COMPAQ
White Paper

Contents

Architectural Comparison of Compaq Evo Workstation W6000/W8000 with Compaq Workstation AP550/SP750

Abstract: Compaq is introducing new world-class computing platforms in its Evo Workstations W6000 and W8000 featuring the leading edge Intel® Xeon processor with the new Intel Hub Architecture i860. This white paper summarizes key technology advancements in the architecture of the new platforms compared with the previous generation Compaq Professional Workstation AP550 and SP750, which featured the i840 chipset.

Notice

The information in this publication is subject to change without notice and is provided "AS IS" WITHOUT WARRANTY OF ANY KIND. THE ENTIRE RISK ARISING OUT OF THE USE OF THIS INFORMATION REMAINS WITH RECIPIENT. IN NO EVENT SHALL COMPAQ BE LIABLE FOR ANY DIRECT, CONSEQUENTIAL, INCIDENTAL, SPECIAL, PUNITIVE, OR OTHER DAMAGES WHATSOEVER (INCLUDING, WITHOUT LIMITATION, DAMAGES FOR LOSS OF BUSINESS PROFITS, BUSINESS INTERRUPTION, OR LOSS OF BUSINESS INFORMATION), EVEN IF COMPAQ HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

The limited warranties for Compaq products are exclusively set forth in the documentation accompanying such products. Nothing herein should be construed as constituting a further or additional warranty.

This publication does not constitute an endorsement of the product or products that were tested. The configuration or configurations tested or described may or may not be the only available solution. This test is not a determination of product quality or correctness, nor does it ensure compliance with any federal, state or local requirements.

©2001 Compaq Computer Corporation. Compaq and Deskpro Registered with the United States Patent and Trademark Office.

Microsoft and Windows are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. Intel, Pentium, and Xeon are trademarks and/or registered trademarks of Intel Corporation in the United States and/or other countries..

Other product names mentioned herein may be trademarks and/or registered trademarks of their respective companies.

Architectural Comparison of Compaq Evo Workstation W6000/W8000 with Compaq Workstation AP550/SP750
White Paper prepared by Workstations Division Engineering

First Edition (April 2001)
Document Number 1527-0501A-WWEN

Introduction

With the introduction of the Compaq Evo Workstations W6000 andW8000, a new level of computing is defined. This white paper discusses the key new processor technologies and the advancements of the system core chipset incorporated in the Evo W6000/W8000 platforms. To better understand the system performance level of the new architecture and to highlight innovative architectural changes, the previous generation Compaq Professional Workstations AP550 and SP750 is used as a baseline. A comparison of key subsystem features and advancements alongside substantiated performance data will help explain why the W6000 and W8000 platforms are a natural choice for demanding workstation applications.

Compaq Professional Workstations AP550 and SP750

The Compaq Professional Workstations AP500 and SP750 feature dual Intel Pentium® III class processors and the Intel i840 chipset. A brief overview of the chipset is presented in the next section, while detailed architecture of the processor and chipset is covered later in this white paper.

Core Chipset Architecture Overview

Ever since the introduction of the BX chipset in 1998, Intel has changed the architecture of its core chipset to meet the demand for high bandwidth and high scalability in the rapidly changing computing environment. The newest architecture is called the Intel Hub Architecture. Compared to the more traditional architecture, often called Bridge Architecture, where processor, memory, graphics (Northbridge) and I/O (Southbridge) interfaces are tightly coupled with limited bandwidth, no consideration for data types, and shared I/O, the Hub Architecture features the following improvements:

- System bandwidth with concurrent data streams, thus allowing better system performance

- Real-time data sharing optimized for streaming and multitasking applications

- Dedicated I/O that enhances the performance of network and high-speed devices and has a greater degree of concurrency, scalability, and expandability

The Intel Hub Architecture consists of the following components:

- Memory Controller Hub (MCH) – interface to processor, memory, and graphics

- I/O Controller Hub (ICH) – interface to Peripheral Component Interconnect (PCI) bus and local accelerators for integrated controllers such as universal serial bus (USB), integrated drive electronics (IDE), Digital Audio, etc.

- RDRAM Memory Repeater Hub (MRH-R) – RAMBUS interface for memory expansion to meet system needs (4GB)

- 64-bit PCI Controller Hub (P64H) – interface to additional 64-bit 33/66MHz PCI bus segment to meet high performance I/O requirements

The Intel i840 Architecture

The Intel i840 architecture was the first Intel Hub Architecture optimized for dual Pentium III class processor systems. This chipset features the 82840 MCH, the 82801 ICH, the 82803 MRH-R, and the 82806 P64H. The MCH includes optimization for the 133MHz Pentium III processor bus, dual RDRAM memory channels, and an AGP-4X interface. Figures 1 and 2 provide an architectural overview of the AP550 and SP750, featuring the i840 architecture, respectively.
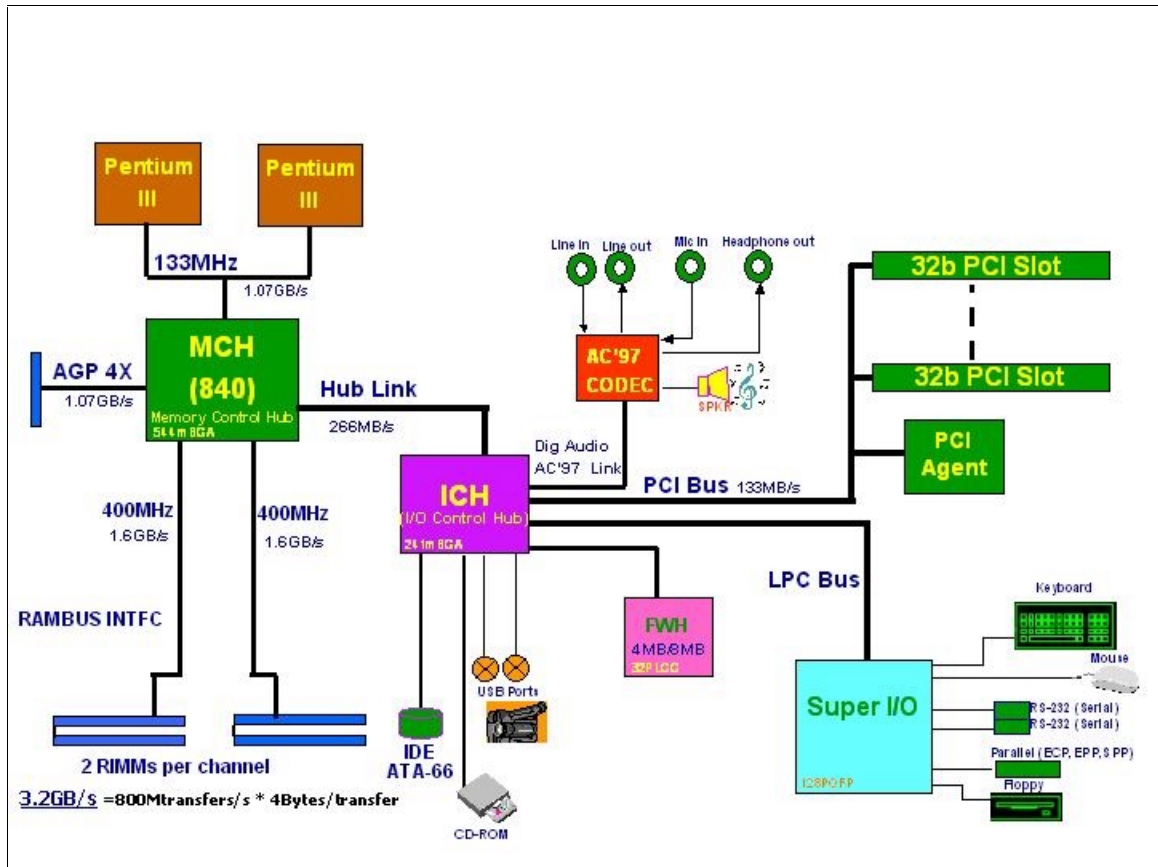

Figure 1: Compaq Workstation AP550 with Dual Pentium III — Architectural Overview
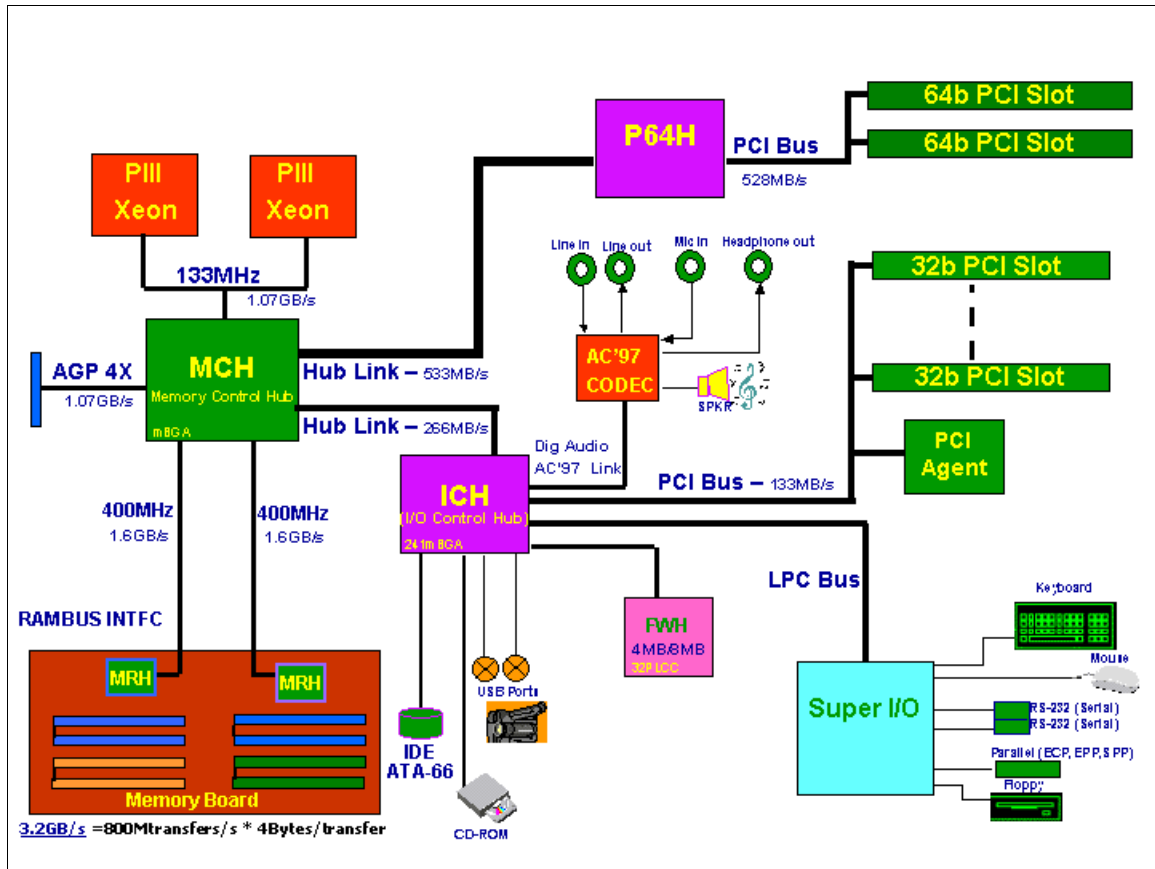
Figure 2: Compaq Workstation SP750 with Dual Pentium III Xeon — Architectural Overview

Compaq Evo Workstations W6000 and W8000

The Compaq Evo Workstations W6000 and W8000 incorporate the new leading-edge Intel Xeon processors and the Intel i860 chipset to break through to the next level of system performance. Detailed discussion of the Xeon processor and the i860 chipset, including performance advantages, follows below.

The Intel Xeon Processor

Representing a breakthrough to a new level of computing, the Xeon processor is a completely redesigned version of the earlier Intel IA32 processor architecture represented by the P6/Pentium III. The Xeon processor, however, maintains backward compatibility with existing applications. The architecture of a processor is the equivalent of a pool of resources that programmers can use to implement specific functions and algorithms — in other words, to build applications. The Xeon processor protects the current investment in existing applications by maintaining the instruction set, registers, and memory-resident data structure (the programmer resource pool) while providing new optimized instructions, registers, and data structures for future applications.

From a high-level perspective, the architecture of this new class of processor looks the same, while the underlying implementation is significantly enhanced to give better levels of performance in terms of frequency and instruction execution per clock. The true level of measured performance is how fast an application executes. This measurement is defined as

Performance = MHz (Frequency) x Instructions executed per clock (IPC)

The Xeon processor addresses the two variables in the performance equation with the new underlying silicon/logic implementation of the NetBurst micro-architecture. Figures 3 and 4, respectively, provide an overview of the micro-architectures of the Pentium III and Xeon processors.
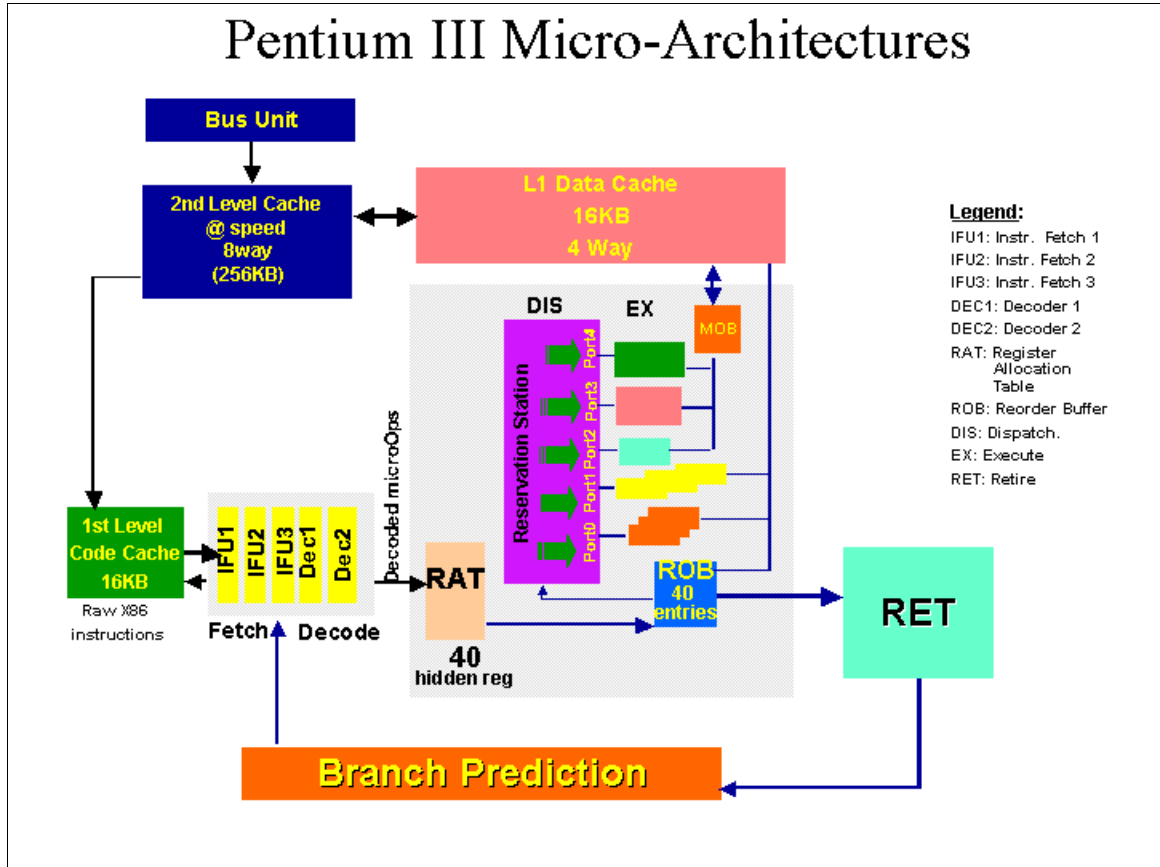
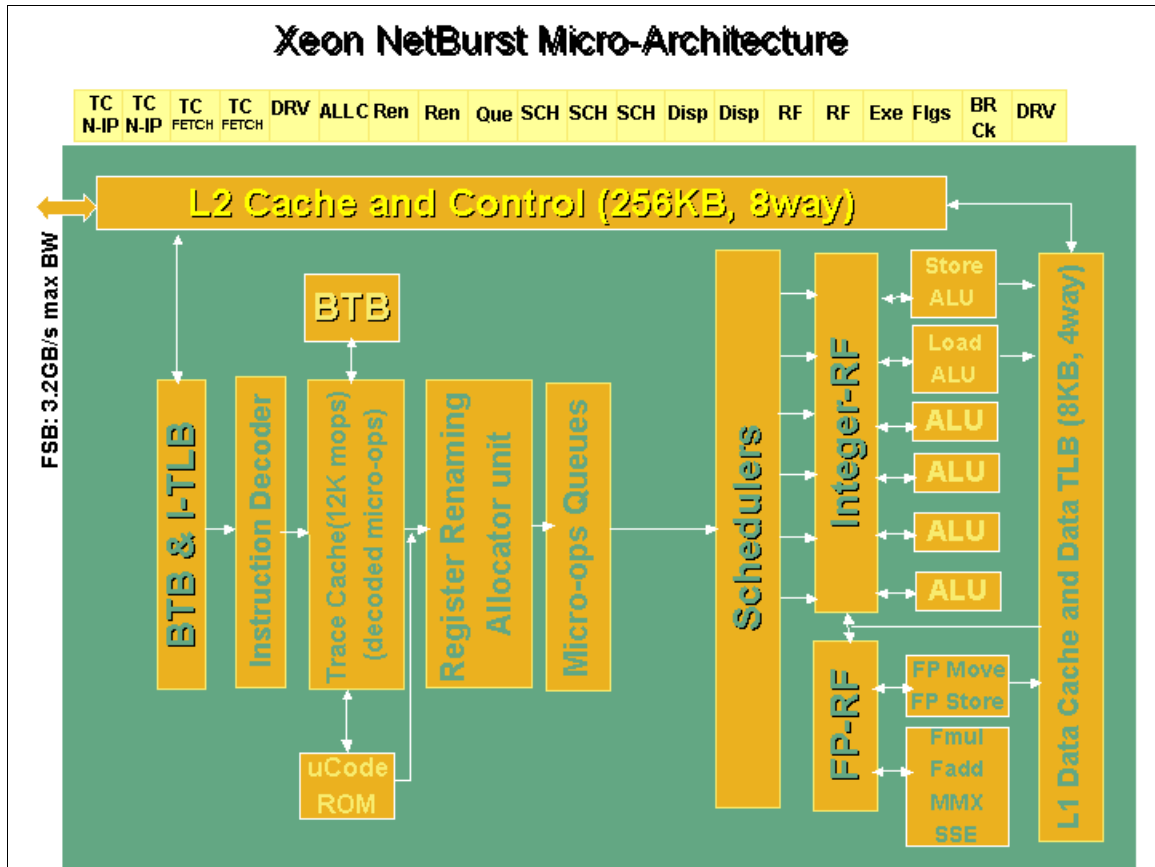**Figure 3: Pentium III Micro-Architecture Overview**

Figure 4: Xeon NetBurst Architecture Overview

The NetBurst micro-architecture attacks the frequency and IPC variables with its advanced $0.18\mu m$ silicon process technology, its redesigned architecture of the complete instruction pipeline, its execution engine, and its extension to the existing instruction set, which is as follows:

- 20-Stage Pipeline as compared to a 10-stage Pipeline in the Pentium III

- Execution Trace Cache to remove the long latency associated with the instruction decoder from the main execution loop in the Pentium III

- Rapid Execution Engine where multiple Arithmetic Logic Units (ALUs) are executed twice as fast as the core frequency, resulting in higher execution throughput, reduced execution latency, and extension of the total of execution ports to seven (7) as compared to five (5) in the Pentium III

- Advanced Transfer Cache with much higher throughput at 54.4GB/s for a 1.7 GHz Xeon (32 bytes x one transfer per clock x 1.7 GHz) to feed the data-hungry execution units as compared to 16GB/s throughput at 1 GHz in the Pentium III

- Advanced Dynamic Execution with very wide windows of instructions (126 instructions versus 42 instructions in the Pentium III) from which the execution units can choose to execute, thus avoiding dependency stalls that would prevent execution units from doing useful work. In addition, 4KB of branch target buffer (as compared to 1KB in the Pentium III), and a multilevel advanced branch prediction algorithm to keep detail on the history of past program branches, thus reducing by approximately 33% the mis-predictions rate as compared to the Pentium III.

- 400MHz System Bus with enhancements to signaling scheme and bus protocols, thus featuring data bandwidth and bus transfer efficiencies much higher than those of the Pentium III, as follows:

  - 200% data bandwidth improvement (3.2GB/s *(8 bytes x 400 Mtransfer/s)* versus 1.06 GB/s *(8 bytes x 133 Mtransfer/s)*)

  - 17% latency improvement for first critical data read

  - 46% latency improvement for 64-byte read

  - 25% latency improvement for data write

  - 64% latency improvement for 64-byte write

  - New cycles every two clocks at 200 MHz versus every three clocks at 133 MHz

  - 200% snoop bandwidth improvement (3.2GB/s *(64 bytes/2 clocks @ 100 MHz)* versus 1.06GB/s *(32 bytes/4clocks @ 133 MHz)*).

  - Higher concurrent requests

  - Faster interrupt servicing (bus message versus I/O cycles)

- Streaming Single Instruction Multiple Data Extension 2 (SSE2) with 144 new instructions that deliver 128-bit SIMD integer arithmetic operation and 128-bit SIMD Double-Precision Floating Point to reduce the number of instructions to complete a task or program, effectively increasing IPCs.

The Intel i860 Chipset

The Intel i860 chipset is a successor to the Intel i840 chipset that was designed in tandem with the new Intel Xeon processor to support up to two processors. Compaq has incorporated this chipset into the design of the Evo Workstations W6000 and W8000 to provide its customers with the many benefits arising from use of the chipset. Chipset features include optimization for the dual Xeon processor 400MHz system bus, dual RDRAM memory channels to balance memory bandwidth with front-side bus bandwidth, and AGP-4X graphics performance. These features are summarized in Figure 5.
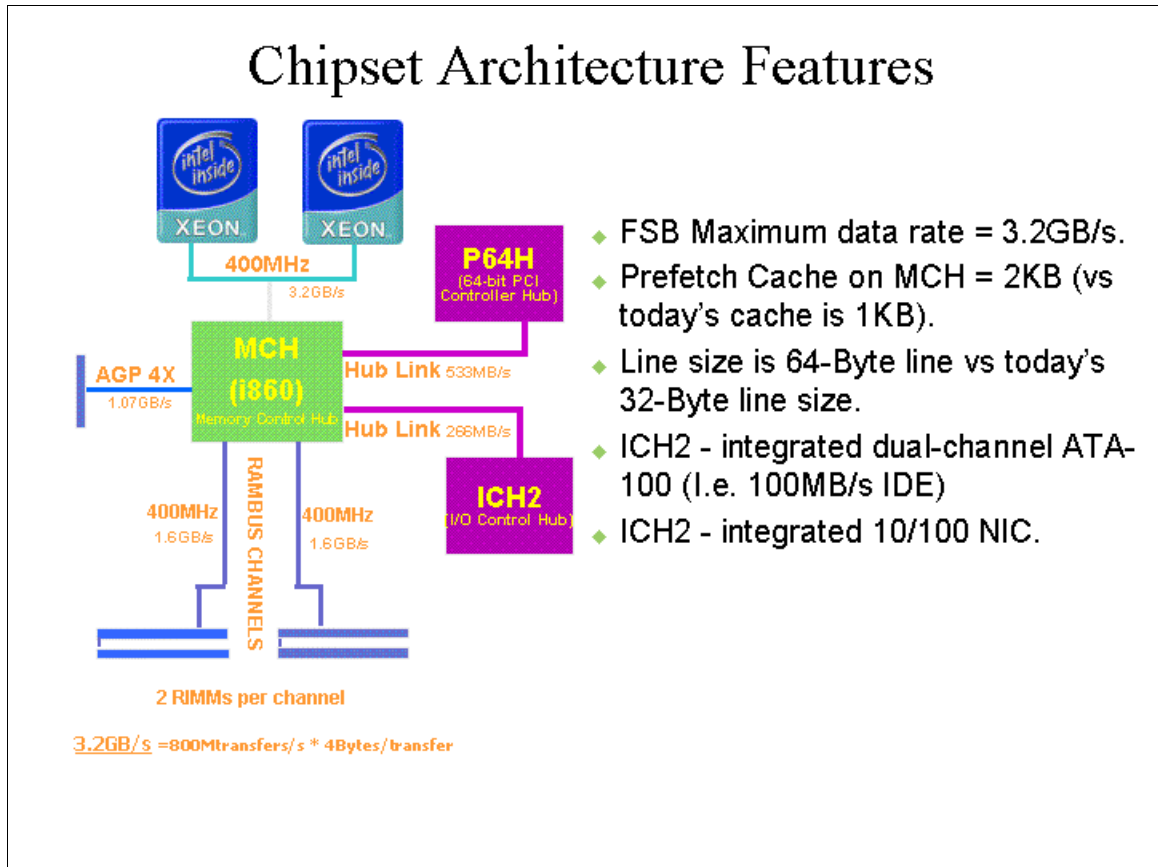
Figure 5: Intel i860 Architecture

To best meet customer needs, the Evo Workstations W6000 and W8000 are designed to have optimal price/performance characteristics while at the same time delivering superior scaling and expandability. Key architectural differences between the W6000 and the W8000 affect total system memory and I/O expansion capabilities. Both the W6000 and the W8000 provide feature-rich integrated devices such as 10/100Mbps LAN, Dual Channel ATA100 IDE, four (4) USB ports, advanced Audio, and seamless system management capabilities. The Workstation W6000 supports 2GB/s of system memory, and has one AGP-4x Pro slot and three PCI expansion slots. The Workstation W8000 supports 4GB/s of system memory, one AGP-4X Pro Slot, and six PCI slots partitioned in two independent segments of 32 bits/33MHz and 64 bits/66MHz. Figures 6 and 7, respectively, provide an overview of the architecture of the Compaq Evo Workstations W6000 and W8000.
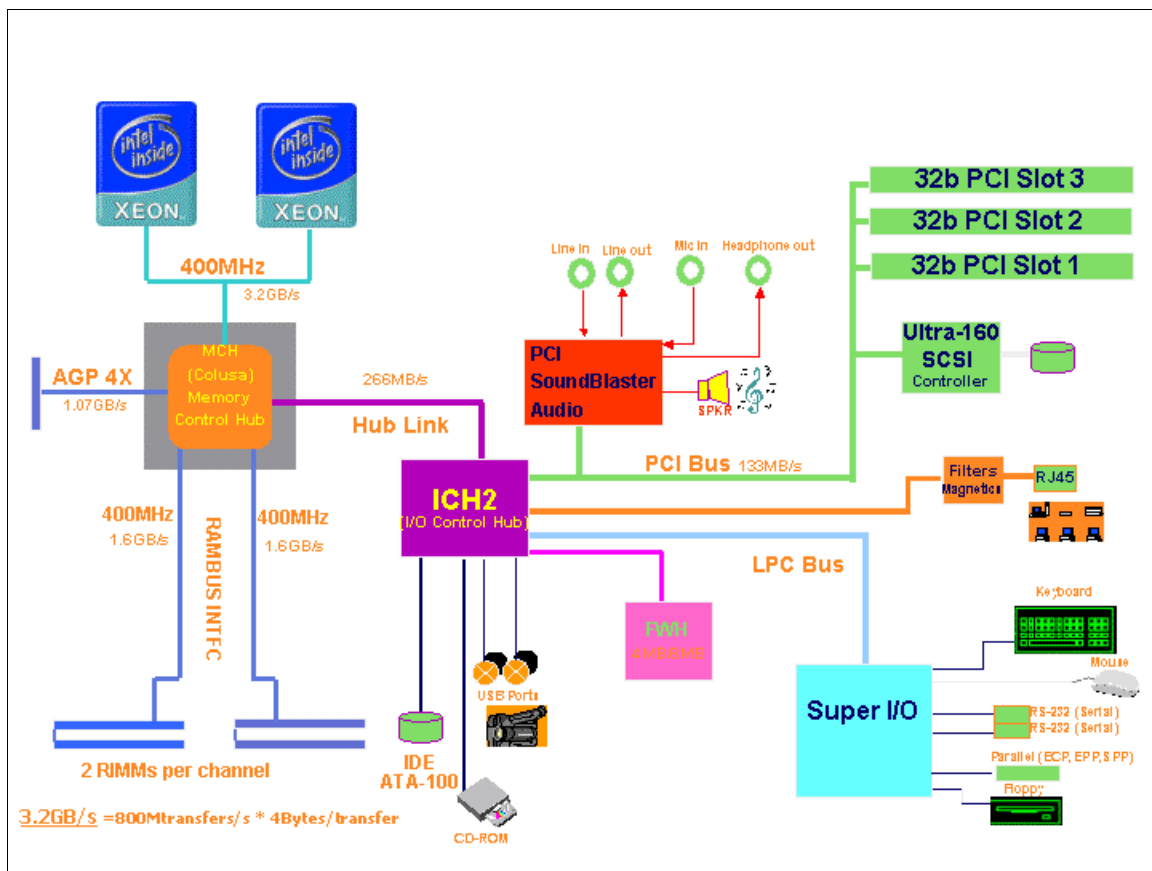
Figure 6: Evo Workstation W6000 with Dual Xeon — Architectural Overview
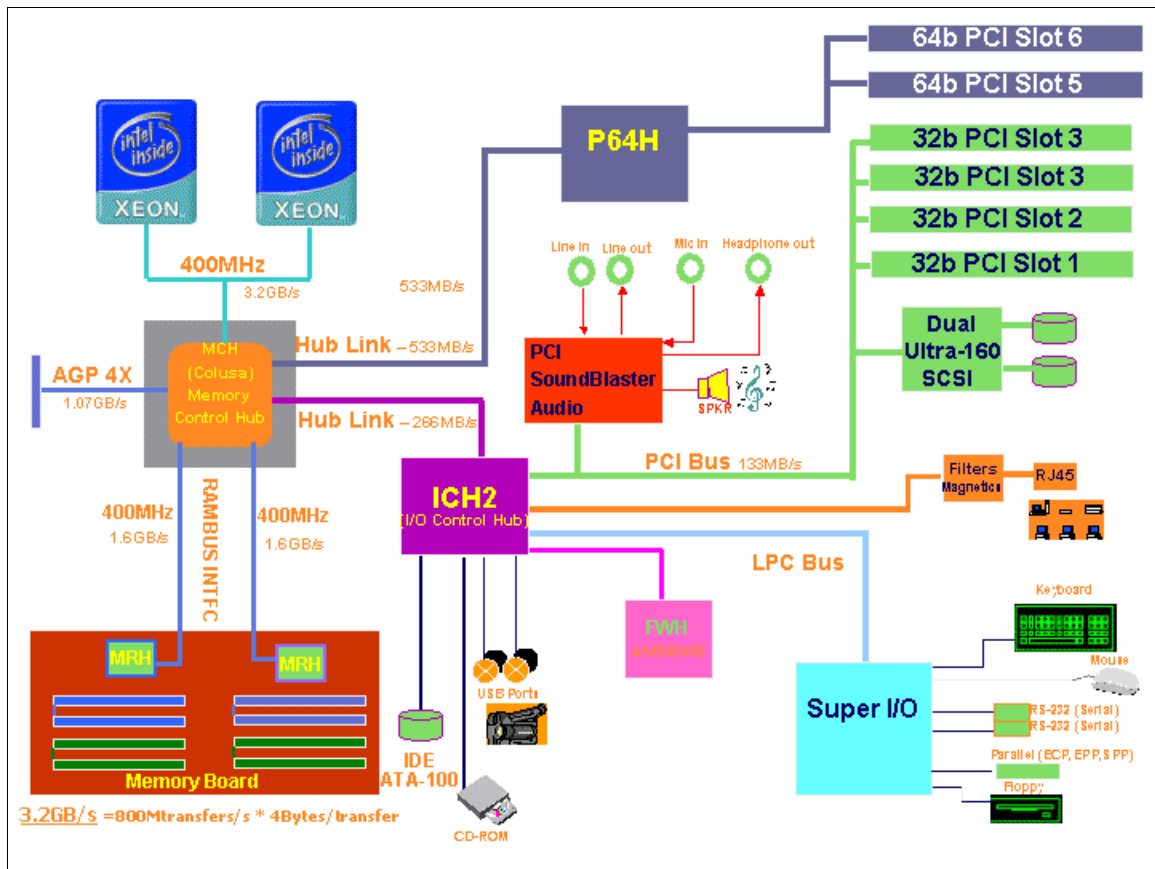
Figure 7: Evo Workstation W8000 with Dual Xeon — Architectural Overview

The Intel i860 chipset adds to the advancements of the Xeon processor micro-architecture an enhanced capability to provide overall balanced system performance with respect both to memory and to I/O.

As processor internal frequency increases, the importance of timely delivery of data increases in order to keep the processor pipeline available for useful work. The i860 MCH is optimized to interface with a dual Xeon processor system bus to deliver 400 Mtransfers/sec on a 64-bit data path. The result is 3.2GB/s bus bandwidth. This bus bandwidth is balanced with a dual-channel RAMBUS memory interface of 3.2GB/s. This high bandwidth delivers great performance for applications with poor data locality on L1 and L2 caches and for applications that require more data transfers from and to memory.

Given the balance of pure bus bandwidth, the system is often limited by how fast data gets to its destination. The governing parameter is called access latency. In the DRAM world, memory cells are not accessible until memory is opened. When a memory page must be opened, the latency in getting the first critical data to the processor is great compared to processor bus speed, thus minimizing processor bus utilization. To reduce this latency, the i860 memory controller is designed to store as much data as possible in its fast temporary buffers so that when the data is needed, it can be provided with a much lower latency. The practice of storing data in a temporary buffer in anticipation of future consumption is called *pre-fetching*, where the memory controller is making intelligent predictions that, if a data location at n is needed, then there is a very good chance that data at location $n+1$ is also needed — a concept similar to that of spatial locality in

cache design. The data delivery time from these pre-fetch buffers to the processor is less than half the time required when the needed data is still in memory (*6 clocks* versus *13 clocks*). The pre-fetch buffer in the i860 MCH is increased to 2KB compared to 1KB in the i840 MCH.

Where I/O is concerned, the partition in the Evo workstations integrates I/O devices through dedicated 266MB/s data pipes to system memory, handling concurrent multiple data streams as follows:

- Dual IDE channel delivery at a peak data rate of 100MB/s (ATA100) as compared to the previous-generation Workstation's rate of 66MB/s (ATA66)

- 24Mbps bandwidth across four USB ports as compared to 12Mbps in the previous-generation Workstation

- Dedicated 10/100Mbps LAN Interconnect Interface as compared to the PCI LAN in the previous-generation Workstation to provide better concurrency and reduce bus bottlenecks associated with PCI protocols

With the Workstation W8000, an additional dedicated 533MB/s data pipe to system memory is incorporated to handle fast I/O on 64-bit, 66MHz PCI expansion buses.

The concept of bringing data closer to consuming devices (pre-fetching) also applies to integrated and PCI expansion devices. Much higher sustainable I/O bandwidth is achieved with the Evo Workstations W6000 and W8000, compared to the earlier Compaq Professional Workstations, by using a more efficient pre-fetching mechanism in which 64 bytes of data (split into two halves) are pre-fetched from memory in anticipation of PCI (I/O) consumption. Pre-fetching starts when consumption of the data begins. This reduces the memory latency associated with waiting for the pre-fetch buffer to empty.

Other Technologies

Audio Enhancement

As compared to the AP550 and SP750 models, equipped with 2-channel audio and AC'97 codec, the W6000 and W8000 models are equipped with SoundBlaster Ensoniq PCI audio controller and CrystalClear SoundFusion Codec, capable of providing 64-voice WaveTable synthesizer, reverb/chorus/bass/treble sound effects, Ensoniq 3D Positional Audio algorithm, and S/PDIF (Sony/Philips Digital Interface) digital audio output for external speakers. This new platform brings true professional-quality audio.

Case for Performance

Applications generally can be divided into two classes: first, integer-based and basic office productivity applications, and second, floating-point-based workstation applications that are memory- and bandwidth-intensive. Recalling the performance equation mentioned earlier in this paper, the IPCs (instructions executed per clock) achievable by the above two classes of applications vary greatly due to the variation of branches in application code. This variation of branches affects the predictability of code flow. A higher probability of correct prediction yields a higher potential IPC. Floating point-based multimedia and vertical workstation applications tend to have branches that are very predictable and thus have a higher IPC potential. In addition, such applications lend themselves well to parallel execution. Using the new SSE2 extension of the Xeon processor, the higher the degree of parallelism achieved, the higher the potential IPC. As a result, these applications scale very well with frequency and benefit greatly from the new

architecture of the Evo Workstations. Integer-based and basic office productivity applications tend to have more random branches in application code, thus are more difficult to predict. The IPC potential is not high, but the frequency variable in the performance equation will benefit these applications. Using industry standard benchmarks, Figures 8 through14 illustrate the gains that the new architecture of the Evo Workstations W6000 and W8000 make possible for both vertical workstation applications and horizontal business applications.

SPEC CPU 2000

SPEC, the Standard Performance Evaluation Corporation, CPU2000 is an industry-standard benchmark for processor performance. This benchmark suite contains two components:

- The Integer Benchmarks (SPECint2000)

- The Floating Point Benchmarks (SPECfp2000).

Performance improvements for the Workstation W6000 compared to the Workstation AP550 against both these benchmarks are illustrated in Figure 8.
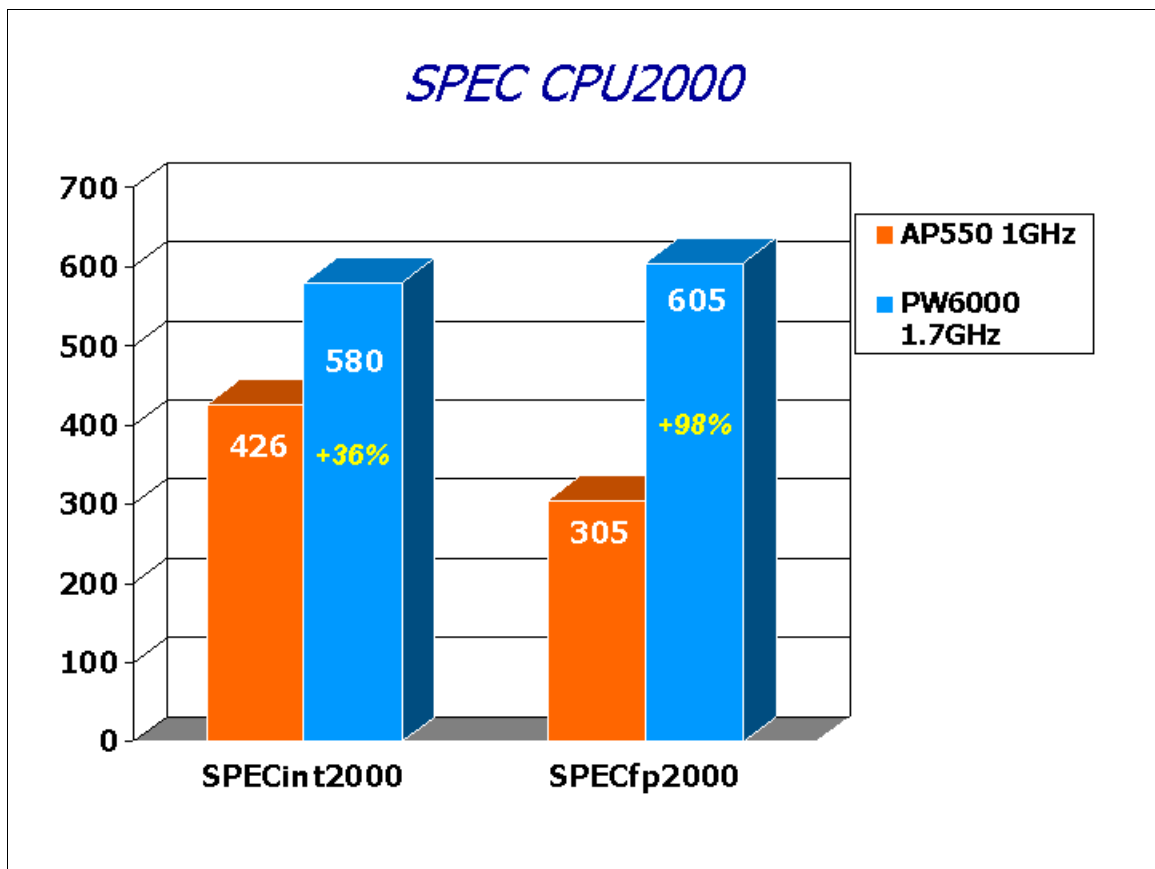

Figure 8: SPEC CPU2000 Xeon computing performance versus Pentium III

The performance gain in the integer benchmark is due to the redesign of the execution units in the Xeon NetBurst micro-architecture. Multiple ALU units (as was shown in Figure 4) are partitioned to handle simple and complex integer operations. Simple operations (Add/Subtract, Logic, Store Data, Branches) are executed at extremely low latency. Breaking down each of these operations into smaller workloads and staggering this workload at two times the main processor clock rate, these operations are completed effectively in ½ of a main clock cycle. 60-70% of typical integer

programs use these fast ALU units. Complex integer operations (Shift/rotate, integer divide/multiply) go to different ALUs for execution with longer latency.

Expanding the data path to 128-bit wide and a deep 128-entries floating-point register file, the two floating-point execution blocks (refer to Figure 4) can execute new operations every clock cycle. Floating point, MMX, SSE, and SSE2 instructions typically have operands from 64 to 128 bits in width. For the Pentium III processor, 128-bit operation is executed by two (2) 64-bit operations. In addition to fast execution units, the deep buffering (wide view) of in-flight instructions (126 micro-ops – see the Advanced Dynamic Execution discussion in the section titled *The Intel Xeon Processor*) allows the out-of-order execution hardware to examine large sections of programs at one time. This ability allows the processor to overlap long-latency instructions (floating point/SSE) and memory instructions by finding independent instructions to work on simultaneously. Floating-point applications usually have either large datasets that do not fit well in L1 data cache but do fit in L2 data cache, or that stream data directly from system memory. The improvement in cache bandwidth to 54.4GB/s and system-bus improvement at 3.2GB/s improves performance of these applications significantly. These improvements in the Xeon micro-architecture result in significant gains in the SPEC floating point benchmark, such as illustrated in Figure 8.

STREAM

STREAM is a synthetic benchmark program that measures sustainable memory bandwidth and the corresponding computation rate for simple vector kernels. The benefits of all the advancements in key technologies incorporated in the Workstations W6000 and W8000 are shown by improvements measured by this benchmark, as illustrated in Figure 9:

- data transfer bandwidth (COPY – measures transfer rate without arithmetic)

- execution units integer/floating point

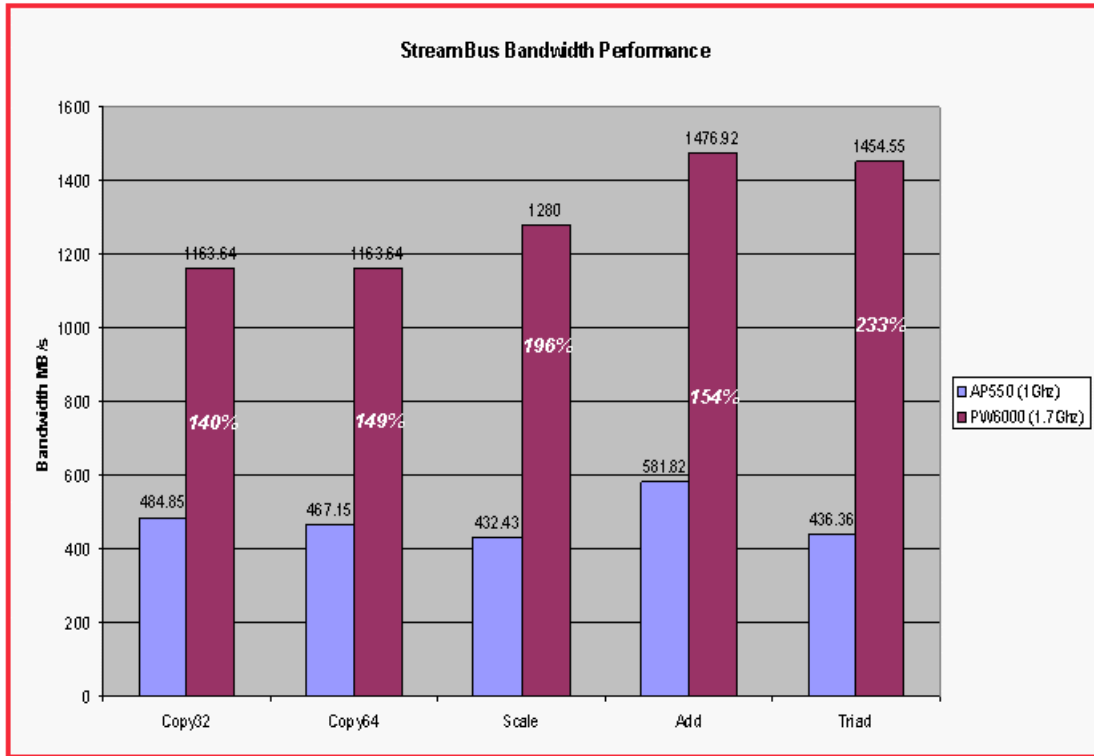- load/store (ADD, SCALE, TRIAD).

Figure 9: W6000 bus bandwidth performance versus AP550 using Stream

SPECapc PRO/E and 3D StudioMax

With an increase in floating point performance, memory bandwidth, and instruction extension for computationally intensive graphics, the benefits to Mechanical Design Automation and Digital Content Creation applications with the W6000/8000 are illustrated with the SPECapc PRO/E (Figure 10) and 3D StudioMax (Figure 11) benchmarks, respectively.
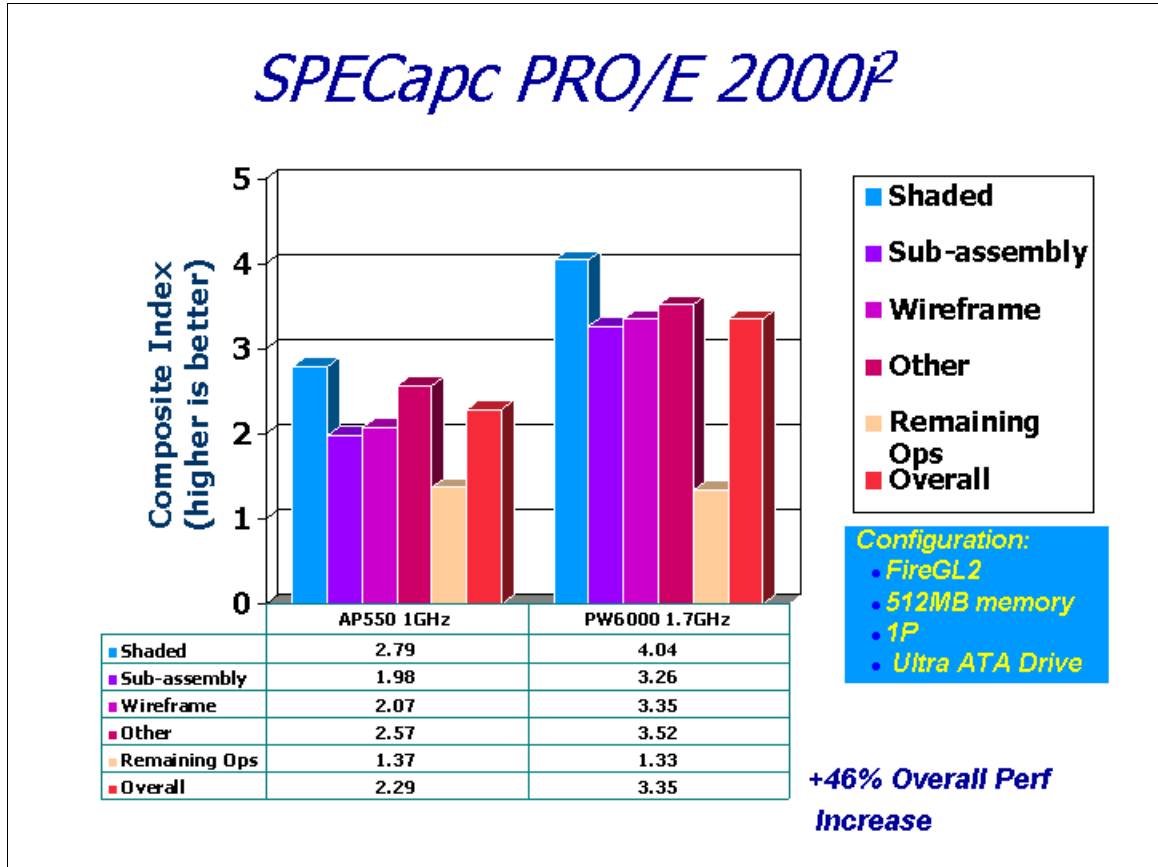
Figure 10: SPECapc (Application Performance Characterization) PRO/E
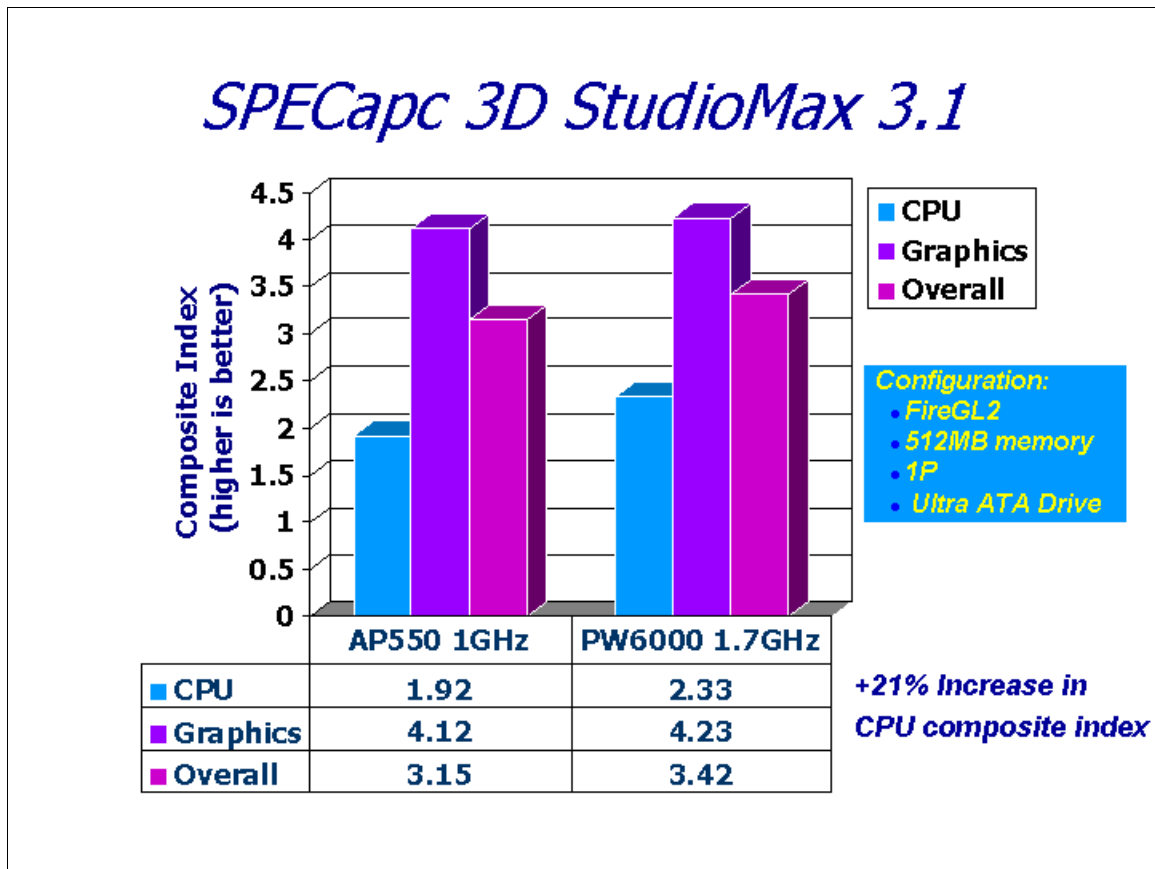
Figure 11: SPECapc 3D StudioMax 3.1

The degree of realism in computer animation is achieved by a great number of computational tasks. Some of these tasks are done in the graphic engine itself, while others utilize system CPU. To display 3D objects on a 2D computer screen, it is much easier to represent 3D objects as a collection of polygons (usually triangles) than as curved surfaces. The larger the number of triangles used to represent the 3D object, the more closely the approximation of the mathematical description resembles the 3D object. The process of breaking up 3D object into triangles is called *tessellation* and involves an enormous number of floating-point vector calculations. Objects in the real world have material properties and reflectivity and these impact how the objects interact with light. The more lighting from various sources and angles, the more realism to the object/scene. Again, calculations of light effects on 3D objects require large numbers of complex floating-point vector calculations. The CPU index performance gain in the SPECapc 3D StudioMax benchmark resulted from the increase in floating-point performance of the Xeon processor.

Impact of DirectX 8.0

Optimized usage of SSE/SSE2 extension and code flow optimization to take advantage of the new NetBurst micro-architecture, allow graphic drivers to make use of DirectX 8.0 programmable vertex and pixel shaders to produce significant performance gains as illustrated in Figure 12.
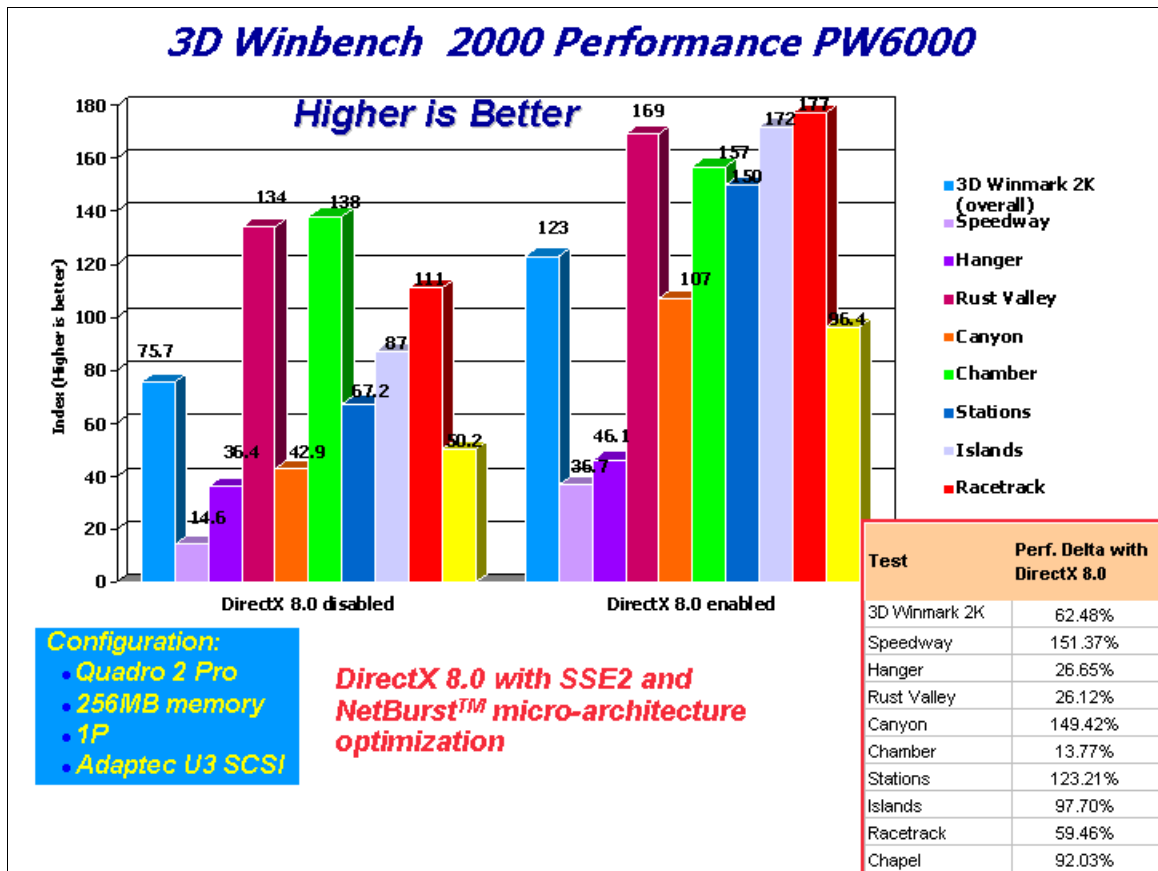
Figure 12: DirectX8 Performance Improvements

Improvements in Productivity Applications

As mentioned earlier, productivity applications have low IPC potential but frequency gains provide a performance boost to these applications as illustrated in the Business Winstone 2001 benchmark illustrated in Figure 13. Equipping systems with multiple processors provides significant benefit to computing and responsiveness applied to these applications as shown in Fig 14.
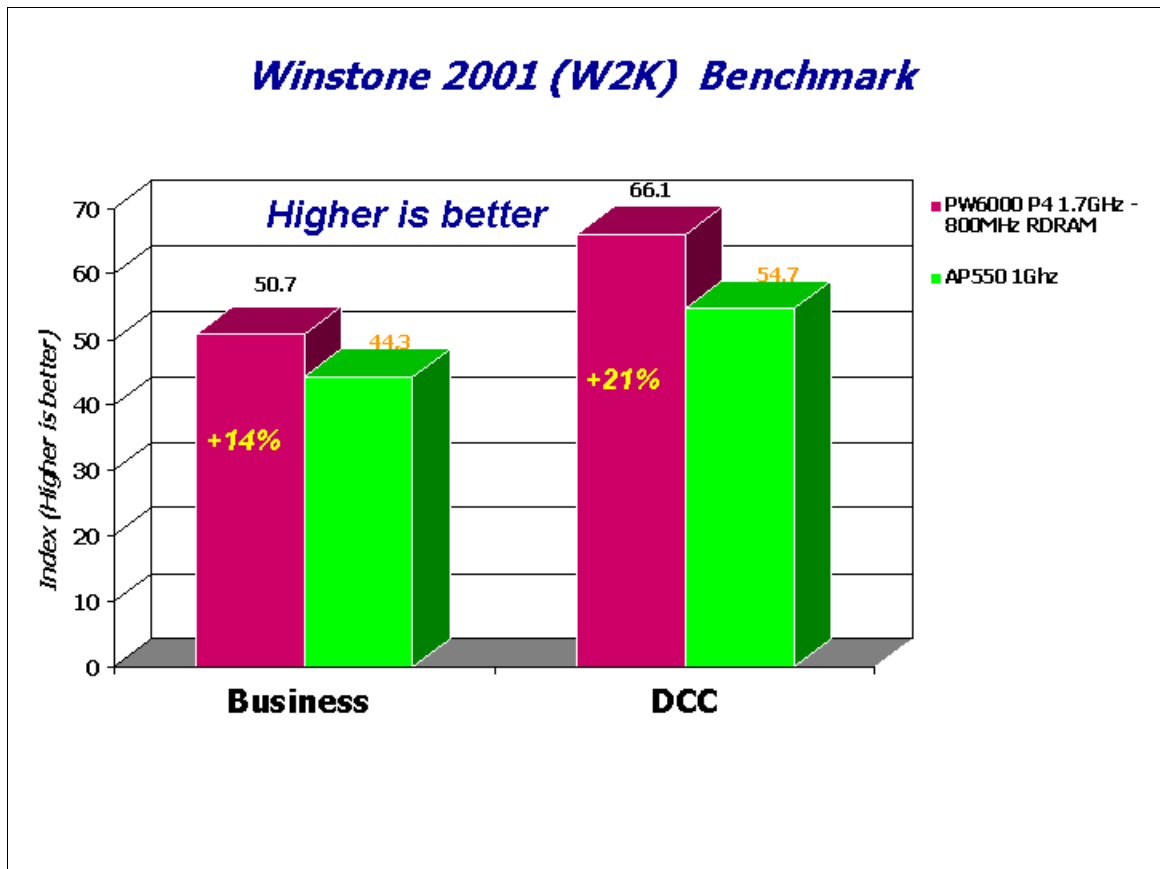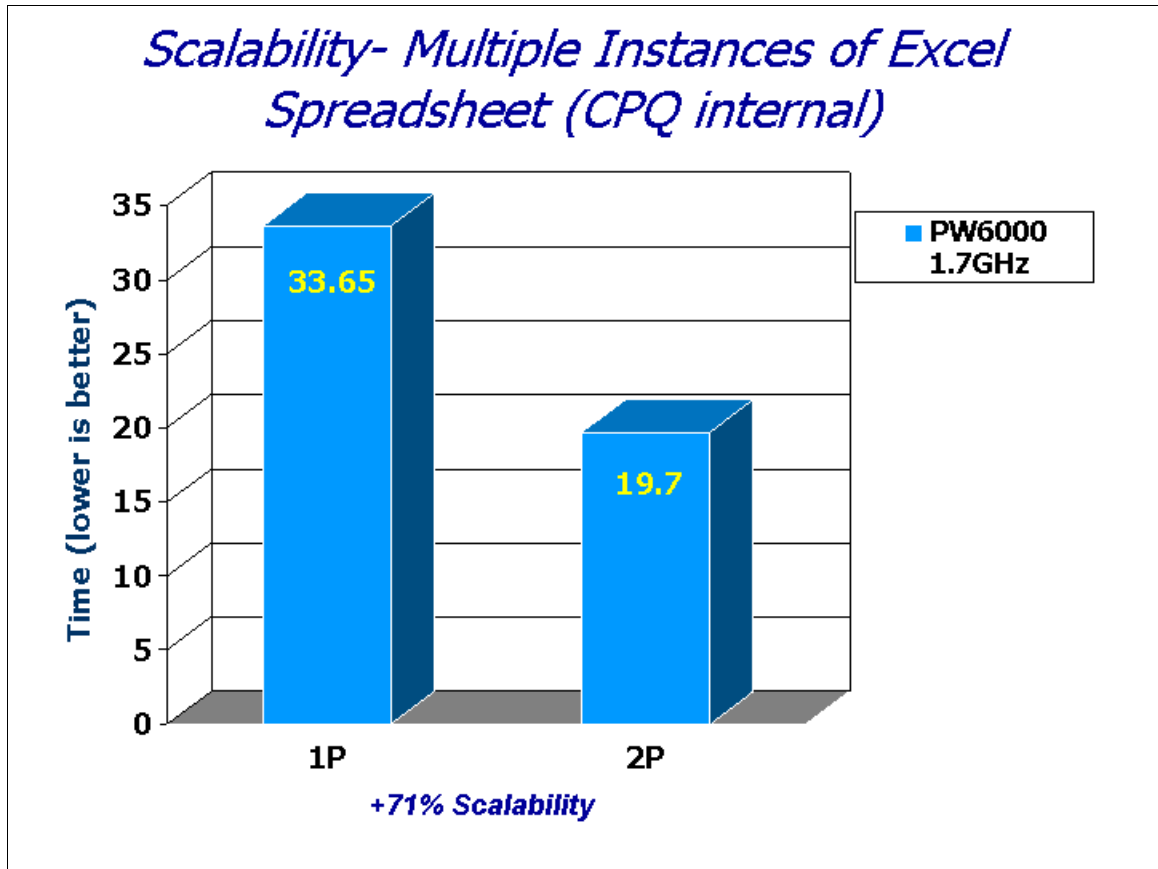
Figure 13: Horizontal Winstone Benchmark

Figure 14: Benefit of dual processor system

Summary

The Compaq Evo Workstations W6000 and W8000 incorporate the new Intel Xeon processor and features that complement innovations in the processor micro-architecture. Equipped with a 400MHz dual Xeon processor system bus and balanced with dual-channel RDRAM, these platforms offer feature-rich I/O with optimized multiple concurrent data streams through dedicated pipes to system memory. These innovations lead to breakthroughs in performance that are measured and substantiated by testing reported in this white paper. The Evo Workstations are the natural choice for the demanding workstation marketplace.