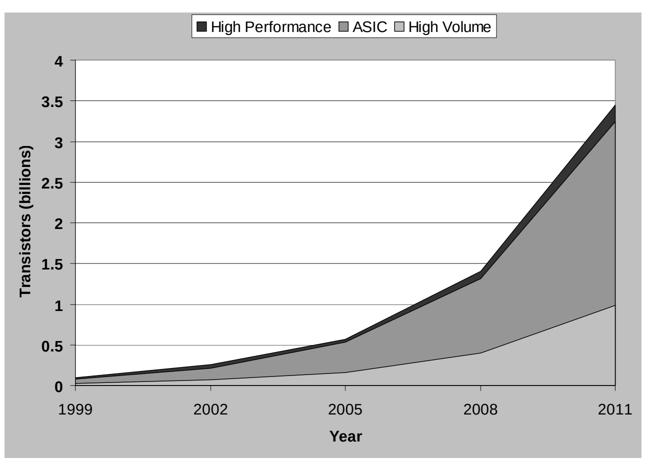
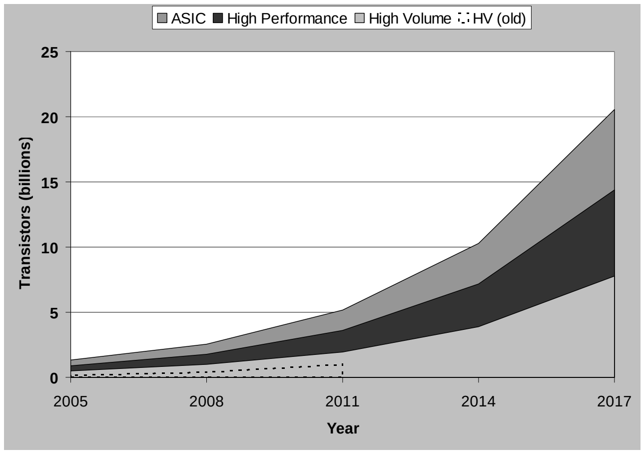
Role of New Technology in Computing Systems of the Future

Ravi Nair
IBM Thomas J. Watson Research Center
Yorktown Heights, NY

Transistors per Chip (Old ITRS Projection)



Transistors per Chip (New)



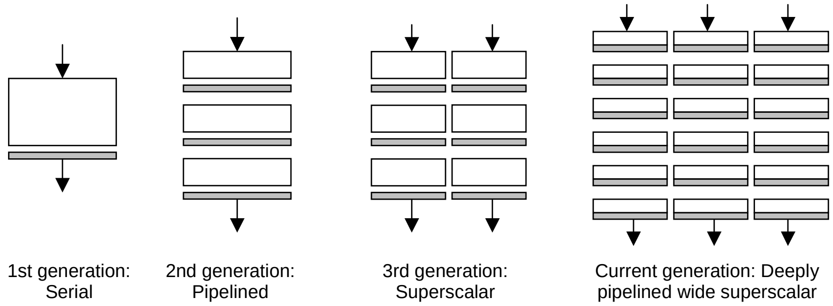
What should we do with all these transistors?

Work on a Chip

- Multiple threads
  - Communicating with each other
  - Communicating with the outside world
- Each thread
  - Going through different stages
  - Performing different types of functions

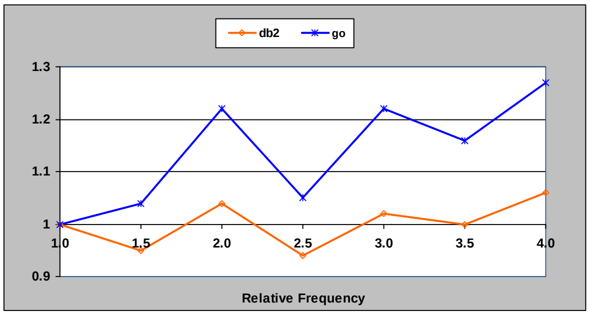
Transistors should be used to improve the useful work done on the chip

Evolution of Single-Thread Parallelism



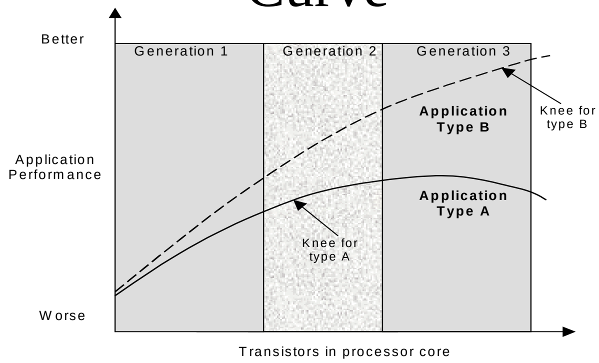
Each new generation needs increasingly greater formatting of instruction stream for maximum benefit

Performance Improvement through Pipelining



Exploitation of each new generation is minimal if done only in hardware

Processor-Application Efficiency Curve

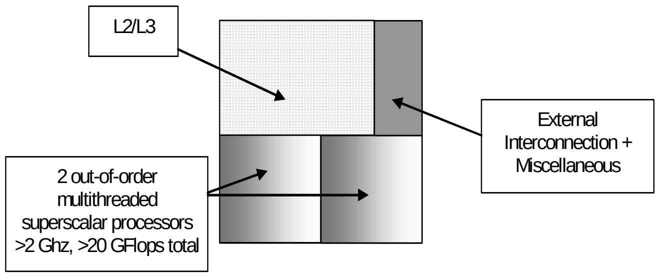


Phenomenal advances in process technology have led to proliferation of software and a shift back towards Type A

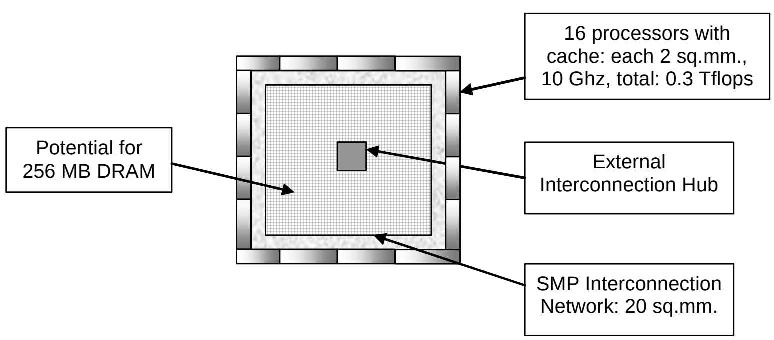
Stagnation of Traditional Systems

- Phenomenal progress in process technology
- Tremendous proliferation of software
- Inefficiency due to layers of software
- Improvement at the processor core level not translating to improvement in performance
- Shift back to Type A

Present-day Multithreaded Chip

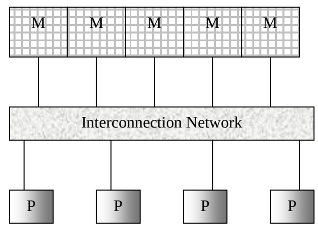


Scaled Chip for 2011



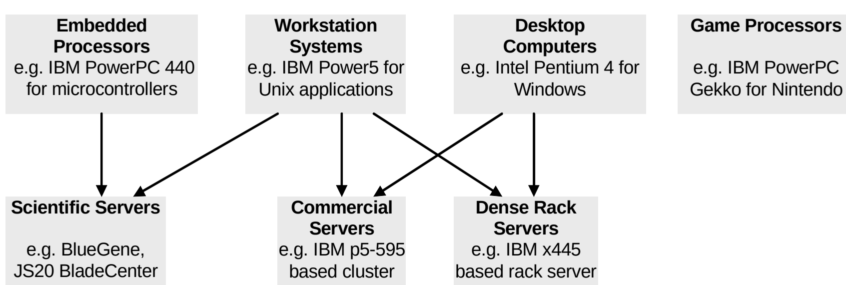
Multiprocessor Configurations

- Variety of configurations have been proposed
- SMP appears to be the survivor
  - ISA support, hardware implementations, software primitives appear to be maturing



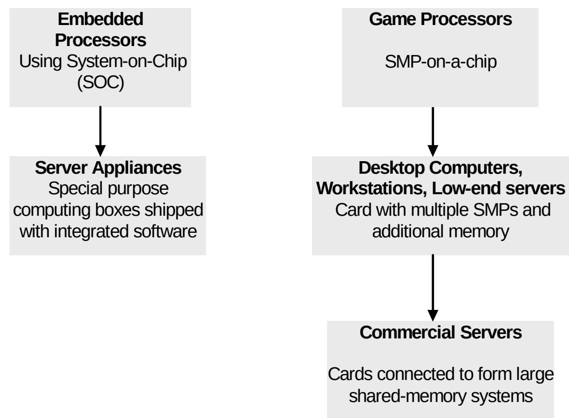
Past and current software and hardware investments in the SMP infrastructure will ensure its continued popularity

Growing Reuse of Microprocessor Cores



With increasing cost of designing, implementing, and testing microprocessors, there will be a move towards common cores

Possible Scenario for the 2010s



Apparent Trend

- No radical departures from current programming paradigms in the next 8-10 years
- Hardware systems will reflect this software inertia
- Hardware systems will migrate functionality into chips with multi-billion transistors
  - All SMPs will migrate into the chip

The "Solution" Paradigm

- Attention will shift towards providing solutions to problems, rather than providing only more computational capability
  - Examples are search problems, complex decision-making
- Combination of heterogeneous computing elements and software spread across the globe to perform a task
- Entire system needs to be examined to provide optimal solution
  - Performance will be constrained more by software aspects of the system, not only by hardware
  - The processor is just a small element in this picture

Computing as an Utility

- Cost of maintaining a system outweighing the cost of purchasing a system
  - Updating the operating system and system utilities
  - Intrusion detection and prevention measures
  - Desire for portability of user's workspace
- Computational capability available on desktops is usually underutilized
- Yet a user often needs even more power than what is immediately available
- Motivation for the Power-Grid model
  - Pay for what you use

The “Appliance” Paradigm

- The Grid will spur the development of “appliances”
  - Special nodes on the grid
- An appliance is a system configured for a special purpose
  - Scientific computing
  - Java virtual machine server
  - Network appliance
- Appliances can take advantage of special purpose accelerators to perform cost-effective, power-efficient, domain-specific computations

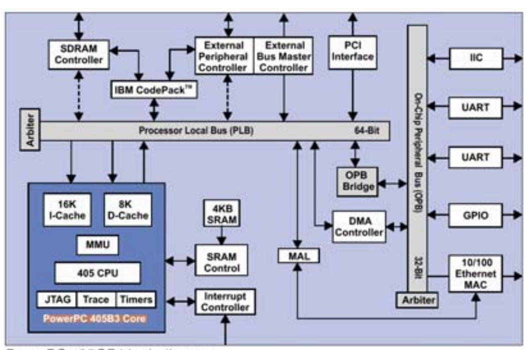
Appliances

- This will potentially be the area of maximum hardware innovation
  - New ISAs, new microarchitectures, new multiprocessor organizations
  - New technologies
- Power-efficient, cost-effective
  - Fewer layers in software
- Compatibility will be less of an issue
  - Hence more flexibility
- Software and hardware designed as a unit
  - Large granularity interface

Appliance Architecture

- Specialized interconnection only
  - Standard processors interconnected and programmed for a domain of applications
  - Example: BlueGene
- Specialized processors
  - Use of domain-specific accelerators
  - Example: SoCs
  - Use of nanotechnology
- System and application software tuned to the hardware

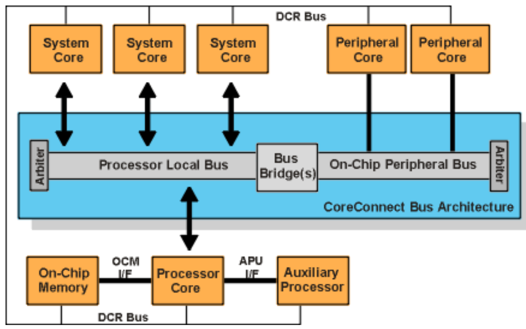
A System on a Chip



PowerPC 405GP block diagram

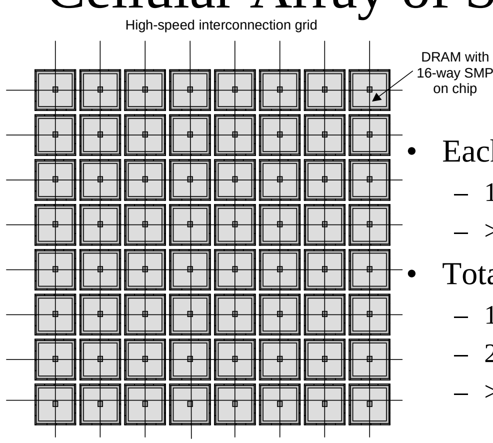
User must be provided the ability to easily snap together items from a library

Architecting the Interface



The challenge for architects is to design buses and interfaces that make this effective

Cellular Array of SMP Nodes



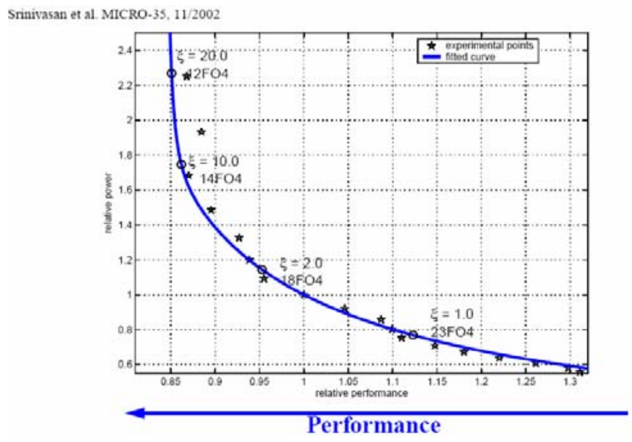
- Each node:
  - 16-way SMP, 4 Gbytes
  - > 10 Ghz
- Total system:
  - 1024 processors
  - 256 Gbytes
  - >20 TFlops

Each node is powerful enough to solve most of the problems of today

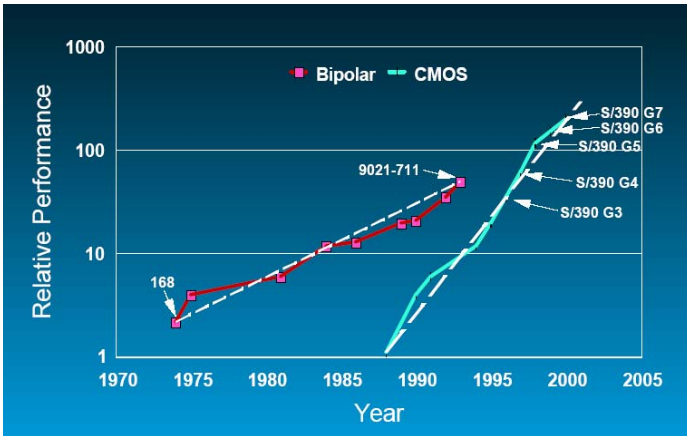
The "Trickle Up" Phenomenon

- Techniques at the low end drive the technology
  - Adopted later at the high-end traditional computing systems
- Examples:
  - Adoption of CMOS in mainframe systems
  - Adoption of SIMD media-computing accelerators
    - Intel i860 (1989) had subword parallelism
    - Allowed 8-bit, 16-bit and 32-bit operations on a 64-bit ALU
  - Adoption of power-saving techniques in workstation systems and blades
  - Adoption of embedded techniques in high performance systems
    - BlueGene, #1 in Top500 Supercomputer rankings, uses embedded technology
      - 500 Mhz, 4 MFlops per chip, 64K chips

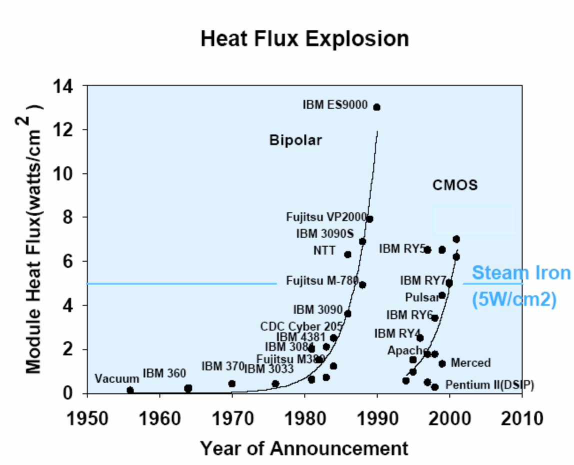
Power Problem



Adoption of CMOS on Mainframes



CMOS is also getting too hot



Problems to work on

- Technology
  - On-chip coherence implementations
  - Bandwidth to/from chip
  - Variability of component performance (a) in space (b) in time
  - Inherently power-efficient technology and techniques
- Architecture
  - SoC Architecture
  - Special purpose engines
- RAS
  - On-chip fault-tolerance
  - System availability
  - Failure prediction
- Ease-of-use
  - Cellular-array programming
  - SoC development tools
  - Application-to-SoC compilation

Concluding Remarks

- Traditional computing systems getting more conservative in hardware
  - Attention shifting to system-wide solutions
- New ideas, new systems, new technologies will drive special-purpose applications
- Nanotechnology in conjunction with new system configurations, new compilers and special OSes could spearhead this appliance revolution
- Software, especially virtualization and compilers, must be included in this multidisciplinary effort