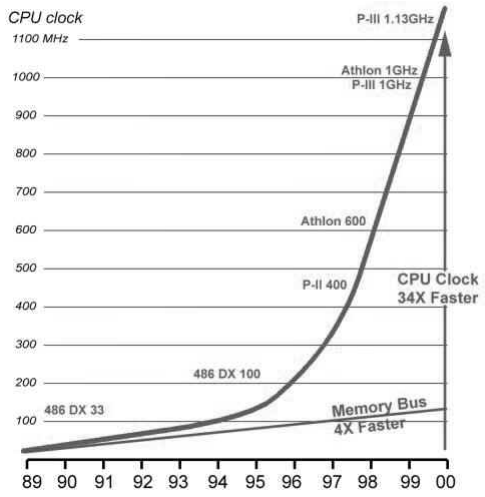
The Big Iron Shootout

- ILP is running out.
- Memory is Too Slow
- Possible Solutions
  - 000 Super-Scalar (21364, Sparc, MIPS?)
  - SMT (Compaq 21464)
  - VLIW (Intel IA64)
  - CMP (IBM Power4, Compaq Piranha)

Memory Is Paramount



000 Super Scalar

- 21364 will be 6 wide
- Pros
  - Tried and True
- Cons
  - Don't scale well
  - lack of IPC

SMT

- 21464 will be 4 context, 6 wide
- Pros
  - Dynamic sharing is a big win.
  - Simple extension of super-scalar.
- Cons
  - Complicated.
  - Non-scalable.

(VLIW) IA64

- Assumes ILP is plentiful.
- Make the compiler find it.
- Pros
  - Very simple pipeline (potentially)
  - Fast clock (potentially)
- Cons
  - May need bells and whistles for performance
  - Itanium is extremely complex
  - Schedule is fixed
  - Needs profiling
  - Different Architecture

Chip Multiprocessor

- Put multiple cores on one die.
- Run lots of threads.
- Pros
  - Simple extension of current technology.
  - No architectural changes.
  - Scales well.
  - Better-than-SMP performance (faster interconnect.).
- Cons
  - Static sharing.

IBM Power4

- Next generation Power implementation.
- Features
  - 2 8 wide 000 cores per chip.
  - 4 chips per package.
  - 3-level integrated cache.
  - Super-fast intra- and inter-chip interconnect.

IBM Power4 (cont.)

- Vital Statistics and Trivia
  - 174 million transistors.
    - Each core is 30 million (3x Pentium III)
  - 5500 pins per chip.
  - 700 lbs to install.
  - Should top out at 2Ghz

Photo Gallery

wafer, giving it roughly 25% faster logic than would be expected from the same process on a bulk substrate.

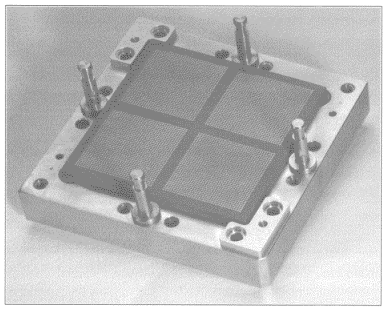


Figure 1. Four POWER4 chips packaged in a glass-ceramic MCM provide the complete processor complex for an eight-way SMP system. Each MCM is 4.5 inches on a side, has 5,200 I/O pads, and dissipates about a half kilowatt of power.

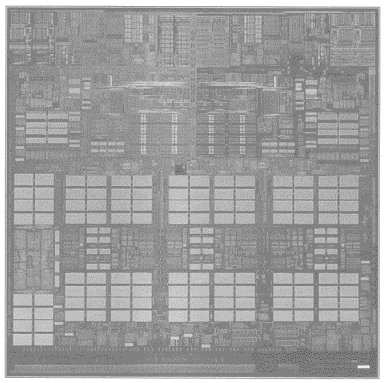


Figure 2. Each POWER4 chip implements 174 million transistors, which occupy about 400 mm^2 in IBM's seven-layer-copper 0.18-micron CMOS-8S2SOI process.

Photo Gallery (Cont.)

tor count, we estimate that each CPU core (including L1

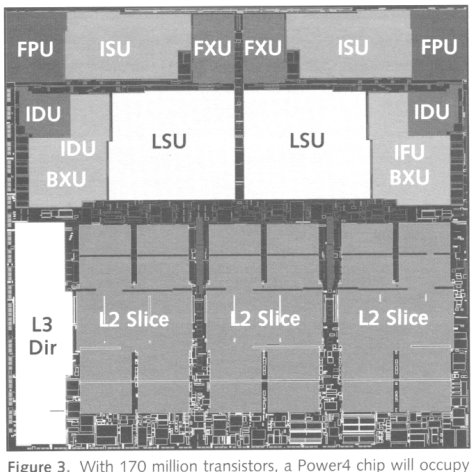


Figure 3. With 170 million transistors, a Power4 chip will occupy

Memory System

- Per-core 32kb L1 data
  - 3 ports, 2 read, 1 write
  - 8 outstanding misses
- Per-core 64k L1 instruction
  - 3 outstanding misses
- Shared, 1.5MB/chip partitioned L2
  - 3 independent "slices"
  - 8-way set associative
  - 100 GB/s to L1

Memory (cont.)

- 128MB (32MB/chip) L3 cache
  - Coherency data is on chip.
  - Actual L3 cache is a custom memory device (may or may not be in the package)
  - 16 GB/sec
- On chip memory controller.

Closer look at the L2

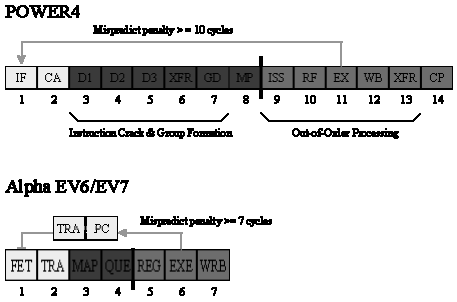
- L2 is shared across the module.
  - Glueless interconnect
  - 16 Byte busses
  - 500Mhz, 35 GB/sec total bandwidth
  - Data moves to last use.

MCM Interconnect

- "Elastic" I/O
  - wave pipelined interconnect busses
  - multiple bits in flight at once.
  - Provides 32GB/s between modules
  - Should scale to 4 MCM (32 cores)

The Pipeline

- 14 Stage pipeline
- 10 cycle misprediction penalty



Front End

- Fetch
  - 8 inst/cycle
- Big Branch Predictors
  - 16k/table entry hybrid predictor.
  - Much larger that 21264.

Instruction Scheduling

- 6 stages to schedule
- Cracking (renaming?)
- Group Formation
  - Combines instructions into groups (a la VLIW)
  - Groups issue at once.
  - Groups commit at once (sort of).
  - Significantly reduces implementation complexity.

Bottom Line

- Power4 Philosophy
  - Current ISA's are fine.
  - Thread level parallelism is abundant.
  - Memory performance is important.
  - Parts of VLIW are good
  - We are IBM, so, darn it, we can do it.

Compaq Piranha

- Advanced Research Prototype
  - built by Compaq Corporate Research and NonStop Hardware.
  - Standard cell ASIC design (mostly).
  - 500Mhz (slow because of ASIC)
- 8 Alpha cores/chip
  - Single issue, in order.
- Should scale to about 100's or 1000's of cores.

Piranha Block Diagram

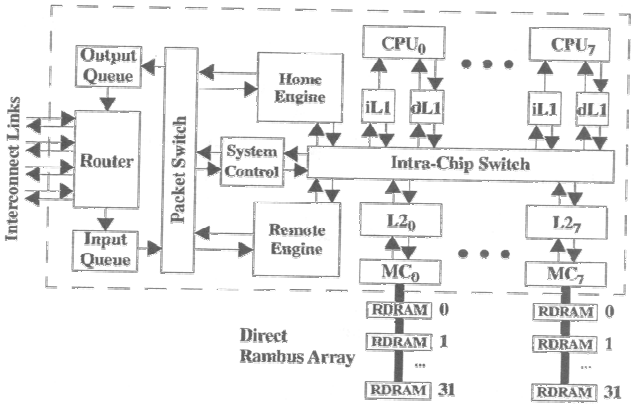


Figure 1. Block diagram of a single-chip Piranha processing node.

Memory Subsystem

- Per-core 64KB Blocking Data/Inst caches.
- Shared L2 1MB cache.
  - No data inclusion (would waste space).
  - L2 maintains duplicate L1 tags.
  - L2 is a 1MB victim cache.
  - Coherence is at L1 level (including I-caches)

Memory Subsystem(cont.)

- Two protocol engines maintain coherence
  - Home engine for on chip data
  - remote engine for off chip data
  - Engines are programmable.
- High performance Coherence protocol
  - Invalidation based directory protocol
  - Cruise missile invalidation

Main Memory

- Direct Rambus
  - Few pins
- One channel per core.
- 2Gigs max memory.

The Pipeline

- High performance in order processor.
- 5 stage in-order single issue.
- Features
  - Branch target buffer
  - Full forwarding
  - Big TLBs (256 entries, 4-way)

Performance

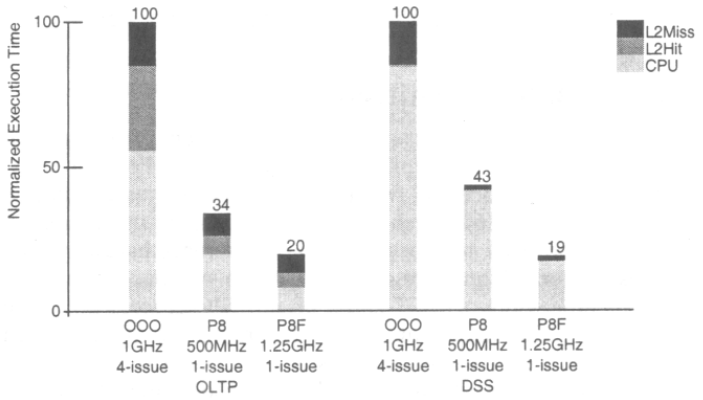


Figure 8. Performance potential of a full-custom Piranha chip for OLTP and DSS.

Bottom Line

- Piranha philosophy
  - Research processor
  - Proof of concept
  - A "limit point" in the design space

Open Questions

- How many cores should there be per chip?
- What should they be (000, SMT, VLIW)?
- Are CMP's viable for workstations?