floating-point registers. A reorder buffer with 16 elements is used as well to support speculative execution. The register file has 12 ports. Although instructions can be executed out-of-order, in-order completion is guaranteed.

The processor uses dynamic branch prediction, by means of a 512-entry, direct-mapped branch history table (BHT). On the basis of recent execution history, the branch is classified as probable or not probable, taken or not taken. The BHT then loads a 64-entry branch target address cache memory, which is fully associative. There is no penalty for a wrong prediction, but there is a benefit for a correct prediction.

The completion unit contains a 6-entry reorder buffer. This unit retires an instruction from the reorder buffer when all instructions ahead of it have been completed and the instruction has finished execution. The completion unit can retire as many as four instructions per clock cycle.

The instruction and data cache memories are each 32 KB, and 4-way set-associative. The LRU replacement algorithm is used. The cache memories are physically indexed, and use the MESI consistency protocol. The write-back or write-through policy is selectable at the page or block level. Software disabling of the cache memories is also provided, and they can be locked against updates. The 604e processor includes a controller for a direct-mapped, unified, external secondary cache memory of up to 128 MB.

The 604e processor has two memory management units. This processor uses an instruction TLB and a data TLB with 128 entries, which are 2-way set-associative. Address translation is provided for 4-KB pages, variable-sized blocks, and 256-MB segments. There are 16 segment registers.

The 604e processor has a 64-bit external data bus and a 32-bit address bus. The interface protocol allows multiple masters to compete for system resources through a central external arbiter. Additionally, there is a logic for maintaining data cache memory coherency for multiprocessor applications. The 604e processor supports burst data transfers for memory accesses and memory-mapped I/O accesses.

6.11.6. PowerPC 740 and 750

The PowerPC 740 and 750 processors are targeted for high-performance, low-power systems. The 740 processor operates at several frequencies between 200 MHz and 500 MHz, while the 750 processor operates at frequencies between 200 MHz and 550 MHz. The 750 processor differs from the 740 primarily in its extensive L2 cache memory support. It contains a bus interface and a controller for L2 cache memory, which are not present in the 740 processor.

This section describes in detail the PowerPC 750 processor. Unless otherwise noted, references to the 750 processor also apply to the 740 processor.

6.11.6.1. Overview

The PowerPC 750 is a superscalar processor that can complete two instructions simultaneously. It incorporates the following six execution units:

- Floating-point unit (FPU);
- Branch processing unit (BPU);
- System register unit (SRU);
- Load/store unit (LSU);
- Two integer units: IU1 executes all integer instructions, and IU2 executes all integer instructions except multiply and divide instructions.

Most integer instructions execute in one clock cycle. The floating-point unit is pipelined, and the tasks it performs are broken into subtasks, implemented in three successive stages. Typically, a floating-point instruction can occupy only one of the three stages at a time. Thus, three single-precision floating-point instructions can be in execution at a time. Double-precision add instructions have a latency of three clock cycles; double-precision multiply and multiply-add instructions have a latency of four clock cycles.

The 750 processor has independent 32-KB, 8-way set-associative, physically addressed cache memories for instructions and data, and independent instruction and data memory management units (MMUs). Each MMU has a 128-entry, 2-way set-associative translation look-aside buffer (DTLB and ITLB) that holds recently used page address translations. Block address translation is done through the 4-entry instruction and data block address translation (IBAT and DBAT) arrays, defined by the PowerPC architecture. During block address translation, effective addresses are compared simultaneously with all four BAT entries.

In the 750 processor, the L2 cache memory is implemented with an on-chip, 2-way set-associative tag memory, and with external, synchronous SRAM memories for data. The external memories are accessed through a dedicated L2 cache memory port that supports a single bank of up to 1 MB of synchronous SRAM memories. The L2 cache memory interface is not implemented in the 740 processor.

The 750 processor has a 32-bit address bus and a 64-bit data bus. Multiple devices compete for system resources through a central external arbiter. The MEI cache-coherency protocol supports the modified, exclusive, and invalid states, a compatible subset of the MESI four-state protocol.

The 750 processor has four software-controllable power-saving modes. When functional units are idle, a dynamic power management mode causes those units to enter a low-power mode automatically without affecting operational performance, software execution, or external devices. The 750 processor also provides a thermal assist unit and a way to reduce the instruction fetch rate for limiting power consumption.

The 750CX version includes an on-chip L2 cache memory, with a size of 256 KB. It is organized as a 2-way set-associative cache memory, with 64-byte line size.

Table 6.1 presents the main features of the 740 and 750 processors manufactured by IBM.

Table 6.1. IBM PowerPC 740 and 750 Processors.

	740 (PID8p)	750 (PID8p)	750CX
Frequency	300, 333, 366, 400, 466, 500 MHz	300, 333, 366, 400, 466, 500 MHz	350, 400, 450, 500, 550 MHz
Technology (CMOS)	0.22 μm	0.22 μm	0.18 μm
Performance	400 MHz	400 MHz	500 MHz
SPECint95	16.0	19.2	20.9
SPECfp95	9.2	13.1	12.8
Dhrystone MIPS	928	928	1160
Typical Power	3.7 W (400 MHz)	4.7 W (400 MHz)	4.0 W (400 MHz)
Voltage (logic / I/O)	3.7 V / 3.3 V	2.0 / 3.3 V	1.8 / 1.8 or 2.5 V
I-Cache / D-cache	32 KB / 32 KB	32 KB / 32 KB	32 KB / 32 KB
L2 Cache	–	–	256 KB
Bus Interface	64-bit data 32-bit address	64-bit data 32-bit address	32/64-bit data 32-bit address

6.11.6.2. Block Diagram

Figure 6.17 shows a block diagram of the PowerPC 750 processor. It contains an instruction unit, a completion unit, two integer units with an integer register file, a floating-point unit with a floating-point register file, a load/store unit, a system register unit, two memory management units for instructions and data, two L1 cache memories for instructions and data, a bus interface unit, and an L2 cache memory interface.

6.11.6.3. Instruction Unit

The instruction unit contains an instruction fetch unit, a 6-entry instruction queue (IQ), a dispatch unit, and a branch-processing unit (BPU).

The instruction fetch unit loads instructions from the instruction cache memory into the instruction queue. The instruction queue holds as many as six instructions and loads up to four instructions from the instruction cache memory during a single processor clock cycle. All instructions except branch instructions are dispatched to their respective execution units from the last two positions of the instruction queue at a maximum rate of two instructions per clock cycle. The dispatch unit checks for source and destination register dependencies, determines whether a position is available in the completion queue, and prevents subsequent instruction dispatching as required.

The BPU extracts branch instructions from the instruction fetch unit. Branch instructions that cannot be resolved immediately are predicted using either the architecture-defined static branch prediction, or the dynamic branch prediction specific to the 750 processor. The BPU folds branch instructions when a branch is taken (or pre-

dicted as taken). Branch instructions that are not taken, or predicted as not taken, are removed from the instruction stream through the dispatch mechanism.

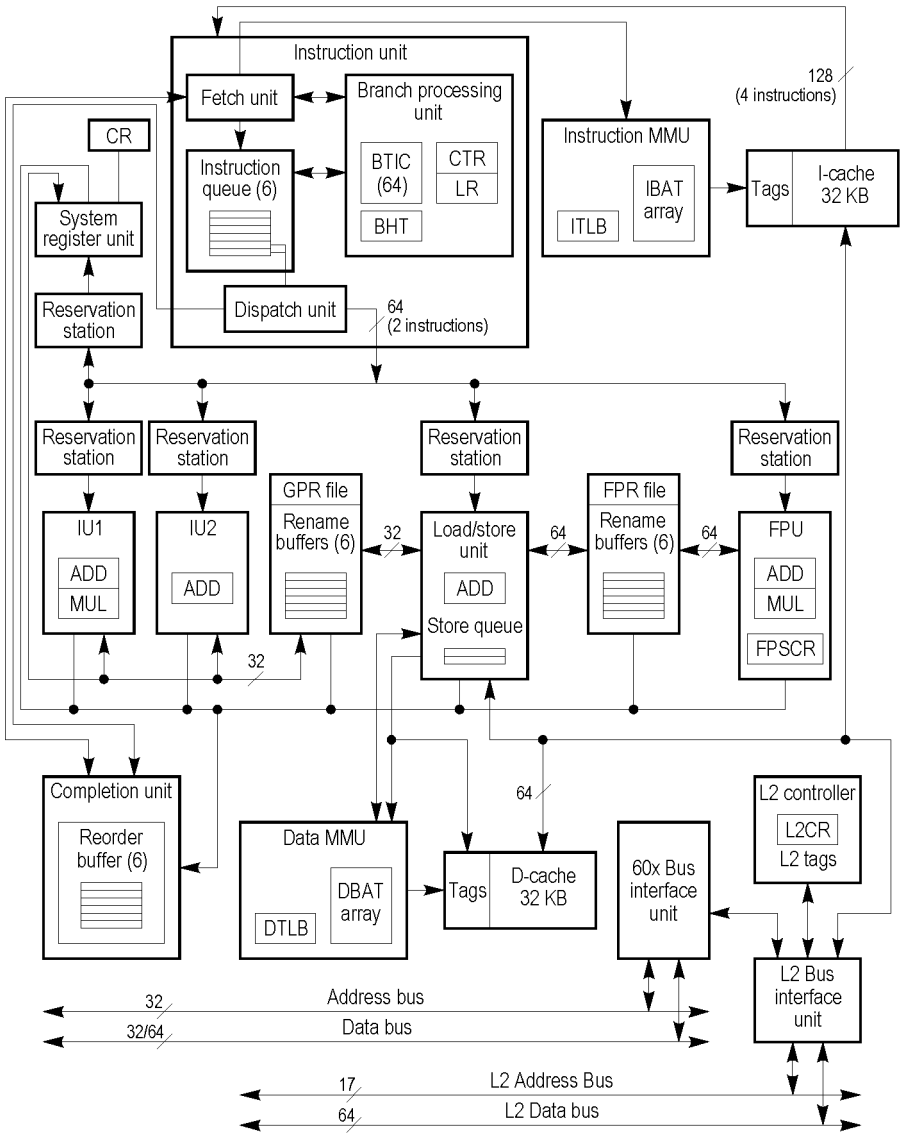

Figure 6.17. Block diagram of the PowerPC 750 processor.

Dynamic prediction is implemented using a 512-entry branch history table (BHT), a cache memory that contains two bits per entry that together indicate four levels of prediction for a branch instruction. When dynamic branch prediction is disabled, the BPU uses a bit in the instruction encoding to predict the direction of the conditional branch. Therefore, when an unresolved conditional branch instruction is

encountered, the 750 processor executes instructions from the predicted target stream, although the results are not written to the registers until the conditional branch is resolved. This execution can continue until a second unresolved branch instruction is encountered.

The branch target instruction cache (BTIC) is a 64-entry, 4-way set-associative cache memory that contains the most recently used branch target instructions. When a target instruction is in the BTIC, it is fetched into the instruction queue in the next clock cycle, a clock cycle sooner than it would be available from the instruction cache memory. Typically, the BTIC contains the first two instructions in the target stream. The BTIC can be disabled through software.

The branch processing unit contains an adder to compute branch target addresses and three control registers: the link register (LR), the count register (CTR), and the control register (CR). The BPU calculates the return address for subroutine calls and saves it into the LR for certain types of branch instructions. The LR also contains the branch target address for the *Branch Conditional to Link Register* instruction. The CTR contains the branch target address for the *Branch Conditional to Count Register* instruction. Because the BPU uses dedicated registers, execution of branch instructions is independent from execution of integer and floating-point instructions.

6.11.6.4. Completion Unit

Instructions are fetched and dispatched in program order. At the point of dispatch, the program order is maintained by assigning each dispatched instruction a successive entry in the 6-entry reorder buffer (completion queue). The completion unit tracks instructions from dispatch through execution and retires them in program order from the reorder buffer.

Instructions cannot be dispatched to an execution unit unless there is a vacancy in the reorder buffer. Branch instructions that do not update the CTR or LR registers are removed from the instruction stream and do not take an entry in the reorder buffer. Branch instructions that update the CTR and LR registers follow the same dispatch and completion procedures as non-branch instructions, except that they are not issued to an execution unit.

Completing an instruction commits execution results to the GPR, FPR, LR, or CTR registers. In-order completion ensures the correct architectural state when the processor must recover after a mispredicted branch or any exception. Retiring an instruction removes it from the reorder buffer.

6.11.6.5. Integer Units

Each integer unit, IU1 and IU2, has a single-entry reservation station that can receive instructions from the dispatch unit and operands from the general-purpose registers or the rename buffers. The IU1 can execute any integer instruction; the IU2 can execute any integer instruction except multiplication and division instructions.

Each integer unit consists of three subunits: a fast adder/comparator, a subunit for logical operations, and a subunit for performing rotates, shifts, and count-leading-zero operations. These subunits handle all one-cycle arithmetic instructions. Only one subunit can execute an instruction at a time. Each integer unit has a dedicated result bus that connects to rename buffers.

The IU1 has a 32-bit integer multiplier/divider as well as the adder, shift, and logical subunits of the IU2. The multiplier supports early completion for operations that do not require full 32x32-bit multiplication.

6.11.6.6. Floating-Point Unit

The floating-point unit (FPU) is designed such that single-precision operations require only a single pass, with a latency of three clock cycles. As instructions are dispatched to the FPU's reservation station, source operands can be accessed from the floating-point registers (FPRs) or from the rename buffers of these registers. Results in turn are written to the rename buffers and are made available to subsequent instructions.

The FPU contains a single-precision multiply-add array and the floating-point status and control register (FPSCR). The multiply-add array allows the 750 processor to efficiently implement multiply and multiply-add operations. The FPU is pipelined so that one single-precision or double-precision instruction can be issued per clock cycle. Thirty-two 64-bit registers are provided to support floating-point operations. Stalls due to contention for FPRs are minimized by automatic allocation of the six floating-point rename buffers. The contents of the rename buffers are written to the appropriate FPR when floating-point instructions are retired by the completion unit.

The 750 processor supports all the floating-point data types of the IEEE 754 standard (normalized, denormalized, NaN, zero, and infinity) in hardware, eliminating the delay implied by the exception routines.

6.11.6.7. Load/Store Unit

The load/store unit (LSU) executes all load and store instructions and provides the data transfer interface between the registers and the memory subsystem. The LSU calculates effective addresses, performs data alignment, and provides sequencing for load/store string instructions.

Load and store instructions are issued in program order; however, some memory accesses can occur out of order. Synchronizing instructions can be used to enforce strict ordering. When there are no data dependencies, at most one out-of-order load operation can execute per clock cycle. Data returned from the cache memory is held in a rename buffer until the completion logic commits the value to a GPR or FPR. Store operations cannot be executed out of order; the data is held in the store queue until the completion logic signals that the store operation can be performed. The 750 processor executes store instructions with a maximum throughput of one per clock. The time required to perform the actual load or store operation depends on the

processor/bus clock ratio and whether the operation involves the on-chip cache memory, the L2 cache memory, the system memory, or an I/O device.

6.11.6.8. System Register Unit

The system register unit (SRU) executes various system-level instructions, as well as condition register logical operations and transfers between special-purpose registers. To maintain system state, the execution of the instructions by the SRU is serialized; that is, the instruction is held for execution in the SRU until all previously issued instructions have executed. Results from these instructions executed by the SRU are not available or forwarded for subsequent instructions until the instruction completes.

6.11.6.9. Memory Management Units

The two memory management units (MMUs) support up to 4 Petabytes (2^{52}) of virtual memory and up to 4 GB (2^{32}) of physical memory for instructions and data. These units also control access privileges for the memory spaces on block and page levels. Referenced and changed status is maintained by the processor for each page to allow the implementation of demand-paged virtual memory systems.

The load/store unit calculates effective addresses for data loads and stores, and the instruction unit calculates effective addresses for instruction fetching. The MMU translates the effective address to determine the correct physical address for the memory access.

The 750 processor supports the following types of memory address translation:

- *Real addressing mode* – In this mode, translation is disabled by clearing bits in the processor state register. When address translation is disabled, the physical address is identical to the effective address.
- *Page address translation* – Translates the page frame address for a 4-KB page size.
- *Block address translation* – Translates the base address for blocks (128 KB to 256 MB).

If translation is enabled, the appropriate MMU translates the higher-order bits of the effective address into physical address bits. The lower-order address bits are directed to the on-chip cache memories where they form the index into the tag memory. After translating the address, the MMU passes the higher-order physical address bits to the cache memory. For accesses that miss in the cache memory, the untranslated lower-order address bits are concatenated with the translated higher-order address bits; the resulting 32-bit physical address is used by the memory unit and the system interface, which accesses external memory.

Instruction and data TLBs provide instruction and data address translation in parallel with the on-chip cache memory access. The TLBs store page address translations for recent memory accesses. For each access, an effective address is presented for page and block address translation simultaneously. If a translation is found in both the TLB and the BAT array, the block address translation in the BAT array is used. When a page address translation is not in a TLB, a hardware search is performed in the page table. The two TLBs are 128-entry, 2-way set-associative cache memories.

6.11.6.10. On-Chip Cache Memories

The 750 processor implements separate instruction and data cache memories. Each cache memory is 8-way set-associative, with a size of 32 KB, and is physically addressed. Each block contains eight contiguous words from memory; a cache block never crosses a page boundary. An entire cache block can be updated by a burst load. Non-aligned accesses across a page boundary can incur a performance penalty. The write policy is write-back.

6.11.6.11. L2 Cache Memory

The L2 cache memory is unified and receives memory requests from the L1 instruction and data cache memories independently. The L2 cache memory is implemented with an on-chip, 2-way, set-associative tag memory, and with external, synchronous SRAM memories for data. The external SRAM memories are accessed through a dedicated L2 cache memory port that supports a single bank of 256 KB, 512 KB, or 1 MB of memory. The L2 cache memory normally operates in write-back mode.

Depending on its size, the L2 cache memory is organized into 64-byte or 128-byte lines, which in turn are subdivided into 32-byte blocks (sectors), the unit at which cache coherency is maintained.

The L2 cache memory controller contains the L2 cache control register (L2CR), which includes bits for enabling parity checking, setting the L2-to-processor clock ratio, and identifying the type of RAM used for the L2 cache memory implementation. The controller also manages the L2 cache memory tag array, 2-way set-associative with 4 KB tags per way. Each block has its own valid and modified status bits.

6.11.6.12. Bus Interface Unit

The main activity of the bus interface unit (BIU) is transferring data and instructions between the processor and system memory. The bus interface is compatible with the 60x bus. There are two types of memory accesses:

- *Single transfers.* These memory accesses allow transfers with sizes of 8, 16, 24, 32, or 64 bits in one bus clock cycle. Single transactions are caused by read

and write operations that access memory directly (when caching is disabled), and store operations in write-through mode.

- *Burst transfers.* Burst transactions always transfer an entire cache memory block (32 bytes). Because the L1 cache memories on the 750 processor use the write-back strategy, burst-read operations are the most common memory accesses, followed by burst-write operations.

Access to the bus interface is granted through an external arbitration mechanism that allows devices to compete for the bus. This arbitration mechanism is flexible, allowing the 750 processor to be integrated into systems that implement various bus parking procedures to avoid arbitration overhead.

Typically, the sequences of memory operations do not necessarily complete in the order they begin. This maximizes the efficiency of the bus without sacrificing data coherency. The 750 processor can dynamically optimize run-time ordering of load/store operations, improving overall performance.

6.11.7. PowerPC 7400

6.11.7.1. Overview

The Motorola 7400 PowerPC processor (MPC7400) is a high-performance, low-power, 32-bit implementation of the PowerPC architecture, combined with a full 128-bit implementation of Motorola's AltiVec instruction set. The MPC7400 processor supports the high-bandwidth MPX bus, and, to maintain compatibility with existing applications, it also supports the 60x bus protocol. The MPC7400 processor provides symmetric multiprocessing (SMP) capabilities, and supports up to 2 MB of external L2 cache memory.

The AltiVec technology expands the capabilities of the MPC7400 processor by allowing high-bandwidth data processing and computationally-intensive operations. It provides high performance for multimedia-oriented desktop computers (scientific, medical, etc.), and digital video processing. The AltiVec technology satisfies the computational demands of networking infrastructure such as multichannel modems and echo cancellation equipment. This technology also enables faster and more secure encryption methods optimized for the SIMD processing model. Other applications are real-time MPEG-2 encode, speech recognition, and high-resolution 3D graphics.

While the MPC7400 processor is software-compatible with existing applications for PowerPC 603e, 740, and 750 processors, to utilize the advantages of the AltiVec technology, some instruction changes in existing source code are required to interface with the vector execution unit.

The MPC7400 is a superscalar processor that can fetch up to four instructions per clock cycle from the instruction cache memory. It can dispatch and complete two instructions simultaneously. As many as eight instructions can execute per clock cycle (including two integer instructions and four AltiVec instructions).