An Overview of the Alpha AXP™ 21164 Micro-Architecture

The World's Highest Performance Microprocessor

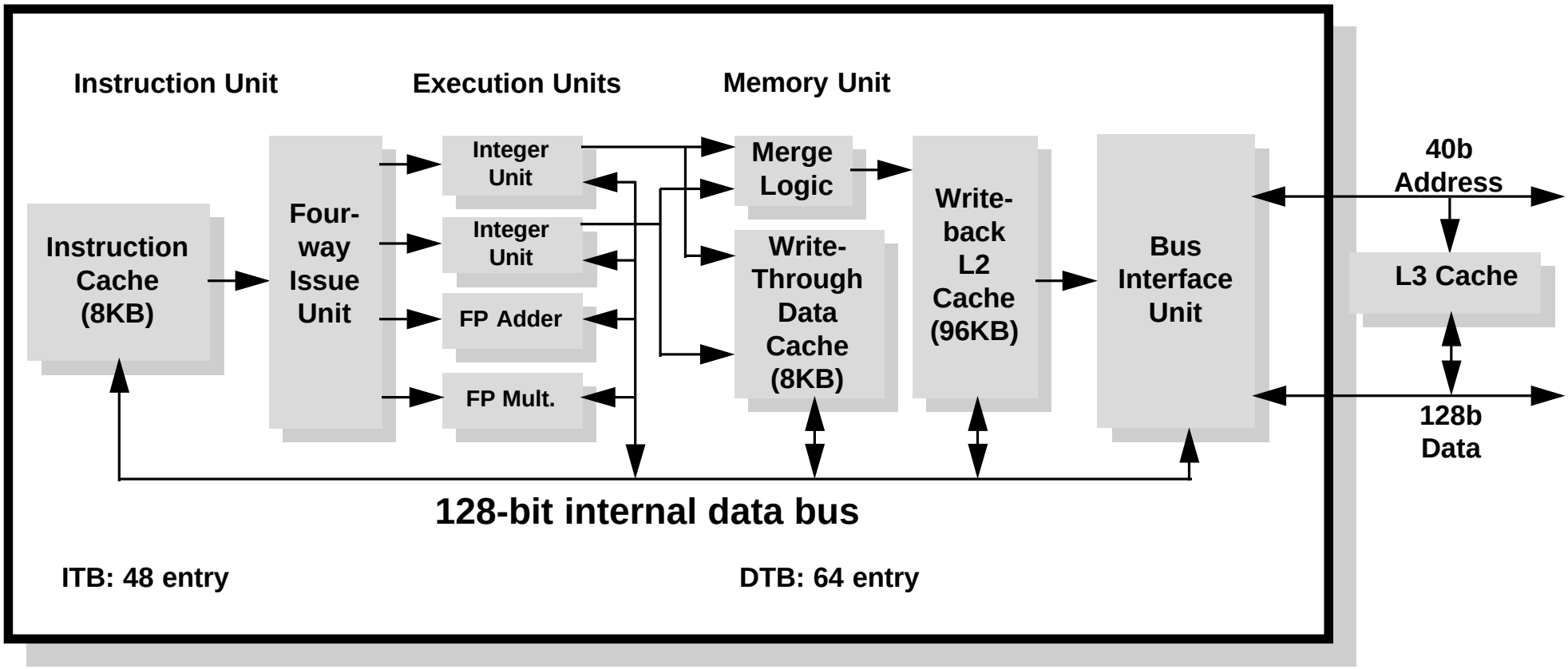
John Edmondson and Paul Rubinfeld
Digital Equipment Corporation
Semiconductor Engineering Group
Hudson, MA

Alpha AXP 21164 Overview

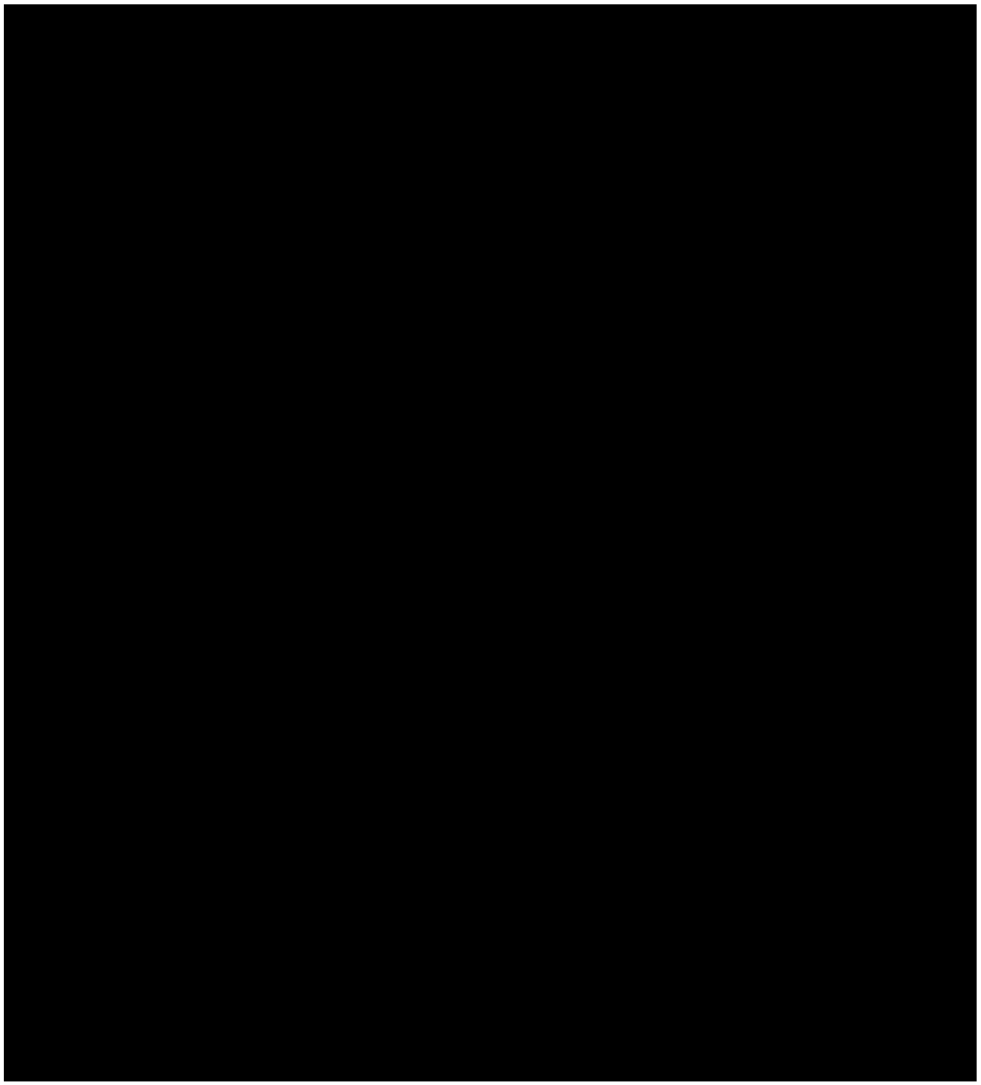
Key Attributes

- 4-way issue superscalar
- Large on-chip L2 cache
- 7-stage integer pipeline
- 9-stage floating point pipeline
- Emphasis on low latency at high clock rate
- High-throughput memory subsystem

Block Diagram

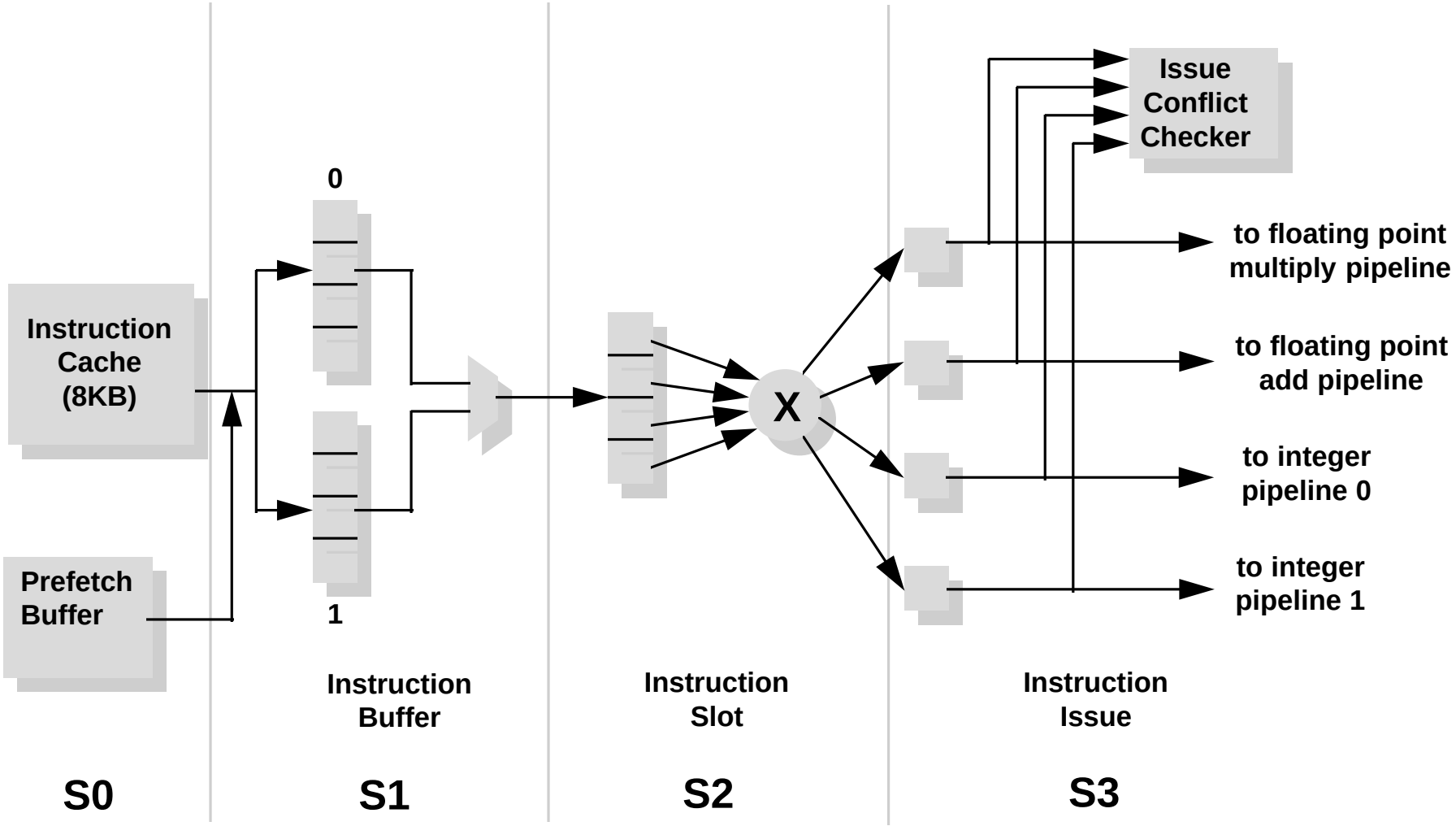


Photomicrograph



16.5 mm x 18.1 mm

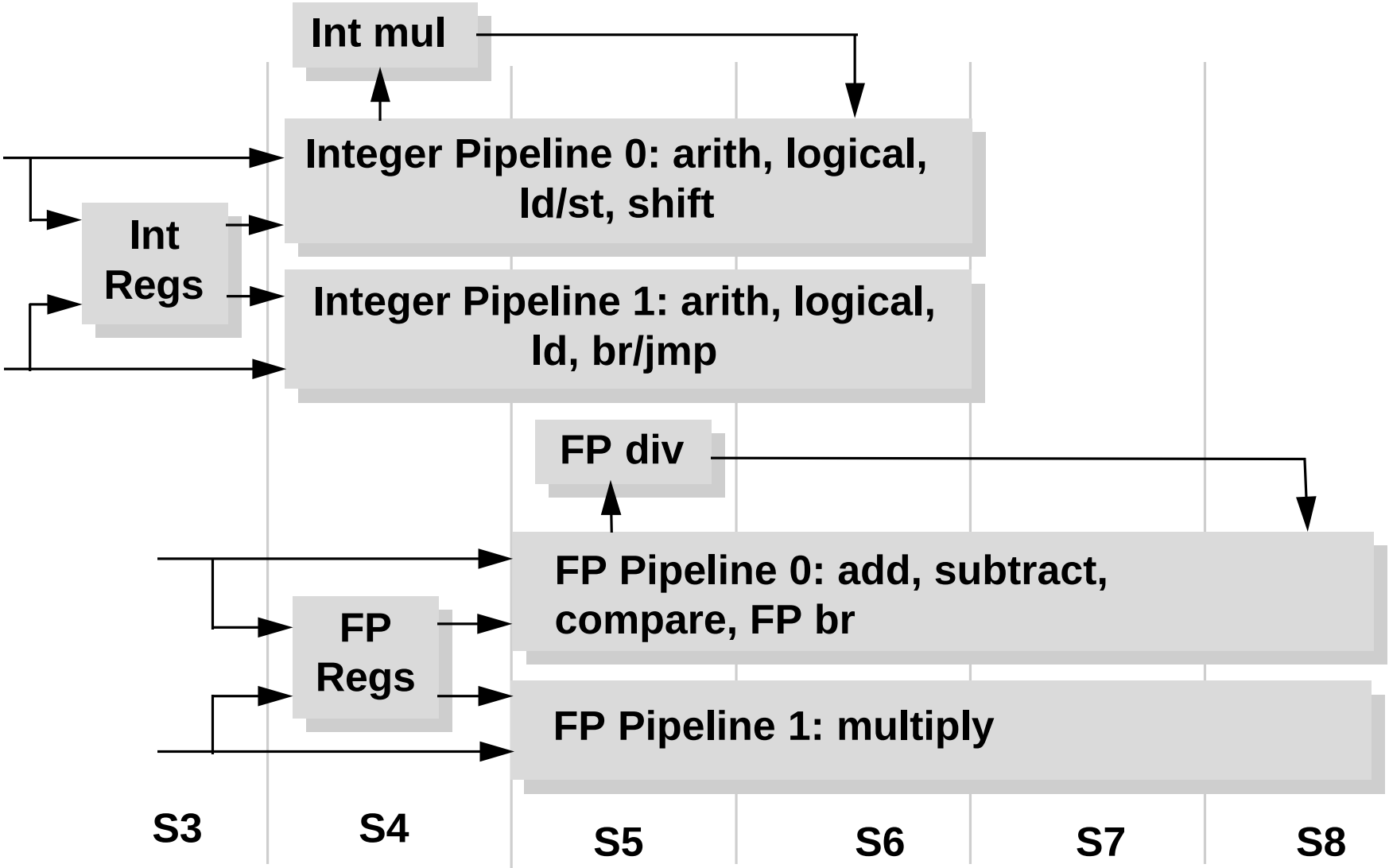
Instruction Issue Pipeline



Instruction Prefetching

- **Aggressive prefetching from L2 cache using high-bandwidth capability**
  - **At least three 32-byte blocks ahead of the current issue point**
  - **Continuous integer instruction issue possible out of L2 cache (2 per cycle)**
  - **60% of peak issue rate possible out of L2 cache (2.4 per cycle)**

Execution Pipeline

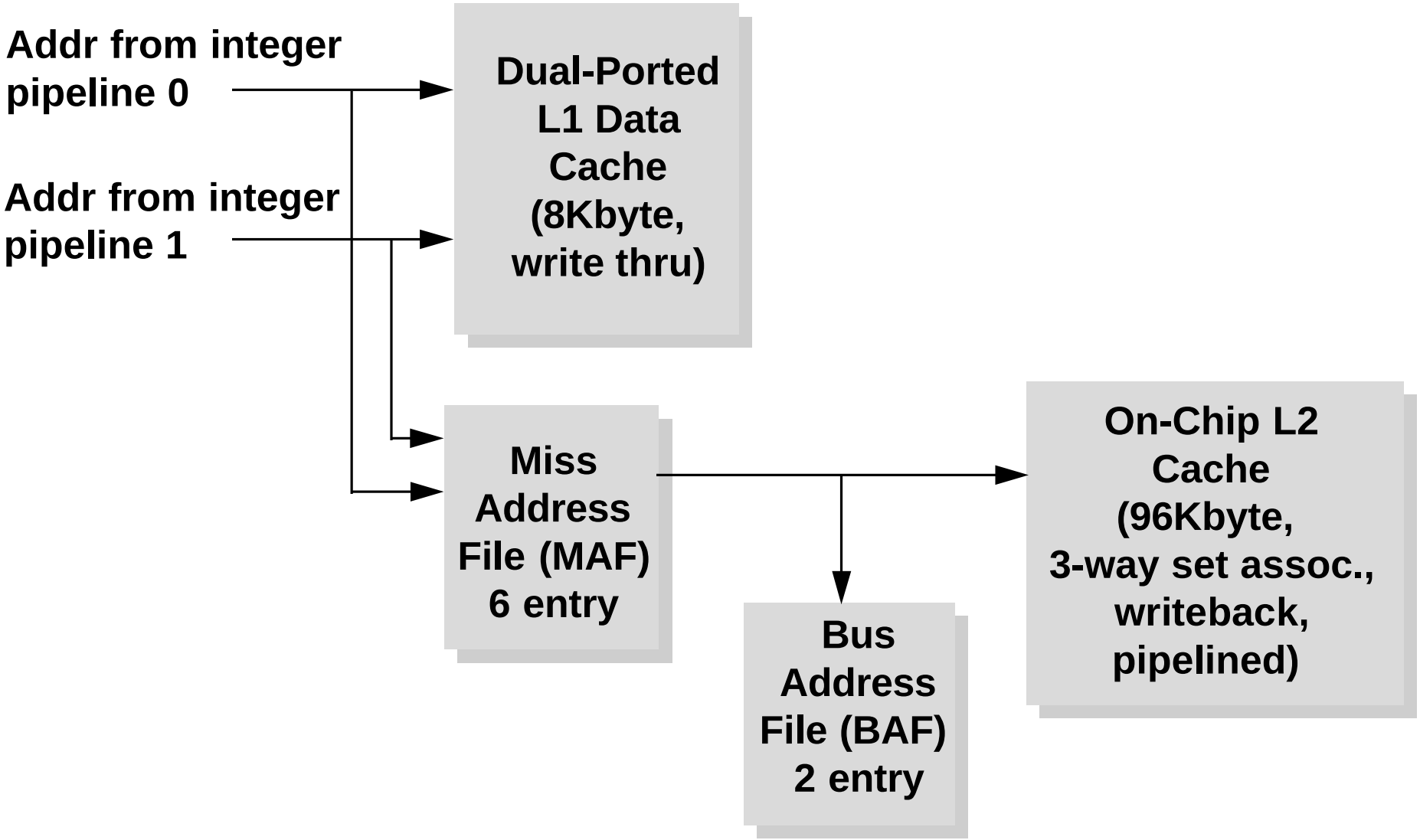


Instruction Latency

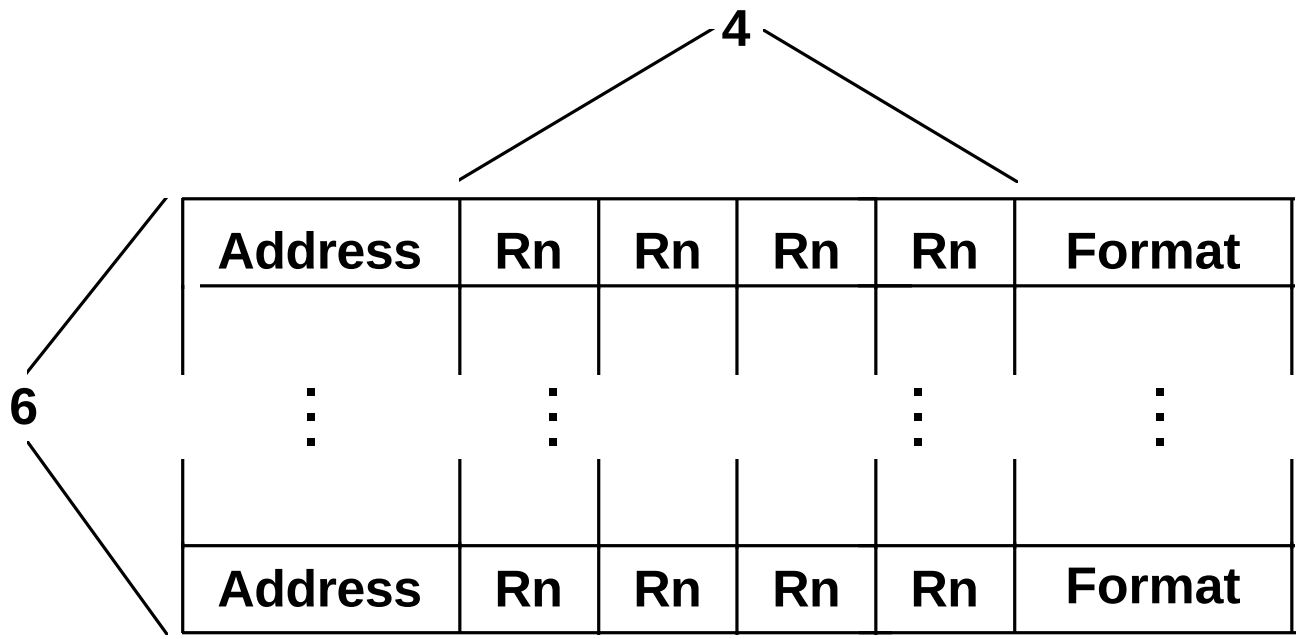
	Latency
Most integer ops	1
CMOV	2
Integer multiply	8-16
Floating point ops	4
Loads (L1 cache hit)	2
Special Case Bypass CMOV or conditional BR dependent on a compare or logical operation	0

Example: CMP R1, R2, R3
BEQ R3, LABEL

High-Throughput Load Execution

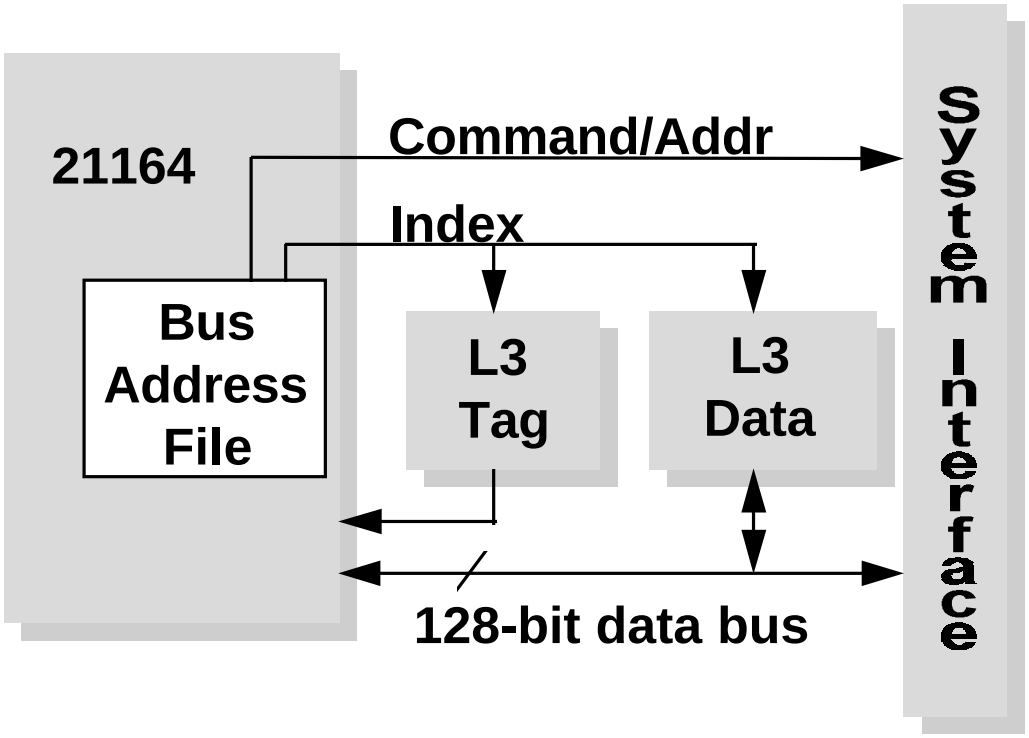


Miss Address File Details



- MAF merges loads to the same cache block
- Up to 21 loads
- Multiple loads merge, regardless of order
- Up to two register file fills per cycle

L3 Cache (off-chip)



- L3 cache is a direct-mapped writeback superset of on-chip L2 cache
- Up to 2 reads (or outstanding read commands) in L3 cache
- Programmable wave pipelining for L3 cache
- L3 cache is optional

Latency & Bandwidth of Memory Operations

	Latency (cycles)	*Bandwidth (bytes/cycle)*
L1 Data Cache	2	16
L2 Cache	8	16
L3 Cache	≥12	≤4

- L1 cache block size is 32 bytes
- L2 and L3 cache block sizes are each 64 bytes (with a 32-byte block size option)

Improvements Over the Previous Generation

- Reduced key latencies

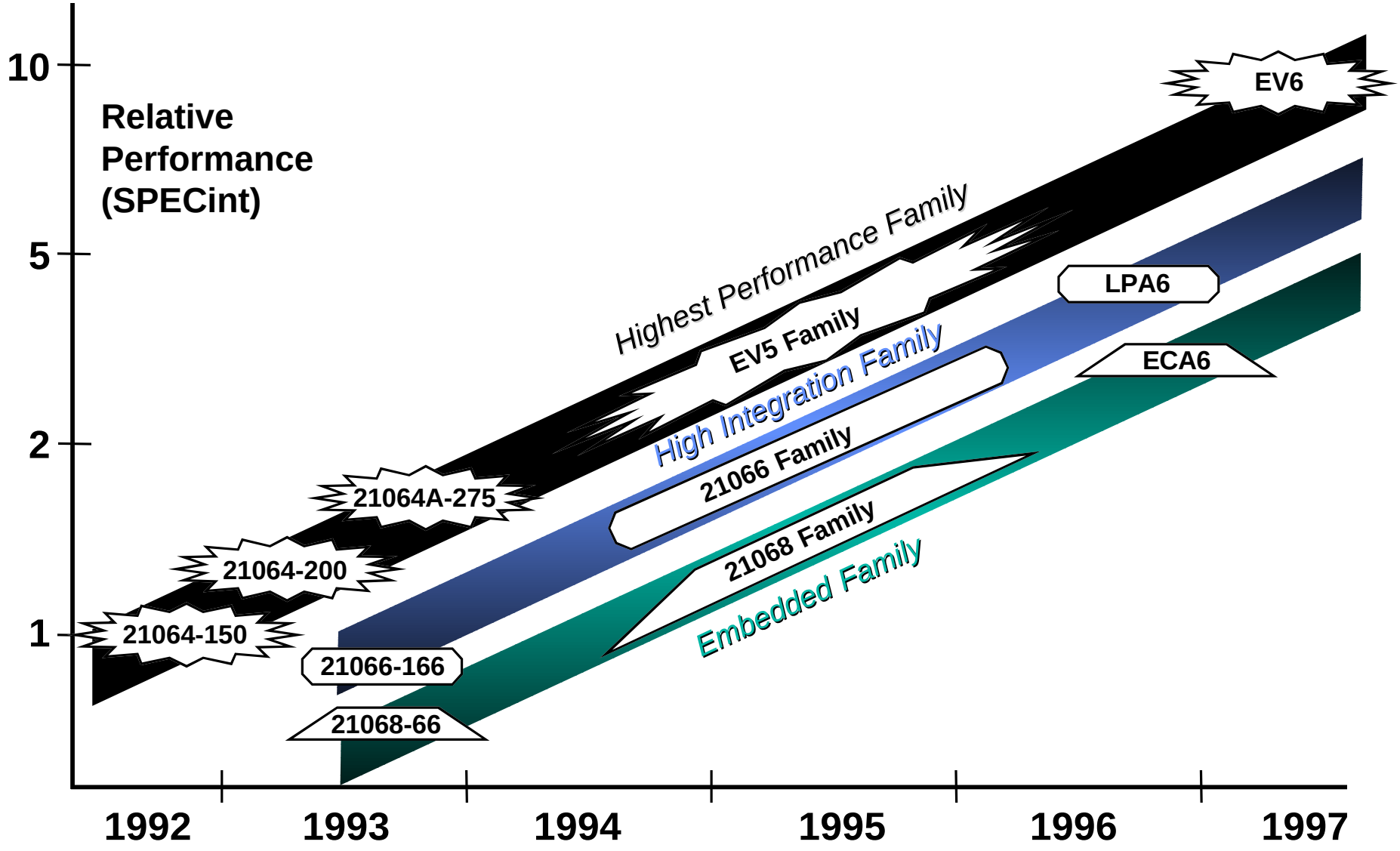
	21164	21064/21064A
Shift/byte ops	1	2
Integer multiply	8-16	19-23
CMP → BR	0	1
FP latency	4	6
L1 data cache	2	3

- Wider issue rate
  - 4 vs. 2
- Cycle time improvement
  - Greater than simple technology scaling

Estimated Performance Results

- **Better than 1 SPECint92 per MHz**
- **Better than 1.5 SPECfp92 per MHz**
- **Better than 2 TPS per MHz**

Alpha AXP Processor Roadmap



Summary

- **The Alpha AXP 21164 is totally new design**
  - **Quad instruction issue**
  - **On-chip secondary cache**
  - **Achieves short latency at a high speed clock**
- **It contains significant micro-architecture and circuit advances over the first implementation**
- **This chip is the world's highest performance microprocessor**