Cell Architecture

Course code: L1T1H1-10
Cell Ecosystem Solutions Enablement

Class Objectives – Things you will learn

- **Cell design motivation**

- **How cell overcomes three important limiters of contemporary microprocessor performance—power use, memory use, and processor frequency**

- **Cell processor organization and components**

  – Power processor element, block diagram, PXU pipeline

  – Synergistic processor element, block diagram, SXU pipeline

  – Memory flow controller and MFC commands

  – Element interconnect bus, command and data topology

  – I/O and memory interfaces

Class Agenda

- **Cell concept**

- **Architecture motivators**

- **Cell synergy**

- **Cell features**

- **Cell processor components**

  – Power processor element

  – Synergistic processor element

  – Memory flow controller

  – Element interconnect bus

  – I/O and memory interfaces

  – Resource allocation management

References

- Jim Kahle, Cell Broadband Engine and Cell Broadband Engine Architecture

Trademarks - Cell Broadband Engine ™ is a trademark of Sony Computer Entertainment, Inc.

Cell Concept

- Compatibility with 64b Power Architecture™
  - Builds on and leverages IBM investment and community

- Increased efficiency and performance
  - Attacks on the "Power Wall"
    - Non Homogenous Coherent Multiprocessor
    - High design frequency @ a low operating voltage with advanced power management
  - Attacks on the "Memory Wall"
    - Streaming DMA architecture
    - 3-level Memory Model: Main Storage, Local Storage, Register Files
  - Attacks on the "Frequency Wall"
    - Highly optimized implementation
    - Large shared register files and software controlled branching to allow deeper pipelines

- Interface between user and networked world
  - Image rich information, virtual reality
  - Flexibility and security

- Multi-OS support, including RTOS / non-RTOS
  - Combine real-time and non-real time worlds

Architecture Motivators

Market Requirements
- ► Natural interaction with the system
- ► Consumer acceptable interaction
- ► Improve Experience
  - – Ease of use
  - – High degree of interaction
  - – Responsive
  - – Realism
  - – Interconnected through network to other devices

Holistic Design Approach
- ► Architecture
- ► Hardware implementation
- ► System structure
- ► Programming Model

Technical Requirements
- ► Dual environment: Real time and conventional
- ► High FLOPS Computational density
- ► High parallelism
- ► Bandwidth & latency controls
- ► Realtime response
- ► Resource reservation
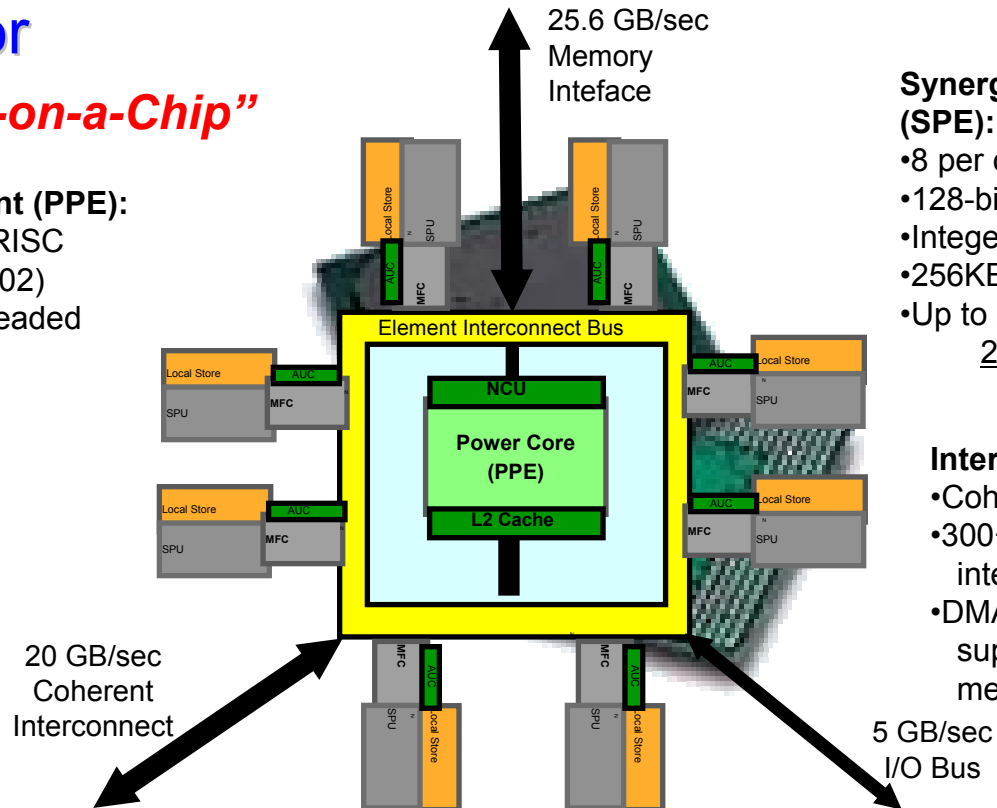- ► High bandwidth
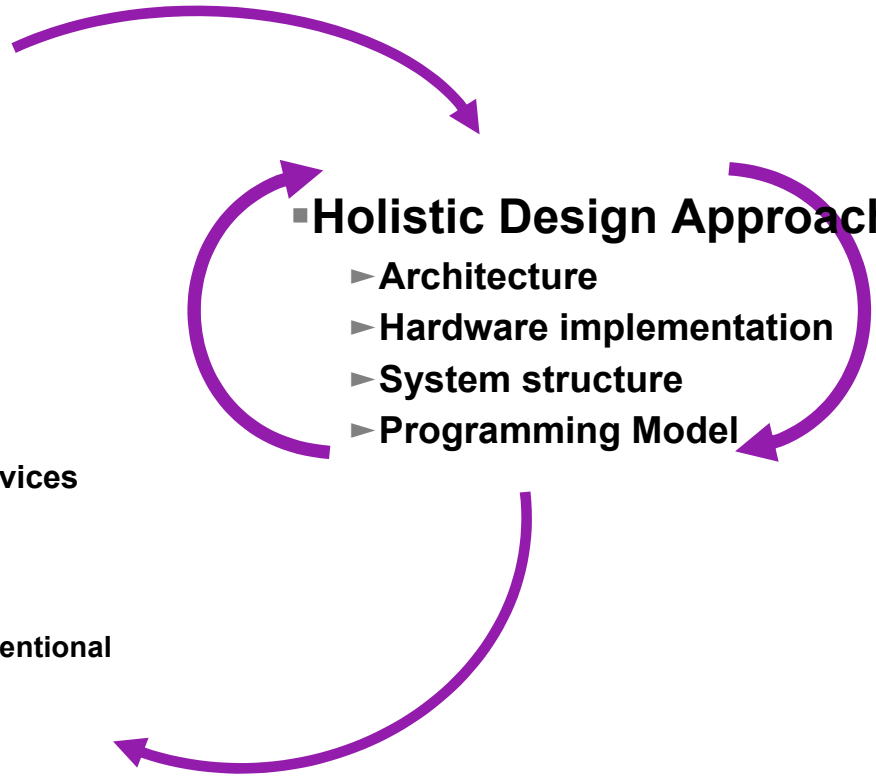
Cell Synergy

- **Cell is not a collection of different processors, but a synergistic whole**
  - Operation paradigms, data formats and semantics consistent
  - Share address translation and memory protection model
- **PPE for operating systems and program control**
- **SPE optimized for efficient data processing**
  - SPEs share Cell system functions provided by Power Architecture
  - MFC implements interface to memory
    - Copy in/copy out to local storage
- **PowerPC provides system functions**
  - Virtualization
  - Address translation and protection
  - External exception handling
- **EIB integrates system as data transport hub**

Cell Processor

"Supercomputer-on-a-Chip"

Power Processor Element (PPE):
- General Purpose, 64-bit RISC Processor (PowerPC 2.02)
- 2-Way Hardware Multithreaded
- L1 : 32KB I ; 32KB D
- L2 : 512KB
- Coherent load/store
- VMX
- 3.2 GHz

25.6 GB/sec Memory Inteface

Element Interconnect Bus

NCU

Power Core (PPE)

L2 Cache

Local Store
AUC
SPU
MFC

20 GB/sec Coherent Interconnect

5 GB/sec I/O Bus

Synergistic Processor Elements (SPE):
- 8 per chip
- 128-bit wide SIMD Units
- Integer *and* Floating Point capable
- 256KB Local Store
- Up to 25.6 GF/s per SPE --- 200GF/s total *

Internal Interconnect:
- Coherent ring structure
- 300+ GB/s total internal interconnect bandwidth
- DMA control to/from SPEs supports >100 outstanding memory requests

External Interconnects:
- 25.6 GB/sec BW memory interface
- 2 Configurable I/O Interfaces
    - Coherent interface (SMP)
    - Normal I/O interface (I/O & Graphics)
    - Total BW configurable between interfaces
    - Up to 35 GB/s out
    - Up to 25 GB/s in

Memory Management & Mapping
- SPE Local Store aliased into PPE system memory
- MFC/MMU controls SPE DMA accesses
    - Compatible with PowerPC Virtual Memory architecture
    - S/W controllable from PPE MMIO
- Hardware or Software TLB management
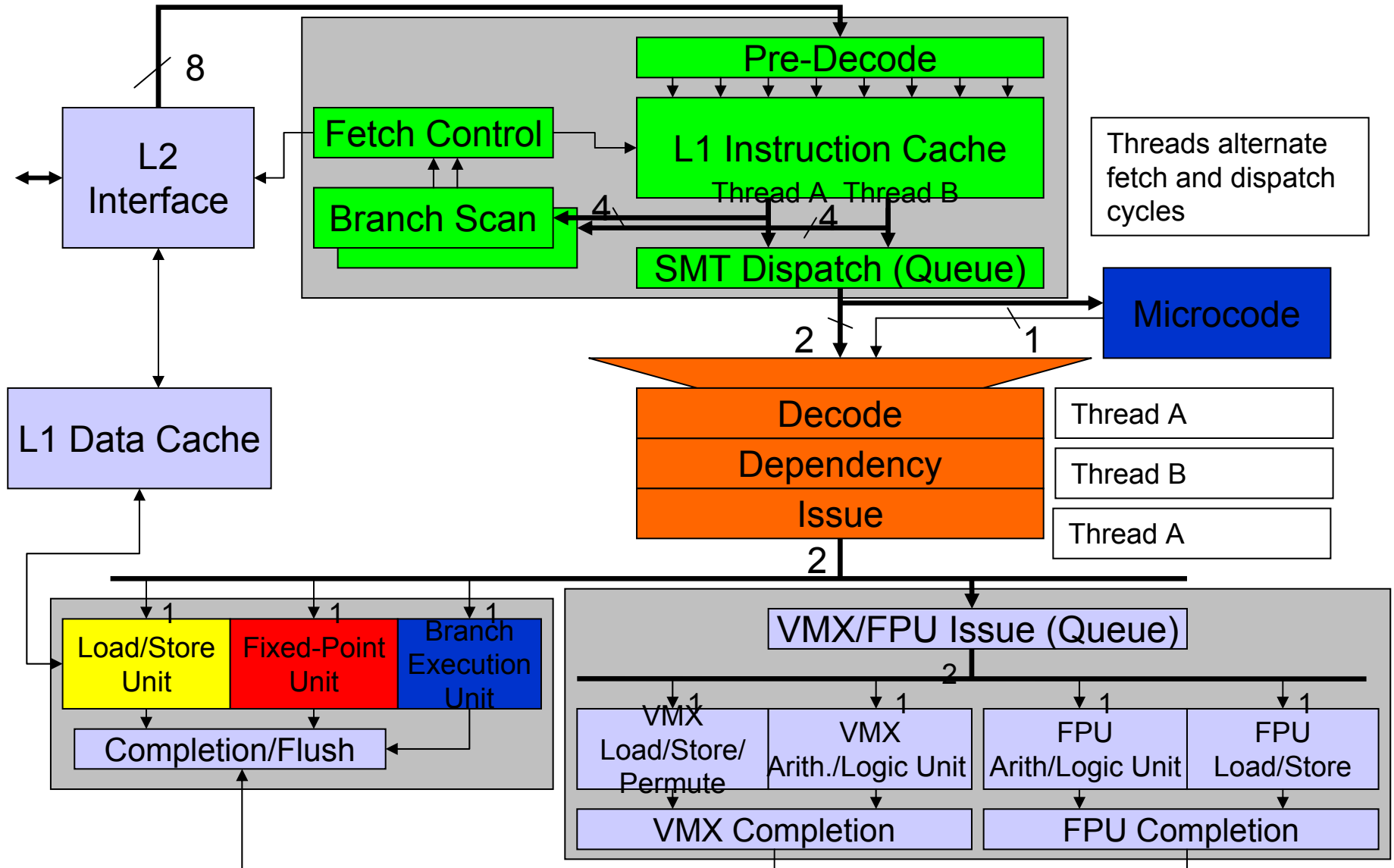- SPE DMA access protected by MFC/MMU

Power Processor Element

- **PPE handles operating system and control tasks**
  - 64-bit Power Architecture^{TM} with VMX
  - In-order, 2-way hardware simultaneous multi-threading (SMT)
 - Coherent Load/Store with 32KB I & D L1 and 512KB L2

PPE BLOCK DIAGRAM

Pre-Decode

Fetch Control

L1 Instruction Cache
Thread A   Thread B

Branch Scan

SMT Dispatch (Queue)

L2 Interface

8

Threads alternate fetch and dispatch cycles

4   4

Microcode

2   1

L1 Data Cache

Decode
Dependency
Issue

Thread A

Thread B

Thread A

2

Load/Store Unit

Fixed-Point Unit

Branch Execution Unit

1   1   1

Completion/Flush

VMX/FPU Issue (Queue)

2

VMX Load/Store/Permute

VMX Arith./Logic Unit

FPU Arith/Logic Unit

FPU Load/Store

1   1   1   1

VMX Completion

FPU Completion

Synergistic Processor Element

- **SPE provides computational performance**
  - Dual issue, up to 16-way 128-bit SIMD
  - Dedicated resources: 128 128-bit RF, 256KB Local Store
  - Each can be dynamically configured to protect resources
  - Dedicated DMA engine: Up to 16 outstanding requests
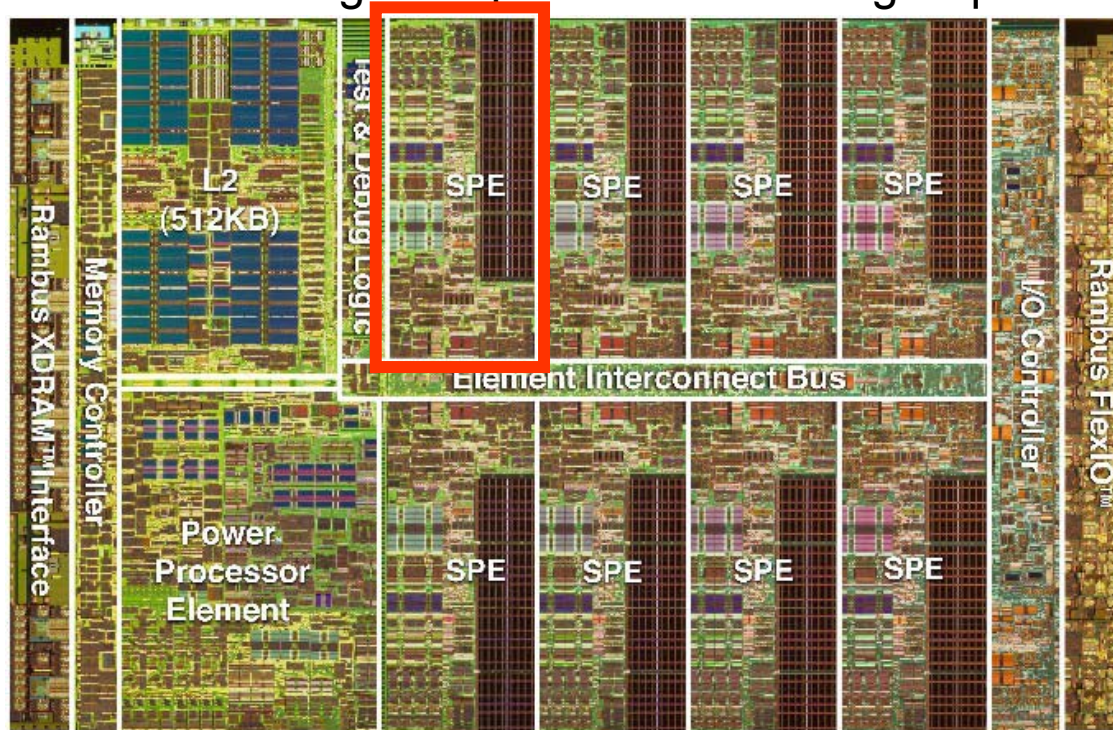
SPE Highlights
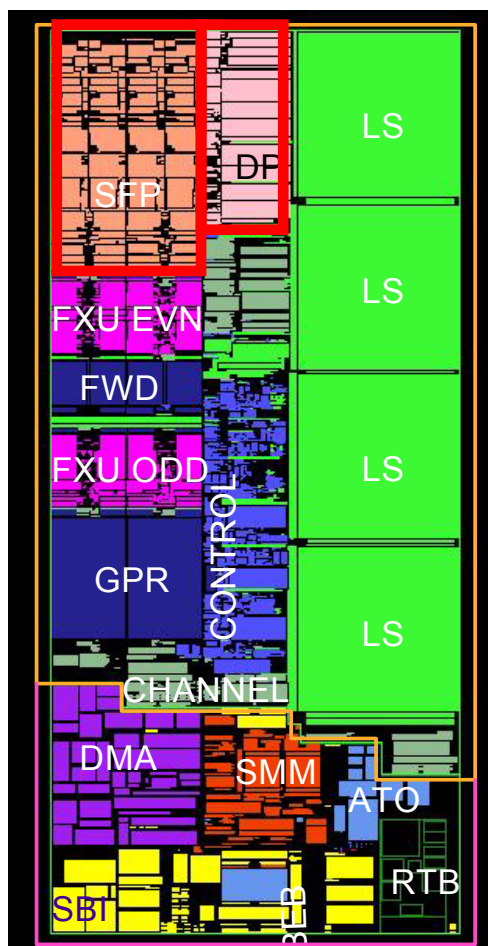
14.5mm² (90nm SOI)

- **RISC like organization**
  - 32 bit fixed instructions
  - Clean design – unified Register file
- **User-mode architecture**
  - No translation/protection within SPU
  - DMA is full Power Arch protect/x-late
- **VMX-like SIMD dataflow**
  - Broad set of operations (8 / 16 / 32 Byte)
  - Graphics SP-Float
  - IEEE DP-Float
- **Unified register file**
  - 128 entry x 128 bit
- **256KB Local Store**
  - Combined I & D
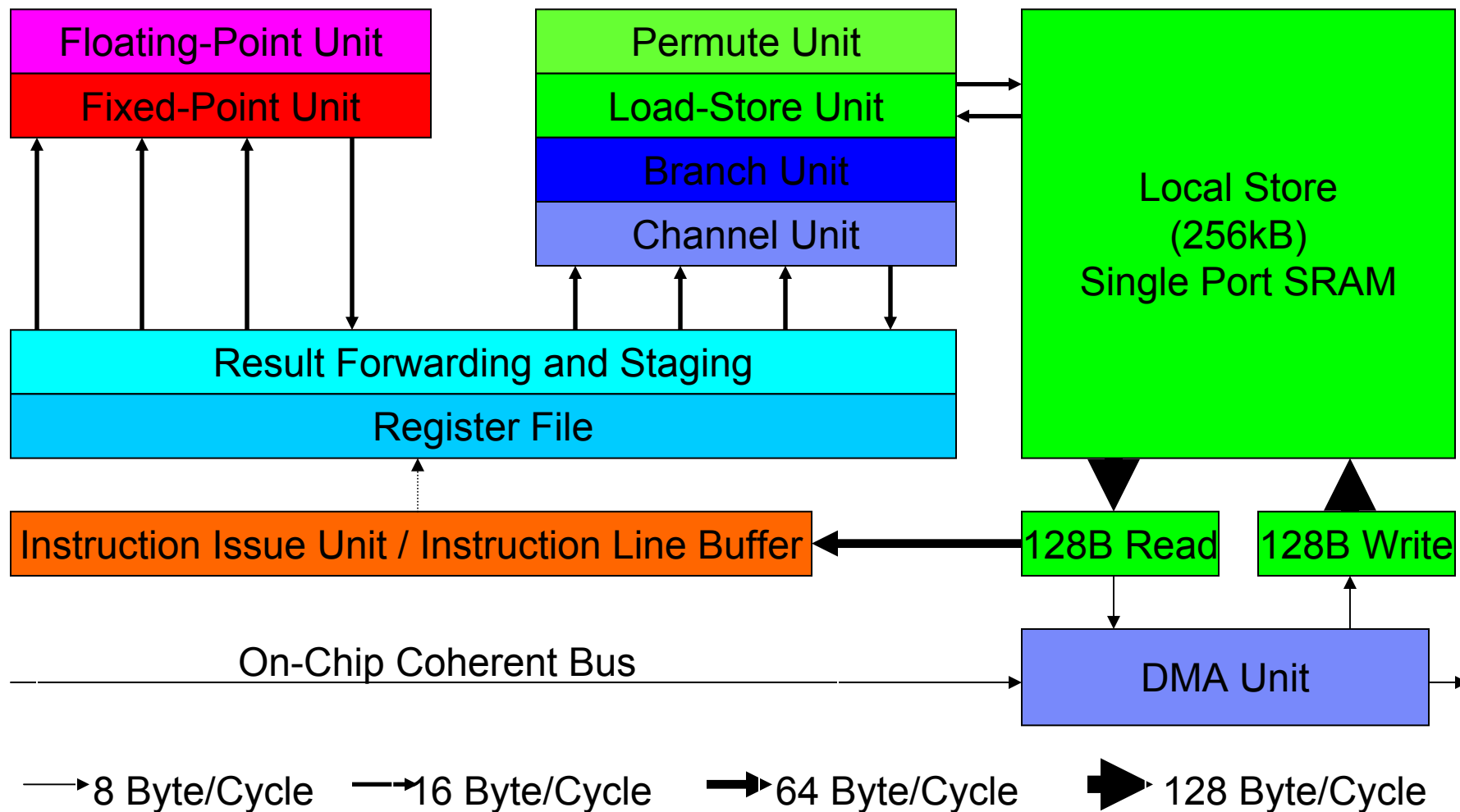 - 16B/cycle L/S bandwidth
  - 128B/cycle DMA bandwidth

What is a Synergistic Processor?
(and why is it efficient?)

- Local Store "is" large 2nd level register file / private instruction store instead of cache
  - Asynchronous transfer (DMA) to shared memory
  - Frontal attack on the Memory Wall
- Media Unit turned into a Processor
  - Unified (large) Register File
  - 128 entry x 128 bit
- Media & Compute optimized
  - One context
  - SIMD architecture
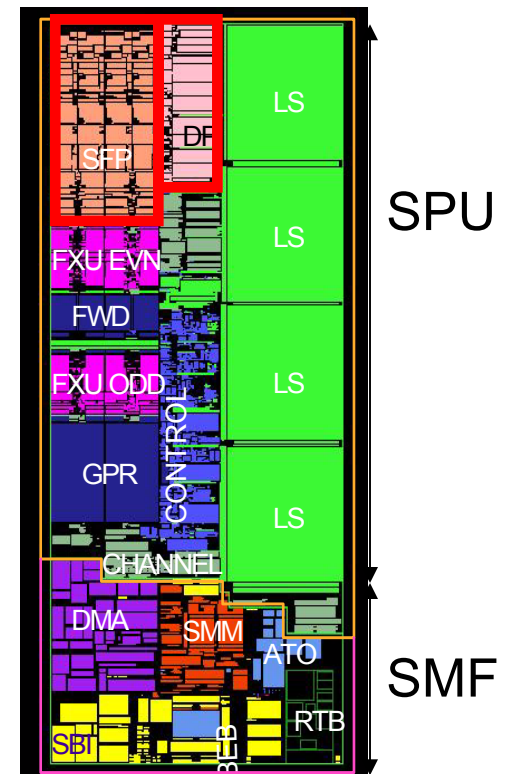
SPU Detail

Synergistic Processor Element (SPE)
- User-mode architecture
  - No translation/protection within SPE
  - DMA is full PowerPC protect/xlate
- Direct programmer control
  - DMA/DMA-list
  - Branch hint
- VMX-like SIMD dataflow
  - Graphics SP-Float
  - No saturate arith, some byte
  - IEEE DP-Float (BlueGene-like)
- Unified register file
  - 128 entry x 128 bit
- 256KB Local Store
  - Combined I & D
  - 16B/cycle L/S bandwidth
  - 128B/cycle DMA bandwidth
- Memory Flow Control (MFC)
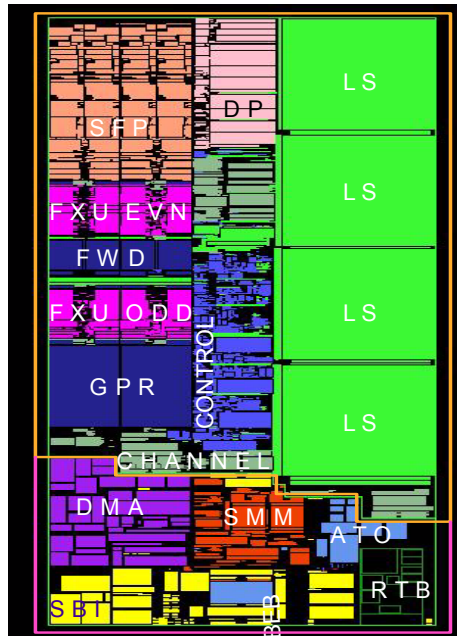
SPU Units:
- Simple (FXU even)
  - Add/Compare
  - Rotate
  - Logical, Count Leading Zero
- Permute (FXU odd)
  - Permute
  - Table-lookup
- FPU (Single / Double Precision)
- Control (SCN)
  - Dual Issue, Load/Store, ECC Handling
- Channel (SSC) – Interface to MFC
- Register File (GPR/FWD)

SPU Latencies

Simple fixed point	- 2 cycles*	
Complex fixed point	- 4 cycles*	
Load	- 6 cycles*	
• Local store size = 256 KB		
Single-precision (ER) float	- 6 cycles*	
Integer multiply	- 7 cycles*	
Branch miss	- 20 cycles	
• No penalty if correctly hinted		
DP (IEEE) float	- 13 cycles*	
• Partially pipelined		
Enqueue DMA Command	- 20 cycles*	

SPE BLOCK DIAGRAM

Floating-Point Unit		Permute Unit	Local Store (256kB) Single Port SRAM
Fixed-Point Unit		Load-Store Unit	
		Branch Unit	
		Channel Unit	

Result Forwarding and Staging

Register File

Instruction Issue Unit / Instruction Line Buffer

128B Read   **128B Write**

On-Chip Coherent Bus

DMA Unit

→ 8 Byte/Cycle   → 16 Byte/Cycle   ➡ 64 Byte/Cycle   ▶ 128 Byte/Cycle

SPE Structure

- **Scalar processing supported on data-parallel substrate**

  – All instructions are data parallel and operate on vectors of elements

  – Scalar operation defined by instruction use, not opcode

    • Vector instruction form used to perform operation

- **Preferred slot paradigm**

  – Scalar arguments to instructions found in "preferred slot"

  – Computation can be performed in any slot

Register Scalar Data Layout

- **Preferred slot in bytes 0-3**
  - By convention for procedure interfaces
  - Used by instructions expecting scalar data
    - Addresses, branch conditions, generate controls for insert

	Preferred Slot														Byte Index
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Registers				BYTE												
			HALFWORD													
	ADDRESS															
	WORD															
	DOUBLEWORD															
								QUAD WORD								
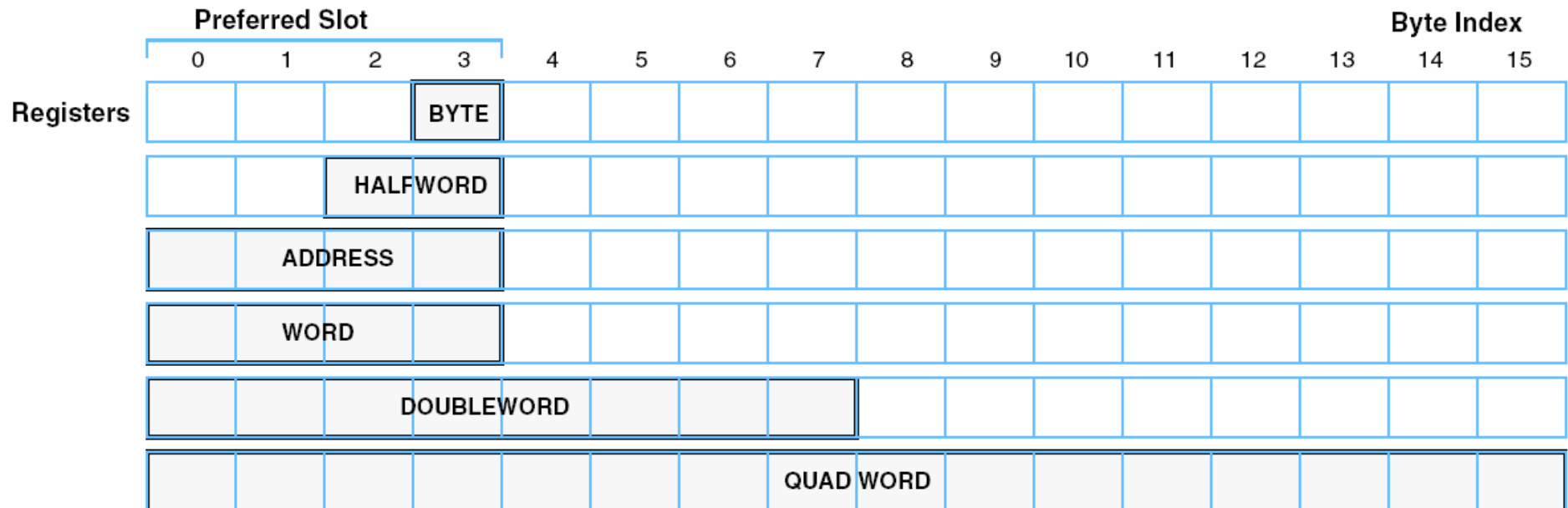
Cell Processor Components

Synergistic Processor Element (SPE):

- Provides the computational performance
- Simple RISC User Mode Architecture
  - Dual issue VMX-like
  - Graphics SP-Float
  - IEEE DP-Float
- Dedicated resources: unified 128x128-bit RF, 256KB Local Store
- Dedicated DMA engine: Up to 16 outstanding requests
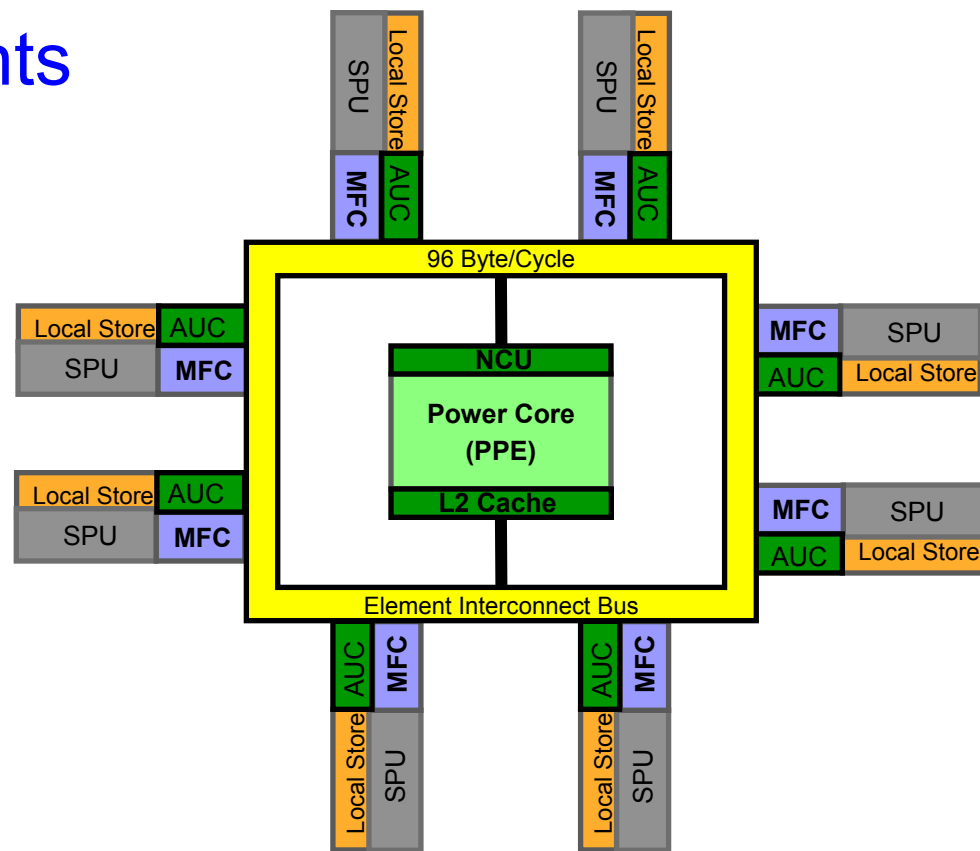
Memory Management & Mapping

- SPE Local Store aliased into PPE system memory
- MFC/MMU controls / protects SPE DMA accesses
  - Compatible with PowerPC Virtual Memory Architecture
  - SW controllable using PPE MMIO
- DMA 1,2,4,8,16,128 -> 16Kbyte transfers for I/O access
- Two queues for DMA commands: Proxy & SPU
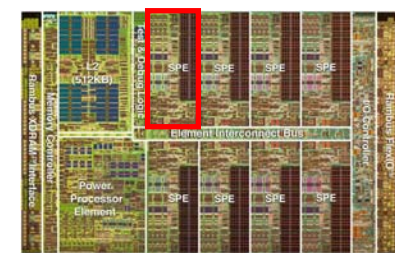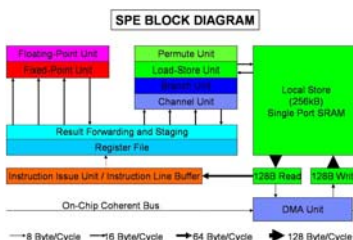
Memory Flow Controller Commands

DMA Commands

Put - Transfer from Local Store to EA space

Puts - Transfer and Start SPU execution

Putr - Put Result - (Arch. Scarf into L2)

Putl - Put using DMA List in Local Store

Putrl - Put Result using DMA List in LS (Arch)

Get - Transfer from EA Space to Local Store

Gets - Transfer and Start SPU execution

Getl - Get using DMA List in Local Store

Sndsig - Send Signal to SPU

Command Modifiers: **<f,b>**

f: Embedded Tag Specific Fence

Command will not start until all previous commands
in same tag group have completed

b: Embedded Tag Specific Barrier

Command and all subsiquent commands in same
tag group will not start until previous commands in same
tag group have completed

Command Parameters

LSA - Local Store Address (32 bit)

EA - Effective Address (32 or 64 bit)

TS - Transfer Size (16 bytes to 16K bytes)

LS - DMA List Size (8 bytes to 16 K bytes)

TG - Tag Group(5 bit)

CL - Cache Management / Bandwidth Class

Synchronization Commands

Lockline (Atomic Update) Commands:

getllar - DMA 128 bytes from EA to LS and set Reservation

putllc - Conditionally DMA 128 bytes from LS to EA

putlluc - Unconditionally DMA 128 bytes from LS to EA

barrier - all previous commands complete before subsiquent
commands are started

mfcsync - Results of all previous commands in Tag group
are remotely visible

mfceieio - Results of all preceding Puts commands in same
group visible with respect to succeeding Get commands

SL1 Cache Management Commands

sdcrt - Data cache region touch (DMA Get hint)

sdcrtst - Data cache region touch for store (DMA Put hint)

sdcrz - Data cache region zero

sdcrs - Data cache region store

sdcrf - Data cache region flush

Cell Processor Components

Power Processor Element (PPE):
- General purpose, 64-bit RISC processor (PowerPC AS 2.0.2)
- 2-Way hardware multithreaded
- L1 : 32KB I ; 32KB D
- L2 : 512KB
- Coherent load / store
- VMX-32
- Realtime Controls
  - Locking L2 Cache & TLB
  - Software / hardware managed TLB
  - Bandwidth / Resource Reservation
  - Mediated Interrupts

Element Interconnect Bus (EIB):
- Four 16 byte data rings supporting multiple simultaneous transfers per ring
- 96Bytes/cycle peak bandwidth
- Over 100 outstanding requests

In the Beginning
 – the solitary Power Processor

96 Byte/Cycle

NCU

Power Core (PPE)

L2 Cache

Element Interconnect Bus
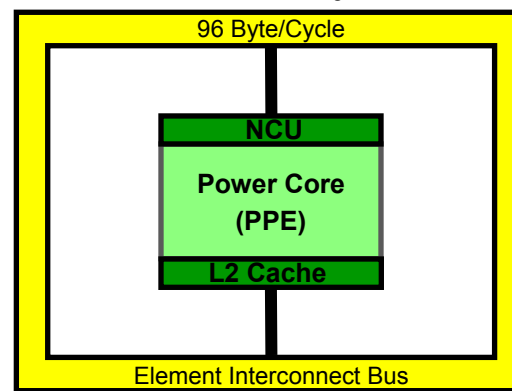
Custom Designed
 – for high frequency, space, and power efficiency

PPE BLOCK DIAGRAM
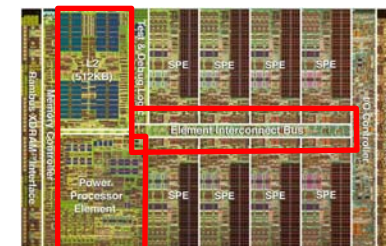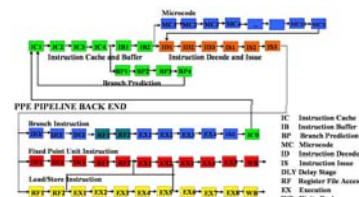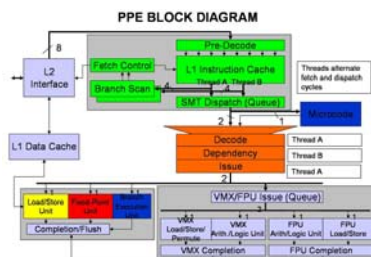
Element Interconnect Bus

- **EIB data ring for internal communication**

  - Four 16 byte data rings, supporting multiple transfers

  - 96B/cycle peak bandwidth

  - Over 100 outstanding requests
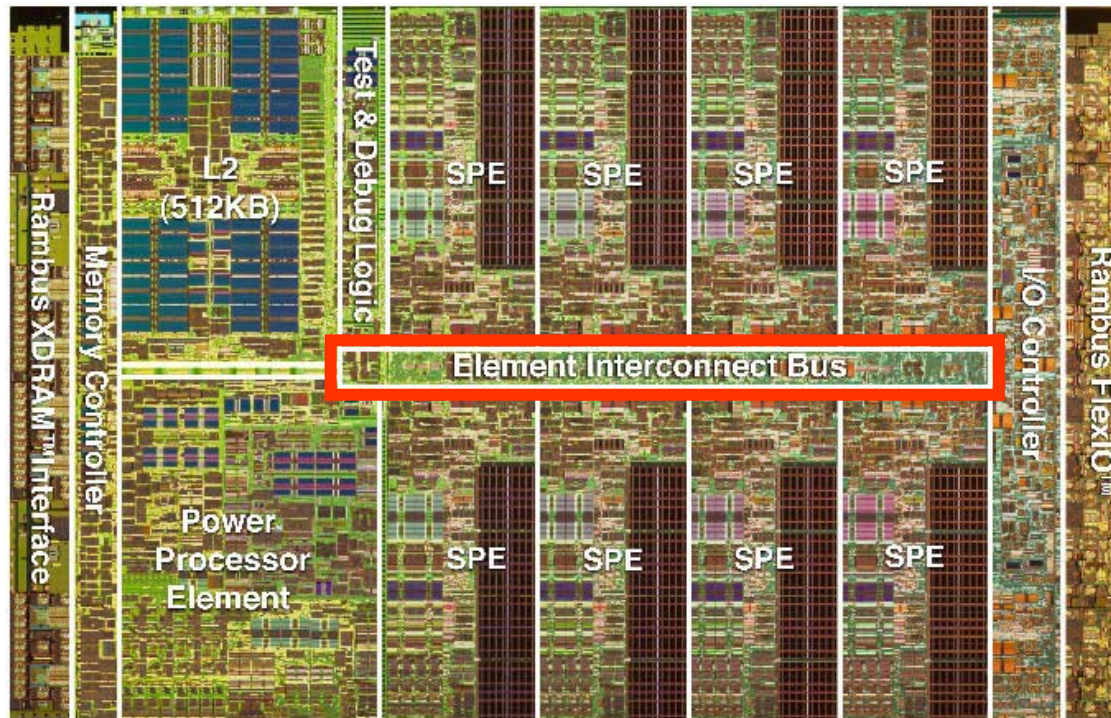
Internal Bandwidth Capability

- **Each EIB Bus data port supports 25.6GBytes/sec* *in each direction***

- **The EIB Command Bus streams commands fast enough to support 102.4 GB/sec for coherent commands, and 204.8 GB/sec for non-coherent commands.**

- **The EIB data rings can sustain 204.8GB/sec for certain workloads, with transient rates as high as 307.2GB/sec between bus units**

* The above numbers assume a 3.2GHz core frequency – internal bandwidth scales with core frequency

Element Interconnect Bus - Data Topology

- **Four 16B data rings connecting 12 bus elements**
  - Two clockwise / Two counter-clockwise
- **Physically overlaps all processor elements**
- **Central arbiter supports up to three concurrent transfers per data ring**
  - Two stage, dual round robin arbiter
- **Each element port simultaneously supports 16B in and 16B out data path**
  - Ring topology is transparent to element data interface
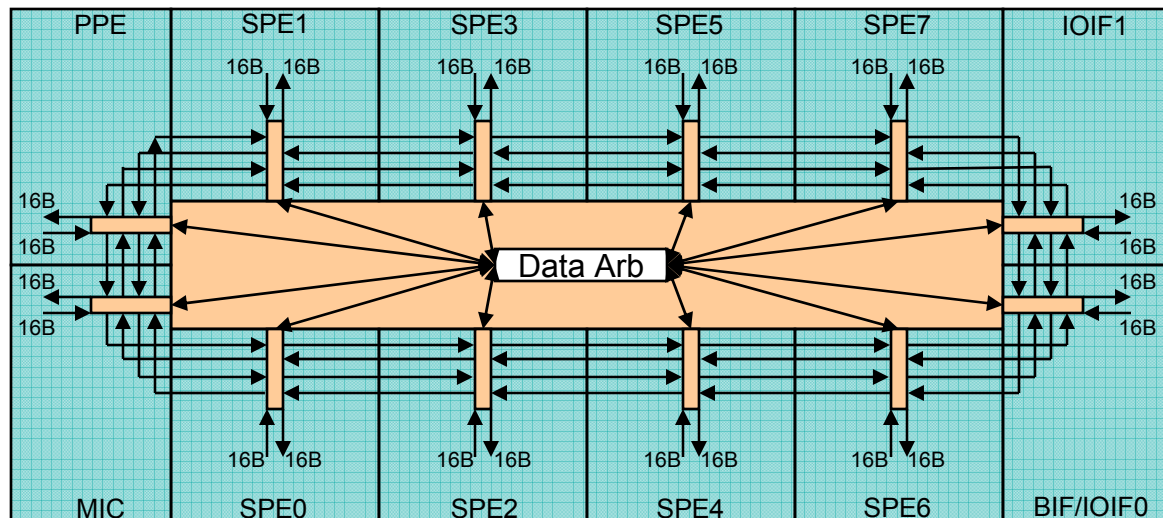
Example of eight concurrent transactions
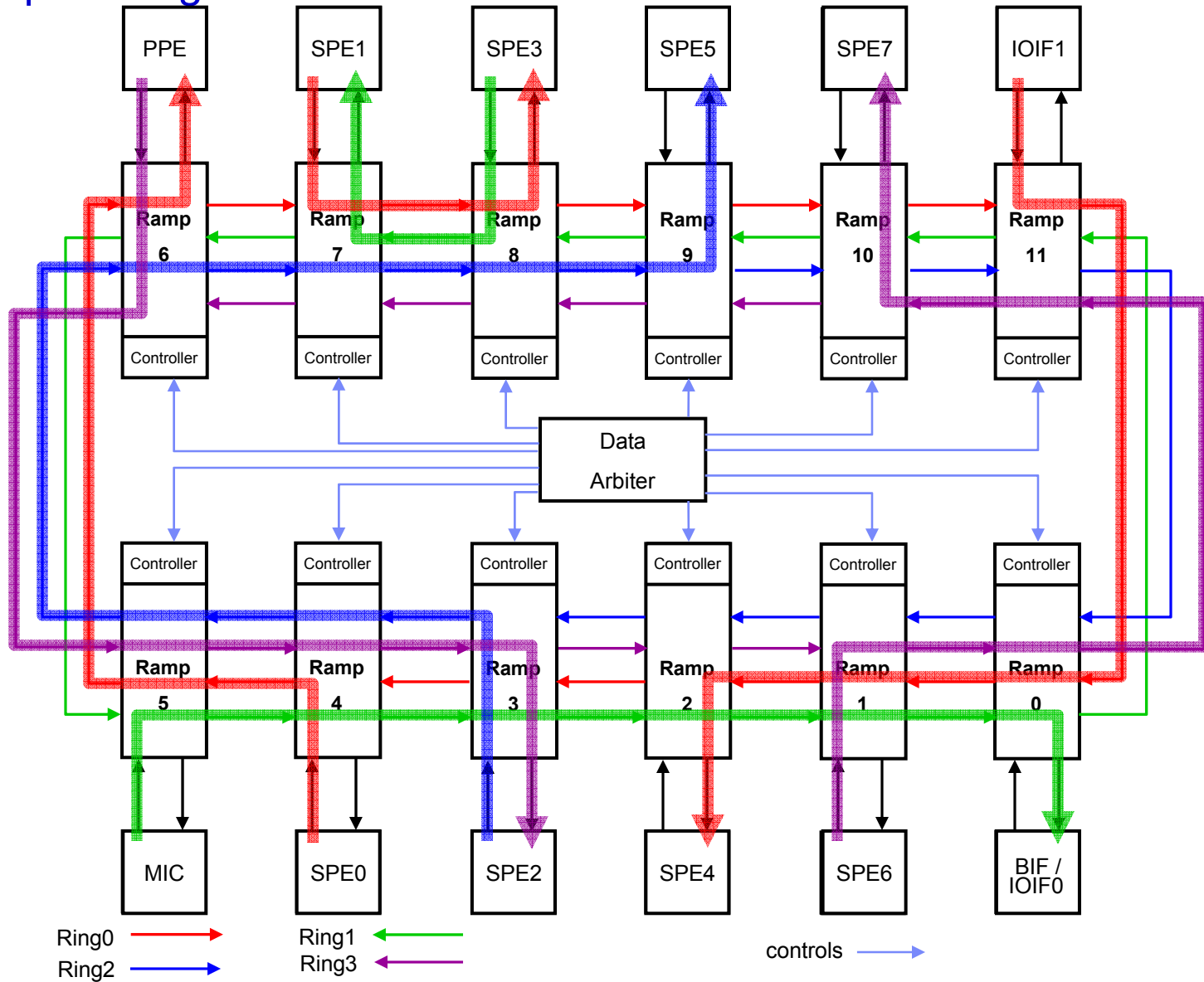
Cell Processor Components

Token Manager (TKM):

- Bandwidth / Resource Reservation for shared resources
- Optionally enabled for RT tasks or LPAR
- Multiple Resource Allocation Groups (RAGs)
- Generates access tokens at configurable rate for each allocation group
  - 1 per each memory bank (16 total)
  - 2 for each IOIF (4 total)
- Requestors assigned RAG ID by OS / hypervisor
  - Each SPE
  - PPE L2 / NCU
  - IOIF 0 Bus Master
  - IOIF 1 Bus Master
- Priority order for using another RAGs unused tokens
- Resource over committed warning interrupt
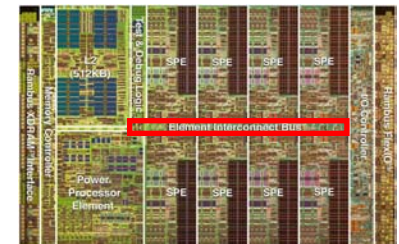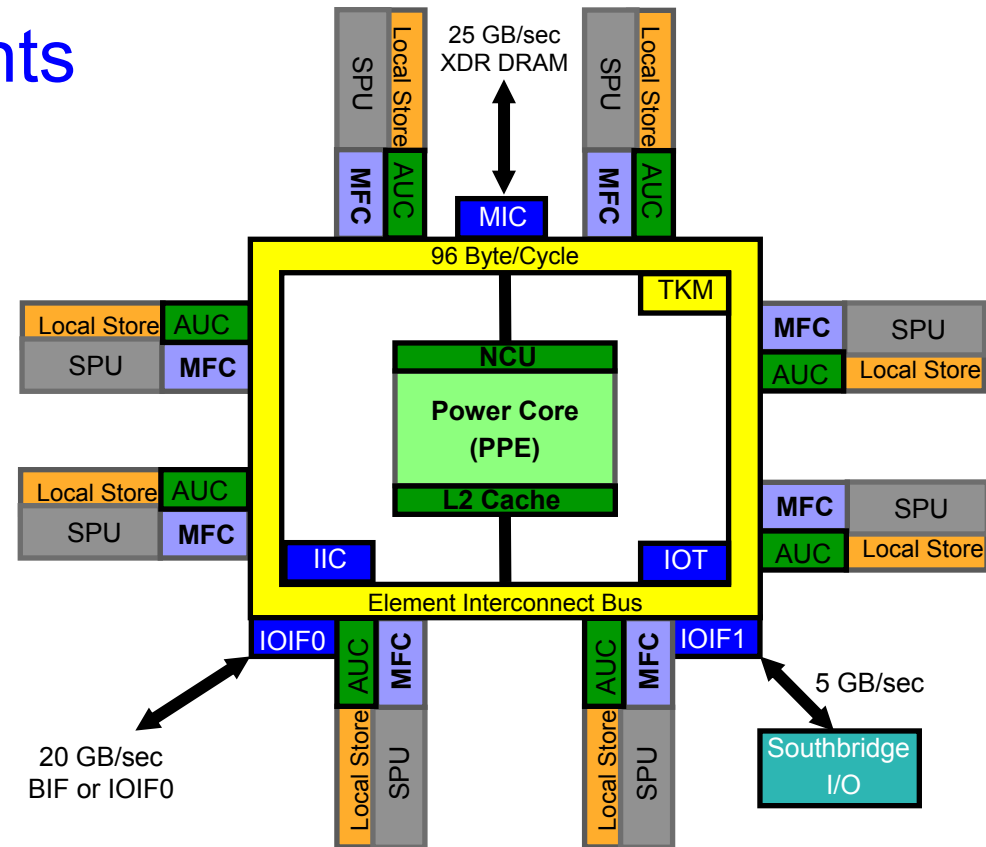
I/O and Memory Interfaces

- **I/O Provides wide bandwidth**
  - Dual XDR^{TM} controller (25.6GB/s @ 3.2Gbps)
  - Two configurable interfaces (76.8GB/s @6.4Gbps)
    - Configurable number of Bytes
    - Coherent or I/O Protection
  - Allows for multiple system configurations
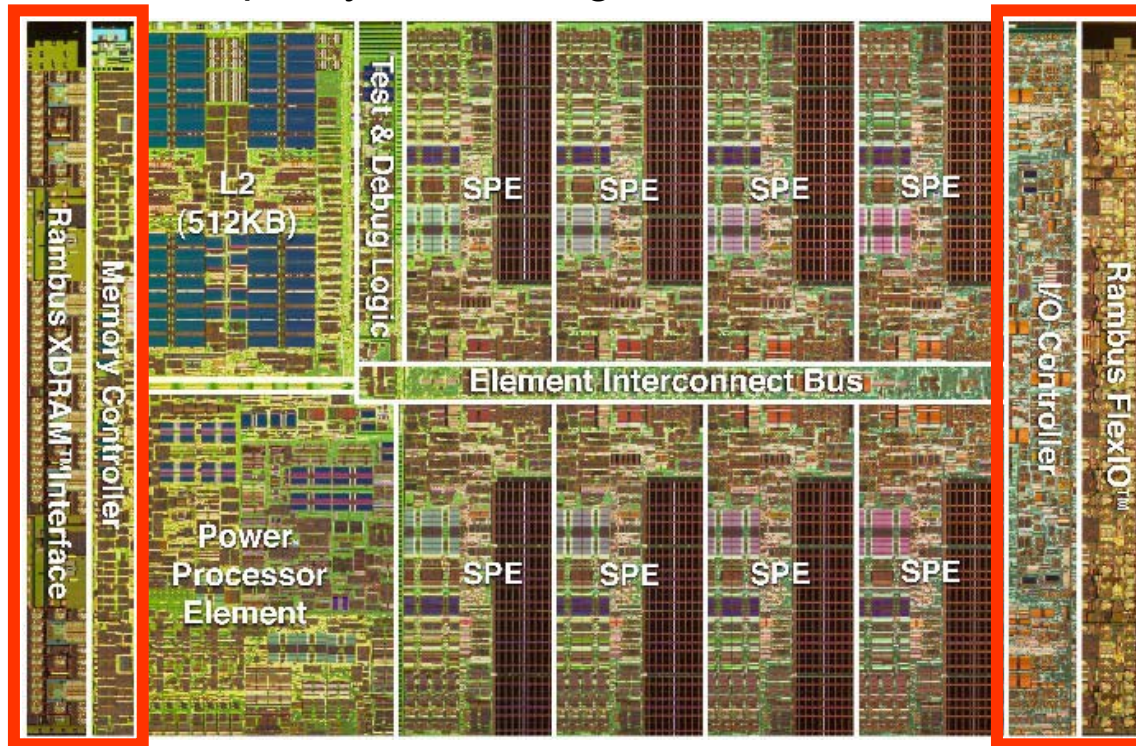
Cell Processor Components

Broadband Interface Controller (BIC):

- Provides a wide connection to external devices
- Two configurable interfaces (60GB/s @ 5Gbps)
  - Configurable number of bytes
  - Coherent (BIF) and / or
    I/O (IOIFx) protocols
- Supports two virtual channels per interface
- Supports multiple system configurations

Memory Interface Controller (MIC):

- Dual XDR^{TM} controller (25.6GB/s @ 3.2Gbps)
- ECC support
- Suspend to DRAM support
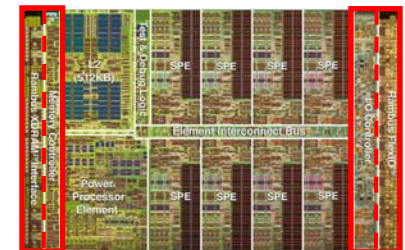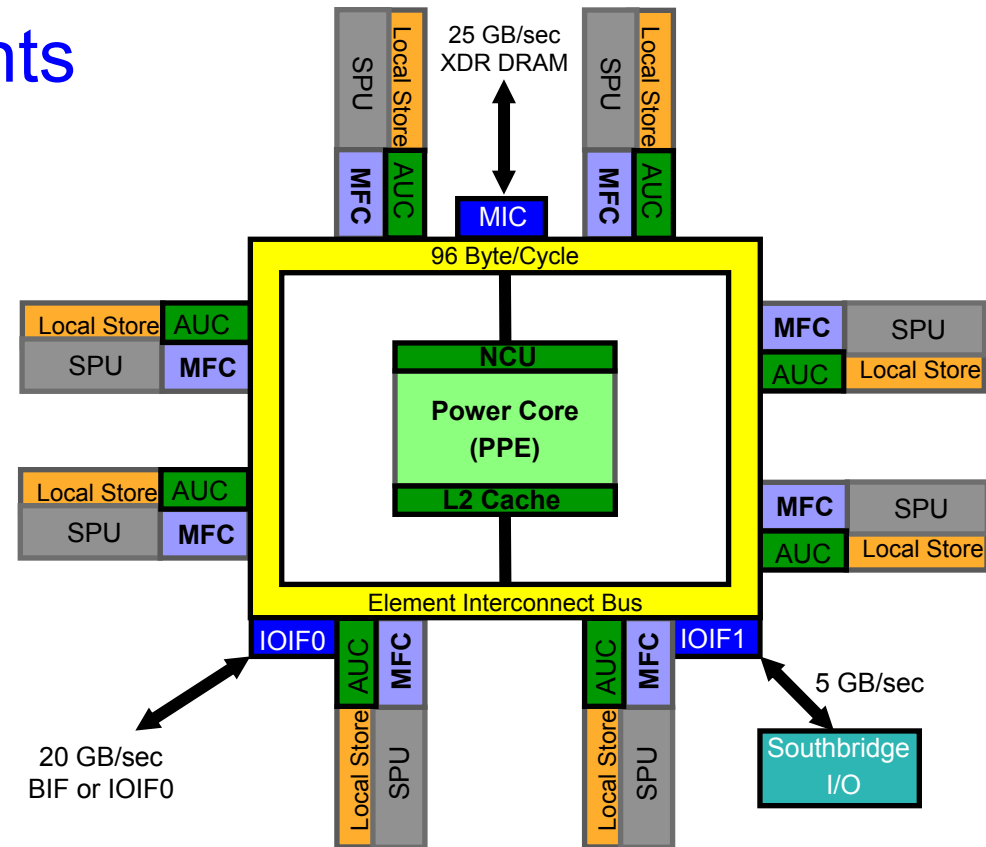
Cell BE Processor Can Support Many Systems

- Game console systems
- Blades
- HDTV
- Home media servers
- Supercomputers