

IBM SP2 Overview

- Scalable POWERparallel 1 (SP1)
 - Development started February 1992, delivered to users in April 1993
- Scalable POWERparallel 2 (SP2)
 - 120-node systems delivered 1994
 - 4–128 nodes, POWER2 processor, 66.7 MHz, 267 MFLOPS
 - POWER2 used in RS 6000 workstations, compatibility with existing software
 - 1997 version:
 - 604 High Node (SMP node): 16 nodes max, 2–8 PowerPC 604, 112 MHz, 224 MFLOPS, 64MB–2GB memory
 - Wide Node: 128 nodes max, 1 P2SC (POWER2 Super Chip, 8 chips on one chip), 135 MHz, 540 MFLOPS, 64MB–2GB memory

1

Fall 1999, Lecture 33

IBM SP2 Overview (cont.)

- Runs various combinations of serial, parallel, interactive, and batch jobs
 - Partition between types can be changed
 - Thin nodes — compute nodes, either shared or dedicated
 - Wide nodes — configured as servers, with extra memory, storage devices, etc.
- A system “frame” contains 16 thin processor or 8 wide processor nodes
 - Includes redundant power supplies
 - Nodes are hot swappable within frame
 - Includes a high-performance switch for low-latency, high-bandwidth communication (multistage network, scales to provide same bandwidth to each processor node)

2

Fall 1999, Lecture 33

IBM SP2 Processors

- POWER2 processor
 - Various versions from 20 to 62.5 MHz
 - RISC processor, load-store architecture
 - Floating point multiple & add instruction with latency of 2 cycles, pipelined for initiation of new one each cycle
 - Conditional branch to decrement and test a “count register” (without fixed-point unit involvement), good for loop closings
- POWER 2 processor chip set
 - 8 semi-custom chips: Instruction Cache Unit, four Data Cache Units, Fixed-Point Unit (FXU), Floating-Point Unit (FPU), and Storage Control Unit
 - 2 execution units per FXU and FPU
 - Can execute 6 instructions per cycle: 2 FXU, 2 FPU, branch, condition register
 - Options: 4-word memory bus with 128 KB data cache, or 8-word with 256 KB

3

Fall 1999, Lecture 33

IBM SP2 Interconnection Network

- General
 - Multistage network, extra stages added as necessary to keep bandwidth to each processor constant
 - Message delivery
 - PIO for short messages with low latency and minimal message overhead
 - DMA for long messages
 - Multi-user support — hardware protection between partitions and users, guaranteed fairness of message delivery
- Routing
 - Packet switched = each packet may take a different route
 - Cut-through = if output is free, starts sending without buffering first
 - Wormhole routing = buffer on subpacket basis if buffering is necessary

4

Fall 1999, Lecture 33

IBM SP2 AIX Parallel Environment

- Parallel Operating Environment — based on AIX, includes Desktop interface
 - Partition Manager to allocate nodes, copy tasks to nodes, invoke tasks, etc.
 - Program Marker Array — (online) squares graphically represent program tasks
 - System Status Array — (offline) squares show percent of CPU utilization
- Parallel Message Passing Library
- Visualization Tool — view online and offline performance
 - Group of displays for communications characteristics or performance (connectivity graph, inter-processor communication, message status, etc.)
- Parallel Debugger

5

Fall 1999, Lecture 33

Kendall Square Research KSR1 Overview

- 6 years in development, 36 variations in 1992 (8 cells for \$500k, 1088 for \$30m)
 - 8 cells, 320 MFLOPS, 256 MB memory, 210 GB disk, 210 MB/s I/O
 - 1088 cells, 43 GFLOPS, 34 GB memory, 15 TB disk, 15 GB/s I/O
- Each system includes:
 - Processing Modules, each containing up to 32 APRD Cells including 1GB of ALLCACHE memory
 - Disk Modules, each containing 10 GB
 - I/O adapters
 - Power Modules, with battery backup
- Modular, scalable, with hot-swappable components

6

Fall 1999, Lecture 33

Kendall Square Research KSR1 Processor Cells

- Each APRD (ALLCACHE Processor, Router, and Directory) Cell contains:
 - 64-bit Floating Point Unit, 64-bit Integer Processing Unit
 - Cell Execution Unit for address gen.
 - 4 Cell Interconnection Units, External I/O Unit
 - 4 Cache Control Units
 - 32 MB of Local Cache, 512 KB of subcache
- 1.2 μm devices, each up to 450,000 transistors, packaged in 8x13x1 printed circuit board
 - 20 MHz clock
 - Can execute 2 instructions per cycle

7

Fall 1999, Lecture 33

Kendall Square Research KSR1 ALLCACHE System

- The ALLCACHE system moves an address set requested by a processor to the Local Cache on that processor
 - Provides the illusion of a single sequentially-consistent shared memory
- Memory space consists of all the 32 KB local caches
 - No permanent location for an “address”
 - Addresses are distributed and based on processor need and usage patterns
 - Each processor is attached to a Search Engine, which finds addresses and their contents and moves them to the local cache, while maintaining cache coherence throughout the system
 - 2 levels of search groups for scalability

8

Fall 1999, Lecture 33

Kendall Square Research KSR1 Programming Environment

- KSR OS = enhanced OSF/1 UNIX
 - Scalable, supports multiple computing modes including batch, interactive, OLTP, and database management and inquiry

- Programming languages
 - FORTRAN with automatic parallelization
 - C
 - PRESTO parallel runtime system that dynamically adjusts to number of available processors and size of the current problem