- Load/store architecture;
- Fixed-length instructions;
- Hardwired instructions;
- Pipelined execution;
- Single-cycle execution.

    Non-RISC Features

- Multiple addressing modes.

6.9. SPARC

6.9.1. Introduction

The *Scalable Processor Architecture* (SPARC) is an open architecture, based on the original RISC processor designed at University of California at Berkeley. SPARC processors are compatible at the instruction-set level and may be hardware compatible. In addition, the *M-bus* specification and the reference MMU specifications are usually followed by implementers of this architecture.

The SPARC architecture is into its fourth generation. This architecture was introduced in 1987. Version 7 of the architecture (1989) had no multiply, divide, or memory management functions. The SPARC version 8 (V8) architecture, announced in 1991, was implemented by 1993. This version introduces integer multiply and divide operations, as well as 128-bit floating-point operations. The V8 architecture has two memory models: total memory ordering and partial memory ordering.

Version 9 (V9) of the SPARC reference architecture was presented in 1994, and is the 64-bit extension of the 32-bit version 8 architecture. This version adds a relaxed memory ordering, which is enabled by a bit in the processor state register. Relaxed memory ordering allows the processor to reorder the memory references, but keep memory consistency. Special instructions are provided to enforce strong memory ordering when necessary. The V9 architecture has features to support optimizing compilers and high performance operating systems. Other characteristics of this architecture are: superscalar implementation, fault tolerance, fast exception handling, fast context switching, big-endian and little-endian byte ordering.

In the V9 architecture, all registers are 64 bits, and there are new condition codes. There is a new static branch prediction bit. The privileged registers are updated and redefined. The register windows and CWP (*Current Window Pointer*) are unchanged. In the V9 architecture, 32 single-precision floating-point registers are added, which can also be grouped as sixteen 64-bit registers, or eight 128-bit registers. There are new instructions for the 64-bit functions, including loads and stores, shifts, conditional branches, conditional moves, and a population count operation, useful for compression, encryption, and finding the first "1" bit. Both little-endian and big-endian byte-ordering modes are defined. One implementation of the V9 architecture was the UltraSPARC processor from Texas Instruments. SPARC-64, designed by Fu-

jitsu, is also a SPARC V9 processor. Both companies claim to have had the first implementation of the SPARC V9 architecture.

The specification for the commercial SPARC processors, their bus structure, and memory management was developed and is maintained by SPARC International. This is an organization open to all component and system manufacturers, as well as software developers who want to design SPARC products. The members of SPARC International are responsible for compatibility and development's direction of the products based on SPARC processors. To ensure that SPARC products are compatible at binary-code level, SPARC International established the SPARC *Compliance Definition* (SCD), a specification of system software and hardware interfaces. By developing systems and software that conform to the SCD, SPARC system vendors and application developers ensure their products are compatible at binary-code level. SPARC International includes Sun Microsystems, Texas Instruments, Fujitsu, LSI Logic, Ross Technology, and others.

SPARC processors address especially the high-end workstations. Sun Microsystems has developed several families of SPARC processors (microSPARC, UltraSPARC, and the new MAJC processor). Sun's first SPARC processor (code-named "Sunrise") operated at a frequency of 16 MHz. This processor was implemented in two Fujitsu gate arrays with 20,000 gates each, totaling 40,000 gates. More recently, the UltraSPARC-III processor was introduced at a frequency of 600 MHz, and contains over 3 millions of gates. 1 GHz versions of the UltraSPARC-IV processor are available from the year 2000. Sun has announced plans for its high-performance UltraSPARC processors. The clock frequencies of the next generation of these processors will increase to over 1.5 GHz. Corresponding to these plans, 1.5 GHz versions of the UltraSPARC-V processor will be available in 2002.

Sun Microsystems also integrates SPARC processors into systems. The Sun workstation, derived from the Stanford University Network workstation, was originally implemented with Motorola's 680x0 family of processors. In 1996, Sun launched the JavaStation, a network computer based on the microSPARC processor. In 1997, Sun was the manufacturer number one of Unix servers. Currently, all of Sun's servers are powered by UltraSPARC processors. They provide the processing power of today's most robust Internet sites, such as Excite, Infoseek, Earthlink, Amazon, Time Warner, and many more.

The Sun UltraSPARC processors are manufactured by Texas Instruments, which is a major manufacturer of SPARC processors, including its SuperSPARC and MicroSPARC products. There are several other manufacturers of SPARC processors. Cypress Semiconductor sold its HyperSPARC production unit to Fujitsu in 1993. Fujitsu, with its own efforts and the acquisition of the HyperSPARC unit from Cypress, has become a major manufacturer. Fujitsu's SPARClite processor features integrated functions for the embedded controller market. This processor is used in approximately 50% of the consumer digital cameras currently on the market.

In the following we present a summary of the main features of the SPARC architecture. These include:

- *Performance and Economy.* With its simplified instruction set, SPARC processors execute a higher number of instructions per second with fewer transistors.

- *Scalability.* SPARC is scalable across a wide range of semiconductor technologies and system configurations. An advantage of SPARC is the flexible integration of cache memory, floating-point unit, and memory management unit, enabling processors to be produced at price/performance ratios suitable for systems ranging from laptop computers to supercomputers.

- *Open Architecture.* A key element of SPARC is its open characteristic, ensuring that users have access to compatible technology from multiple vendors. To encourage SPARC product development and compatibility, customers have access to complete system level and CPU component technologies.

- *Powerful Development Tools.* The Solaris operating environment, Sun's implementation of the Unix operating system, is running on SPARC-based systems. Solaris is a reliable operating environment that runs the critical applications of a large number of companies. A complete set of development tools is available for developing hardware and software products. Real-time operating systems are also available.

6.9.2. HyperSPARC

The superscalar, superpipelined HyperSPARC processor is a version 8 architecture, and it was developed by Ross Technology. It includes a 16-KB two-way set-associative instruction cache memory, with provision for a 128-KB or 256-KB unified external cache memory. Up to two instructions per clock cycle may be issued. Both integer and floating-point paths are six-stage pipelines. In the integer pipeline, results of an operation are made available to succeeding instructions by a process called *data forwarding.* This eliminates the need to store the result in a register, and then retrieve it immediately. There are five execution units on-chip, for load/store, branch/call, integer operations, floating-point add, and floating-point multiply.

The original HyperSPARC chipset consisted of the 7C620 processor, the 7C625 cache memory controller/MMU, and two or four 7C627 SRAM memories. It was available on a module and supported the level 2 *M-bus.*

6.9.3. SuperSPARC

Texas Instruments' implementation of a third-generation version 8 SPARC architecture is the TMS390Z50 SuperSPARC processor. It is capable of issuing three instructions per clock cycle (one floating-point, and two integer). This processor implemented in BiCMOS technology incorporates 3.1 million transistors, does all load and store operations in 64-bit mode, and can optimize most branches to operate without delays, whether taken or not. It incorporates a high-performance floating-point unit, and has two on-chip cache memories. The instruction cache memory has a

size of 20 KB, and is accessed by a 128-bit wide internal bus. This memory is five-way set-associative. The data cache memory has a size of 16 KB, is four-way set-associative, and is accessed by a 64-bit wide data path.

The SuperSPARC processor supports the level-2 interface of the *M-bus*. The integer multiply and divide operations use the floating-point hardware, and take four and five clock cycles, respectively.

6.9.4. MicroSPARC and MicroSPARC-II

The TMS390S10 MicroSPARC manufactured by Texas Instruments addresses the desktop market, and it is a highly integrated, low-cost processor. Incorporating a 4-KB instruction cache memory, a 2-KB data cache memory, and a floating-point unit, the MicroSPARC processor can be interfaced easily to the *S-bus* architecture. Memory management and DRAM control circuitry is included on-chip. A data bus of 32 or 64 bits is provided, and up to 128 MB of system memory can be addressed. The MicroSPARC processor has a 32-entry, fully-associative TLB. It uses a five-stage pipeline, and is compatible with version 8 SPARC architecture. It supports seven register windows. The MicroSPARC processor contains static circuits that can allow the system clock to be stopped with no loss of state.

The MicroSPARC-II processor is also a version 8 architecture, and includes the MMU, DRAM controller, *S-bus* controller, and power management circuitry. This processor implements eight register windows, and has a 16-KB fully-associative, virtually addressed and tagged instruction cache memory, and an 8-KB direct-mapped data cache memory. There is a 4-entry write buffer, and a 64-entry fully-associative TLB. The MicroSPARC-II processor uses 32-bit virtual addresses, 31-bit physical addresses, and eight address spaces.

6.9.5. SPARClite

SPARClite is a family of embedded processors manufactured by Fujitsu. The five-stage pipeline was maintained, but an eighth register window was added, for a total of 136 32-bit registers. Using a static CMOS technology, the power consumption at idle is minimal. The clock can be stopped to conserve power.

The initial device, MB86930, included instruction cache and data cache memories of 2 KB each. These cache memories are 2-way set associative. The processor has an internal Harvard architecture. It has fast 32 by 32 multiply, and 64 by 32 divide operations. A new SCAN instruction can check how many leading bits of a value are the same. No floating-point hardware is included. The MB86931 processor resembles the MB86930. It has an interrupt controller for 15 channels, four 16-bit timers, and dual synchronous/asynchronous data ports.

The MB86932 model expands the on-chip instruction cache memory to 8 KB. There is a 16-entry TLB and a DMA controller for two channels. Write operations are buffered. Although virtual memory is usually not needed in embedded appli-

cations, the protection mechanisms provided by the MMU can be useful. The write policy for the cache memory is write-through.

The MB86933 model has a 1-KB, direct-mapped instruction cache memory, and six register windows, for a total of 104 registers. It supports external memory with minimum interface logic. The MB86933 processor has a DRAM controller and can use 8-, 16-, or 32-bit wide memory. It is a fully static design. The hardware multiplier operates in five clock cycles for 32-bit data. There is a one word instruction prefetch buffer, and a one word write buffer. An on-chip interrupt controller handles the four interrupt lines. A companion chip, the MB86940, provides priorities for 15 interrupts, four 16-bit counter/timer circuits, and two 8251 USART ports.

The MB86934 processor includes floating-point operations compatible with the IEEE 754 standard. It has an 8-KB instruction cache memory and a 2-KB data cache memory, both 2-way set associative. The cache memories can be filled in burst mode. The processor contains a DMA controller for two channels, and eight register windows.

6.9.6. UltraSPARC-I

The UltraSPARC family started with the UltraSPARC-I processor in 1994. It was developed by Sun Microsystems according to the SPARC V9 architecture, and manufactured by Texas Instruments. The UltraSPARC-I processor includes a nine-stage pipeline, and has a speculative superscalar architecture that can issue up to four instructions per clock cycle. Up to 18 instructions may be executed speculatively before branch resolution. The integer unit has two ALUs, and the floating-point unit has three floating-point and two graphics functional units. Graphics operations are provided for image scaling, rotating, and smoothing. Dynamic branch prediction is used, and a power-down mode is included, reducing power consumption to 20 mW. There is a 16 KB instruction cache memory, with secondary cache memory support. The MMU contains an instruction TLB and a data TLB, with 64 entries each.

Sun's UltraSPARC-I was the first general purpose microprocessor which introduced multimedia and networking extensions. These extensions are called VIS (*Visual Instruction Set*), representing accelerating instructions for multimedia applications. Intel announced its MMX extensions 18 month later.

6.9.7. UltraSPARC-II

The UltraSPARC-II processor represents the second generation in the UltraSPARC family. A complete 64-bit implementation of the SPARC V9 architecture, the UltraSPARC-II processor is compatible at the code level with the previous versions of SPARC processors. The main characteristics of this processor are: scalability from common workstations up to supercomputers, high integer and floating-point performance, and high reliability.

The UltraSPARC-II processor microarchitecture has built-in multiprocessing capability for systems with up to 64 processors, therefore ensuring scalability at the

system level. The processor supports multiple speeds and sizes (256 KB to 16 MB) for the L2 cache memory. The operating frequency range of the processor is between 250 MHz and 480 MHz.

A high memory bandwidth is ensured by using the high-performance UPA (*Ultra Port Architecture*) bus. The peak bandwidth to main memory is 1.6 GB/s with an UPA bus at 100 MHz. The VIS multimedia extensions significantly improve the performance of multimedia applications with intensive graphics operations, and thus reduce overall system costs by eliminating the need for a special-purpose multimedia processor. The UltraSPARC-II processor delivers superior computing performance by using a superscalar and superpipelined RISC architecture, and a 0.25 micron technology. The block load and store instructions ensure a high performance access to large data sets across the network. The performance of the processor at 450 MHz is 17.4 SPECint95 and 25.7 SPECfp95.

The UltraSPARC-II processor helps the system designer by implementing industry-standard testing interfaces. It also uses Error Correction Code (ECC) and parity bit to increase system reliability.

6.9.8. UltraSPARC-IIi

6.9.8.1. Overview

Sun's UltraSPARC-IIi processor incorporates a CPU, PCI bus interface, and memory controller in a single chip, and was designed to deliver a high performance at a low system implementation cost. Its high performance makes this processor suitable for network servers. The UltraSPARC-IIi is also an optimal solution for the embedded applications which include new generations of Internet communication systems and advanced image processing technology. The UltraSPARC-IIi processor is available at frequencies of 270, 300, 333, 360, 440, and 480 MHz.

The UltraSPARC-IIi processor series addresses traditional system performance degradation by integrating the complex, high speed communication subsystem within the processor. By minimizing memory-access latency and maximizing I/O throughput, the overall system performance is significantly improved. System design is simplified by integrating all high-speed interconnects into the processor. Sun also offers a family of UltraSPARC-IIi modules with fully integrated external cache memory.

The UltraSPARC-IIi processor incorporates a high-performance dual-mode 33/66 MHz PCI bus at 3.3 V that is version 2.1 PCI bus-compliant, and has a data throughput of up to 200 MB/s in PIO (*Programmed Input/Output*) mode. This interface can be connected to Sun's Advanced PCI Bridge (APB) chip to provide two fully independent 33 MHz PCI buses at 5 V that support a broad family of devices, including Super IO interface cards, SCSI network adapters and graphics cards. Up to four of the optional APB chips can be used to allow the connection of up to 32 PCI devices.

The UltraSPARC i-series processors allow a dual-mode operation of the cache memory. The standard performance mode enables the use of widely-available, lower cost cache memory, while the high performance mode supports the use of high-

speed cache memory to optimize overall system performance. The UltraSPARC-IIi processor contains a data cache memory of 16 KB and an instruction cache memory of the same size. The processor uses a 144-bit DRAM data bus with 8-bit ECC on each 64-bits of data.

The performance of the UltraSPARC-IIi processor at 333 MHz with 2 MB external cache memory is estimated to 14.2 SPECint95 and 16.9 SPECfp95.

6.9.8.2. Block Diagram

Figure 6.6 presents the functional block diagram of the UltraSPARC-IIi processor.
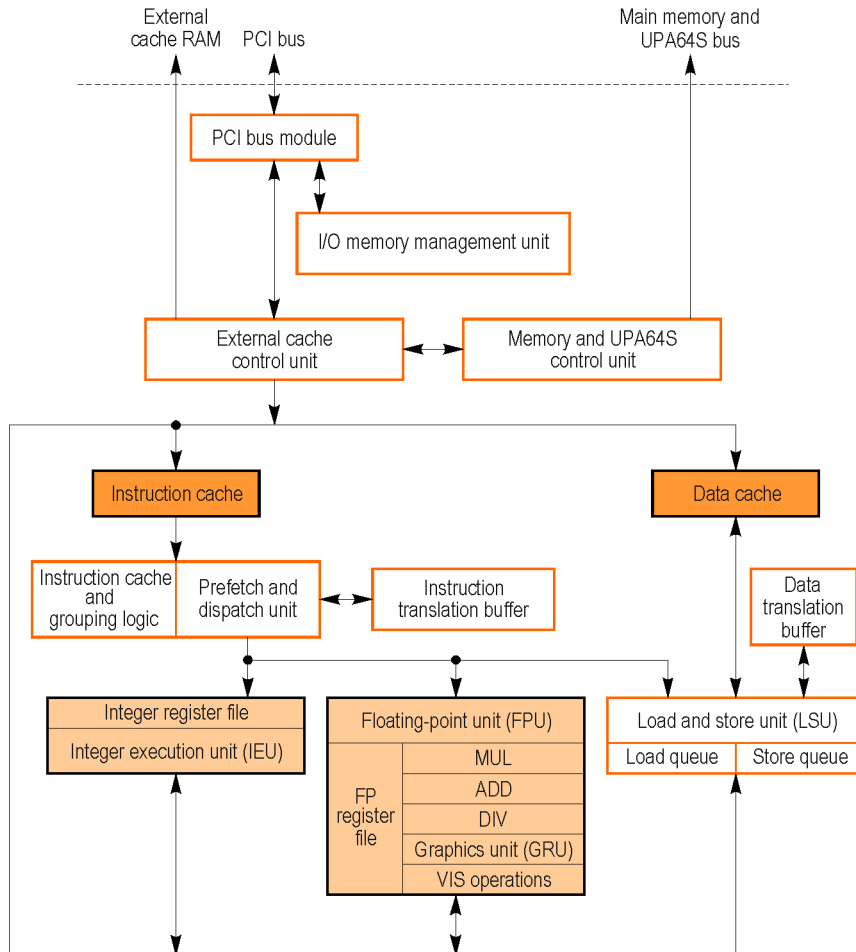

Figure 6.6. Functional block diagram of the UltraSPARC-IIi processor.

The processor consists of a Prefetch and Dispatch Unit, an Integer Execution Unit, a Floating-point Unit, an I/O Memory Management Unit, a Memory Control Unit, a Load and Store Unit, an External Cache Unit, a Graphics Unit, an Instruction Cache and a Data Cache.

6.9.8.3. Prefetch and Dispatch Unit

The *Prefetch and Dispatch Unit* (PDU) ensures that all execution units remain busy by fetching instructions before they are needed in the pipeline. Instructions can be prefetched from all levels of the memory hierarchy, including the instruction cache memory, external cache memory, and main memory.

The PDU was designed with several features to support the high performance requirements of the architecture:

- A 12-entry prefetch buffer isolates instruction prefetching from instruction dispatch and prevents pipeline stalls;
- A 16 KB 2-way set-associative instruction cache memory that is physically addressed and tagged;
- Pre-decoded instructions in the instruction cache memory;
- A 9-stage instruction pipeline to minimize latency;
- Dynamic branch prediction to allow for greater prediction accuracy.

The UltraSPARC architecture uses a pipeline with nine stages. In the first stage of the pipeline, the instruction is fetched from the instruction cache memory. In the second stage, instructions are decoded and placed in the instruction buffer. The third stage groups and dispatches up to four instructions. Next, integer instructions are executed and virtual addresses calculated during the execution phase. In the fifth stage, the data cache memory is accessed. In the next stage, cache hits and misses are determined, and branches are resolved. If a cache miss was detected, the missing line is stored into the load buffer. In the seventh stage, the integer pipeline waits for the floating-point and graphics pipeline to fill. In the next stage, traps are resolved. In the final stage, all results are written to the register files.

The UltraSPARC architecture uses dynamic branch prediction to speed up the processing of branches. The branch prediction mechanism is based on a two-bit state machine that performs the prediction based on the most recent history of the specific branch. To implement dynamic branch prediction, a two-bit prediction information is maintained for every two instructions in the instruction cache memory. This structure ensures that every branch in the instruction cache memory has a prediction information. The UltraSPARC-IIi processor maintains state information for up to 2,048 branches, a number that exceeds the needs of most applications.

In order to optimize superscalar processor performance, the *branch following* technique is used. This technique represents the ability to rapidly fetch predicted branch targets. The UltraSPARC architecture provides single-cycle branch following through the target prediction mechanism in the instruction cache memory.

The dynamic branch prediction and branch folding mechanisms employed by the UltraSPARC architecture in conjunction with Sun's WorkShop optimizing compilers provide superior performance with less complexity of the circuits. For example, with the UltraSPARC-IIi processor, 88% of integer instruction branches and 94% of floating-point instruction branches are successfully predicted.

6.9.8.4. Integer Execution Unit

The *Integer Execution Unit* (IEU) is designed to maximize performance while maintaining full software compatibility. This unit comprises the following:

- Two ALUs for arithmetic, logical, and shift operations;
- An integer multiplier;
- An integer divider;
- 8-window register file.

The UltraSPARC processors attempt to execute all instructions in a single cycle, with all instructions issued to the instruction pipeline at the highest possible rate. Because the processors are pipelined, latencies may occur. Once started, execution proceeds at a rate of up to four instructions per clock cycle and per execution unit. Load and store operations may be executed in one clock cycle, double word load and store operations in two clock cycles, and all floating-point store operations, both single and double precision, in one clock cycle.

6.9.8.5. Floating-point Unit

The *Floating-point Unit* (FPU) is a pipelined floating-point processor that conforms to SPARC-V9 architecture specifications. It is IEEE-compliant and consists of five separate functional units for floating-point and multimedia operations. The separation of execution units enables the UltraSPARC processors to issue and execute two floating-point instructions per clock cycle. Source data and results are stored in a 32-entry register file. Most floating-point instructions have a throughput of one clock cycle and a latency of three clock cycles. The FPU is able to execute operations in hardware on both single-precision (32-bit) and double-precision (64-bit) numbers, normalized or denormalized, and in software on quad-precision (128-bit) operands.

The IEU and the FPU have a dedicated control interface which performs the dispatch of instructions fetched by the PDU to the FPU. Once in the instruction queue, the PDU is responsible for distribution of instructions to the FPU. The IEU controls the part of the operation that implies the data cache, while the FPU decides how to manipulate the data. The FPU performs all floating-point operations and implements a 3-entry floating-point instruction queue to improve overall performance.

6.9.8.6. I/O Memory Management Unit

The *I/O Memory Management Unit* (IOM) handles all memory operations, implements virtual memory and translates virtual addresses of each running process to physical addresses in memory. The logical (virtual) address space is partitioned into pages which are mapped into physical (real) memory. The operating system translates a 64-bit address into a 44-bit address space supported by the processor. The IOM provides the translation of a 44-bit virtual address to a 41-bit physical address through the use of a Translation Look-aside Buffer (TLB).

The IOM also provides memory protection so that a process can be prohibited from reading or writing to and from the address space of another process. Access protection is also supported to ensure that any given process does not gain unauthorized access to memory. For example, a process will not be allowed to modify areas that are marked as read-only or reserved for supervisory software.

Finally, the IOM performs the arbitration function between references to memory by I/O devices, data cache memory, instruction cache memory, and TLB. At any given time, a contention for memory access may arise between an I/O access involving the bus as well as internal accesses requested by references to the instruction cache, data cache, and TLB.

The TLB is a 64-entry, fully associative cache memory that contains page descriptors, used for virtual to physical address translations. The TLB also stores the page protection and usage information. For the address translation, the virtual address to be translated by the memory management unit is compared to each entry in the TLB. If the address to be translated is found in the TLB, the corresponding physical address is read from the table. If the address is not found, the memory management unit generates a trap and the missing address is loaded by the software.

6.9.8.7. Memory Controller Unit

The *Memory Controller Unit* (MCU) implements the function of access control to main memory, establishing priorities. All transactions to the system, such as external cache memory misses, interrupts, and cache memory writebacks to main memory are handled by the MCU. Additionally, this unit manages I/O transactions between the processor and the memory. These transactions include read and write operations from the processor to PCI devices, as well as DMA read and write operations from the devices to the processor.

All transactions with the DRAM and UPA64S subsystem are also handled by the MCU. This unit communicates with the memory system at a frequency lower than the frequency of the processor (the ratio is typically 1/4). The MCU provides support for multiple outstanding read and write requests to the *Ultra Port Architecture* (UPA) bus.

6.9.8.8. Load-Store Unit

The *Load-Store Unit* (LSU) is responsible for generating the virtual address of all load and store operations and for accessing the data cache memory. The LSU uses a load queue and a store queue. One load or one store operation can be issued per clock cycle. To further optimize data stores, two or more store operations are combined if they are in the same 8-byte block, so that a single data transfer occurs between the processor and the second-level cache memory. This frees up the data bus allowing transfers with this cache memory to be processed more rapidly.

6.9.8.9. Data and Instruction Cache Memories

The UltraSPARC architecture employs separate instruction and data cache memories. The data cache memory (D-cache) has a capacity of 16 KB, is direct-mapped and uses the write-through strategy. This memory is organized as 512 lines with two 16-byte sub-blocks of data per line. Each line has a cache tag associated with it. On a data cache miss, 16 bytes of data are written into the cache memory from main memory.

The D-cache is virtually-indexed and virtually-tagged. Virtual addresses delivered to the data cache memory are used when attempting to find the required word in the cache memory. The least significant part of the virtual address is used to access one line of the cache (direct-mapped) that may contain the required word. The most significant part of the virtual address is then compared with the tag address bits for a possible match. This method ensures that cache misses are quickly detected, and that addresses are translated only on a cache miss.

The instruction cache memory (I-cache) has a capacity of 16 KB and it is 2-way set-associative. It is organized as 512 lines of 32 bytes of data, with each line having an associated cache tag. On an I-cache miss, 32 bytes of data are written into the cache memory from main memory. The I-cache is physically indexed and physically tagged. It returns up to 4 instructions from an 8 instruction-wide line.

6.9.8.10. External Cache Unit

The *External Cache Unit* (ECU) handles I-cache and D-cache misses. The ECU can perform one access to the external cache memory (E-cache) every other cycle. The size of the external cache memory can be 256 KB, 512 KB, 1 MB, or 2 MB, where the line size is always 64 bytes.

The ECU also provides support for the 2-2-2 and 2-2 modes of external cache memory operation. In 2-2-2 mode, accesses to the external cache memory are pipelined and take six clock cycles (two to send the address, two to access the SRAM, and two to return the data), providing 16 bytes of instructions or data per clock cycle. In 2-2 mode, the latency is four clock cycles. These low latencies can effectively make the external cache memory a part of the pipeline. For programs with large data sets, this means that data can be maintained in the external cache memory and instructions

scheduled with load latencies. Floating-point applications can use this feature to "hide" D-cache misses.

6.9.8.11. Graphics Unit

The *Graphics Unit* (GU) ensures optimal performance for advanced multimedia applications, such as video conferencing, 2D and 3D image processing, image recognition, animation, and virtual reality applications. The UltraSPARC processors were the first general-purpose processors that delivered the computing power needed for the advanced graphics and real-time video applications. Allowing to decompress and process video information with high resolution in real time, these processors can eliminate the need for a dedicated video processor.

The GU uses the integer registers for addressing image data, and the floating-point registers for processing that data. This division of operations enables the processor to make full use of all available internal registers, maximizing graphical throughput. The GU uses 64-bit data paths which allow 8-, 16-, and 32-bit fixed-point data to be operated on in parallel. It accelerates graphics operations by performing up to 10 operations per clock cycle. To quickly manipulate images, the GU uses block load and store operations.

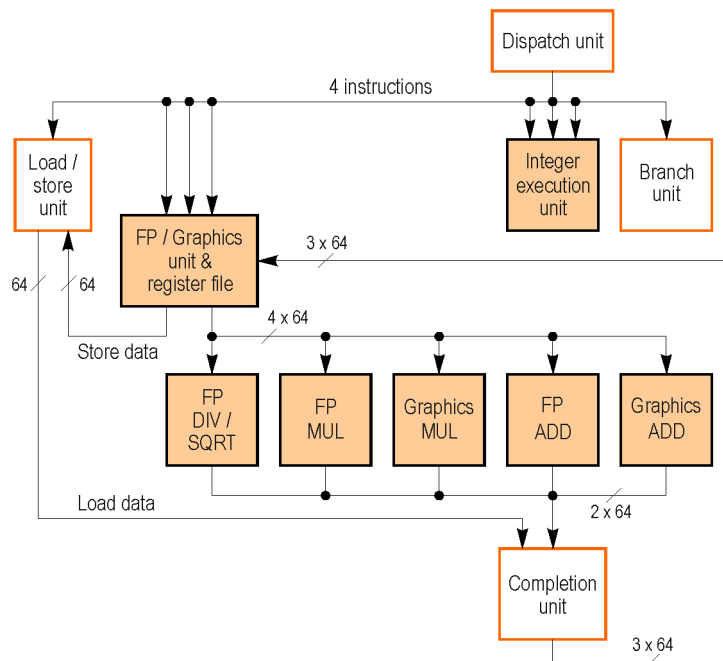The block diagram of the graphics unit is presented in Figure 6.7.

Figure 6.7. The graphics unit of the UltraSPARC processors.

6.9.8.12. The Visual Instruction Set

Specialized instructions that execute complex multimedia and networking operations can significantly increase a processor's performance. The *Visual Instruction Set* (VIS) enables the execution of complex graphics operations to be performed in a single clock cycle. VIS instructions can be categorized as follows:

- *Image processing.* VIS includes image processing instructions to accelerate the scaling and rotating of images, pixel interpolation, filtering, and volumetric rendering. In particular, the *Partitioned Add/Subtract* instruction allows the addition or subtraction of four 16-bit or two 32-bit values to be completed in one clock cycle. The *Partitioned Multiply* instruction allows four 16-bit values to be multiplied by four 8-bit values in a single clock cycle. The *Array Addressing* instruction is useful for efficiently traversing volumetric data sets used in 3D visualization. Each pixel in a 3D image has an x, y and z component. The *Array Addressing* instruction converts the 3D fixed-point addresses into a blocked-byte address, making it possible to easily move along any line or plane. The instruction specifies an element size of 8, 16 or 32 bits. This operation, which typically requires 24 instructions, is accomplished with one instruction by the graphics unit.

- *Video compression.* VIS incorporates special instructions to accelerate video compression and decompression, including the H.261, JPEG, MPEG-1, and MPEG-2 algorithms. Other VIS instructions perform efficient color space conversion, eliminating the need for separate external hardware. The advantages of VIS can be illustrated by examining the benefit of motion estimation on MPEG compression. Motion estimation compares the values of blocks of pixels between two video frames and computes a motion vector. This process typically requires eight pixel subtracts, eight absolute values, eight pixel adds, eight loads, eight aligns, and eight additional adds. The inclusion of the motion estimation instruction reduces these 48 instructions to only one.

- *Pixel format and conversion.* Pixel format and conversion instructions allow the processor to operate directly on pixel data. The pixel data can be stored as 8-bit or 16-bit values. In particular, the *Pixel Expand* instruction performs the conversion of four 8-bit integer values to four 16-bit integer values and stores the results as one 64-bit word. The *Pixel Pack* instruction performs the conversion of four 16-bit or two 32-bit values to four 8-bit or two 16-bit values, and also stores the results as one 64-bit word. Use of these instructions can result in pixel manipulation operations typically taking up to 34 instructions being reduced to only one, greatly improving application performance.

- *Data transfer and animation speed-up.* VIS takes advantage of the block load/store features of the UltraSPARC processors that allows to perform 64-byte load and store operations directly from the CPU to main memory, from main memory to the video memory, or from the video memory to main memory.

6.9.9. UltraSPARC-III

The third generation of the UltraSPARC family of processors is represented by the UltraSPARC-III processor. Initially at 600 MHz, and later also available at 750 MHz, the UltraSPARC-III processor is highly scalable, enabling systems with up to 1000 processors, such as those that will be used in future generations of high-performance servers in the enterprises and the telecommunication market. The processor maintains complete Solaris operating system and application software compatibility.

Using Sun's VIS instruction set for accelerating networking, multimedia, imaging, and Java applications, the UltraSPARC-III processor offers a two- to three-times performance improvement, without recompilation, compared to previous processor generations. This performance is possible due to the processor's memory bandwidth of 2.4 GB/s, which is more than twice as high as previous processors. With 8 MB of second level cache memory, the 600 MHz UltraSPARC-III processor delivers 6 billion operations per second, having an estimated SPECint95 of 35 and SPECfp95 of 60. The floating-point performance indicator of the processor is a measure of its superior performance for computing on large data sets, such as in computer aided design, computer aided manufacturing, electronic design automation and scientific modeling applications.

The UltraSPARC-III processor will make an important contribution to the performance of Sun's new Solaris 8.0 Operating Environment. This processor offers significant design innovations in massive multiprocessing, *Non-Uniform Memory Access* (NUMA), and *Scalable Shared Memory* (SSM) architectures. The CPU is able to handle non-uniform memory accesses that allow the operating system to take advantage of the CPU's multi-processing and multi-threading features. The CPU has a built-in memory controller allowing high-bandwidth memory and I/O device access, which contributes substantially to optimizing system-level performance. The *Scalable Shared Memory* feature of the UltraSPARC-III processor provides coherency across all processors in the system. With the SSM feature, systems with the UltraSPARC-III processor can quickly expand from just a few processors to hundreds of processors without having to rewrite applications or use additional circuitry. The Solaris Operating Environment is the ideal operating environment to take full advantage of SSM.

The UltraSPARC-III processor is targeted for network servers as part of the Internet infrastructure, for high-performance desktop computers and for enterprise and workgroup servers. Due to its large memory bandwidth and scalability for multi-processing, the UltraSPARC-III offers significant advantages for applications such as electronic commerce, scientific computation and data bases.

6.9.10. MAJC

MAJC (*Microprocessor Architecture for Java Computing*, pronounced "Magic") is a new architecture introduced in 2000 by Sun Microsystems. The MAJC architecture is designed to meet the need for high-speed processing demanded by on-line multimedia

applications, telecommunication networks, electronic commerce, digital voice processing, and advanced image processing. The MAJC architecture supports high-speed processing and transmission of digital audio information, video images, voice, and data over digital networks. Advanced technologies which can be based on the MAJC architecture include interactive TV, digital telephony, 3D animation, and virtual reality simulations.

The designers of the MAJC architecture employed the advances in both hardware and software technologies to address the new computational challenges. These challenges come from the increase in communication bandwidth and the techniques used to transfer and process the data. The advances in the semiconductor technology increased the number of transistors in a chip and the operating frequency. Software technology improvements in compilers, Java applications, and multi-threaded applications make it possible for high level language development of multimedia and telecommunication infrastructure-based applications to meet the rapid devel1opment time requirements.

The MAJC architecture is scalable and uses parallelism at several levels. At the highest level of parallelism, the MAJC architecture provides support for multiple processors on a chip. The next level consists in the possibility of using the technology of vertical *micro-threading*, which is achieved through hardware support for rapid context switching. The next level comes from the instruction-level parallelism via an improved very long instruction word (VLIW) architecture. The instruction packets can vary in length, up to 128 bits, with a maximum of four instructions each of 32 bits per packet. The lowest level of parallelism comes from single instruction multiple data (SIMD) parallelism.

Sun's MAJC-5200 processor is the first implementation of the new MAJC architecture. It is the first multiprocessor system on a chip, suitable for multimedia applications. The processor targets communication devices, client platforms, and application servers delivering digital data and running Java applications. The MAJC-5200 processor includes on a single chip two CPUs, a memory controller, a PCI controller, two fast I/O controllers, a graphics preprocessor, a data transfer engine (DTE), and a centralized crossbar switch.

The two CPUs of the MAJC-5200 processor are running at 500 MHz which provide a performance of 6.16 GFLOPS and 12.33 GOPS. The processor includes the main features of the MAJC architecture, which allows to obtain a high performance multimedia processor. The main design principles of the MAJC-5200 processor are the following:

- Exploiting a new level of parallelism for the current generation of processors, namely the thread-level parallelism.

- Mapping the most important multimedia algorithms onto a high-speed general-purpose microprocessor architecture with digital signal processing (DSP) capabilities.

- Dedicating a higher percentage of the chip area to information processing (visual and audio data) rather than maintaining complex state information

(through large reorder buffers and register renaming structures). On the MAJC-5200 processor, a large portion of the chip is dedicated to adders, multipliers, shifters, bit and byte operations, etc.

- Supporting a multi-threaded processor architecture that provides multi-threading at all levels (operating system, Java Virtual Machine, application, hardware).

- Simplifying the hardware design where possible except when performance would be compromised.

The block diagram of the MAJC-5200 processor is illustrated in Figure 6.8.
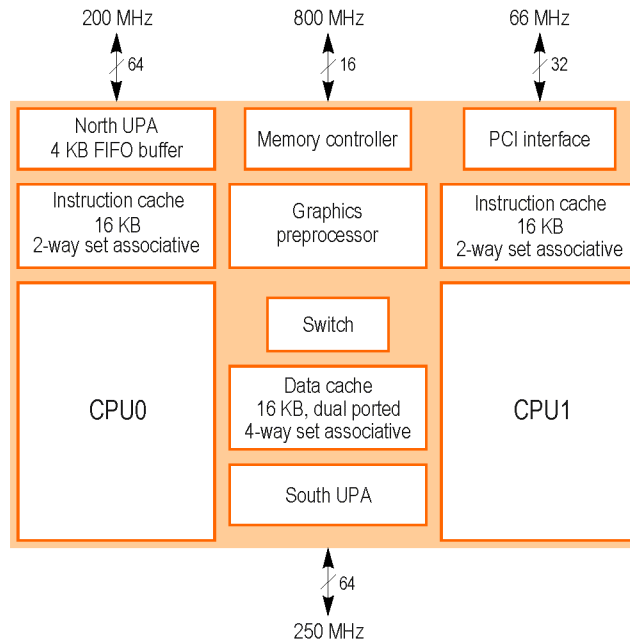

Figure 6.8. Block diagram of the MAJC-5200 processor.

Each of the two CPUs of the MAJC-5200 processor can issue four instructions in a clock cycle, and contains its own 2-way set-associative instruction cache memory of 16 KB. The two CPUs share a 4-way set-associative data cache memory of 16 KB and common external interfaces.

A high throughput bandwidth is obtained by a multitude of interfaces with built-in controllers. The main memory is a Direct Rambus DRAM (DRDRAM) with an interface with a peak transfer rate of 1.6 GB/s. A 32-bit PCI interface at 66 MHz provides DMA and programmed I/O (PIO) transfers with a rate of up to 264 MB/s. There are other two 64-bit interfaces with a transfer rate of up to 4 GB/s at 250 MHz: North and South UPA (*Universal Port Architecture*). The North UPA block contains a 4 KB input FIFO buffer that can be accessed by both CPUs.

A Data Transfer Engine (DTE) provides DMA transfer capabilities among the various memory and I/O devices, with the bus interface unit acting as a central switch. Another specialized block in the chip is the graphics preprocessor (GPP). The GPP has built-in support for real-time 3D geometry decompressing, and load balancing between the two CPUs.

The MAJC-5200 processor can be used in a variety of products that include high-end graphics systems, telecommunication infrastructure and document processing. For 3D graphics processing, MAJC-5200 has two features that significantly enhance its performance: the GPP and the two CPUs. The GPP decompresses polygon information and distributes the uncompressed information to the CPUs using a load balancing mechanism. The geometry transformation and lighting are then performed using the CPUs. This pipelined architecture delivers a performance between 60 and 90 million triangles per second.

The performance in video and image processing applications is mainly achieved using the instruction set that is suitable for several operations: Direct Cosine Transform (DCT), MPEG-2 decompression, motion estimation, convolution, color conversion. The versatile bit and byte operations are useful for the variable length decoding, common for video and image decompression. Motion estimation used by a video encoder is significantly sped up via the byte permutation and pixel distance operations. The large register file aids in convolution operations since the filter coefficients, image data, and the intermediate values can be easily stored in registers avoiding memory operations.

6.9.11. Summary

The main features of the SPARC architecture are the following:

RISC Features

- Large register set – register windows;
- Load/store architecture;
- Hardwired instructions;
- Small instruction set;
- Single-cycle execution.

    Non-RISC Features

- Multiple addressing modes.