TECHNOLOGY BRIEF

March 1999

Compaq Computer Corporation

ISSD Technology Communications

CONTENTS

Compaq 8-Way Multiprocessing Architecture

EXECUTIVE OVERVIEW

The next logical trend in the performance curve for high-end, standards-based servers is an 8-way architecture coupled with the next-generation Intel 32-bit (IA-32) processor: the Pentium® III Xeon™ processor. Compaq carefully evaluated the options for an 8-way symmetric multiprocessing (SMP) architecture and joined efforts with Intel/Corollary to develop an optimum solution for next-generation, standards-based servers. The combination of Compaq's input/output (I/O) controller technology with the Profusion 5-point crossbar switch, memory controller, and Pentium® III Xeon™ processors will yield a balanced, high-performance system architecture for the future. Intel's acquisition of Corollary further ensures that the 8-way SMP architecture will become the industry standard for 8-way servers. This brief explains Compaq's current architecture for 8-way multiprocessing servers and the unprecedented performance and scalability delivered with this emerging industry-standard technology.

Please direct comments regarding this communication to the ISSD Technology Communications Group at this Internet address: TechCom@compaq.com

COMPAQ

NOTICE

Compaq 8-Way Multiprocessing Architecture
Third Edition (March 1999)
Document Number ECG058/0399

INTRODUCTION

Not long ago, 4-way SMP servers emerged as the leading edge in standards-based servers. Currently, certain applications and environments demand greater performance than four processors can supply. For SMP servers, the next logical trend in the performance curve will be 8-way servers that use the next-generation Intel 32-bit (IA-32) processor: the Pentium® III Xeon™ processor. The Pentium® III Xeon™ processor is the right choice for 8-way implementation for two reasons. First, because of its improved architecture and higher clock speeds, it will substantially outperform the Intel Pentium® II Xeon™ processor in 8-way implementations. Second, the Pentium® II Xeon™ processor is nearing the final version of its generation while the Pentium® III Xeon™ processor is the first member of the next generation in 32-bit microprocessors from Intel. Compaq will not ask customers to invest in short-lived technology.

This brief explains Compaq's 8-way SMP server architecture. It assumes that the reader understands current SMP designs.

8-WAY ARCHITECTURE OVERVIEW

Compaq and Corollary have been working together on the Profusion 8-way chipset technology for more than two and one half years. Corollary designed the Profusion 5-point crossbar switch that is built into the architecture and the memory controller. Compaq has applied its experience with industry-leading I/O subsystems, including peripheral component interface (PCI) technology, to complete the overall design for a balanced 8-way architecture.

In the history of standards-based SMP machines, Compaq and Corollary have demonstrated a consistent vision and commitment to advancing SMP technology. Compaq will use the industry-standard 8-way SMP architecture with its built-in Profusion chipset in future system designs. The jointly developed Profusion chipset including the Compaq I/O controller have been made available to other system providers.

Figure 1 shows a block diagram of the 8-way SMP architecture. The essential features of the architecture are:

- Dual 100-megahertz (MHz) processor buses
- Dedicated 100-MHz I/O bus
- 8-way multiprocessing with Pentium® III Xeon™ processors
- Multiported system architecture (5-point crossbar switch)
- Dual-ported, interleaved memory
- Uniform memory access for all eight processors
- Dual cache accelerators
- Up to four Compaq-designed host-to-PCI bridges
- Up to 32 gigabytes (GB) of synchronous dynamic random-access memory (SDRAM)

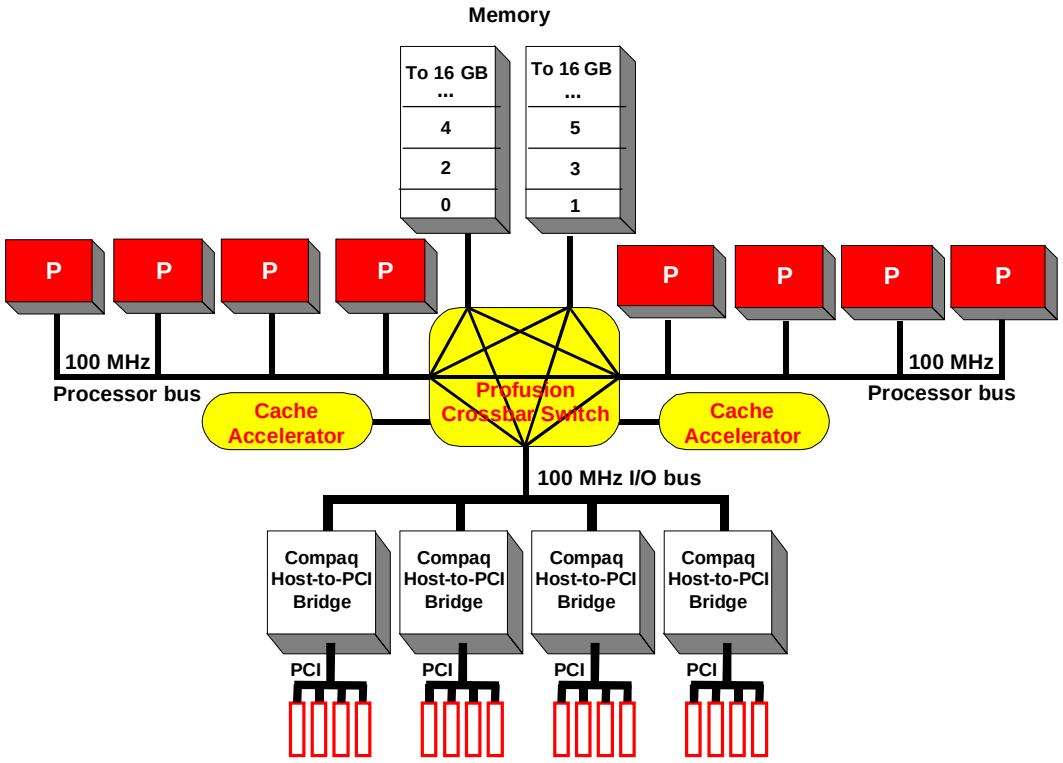

Figure 1: Compaq 8-way SMP system architecture

Processor and I/O Bus Design

The two processor buses and the dedicated I/O bus are Gunning transceiver logic (GTL+), 64-bit buses with a theoretical maximum bandwidth of 800 megabytes per second (MB/s). The use of GTL+ allows the buses to operate at higher clock speeds without severely reducing the bus length or number of electrical loads. All three buses operate at 100 MHz.

The Profusion chipset joins the two processor buses, the I/O bus, and two memory ports together through a crossbar switch. The otherwise independent processor and I/O buses are joined by a logical connection that is made only when required to transfer data. The GTL+ bus running at 100 MHz can support a maximum of five loads per bus. This allows four processors and one connection to the memory controller on each processor bus and up to four host-to-PCI bridges with a connection to the memory controller on the I/O bus. Each of the three GTL+ buses has independent access to the two memory controller buses. This architecture prevents I/O traffic from consuming bandwidth on the processor buses.

Processor Technology

The Pentium® III Xeon™ processor is based on the Pentium® II Xeon™ processor core and cartridge form factor. The Pentium® III Xeon™processor provides a higher internal core frequency than the Pentium® II Xeon™ processor. It also includes 70 new streaming instructions that permit more efficient access to memory above 4 GB.

100-MHz Bus

Compaq 8-way SMP architecture takes full advantage of the 100-MHz capabilities of both processor buses. Because of the increased frequency of the bus, the number of electrical loads must be limited so signals do not degrade and cause a drop in bus performance. The Profusion chipset limits the number of electrical loads on each bus to five. This is an optimum solution for the Pentium® III Xeon™ processor that will operate at bus frequencies up to 100 MHz.

Core Frequency

The Pentium® II Xeon™ processor operates at a maximum core frequency of 450 MHz on a 100-MHz bus. On the other hand, the Pentium® III Xeon™ processor operates at a minimum core frequency of 500 MHz on a 100-MHz bus (550 MHz processor expected soon).

Level-2 Cache

Pentium® III Xeon™ processors will have a Level-2 (L2) cache size of 512 kilobytes (KB), 1 MB, or 2 MB. As with Pentium® II Xeon™ processors, the cache will operate on a full-speed backside bus. The large cache size of the Pentium® III Xeon™ processor combined with the full-speed backside bus will allow very efficient access to the most frequently used data. This allows the processors to use less bandwidth on the host bus and notably enhances the performance of an 8-way SMP server. Table 1 compares basic features of the Pentium® II Xeon™ processor and the Pentium® III Xeon™ processor.

Table 1: Comparison of Pentium® II Xeon™ processor and Pentium® III Xeon™ processor

Feature	Pentium® II Xeon™	Pentium® III Xeon™
Core Speed	400 MHz, 450 MHz	500 MHz and above
System Bus Speed	100 MHz	100 MHz
L2 Cache Bus Speed	Full Speed	Full Speed
L2 Cache Size	512 KB, 1 MB, 2 MB	512 KB, 1 MB, 2 MB
MMX™ Technology	Yes	Yes
Additional Streaming Instructions		70 instructions

Profusion 5-point Crossbar Switch

An important element of the 8-way SMP architecture is Corollary's system controller design, which is based on a crossbar switch. The 5-point (multiport) crossbar switch is schematically represented in Figure 2. The static random-access memory (SRAM) actually has ten unidirectional ports that appear as five bi-directional ports to the user.

The 5-point crossbar switch is nonblocking. It allows simultaneous read and write paths. There are ten possible information paths between the processors, memory, and I/O. These direct paths minimize the time it takes to move data from one port to another.

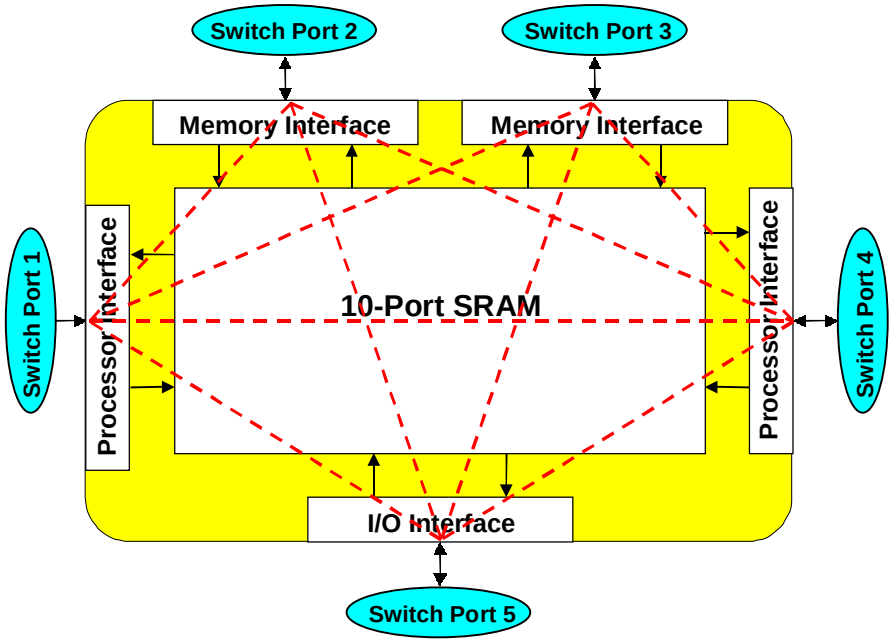

Figure 2: Profusion 5-point (multiport) crossbar switch

Figure 3 shows partitioning of the application-specific integrated circuit (ASIC) in the Profusion crossbar switch. While the crossbar switch consists of two physical chips, the functions are partitioned so that for every transaction from a processor or the host-to-PCI bridge, the address and command portions are routed through the memory address controller (MAC). The data is routed through the data interface buffer (DIB). The MAC manages the external cache accelerators and tracks the information stored in the DIB.

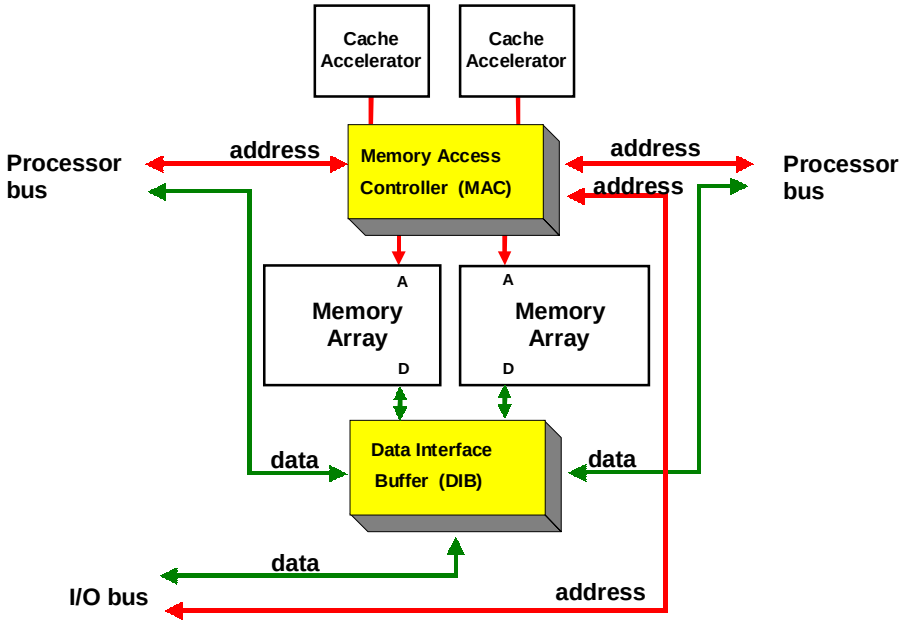

Figure 3: ASIC partitioning in the Profusion chipset

When data travels through the 5-point crossbar switch, it goes through the DIB. The DIB allows simultaneous data transfer on all five ports, has 64-cache-line buffers, and uses error-correcting code to maintain data integrity. The cache-line buffers can be used by any transaction for any device on any bus. The efficiency of the buffers is high because there are no dedicated queues between buses.

Finally, with the 100-MHz bus, there is a peak throughput on each port of 800 MB/s. With all five ports, the crossbar switch allows a maximum peak throughput of 4.0 GB/s.

Dual Interleaved Memory and Uniform Memory Access

As noted in Figures 1, 2, and 3, the crossbar switch has two ports to main memory. Having two memory ports increases memory bandwidth, reduces access conflicts, and increases the maximum memory supported. The two memory banks are cache-line interleaved; they share a common address range. One memory port responds to even-numbered cache lines, and the other port responds to odd-numbered cache lines. This configuration has the highest performance because it allows both memory buses to be used simultaneously, theoretically doubling throughput. It is especially advantageous for applications that access memory in a random manner. In random accesses, roughly half the requests at any one time are even-numbered lines, while the other half are odd-numbered lines.

Due to the design of the dual-ported memory controller, processors have equal access times to either memory port. This uniform memory access reduces latency. In nonuniform memory access (NUMA) architectures, a processor has quick access to one memory bank but a lag time (or latency) to a second memory bank.

Cache Accelerators

Compaq 8-way SMP architecture uses cache accelerators to minimize unnecessary traffic on the processor buses when maintaining data coherency between the caches on all three buses. One of the design challenges of SMP systems is to maintain a consistent view of memory by all the processors and the I/O subsystems. This is typically referred to as maintaining data coherency. Because data is shared between multiple caches on the processor buses and main memory, there is the possibility that copies of the same address of data may be inconsistent. Intel architecture processors support a snooping protocol to solve the cache-coherency problem. In a typical two-bus design, a memory transaction from one processor bus would have to "snoop" the remote processor bus to make sure that the most recent data is used. Every snoop cycle consumes bandwidth on the remote bus and limits the performance of the system.

The Profusion chipset uses two cache accelerators to minimize the number of snoop cycles that occur between the processor buses when supporting processors on both buses. Cache accelerators are not required when using only one processor bus. Each cache accelerator holds the addresses of data stored in all of the L2 processor caches on its corresponding bus. A cache accelerator also holds information about the state of the data—for example, whether the data is shared between multiple caches or whether the data is modified. By holding the state of the data in addition to the address, the cache accelerator can direct snoop traffic to the other bus only when it is required to maintain data coherency. This filtering reduces traffic on both buses, therefore improving overall architecture performance.

I/O Filter

The 8-way SMP architecture also includes Compaq host-to-PCI bridges with prefetch buffers that make them caching bridges. The Profusion chipset contains a built-in I/O filter for the caching bridges on the I/O bus. This I/O filter is specifically designed to work with up to four Compaq host-to-PCI bridges. When a processor requests a cache line with the intent to modify, the MAC

performs a look-up into the I/O filter to determine if that cache line resides in one of the host-to-PCI bridges. If it does reside there, the MAC initiates a transaction on the I/O bus to invalidate that cache line. If the cache line is not present in one of the bridges, then no transaction is run on the GLT+ bus. This reduces snoop traffic on the I/O bus whenever a processor requests data. The cooperative design of the memory and I/O subsystems is a clear example of the performance advantage to be gained through the technology exchange between Corollary and Compaq.

I/O Technology

Compaq has designed a host-to-PCI bridge (or I/O controller) to address the needs of next-generation servers. It is specifically engineered to enhance performance of the Profusion chipset.

Compaq's host bridge includes the following important features:

- Compliant with the PCI 2.2 Specification[1]
- 64-bit, 66-MHz PCI support
- Asynchronous design to accommodate multiple bus frequencies
- Multiple prefetch buffers
- PCI Hot Plug controller integrated into the bridge
- Peer-to-peer operations supported on a single PCI bus segment and across the I/O bus to other PCI segments

Delayed Transactions

One of the most important features of Compaq's host bridge is that it supports delayed PCI transactions, which improve bus performance. Other host-to-PCI bridge devices do not support delayed transactions at all. These devices often can slow transactions when both a read and a write request occur simultaneously. A delayed transaction is a modified version of a split transaction. Splitting one transaction into two separate transactions frees the bus and the processor for other activities while data requests are being completed.

In a delayed PCI transaction, however, the processor must poll the host bridge to determine when its data is there, rather than waiting for an interrupt as with a split transaction. Compaq built in additional features to reduce the amount of processor polling required by the delayed PCI transaction, further maximizing bus efficiency.

PCI Bus

The Compaq host-to-PCI bridge supports 64-bit, 66-MHz PCI transactions. Therefore, it provides four times the maximum theoretical bandwidth of today's standard 32-bit, 33-MHz PCI bus.

Asynchronous Design

The host-to-PCI bridge is split into two sections: upstream (host I/O side) and downstream (PCI side). All the functions on the upstream side are in the host processor clock domain at 100 MHz. All functions on the downstream side are in the PCI clock domain at up to 66 MHz. This asynchronous design will provide maximum I/O performance as clock rates increase with new processor releases.

---

[1] *PCI Local Bus Specification*, Revision 2.2, announced January 25, 1999, PCI Special Interest Group

Multiple Prefetch Buffers

The Compaq host-to-PCI bridge is designed with multiple prefetch buffers to ensure optimum I/O-to-processor performance. Each buffer can hold multiple cache lines. These buffers have been sized to provide optimal performance at a reasonable and cost-effective die size. Because of the delayed transaction support, multiple PCI devices can request data and the bridge can get the data concurrently. Other host-to-PCI bridges can only hold a single cache line and can only handle a single request at a time.

PCI Hot Plug Technology

The host-to-PCI bridge supports PCI Hot Plug technology, pioneered by Compaq. PCI Hot Plug technology is also PCI 2.2 compliant. This technology allows a PCI adapter to be added or removed while the server is up and running. Slots are powered down individually, impacting only the desired slot. The host-to-PCI bridge includes the electronics to control the PCI bus during these functions.

Peer-to-Peer Transactions

The host-to-PCI bridge also supports PCI peer-to-peer transactions. The host bridge allows communications between two devices on the same PCI bus segment. It also allows peer-to-peer communication across the I/O bus to PCI devices on other PCI bus segments (Figure 4).

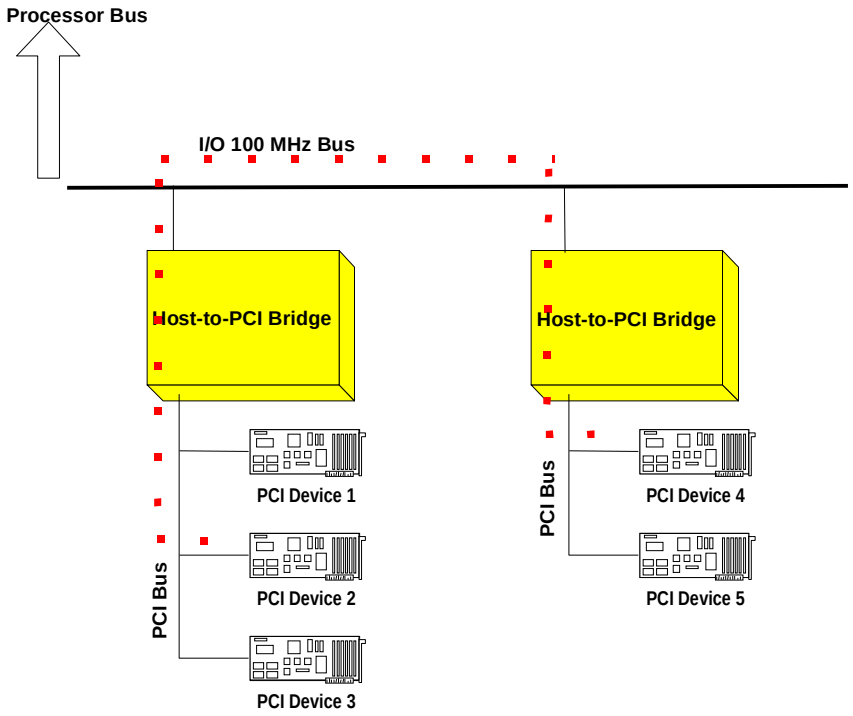

Figure 4: Schematic of peer-to-peer communications between PCI devices on different I/O bus segments

Microprocessors embedded in the PCI devices make these devices “intelligent” and perform the communication tasks previously handled by the processors. Data requests that normally go to the processor bus are handled by the host-to-PCI bridges. Off-loading work from the processors delivers overall performance improvements to the system.

PERFORMANCE AND SCALABILITY

The 8-way SMP architecture was designed to eliminate the bottlenecks of previous architectures, and to run on commercially distributed operating systems. Previous architectures were constrained by limited memory, inadequate IO bandwidth, or insufficient computing power. This architecture is capable of growing as the needs of the user increase. It will support up to 32 GB of SDRAM; four PCI buses, each with a maximum throughput of 533 MB/s; and one to eight Pentium® III Xeon™ processors.

Figure 5 shows an estimate of the relative online transaction processing (OLTP) performance of an SMP system with the Profusion architecture, based on the number of processors. With eight processors, the performance is expected to be about five times the performance of a single processor.

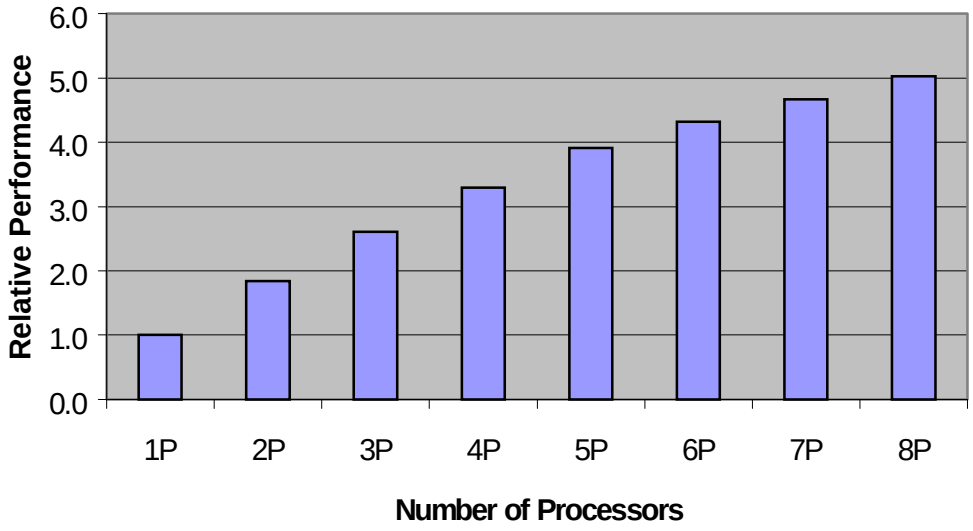

Figure 5: Relative OLTP performance of a Compaq 8-way SMP system, based on number of processors

The 8-way architecture offers the greatest expandability of Compaq architectures. One to eight processors can be installed in the same 8-way system. It can support up to 32 GB of SDRAM and multiple generations of Intel processors, including Intel's next-generation 32-bit processors.

Compaq 8-way SMP systems with the Profusion chipset provide better scalability of business applications such as SAP R/3 and Microsoft SQL Server for Windows NT. Figure 6 charts the scalability of a Compaq 8-way server with the Profusion chipset under an SAP/SQL7 workload. It compares a *Compaq ProLiant* 8-way server with the Profusion chipset and Pentium® III Xeon™ processors, to a *Compaq ProLiant* 7000 4-way server with the Intel 450NX chipset and Pentium® II Xeon™ processors. Performance of the new 8-way server scales 363 percent from one to four processors. Comparatively, performance of the *ProLiant* 7000 server scales 340 percent.

The increased throughput of the Profusion chipset in the 8-way system enables high-end business applications to scale and perform at levels never before realized under Windows NT. By keeping the number of coherency transactions to a minimum, Compaq can deliver an unprecedented 60 percent performance scaling from four to eight processors under an SAP/SQL7 workload.

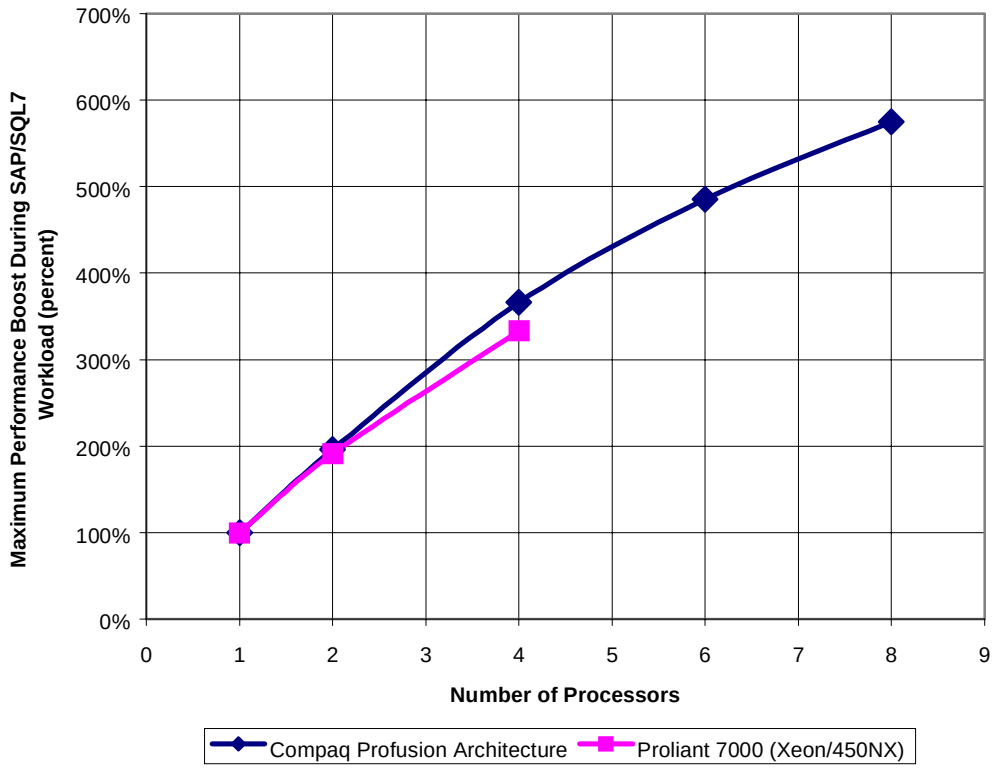

Figure 6: Comparison of processor scalability with SAP applications under Microsoft SQL Server for Windows NT

CONCLUSION

Compaq carefully evaluated the options for designing an 8-way SMP architecture and chose the optimum solution for the next-generation standards-based server. Through their technology exchange agreement, Compaq and Corollary are able to take full advantage of the Pentium® III Xeon™ processor in SMP servers. The combination of Compaq's I/O design, Corollary's Profusion system controller, and Intel's next-generation Pentium® III Xeon™ processor will yield a balanced, high-end performance system architecture for applications today and long into the future.

The Profusion chipset in 8-way systems enables high-end business applications and databases to scale and to perform at levels never before realized under commercially distributed operating systems. The increased performance and scalability are the result of improved architectural design, faster memory access with SDRAM, larger caches, and improved handling of coherency transactions.