

Roadrunner

By

Diana Lleva

Julissa Campos

Justina Tandar

Overview

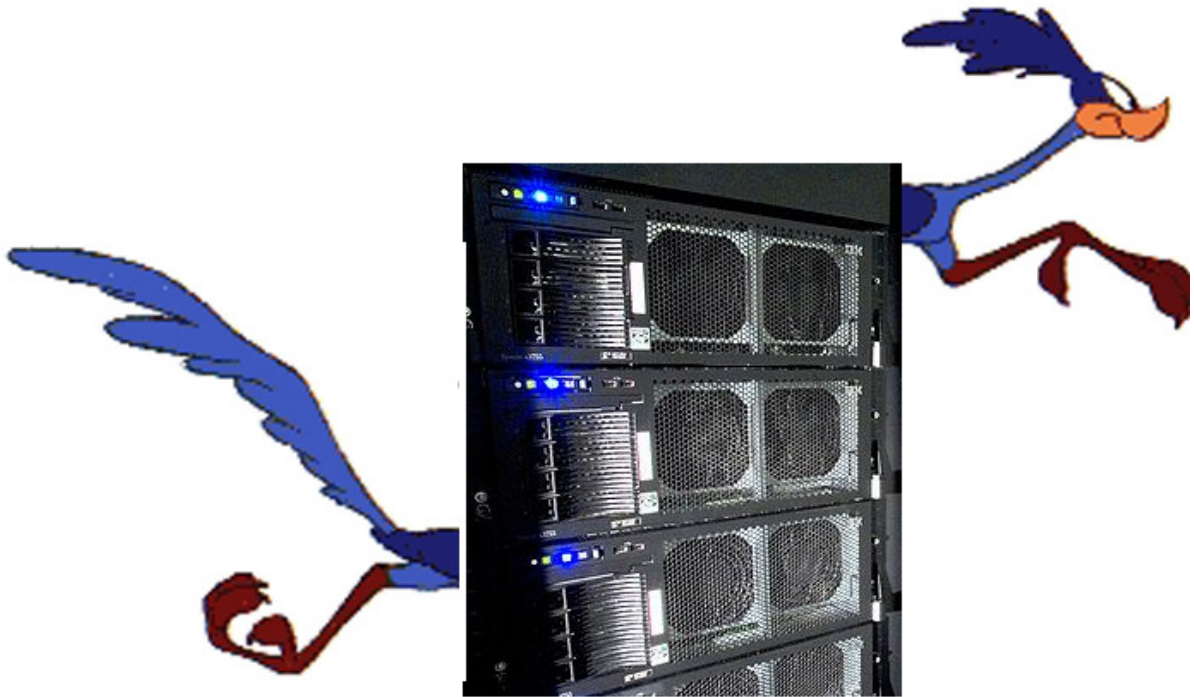
- Roadrunner background
- On-Chip Interconnect
- Number of Cores
- Memory Hierarchy
- Pipeline Organization
- Multithreading Organization

Roadrunner Background

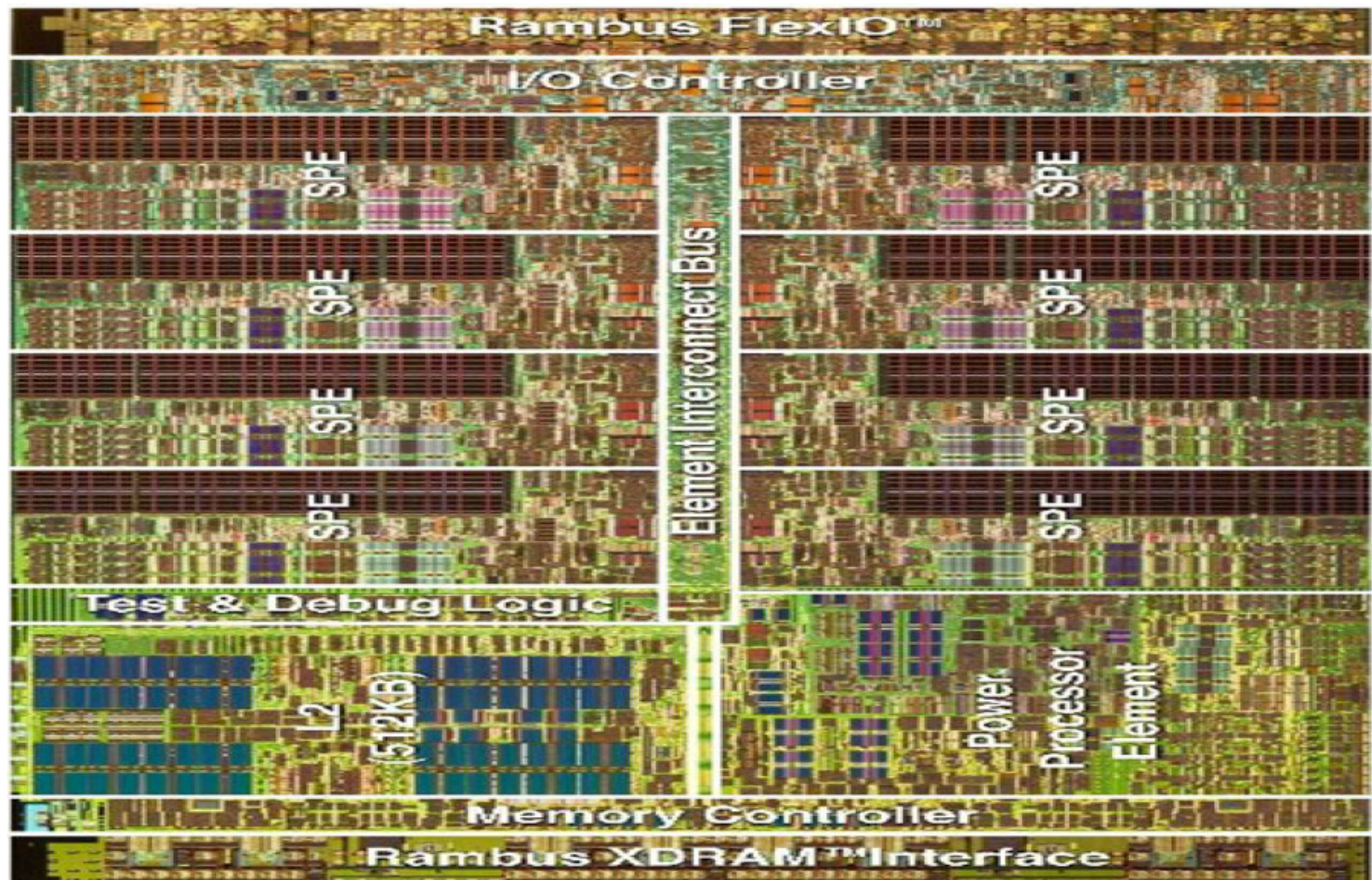
- Currently fastest supercomputer in the world
- Completed in 2009
- 3 Phase Project:
 - 1 – Base System – 76 Teraflops
 - 2 – Advanced algorithms – evaluate cell
 - 3 – Roadrunner – 1.3 Petaflops, cell clustered, Opteron Cluster

Roadrunner at a Glance

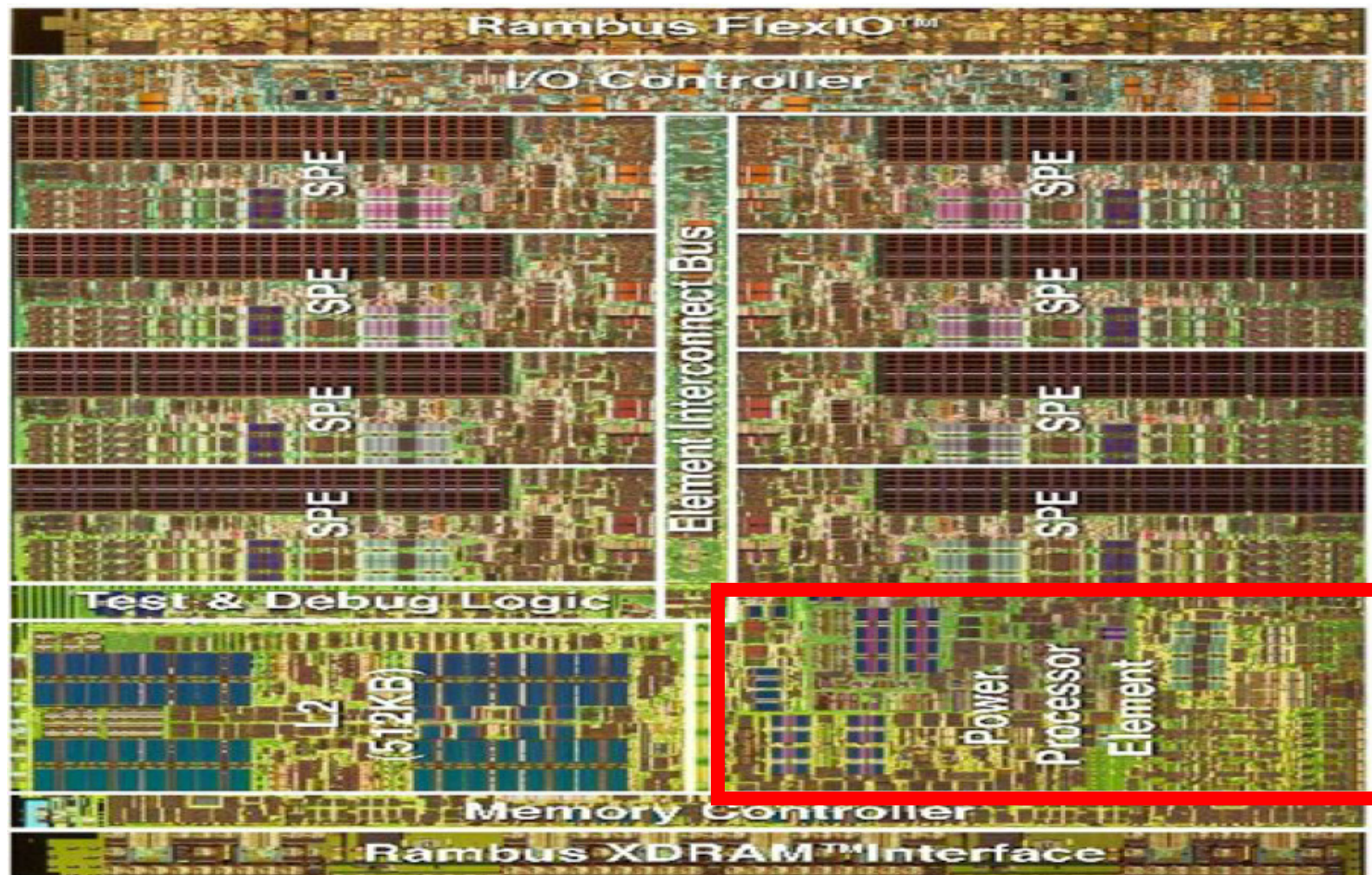
- Structure:
 - 6,912 AMD Opteron 2210 Processors
 - 12,960 IBM PowerXCell 8i cells
- Speed: 1.71 petaflops (peak)
- Rank: 1 (Top500) as of June 2008
- Cost: \$133 M
- Memory: 103.6 TiB
- Power: 2.35 MW



On-Chip Interconnect



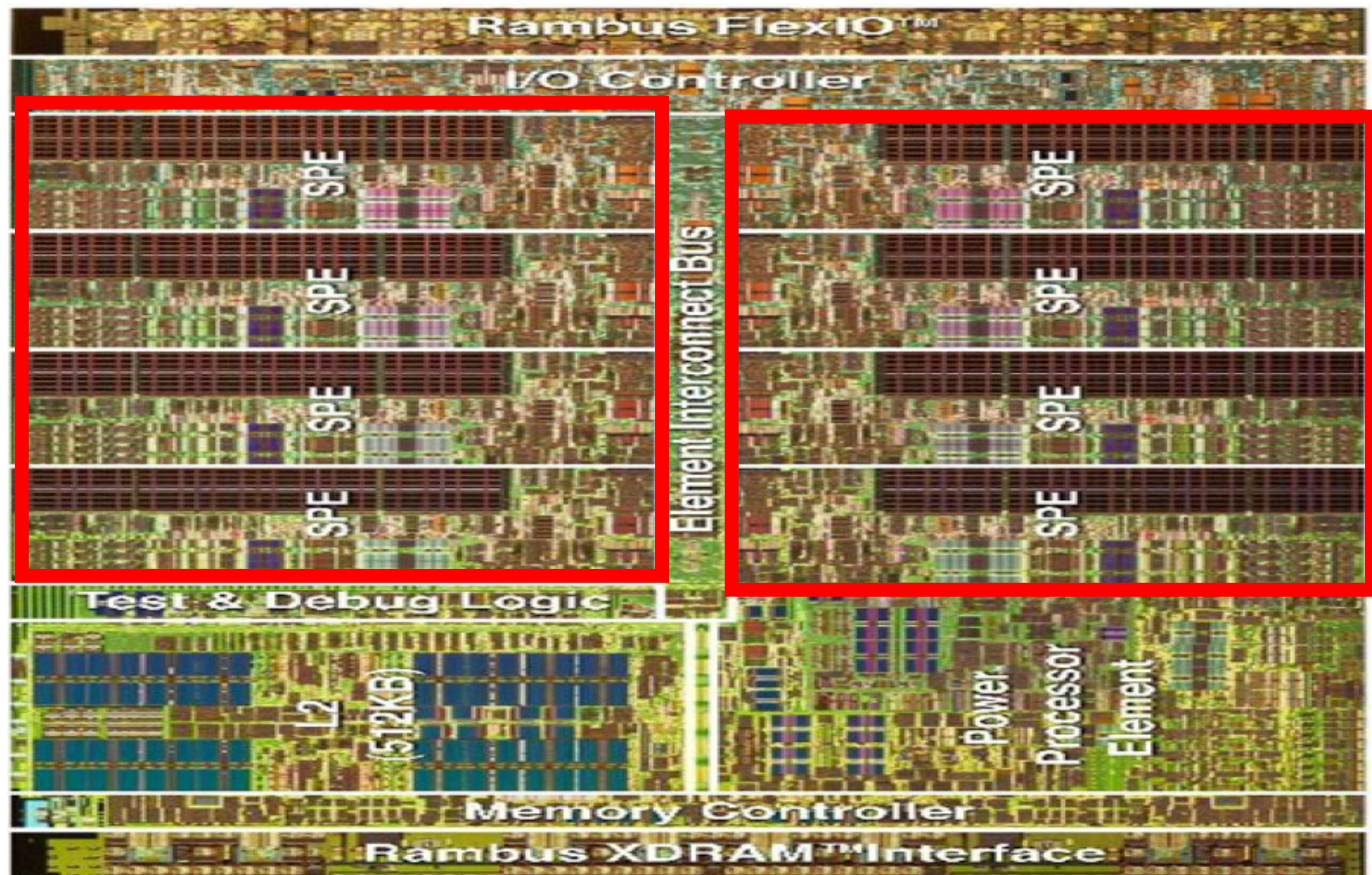
On-Chip Interconnect



Power Processor Element (PPE)

- 1 PPE/chip
- 1 IBM's VMX unit
- 32k L1 Cache
- 512k L2 Cache
- 2 way simultaneous multithreading

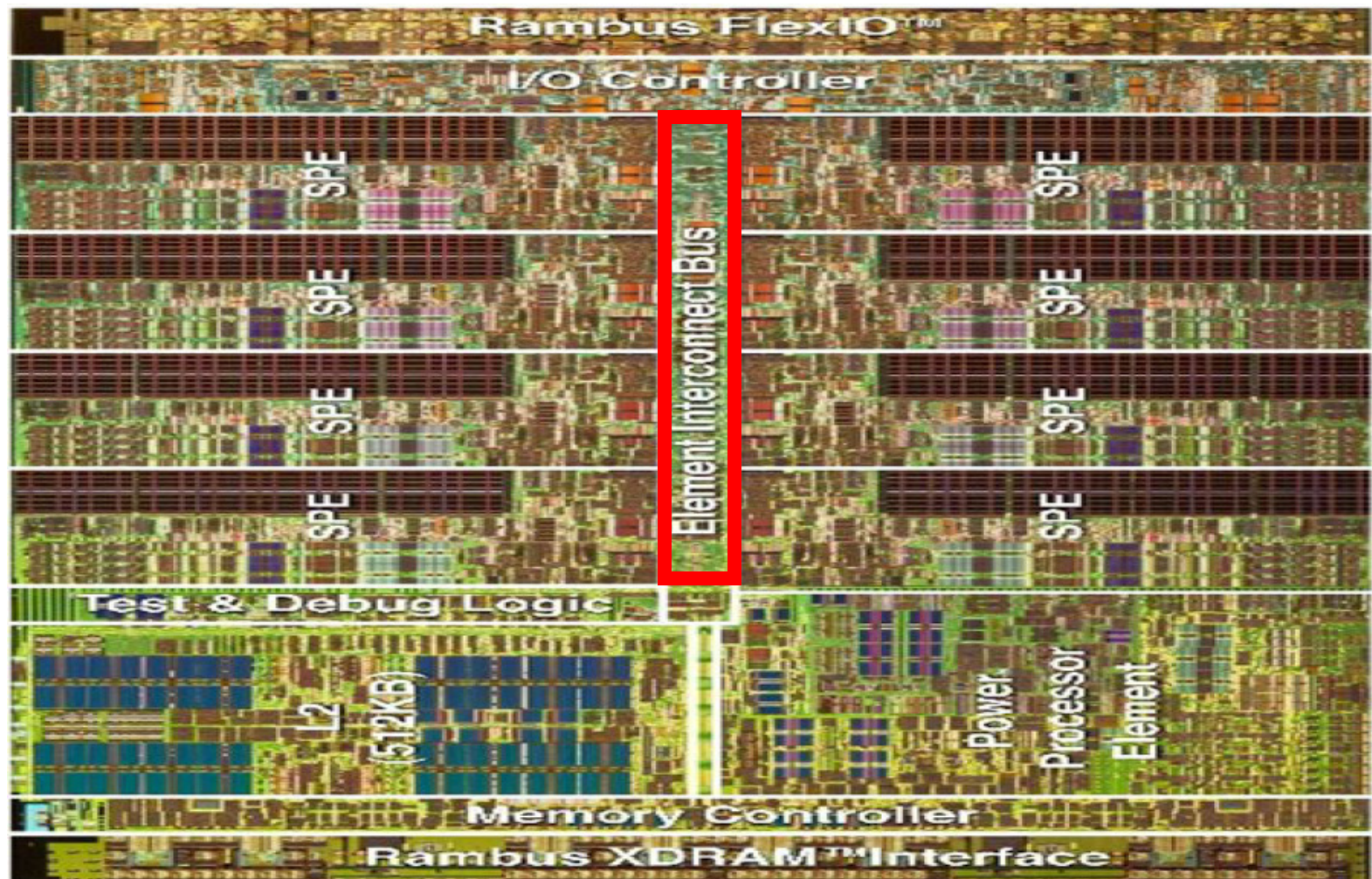
On-Chip Interconnect



Synergistic Processing Elements (SPE)

- 8 SPE's/chip
- 128 bit SIMD ISA
- 128x128 bit Register File
- 256k Local Store
- 1 Memory Flow Controller
- Isolation Mode (Security)

On-Chip Interconnect

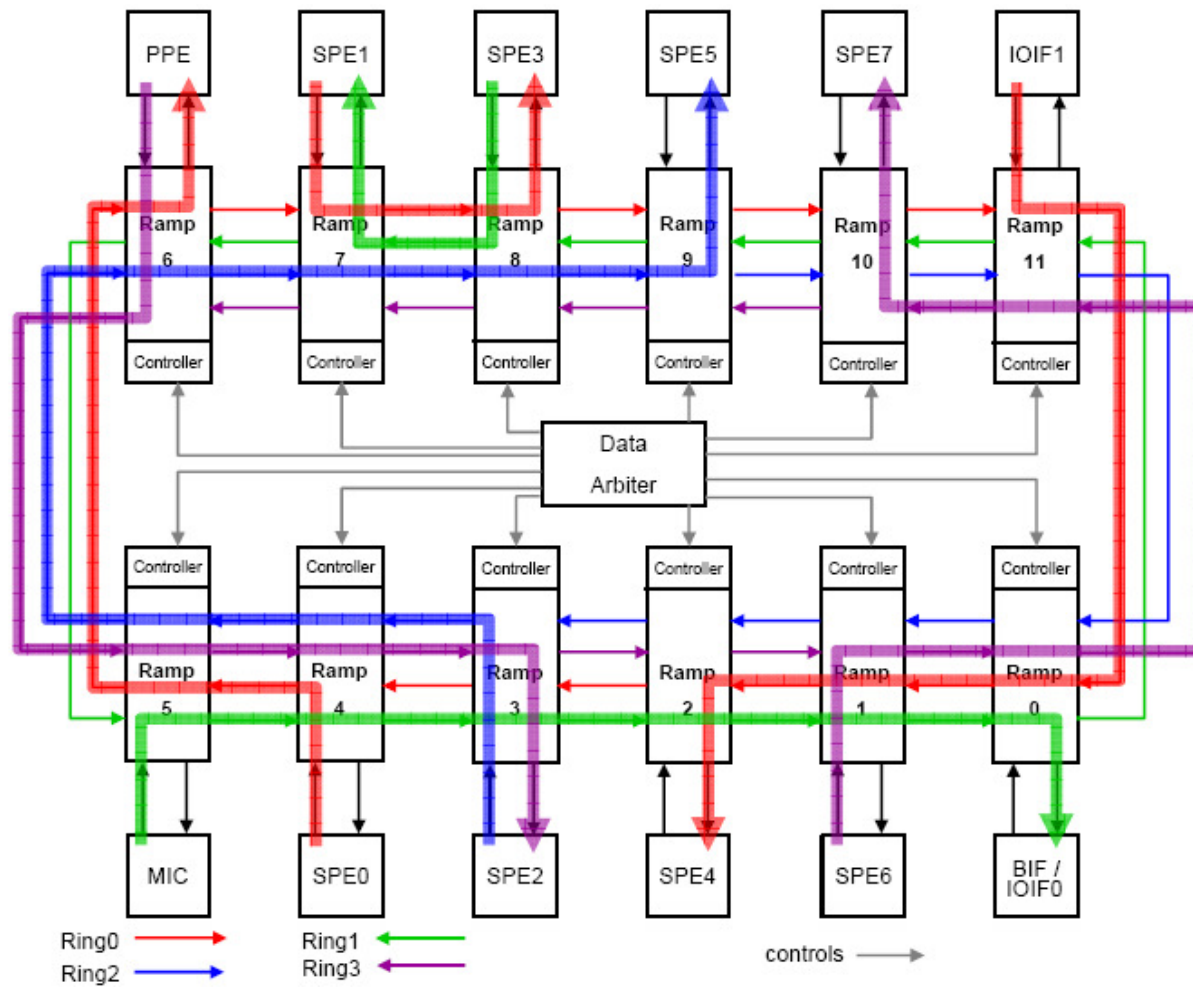


Element Interconnect Bus (EIB)

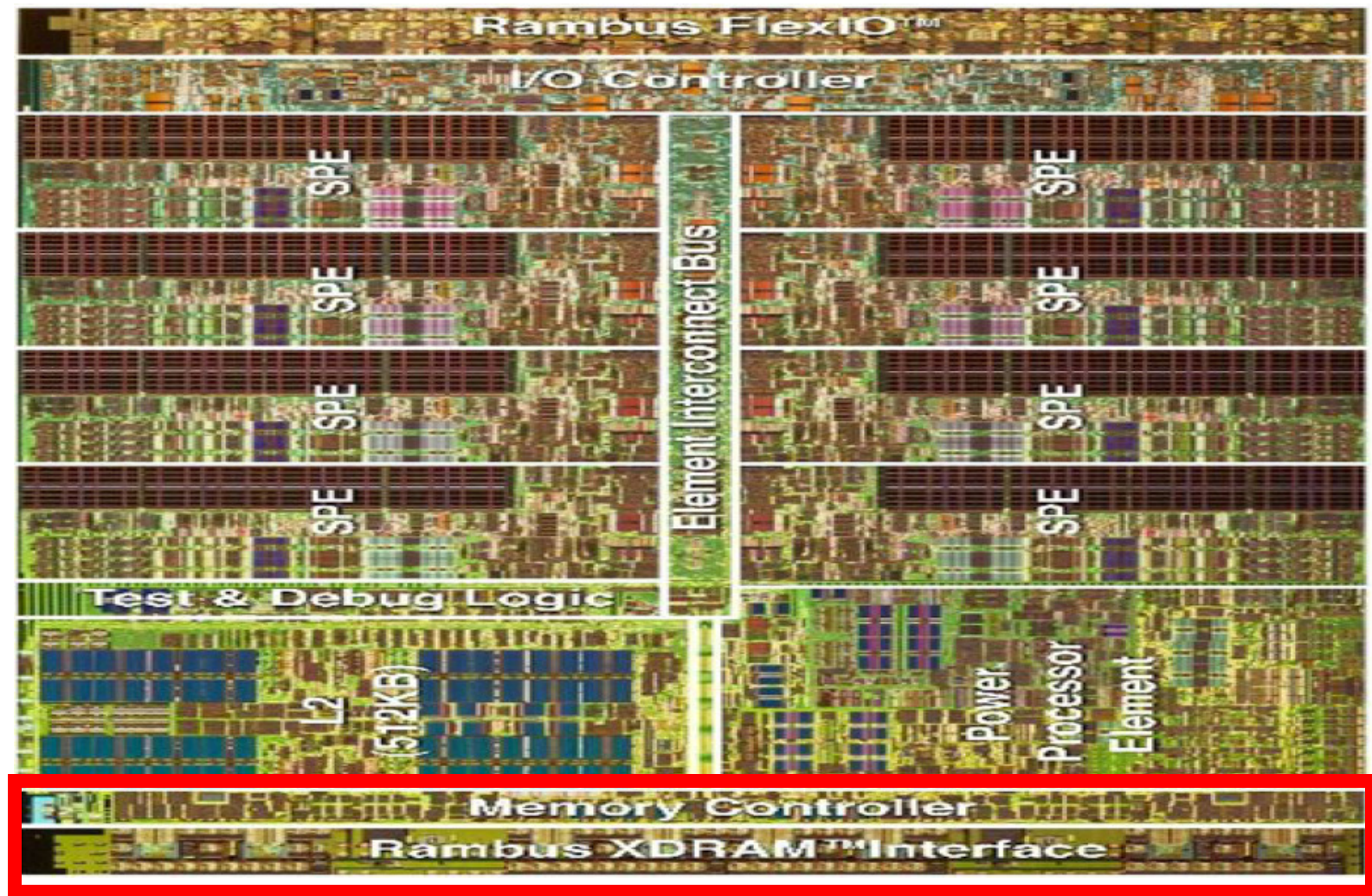
- 1 bus/chip
- Bandwidth: 96 B/cycle
 - Four 16 B data rings for internal communication
- Data port supports 25.6 GB/sec in each direction
- Command bus supports:
 - 102.4 GB/sec – coherent commands
 - 307.2 GB/sec – between units

EIB

Example of EIB running 8 transactions concurrently



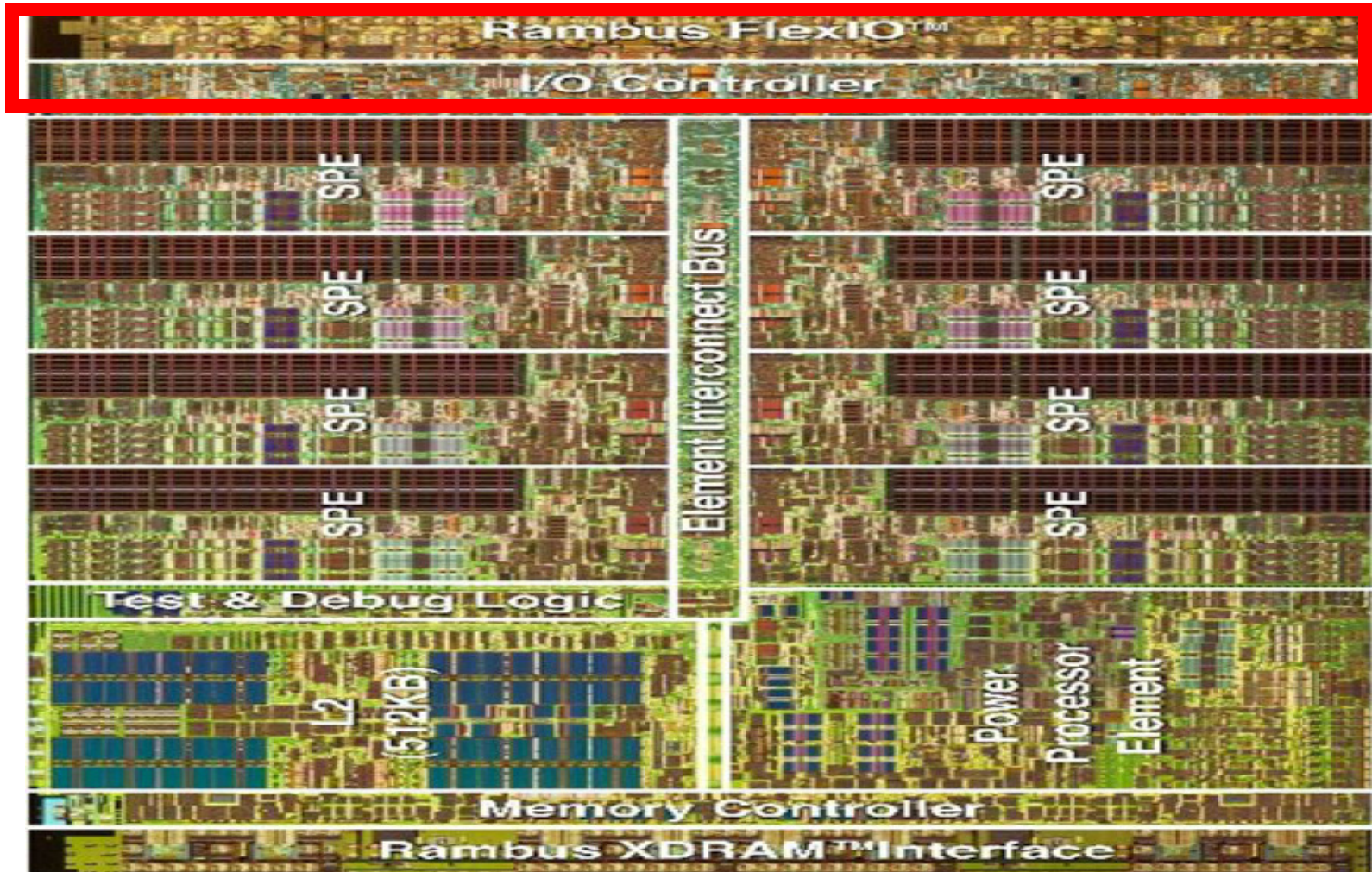
On-Chip Interconnect



System Memory Interface

- Memory Controller & Rambus XDRAM Interface
- Bandwidth: 16 B/cycle
- Speed: 25.6 GB/s

On-Chip Interconnect



Input/Output Interface

- I/O Controller & Rambus FlexIO
- Bandwidth: $2 * (16 \text{ B/cycle})$

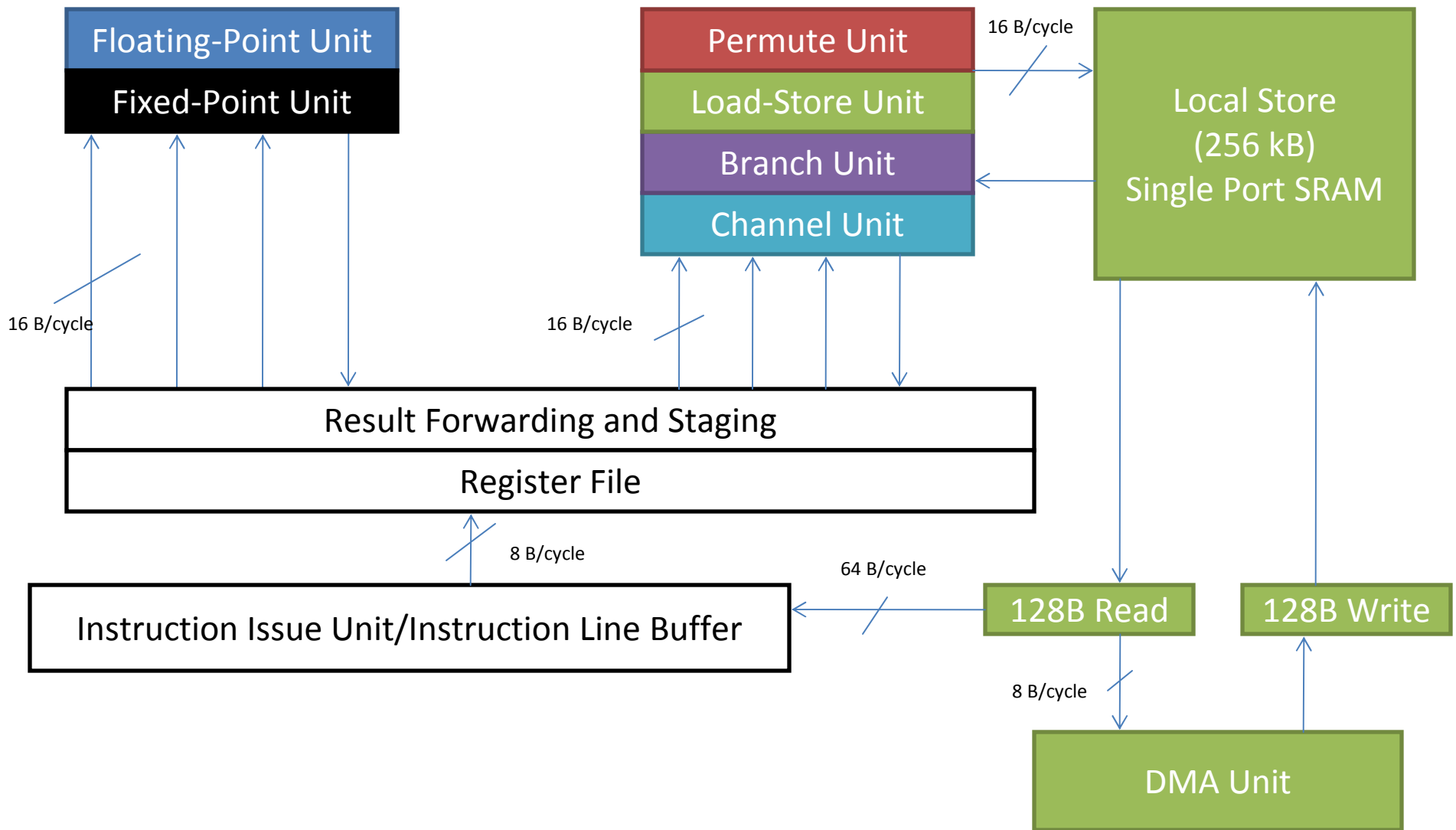
Number of Cores

- 6,912 dual-core AMD Opteron processors and 12,960 of IBM's Cell eDP accelerators
- Not including the cores in the operations and communication nodes:
 - 6,120 Opteron (2 cores) + 12,240 PowerXCell 8i (9 cores) = 122,400 cores

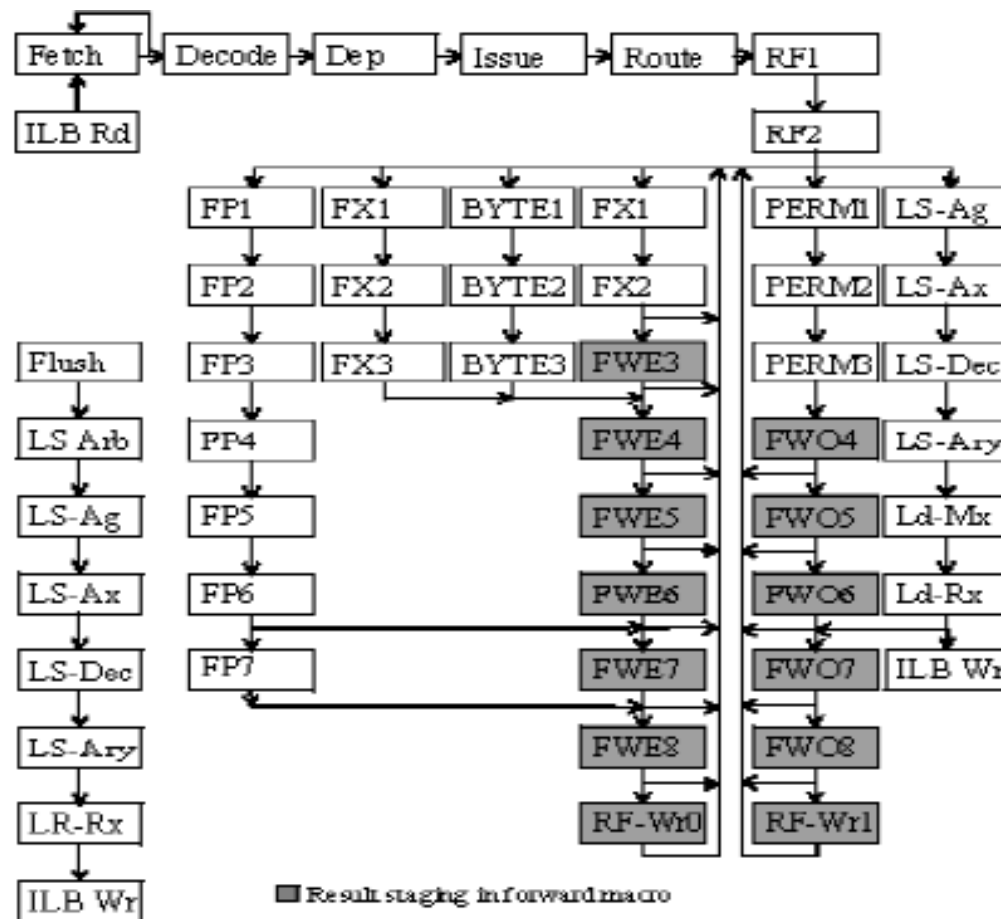
Pipeline Organization

- 6 Stages
- In-order Execution
- Dual Issue Pipeline
 - 2 Pipes
 - In case of instruction swap → Single issue
 - 9 Units per Pipeline

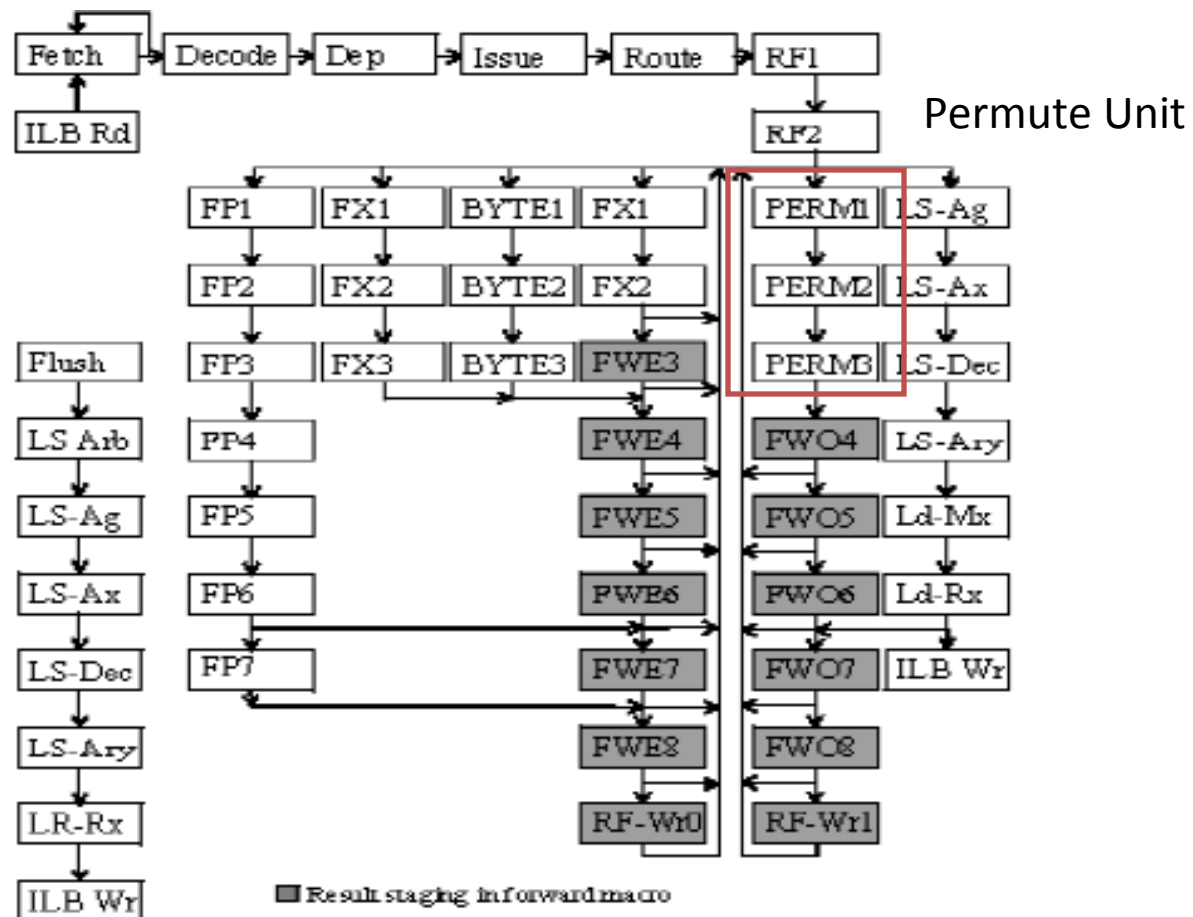
Pipeline Organization



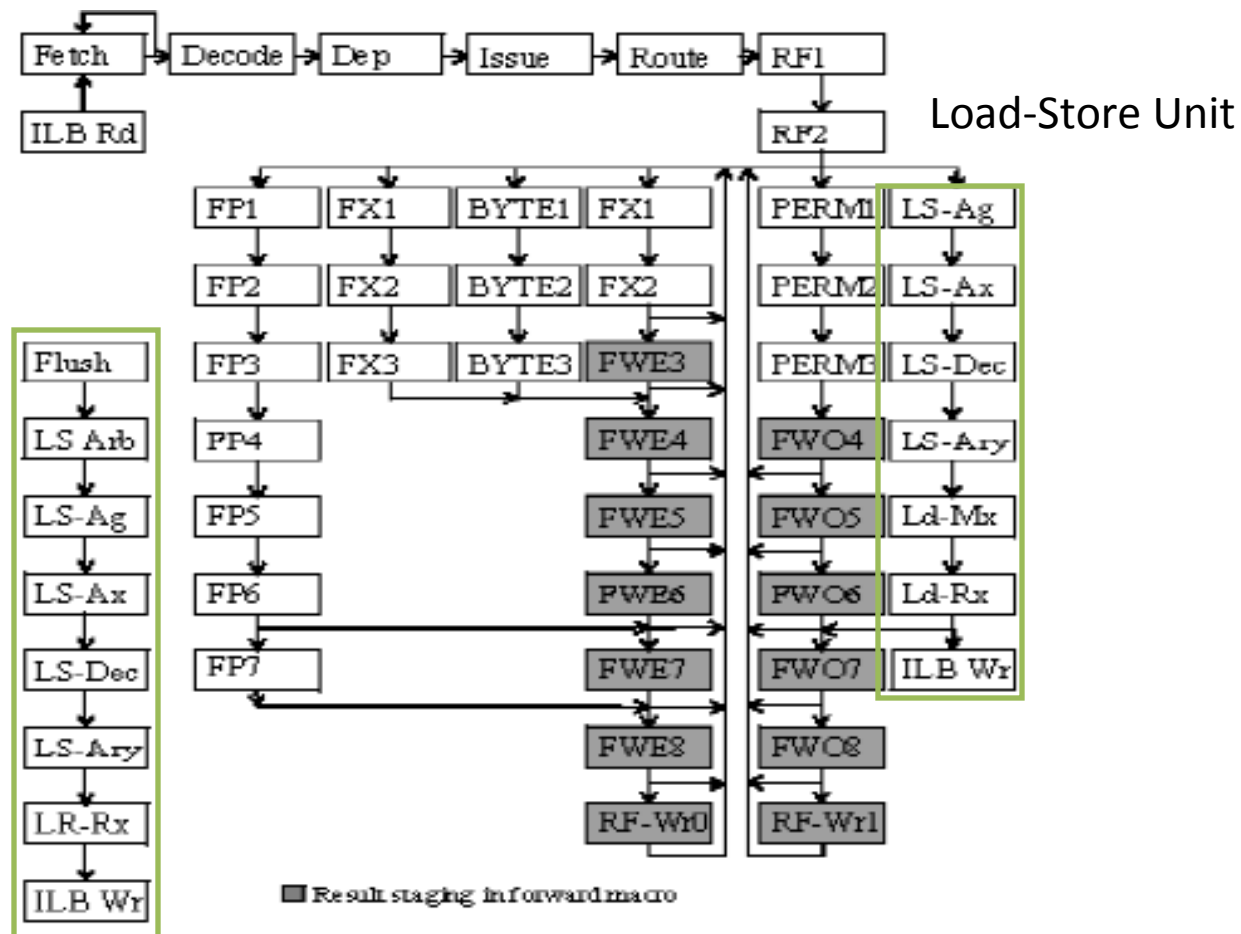
Pipeline Organization



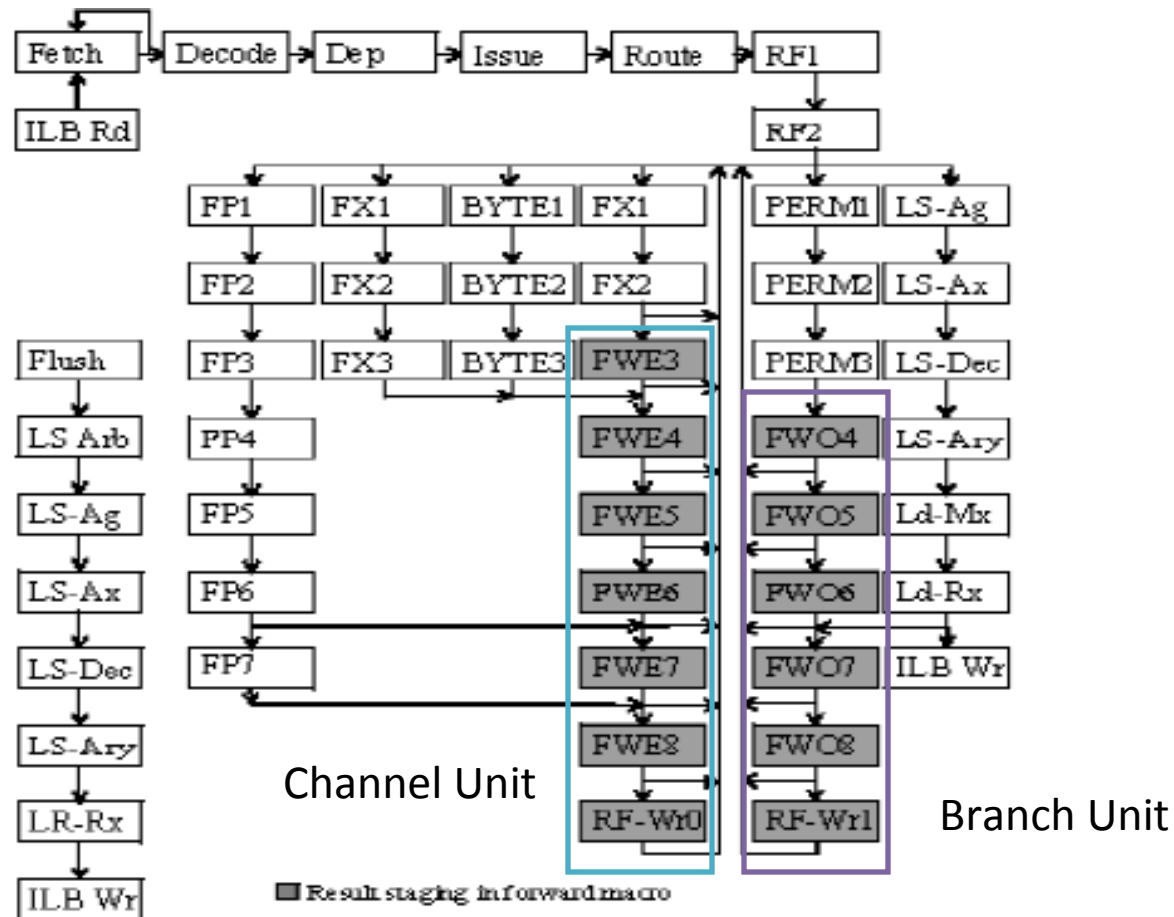
Pipeline Organization



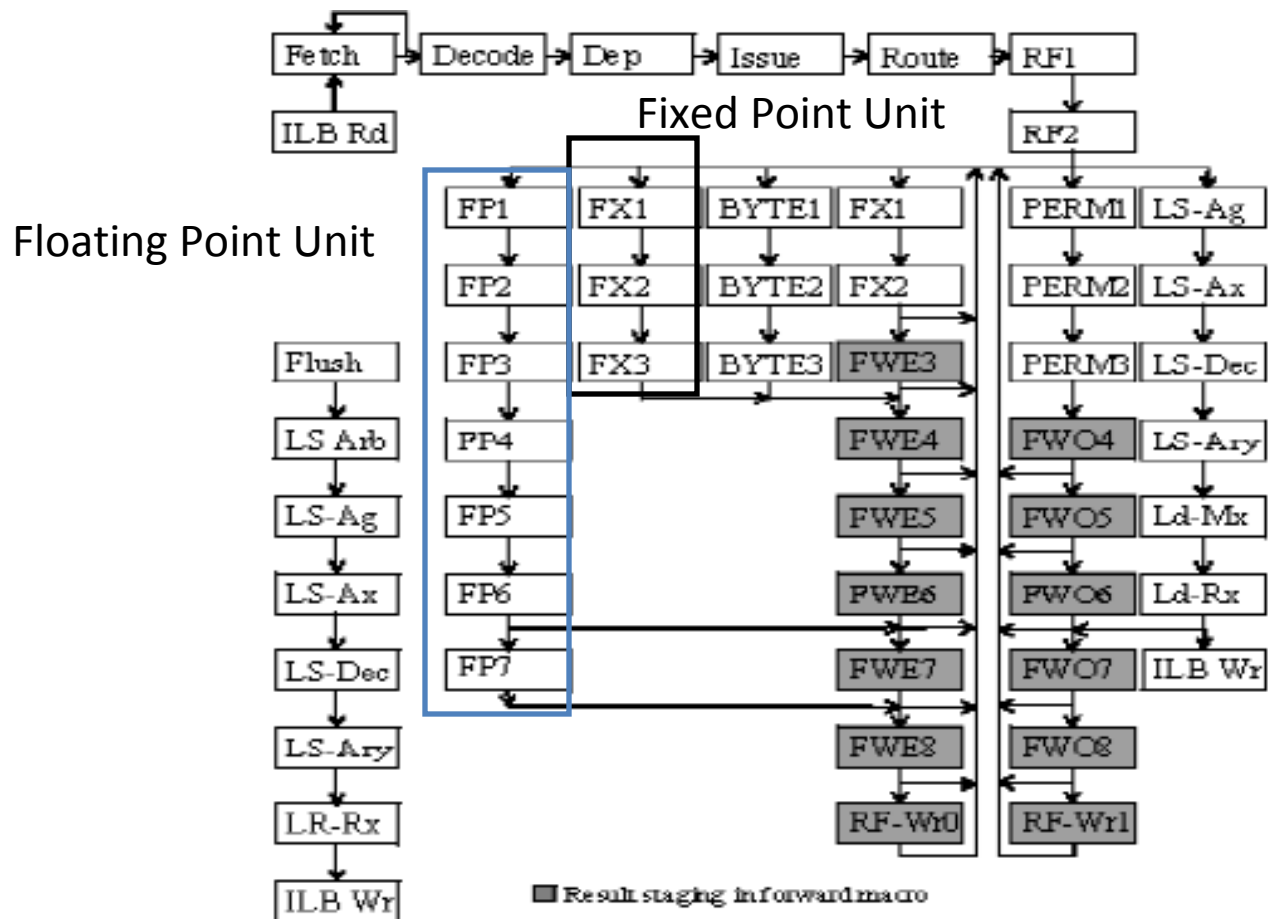
Pipeline Organization



Pipeline Organization



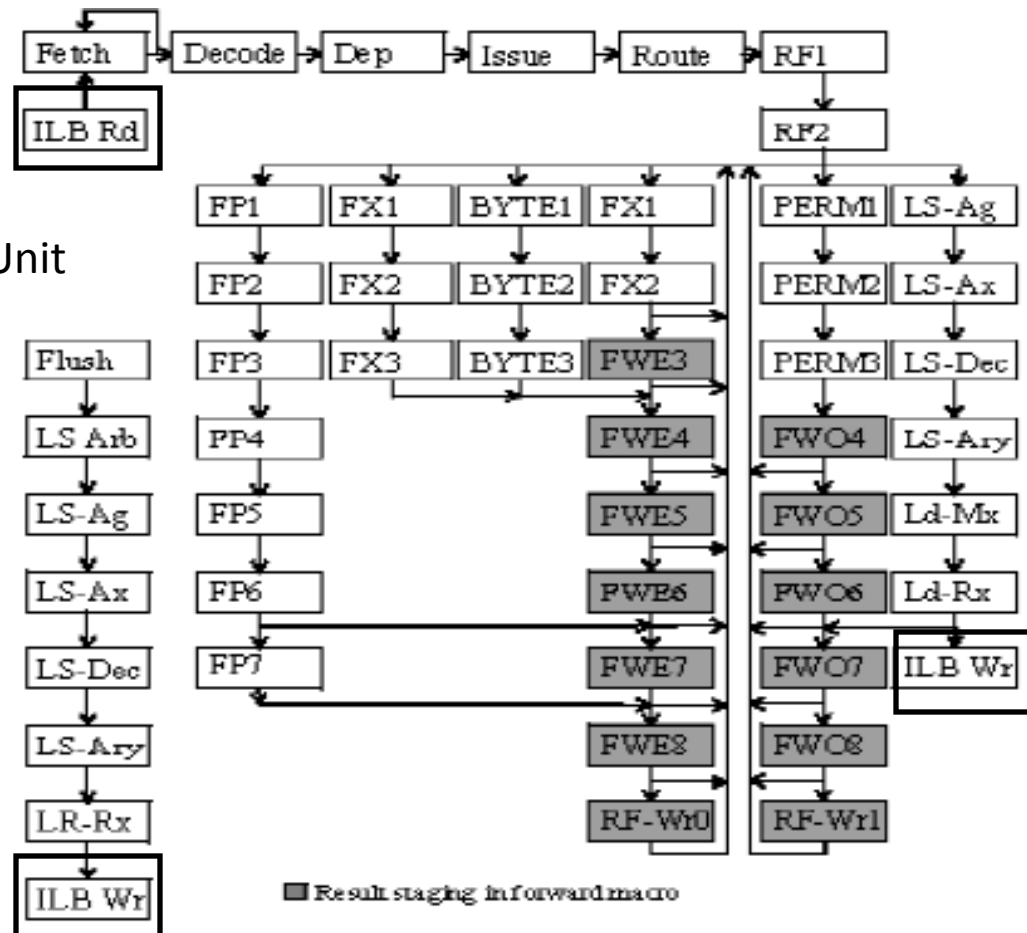
Pipeline Organization



Note: These are not the only locations for these units

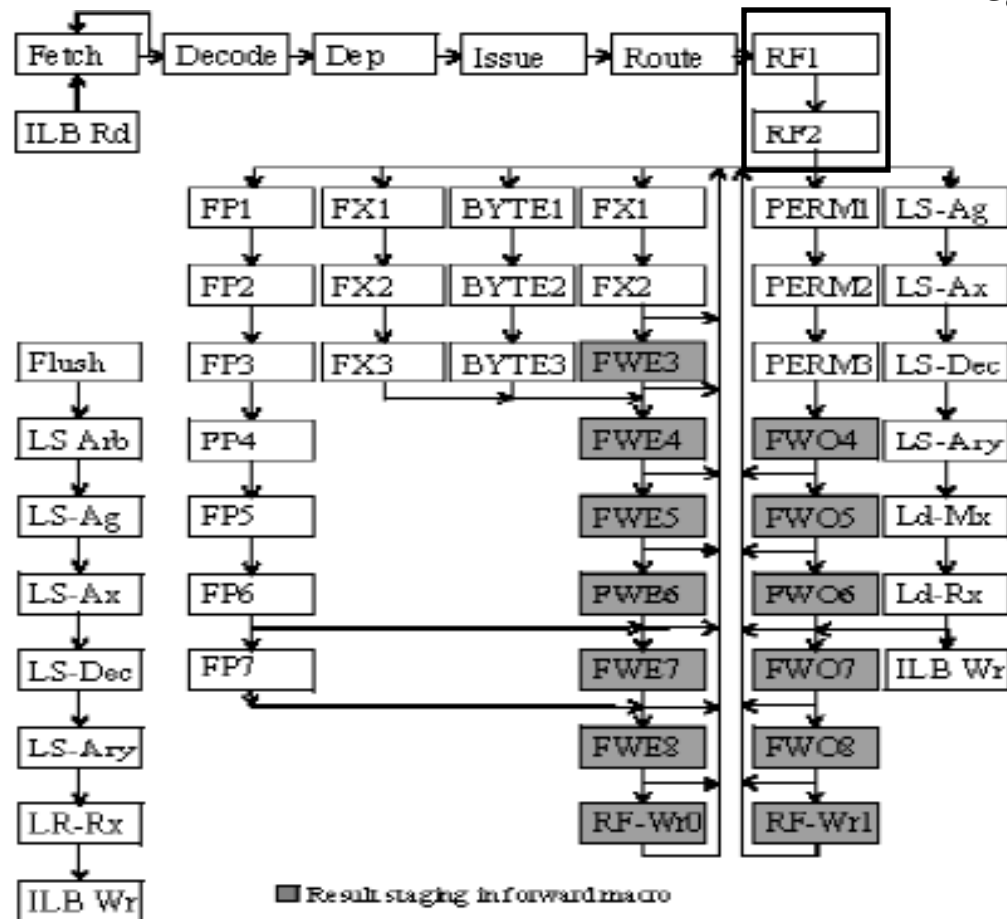
Pipeline Organization

Instruction Issue Unit



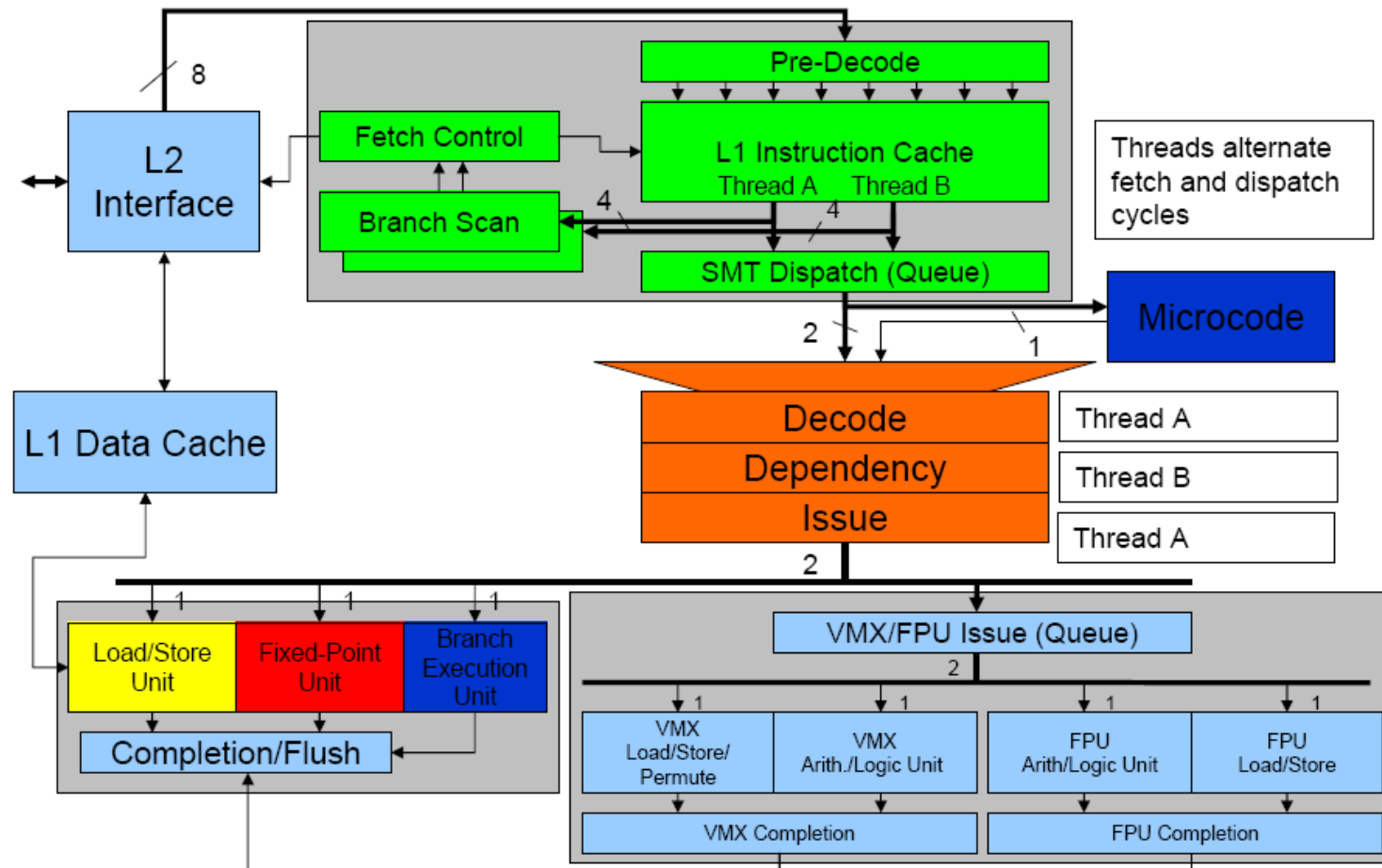
Pipeline Organization

Register File Unit

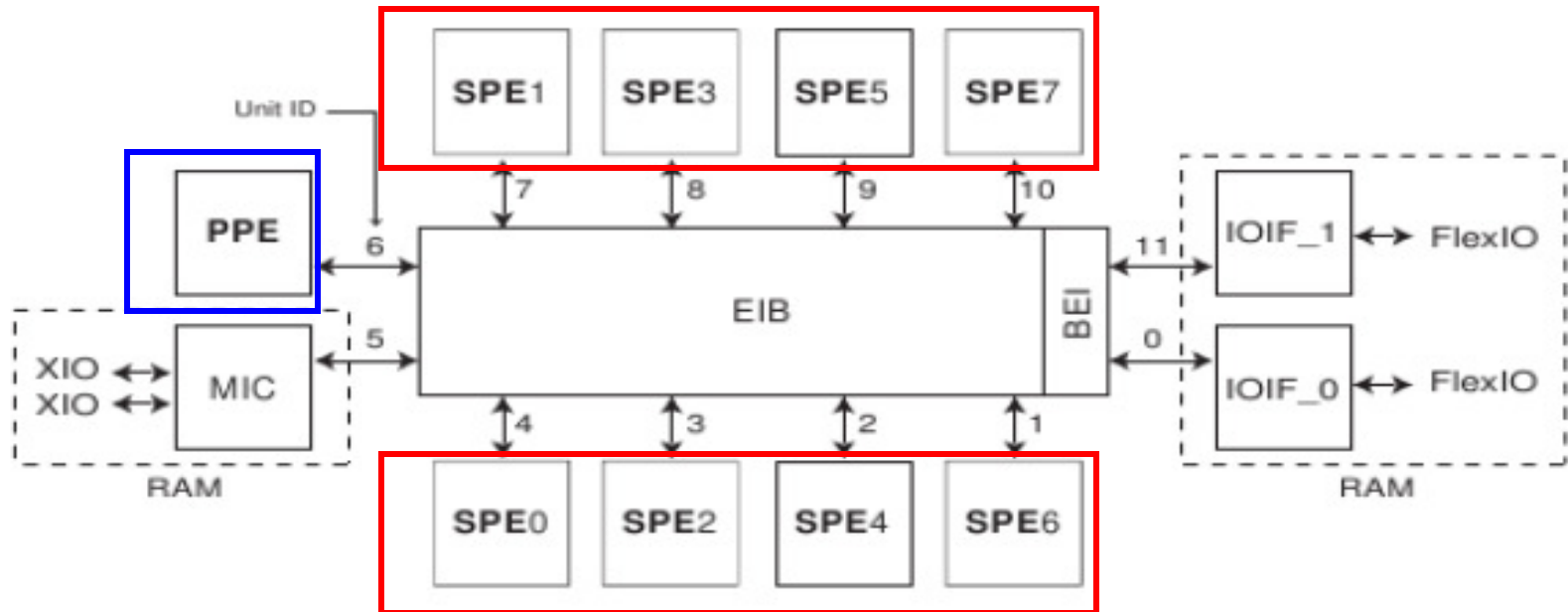


Pipeline Organization

*More detailed look into the pipeline



SPE & PPE



Cell/B.E.

SPE & PPE

- This implementation was based on issues of memory latency
- PPE has 32 KB L1 instruction and data caches
- PPE has 512 KB L2 unified (instruction and data) cache
- Each SPE's DMA controllers can maintain up to 16 DMA transfers in flight simultaneously

PPE vs. SPE

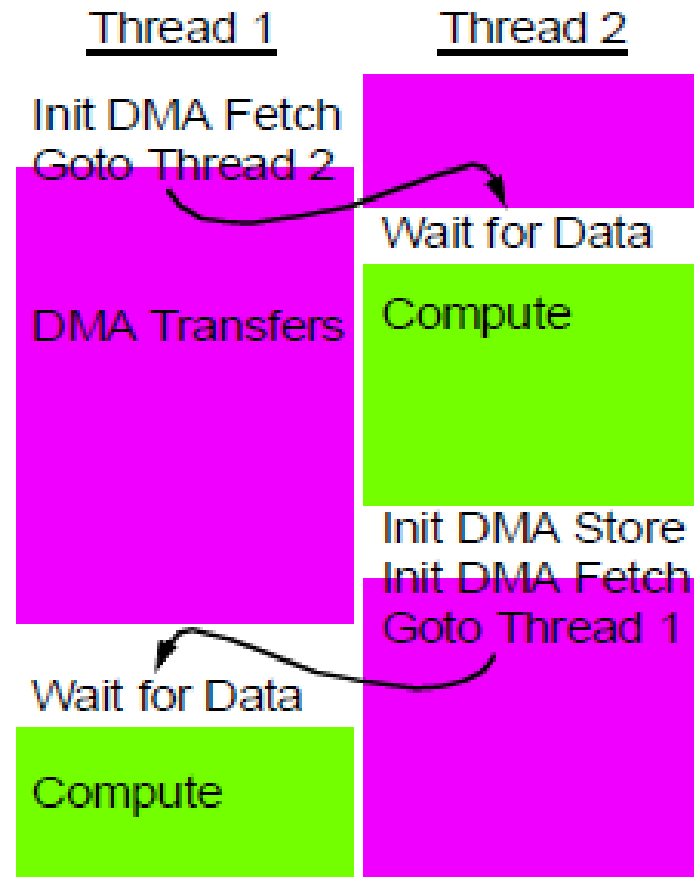
- PPE:
 - Accesses memory like a conventional processor
 - Uses load and store commands
 - Moves data between main storage and registers
- SPE:
 - Uses direct memory access commands
 - Moves data between main storage and a private local memory

PPE vs. SPE

- PPE:
 - Faster at task switching
- SPE:
 - More adept at compute intensive tasks
- Specialization is a factor for improvement in performance and power efficiency

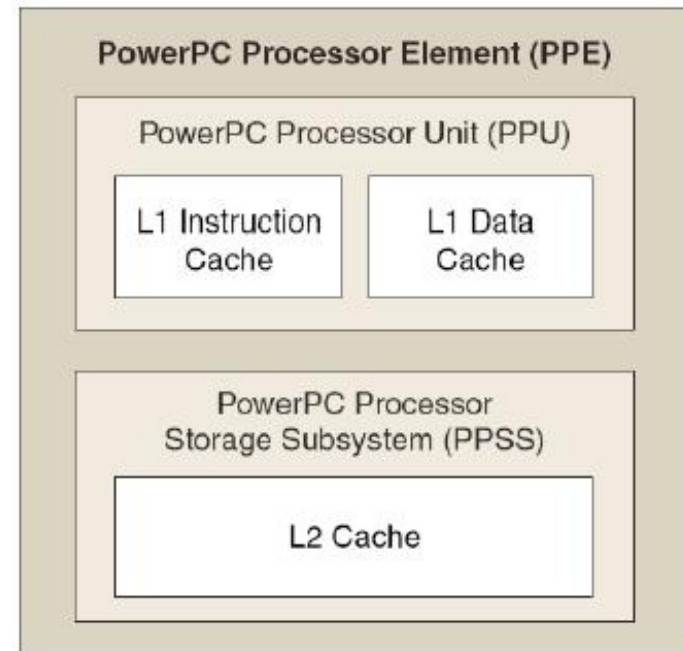
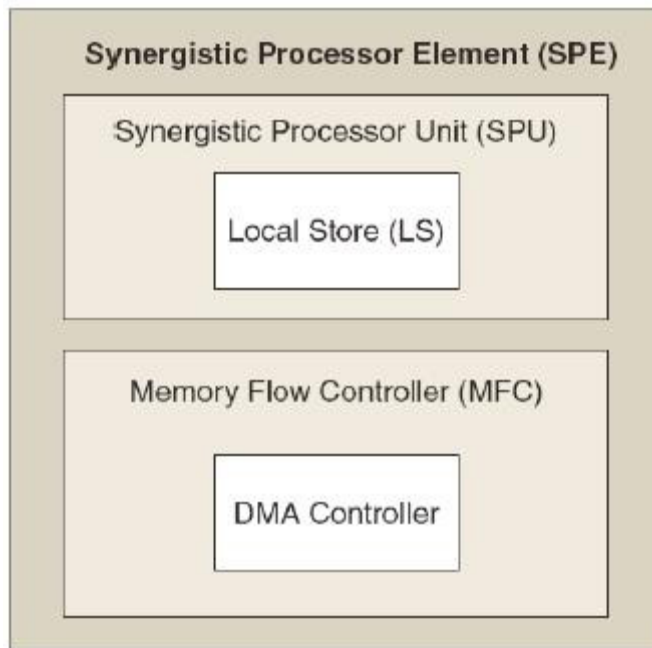
Double Buffering

- DMA commands are processed in parallel with software execution



SIMD Organization

- IBM PowerXCell 8i:
 - 8 Synergistic Processor Elements (SPE) + 1 PowerPC Processing Elements (PPE)



Synergistic Processing Elements

- Optimized for computing intensive code using SIMD
- Lots of SIMD registers (128 x 128-bit)
- Operations (per cycle):
 - sixteen 8-bit integers
 - eight 16-bit integers
 - four 32-bit integers
 - four single-precision floating-point numbers

PowerPC Processing Elements

- Optimized for control intensive code
- VMX SIMD unit: 128-bit SIMD register
- Operations (per cycle):
 - sixteen 8-bit integers
 - eight 16-bit integers
 - four 32-bit integers

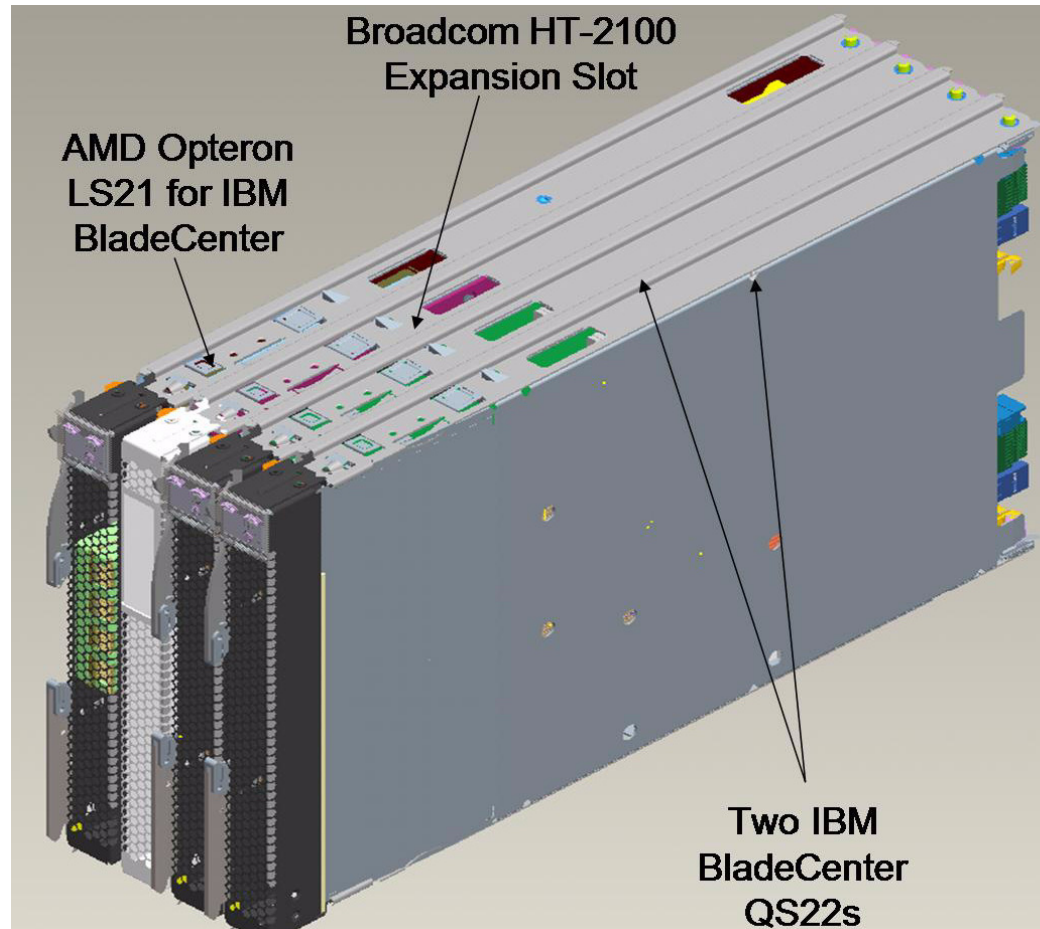
SIMD in VMX and SPE

- 128 bit-wide datapath
- 128 bit-wide registers

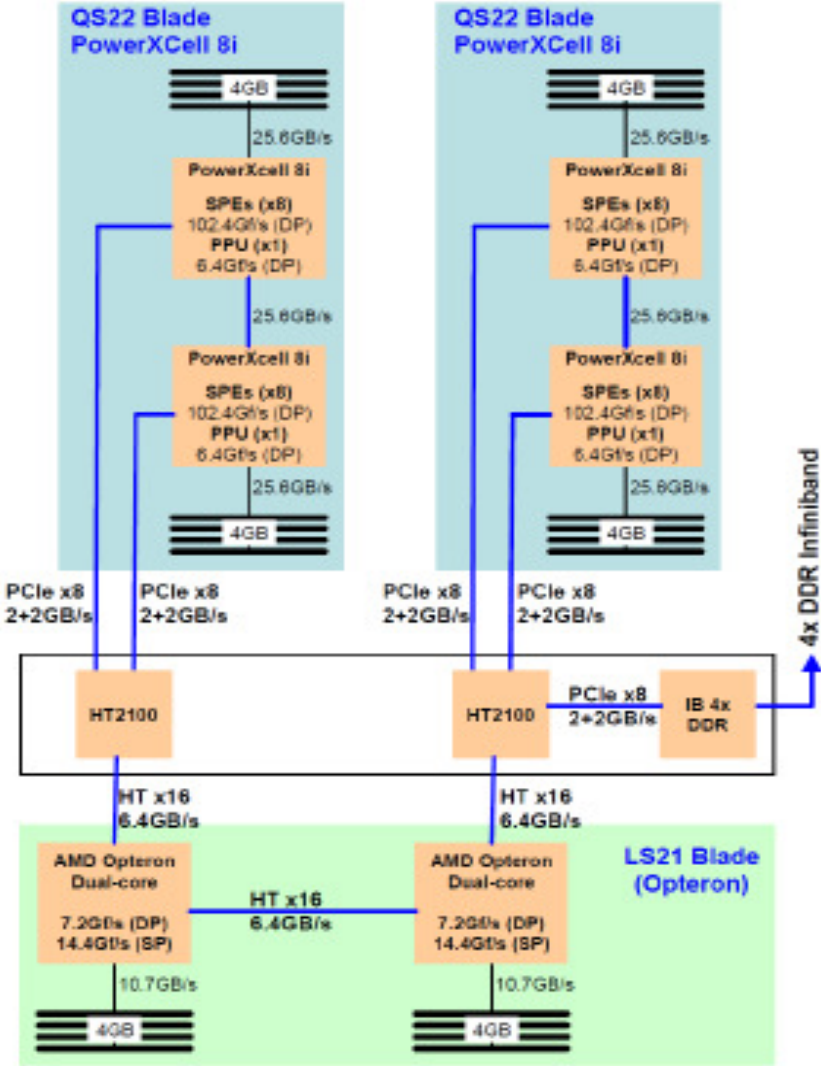
TriBlade

- One IBM LS21 Opteron blade
- Two IBM BladeCenter QS22 Cell/B.E. blades
- One blade that houses the communications fabric for the compute node

TriBlade



TriBlade



TriBlade

- One Cell/B.E. chip for each Opteron core
- Hierarchy: Opteron processors establish a master-subordinate relationship with the Cell/B.E. processors.
- Each TriBlade node is capable of 400 gigaflops of double precision compute power

Multithreading

- Threads alternate fetch and dispatch cycles
- To run multithreaded programs, multiple SPEs run simultaneously
- Usually, these programs create and organize the SPE threads as needed
- Each SPE supports one thread
- The PPE can support two threads without a program creating them

Multithreading

ALF -
Synchronization
constructs

