When a fatal error is detected while the system is running, it is logged and the system automatically resets. POST is invoked and tests all the hardware resources to determine if the error is permanent or transient. POST will then make the judgment on the resulting configuration.

In case of an XDBus failure, the system can be rebooted with a single bus without losing any functionality but the overall performance will be less as half the memory bandwidth is lost.

In the event of a processor failure, the failed Processor Unit is isolated from the rest of the system. This does not affect functionality as long as there are other processors in the system.

If an MQH fails, the Memory Unit is disabled. When a DRAM chip fails the corresponding memory group can be disabled and the Memory Unit remains functional if the second group is also populated.

Failure in the I/O subsystem are more difficult to handle as in general it is important to avoid loss of connectivity to I/O resources. If an IOC fails, POST will reconfigure the system with a single XDBus if no alternate path exists to critical I/O devices.

### Serviceability

The SPARCcenter has numerous features designed to enhance the serviceability of the system.

There is an extensive failure/error logging and reporting logic in all the ASICs.

There are no jumpers for configuration, no slot dependency in the backplane and all connectors are keyed which makes installation less prone to human errors.

There is a very small number of FRUs which contributes to a very low maintenance cost.

The front panel, the System Board and the Control Board all have status/error lights.

### Solaris

The SPARCcenter 2000 is compatible with all other SPARC platforms and runs the Solaris 2.X software environment. Solaris offers the largest Unix application base in the industry.

The latest version of Solaris 2.X offers several new features to improve performance and makes system management and security easy and cost-effective.

Solaris 2.X has full symmetric multiprocessing and multithreading capabilities to improve the performance of both commercial and technical applications.

It features installation tools to configure and upgrade systems over the network.

System administration is made easy through: automatic backup over the network, graphical tools to set-up new accounts and print servers, a tool to configure client systems and an icon based tool to install third-party software.

All these new features make the SPARCcenter 2000/

Solaris combination the ideal rightsizing solution.

### Conclusion

With 20 SuperSPARC processor at 40 MHz, the SPARCcenter 2000 delivers 2.19 GIPS and 269 MFLOPS[1]. In an 8 way configuration, the system delivers 8,047 SPECrate_int92 and 10,600 SPECrate_fp92[2].

The entry price with two SuperSPARC processors is below $100,000 which gives an unprecedented price performance ratio for this type of system.

The combination of large scale multiprocessing, extensive main memory and I/O capacity, reliability and availability and a very attractive price/performance make the SPARCcenter 2000 unique.

The multidimensional scalability and the ability to accommodate future processor upgrades will protect customers' investment and make the SPARCcenter 2000 the computing platform of the 90s.

### References

[1] "XDBus: A High-Performance, Consistent, Packet-Switched Bus". Pradeep Sindhu, Jean-Marc Frailong, Jean-Gastinel, Michel Cekleov, Leo Yuan, Bill Gunning and Don Curry. Proceedings COMPCON 1993.

[2] "The Next-generation SPARC Multiprocessing System Architecture". Jean-Marc Frailong, Michel Cekleov, Pradeep Sindhu, Jean gastinel, Mike Splain, Jeff Price and Ashok Singhal. Proceedings COMPCON 1993.

[3] "A Three-Million-Transistor Microprocessor". Fuad Abu-Nofal, Rick Avra, Kanti Bhabuthmal, Rom Bhamidipaty, Greg Blanck, Andy Charnas, Peter DelVecchio, Joe Grass, Joel Grinberg, Norm Hayes, George Haber, Jim Hunt, Govind Kizhepat, Adam Malamy, Al Marston, Kaushal Mehta, Sunil Nanda, Hoa Van Nguyen, Rajiv Patel, Andy Ray, Jim Reaves, Alan Rogers, Stefan Rusu, Tom Shay, Irwan Sidharta, Terry Tham, Peter Tong, Richard Trauben, Anthony Wong, David Yee, Naeem Maan, Don Steiss, Lynn Youngs. Proceedings IEEE International Solid-State Circuits Conference 1992.

[4] "A CMOS Low-Voltage Swing Transmission-Line Receiver". Bill Gunning, Leo Yuan, Trung Nguyen, Tony Wong, Proceedimgs IEEE International Solid-State Circuits Conference, 1992.

[5] "A BiCMOS 50 MHz Cache Controller for a SuperScalar Microprocessor". Balakrishna Joshi, R. K. Anand, Curt Berg, Jorge Cruz-Rios, Ashok Krishnamurthi, Nyles Nettleton, Sophie Nguyen, Jim Reaves, John Reed, Alan Rogers, Stefan Rusu, Chuck Tucker, Chung Wang, Ming Wong, David Yee, Jung-Herng Chang. Proceedings IEEE International Solid-State Circuits Conference 1992.

[6] "IEEE Standard Test Access Port and Boundary-Scn Architecture, P1149.1". IEEE Computer Society, Test technology technical Committee, January 1989.

---

1. With the Apogee 0.82 compiler and the SUN Fortran compiler 3.0α respectively

2. With the SUN C and Fortran compilers 3.0α

supports an explicit flow control mechanism which is activated when one of the XDBus client detects that one of its incoming queue is becoming congested.

## JTAG

JTAG (acronym for Joint Test Action Group) is a boundary-scan test standard adopted by IEEE [6]. In the SPARCcenter 2000, JTAG is implemented on the System Board, Control Board, backplane and all the ASICs. There is a JTAG control port accessible from each processor on the BootBus. An optional Service Processor may also be connected to the Control Board. In this case, the System Board JTAG ports are overridden.

JTAG provides a serial path to shift data and commands from the JTAG control port to a particular chip in the system. Rather than a single long serial path on the System Board, JTAG provides a way for branching to a particular scan path which contains a limited number of ASICs. There are six such paths per System Board which are called scan rings. It is also possible from a processor to scan devices on another System Board through the backplane JTAGs scan bus.

Within each ASIC there are also multiple scan paths. Dedicated component ID, status and control registers as well as all the flip-flops in the chip can be accessed through JTAG. This is a very important feature of the SPARCcenter 2000 which has multiple applications.

During self-test and debugging it is possible to read and write all registers inside a chip through full scan. This allows testing of all the logic in the ASICs by using Automatic Test Pattern Generation. In all the SPARCcenter 2000 ASICs the test coverage is more than 95%.

A second important use of JTAG is to test board connectivity. A data pattern is scanned in the output register of one of the chips and then the contents of the input register of another chip, connected to the former one, is read and checked.

The SPARCcenter 2000 ASICs also provides Shadow scan chains which can be read without perturbing the chip's behavior. This may be used to monitor hardware statistics and check for errors while the system is running.

JTAG is also used to initialize and configure the SPARCcenter 2000 after a power-on reset. By reading the identifier of the ASICs, the current system physical configuration can be easily determined. The configuration software then loads in various control registers basic information necessary to uniquely identify each device in the system.

## Automatic System Configuration

The SPARCcenter 2000 uses an extensive set of diagnostics to determine which devices are fully functional. These diagnostics are part of the Power-On Self Test (POST) firmware which is stored in the BootBus EPROM. POST runs before the kernel is booted and configures automatically the system. POST also runs before each reboot and when recovering from fatal errors.

POST runs in several phases. First each processor tests itself and then the rest of the Processor Unit (CC and BWs) is also tested using JTAG. At that time the System Board are in loopback mode and the access to the backplane segment of the XDBuses is disabled. The other devices are also tested while the System Board is in loopback mode. The processors then elect one of them as master processor through JTAG. Next, all XDBus interface are checked and the master processor configures the system, including mapping the memory banks.

In case of failure there are multiple alternatives. A series of heuristics are used to decide which configuration to be used. There is often more than one way to reconfigure the system and it is generally a tradeoff between performance and functionality. This is discussed in the next section.

## Reliability, Availability and Serviceability

The SPARCcenter 2000 incorporates many features that improve the reliability, availability and serviceability of the system.

### Reliability

By using a very high level of integration and a conservative design methodology the probability of errors is greatly reduced. The extensive diagnostics performed by POST also ensure that only fully functional devices are enabled so that errors cannot occur because of failing components.

The SPARCcenter 2000 can also tolerate and recover from some errors like memory errors. The SIMM implementation guarantees that a Memory Unit will continue to operate even in case of the permanent failure of a DRAM chip. The failing memory group can then be deconfigured.

All data paths are parity protected in the SPARCcenter 2000 to ensure data integrity. XDBus, XBus, SBus, the arbitration signals and the SuperSPARC processor bus are parity protected. The external cache is also parity protected. There is also extensive checking of the XDBus transactions consistency by all the client devices.

The SPARCcenter 2000 is equipped with a number of environmental sensors to protect the system from hazardous conditions. A sensor checks the temperature and another one detects a failure from the fan. Abnormal AC or DC conditions are also detected.

### Availability

The SPARCcenter 2000 is not a fault tolerant system in the sense of being capable of "non-stop" operation. However, it has extensive capabilities to detect and identify errors and the system can be reconfigured automatically without human intervention.

BootBus
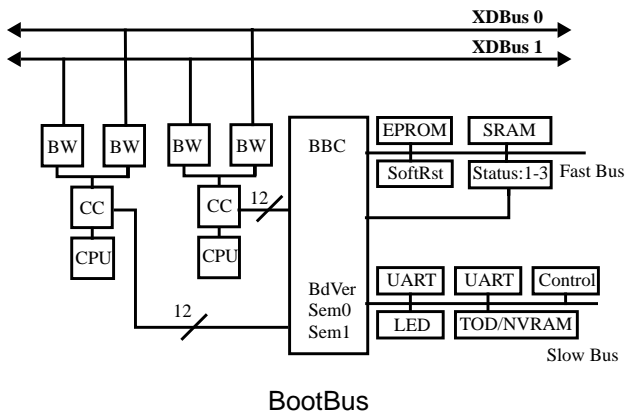
EPROM, 16 KB of SRAM, three Status registers, two Semaphore registers, the System Software Reset register and the System Board version register. The EPROM contains the boot code and the SRAM is used as a scratch pad and for the stack.

The devices located on the slow bus are: the LED diagnostic register, a Control register, a UART which provides two RS232 ports, a UART for a keyboard/mouse interface, the JTAG Master Interface register and a Time-Of-Day/ Non Volatile RAM chip.

The slow devices are shared using one of the semaphore register.

## Control Board

The Control Board provides the System Clock Generation, Central Arbitration, Power-on reset generation and the JTAG port for a Service Processor connection. It also has some LEDs to indicate the status of power and some system signals.

The figure below details the major functional units of the Control Board.
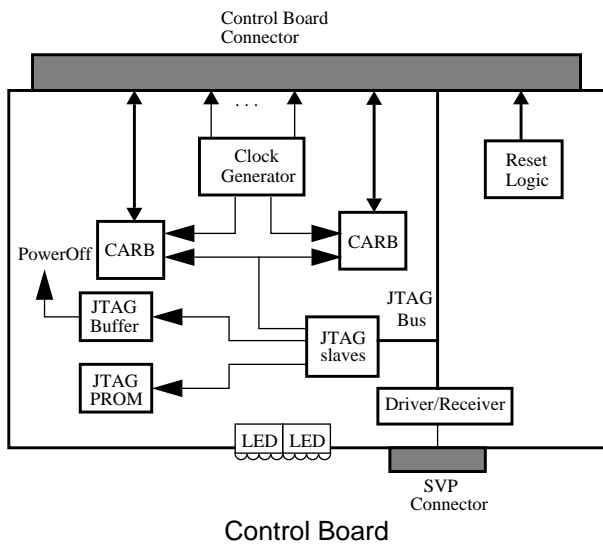


Control Board

The Control Board supports the clock generation logic.

The 40 MHz system clock is generated as a PECL signal. All clock traces on the distribution path have the same length and an equal number of loads to minimize the skew. The clock received by the Central Arbiters (CARB) has the same clock path as the XDBus client chips on the System Board.

The reset logic generates a general system reset signal which is forwarded to all System Boards. A reset is generated when either the reset switch on the front panel is activated, a power-on condition is detected, the optional Service Processor requests a system reset, a fatal error is detected in the system or one of the processor requests a reset by setting a bit in one of the BootBus register.

The Control Board also includes a JTAG PROM which contains the system ID and the Ethernet address.

## Arbitration

The SPARCcenter 2000 uses a two level arbitration scheme to grant access to the XDBus. Because the two XDBuses operate in parallel, the arbitration logic is duplicated for each of them. The arbitration is implemented with two types of ASIC: the Board Arbiter (BARB) and the Central Arbiter (CARB). Arbitration requests generated by XDBus client devices are collected by BARBs and forwarded to the CARB. The CARB then selects a client device as XDBus master by issuing a grant signal to that client's BARB. This selection is made according to specific requirements of priority and fairness. Finally the BARB forwards the grant signal to the device. The architecture of this hierarchical arbitration scheme is depicted below.
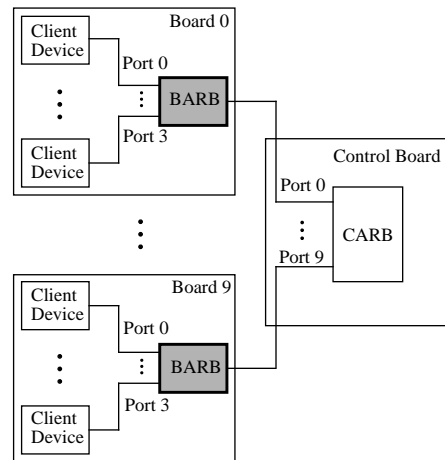


Two Level Arbitration Architecture

The BARB and the CARB also participate in the data consistency protocol by merging the "Shared" and "Owner" signals issued by the BWs when they respectively detect that a block is present in the cache and modified.

The arbitration algorithm uses multiple level of priorities to provide an implicit flow control mechanism. It also

## I/O Unit

### Overview

The SPARCcenter 2000 System Board supports a complete SBus which is used as I/O bus. Each SBus has four slots and is clocked at 20 MHz. All peripheral devices are connected to the SBus.

The I/O Unit provides the bridge between the SBus and the XDBus complex. The I/O Unit is composed of two I/O Cache chip (IOC), an SBus Interface chip and an External Page Table (XPT). The IOCs and the SBI are interconnected with an XBus. The following figure depicts the architecture of the I/O Unit.
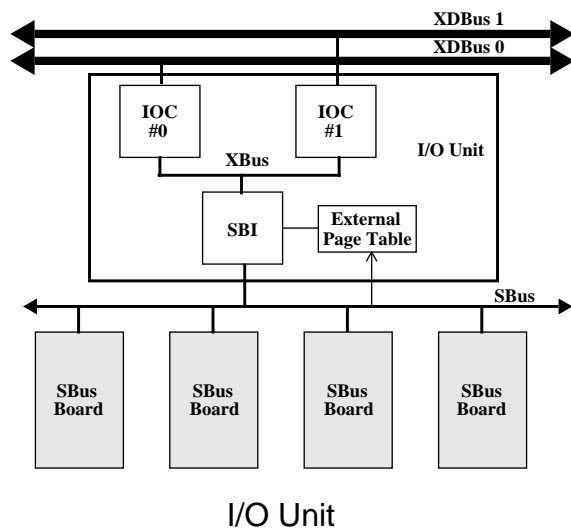


I/O Unit

### I/O Model

The I/O Unit provides three different I/O models:

- Programmed I/O where the processor directly reads and writes the I/O devices.

- Consistent DVMA I/O. DVMA stands for Direct Virtual Memory Access. In this mode, the data is moved directly between the SBus and the XDBus complex and the SBus address is translated into a physical XDBus address by the XPT. This mode is called consistent because the data is moved into the "Shared Memory Image" in the I/O Cache. In this mode, there can be only one pending transaction between an SBus board and the memory system.

- Stream Mode DVMA. As for the previous mode, SBus addresses are translated by the XPT but the data is not moved directly between the SBus and the memory system. Instead it go through pairs of buffers. These double buffers are not part of the shared memory image and are not kept consistent until they are flushed or invalidated by software.

### Implementation

The IOC contains a small fully associative write-back cache which is kept consistent. Data is read from this cache or written into this cache when I/O transfers are done in consistent DVMA mode. The IOC also provides simulta-

neous I/O accesses to XDBus for each SBus slot. Even if an SBus board has an XDBus transaction pending, the other SBus boards can still access the XDBus. The IOC provides the interface between XDBus and the local XBus.

The SBI contains the read and write buffers used in Stream Mode DVMA. Each SBus slot has its own pair of double buffers and they are managed under software. The Stream Mode is the most efficient of the two DVMA modes. In stream mode, each SBus can sustain 50 MB/s when using 64-byte bursts and 30 MB/s when using 16-byte bursts The peak bandwidth is 80 MB/s. The transfer mode is selected on a slot basis.

The External Page Table implements a single-level page table through a set of SRAM chips. It can map up to 64 MB of DVMA address space. Each entry maps a 4 KB page. The XPT is controlled by the SBI and is maintained consistent by the kernel.

The SPARCcenter 2000 supports the Revision B.0 of SBus. All SBus transfer sizes (2,4, 8, 16, 32 and 64 bytes) are supported in both DVMA I/O modes. All SPARC addressable quantities are supported in programmed I/O mode. Each slot supports the full 28-bit address space. The SBus clock is independent of the system clock. An asynchronous boundary is implemented inside the SBI. An important feature of the SPARCcenter 2000 implementation of SBus is the parity extension support for data integrity. Parity can be enabled on a slot basis, so that devices which are not supporting parity can still be used.

### Peripherals

The principal advantage of using a standard bus is the wealth of available peripheral boards. This is specially true for SBus since a large number of systems use it as an I/O bus and there is a large market for independent board manufacturers.
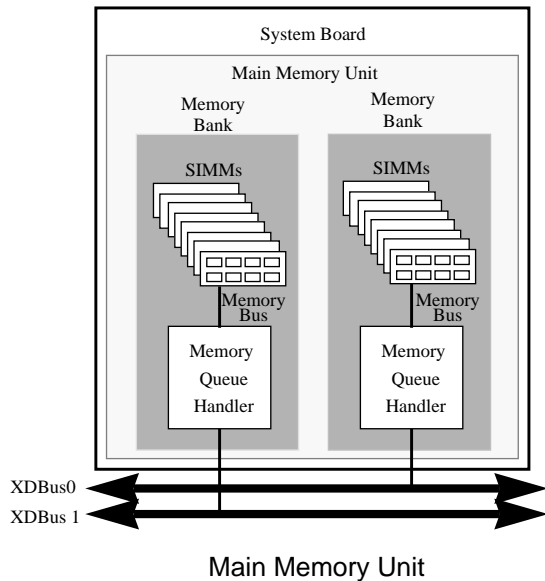
The following common interfaces are available: FDDI, Token Ring, HSI (4 synchronous lines), DSBE (Differential SCSI with buffered Ethernet controller), ISDN, GX frame buffer.

### BootBus

Each SPARCcenter 2000 System Board also supports an 8-bit local bus called the BootBus. This bus is shared by the two Processor Units and is used to access system support devices. The BootBus is controlled by the BootBus Controller chip (BBC). The Cache Controllers are connected to the BBC through a 12 signals interface.

There are two types of devices connected to the Boot-Bus: fast and slow devices. The SuperSPARC processors can access the fast devices simultaneously. However, the slow bus can only be accessed by one processor at a time. The following figure illustrates the connection between the Processor Units and the BootBus devices.

The devices connected to the fast bus are: 512 KB of

System Board

Main Memory Unit

Memory Bank     Memory Bank

SIMMs     SIMMs

Memory Bus     Memory Bus

Memory Queue Handler     Memory Queue Handler

XDBus0
XDBus 1

Main Memory Unit

transfer on the XDBus). The MQH does not support writes on smaller quantities and the main memory in the SPARC-center 2000 is always cached.

A memory bank is the unit of interleaving. A memory bank consists of one or two groups of 4 custom SIMMs. These are the two possible configurations (in addition, of course, to the case where the bank is not populated at all).

The MQH can handle DRAM densities from 1 Mbit to 256 Mbit. The first generation of SPARCcenter 2000 uses 4 Mbit and 16 Mbit DRAMs.

The MQH is connected to the SIMM through a Memory Bus. This is a 72 bit wide TTL bus also clocked at 40 MHz. However, the timing access to the SIMM is fully program-mable allowing DRAM with different timing to be used.

The memory is protected by an Error Correcting code which detects and corrects single-bit errors and detects all double-bit errors. It can also detect triple and quadruple-bit errors if the erroneous bits are in the same nibble.

The MQH is implemented as a 100K CMOS gate array.

## Interleaving

Because of XDBus interleaving, for each memory group on a given XDBus there must be an identical group on the other XDBus. The memory size increment is the memory capacity of two groups of 4 SIMMs.

The physical memory address space is entirely pro-grammable and each memory group is controlled by a dis-tinct address decode register. The memory bank on the same XDBus can be configured for no interleave, 2-way interleave or 4-way interleave by programming the address decoding registers of the MQHs.

In a shared-memory symmetric multiprocessor system the motivation for an interleaved main memory is to allow multiple independent accesses. With the XDBus packet-switched protocol, multiple memory transactions issued by

different processors and I/O devices can be pending at the same time. In large configurations, memory bank interleav-ing reduces the probability of having "hot spots" develop-ing on a given MQH.

## SIMM

The SPARCcenter 2000 uses custom SIMMs. The SIMM is composed of Crossbar ASIC and memory chips on a small board. Each SIMM is organized as 18 bits wide (16 bits of data and 2 bits for ECC), so 4 SIMMs operating in parallel are necessary to interface to the memory bus.

Each SIMM contains 18 x4 DRAM chips and four crossbar ASICs. Since the DRAM access time is nominally four system clock cycles, the DRAM data path on the SIMM was designed to be four times the memory bus width. The Crossbar's function is to time multiplex the data and transform a single DRAM access into four dou-ble-word transfers on the memory bus. Since the memory is always accessed in 64 byte blocks, two sequential accesses using page mode are made on the memory bus.

With 16 Mbit DRAM chips a SIMM has a capacity of 32 MB, and a memory group a capacity of 128 MB. The minimum memory configuration for the SPARCcenter 2000 is 256 MB. A fully populated backplane SPARC-center 2000 system with 10 System Boards provides a maximum main memory size of 5 GB.

## NVRAM

The SPARCcenter 2000 also supports a battery-backed non-volative RAM (NVRAM). The main purpose of using non-volatile memory is to accelerate synchronous disk writes. Write accesses to disk can take place in two phases. First the data is copied into non volatile memory and the write is acknowledged. Later, in a second phase the actual write to the disk takes place when a scheduling algorithm determines it is time to do so. The data stored in the NVRAM has the same property as data stored on disk drives, it can survive system crashes and power failures.

Synchronous disk writes can be done much faster because accessing the NVRAM is much faster than access-ing current technology disks. In the SPARCcenter 2000, the NVRAM bandwidth is comparable to the regular main memory bandwidth.

Applications like NFS and DBMS, where data integrity is crucial, can easily take advantage of NVRAM and pro-vide a better response time.

The NVRAM is implemented with non-volatile SIMMs (NVSIMM). The NVSIMM is composed of SRAM chips, Crossbar ASICs and a small lithium coin battery. Each NVSIMM has a capacity of 1 MB and the minimum con-figuration is two group of 4 NVSIMM, i.e. 8 MB.

be plugged in any slot of the cardcage and the BoardID can be accessed at system configuration time through a pseudo register. There are no jumper or switch settings required for any configuration. This avoids the likelihood of installation errors.

A minimum system consists of at least one System Board with a single processor and one Memory Unit. Because all units are equally accessible, the specific locations of the memory, processors, and I/O devices are not fixed. If an application requires only 4 processors but 3 GB of memory, the system can be configured with 6 System Boards with fully populated Memory Units but only two SPARCmodules. The boards may be plugged in any slot.

## Processor Unit

The Processor Unit consists of a SuperSPARC processor, an external cache and system support devices connected to the BootBus. The figure below illustrates the main components of the Processor Unit and their interconnections:
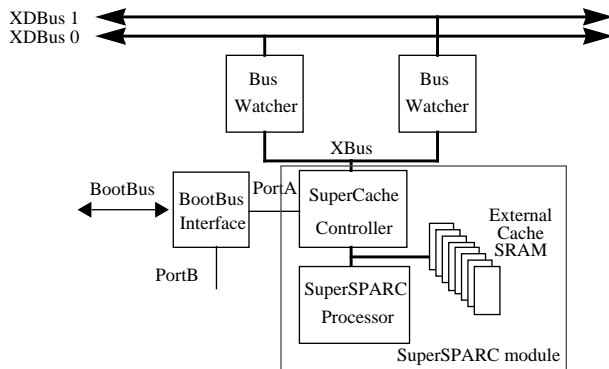


Processor Unit

The external cache includes the Cache Controller and two Bus Watcher ASICs and 1 MB of parity protected SRAM. The SuperSPARC processor, the Cache Controller and the SRAM are located on the SuperSPARC module. The module is a small daughter card which plugs through a 100 pin connector onto the System Board. The Bus Watchers are located directly on the System Board.

The SuperSPARC processor is a highly integrated chip which includes two integer units, a floating point unit, a branch processor, a SPARC reference MMU and a 36 KB on-chip set-associative cache [3]. This processor is capable of executing up to 3 instructions per cycle if they are fetched from the internal cache and scheduled properly. The processor is a 3 Million transistor custom BiCMOS chip.

The Cache Controller controls 1 MB of combined instruction and data external cache. The cache is physically accessed and is managed as a write-back cache. The current SRAM technology limits the size to 1 MB but the

Cache Controller can support up to 2 MB of cache. The Cache Controller is a 2.2 Million transistor BiCMOS chip [5].

The Cache Controller and the two Bus Watchers implement the cache consistency protocol. The Bus Watcher chips contain a copy of the cache tags to minimize contention between the processor and the XDBus accesses. The cache consistency protocol relies on snooping the XDBus traffic. The Bus Watcher basically filters out almost all bus transactions leaving the processor free to access the cache most of the time. When sharing is detected, the Bus Watcher updates some state bits or retrieves the requested data from the cache. Each Bus Watcher contains half of the duplicated tags and are interleaved on a 256 byte boundary.

The Bus Watchers and the Cache Controller are interconnected by a local packet-switched bus known as the XBus. The XBus is very similar to the XDBus. Transactions supported by the XBus are similar to XDBus transactions. The difference is in the use of dedicated commands to maintain the two copies of cache tags consistently. The use of a packet-switched protocol on the XBus is also key to the performance of the processor in a multiprocessor environment. Although the external cache handles only a single miss at a time, multiple requests may be outstanding. For instance multiple requests for block invalidation or update can be issued from the same Processor Unit. This guarantees that the processor does not stall even when there is lots of data sharing with other processors.

The XBus is also uses GTL transceivers which allow a direct pin to pin connection between the Cache Controller and the Bus Watcher. The XBus is also clocked at 40 MHz.

Although the first generation of the SPARCcenter 2000 utilizes 40 MHz SuperSPARC processors, faster Super-SPARC processor modules can be used to upgrade existing SPARCcenter 2000 systems. Most of the Cache Controller operates synchronously to the processor clock. The XBus interface operates synchronously to the system clock and there is an asynchronous boundary inside the Cache Controller.

The Cache Controller is also connected to a local bus called the BootBus. Support devices like a time-of-day-clock, UART, scratch-pad memory, EPROM are attached to the BootBus. The BootBus is shared by the two processors on the same System Board.

## Memory Unit
### Overview

The SPARCcenter 2000 Memory Unit consists of two memory banks, one per XDBus. A memory bank is defined as a memory array controlled by a Memory Queue Handler (MQH) ASIC. The MQH interfaces directly to the XDBus and controls the memory array. The MQH supports read and write operations on 64-byte blocks only (unit of block

Queue Handlers (MQH), an SBI Interface (SBI) and two I/O Caches (IOC) and ten smaller ASICs: two Board Bus Arbiters (BARB) and eight bit-slice pipeline registers to connect the on-board XDBus segments to the backplane (BIC). The System Board also supports two SPARCmodules consisting of a 3 Million transistor SuperSPARC processor and a 1MB direct-map cache memory. The cache memory is composed of a 2 Million transistor Cache Controller (CC) and the SRAM array. The SPARC modules are not actually part of the System Board. Instead, the SPARCmodules are mounted on the System Board through dedicated connectors. Similarly, the memory DRAM memory chips are not on the System Board but on SIMMs which plug vertically.

The following block diagram shows the location of the major components.
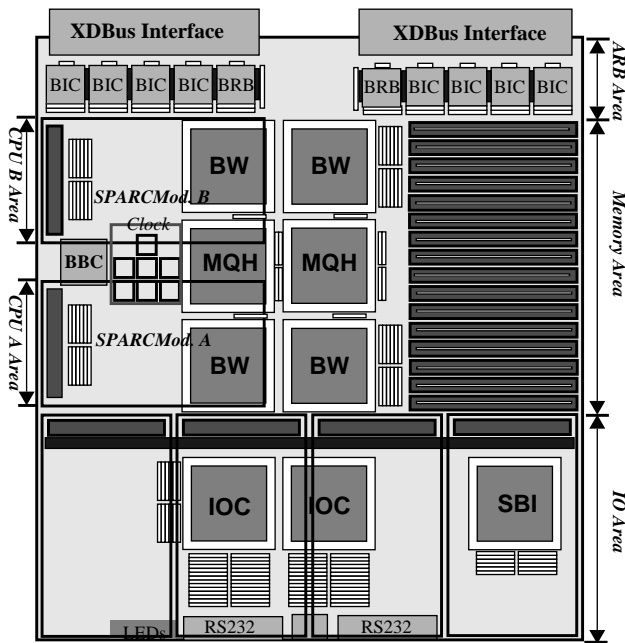


SPARCcenter 2000 System Board Layout

This highly integrated design is key to system reliability. It also simplifies design without compromising performance and configuration flexibility.

The System Board has fourteen layers, eight for signals and six for power and ground.

## XDBus Backplane

### Pipelined Implementation

The backplane connects up to 10 System Boards. It consists of two bidirectional segments of the XDBuses.The System Board supports unidirectional segments for each of the XDBuses. The figure below shows the architecture of the XDBus complex.

The on-board XDBuses are each composed of the unidi-
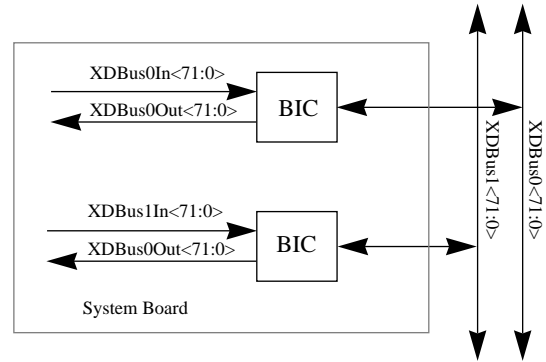


Bidirectional Backplane XDBuses
and Unidirectional on-board XDBuses

rectional segments XDBusIn and XDBusOut. Each of these segments consists of 64 bits for address and data and 8 bits for parity. The Bus Interface chips (BIC) separate the unidirectional on-board XDBuses and the bidirectional backplane XDBuses. The BIC is composed of two pipeline registers, one between the outbound on-board segment and the backplane segment and the other between the backplane segment and the inbound on-board segment. Each BIC has an 18 bit wide data path including 16 bits of data/address and 2 bits of parity. Four BICs are necessary to provide the interface between the on-board and backplane segments of a single XDBus.

### Packet-Switched Protocol

The XDBus uses a packet-switched protocol (also known as a split transaction protocol) to transfer data between clients. A packet-switched protocol offers a larger overall throughput than the more conventional circuit-switched protocol. In a packet-switched protocol, the requestor arbitrates for the control of the bus and as soon as it is granted it sends a request packet and immediately releases the bus. The bus is free to be used by other clients while the request is being processed. When the requested data is available a reply packet is issued. Reply packets are tagged so that they can be matched with the corresponding request. A packet-switched protocol permits an optimal utilization of the raw bandwidth.

### Low-Power Electrical Power Implementation

The XDBus runs at 40 MHz and uses low voltage-swing technology called GTL (Gunning Transceiver Logic) [4]. GTL uses a 0.8V voltage swing between 0.4 and 1.2V. This technology is specially designed for high-speed, high density CMOS gate arrays. GTL permits CMOS to be used in a terminated transmission line environment. Because the power dissipation is very low, a wide bus like the XDBus can be driven directly from an ASIC without having to use costly external drivers.

### Configuration

The backplane also provides identifiers (BoardIDs) to uniquely identify each System Board. A System Board can

SPARCstation 10 systems.

The main memory is configured in multiple Memory Units. All these units have the same access time from every processor and I/O device regardless of their physical locations in the system. Physical memory addresses are interleaved between the two XDBuses on a 256-byte boundary and memory banks attached to the same bus can also be interleaved to avoid bottlenecks. A Memory Unit can have a memory capacity between 64 MB with 4 Mbit DRAM chips and 512 MB with 16 Mbit DRAM chips. A fully configured system can support 5 GB of main memory.

The SPARCcenter 2000 offers incrementally expandable I/O with up to 10 SBuses. Each SBus supports 4 SBus slots for a maximum configuration of 40 SBus peripheral boards. Each SBus is connected to the XDBuses through an I/O Unit. Like memory, all SBuses are accessed with the same latency from every processor. Each SBus delivers 50 MB/s of sustainable data throughput. A SPARCcenter 2000 can be configured with up to 18 2.1 GB DSCSI-2 disks in the system rack for a maximum of 38 GB internal capacity. Expansion racks with 48 drives for a capacity of 100 GB are also available. The SPARCcenter 2000 can also be configured with a multitude of independent network interfaces. This exceptional I/O capacity and configurability makes the SPARCcenter 2000 suitable for very large applications.

Although the architecture of the SPARCcenter 2000 is similar to other large-scale symmetric multiprocessing systems, the expansion capability, the overall system balance with commensurate memory bus bandwidth makes this system unique in the industry.

### Packaging and Power

The SPARCcenter 2000 consists of 10 System Boards configured in a 10 slot XDBus backplane and a Control Board mounted on the other side of the backplane. The SPARCcenter 2000 uses a 9U format for the System Board. The basic system is packaged in the standard 56" SunRack with up to 18 5.25" disk drives, a CDROM, 1/4" tape and up to three 8mm 5 GB tape drives for backup.

The power requirements do not exceed 180 W per slot, including 10 W per SBus slot. Most power is drawn from the +5V supply. There is also a +1.2V supply for XDBus termination and a +/-12V for the SBus boards. The power supply is resistant to most kind of power fluctuations. The system can continue functioning during a brownout of 160 VAC for at least 15 minutes. It can also tolerate complete AC brownout if they are limited to a single cycle.
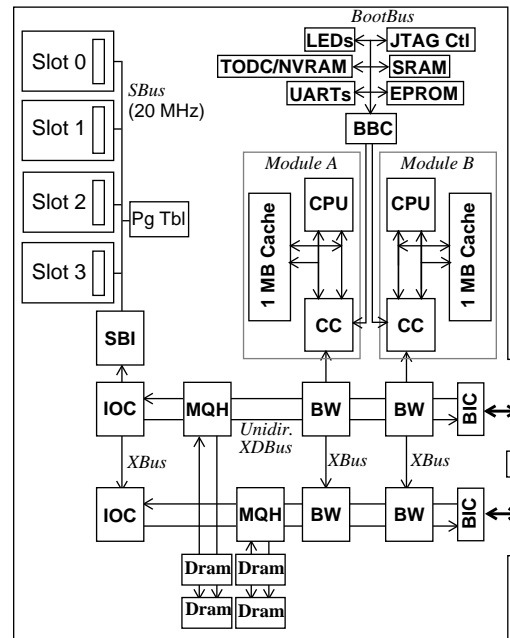
### System Board

The System Board is the primary component in the implementation. A system may contain up to 10 System Boards. Each board contains a backplane interface connection to dual XDBuses, two sockets for the SPARCmodules, sockets for 16 SIMMs, an SBus with four slots, the local
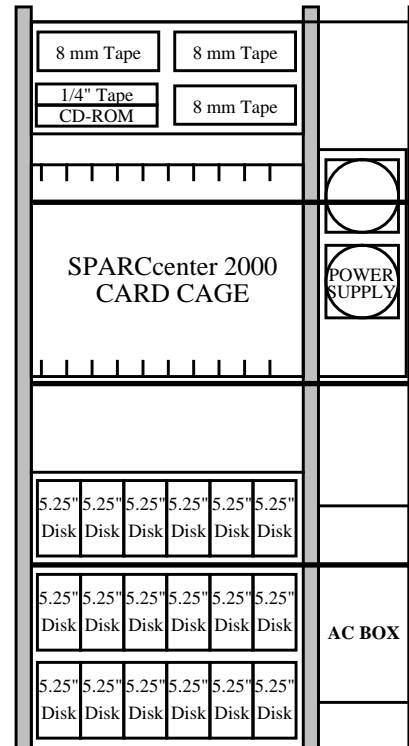


SPARCcenter 2000 system cabinet configuration

XDBus segments and a JTAG interface for diagnostics and configuration. Each System Board contains two Processor Units, a Memory Unit and an I/O Unit.



System Board Logic Diagram

The actual implementation is highly integrated: the System Board consists primarily of nine large 100K gate CMOS ASICs: four Bus Watchers (BW), two Memory

# SPARCcenter 2000: Multiprocessing for the 90's!

Michel Cekleov, David Yen, Pradeep Sindhu*, Jean-Marc Frailong*, Jean Gastinel*, Mike Splain, Jeff Price, Gary Beck, Bjorn Liencres, Fred Cerauskis, Chip Coffin, Dave Bassett, David Broniarczyk, Steve Fosth, Tim Nguyen, Raymond Ng, Jeff Hoel*, Don Curry*, Leo Yuan, Roland Lee, Alex Kwok, Ashok Singhal, Chris Cheng, Greg Dykema, Steve York, Bill Gunning*, Bill Jackson*, Atsushi Kasuya*, Dean Angelico, Marc Levitt, Medhi Mothashemi, David Lemenski, Lissy Bland*, Tung Pham*.

| | |
|---|---|
| Sun Microsystems Computer Corporation<br>2550 Garcia Avenue<br>Mountain View, CA 94043-1100 | * Xerox Palo Alto Research Center<br>333 Coyote Hill Road<br>Palo Alto, CA 94304 |

### *Abstract*

The SPARCcenter 2000 is the first implementation of a new generation of symmetric high-performance, highly configurable SPARC multiprocessor systems. With up to 20 processors, extensive main memory, expansibility and large IO capacity, the SPARCcenter 2000 is designed to meet the computing needs of most corporate data centers. A high throughput system interconnect, composed of two interleaved XDBuses, combined with multiprocessor scalability, make the SPARCcenter 2000 the right platform for compute intensive tasks. Reliability and availability features, full SPARC binary compatibility, IO scalability and Solaris 2.X MP capability also make the SPARCcenter 2000 the ideal rightsizing platform for commercial and RDBMS applications.

## Introduction

The SPARCcenter 2000 defines a new breed of shared-memory multiprocessor computers designed to accommodate the needs of most organizations, from large department to medium-scale enterprises. It uses a modular architecture composed of three types of units: the Processor Unit, the Memory Unit and the I/O Unit. All these units are interconnected through two XDBuses. The XDBus is a high-speed consistent packet-switched bus [1].

The SPARCcenter 2000 provides unparalleled expansion capability in combination with excellent scalability in three dimensions: compute power, main memory capacity and I/O bandwidth. The XDBus is flexible enough and fast enough to deliver outstanding performance and maintain scalability to levels which are unheard-of in today's marketplace.

This paper describes the implementation of the SPARCcenter 2000. It begins with an overview of the system architecture, followed by a description of the implementation of the System Board, and its various components: the Processor Unit, the Memory Unit, the I/O Unit, the Boot-Bus. These sections focus on the salient details of the implementation. For a description of the architecture the reader can refer to [2]. The following sections describe the implementation of the Control board, the Arbitration, and the support of the JTAG scanning logic. Finally the paper closes by discussing some less visible but potentially important features of the SPARCcenter 2000: reliability, availability, and serviceability. A few performance numbers are quoted in the conclusion.

## System Overview

### System Architecture

The system architecture of the SPARCcenter 2000 is depicted by the following picture:
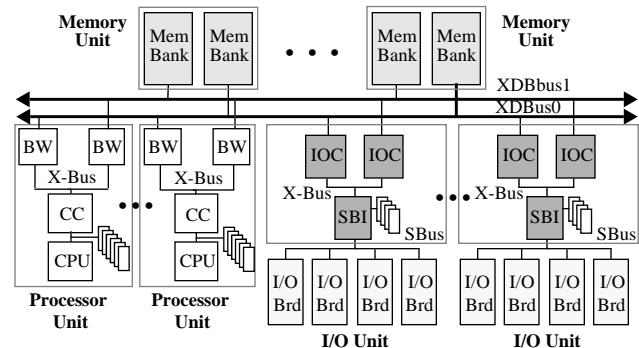


SPARCcenter 2000 system Architecture

The heart of the SPARCcenter 2000 is a high-speed packet-switched bus complex which provides a very high data bandwidth. The backplane consists of two XDBuses each providing 320 MB/s of data throughput with at a 40 Mhz clock rate. The XDBuses operate in parallel and the system can be rebooted with a single XDBus in case of a permanent failure of one of them. The system's functional units are connected to both XDBuses. Memory banks are attached to individual XDBuses and a memory unit is composed of two interleaved memory banks.

The SPARCcenter 2000 can use up to 20 SuperSPARC processors [3]. The SPARCcenter 2000 uses processor modules compatible with the SPARCserver 600MP and