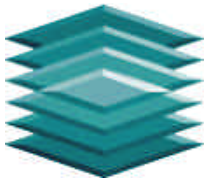
Unisys ES7000 Challenges Sun E10000



September 2000

A Special Evaluation Report
Prepared for Unisys by:

D.H. Brown Associates, Inc.
222 Grace Church Street
Port Chester, NY 10573
(914) 937-4302
Fax: (914) 937-2485

TABLE OF CONTENTS

Unisys ES7000 Challenges Sun E10000

EXECUTIVE SUMMARY

With the availability of Microsoft's Windows 2000 Datacenter Server operating system, the Unisys e-@ction Enterprise Server ES7000 now supports its largest configuration of thirty-two processors. This achievement invites an architectural comparison to one of the leading UNIX-based SMP servers, Sun's Enterprise 10000 (Starfire).

With its compelling combination of advanced features and industry-standard Intel hardware, D.H. Brown Associates, Inc. (DHBA) rates the ES7000 above the E10000. In particular, the ES7000 exhibits its greatest strength in price/performance, but also excels in memory and I/O capacity, while the E10000 leads in scalability. The following table summarizes the results on a function-by-function basis. (The ++ rating represents a strong advantage. The + represents a modest advantage, while = represents parity).

TABLE 1: Unisys ES7000 vs. Sun E10000

Functional Area	Unisys ES7000	Sun E10000
Processors	+	
Cache Coherency	+	
Interconnect	=	=
Memory	+	
Partitioning	+	
Upgradability	=	=
Scalability		+
I/O Capacity	+	
RAS	=	=
Price/Performance	++	
Overall	**+**	

The Unisys ES7000 hardware design and implementation demonstrates superiority over the E10000 in six out of ten categories as reviewed in detail in the body of this report. The advantages can be attributed to the large systems experience of the Unisys engineers, complemented by recent technological improvements.

Some salient questions remain with regard to the ES7000, including the dependency of some important functions on software that is not yet available. The Itanium processor may provide a boost in performance, but there will be a

considerable time lag before many applications are ported over to its full capabilities, since the porting of customer applications to 64 bits will consume time and resources. Some ES7000 functions, such as dynamic partitioning, dynamic CPU, and memory detach and add depend on future releases of the Windows 2000 operating system.

Further, with many installations, the Sun E10000 and its Solaris operating system are known entities and offer a proven track record. Performance is a matter of record, and many applications are available. It is our opinion that Sun will improve price/performance of the E10000 with the UltraSPARC III and other enhancements in the future. However, at this time, the features of the Unisys ES7000 entice a close look by prospective users.

DHBA has examined the ES7000 system and the underlying Cellular Multiprocessing (CMP) architecture since 1998. DHBA's October, 1999 evaluation compared the ES7000 against Windows NT clusters and SMPs, as well as UNIX clusters and SMPs.[1] This report updates DHBA's previous assessment by concentrating on architectural comparisons with Sun's E10000.

[1] *Unisys CMP Blends SMP and Clustering for High-End Windows*, D.H. Brown Associates, Inc., October 1999

SYSTEM ARCHITECTURE

The Unisys ES7000 and Sun 10000 both represent powerful systems suitable for enterprise computing requirements. Since many of the differences between them relate to choices in the underlying system architecture as well as implementation decisions, it is useful to provide a basic overview of the design points for these two high-end servers.

ES7000 SYSTEM ARCHITECTURE

Figure 1 depicts a rough block diagram of the ES7000 design. A pair of Intel processors share a bus; two pairs of processors share a third level (L3) cache. The L3 is 16 MB for the IA-32 processors and 32MB for the IA-64 processors. This subassembly of four processors plus L3 cache is a configurable unit that can be assigned independently to a partition. The configurable I/O subassembly contains 3 PCI buses, each with four slots, for a total of 12 PCI slots per direct I/O bridge connection to the crossbar. Memory units, physically packaged in four subassemblies, each contain multiple SDRAM banks to provide high bandwidth. A large crossbar switch interconnects the processors, I/O, and memory. Note that Unisys' crossbar is implemented with a passive midplane, without active circuitry on this crucial portion of the system.

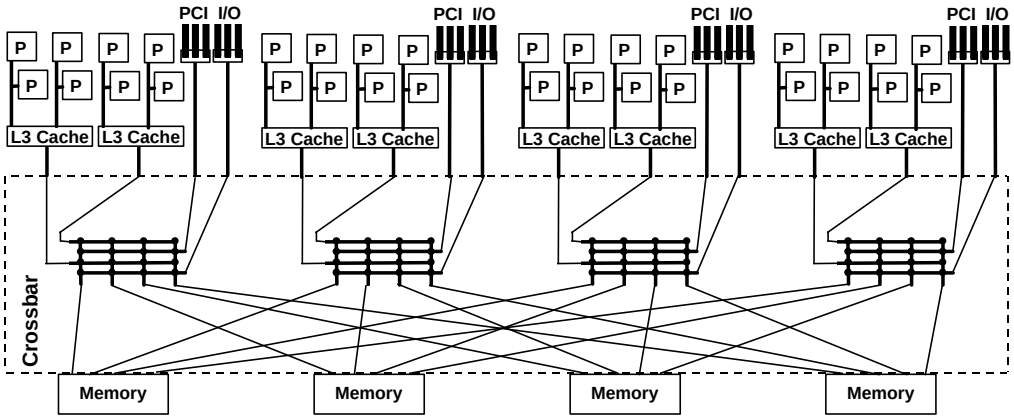


FIGURE 1:
ES7000 Architecture Diagram

E10000 SYSTEM ARCHITECTURE

As seen in Figure 2, the E10000 "system board" contains four 400 MHz UltraSPARC II processors, each with 8 MB L2 cache, up to 4 GB of memory and two 64-bit SBuses, each supporting two SBus connectors. On-board address and data paths implement the Ultra Port Architecture (UPA), providing the "local" connection among the processors, memory and I/O channels. Up to sixteen system boards can be configured in a system cabinet giving a maximum of sixty-four processors, 64 GB of memory and thirty-two I/O processors (sixty-four I/O controllers).

To access the memory or I/O of other boards, the E10000 uses the next level of the interconnect hierarchy, the Gigaplane-XB, a combination of interleaved address buses and a crossbar data switch. For a system as large as the E10000, reliance on a single bus for data and addresses could limit scalability. Note that Sun's crossbar design employs active circuitry on its centerplane, potentially a source of component failure.

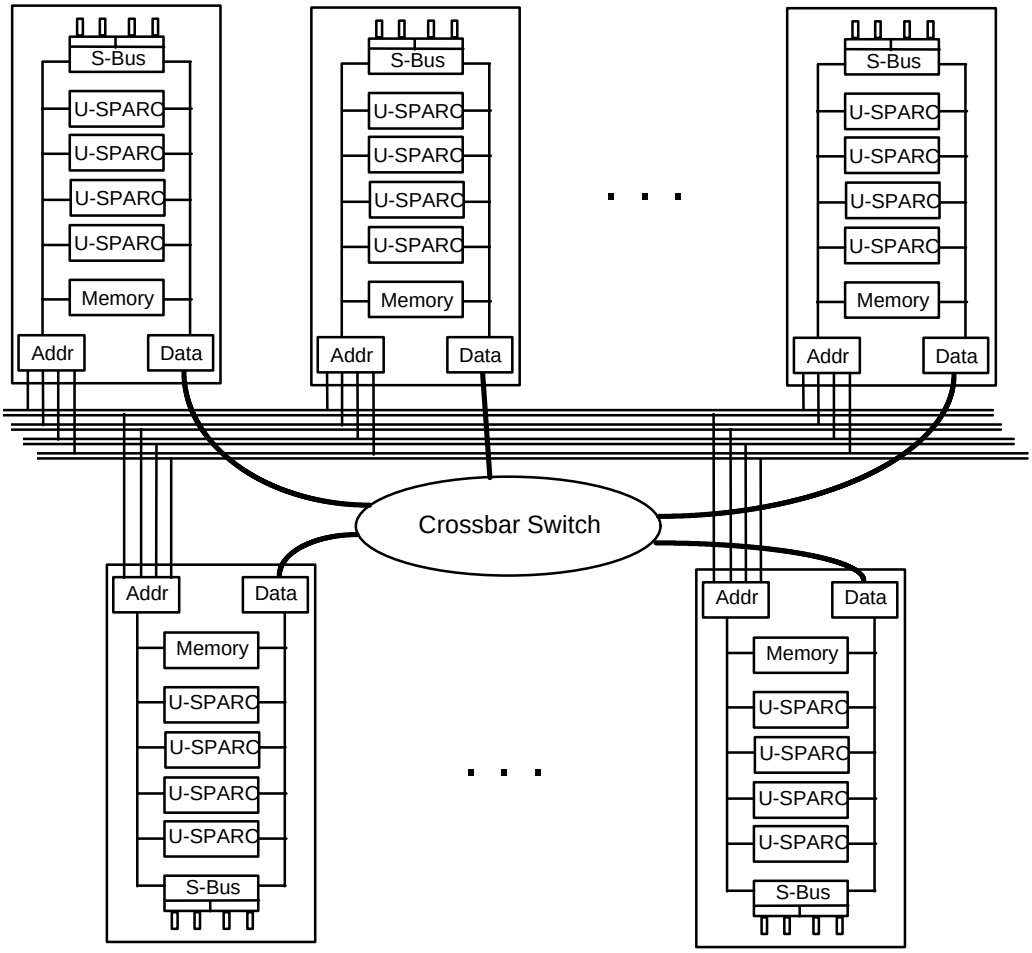


FIGURE 2: Starfire Bus/ Switch Hierarchy

PROCESSORS

The ES7000 has a lead over the E10000 for processor technology due to current performance advantages and rapid upgrades typically offered by Intel technology. The E10000 is powered by Sun's proprietary microprocessor, the UltraSPARC II, which currently operates at 400 MHz. Upgrades of the UltraSPARC have been relatively infrequent, causing a lag versus most other microprocessors. (Note, Sun now presents a roadmap for the future of UltraSPARC that is much more aggressive and promises to equal the plans of the other vendors.) Unisys, on the other hand, embraced Intel technology not only for its rapid advances, but also for the abundant software available for Windows NT. A drawback for Windows NT has historically been that it could not support large SMP hardware. However,

the availability of Windows 2000 Datacenter Server largely neutralizes this limitation. The ES7000 currently incorporates powerful 550 MHz, 700 MHz Pentium-III Xeon processors and plans to incorporate 900 MHz Xeon and faster IA-32 processors, such as the Foster. The ES7000 was designed to upgrade readily to the 64-bit Itanium processor anticipated to be available in volume early in 2001. (Performance implications are discussed later in this report.)

The ES7000 supports a mix of processors. For instance, different frequency Pentium processors, or Pentium and Itanium processors can coexist in a single system. This permits gradual upgrades to newer technology on an as needed basis and at a reasonable price. To upgrade to faster processors, the E10000 must upgrade all processors at the same time – an expensive proposition for a large E10000 system.

Both the ES7000 and E10000 processors have a 16KB instruction cache and a 16KB data cache. However, the E10000 Level 2 (L2) cache at 8MB exceeds the maximum L2 option of 2MB for processors in the ES7000. To compensate, the ES7000 shares a Level 3 cache among four processors. Pentium-III Xeon L3 cache size equals 16MB, while the Itanium L3 cache size expands to 32MB.

CACHE COHERENCY

The ES7000 enjoys a slight advantage over E10000 for cache coherency. All SMP systems must implement a cache coherency mechanism to avoid corrupting data which are accessed by multiple processors at the same time. Executing these procedures requires resources that can affect performance, particularly as the number of processors scale up. The ES7000 employs a directory-based coherency scheme for high performance. This eliminates the need to broadcast each address to all processors and wait for responses as required by snoopy protocols. Its efficiency allows for scalability without performance degradation. By contrast, the E10000 implements the “snoopy” protocol by using the four address buses of its Gigaplane-XB interconnect to broadcast addresses to all the memory banks. Although the addresses are rapidly propagated, the time lag until all responses are received likely affects performance as the number of processors approaches the maximum of sixty four.

INTERCONNECT

The ES7000 and E10000 have equally strong interconnects. With almost the same data bandwidth, both systems' interconnects adequately supply data transfer needs and maintain constant latency. Without a rapid interconnect with sufficient throughput to supply data to the processors and I/O, the system may become bottlenecked.

As shown in the ES7000 diagram, the interconnect consists of four 4X4 crossbar switches with memory on one side and processor caches and I/O on the other. The 1.6GB/second peak bandwidth of each switch path gives 25.6GB/second total peak and 20GB/second sustained throughput. The eight paths between the switch and L3 caches sustain a rate of 10GB/second for the Pentium Xeon processor and 13 GB/second for the Itanium processor, while the I/O capacity is about half at a sustained five GB/second (800MB/second peak for each set of three PCI busses). The crossbar switch can supply these maximum capacities simultaneously. The point-to-point nature of the crossbar provides fast, consistent latency for all transfers.

On the E10000, the Gigaplane-XB, an active centerplane, implements a crossbar (XB) switch to provide the high data bandwidth of a direct point-to-point interconnect. Sixteen bytes wide, the crossbar operates at 100 MHz to synchronize with the processor clock rate (4 * 100 MHz = 400 MHz.). Sustained data rate has been measured as 10.5 GB/sec. The address bus, not requiring quite as high a bandwidth, takes a different approach. Since "snoopy" protocols are used to maintain cache coherence, all system boards need to observe all addresses. Consequently, four independent interleaved address buses, associated with the four memory banks, spread the address traffic while enabling fast bus snooping at the rate of two cycles per address on each bus.

MEMORY

The ES7000's memory implementation is somewhat superior to that of the E10000. Applications today demand more and more memory to operate efficiently. Lack of sufficient memory limits a system's ability to respond in a timely fashion, especially in the heavy on-line environment prevalent in today's data environment.

Both ES7000 and E10000 systems support a maximum memory of 64GB, but the E10000 must service sixty-four processors rather than the thirty-two of the ES7000. In a maximum configuration, the ES7000 thus provides twice as much memory per processor. With memory independent of the processor modules, the ES7000 readily adds as much memory as required regardless of how few processors are present in the configuration. The E10000 carries the memory on the system board with the processors. Memory can be placed on a system board with no processors, but this option increases the cost of additional memory considerably.

PARTITIONING

Due to its more flexible partition configurability, the ES7000 has a slight advantage over the E10000. The ability to split a large system into smaller isolated portions permits multiple versions of software, operating systems, and applications to run concurrently. This allows users to test new operating systems and applications without affecting production partitions. In addition, it provides an excellent vehicle for server consolidation.

The E10000 led the way in introducing dynamic partitioning, long a mainframe feature, to UNIX SMP systems. Labeled Dynamic System Domains, the basic unit of a domain is a system board. Domains can contain as few or as many boards as required: configuring all boards in a domain results in a single domain or standard system. The E10000 supports up to sixteen domains. The domains achieve complete isolation while sharing the interconnect by having registers on the system board define the domain. All the components controlled by the system board become part of the domain it occupies, including processors, memory and I/O. Failures within the domain do not affect other domains, but failure of a systemwide component such as the control board, or some crossbar failures, will bring the whole system down. Domain configurations may be reconfigured dynamically with the system running, a leading edge feature. Data can be shared between domains using the interconnect for rapid transfer. Note that some inter-domain interconnect failures could affect both communicating processors and so this mechanism is not intended for high availability failover configurations.

Partitioning of the ES7000 allows great flexibility. Up to eight partitions may be configured. The quantity of processors, memory and I/O in a partition can be selected by the user with only the limitations that the four processors sharing an L3 cache must all reside in the same partition and a PCI subassembly of three buses must remain in the same partition. If multiple programs are running in a partition, their resources may be adjusted dynamically (within the partition). Support for dynamically changing resources among partitions will be achieved by the next major release of Windows 2000 Datacenter Server. Another significant feature of the ES7000 partitioning is selective memory sharing. The ES7000 memory system is designed to support a combination of memory that is private to a partition, while at the same time allowing a portion of memory to be shared among partitions. Selective memory sharing allows memory-coupled applications that depend on sharing data. Initial release of this feature limits the shared area to one 256MB section of memory accessible by two specific partitions using messaging. As noted above, the E10000 communicates rapidly over the Gigaplane between domains by means of messaging. This allows any domain to contact any other domain. Future releases of ES7000 software will expand the memory sharing function flexibility and allow data access by direct memory addressing.

UPGRADABILITY

The ES7000 and E10000 have roughly equivalent upgradability. When choosing a system, users frequently consider the ease and cost of upgrades, which involves adding processors, memory, and I/O to an existing system. However, it can also involve moving to newer, faster equipment. If the existing system accommodates a component upgrade, the user benefits accordingly.

The absence of configuration linkage between memory and processors on the ES7000 permits adding more of each easily and economically. With memory and processors both on the system board of the E10000, memory intensive applications could require adding system boards with no processors on them just to increase the memory. This increases the price of the memory significantly.

The ES7000 supports the coexistence of various speeds or even generations of processors. Thus, both Pentium Xeon and Itanium processors can coexist in the same system. Users may add a partition to test and develop for the Itanium processors while production on Pentium Xeon processors continues in another partition. In the E10000, by contrast, all processors must be at the same frequency. To upgrade to faster processors, all processors in the system must be replaced. Because the system clock and the processor clocks are synchronized, the processor clocks must be an integer multiple of the system clock. Currently, the E10000 system clock runs at 100 MHz to support 400 MHz processors. If faster UltraSPARC IIs clock at a rate that is not a multiple of 100 MHz, the crossbar clock may need to be slowed down to remain in synchronism.[2] Furthermore, to take best advantage of the next generation UltraSPARC III, a new server design may be required. If so, upgrading from an UltraSPARC II system to an UltraSPARC III system would require a box swap.

The ES7000 requires addition of processors in increments of four. By contrast, the E10000 allows addition of individual processors, allowing greater granularity. To facilitate upgrades of the E10000, Sun also offers a "Capacity On Demand" (COD) program. For a small premium, additional resources are included in the initial installation but are not used or paid for until required by the customer. This "terms and conditions" approach to upgrades has been popular with E10000 customers and apparently meets their needs. If this strategy becomes desirable for ES7000 users, Unisys could implement such a plan in the future.

[2] A recent TPC-H submission indicates that Sun plans 466 MHz processors within the next six months (to meet TPC submission rules). While Sun has not revealed the speed of its crossbar with the future 466 MHz UltraSPARC, the new chips might force the crossbar to slow to 93.2 MHz (93.2x5=466).

SCALABILITY

The E10000 scalability surpasses that of the ES7000, providing Sun with an advantage against this criterion. Growth to sixty-four processors gives the E10000 double the scalability of the ES7000. Furthermore, if Sun equalizes processor performance in the near future, E10000 capacity should exceed the ES7000.

I/O CAPACITY

The ES7000 holds an edge over the E10000 in terms of I/O capacity. Both systems have an abundance of I/O channels, but ES7000 offers ninety-six PCI, while the E10000 has sixty-four proprietary SBus or up to a maximum of thirty-two industry standard PCI. The advent of huge disk farms and ever-faster online connections increases the importance of a large number of I/O channels for enterprise class servers. Configuring dual channels for disks to increase availability also creates demand for more I/O channels. The processing power of large SMPs requires sufficient I/O flow in order to avoid restricting the rest of the system.

RELIABILITY, AVAILABILITY AND SERVICEABILITY (RAS)

ES7000 and E10000 have roughly equivalent RAS capabilities. Many systems today operate in a 24x7 mode making RAS of primary importance. When errors and failures are detected, the malfunctioning components must be deactivated and replaced quickly while the system continues running. To achieve this, systems must be able to provide redundant components. The design of both ES7000 and E10000 systems incorporate an abundance of these features. The inclusion of partitioning greatly enhances the availability of both systems. However, certain features favor one or the other of the systems.

The ES7000 includes ECC throughout the whole data path, while the E10000 limits checking to parity in the processor caches. Recent articles reported problems due to the L2 caches in the UltraSPARC II, but no cases were cited specifically for the E10000. Sun plans to install mirrored L2 caches by the end of the year to alleviate the problem. Also, the ES7000 supports the memory-write protection provided by the Intel chipset to avoid corrupting memory. This feature is not available on the E10000.

When failure is detected, the E10000 automatically inactivates memory or CPUs. Additional memory and CPUs are recognized dynamically and used after the

operator designates their domain assignment. The ES7000 requires operator intervention for both procedures.

Sun's E10000 design employs an active centerplane containing the circuitry for its Gigaplane-XB data crossbar and quad address buses. The design does allow degraded operation after failure – half the data crossbar and one or more of the address buses can be disabled at reboot. However, repair of this critical component will affect the entire system. Unisys' ES7000 employs a passive midplane that is less likely to fail.

PERFORMANCE AND PRICE/PERFORMANCE

On a pure price/performance basis, the ES7000 proves to be much less expensive than the E10000. Industry standard benchmarks such as TPC or SAP R3 provide the best means of comparing system performance. Even though these are imperfect and reflect specific environments, they are the best metrics available. Unisys has not reported results for any standard benchmarks at this time and bases its comparisons on internal modeling. Most vendors adopt the same technique. The only hint of a comparison entails looking at SPECint, a benchmark that measures the CPU alone. Based on Intel processors relying on a smaller L2 cache, the ES7000 clearly provides greater CPU performance than the UltraSPARC II, although the exact amount cannot be determined.

To assess price/performance, DHBA considers a range from equal to double for the Pentium processor versus the UltraSPARC. Our analysis indicates that the Pentium processor provides more performance. The Tables below show a performance range based on number of processors. Table 2 assumes similar performance and configures both systems with the same number of processors. Table 3 reflects the power of the Pentium and requires the Sun system to be configured with twice the number of processors for equal performance.

TABLE 2: Equal Processor Performance

System	Processors	Memory	I/O Ports	Price
ES7000	32	32GB	96	$815,000
E10000	32	32GB	64	$1,631,000

TABLE 3: Equal Performance with Expanded Sun Configuration

System	Processors	Memory	I/O Ports	Price
ES7000	32	32	96	$815,000
E10000	64	32	64	$2,634,000

These estimates produce a range where the E10000 varies from 2.0 to 3.2 times the price of the ES7000 for similar performance. They also appear to validate the better price/performance achievable with commodity components.

SYSTEM SOFTWARE

Microsoft will reach a critical milestone in its vision of distributed computing with the expected September 26 announcement of Windows 2000 Datacenter Server, the high-end version of its server operating system. With Datacenter Server, Microsoft plans to attack the heart of commercial server space, targeting its competitive efforts in particular on leading UNIX implementations such as Sun's Solaris. A head-to-head comparison of Windows 2000 Datacenter Server and Solaris functions is beyond the scope of this report, however, several key tradeoffs are noted below.

Users have historically been attracted to Windows NT's affinity with PC desktops, specifically, its consistency compared with the relatively fragmented UNIX landscape, and its rich set of highly integrated functions. However, UNIX vendors were persistently able to blunt Microsoft's enterprise efforts by proving that its systems could offer superior reliability and scalability. Thus, Microsoft made scalability and reliability improvements primary design points in Windows 2000.

As a result, Windows 2000 Datacenter Server introduces support for thirty-two processor SMP systems and 64 GB of physical memory. This dramatically increases the scalability headroom for Windows 2000 applications and helps enable the deployment of Windows 2000 solutions in enterprise environments using advanced hardware such as the ES7000. Windows 2000 Datacenter Server also introduces useful administrative tools for datacenter environments, such as the Process Control Manager, which provides additional flexibility in managing critical workloads.

In Windows 2000 Datacenter Server, Microsoft has also focused specifically on improved reliability, which had represented one of the most pressing limitations in earlier versions of Windows NT. To address the yet higher levels of robustness required in datacenter environments, Windows 2000 Datacenter Server builds on the basic reliability improvements introduced in the initial release of Windows 2000, which sharply reduced the number of administrative restarts required and implemented more robust management of shared DLLs. In particular, Microsoft recognized that many of Windows NT's earlier instabilities were due to the extraordinarily broad array of hardware it tried to support. Thus, Windows 2000 Datacenter Server limits the hardware platforms supported to a set of high-end systems specifically itemized by Microsoft and key OEMs. The constrained set of platforms supported allows Microsoft to work more closely with partners such as Unisys to arrive at far more robust solutions, without sacrificing the flexibility gained by the use of industry-standard components.