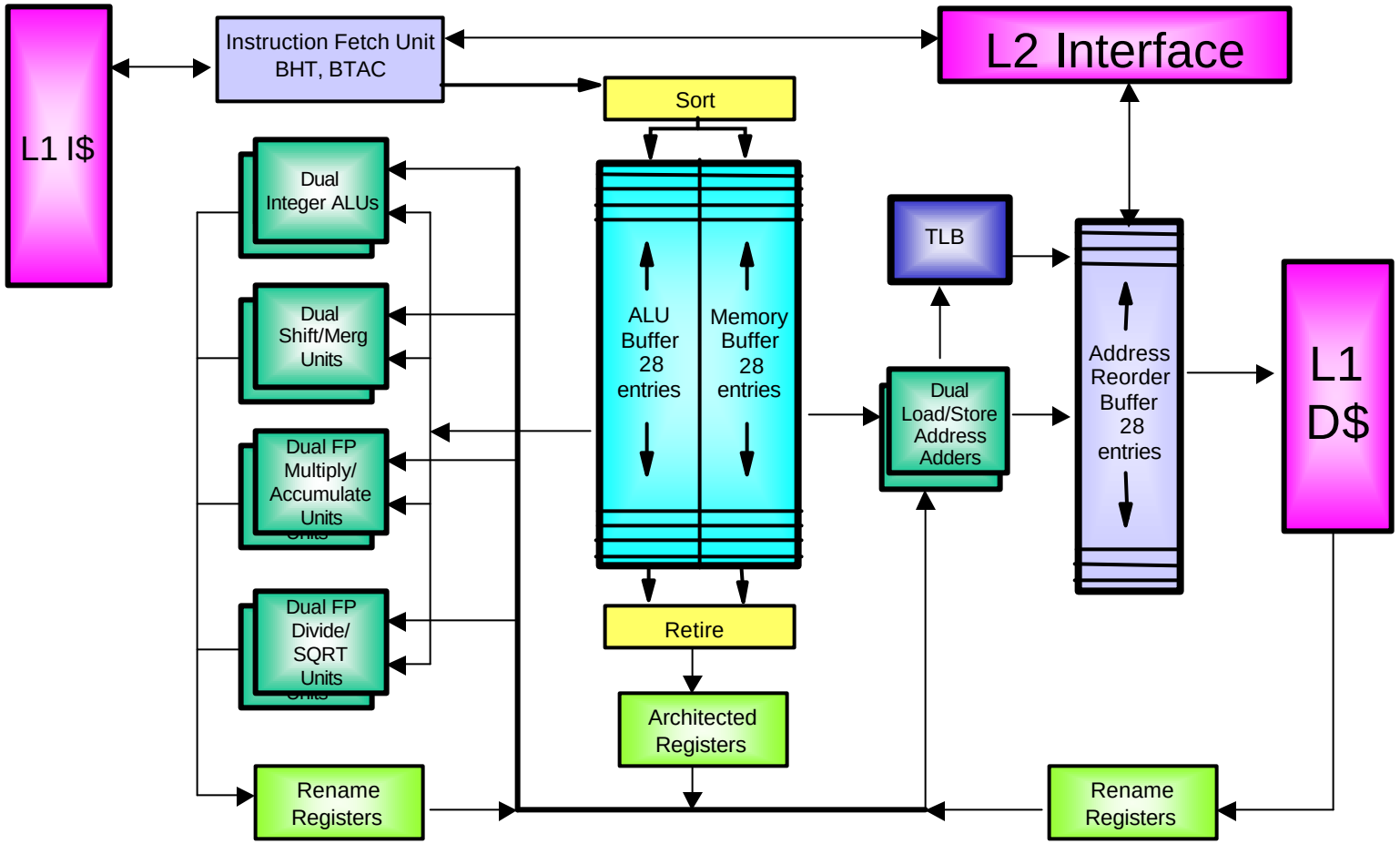
HP's Mako Processor

David J. C. Johnson
Fort Collins Microprocessor Lab
October 16, 2001

The Next Generation of PA-RISC

- Two entire PA-8700 CPU cores
- Improved branch prediction
- ¾ MB L1 D-cache per core
- ¾ MB L1 I-cache per core
- 32 MB off-chip DRAM Level-2 cache
- New high-bandwidth system bus
- .13um SOI, copper, low-k dielectrics
- 8 metal layers & local interconnect
- 300 million FETs, 25 million logic
- 1 GHz clock frequency
- Die size of 23.6mm X 15.5mm

Processor Core



Dual Cores

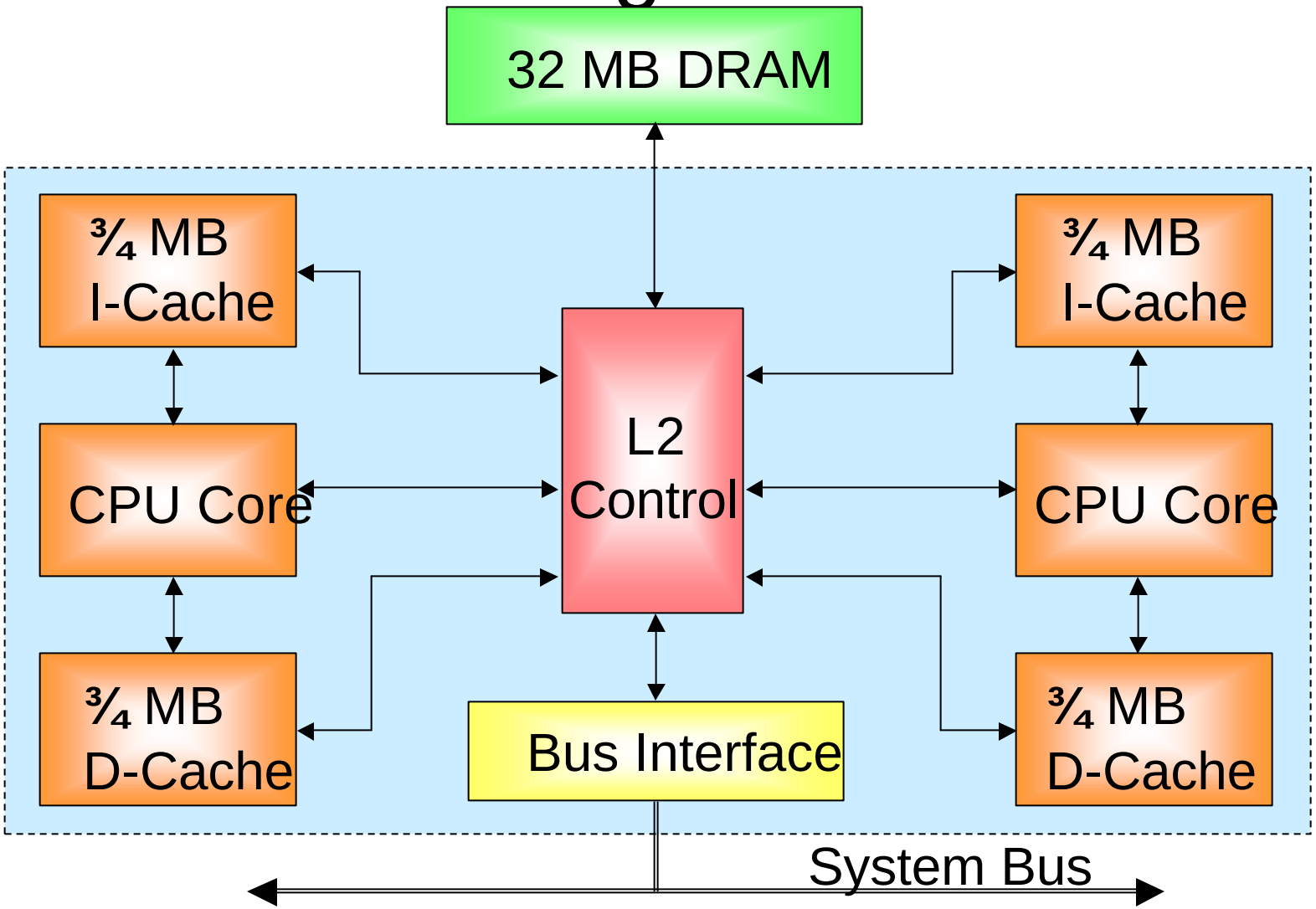
Physical

- Two exact copies of a single core
- Re-use of PA-8700 core helps schedule
- Scale long routes by repeater insertion
- Extra metal layers for VDD/GND
- Micro-architectural changes made for power reduction
- Leakage – use high-Vt FETs in caches

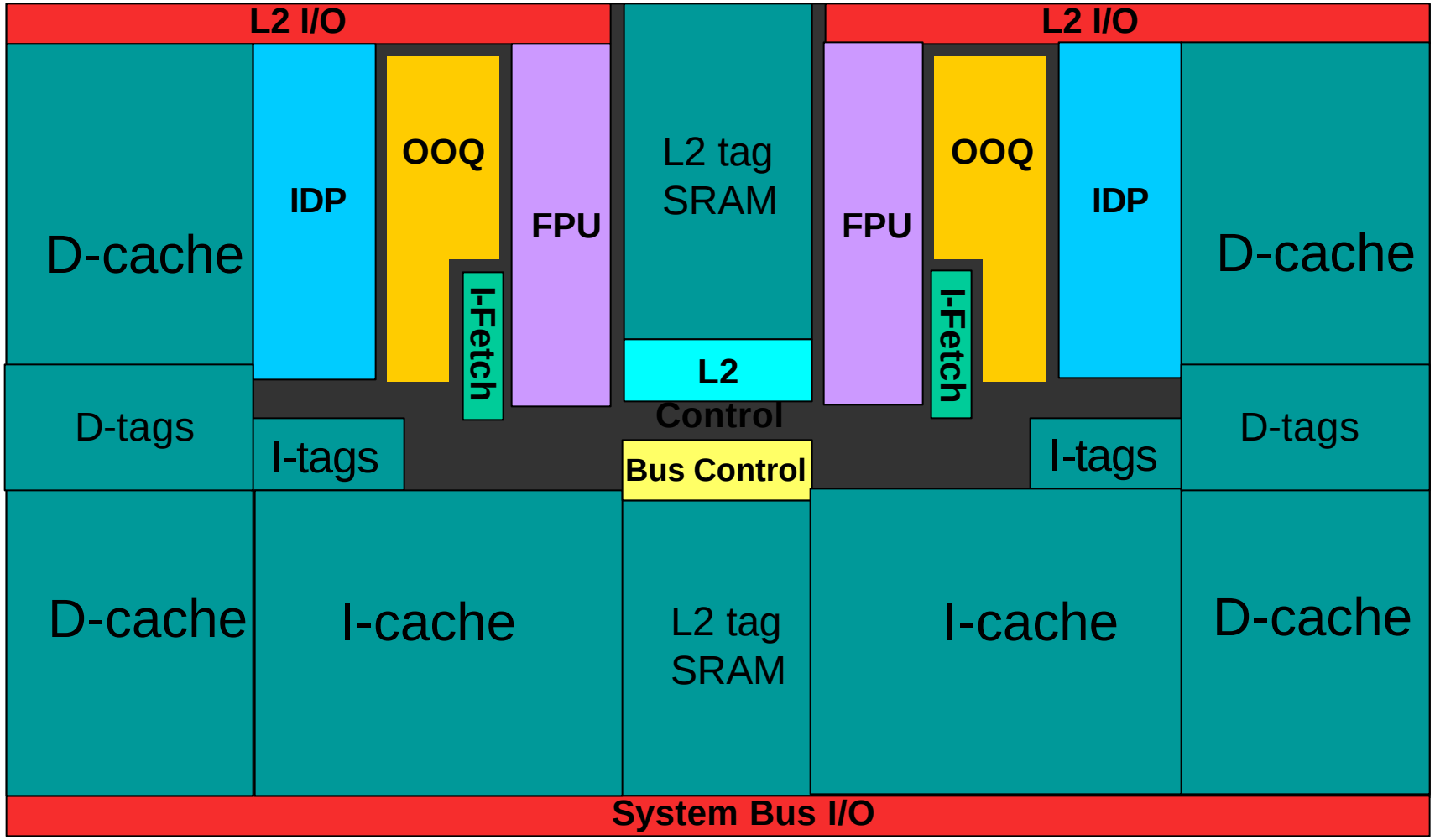
Functional

- Enables 2X MP in the same box
- Cores arbitrate between themselves for L2 cache access
- Coherency lookups in both of the L1 data caches and in L2 simultaneously
- No direct L1-to-L1 data passing

Chip Block Diagram

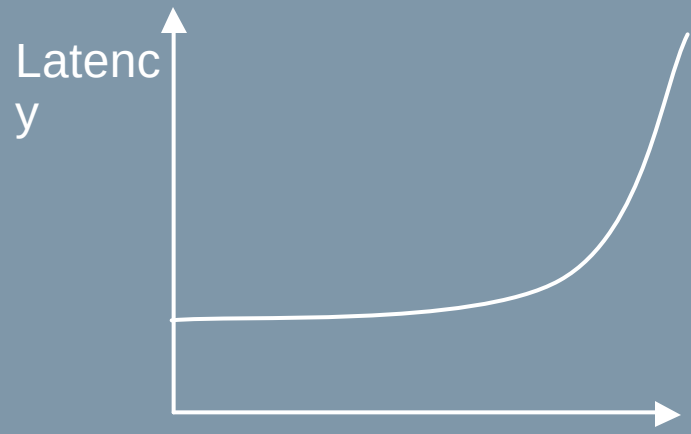


Floorplan



System Performance

Significant capacity increase needed to improve cache hit rate, which in turn reduces congestion in the memory system. This benefits all system transactions.



Commercial Applications

Latency intolerant, linked data structures, random access, shared data

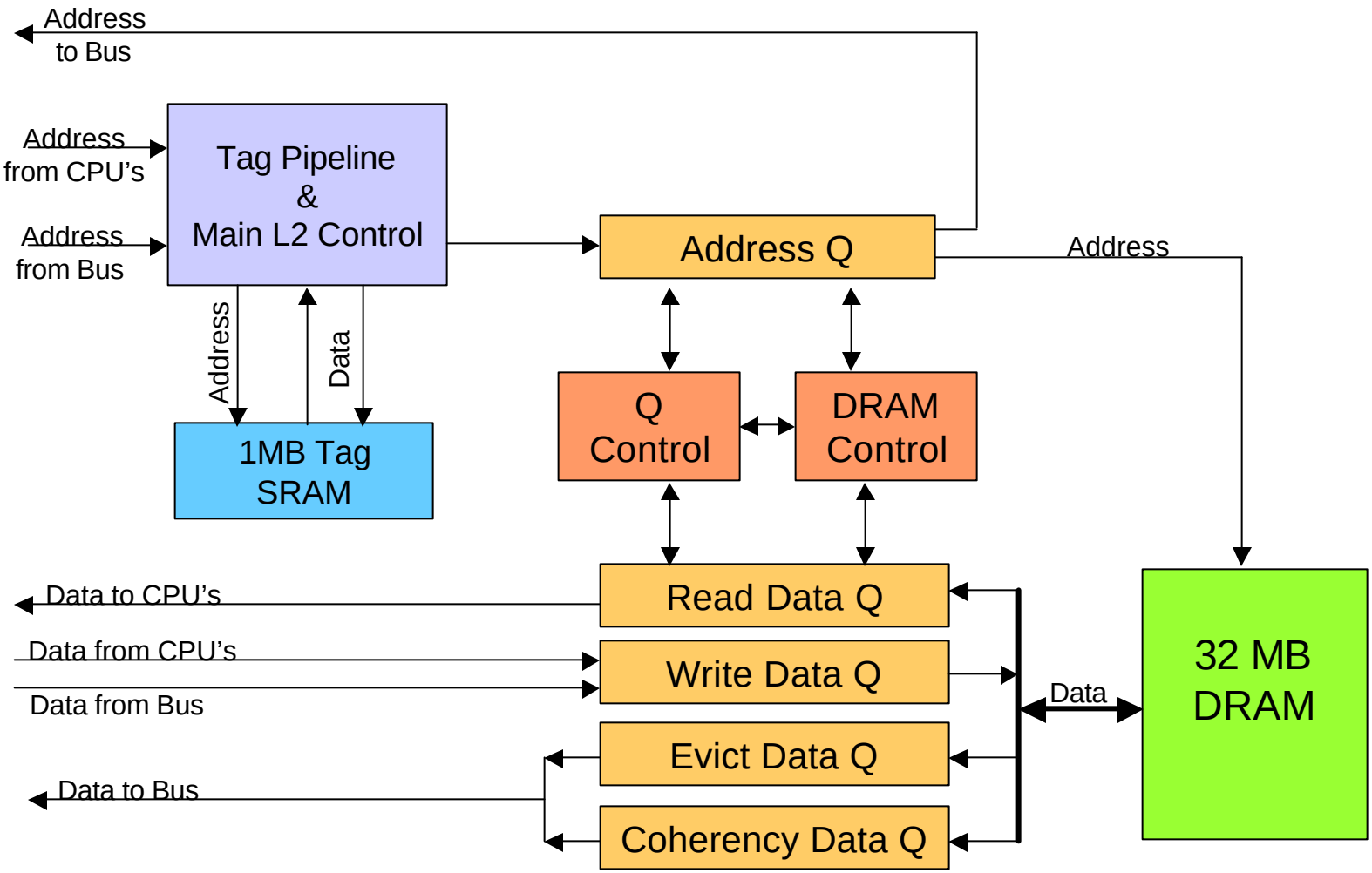
Scientific/MDA Applications

Latency tolerant, strided vector

L2 Cache Design

- 32 Megabytes of capacity
- 40 cycle latency
- 10+ GB/s bandwidth
- Data stored off-chip in fast DRAM
- 1 MB of tags on-chip in SRAM
- ECC coverage on tags and data
- 4-way associative, with tag access done before data in order to minimize # of IO pins
- Shared between the 2 cores
- Unified instruction and data
- Line size is 128 bytes
- Physically indexed and tagged
- No inclusion, exclusion guaranteed on private lines

L2 Cache Block Diagram



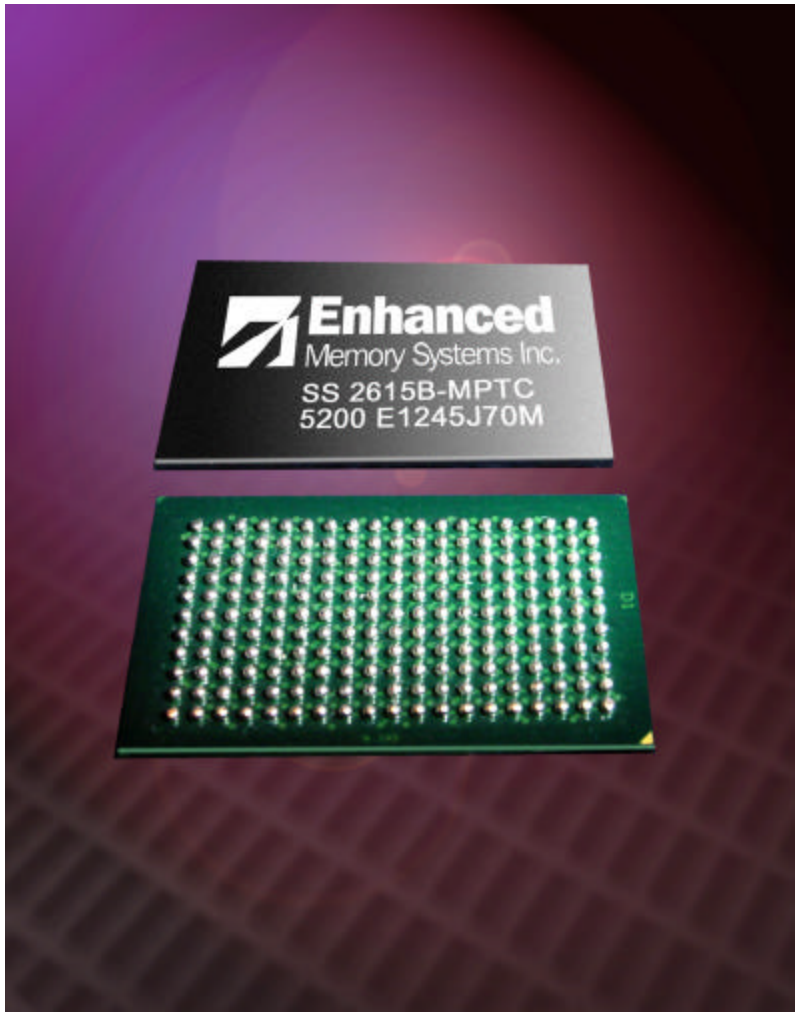
Cache Technology

- DRAM solves lots of problems
  - area, power, cost, capacity, IO speed
- Low-latency is critical, bandwidth alone is insufficient
- No significant increase in coherency traffic even with large L2
- Chip and board area are both expensive
- Use the right technology for the job! Don't try to solve all problems with a high-speed logic process.

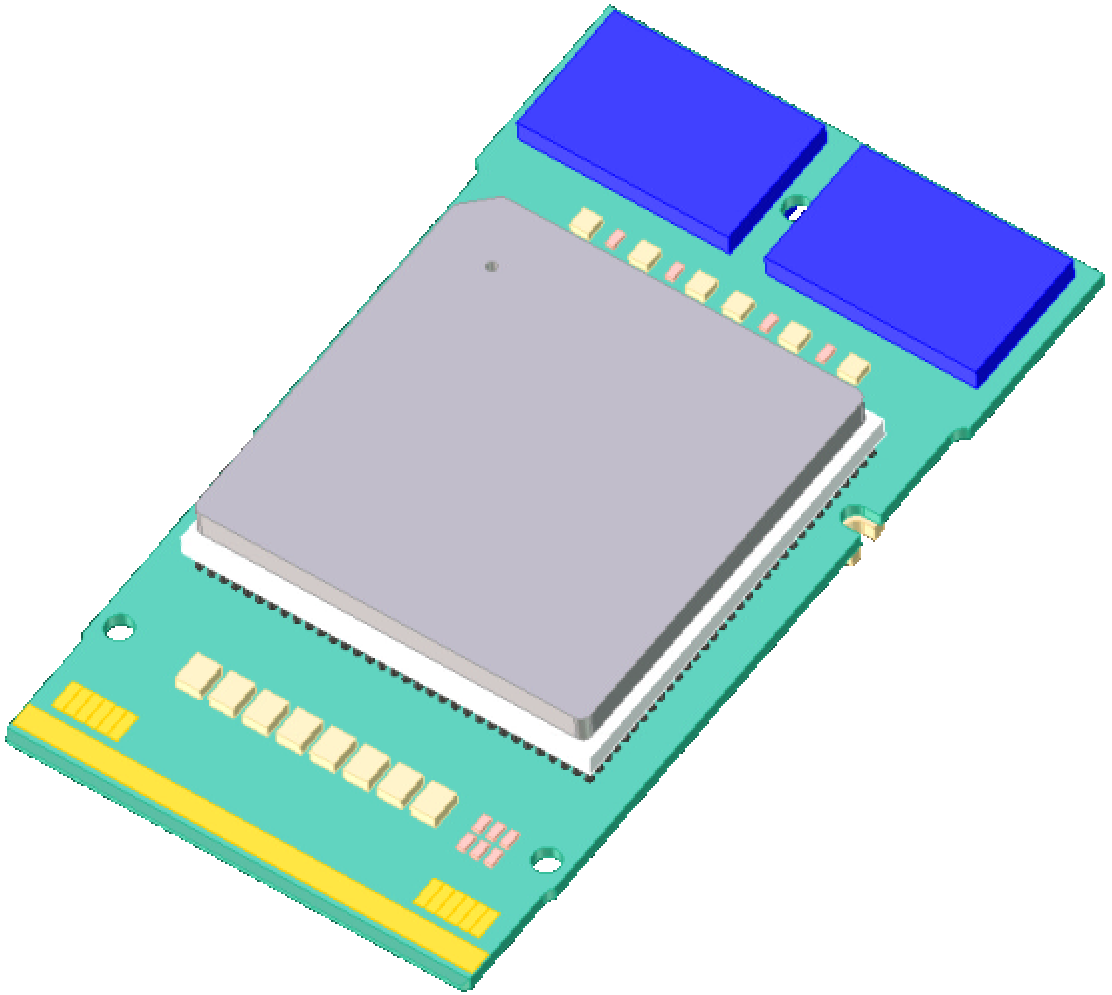
L2 DRAM Chip

- 72 Mbit DDR ESRAM organized 2Mx36
- 13.3ns latency
- Broadside addressing
- Coherent late write feature
- 300 MHz Clock, 600 Mbps data rate/pin
- 2.7 GB/sec Bandwidth
- 1.2V Source Terminated HSTL I/O
- 2.5V Power supply
- Under 2 watts power dissipation
- Standard 209-pin PBGA

Enhanced Memory Systems Inc.



Processor Module



System Bus

- Compatible with McKinley's Itanium Processor Family (IPF) bus
- Compatibility provides easy transition from PA-RISC to IPF
- 200 MHz for 5-drop bus
- Separate address and data buses
- 128-bit, double pumped, source synchronous data bus
  - 6.4 gigabytes/sec @ 200MHz
  - 128 byte lines only
- ECC on the data bus, parity protection on other signals
- Supports pipelined transactions and out of order completion

Processor Performance

- Modern process technology enables massive compute power on a single chip
- Caching and bus improvements are necessary to utilize computational units effectively and minimize system congestion
- Optimum results achieved by appropriate use of available circuit technologies
- Balanced design delivers high performance. SPEC2000 int/fp estimates per core are 900 and 1000.

Summary

- Hewlett-Packard will continue to support our PA-RISC customers
- Hewlett-Packard provides a clear path forward into IPF with commonality in system designs
- The Mako CPU chip will be used across the full range of HP’s server products and is expected to be available in workstations as well