

## Fujitsu Boosts PRIMEPOWER Performance

Richard Partridge, VP and Research Director • October 22, 2002

This month, Fujitsu announced new models of its PRIMEPOWER servers, scaling to 128 SPARC-compliant processors. Fujitsu already markets up to 128-way systems; the new models are packaged more densely, with eight processors per board compared to the four processors per board of the original design. Furthermore, the new systems will employ Fujitsu's latest SPARC64 V processors, operating up to 1.3 GHz. In addition to higher performance, the SPARC64 V processors incorporate advanced error handling features leveraged from Fujitsu's mainframe design expertise.

Table 1 highlights the capacities of PRIMEPOWER models that offer SPARC64 V processors, scaling as large as 128-processor SMP configurations. Competitive systems typically scale to about half as many processors, or degrade I/O performance by replacing PCI cards with auxiliary processors. PRIMEPOWER's balanced design, on the other hand, scales up the number of processors, memory capacity, and I/O capabilities without tradeoffs.

**Table 1: New/Enhanced PRIMEPOWER Models Offering SPARC64 V**

PRIMEPOWER Models	Number of Processors	Chip Clock Rate	Memory	PCI Slots
<i>Compute Nodes (Midrange)</i>				
650	8	1.08 GHz	32 GB	20
850	16	1.08 GHz	64 GB	40
<i>Enterprise Servers</i>				
900	16	1.35 GHz	64 GB	36
1500	32	1.35 GHz	128 GB	72
2500	128	1.35 GHz	512 GB	320

Not shown in this table are Fujitsu's SPARC64 GP platforms. Extending downward to workgroup servers, the PRIMEPOWER 200 and 400 offer, respectively, two and four 700 MHz SPARC64 GP processors. Existing PRIMEPOWER models 600, 800, and 1000 scale to 128 SPARC64 GP processors, at up to 788 MHz. These models continue to be supported, although they are being supplanted by the new SPARC64 V systems. The recently announced models 900, 1500, and 2500 offer not only more performance but also a more compact packaging than their predecessors and thus will be the preferred solution for enterprise customers.

Worldwide, Fujitsu has been a long-time player in high-end computing. It is well known throughout the Asia/Pacific region, having offered mainframe and supercomputing systems for nearly fifty years. In Europe, the Fujitsu-Siemens joint venture has enhanced the visibility of Fujitsu's SPARC/Solaris platforms. Stretching back to the 1970s with its Amdahl mainframes, Fujitsu has also had a presence in the United States. However, within the U.S., its UNIX servers have achieved only minor penetration compared to the strong market share of Sun's SPARC/Solaris offerings. Now, with the introduction of the new PRIMEPOWER models, the U.S. arm, Fujitsu Technology Solutions, Inc. (FTSI), is intensifying its marketing efforts with the goal of significantly growing its U.S. customer base.

**D.H. Brown Associates, Inc.**

[www.dhbrown.com](http://www.dhbrown.com)

*Our research program in Enterprise Servers makes this Technology Trends available to all our subscribers. Those interested in this program should contact [marketing@dhbrown.com](mailto:marketing@dhbrown.com).*



# Fujitsu Boosts PRIMEPOWER Performance

October 22, 2002

The current economic slowdown might seem an inopportune time to challenge entrenched vendors with a new product. However, PRIMEPOWER does not require customers to adopt a new RISC architecture or a new operating system. Instead, PRIMEPOWER runs the widely used Solaris Operating Environment (OE) on processors that comply with the open SPARC V9 standard. That is, PRIMEPOWER offers a price/performance alternative for SPARC/Solaris customers, a feature that customers do welcome during tight economic times.

## SPARC64 V

The Fujitsu-designed SPARC64 processor is an important piece of PRIMEPOWER differentiation. Licensees of SPARC, an IEEE Standard overseen by a non-profit organization, have implementation freedom, such as incorporating circuitry that increases performance or enhances reliability. For SPARC64, Fujitsu focuses on increasing instruction-level parallelism, in order to get more work done per clock cycle. Specialized instruction decode and execution-scheduling hardware analyzes data dependencies and potential instruction parallelism for code segments, allowing up to six execution units to remain busy in SPARC64 V's out-of-order superscalar design.

Table 2 illustrates Fujitsu's processor design emphasis by comparing recent SPARC64 chips as well as UltraSPARC III from Sun. By using a design that executes more instructions in parallel, SPARC64 can achieve high scores on the SPECint2000 (847) and SPECfp2000 (1205) benchmarks without resorting to extreme clock frequencies. Note that Fujitsu focused on enhancing its floating point performance and nearly doubled its SPECfp2000 result in moving from SPARC64 GP to SPARC64 V. Fujitsu had long offered a custom vector design VPP Series of supercomputers. By enhancing SPARC64 V floating point performance, Fujitsu will now target large PRIMEPOWER configurations, instead of its VPP Series, for the High Performance Computing (HPC) market.

**Table 2: SPECint/fp and SPEC/MHz**

Processor	Fujitsu SPARC64 V	Fujitsu SPARC64GP	Sun UltraSPARC III	Sun UltraSPARC III
Frequency	1,350 MHz	810 MHz	1,050 MHz	900 MHz
System	PRIMEPOWER 900	PRIMEPOWER 850	Sun Blade 2050	Sun Fire V880
SPECint2000	847	617	610	507
SPECint/MHz	0.63	0.76	0.58	0.56
SPECfp2000	1,205	613	827	713
SPECfp/MHz	0.89	0.76	0.79	0.79
Test Date	9/02	6/02	11/01	5/02

Certainly, benchmark bragging rights often change hands as vendors release ever faster parts using the latest fabrication technology advances. Sun has already announced a 1,250 MHz UltraSPARC III (but has not yet formally measured SPECint/fp) and indicates that

even faster chips are coming. Fujitsu's chip road map also promises periodic speed increases.<sup>1</sup> Even though the two vendors' SPARC implementations will likely leapfrog each other's benchmark results, the fact remains that Fujitsu's SPARC64 implementations will continue to offer a legitimate alternative for SPARC/Solaris computing.

## PRIMEPOWER Performance

Powerful processors do form the foundation for high-end systems. At the same time, a high-performance interconnect is needed to join those processors to create a large, shared-memory, multiprocessing server. PRIMEPOWER engineers leveraged Fujitsu's mainframe and supercomputer interconnect skills to devise such an interconnect – a high-bandwidth, low-latency crossbar that scales to efficiently accommodate midrange to high-end systems. For its top end 128-way model 2500, Fujitsu claims aggregate bus bandwidth of 133 GB/second. While such a figure sounds impressive, customers are more concerned with how such “speeds and feeds” translate into application performance.

Considerable effort is required to obtain performance measurements on benchmarks that approximate large application environments. Such benchmark evidence is often not available until well after the announcement of a new platform. Fujitsu indicates it is working on benchmarking the new SPARC64 V-based PRIMEPOWER models, but does not have results as yet. There are formal benchmark results from previous PRIMEPOWER models that did showcase the scalability of the prior design. Tables 3 and 4 highlight results for existing high-end PRIMEPOWER models as measured on SAP SD and TPC-C. Note that the tests of Tables 3 and 4 were run using 563 MHz and 675 MHz chips, which offered about half the integer performance compared to the latest SPARC64 V processors.<sup>2</sup> Table 5, SPECweb99, shows that an 810 MHz SPARC64 GP reports best performance for an eight-way system. Given the increased performance of SPARC64 V, measurements on the new PRIMEPOWER models should prove to be quite competitive on these commercially oriented tests. (In addition, with the dramatic jump in floating point performance, SPARC64 V-powered PRIMEPOWER servers will likely fare well on technical benchmarks.)

**Table 3: SAP R3 Two-Tier Sales and Distribution (SD)**

Vendor	System No. of Server CPUs x Proc. (MHz)	Steps/Hr.	Database	Version	Date
Fujitsu	PRIMEPOWER 2000 128x SPARC64 GP (675)	2,345,000	Oracle9i Solaris 8	4.6C	1/02
Fujitsu	PRIMEPOWER 2000 64x SPARC64 GP (675)	1,263,000	Oracle9i Solaris 8	4.6C	2/02
IBM	p690 32x POWER 4 (1300)	1,250,000	DB2 7.2 AIX 5.1	4.6C	4/02
Sun	Fire 15K 76x UltraSPARC III (900)	1,242,000	Oracle 8.1.7 Solaris 8	4.6C	9/01

# Fujitsu Boosts PRIMEPOWER Performance

October 22, 2002

Table 4: TPC-C, Non-Clustered

Vendor	System Database	CPUs x Proc. (MHz)	tpmC \$/tpmC	Date
Fujitsu	PRIMEPOWER 2000 SymfoWARE 3.0.1	128x SPARC64 GP (563)	455,818 \$28.52	8/01
HP	Superdome Oracle9i	64x PA8700 (875)	423,414 \$15.64	8/02
IBM	p690 Oracle9i	32x POWER 4 (1300)	403,255 \$17.80	8/02
HP Compaq	AlphaServer GS320 Oracle9i	32x Alpha 21264A (1001)	230,533 \$44.62	6/01

Table 5: SPECweb99

Vendor	System No. of Server CPUs x Proc. (MHz)	Steps/Hr.	Software	Date
IBM	p690 16x POWER 4 (1300)	21,000	Zeus 4.0 AIX 5.1	10/01
HP	rp8400 16x PA-8700 (750)	15,000	Zeus 3.4.3 HP-UX 11i	8/01
Fujitsu	PRIMEPOWER 800 12x SPARC64 GP (675)	11,223	Zeus 4.0.1 Solaris 8	11/01
Fujitsu	PRIMEPOWER 850 8x SPARC64 GP (810)	10,110	Zeus 4.0.1 Solaris 8	9/02
IBM	p660 6M1 8x RS64 IV (750)	10,000	Zeus 4.0 AIX 5.1	3/02
HP	rp8400 8x PA-8700 (750)	9,186	Zeus 3.4.3 HP-UX 11i	9/01
Sun	Fire 4800 12x UltraSPARC III (750)	8,739	iPlanet 6.0 Solaris 8	4/01

## PRIMEPOWER Partitioning

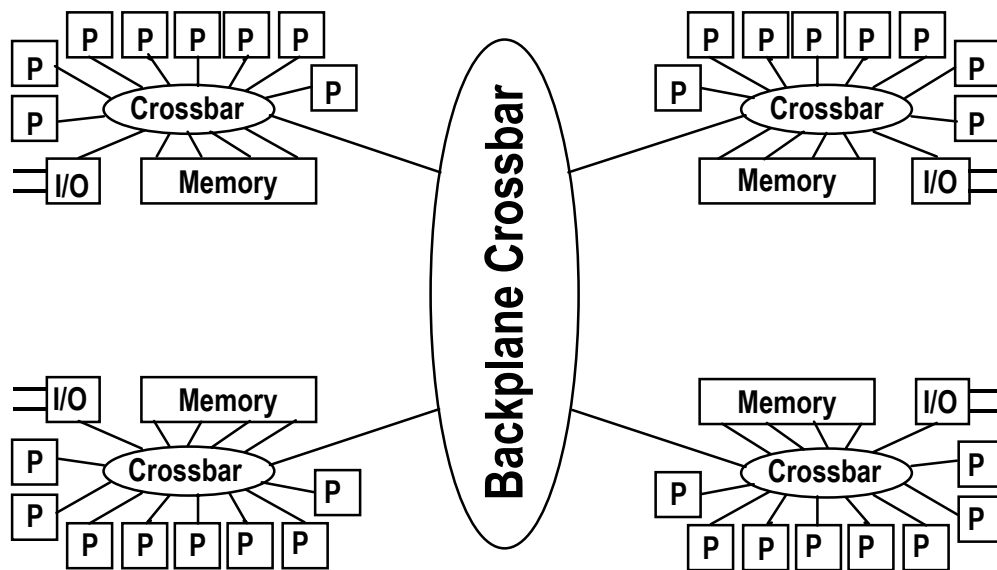
Of course, some individual applications do need a system as large as the 128-way PRIMEPOWER 2500. At the same time, customers value large systems as vehicles for consolidating the workloads of multiple servers into an easy to manage, central server. Large workload consolidation servers, often called throughput servers, collect diverse applications under the control of a single operating environment. Since Solaris supports up to 128 processors, such workload consolidation could be efficiently accomplished on a PRIMEPOWER 900, 1500, or even 2500.

Some applications do not “play together well” as they may have conflicting software stacks; for example, they may depend upon different version levels of middleware. In such a case,

the large consolidation server needs to be partitioned into smaller operating system images until the eventual convergence onto a common software stack. The error isolation afforded by partitioning is also very valuable to assure that failures in software under development or test cannot affect production workloads.

Solaris has been a leader in dynamic partitioning, allowing computing, memory, and I/O resources to be reassigned among partitions to meet changing workload needs. On the new PRIMEPOWER Enterprise Servers (models 900, 1500, 2500), Fujitsu has enhanced its hardware-enforced partitioning capabilities to provide additional flexibility to customers.<sup>3</sup> To understand the additional flexibility of Fujitsu's new design, it is necessary to delve a bit into the hardware design.

Figure 1: PRIMEPOWER 1500 Block Diagram



PRIMEPOWER systems are constructed from modular building blocks. Each of the system boards contains up to eight processors,<sup>4</sup> up to 32 GB of memory (with currently available technology), and connections to PCI I/O. Figure 1 represents a block diagram of a 32-processor PRIMEPOWER 1500 with its four system boards.<sup>5</sup> With processors, memory and I/O, each system board essentially offers a self-contained server. Independent partitions can be created within PRIMEPOWER by instructing the backplane crossbar switch to isolate sets of system boards from each other. Each independent partition runs a different instance of the operating system and is electrically isolated from hardware or software failures in other partitions.

Reconfigurable system board partitioning is not unique to PRIMEPOWER. Similar capabilities can be found on other UNIX/RISC platforms. What is unique in the new PRIMEPOWER models is a partitioning within each system board that Fujitsu calls Extended Partitioning (XPAR).

# Fujitsu Boosts PRIMEPOWER Performance

---

October 22, 2002

The electrically enforced isolation fundamental to hardware partitioning ensures that faults in one partition do not affect other partitions. However, the granularity of hardware partitioning typically has aligned along physical packaging boundaries, such as an entire system board with all its processors and memory. On the other hand, software partitioning usually allows individual processors to be assigned to different partitions but depends on a software layer to attempt fault containment. Customers trust the isolation provided by hardware partitioning, but often find that a full system board's resources are too coarse a granularity.

With XPAR, Fujitsu has retained the rigorous isolation of hardware partitioning, but it has also allowed a reconfigurable granularity of less than a full system board. Just as the backplane crossbar switch can be designed to isolate the different system boards connected to the backplane, so too can the on-board interconnect be designed to isolate the different processors that connect to that switch. On the new PRIMEPOWER models, each eight-processor capable system board is subdivided into two or four XPARs, depending upon the particular model.<sup>6</sup> By extending hardware isolation to the system board interconnect, Fujitsu has blended the best characteristics of hardware and software partitioning.

## Enhanced RAS (Reliability, Availability, and Serviceability)

The 0.13 micron process technology used for SPARC64 V offers chip designers a large number of circuits not only for increasing performance but also to directly enhance error detection for fault isolation. In fact, error handling becomes ever more important as RISC systems are depended upon for the most critical tasks. Furthermore, as chip circuitry gets ever smaller with each process generation, more electrical "noise" shows up, often caused by alpha particle radiation.

From their mainframe design expertise, Fujitsu's processor engineers have extensive experience in designing error detection and correction circuits. Hardware instruction retry was a technique used in mainframes beginning in the 1970s. Many error conditions are intermittent and the failing instruction can be re-executed successfully, as long as proper machine-state has been saved to re-initialize for that retry. In its SPARC64 V implementation, Fujitsu offers mainframe-class hardware instruction retry in a RISC processor. Hardware instruction retry is transparent to software and does not require any operating system or application support. Except for a minuscule performance hiccup when the instruction is repeated, hardware instruction retry is not noticed, other than the crucial but positive fact that the system keeps running and does not crash from the intermittent failure.

Fault isolation was also an important concern in early mainframe systems. However, typical RISC chip designs have focused on enhancing the performance of their execution units, and have not been inclined to insert error-checking circuits that could slow performance. Although failures may be relatively rare, the lack of error-checking circuits can mean that a hardware failure may not be detected and could lead to data corruption. Even if the bad

data were discovered later, by that time it would be too late to reverse the erroneous execution and data corruption. The entire system would be typically brought down to avoid unknown data integrity exposures.

Certainly, SPARC64 V offers error detection and correction for cache data and memory bus, as do many other current RISC processors. What sets SPARC64 V apart from other chips is a mainframe-influenced level of error checking that avoids data integrity problems. Fujitsu has added error checkers to processor registers, internal data paths, Arithmetic Logic Units (ALUs), data caches and tag arrays, and the Translation Lookaside Buffer (TLB). If the error results from a solid failure, the SPARC64 V can disable portions of itself without intervention of software, such as parts of the cache, and can continue operating in a degraded mode.

From the system design perspective, PRIMEPOWER includes dual power supplies and redundant fans, as is typical of enterprise servers. In addition, the backplane crossbar is actually constructed in two sections. If one section fails, no connectivity is lost – all system boards can continue to communicate with each other, albeit with reduced data bandwidth. The Error Correcting Code and the layout of memory chips allows recovery not only from single bit errors, as is typical, but also from a group of errors that come from failure of an entire memory chip.

The collection of reliability and availability features in both chip and system will minimize system crashes due to failures. Should an unrecoverable error occur, the dynamic partitioning capabilities allow the failing components to be isolated in their own partition for further diagnosis or to be hot-swap replaced.

## Final Thoughts

Although benchmark evidence is not yet available to confirm the performance capabilities of the new models, the strong performance of existing PRIMEPOWER systems suggests that the faster processors and improved interconnect will result in the new servers being quite competitive.

The new PRIMEPOWER models scale to 16-way, 32-way, and even 128-way, while maintaining a balance of processors, memory, and I/O thanks to a modular system board approach. This system board building block forms the fundamental unit of hardware partitioning. Beyond that, XPARs introduce a reconfigurable granularity that is a subset of the system board.

Complementing the performance, scalability, and flexibility attributes, Fujitsu designers have engineered mainframe-like RAS characteristics into both the SPARC64 V chip and the PRIMEPOWER system design. These enterprise-level characteristics, combined with the broad application portfolio available for Solaris OE, should bring PRIMEPOWER to the attention of IT organizations seeking a robust UNIX server.

# Fujitsu Boosts PRIMEPOWER Performance

October 22, 2002

- <sup>1</sup> SPARC64 V employs 0.13 micron semiconductor technology to run at a 1.3 GHz clock rate. Future implementations will take advantage of improvements in semiconductor technology to run at even higher frequencies. Fujitsu has indicated that it anticipates shipping 1.8 GHz SPARC64 in the 2003 timeframe and targets 2.5 GHz chips in 2004 that use 0.10 micron technology.
- <sup>2</sup> The 675 MHz SPARC64 GP delivered 475 SPECint2000 and 493 SPECfp2000. The 563 MHz SPARC64 GP provided 395 SPECint2000.
- <sup>3</sup> The midrange models 650 and 850 do not support partitioning, either full system board or XPAR.
- <sup>4</sup> Note that different system boards may be populated with different speed processor chips and yet, can still be joined into the same partition. For example, a PRIMEPOWER server could retain its original system boards while upgrading to faster speed processors that become available in the future. By allowing mixed processor speeds within a partition, Fujitsu provides investment protection.
- <sup>5</sup> The 128-way model 2500 contains 16 system boards connected via a larger backplane switch. The 16-way model 900 has two system boards that can be directly connected without the need for a backplane switch.
- <sup>6</sup> PRIMEPOWER 900 and 1500 offer up to four XPARs on their eight-processor system boards. The PRIMEPOWER 2500 supports two XPARs per system board. Note that the PRIMEPOWER 2500 employs slightly different interconnect chips that allow up to 128 of the 128-way servers (calculating to a maximum of 16,384 processors) to be joined into an HPC cluster with a hardware synchronization mechanism.

This document is copyrighted © by D.H. Brown Associates, Inc. (DHBA) and is protected by U.S. and international copyright laws and conventions. This document may not be copied, reproduced, stored in a retrieval system, transmitted in any form, posted on a public or private website or bulletin board, or sublicensed to a third party without the written consent of DHBA. No copyright may be obscured or removed from the paper. D.H. Brown Associates, Inc. and DHBA are trademarks of D.H. Brown Associates, Inc. All trademarks and registered marks of products and companies referred to in this paper are protected.

This document was developed on the basis of information and sources believed to be reliable. This document is to be used "as is." DHBA makes no guarantees or representations regarding, and shall have no liability for the accuracy of, data, subject matter, quality, or timeliness of the content.