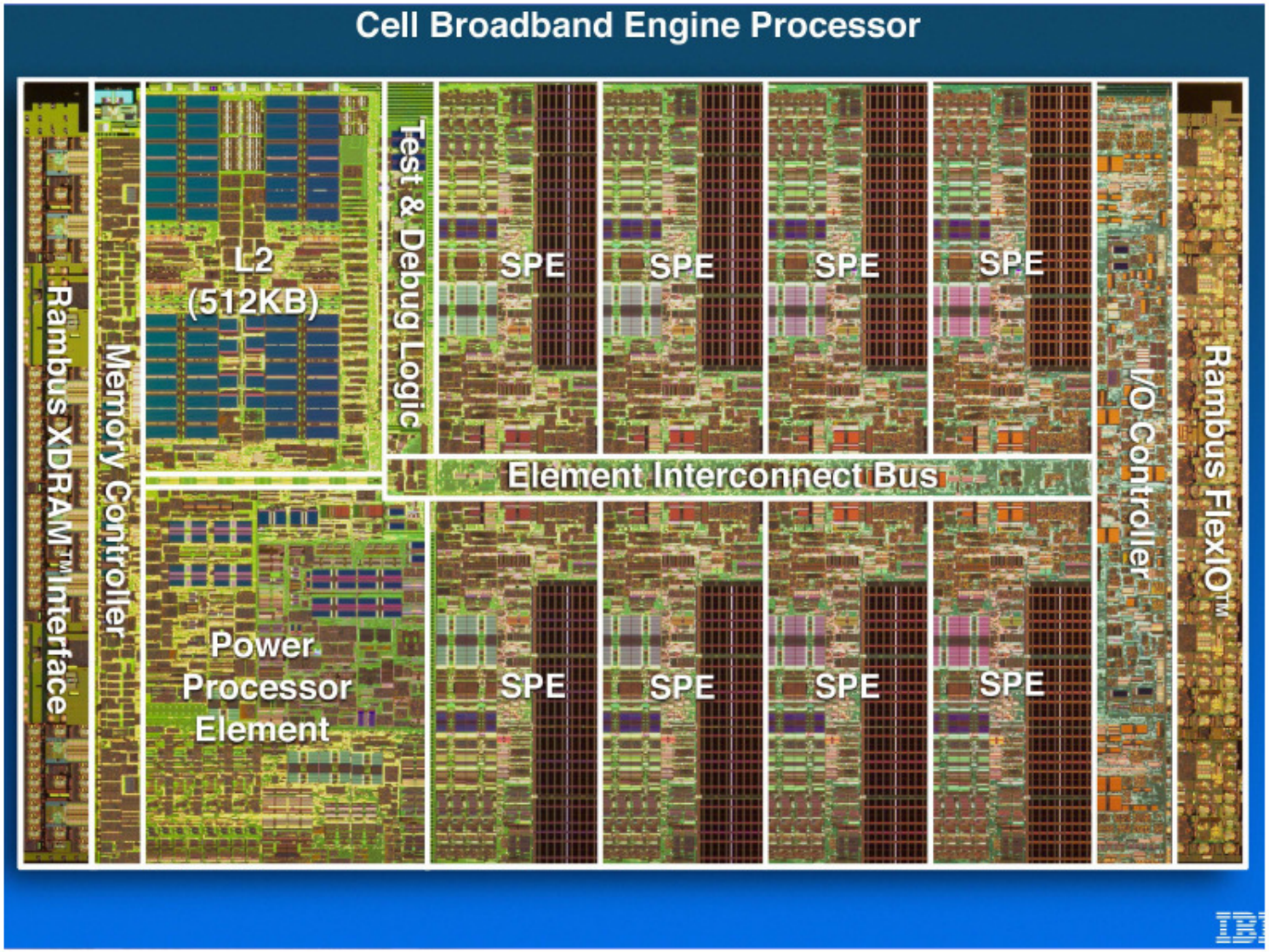
64 vertices or pixels are operated together (SIMD)

32 vertex threads or 64 pixel threads can be active

24,576 registers

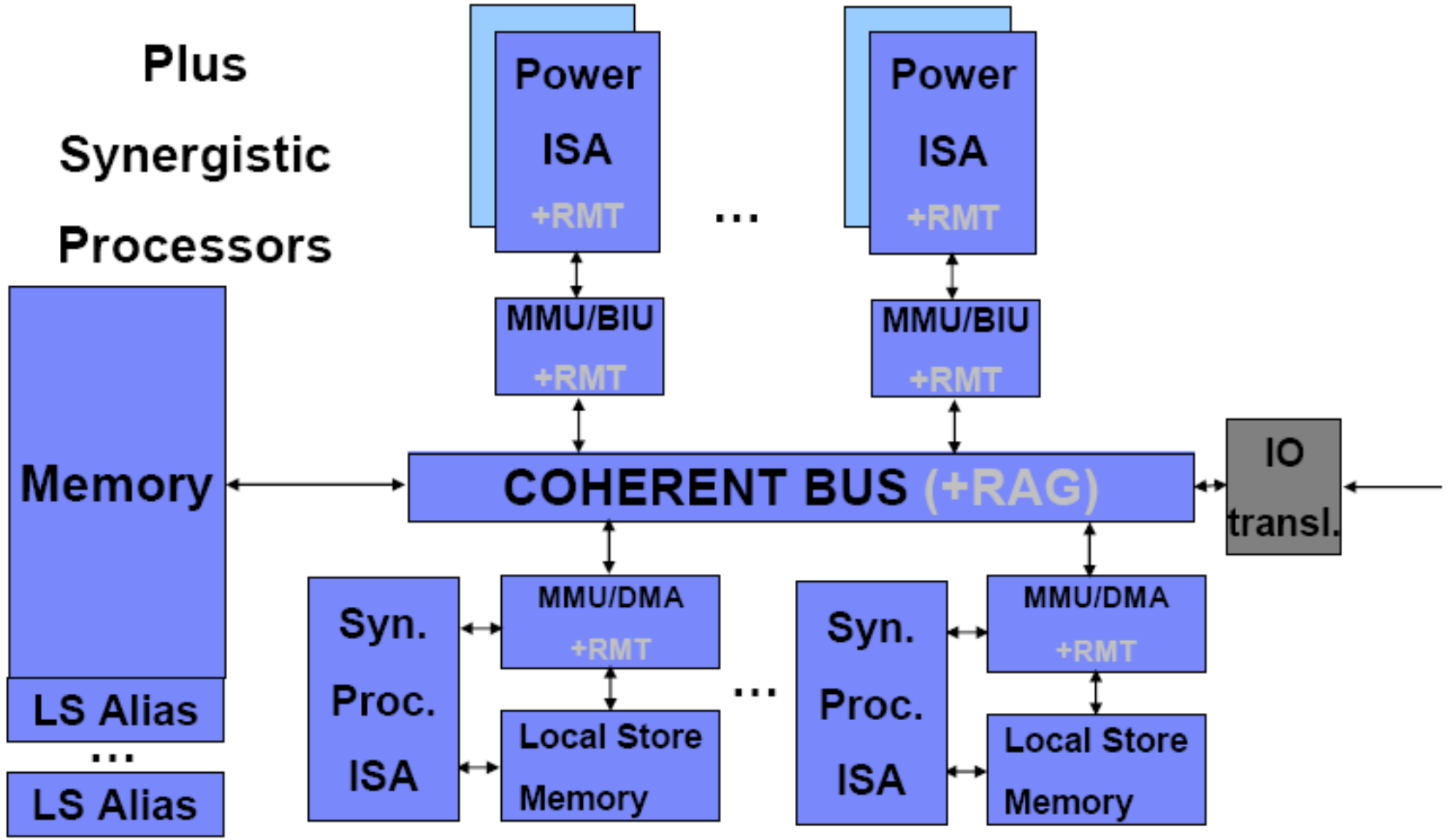
ALU:16 small shader cores
32Kb texture cache
Texture pipes: 16 bilinear filters per cycles

Cell

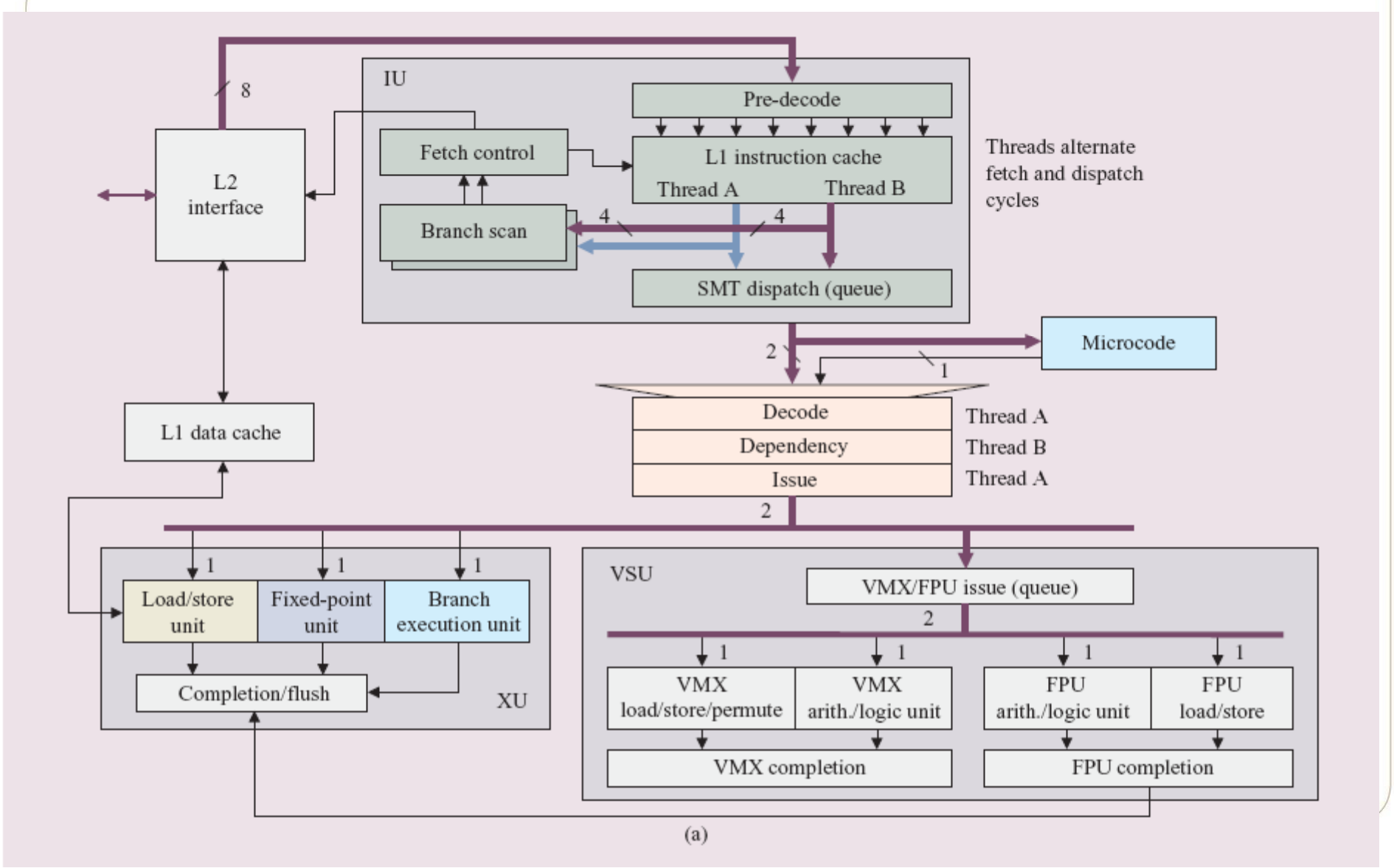

Cell Broadband Engine Processor
Rambus XDRAM™ Interface
Memory Controller
L2
(512KB)
Test & Debug Logic
SPE
SPE
SPE
SPE
Element Interconnect Bus
I/O Controller
Rambus FlexIO™
Power
Processor
Element
SPE
SPE
SPE
SPE



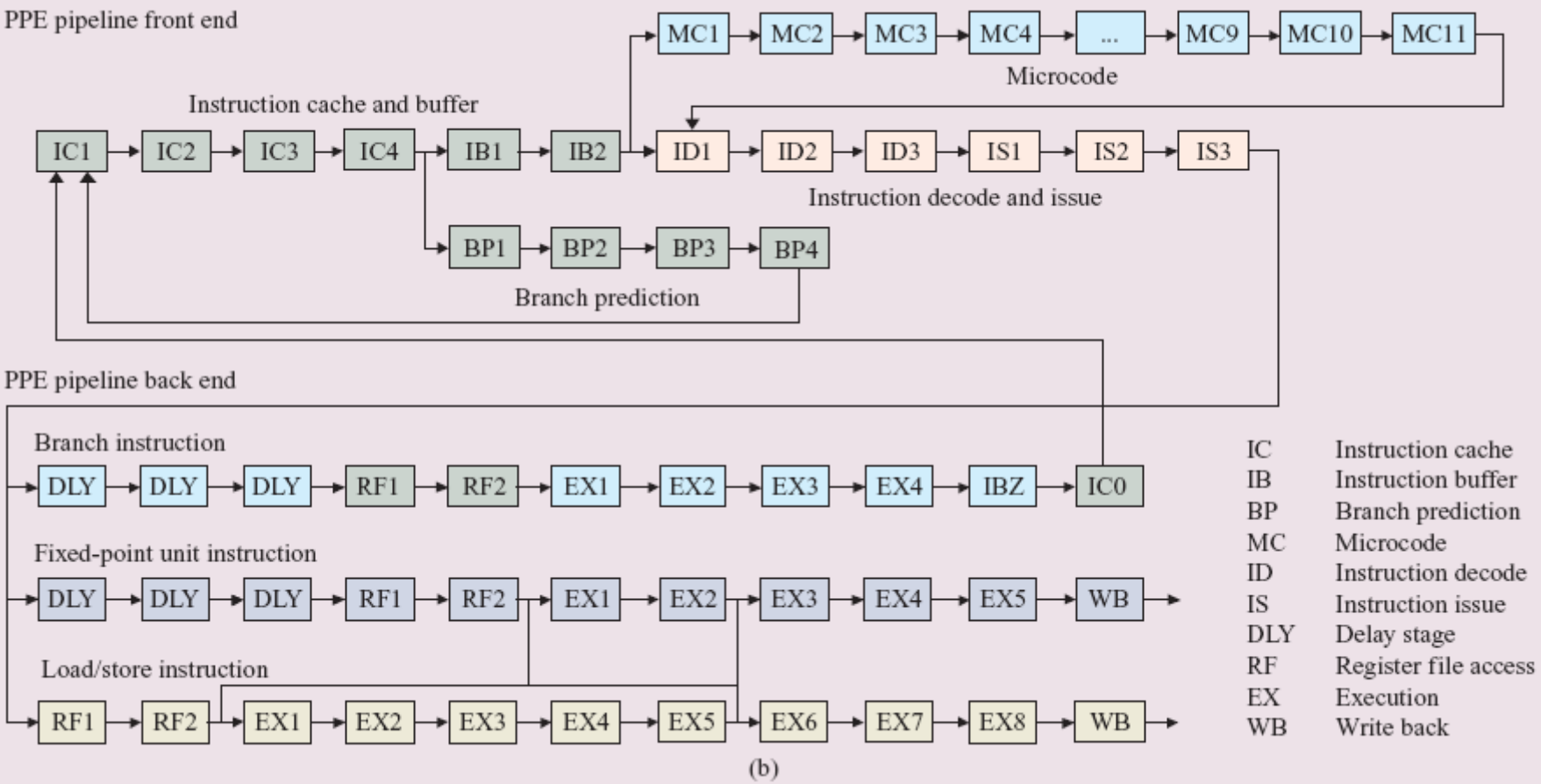
Cell Architecture is … 64b Power Architecture™+ MFC



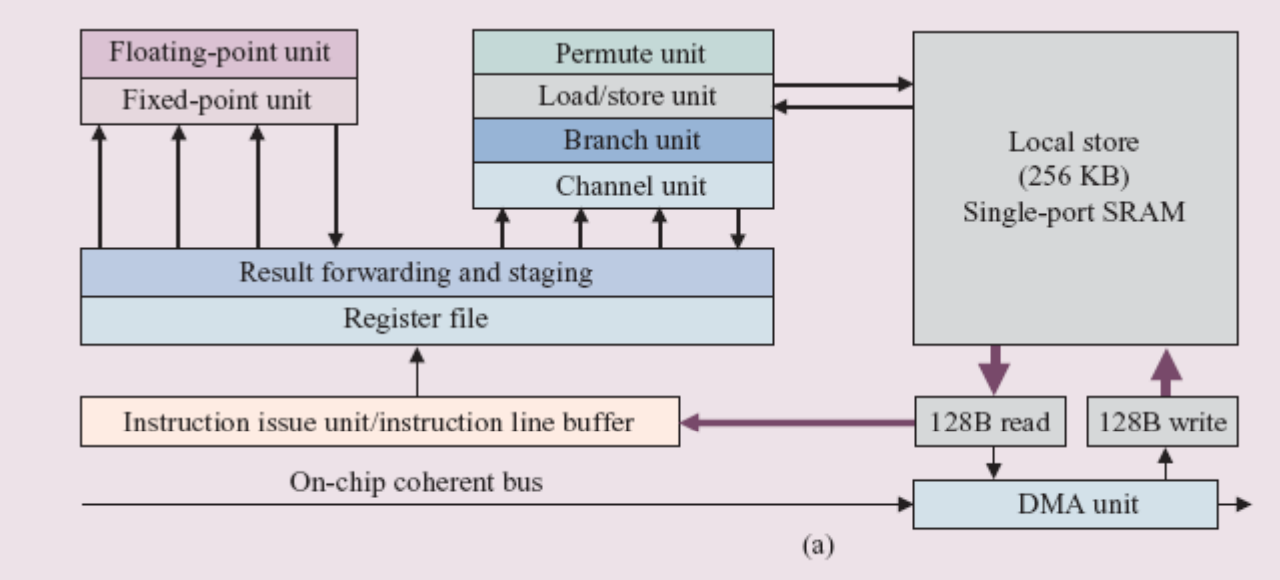
PPE Major Units

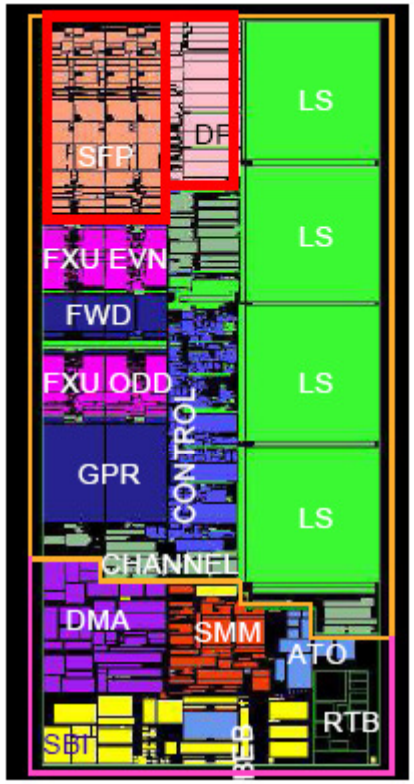


PPE Pipeline



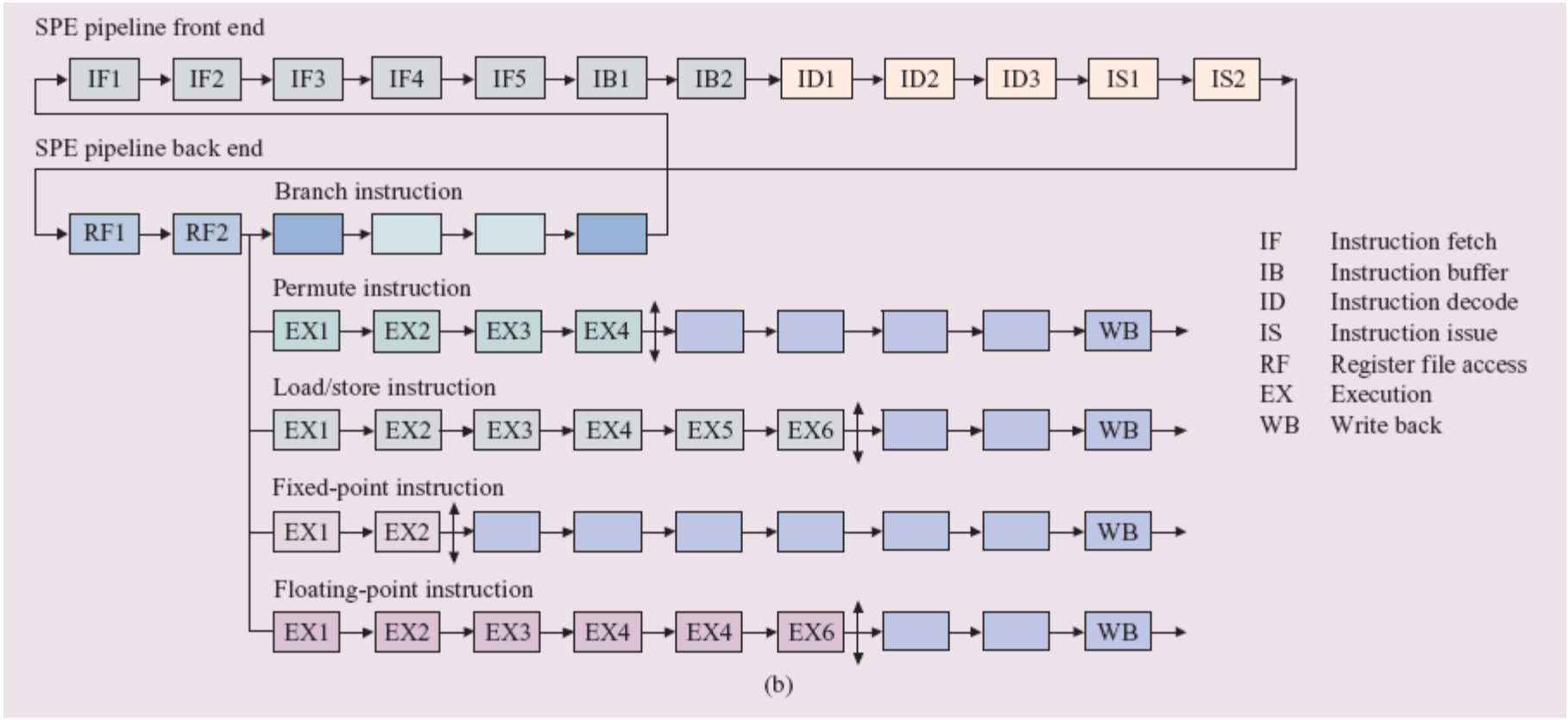
SPE Major Units





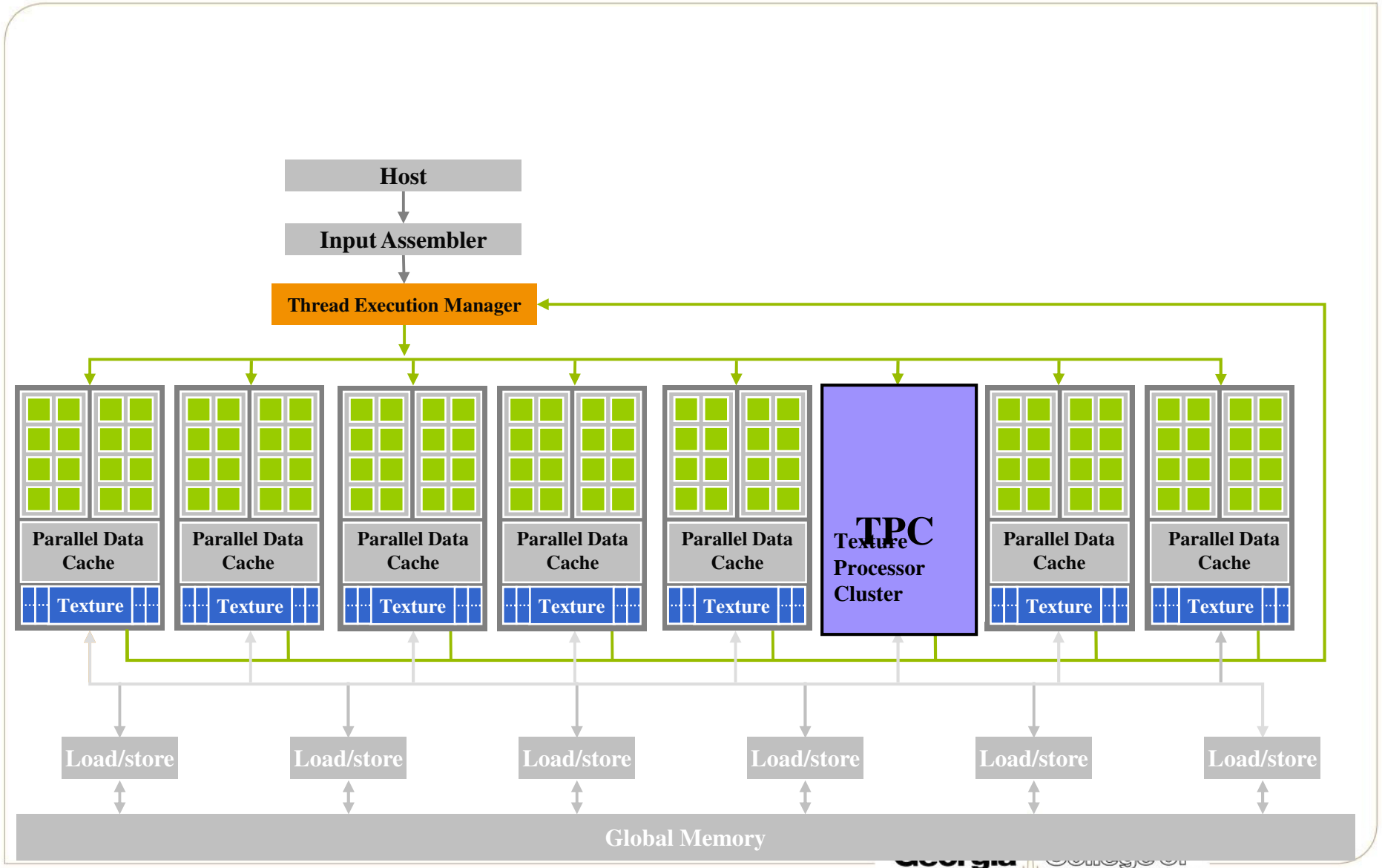
14.5mm^2 (90nm SOI)

SPE Pipeline

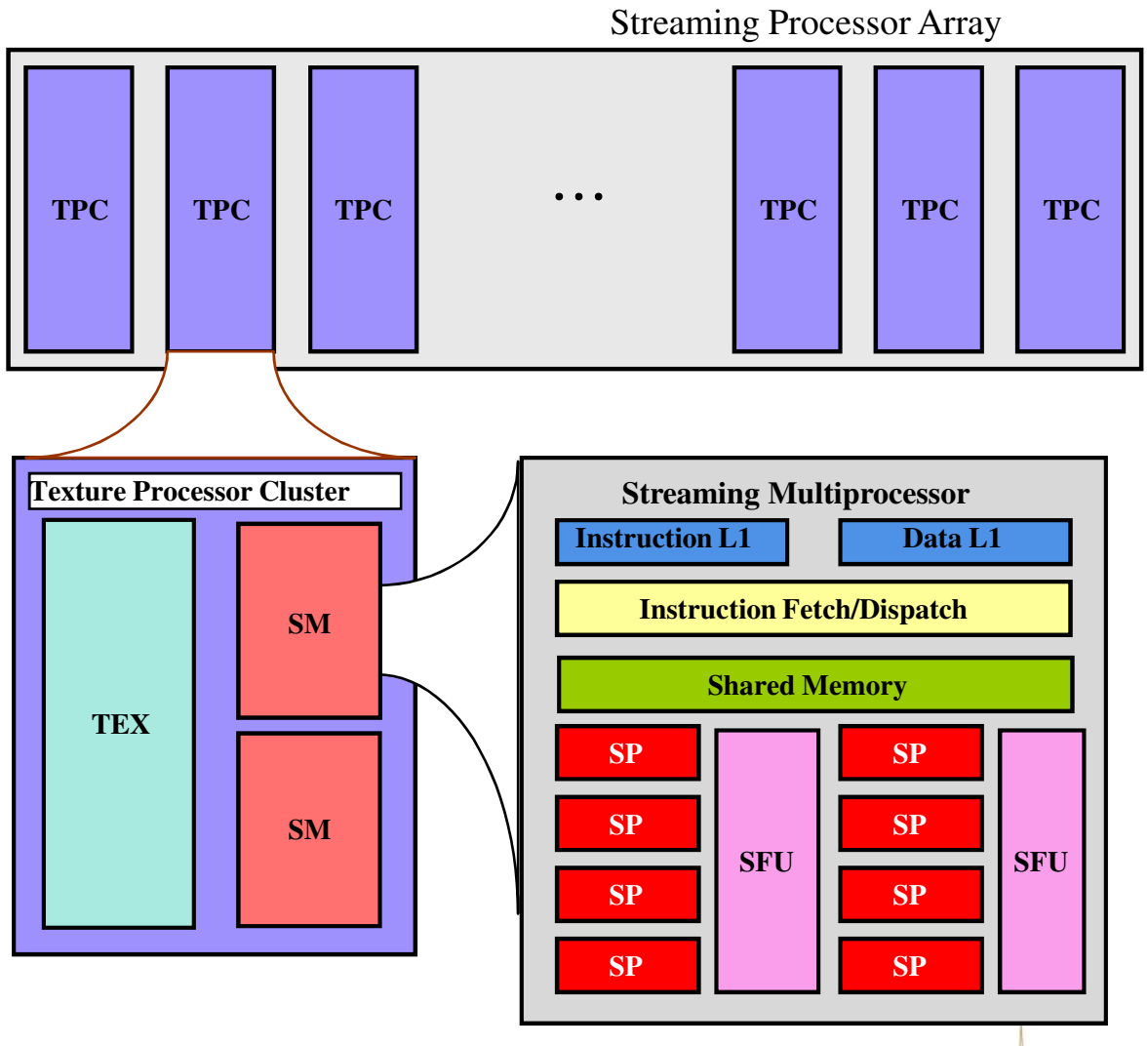


(b)

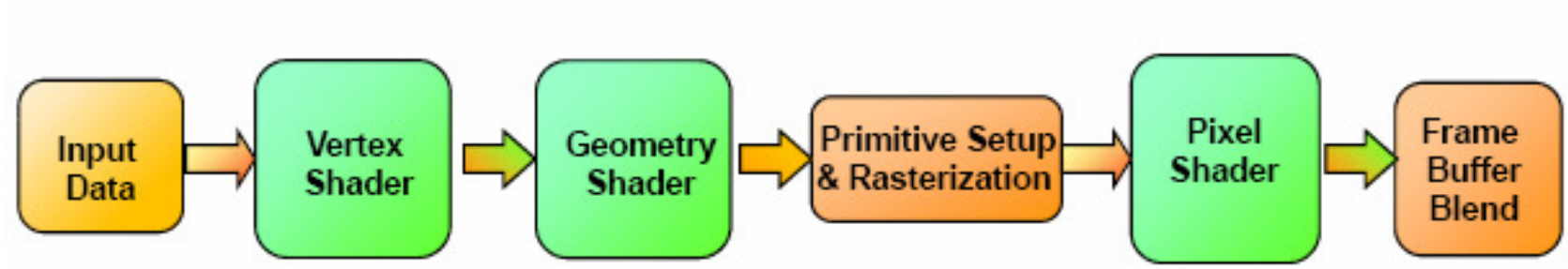
GeForce 8800 GTX



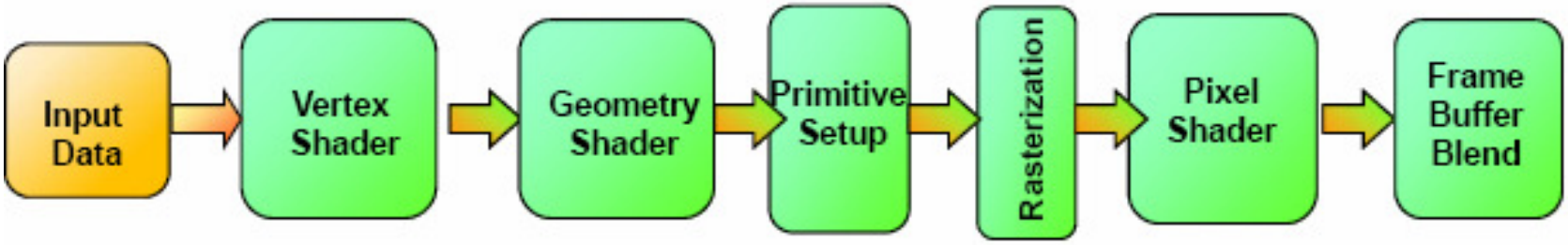
GeForce-8 Series HW Overview



Programmable Pipeline Comparison



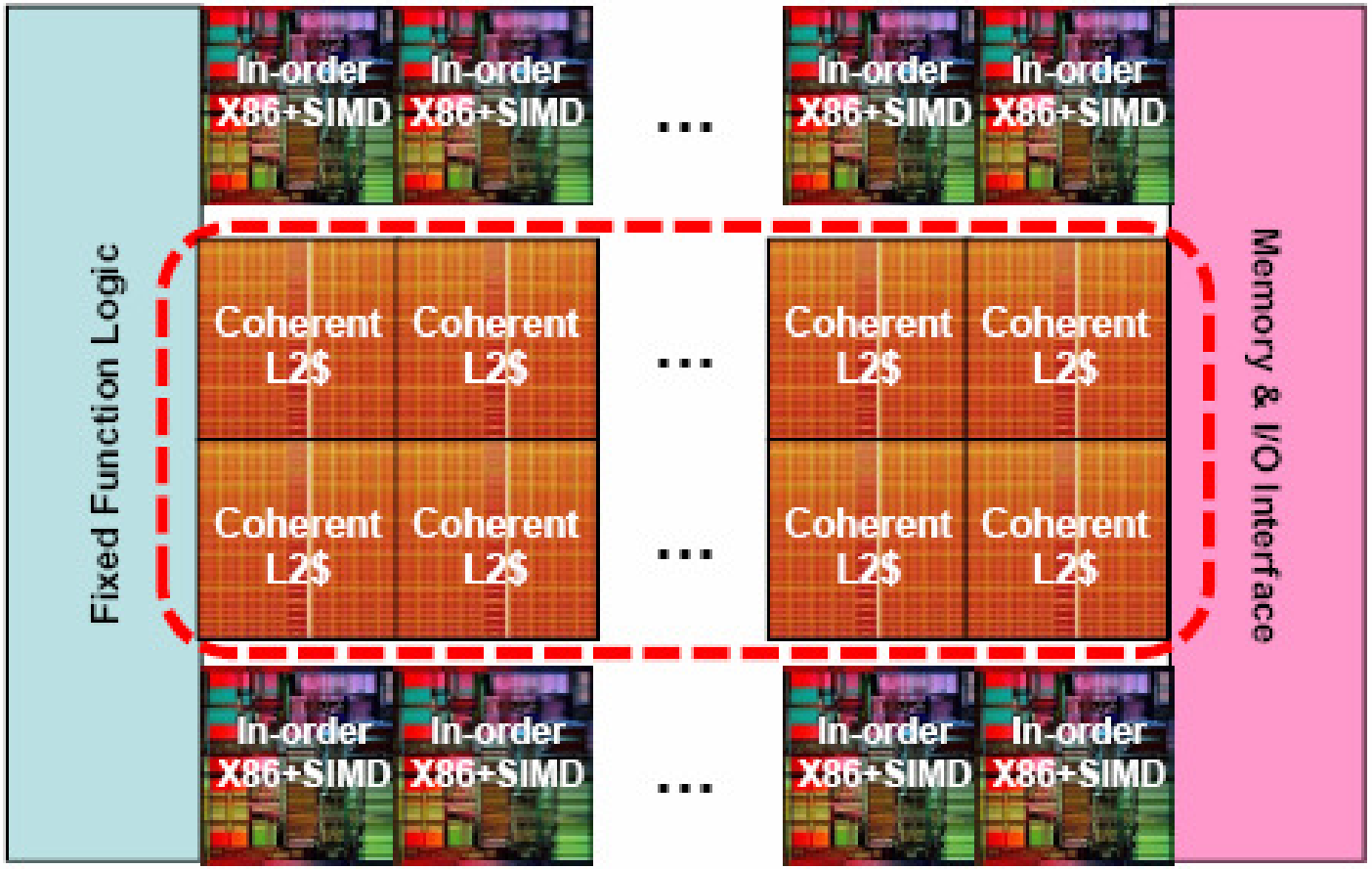
Conventional GPGPU pipeline (base on DirectX10)



Larrabee's fully programmable pipeline

http://users.ece.gatech.edu/~lanterma/mpg08/Larrabee_ECE4893.pdf

Larrabee Architecture

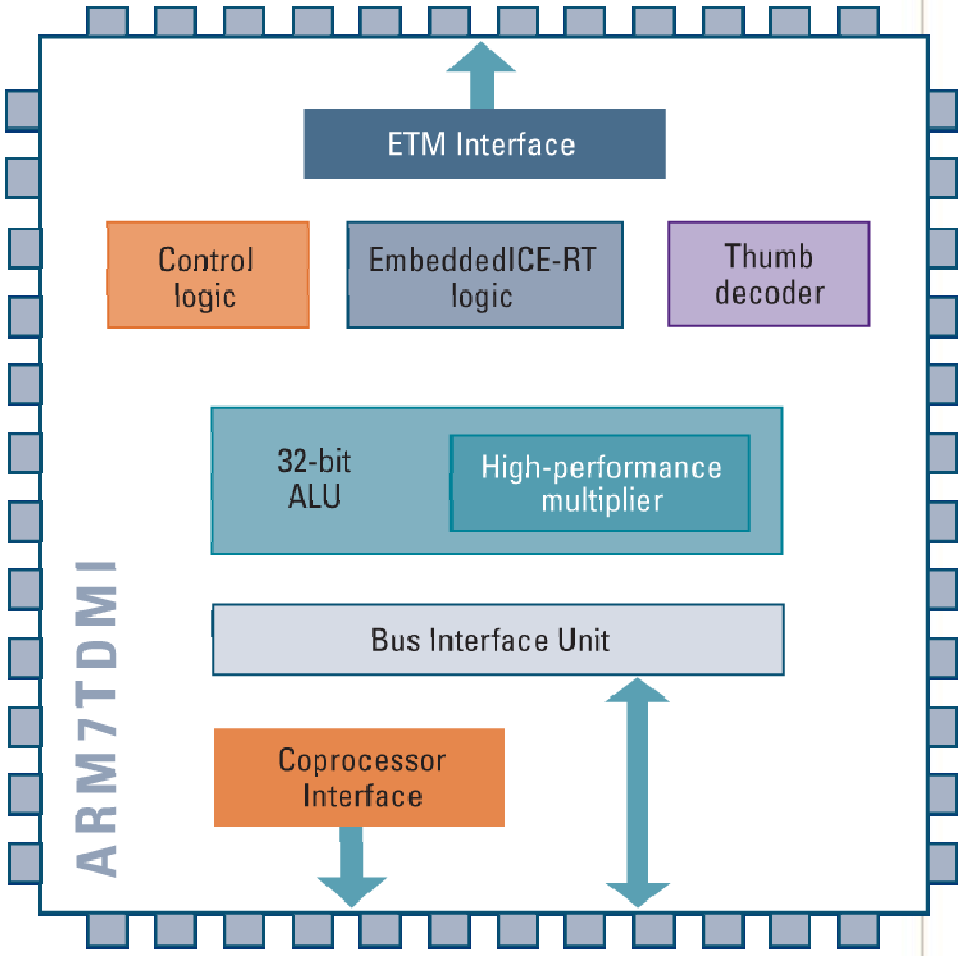


- 4-way SMT in order processor with cache coherence
- Extended X86 ISA
- Fixed functions: texture filtering



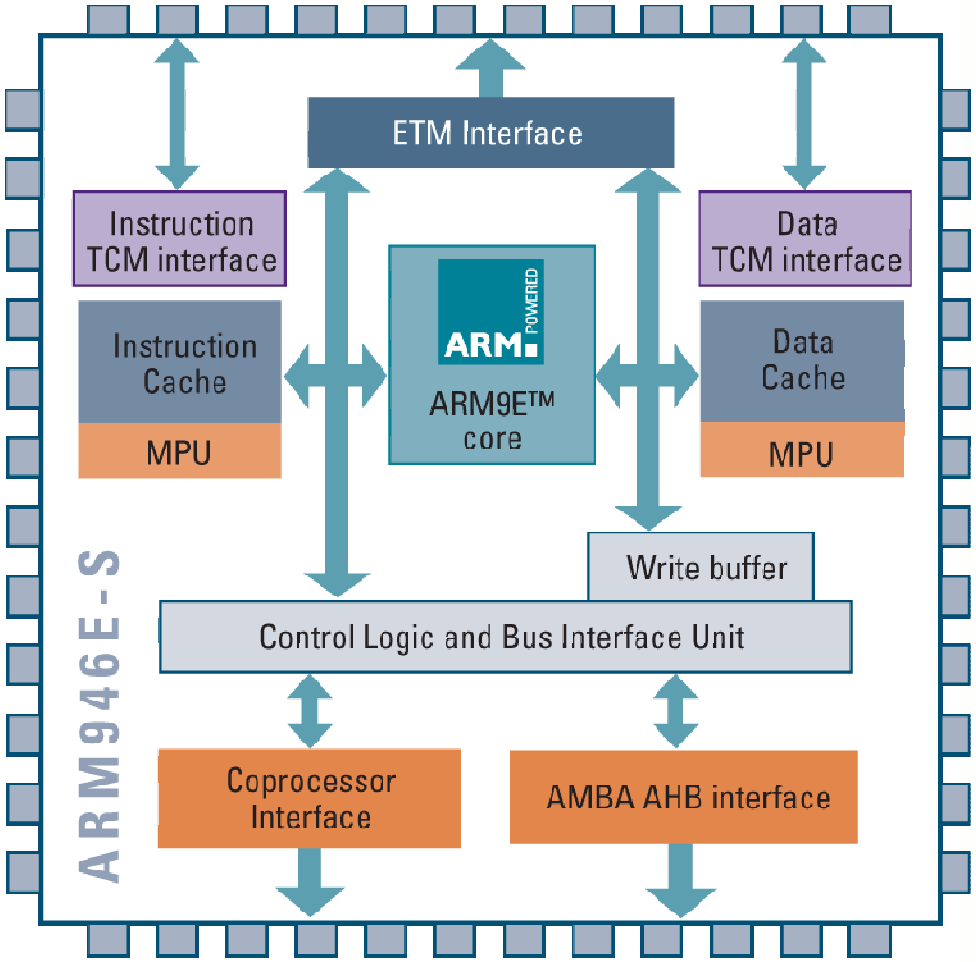
ARM7 TDMI

- 32/16-bit RISC
- 32-bit ARM instruction set
- 16-bit Thumb instruction set
- 3-stage pipeline
- Very small die size and low power
- Unified bus interface (32-bit data bus carries both instruction, data)



ARM9

- 5-stage pipeline
- I-cache and D-cache
- Floating point support with the optional VFP9-S coprocessor
- Enhanced 16 x 32-bit multiplier capable of single cycle MAC operations
- The ARM946E-S processor supports ARM's real-time trace technology





Conclusion

- Game architecture is a leading architecture
- Yesterday's scientific workload→ today's Game workload → tomorrow's daily user workload



Final Report

- Guideline is posted.



Final Exam Guidline

- Buzzword, sample exam will be posted during the weekend.
- Office hours on Tuesday 3:30-5:30

- Performance evaluation
- Comprehensive
  - Review mid-term exams

- **Final exam starts at 8:30!!**