Design of the VAX 4000 Model 400, 500, and 600 Systems

By Jonathan C. Crowell, Kwong-Tak A. Chui, Thomas E. Kopec, Samyojita A. Nadkarni, and Dean A. Sovie

1 Abstract

The design of Digital's NVAX CPU chip provided the opportunity to bring RISC-class performance to deskside CISC VAX computer systems. The new VAX 4000 Model 400, 500, and 600 low-end systems take full advantage of the performance capabilities of the NVAX microprocessor. The three systems offer from two to four times the performance of the previous top-of-the-line VAX 4000 Model 300 system in the same deskside enclosure. To achieve this increased performance, Digital's systems engineers designed a new high-performance memory controller chip as part of the CPU module, whose basic design is shared by the three systems. In addition, a high-performance memory module and a VLSI bus adapter chip were designed.

2 Introductin

The design of Digital's NVAX high-performance microprocessor offered systems engineers the opportunity to design computer systems with significantly improved performance.[1] The project structured to use the NVAX CPU chip to upgrade the VAX 4000 line of low-end deskside computers resulted in a family of three new systems and the associated CPU modules. These systems share a basic CPU module design but offer a range of performance capabilities. This paper first presents the goals of the development project and then describes the architecture, design, and implementation of the resulting new VAX 4000 Model 400, 500, and 600 systems.

3 Goals of the VAX 4000 Project

Schedule and performance goals were of prime concern to the engineers committed to upgrading the VAX 4000 family of computers. Time-to-market was a key goal of the development project. Consequently, fully qualified systems were ready to be shipped to customers when the NVAX CPU chip was released and available in volume. The initial goals for performance specified that the new systems provide three times the performance of the VAX 4000 Model 300 system. Ultimately, the performance of the NVAX CPU chip exceeded its design goals. As a result, the new top-of-the-line VAX 4000 Model 600 system performance is four times that of the Model 300.

Achieving the performance goals required designing a new, high-performance memory controller chip called the NVAX data and address lines (NDAL) pin bus memory controller (NMC). The objectives were to double the memory bandwidth of the VAX 4000 Model 300 system and to provide a total system memory capacity of 512 megabytes (MB). To support the NMC specifications, a

high-performance memory module called the MS690 was designed.[2]

To reduce hardware and software development cost and the risk of failing to meet project schedules, the system design incorporated all existing high-performance I/O adapter chips. These devices include the second-generation Ethernet controller chip (SGEC), the Digital Storage Systems Interconnect (DSSI) shared-host adapter chip (SHAC), the CVAX Q22-bus interface chip (CQBIC), and the system support chip (SSC).[2,3,4] A very large-scale integration (VLSI) bus adapter chip was required to provide CVAX pin (CP) buses to connect these I/O devices.[5] The NDAL-to-CP bus adapter chip (NCA) was designed to meet this need.

The three CPU modules designed to upgrade existing VAX 4000 Model 300 systems retain the same BA440 system enclosure used for these systems. The upgrade requires that the older MS670 memory module used in the Model 300 be replaced with the new, higher-performance MS690 memory module. The new systems had to support all of the Q-bus option modules that were supported on the VAX 4000 Model 300.

4 System Overview of the VAX 4000 Models 400, 500, and 600

The BA440 system enclosure shown in Figure 1 supports the VAX 4000 Models 300, 400, 500, and 600. This pedestal enclosure was designed to operate in an open office environment. To allow the systems to operate quietly, the cooling fans are speed controlled, based on the ambient temperature. The enclosure power supply provides 644 watts of direct current from a standard 15-ampere wall circuit. The system was designed and qualified to operate in an environment with a temperature range of from 10 to 40 degrees Celsius.

The new CPU modules, differentiated only by the part numbers KA675, KA680 and KA690, are utilized as the engines for the VAX 4000 Models 400, 500, and 600, respectively. All three CPU modules, henceforth referred to as the CPU module, provide the same I/O functionality, including two DSSI buses, a thick-wire and ThinWire Ethernet adapter, a Q-bus adapter, and the console serial line. The CPU module performance is 16, 24, and 32 times the performance of the well-known VAX-11/780 system, for the three new VAX 4000 systems, respectively. The CPU cycle clock speed and cache sizes determine the product performance, as discussed in the section CPU-cache Subsystem.

The backplane in the system enclosure provides the signal interconnection and power distribution between system components. There are connectors and slots for the CPU module, four slots for MS690 memory modules, and seven Q-bus slots. The CPU module has a 270-pin connector that receives the module power and connects the CPU to the NVAX memory interconnect (NMI) bus, the system DSSI bus, and the Q-bus. The backplane was modified to support the wider 72-bit data path of the new MS690 memory modules. This new backplane was phased into the BA440, enabling most VAX 4000 Model 300 systems to be upgraded without requiring a backplane change.

The system enclosure supports up to four DSSI or small computer system interface (SCSI) tape integrated storage elements (ISEs). These cableless bricks support either one 5.25-inch, full-height drive or two 3.5-inch drives. The ISEs are available in variants that support the 2-gigabyte (GB), RF73 DSSI disk drive and dual 85MB, RF35 DSSI disk drives. The single-system pedestal can support six RF35 devices and a tape drive, providing 4.8GB of storage for applications that require high I/O rates. This RF35 configuration can provide over 360 queued I/Os per second for random I/Os. If RF73 drives are used, the single-system box can provide 8GB of storage.

There are several ways to expand the base VAX 4000 system; the most common way is to expand to another DSSI-based system and create a two- or three-node DSSI VAXcluster. The Q-bus in the VAX 4000 system can be expanded to provide 10 additional Q-bus slots to each system using the B213A Q-bus expansion enclosure. The DSSI expansion enclosures together with the Q-bus DSSI adapter (KFQSA) can expand the total available disk storage to 28 DSSI disks. Using the RF73 disk allows up to 56GB of disk storage.

5 CPU Module

The CPU module common to the three new VAX 4000 systems is based on a highly integrated CPU and I/O system built on the single 21.6-by-26.7-centimeter (8.5-by-10.5-inch) module shown in Figures 2 and 3. The CPU module printed wiring board (PWB) consists of the following subsystems: a central processor and its associated three-level cache; a pin bus, bus adapter, and memory controller; and an I/O system with integrated controllers for DSSI and Ethernet buses. The CPU module also contains a CQBIC and 512KB of field erasable programmable read-only memory (FEPROM) for console code.

NOTE

Figure 2 (CPU Module) is a photograph and is unavailable.

CPU-cache Subsystem

The CPU-cache subsystem is built around the single-chip NVAX CPU, which provides a three-level cache architecture. The first two levels of cache, which are contained on the chip, include a 2KB virtually addressed instruction cache and an 8KB physically addressed instruction and data cache. The third level of cache, the backup cache, is constructed using static random-access memories (SRAMs) on the module and is completely controlled by the NVAX CPU chip. The backup cache was designed to support a CPU cycle time as low as 10 nanoseconds (ns) with a slip cycle, i.e., a two-cycle read (20 ns) using 8-ns SRAMs. This write-back caching architecture significantly reduces the demands on main memory by caching

both reads and writes without the need for a memory access. On all previous VAX 4000 systems, the caches required that all write operations continue through to main memory, i.e., write through.

The NVAX CPU is clocked by a differential emitter-coupled logic (ECL) surface acoustic wave oscillator. This oscillator runs at 250 megahertz (MHz) (16-ns cycle time), 286 MHz (14-ns cycle time), or 333 MHz (12-ns cycle time) on the KA675, KA680, and KA690 CPU modules, respectively. The NVAX chip produces a four-phase internal clock directly from this input and generates system clocks at one-third the internal clock rate.

The new CPU module design supports either a 128-kilobyte (KB) or a 512KB backup cache. (512KB for the KA690 module and 128KB for the KA680 and KA675.) The tag store for the two cache sizes can be constructed from SRAMs that have the same 24-pin package with compatible pinouts. This packaging makes it easy to design the PWB to support either cache. However, the data store, which is constructed from parts whose packages have incompatible pinouts, required a special dual footprint design to accommodate either 16K-by-4K or 64K-by-4K SRAMs. This footprint is designed with the decoupling capacitors and series edge-limiting resistors carefully placed in the outline of the footprint to allow a dense, geometric packing of the 18 RAMs necessary for the data store.

The backup cache is a write-back cache with memory coherence maintained through a directory-based broadcast coherence protocol.[1] When the NVAX CPU needs to write data to memory, the data is first transferred into the backup cache with a request for write privilege command. Once a cache row is stored in the cache as "written," any direct memory access (DMA) read to that memory address of displacement of that cache row will result in a release write privilege transaction, even if the data was never actually written. The cache controller inside the NVAX CPU is responsible for all activities related to cache maintenance.

NDAL Interconnect, Memory Controller, and Adapter

The NDAL bus is a synchronous, multiplexed address and data, pended interconnect. Each device on the NDAL bus may be a commander (request data transfer), a responder (respond to commander requests), or both. In the new VAX 4000 systems, the NVAX CPU chip is only a commander, the NMC is only a responder, and the NCA I/O adapter is both a commander and a responder.

Arbitration of the NDAL interconnect is performed by the NMC, which is also the default master responsible for driving valid no-operation bus cycles when there is no master activity. The NVAX CPU is responsible for watching all NDAL traffic and performing any invalidates or write backs of primary and backup cache data required to maintain cache coherence with memory.

I/O Buses

The CPU module uses a custom third-generation complementary metal-oxide semiconductor (CMOS-3) process I/O adapter (the NCA chip) to interface

between the NDAL bus and a pair of 32-bit CP buses. Two CP buses are used to prevent long response latencies on the Q-bus from interfering with buffer management on the Ethernet interface. One CP bus, CP1, operates under the synchronous CP bus protocol, and the other, CP2, uses the asynchronous protocol. The faster peripherals, the Ethernet and the two

DSSI adapter chips, reside on CP1 and can take advantage of optimizations in the NCA to reach a peak bus bandwidth of 33MB per second (MB/s). CP2 has only one peripheral DMA master, i.e., the CQBIC, which also connects the SSC and the console FEPROM to the system. The NCA acts as a master on both CP1 and CP2. Since only one of the buses is asynchronous, the system uses only one CP bus clock (CCLK) chip for signal synchronization and CP clock distribution.

The two CP buses share the same clocks. Consequently, arbitration is performed by a single programmable sequencer, which serves both buses. The sequencer is clocked at 35 ns (KA680 and KA690) or 40 ns (KA675); this is one-half the CP cycle time. The arbitration for CP2 is a simple two-priority scheme with the CQBIC at the higher priority. When more than one master is requesting the bus, the minimum DMA request deassertion times on both the CQBIC and the NCA effectively make this scheme behave like a round-robin arbiter. The internal state does not have to keep track of the previous master. No special treatment is required for lock cycles because the CQBIC will never perform a lock on behalf of the Q-bus.

6 Signal Integrity

Signal integrity work began very early in the project, and the effort was a close collaboration between the CPU module design team, the design teams for the three VLSI devices, and the VAX 6000 Model 600 CPU module design team. Because some CPU module team members had experience with designing CP bus modules, the signal integrity issues on the CP buses were generally well understood. The CP bus data lines were routed with only a length constraint. The control signals on CP1 required significant analysis and SPICE modeling to meet both settling time and waveform requirements.

The most critical signals in the backup cache are the output enable and write enable signals. The output enable deasserting edge must be transitioned quickly to avoid tri-state contention on the cache data lines. The write enable signal must be perfectly monotonic through the threshold region, because it is an edge-sensitive signal. Both of these requirements were met by using a strong driver in the NVAX chip and a parallel R-C termination at the far end of two of the three stubs. The R-C terminations absorb some of the incident energy, reducing the reflections to an acceptable amount and allowing incident-wave switching without the reflected wave reentering the threshold region.

The backup cache data store was designed with strong drivers and incident-wave switching on the address lines. The routing of a representative cache address signal is shown in Figure 4. The stubs were arranged in such a way that the unavoidable reflection from the far end of the lines was reduced by a partial reflection from the center junction. Thus, signals traverse the threshold region fairly cleanly and settle rapidly outside the

threshold region, i.e., approximately 3 ns elapse from the beginning of the transition at the driver until the time when a valid signal arrives at the receiver.

7 Printed Wiring Board Physical Design

The NVAX CPU chip draws several amperes of current from the 3.3-volt power supply. This current draw has significant high-frequency components. The integrated decoupling capacitor on the NVAX die helps eliminate some of the high-frequency current pulses, but much of this current must be supplied by the module-level decoupling capacitors.[1]

Charge stored on the module in these decoupling capacitors supplies this current. Any inductance in the path of the current reduces the effectiveness of the capacitors by limiting the rate-of-change of the current. In addition, the larger the physical area enclosed by the current path, the more radio frequency (RF) energy will be radiated into space that must be contained by the enclosure in order to meet regulatory radiation requirements. These two issues led to the exploration of how to minimize both the inductance of the decoupling path and the physical area of the RF current spread.

The internal PWB standard, as it existed when the new CPU module was being designed, required a minimum of 25 mils (0.025 inch) of surface etch on any device before a via could be dropped into an inner layer. Traditionally, the inductance of this connection was reduced by using a wide (i.e., 25-mil) etch for this connection. The inductance of surface etch on the module lay-up used on the new CPU module (calculated with two-dimensional transmission line [TDTL]) is shown in Table 1.

Table 1 provides the data to calculate the total inductance of a pair of 25-by-25-mil etch segments, i.e., approximately 0.4 nanohenrys (nH). (This measurement is approximate, due to the short dimensions of the segments.) The effective series inductance of the high-quality, 1,000-picofarad (pF) RF capacitor used on the CPU module is approximately 1 nH, including the inductance of the vias connecting the capacitors to the power planes. The reactance as a function of frequency for the inductive component of this decoupling system is shown in Table 2.

Because the resulting impedance is still quite high at upper frequencies, multiple capacitors are used in parallel. The 1,000-pF capacitors are chosen from two different case styles to ensure that the parasitic inductance of the capacitors is not identical for all the high-frequency decoupling capacitors. This method of selection staggers the frequency of the parasitic resonances.

8 Cycle Design Goal and Testing

Although the original design goal for the CPU module was a 12-ns CPU cycle time, as knowledge about Digital's fourth-generation complementary metal-oxide semiconductor (CMOS-4) process increased, the design teams

investigated the critical paths for a 10-ns operation. At this time, the CPU module had not been routed, so a 10-ns cycle time was set as the layout goal. The cache loop layout resulted in a measured requirement that RAMs

have an access time of approximately 8.5 ns to meet worst-case timing with no added slip cycles.

RAMs meeting this specification were not readily available when the CPU module was designed. However, several vendors are beginning to ship devices at this speed today. The NDAL data lines were capable of running at the 30-ns NDAL cycle that is generated with a 10-ns CPU clock. The point-to-point NDAL arbitration signals are the tightest timing path on the NDAL. The combination of analysis by the NMC team and a careful hand-routing of these signals by the module team allowed appropriate NMC speed binning (i.e., sorting the chips based on correct operation at the fastest possible speed) to meet the timing requirements for a 30-ns NDAL cycle goal.

Very few NVAX CPU chips were available that would function at 10 ns over the full range of voltage and temperature. However, empirical signal-delay measurements and limited-range module testing have proven that the NMC and the NCA are ready to operate on the CPU module at this speed. An NVAX CPU that functions at this 10-ns speed can operate the cache with no slip cycles using the faster SRAMs. Future products may be based on an NVAX running at a 10-ns cycle, as sufficient yields at this speed bin are reached.

9 Module Testing

The module and chip teams considered more than one approach when determining what module-level testability features to implement in the VLSI devices Digital was building for VAX 4000 Model 500 computers. The scan-based Test Access and Boundary Scan Architecture (JTAG) proposal was coming into its own, and the teams desired to follow that specification, if scan-based test features were to be used.[6] However, no other devices on the module would have scan capabilities. Thus, the overall module test strategy could not be based entirely on the JTAG specification.

As the teams reviewed the overall module design, certain issues appeared to show promise for the application of a scan-based test:

1. The automatic test equipment (ATE) pin density was likely to be very high in the areas of the three 339-pin pin grid arrays (PGAs). Reducing this high density would improve the reliability of the test fixturing.

2. Previous experience with module manufacture in Digital's plants showed that the risk for solder defects would be high in the cache area, because of the J-lead SRAMs. A fast way to isolate these problems would help reduce debug time.

3. Providing at least a driver or a receiver for use by the JTAG scan ring would eliminate the need to use continuity structures to verify bonding

and solder integrity on the VLSI parts.

Eventually, the module and chip teams settled on a subset implementation of the JTAG scan-based test. The NVAX CPU chip implements scan latches on all data and control pads (input and output, in the case of bidirectional pads). Thus, the cache SRAMs can be tested using only the JTAG port, and the NVAX CPU can act as the driver for scan testing of the NDAL interface. The NMC and the NCA implement receive-only scan, so that the NDAL interface could be tested with no ATE pins required. The external tester was able to test the remaining pins solely by driving signals that could be scanned out of the pad latches. This subset implementation provided the same module-level coverage as would have been possible using a full JTAG implementation. In addition, the implementation removed some design obstacles that were causing implementation problems in the chips.

The ability to use scan-based testing is advantageous to the manufacturing process in the following two areas:

1. The scan tester can find open circuit defects in the cache area where the bed-of-nails tester could not resolve whether the problem was a fixture contact problem or an actual open circuit.

2. The ability to create "virtual test points" on scanned nets has allowed the test coverage of the bed-of-nails tester to be expanded without having to purchase an expensive tester upgrade.

Unfortunately, the module and chip teams' previous experience with scan-based testing at the module level had been spotty at best. The ability of this test to reduce the pin density, therefore, was not used to full advantage in the problem areas under the large PGA devices. Based on the experience testing the CPU module for the three new VAX 4000 systems, follow-on products have been able to use this testing feature to good advantage. The teams now have a firm base of experience on which to base future test strategies.

10 The NVAX memory Controller

The NMC is a 520-by-500-mil custom chip fabricated using Digital's 1-micrometer CMOS-3 process and contains 148,000 transistors packaged in a 339-pin PGA. The NMC provides the interface between the NDAL bus and up to 512MB of main memory by means of the NMI. The NDAL bus supports three other nodes on the NDAL - the NVAX CPU and up to two I/O adapters (IO1 and IO2). In the new VAX 4000 systems, the NCA serves as both I/O nodes on the bus. The NMC contains the arbiter for the NDAL and also helps the NVAX CPU maintain cache-memory coherency in the system by interfacing with a separate O-bit memory.

The NMI can operate with either a 32- or 64-bit-wide data path and supports single error correction, double error detection, and nibble error

detection, and runs synchronous with the NDAL clock. The NMI timing scales with the NDAL clock cycle time.

In this section, we describe the architecture of the NMC, the objectives of the NMC project, and the results of the effort.

NMC Architecture

As shown in Figure 5, the NMC is partitioned into six major sections: the NDAL arbiter, the NDAL interface, the transaction handler, control and status registers (CSRs), the memory interface, and the O-bit interface. The NMC responds to all memory space addresses when NDAL address bit 29 is equal to 0 and responds to I/O space addresses in its allocated range, i.e., 2101 0000 .. 2101 FFFF (hexadecimal).

The NDAL arbiter gives highest priority to the NMC for returning read data. The two I/O nodes have second priority; their requests are handled in a round-robin fashion. The CPU has lowest priority.

The NDAL interface consists of an input section and an output section. The input section monitors the NDAL for a new transaction every cycle. A valid transaction that has been decoded by the NMC is put into one of four transaction queues (INQUEUEs). There is one queue for each of the NDAL nodes: CPU, IO1, IO2, and the fourth, which stores release write privilege transactions. Commander nodes on the NDAL initiate release write privilege transactions to release the write privilege of blocks in memory. The NMC must accept a release transaction from a node, irrespective of the state of its INQUEUE; otherwise, there is a potential for deadlock. Consequently, the NMC has a separate queue for release write privilege transactions. The output section of the NDAL interface buffers up to six quadwords (i.e., six groups of four contiguous 16-bit words for a total of 384 bits) of read data that must be returned to the NDAL. In a normal functioning system, a buffer depth of six quadwords together with an arbitration scheme that gives the NMC the highest priority will never result in a full output queue. Therefore, there was no need to check for a full queue and to stall while loading the queue.

The transaction handler arbitrates between the four INQUEUEs and stores selected transactions in a current transaction buffer. The current transaction buffer serves as a pipeline stage; this buffer allows the corresponding INQUEUE to be loaded with the next transaction while the current transaction is being serviced. The NMC can service back-to-back transactions with no stall cycles on the NMI.

The CSR section of the NMC contains memory configuration registers, error status registers, and mode and diagnostic registers.

The memory interface contains the data path, address path, and control for up to four memory modules on the NMI. The data path contains all the error correction and detection logic. The address path contains the row and column address multiplexers and a refresh address counter. The control is provided by a state machine that can perform multitransfer read operations, multitransfer write operations, and read-modify-write operations.

The O-bit interface directly controls the O-bit dynamic random-access memories (DRAMs), which are housed on the CPU module. For every memory transaction, the NMC reads the corresponding O-bit in parallel with the memory access. If the block of memory is written, the memory transaction is

aborted until the corresponding release transaction is received by the NMC. If the block is unwritten, the memory transaction is allowed to complete. Initiating the memory transaction in parallel with the O-bit access reduces the transaction latency on transactions that are not written. Since most memory accesses are to unwritten locations, using this scheme improves memory performance considerably. In addition, the system design engineers were able to use inexpensive DRAMs to implement the O-bit memory instead of faster, more expensive SRAMs.

NMC Project Objective and Results

The NMC project objective was to create a high-performance memory design that would be compatible with the VAX 4000 Model 300 memory subsystem and could provide two to three times the performance of that subsystem. (The VAX 4000 Model 300 has a bandwidth of 47.4MB/s, at a cycle time of 28 ns, using 100-ns DRAMs with a 32-bit memory data bus.) This goal was achieved by using a 64-bit memory data bus and an interconnect that operates at a cycle time as low as 36 ns, using 100-ns DRAMs, and at an NDAL cycle as low as 30 ns, using 80-ns DRAMs. The asymptotic bandwidth on the NMI using the 100-ns DRAM technology and a 64-bit data path is 111.11MB/s, i.e., 2.3 times the bandwidth of the VAX 4000 Model 300. Using faster 80-ns DRAMs, the bandwidth is 133.33MB/s, i.e., 2.8 times the bandwidth of the VAX 4000 Model 300.

The NMC interface is efficient from the moment it receives a transaction on the NDAL until it starts a transaction on the NMI. This timing path was extremely tight and results in real memory read bandwidth of 63.6MB/s and 76.32MB/s at 36-ns and 30-ns cycle times, respectively.

The NMC chip was designed to meet an NDAL cycle time of 36 ns, which made the timing very critical. Most of the NMC chips produced can exceed this goal and will run at an NDAL cycle time of 30 ns. Future designs based on the 10-ns NVAX chips will require this cycle time.

To meet the performance goal, we chose to use a 64-bit memory interface. However, achieving compatibility with the Model 300 memory modules presented a challenge with respect to the ECC generation and checking mechanism. A simple approach would have been to include two separate ECC trees, one for 64-bit operation and the other for 32-bit operation. This design would have been very area-intensive, so we chose 64-bit ECC code such that 32-bit ECC was a subset. 64-bit ECC requires eight check bits, and 32-bit ECC requires seven check bits. In our 64-bit code, the eighth check bit depends solely on the upper 32 bits. In 32-bit mode, we force the upper 32-bits to a known value; therefore, that check bit is always a fixed value in 32-bit mode.

The VAX 4000 systems do not allow the use of 32-bit memory modules, because

it is difficult to meet Q-bus latency requirements with the slower memory. This system constraint indirectly affected the NMC. The CQBIC and SGEC devices, for example, had stringent low latency requirements. The Q-bus latency problem had to be solved without causing SGEC latency problems. Thus, the latency issue was addressed in the following way:

- The NMC has a mode that indicates whether or not the Q-bus is present in the system. During this mode, the transaction handler gives the Q-bus node (IO2) the highest priority. However, to keep the SGEC latency within required limits, the transaction handler must service transactions from the IO1 node at strategic times.

- To minimize the latency seen by any one node, the three NDAL nodes require separate queues. The simplest implementation of the NDAL input interface would have been to have two queues, one for release write privilege transactions and one for all other transactions. Thus, preserving the order of NDAL transactions would have been very easy. With three queues, it is necessary to compare queue addresses to preserve transaction ordering.

- The NMC combines the CPU request for write privilege, the DMA read, and the DMA write into a single transaction before retiring the data to memory. This optimization reduces the latency of written transactions.

Although the features that were added to the NMC chip to reduce Q-bus and SGEC latency increased the complexity of the chip, these features successfully keep the Q-bus latency below 8 microseconds.

11 NDAL-to-CP Bus Adapter Chip

NCA is a full-custom, high-performance I/O controller chip that provides the electrical and functional interface between the 64-bit NDAL bus and the 32-bit CP bus. In the new VAX 4000 systems, the NDAL supports three chips: the NVAX CPU, the NCA, and the NMC. On the CP bus, the NCA supports the SHAC, SGEC, CQBIC, and SSC chips. The NCA is fabricated in Digital's CMOS-3 process, contains 155,000 transistors, and is packaged in a 339-pin PGA. The design goals for the NCA project were high quality, improved performance, optimized time-to-market, and leveraged use of existing CP bus chips. The NCA team achieved all design goals and completed the project by the scheduled manufacture release date.

NCA Architecture Overview and Partitioning

Although the original concept of the NCA was motivated by a memory and I/O controller chip called the G chip (used in the VAX 4000 Model 300 systems), the NCA chip was a new design. To optimize the DMA to memory bandwidth and the bus access latency, the NCA provides the two CP bus interface ports (CP1 and CP2, mentioned previously), which operate independently. This strategy has three advantages. First, the Q-bus adapter and the system read-only memory (ROM)/console are connected to the CP bus separately from the Ethernet and the mass storage devices. This arrangement allows the system to tune and optimize throughput on the Ethernet and mass storage devices without degrading the bus access latency seen on the Q-bus.

Second, the loading on the two CP buses is reduced. Therefore, each bus can operate at a higher frequency and without external buffers; this also saves module area. Third, the dual-bus structure allows the use of a simpler bus arbitration scheme.

In addition to using the dual-bus strategy, the NCA uses write buffer and read prefetch transactions to allow DMA devices, particularly the SHAC and SGEC chips, to operate efficiently in double octaword mode, where a two-octaword (32-byte) burst of data is transferred within a single bus grant. For write transactions, the NCA buffers up to two octawords of data. Thus, the bus can operate without stalling, while the NCA arbitrates for the NDAL bus for the buffered write transactions. For read transactions, the NCA contains a hexword-size (32-byte) prefetch buffer. Whereas the maximum burst length is only an octaword on the CP bus, the NCA requests up to a hexword of data during DMA memory read operations. The extra data is stored in the prefetch buffer and is immediately available if the subsequent CP bus read transaction targets the same address as the prefetched data. For NVAX initiated I/O, up to four operations can be buffered simultaneously.

The NCA is partitioned into four major sections: the NDAL, CP1, CP2, and registers. The NDAL interface and each CP bus interface contain the master and slave sequencer and controls for the corresponding bus. The NCA chip also has I/O queues and an internal arbiter to select operations from CP1, CP2, and NCA register read transactions to the NDAL bus, based on a predetermined priority.

The register section contains the control and status registers and the interval clock timer registers. The interval clock is a software-programmable timer used by the operating system to account for time-dependent events.

The NCA supports parity check and detection on both the NDAL and the CP buses. The NCA also supports all interrupts defined by the NDAL and CP bus protocols for other types of error conditions. When an error occurs, an error status bit is set in the NCA error status register. Depending on the type of error, an address may be available for diagnostics. The NCA also provides a mechanism to force a parity error condition on any of the buses to help debug the interrupt routines of the operating system software.

Q-bus Latency Support

To reduce latency seen by the Q-bus devices, the NCA provides special logic to gain priority from the NDAL arbiter. The NCA informs the arbiter of the imminent Q-bus operation, for which latency is a concern. When a Q-bus is present in the systems, the NCA is programmed to use the two IDs mode on the NDAL and to enable the Q-bus present bit of the control register. Upon detection of the Q-bus "map" read transaction on the CP bus, the NCA immediately asserts a signal to the NDAL arbiter. The arbiter will not grant the bus to other devices until after the Q-bus read transaction is accepted by the NMC or until the signal is deasserted. Requests from the buffered write at the same interface are masked off until the signal is deasserted. Using this scheme, the Q-bus latency in the new VAX 4000

systems was never more than 8 microseconds.

Enhanced CVAX Pin Bus

The NCA supports the standard bus protocol in both synchronous and asynchronous modes. The existing CP bus protocol does not utilize the maximum bus bandwidth possible with the standard CP bus protocol. The fastest data transfer rate is two cycles per four-byte (i.e., 32-bit) longword, because the two primary signals for the handshake use the same clock phase for the assertion. When the sent signal is detected, it is already too late to generate the received signal within the same cycle. To achieve the one-cycle transfer rate, a modified protocol is used. The received signal is changed to represent a ready-to-receive signal. The received signal is asserted regardless of whether or not the sent signal is asserted. When the sent and received signals are asserted at the same time, both the master and slave devices know that the data was successfully transferred.

This protocol works with the existing CP bus chips and has increased the theoretical bandwidth by up to 66 percent. For an 80-ns CP bus cycle time, the maximum bandwidth is 33.33MB/s.

Testability

To assist module testing, the NCA contains features that comply with the IEEE Standard P1149.1 JTAG testability.[6] At the pin level, five special pins are provided and work in combination with the internal test access port controller inside the NCA and a bed-of-nails tester to perform short- and open-circuit interconnection tests.

12 MS690 Memory Module

The MS690 family of CMOS memory modules was designed to support the memory requirements set forth by the NVAX memory controller.[2] The NMC requires the MS690 memory module to provide a two-way, bank-interleaved, 72-bit data path. In addition, a self-test feature is provided that was used on the VAX 4000 Model 300 memory subsystem. The MS690 module returns a unique board identification signature when polled by the NMC. The module used existing qualified parts and fits on a quad-sized PWB.

A common goal of Digital's Electronic Storage Development (ESD) teams is to utilize a single PWB design to accommodate as many memory sizes as possible. The ESD teams routinely stretch the boundaries of Digital's manufacturing processes to provide world-class memory subsystems. Because memory subsystems form the core of the ESD charter, the ESD teams are uniquely tuned into, and actively shaping, present and future device specifications for all types of random-access devices. This advance and intimate knowledge allows us to build current technology products with the hooks necessary to capitalize on the next generation of storage devices.

The MS690 options are available in 32MB, 64MB, and 128MB sizes and are self-configuring. The MS690 memories communicate with the NMC by way of the private NMI. All control and clocks signals originate off-board via the NMI from the NMC. Up to four memory modules of any density mix may coexist on the NMI with a maximum memory size of 512MB.

The MS690 is an extension of the existing 39-bit MS670 memory product designed for the VAX 4000 Model 300 product line. The DC562 GMX was designed and produced in Digital's Hudson, Massachusetts, plant for the MS670 32MB memory. This GMX is a semi-intelligent, 20-bit-wide, 4-to-1 and 1-to-4 transceiver, with internal test/compare/error logging capabilities for its five I/O ports. The MS670 required eight banks of 39 bits of data, hence the requirement of two GMX chips per module.

The KA670 CPU module used in the VAX 4000 Model 300 transfers 32-bit longwords of data. For every longword, 7 bits of ECC must be allocated, i.e., 8 x (32 + 7) = 312 DRAMs. The CPU module used in the VAX 4000 Model 500 transfers 64-bit quadwords of data. For every quadword, 8 bits of ECC must be allocated, i.e., 4 x (64 + 8) = 288 DRAMs. The MS690 memory is configured as two interleaved bank pairs, each 72 bits wide (64 bits of data plus 8 bits of ECC); all transactions are 72 bits. The memory module supports quadword, octaword, and hexword read/write/read-modify-write transactions. Transactions less than 72 bits, i.e., bytes, words, and longwords, are not supported.

Doubling the data word length is advantageous in two ways: the I/O bandwidth effectively doubles, and 24 fewer DRAMs are required. This last benefit results from the fact that only one additional bit is required to protect 64 bits of data as compared to protecting 32-bit data. The available PWB space allowed room for two additional GMXs to handle the 33 additional data bits. The ability to use the existing GMX integrated circuit eliminated the need for a new, 40-bit-wide, GMX-type VLSI development.

Because DRAMs are edge-sensitive devices, module layout, balanced etch transmission lines, and signal conditioning are extremely important to a quality product. The MS690 design team used a combined total of 18 years of memory design experience along with extensive use of SPICE modeling to determine the optimal PWB layout. The result was a double-sided, surface-mount PWB panel that can accommodate all density variations of the MS690 memory option and thus help control costs by reducing product-unique inventory. All parts, except the bare PCB, are used on products already produced in volume at Digital's Singapore and Galway, Ireland, manufacturing plants.

The MS690-BA memory module, which uses 100-ns 1M-by-1M DRAMs, can support NMC cycle times of 36 ns and 42 ns, respectively, for the VAX 4000 Model

400 and 500 systems. The MS690-CA/DA modules use 80-ns 4M-by-1M DRAMs and can accommodate 30-ns, 36-ns, and 42-ns NMC cycle times.

13 Performance

The CPU I/O subsystems on all three products provide exceptional performance, as shown in Table 3. The pair of DSSI buses on the CPU modules for the VAX 4000 Models 500 and 600 were tested under the VMS operating system performing single-block (512-byte) reads from RF73 disk drives. The read rate was measured at over 2,600 I/Os per second with both buses running. The Ethernet subsystem, based on the SGEC adapter chip, is also very efficient. It has been measured transmitting and receiving 192-byte-long packets at a rate of 5,882 packets per second. Packets 1,581 bytes long can be transmitted at a rate of 9.9 megabits per second.

The performance of the CPU subsystem has traditionally been measured using a suite of 99 benchmarks.[7] Scaling the results against the performance of the VAX-11/780 processor and taking the geometric mean yields the VAX unit of performance (VUP) rating. The processor VUP rating for the new VAX 4000 system with the lowest performance, the Model 400, is twice the VUP rating of the system it is replacing, the Model 300. The two new high-end systems provide three and four times the performance of the Model 300-an impressive performance increase.

The system performance in multistream and transaction-oriented environments was measured with TPC Benchmark A.[8] This benchmark, which simulates a banking system, generally indicates performance in environments that are characterized by concurrent CPU and I/O activity and that have more than one program active at any given time. The performance metric is transactions per second (TPS). The measured performance of the VAX 4000 Model 600 system was more than 100 TPS, tpsA-local. As shown in Table 3, the performance of the new VAX 4000 Model 400, 500, and 600 systems is impressive, even compared to RISC-based systems.

14 Acknowledgements

The authors would like to acknowledge the contributions of the following members of the design teams: Matt Baddeley, Roy Batchelder, Abed Chebbani, Harold Cullison, Todd Davis, Joe Ervin, Nannette Fitzgerald, Avanindra Godbole, Judy Gold, Bryce Griswold, Thomas Harvey, Jim Kearns, John Kumpf, Steve Lang, Dave Lauerhass, Bill Munroe, Chun Ning, Jim Pazik, Cheryl Preston, Matt Rearwin, Dave Rouille, Prasad Sabada, Stephen Shirron, Adam Siegel, Jaidev Singh, Emily Slaughter, Michael Smith, Steve Thierauf, Bob Weisenbach, and Pei Wong.

15 References

1. G. Uhler et al., "The NVAX and NVAX+ High-performance VAX Microprocessors," Digital Technical Journal, vol. 4, no. 3 (Summer 1992, this issue): 11-23.

2. KA680 CPU Module Technical Manual, (Maynard, MA: Digital Equipment Corporation, Order No. EK-KA680-TM-001, 1992).

3. B. Maskas, "Development of the CVAX Q22-bus Interface Chip," Digital Technical Journal, vol. 1, no. 7 (August 1988): 129-138.

4. J. Winston, "The System Support Chip, a Multifunction Chip for CVAX Systems," Digital Technical Journal, vol. 1, no. 7 (August 1988): 121-128.

5. P. Rubinfeld et al., "The CVAX CPU, a CMOS VAX Microprocessor Chip," ICCD Proceedings, (October 1987): 148-152.

6. IEEE Standard Test Access Port and Boundary Scan Architecture, IEEE Standard 1149.1-1990 (New York, NY: The Institute of Electrical and Electronics Engineers, Inc., 1990).

7. VAX Systems Performance Summary (Maynard, MA: Digital Equipment Corporation, Order No. EE-C0376-46, 1991).

8. Transaction Processing Performance Council, TPC Benchmark A Standard Specification (Menlo Park, CA: Waterside Associates, November, 1989).

16 General Reference

DEC STD 032-0 VAX Architecture Standard (Maynard, MA: Digital Equipment Corporation, Order No. EL-00032-00, 1989).

17 Biographies

Jonathan C. Crowell An engineering manager in the Entry Systems Business Group, Jon Crowell was the project leader and system engineer on the VAX 4000 Models 100, 400, 500, and 600 and the MicroVAX 3800, 3900, and 3100 Model 90 systems. He is now working on the design of the next generation of VAX 4000 systems. Previously, Jon worked in the Systems Integration Group qualifying Q-bus devices and DSSI adapters and storage devices. He joined Digital in 1986. Jon received a B.S.E.E. (1981) and an M.S.E.E. (1986) from Northeastern University. He hold six patents and is an active member of IEEE.

Kwong-Tak A. Chui Kwong-Tak Chui is a principal hardware engineer in the Semiconductor Engineering Group. He is working on the design of the I/O controller section of a new CPU. Kwong was the project leader for the I/O controller chip of the NVAX chip set used in the VAX 4000 Model 400, 500, and 600 systems. Since joining Digital in 1985, he has worked on four other VLSI chip design projects for the VAX 3000 and VAX 4000 series computers. Kwong holds a B.S. in computer engineering (1985) from the University of Illinois at Urbana-Champaign and an M.S.E.E. (1989) from Cornell University.

Thomas E. Kopec Principal engineer Tom Kopec was a member of the Entry Systems Business Group that designed the VAX 4000 Models 200 through 600, and the MicroVAX 3500 and 3800 systems. He also led an Alpha AXP development project. Recently, Tom joined the Assistive Technologies Group to work on compact speech-synthesis and text-to-voice systems. He received

a B.S.E.C.E. (1980, with honors), concentrating in microwave engineering and signal processing, and an M.S.E.C.E. (1985), concentrating in image processing and computer graphics, both from the University of Massachusetts at Amherst.

Samyojita A. Nadkarni Sam Nadkarni, a principal hardware engineer in the Semiconductor Engineering Group, is currently the project leader for a set of support chips for the DECchip 21064 processor. She was the project leader for the NMC chip used in the VAX 4000 Model 400, 500, and 600 systems. She also worked on memory controller/bus adapter chips for the VAX 4000 Model 300 and MicroVAX 3500 systems. Sam joined Digital in 1985 and holds a Bachelor of Technology (1983) from the Indian Institute of Technology and an M.S. (1985) from Rensselaer Polytechnic Institute.

Dean A. Sovie Design engineer Dean Sovie joined Digital in 1981 and is a member of the Electronic Storage Development Group. He is currently involved in the battery backup laser memory design. Prior to this work, he contributed to the design of the high-performance memory module used in the VAX 4000 Model 400, 500, and 600 systems. Dean also helped design memories for the VAX 6000, VAX 4000 Model 200, and PDP-11 systems. In addition to his present responsibilities, he is pursuing a degree in electrical engineering at Northeastern University.

18 Trademarks