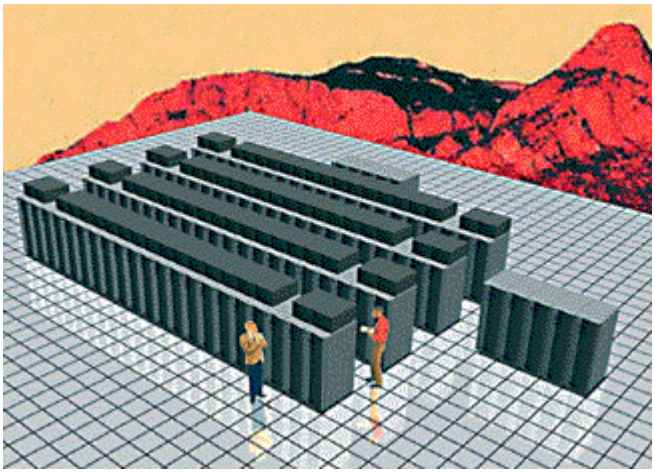
The ASCI Red TOPS Supercomputer

Introduction

The ASCI Red TOPS Supercomputer is the first step in the ASCI Platforms Strategy, which is aimed at giving researchers the five-order-of-magnitude increase in computing performance over current technology that is required to support "full-physics," "full-system" simulation by early next century. This supercomputer, being installed at Sandia National Laboratories, is a massively parallel, MIMD computer. It is noteworthy for several reasons. It will be the world's first TOPS supercomputer. I/O, memory, compute nodes, and communication are scalable to an extreme degree. Standard parallel interfaces will make it relatively simple to port parallel applications to this system. The system uses two operating systems to make the computer both familiar to the user (UNIX) and non-intrusive for the scalable application (Cougar). And it makes use of Commercial Commodity Off The Shelf (CCOTS) technology to maintain affordability.

Hardware

The ASCI TOPS system is a distributed memory, MIMD, message-passing supercomputer. All aspects of this system architecture are scalable, including communication bandwidth, main memory, internal disk storage capacity, and I/O.



Artist's Concept

The TOPS Supercomputer is organized into four partitions: Compute, Service, System, and I/O. The Service Partition provides an integrated, scalable host that supports interactive users, application development, and system administration. The I/O Partition supports a scalable file system and network services. The System Partition supports system Reliability, Availability, and Serviceability (RAS) capabilities. Finally, the Compute Partition contains nodes optimized for floating point performance and is where parallel applications execute. The system hardware parameters are summarized in Table 1.

Table 1. System hardware parameters

Compute Nodes (Red - Red / Black - Black)	4,640 (1,168 - 2,304 - 1,168
Service Nodes (Red / Black)	16 (8 / 8)
Disk I/O Nodes (Red / Black)	74 (37 / 37)
System Nodes (Red / Black)	2 (1 / 1)
Network Nodes - Ethernet/ATM (Red / Black)	20 (10 / 10)
System Footprint	~2500 Square Feet
Number of Cabinets (Computer / Switch / Disk)	104 (76 / 8 / 20)
System RAM (Compute Nodes / I/O Nodes) (Will double when upgrade is complete)	606 GB Total (128 MB / 256 MB)
Topology	Mesh (38 X 32 X 2)
Node Link Bandwidth - Bi-directional	800 MB/s
Cross Section Bandwidth - Bi-directional	51.2 GB/s
Total Number of Pentium II Xeon Core Processors	9536
Processor to Memory Bandwidth	533 MB/s
Compute Node Peak Performance	666 MOPs
System Peak Performance	3.15 TOPs
Linpack Performance - Full System (Center + Red or Black / Red or Black)	2.1213 TOPs (1.5336 TOPs / 5.22.5 TOPs)
RAID Disk Storage - Total / per Color	12.5 TB / 6.25 TB
RAID I/O Bandwidth - Total per Subsystem	4.0 GB/s 1.0 GB/s

Software

Software on the TOPS Supercomputer is a combination of operating systems tailored for specific tasks and standard programming tools to make the computer both familiar to the user and non-intrusive for the scalable application. To the application programmer, the system looks like a UNIX-based supercomputer. All the standard facilities associated with a UNIX workstation will be available to the user.

The operating system used for the Service, I/O, and System Partitions is Intel's distributed version of

UNIX (POSIX 1003.1 and XPG3, AT&T System V.3 and 4.3 BSD Reno VFS) developed for the Paragon XP/S Supercomputer. The Paragon OS presents a single system image to the user. This means that users see the system as a single UNIX machine despite the fact that the operating system is running on a distributed collection of nodes.
The operating system in the Compute Partition is Cougar. Cougar is Intel's port of Puma, a light-weight operating system for the TOPS, based on the very successful SUNMOS system for the Paragon. (SUNMOS, and subsequently Puma, were developed by Sandia National Laboratories and the University of New Mexico.) System services and support for the interactive user are provided by a host OS (in this case, the Paragon OS running in the Service Partition). All access to hardware resources comes from the Q-Kernel, the lowest-level component of Cougar. Above the Q-Kernel sits the process control thread (PCT), which runs in user space and manages processes. At the highest level is the user's applications. As with most MPP systems, the basic programming model in Cougar is based on message passing. FORTRAN77, FORTRAN90, C and C++ will be supported. The interactive debugger and performance analysis tools will understand these languages and map onto original source code. A checkpoint/restart capability, required for the huge applications to be run on this computer, will be provided. The I/O bandwidth will be sufficient to checkpoint the entire system memory in approximately five minutes.

Conclusion

The ASCI platform effort will bridge the gap between giga-scale and tera-scale computing to accommodate the five-order-of-magnitude increase in performance required by "full-physics", "full-system" simulation. The first deliverable in this effort is the 1.8 TOPS ASCI Red Supercomputer, a distributed memory, message passing, MIMD system that Intel is building and installing at Sandia National Laboratories.

For more information, contact: **SAN**

James L. Tomkins, SNL, 505-845-7249, jltomkin@sandia.gov

Last modified: October 15, 1999