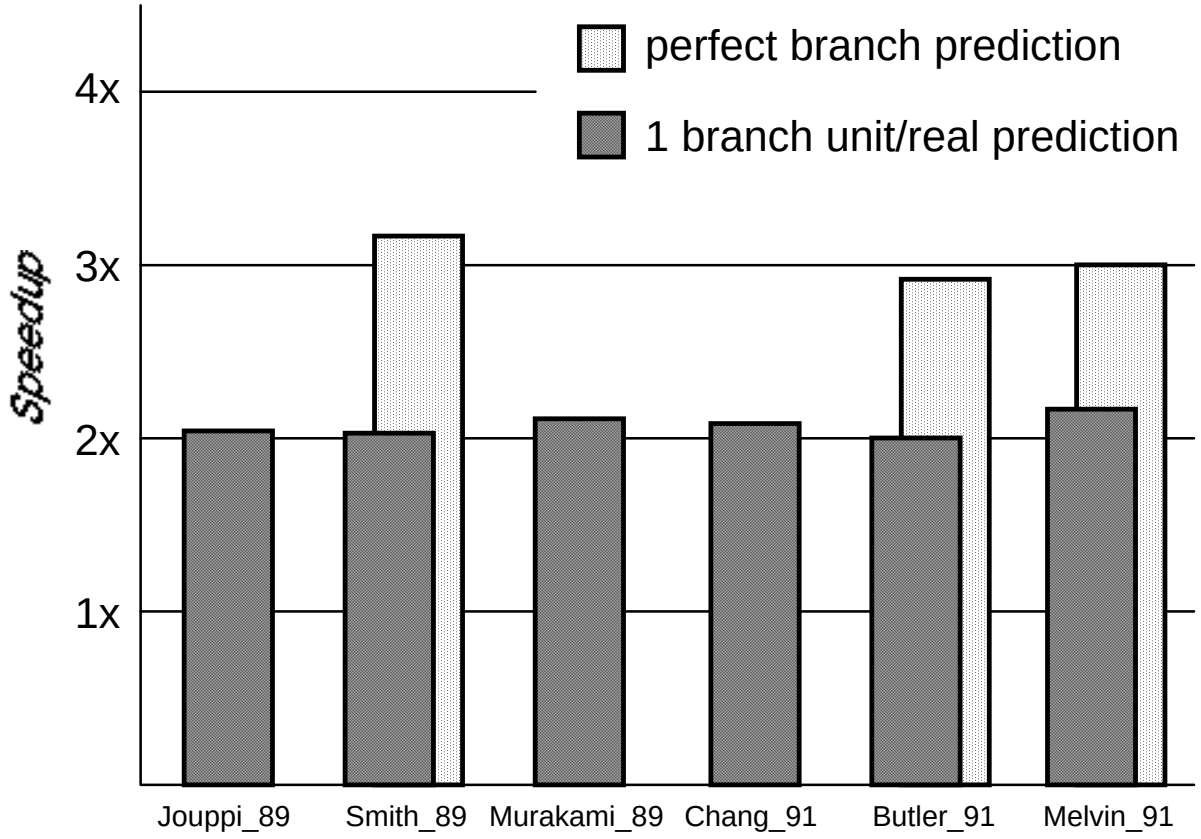
Results of ILP Studies

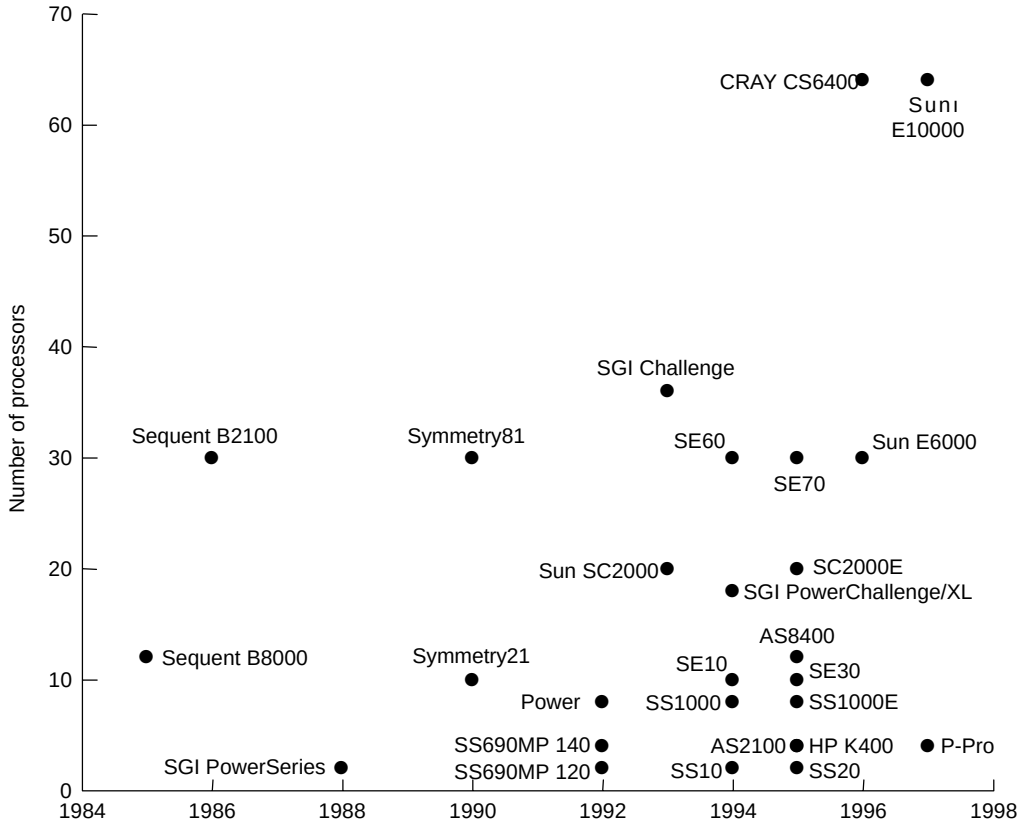
- Concentrate on parallelism for 4-issue machines



- Realistic studies show only 2-fold speedup
  - Recent studies show that more ILP needs to look across threads

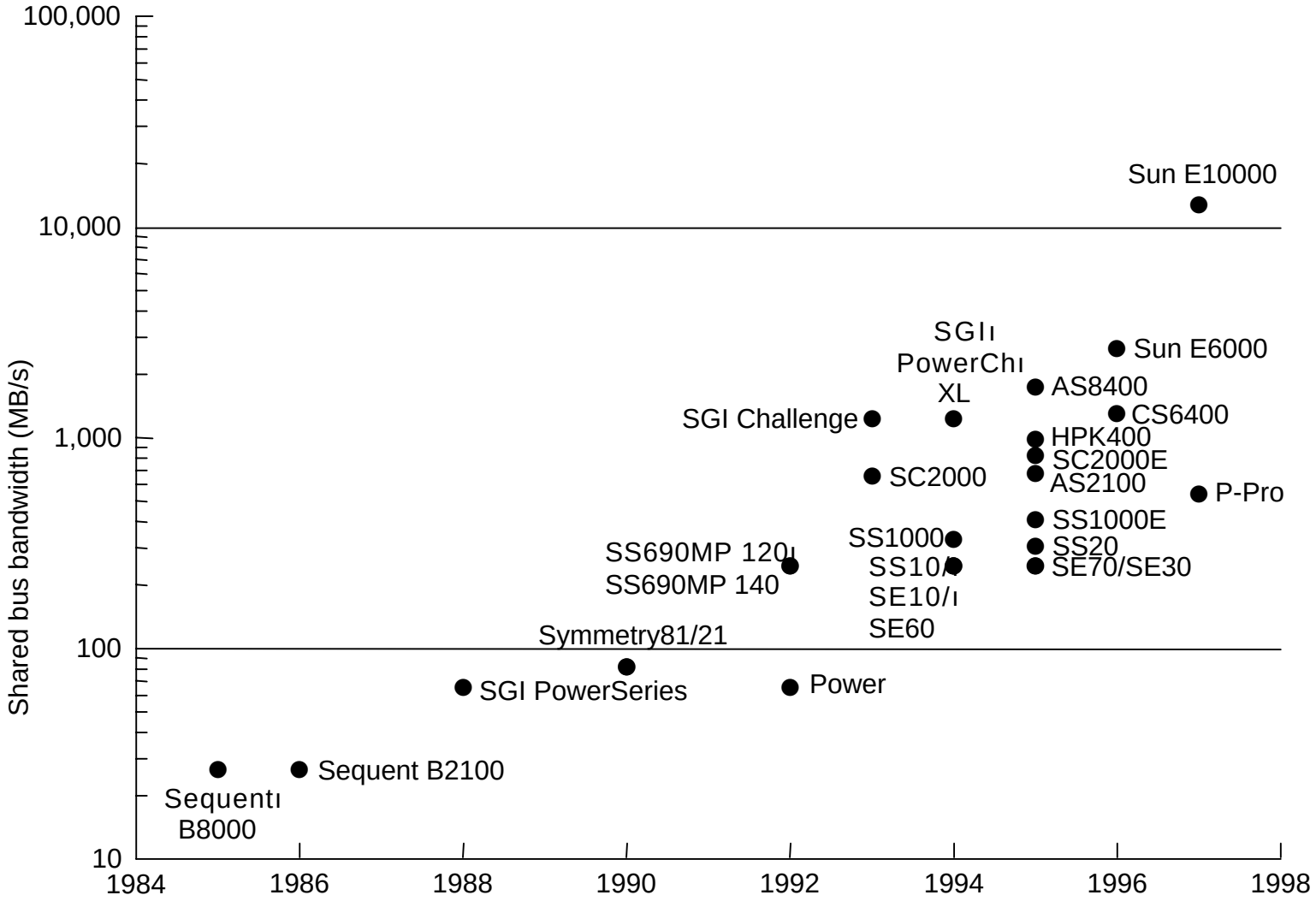
Architectural Trends: Bus-based MPs

- Micro on a chip makes it natural to connect many to shared memory
  - dominates server and enterprise market, moving down to desktop
- Faster processors began to saturate bus, then bus technology advanced
  - today, range of sizes for bus-based systems, desktop to large servers



No. of processors in fully configured commercial shared-memory systems

Bus Bandwidth



Economics

Commodity microprocessors not only fast but CHEAP

- Development cost is tens of millions of dollars (5-100 typical)
- BUT, many more are sold compared to supercomputers
- Crucial to take advantage of the investment, and use the commodity building block
- Exotic parallel architectures no more than special-purpose

Multiprocessors being pushed by software vendors (e.g. database) as well as hardware vendors

Standardization by Intel makes small, bus-based SMPs commodity

Desktop: few smaller processors versus one larger one?

- Multiprocessor on a chip

Consider Scientific Supercomputing

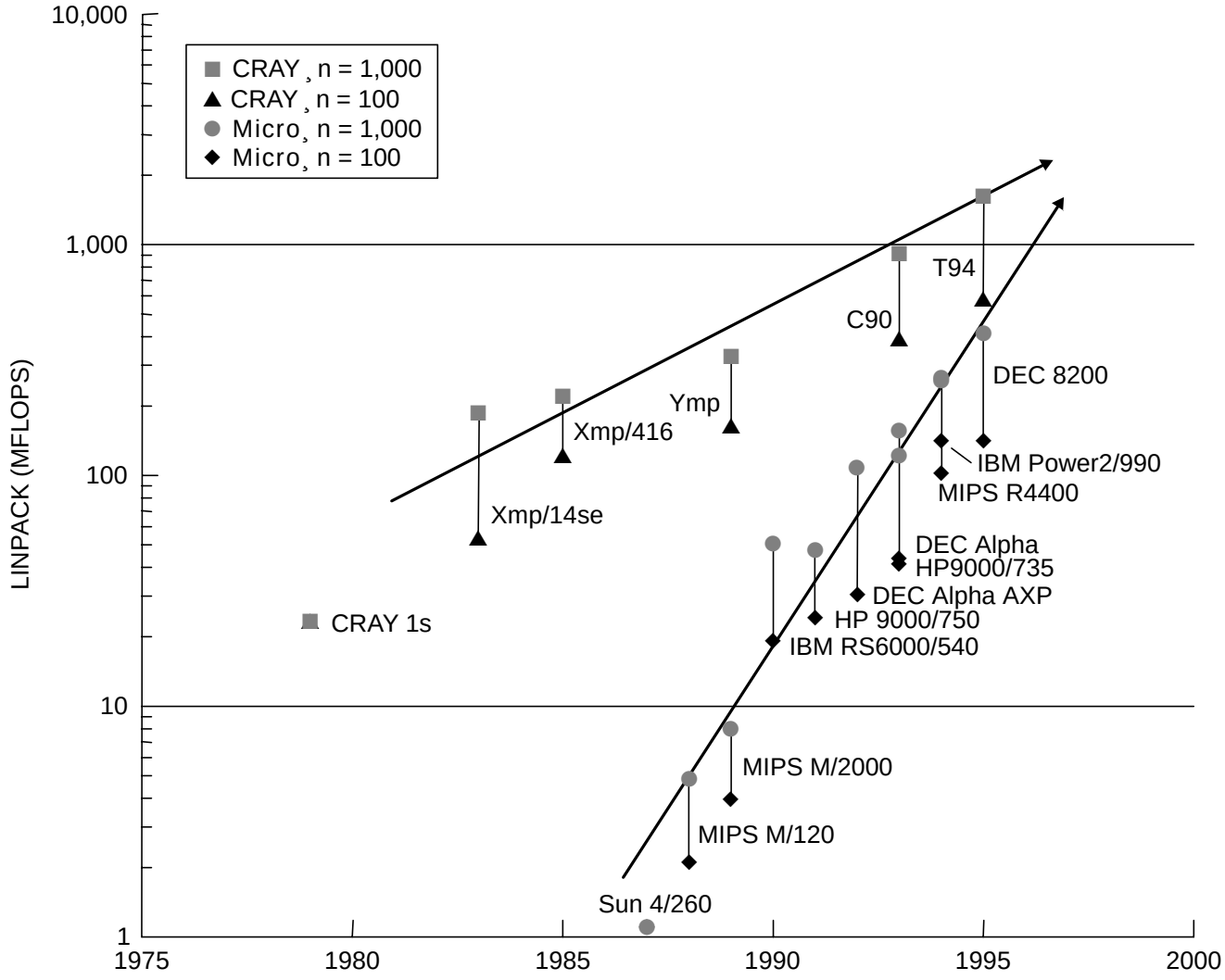
Proving ground and driver for innovative architecture and techniques

- Market smaller relative to commercial as MPs become mainstream
- Dominated by vector machines starting in 70s
- Microprocessors have made huge gains in floating-point performance
  - high clock rates
  - pipelined floating point units (e.g., multiply-add every cycle)
  - instruction-level parallelism
  - effective use of caches (e.g., automatic blocking)
- Plus economics

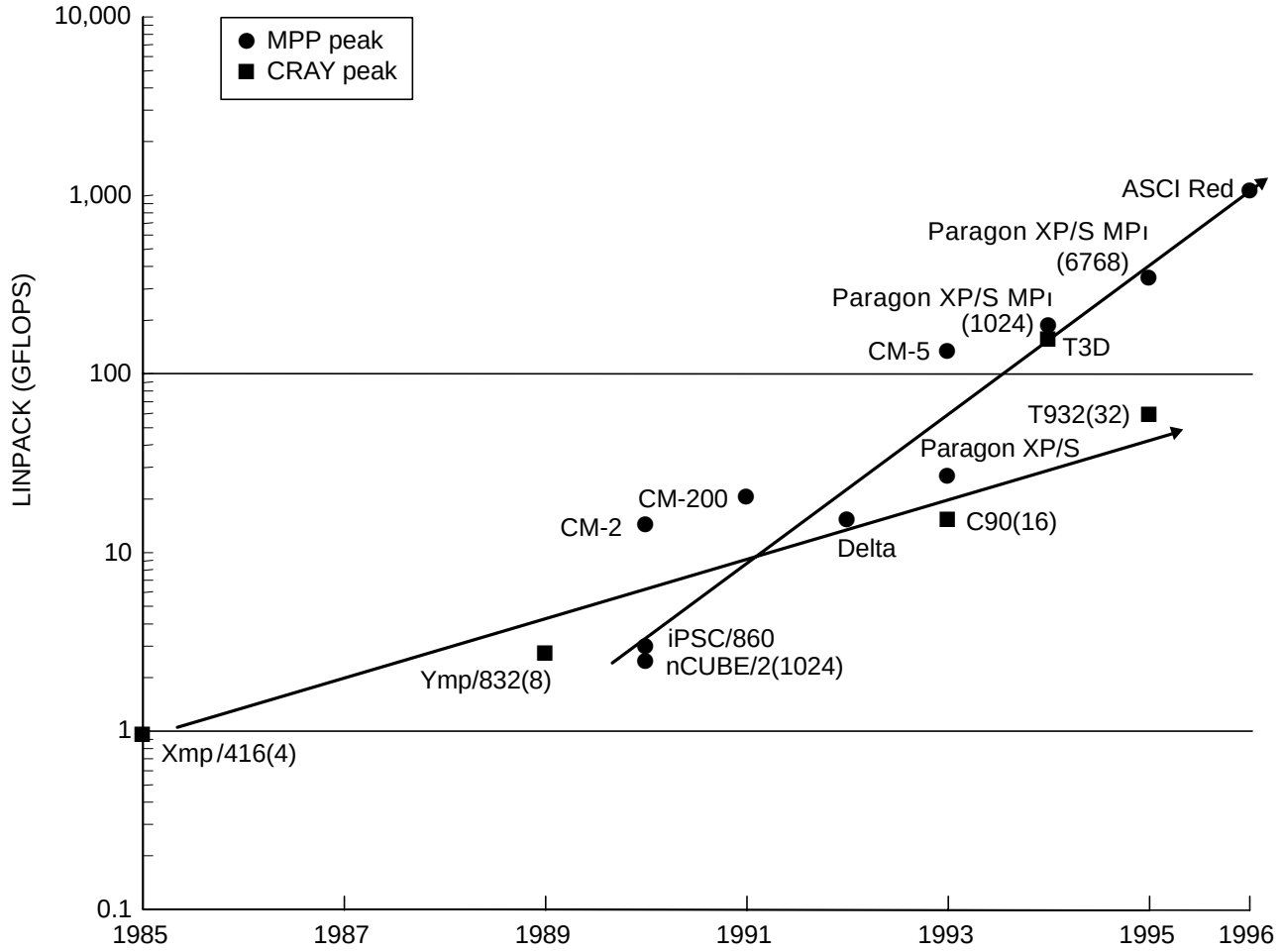
Large-scale multiprocessors replace vector supercomputers

- Well under way already

Raw Uniprocessor Performance: LINPACK

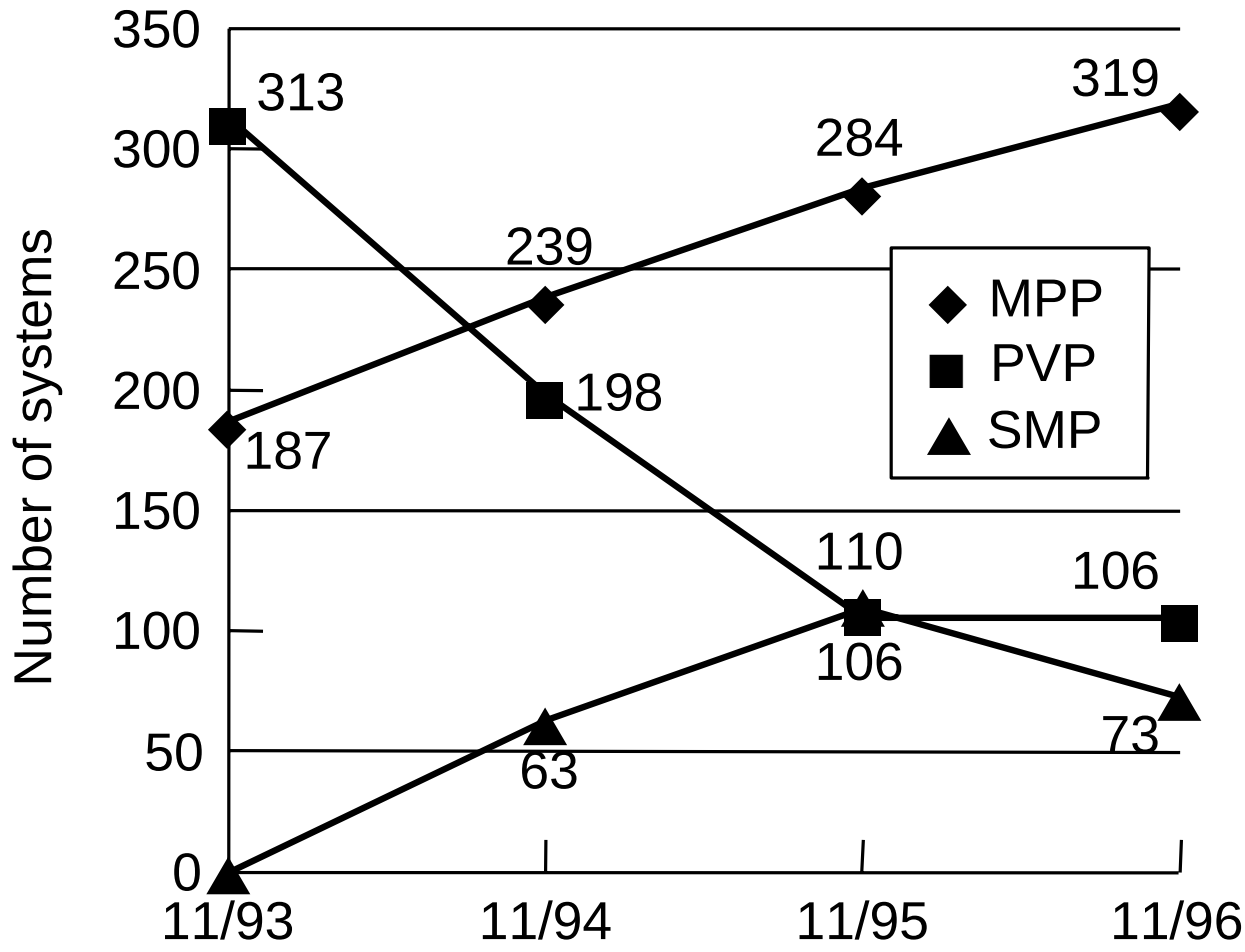


Raw Parallel Performance: LINPACK



- Even vector Crays became parallel: X-MP (2-4) Y-MP (8), C-90 (16), T94 (32)
- Since 1993, Cray produces MPPs too (T3D, T3E)

500 Fastest Computers



Summary: Why Parallel Architecture?

Increasingly attractive

- Economics, technology, architecture, application demand

Increasingly central and mainstream

Parallelism exploited at many levels

- Instruction-level parallelism
- Multiprocessor servers
- Large-scale multiprocessors (“MPPs”)

Focus of this class: multiprocessor level of parallelism

Same story from memory system perspective

- Increase bandwidth, reduce average latency with many local memories

Wide range of parallel architectures make sense

- Different cost, performance and scalability