Sun

microsystems

UltraSPARC III:

A 600 MHz 64-bit Superscalar Processor for 1000-Way Scalable Systems

Bill Lynch

Gary Lauterbach

Sun Microsystems, Inc.

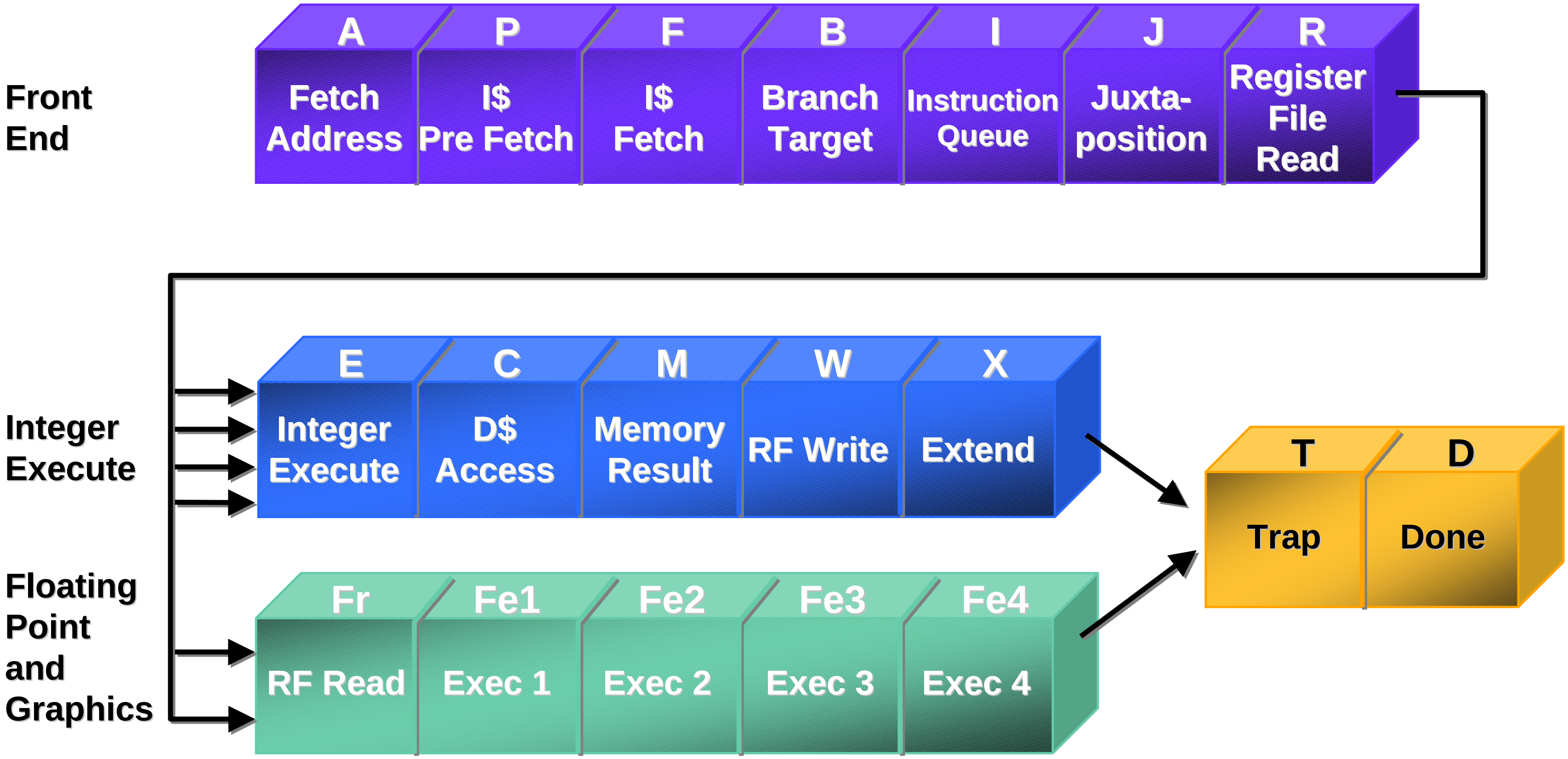
Presentation Overview

- UltraSPARC III overview
- Pipeline
- Memory System
- Scalability
  - Technology
  - Systems
  - Applications

UltraSPARC III Overview

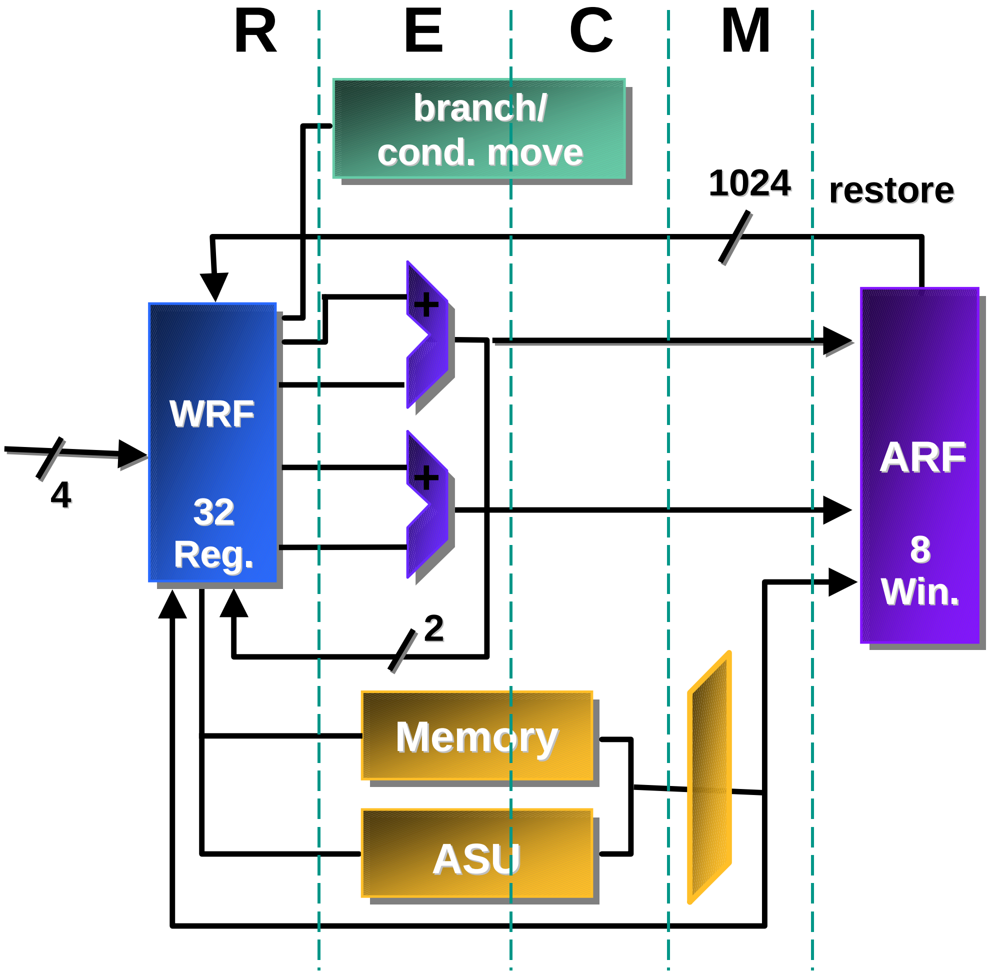
- .25 micron 6-layer metal process
- 330 mm^2
- 600 MHz
- 70W @ 1.8V
- 35 SPECint95, 60 SPECfp95
- Code schedule and system software compatible with UltraSPARC I, II

Pipeline Diagram



Integer Execute

- 4 integer issue
- Arithmetic/Special Unit (ASU)
- Pipelined predicated execution
- 7R3W port Register File



Floating Point/Graphics Execute

- 2 instruction issue from:
  - FP/graphics add
  - FP/graphics mul/div/sqrt
- Fully pipelined FP/graphics add/mul
- Concurrent FP div/sqrt unit
- 5 Read, 4 Write port register file

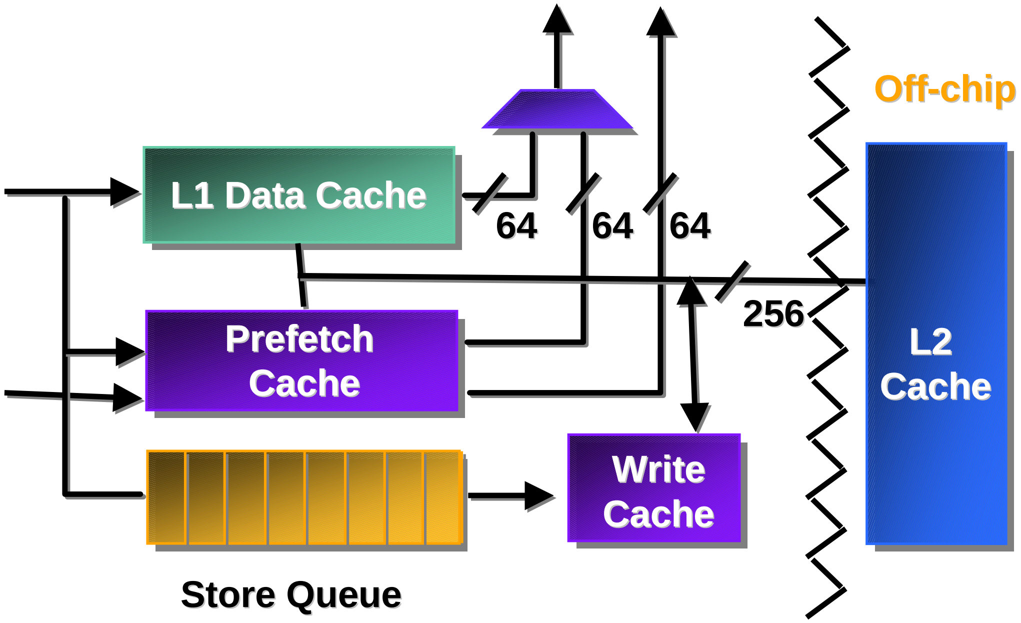
Instruction Latency

Instruction	Latency	
Graphics Add	4 cycles	
Graphic Multiply	4 cycles	
	Single	Double
FP Add	4 cycles	4 cycles
FP Multiply	4 cycle	4 cycle
FP Divide	17 cycle	20 cycle
FP Square Root	24 cycle	24 cycle

Memory Subsystem

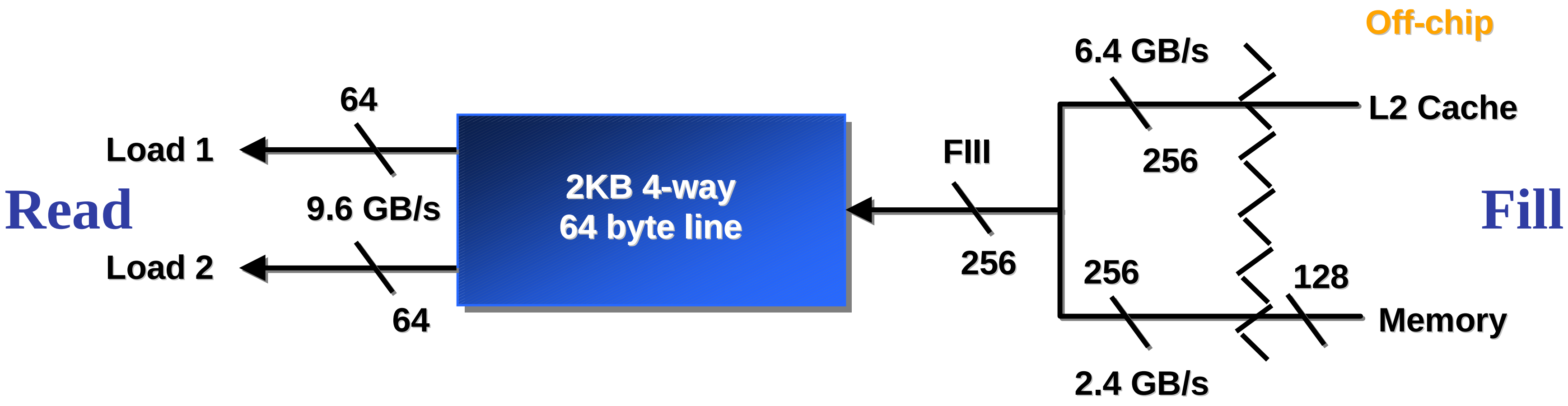
- **L1 data cache :**
  64KB 4-way associative
- **L1 prefetch cache:**
  2KB 4-way associative
- **L1 write cache:**
  2KB 4-way associative
- **L2 cache:**
  1-8 MB direct-mapped
- **8 entry RAW forwarding store queue**
- **On-chip memory controller**

Cache	Latency	Bandwidth
L1 Data	2 cycles	9.6 GB/s
L1 Prefetch	3 cycles	18.4 GB/s
L1 Write	1 cycles	13.6 GB/s
L2 External	12 cycles	6.4 GB/s



Prefetch Cache

- Concurrent fill and dual reads
- 8 outstanding software prefetches
- Autonomous hardware stride prefetch
- Fully coherent cache



Technology Scaling

- Wires
  - Significant fraction of critical paths
  - >1 cycle to cross chip
  - Heavy use of repeaters
- Architecture for technology scaling
  - “Special” instructions
  - Deep pipeline
  - Reasonable width

System Scaling

- On-chip memory controller
  - Bandwidth scales with added processors
  - Low latency
- On-chip snoop tags
- “No snoop” pages
- 15 outstanding transactions
- Tagged transactions
- SPARC V9 TSO compilant

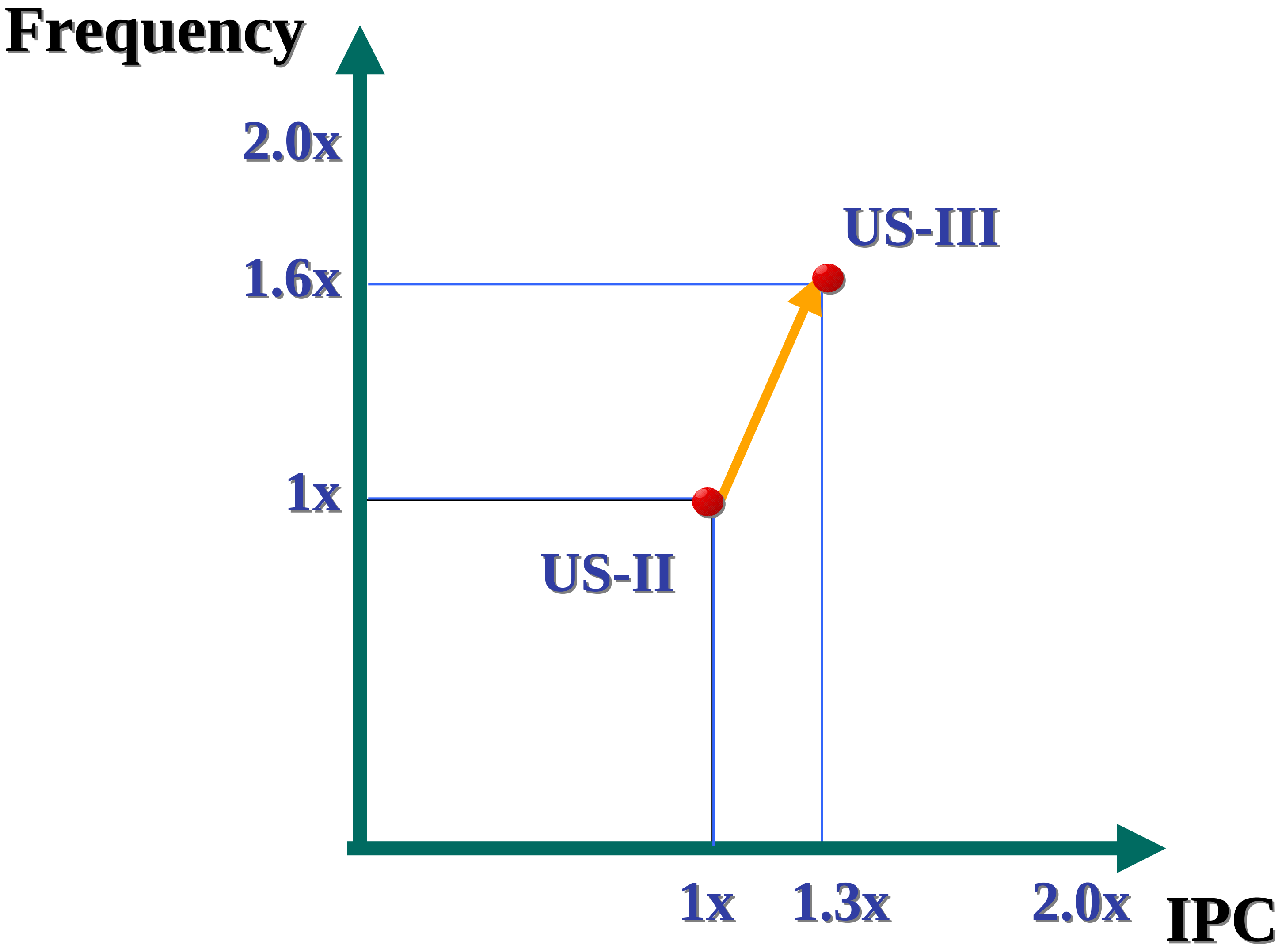
Application Scaling

- Increased frequency
  - Lower instruction latency
  - Higher instruction bandwidth
- Improved memory system
  - Lower average memory latency (in cycles!)
- Existing SPARC v9 binaries scale
  - Consistent scheduling with existing binaries

Increasing Performance

- Increase frequency
  - Works for all programs (SPEC, Oracle,
    CAD, etc.) if memory systems scales
- Increase IPC
  - Reduce execution time down to critical dependency path
  - More IPC requires speculation

Increasing Performance



Speculation

	+	-
Static, compiler	More instructions available Convert control to data dependencies	Static trace prediction
Dynamic, hardware	Dynamic trace prediction Dynamic memory disambiguation	Smaller “window”

Summary

- Scales with technology
  - Architecture and implementation attention to wires
- Scales system performance
  - On-chip memory controller
  - On-chip snoop tags
- Scales 64-bit application performance
  - Increased clock rate
  - Static speculation
  - Optimized memory system

Sun

microsystems