



# **Increasing Performance in Network Storage with Multi-Processors and High-Speed I/O**

# REVISION HISTORY

<i>Revision #</i>	<i>Date</i>	<i>Change Description</i>
1250-WP100-R	09/26/02	Initial release.

Broadcom Corporation  
16215 Alton Parkway  
P.O. Box 57013  
Irvine, California 92619-7013  
© 2002 by Broadcom Corporation

All rights reserved  
Printed in the U.S.A.

Broadcom<sup>®</sup> and the pulse logo<sup>®</sup> and SiByte<sup>™</sup> are trademarks of Broadcom Corporation and/or its subsidiaries in the United States and certain other countries. All other trademarks are the property of their respective owners.

---

# TABLE OF CONTENTS

<b>Section 1: Introduction</b> .....	1
<b>Section 2: Network Storage Design Challenges</b> .....	2
Processing Performance .....	2
Flexibility .....	3
I/O and Memory Bottlenecks .....	3
Fixed Power and Space Constraints .....	3
<b>Section 3: Multi-Processor for Network Storage Designs</b> .....	4
Unparalleled Processing Performance .....	5
ZBbus .....	6
Advanced Memory Access .....	6
Integrated I/O .....	6
HyperTransport for Faster Chip-to-Chip Data Transfer .....	6
On-Chip Debug Features .....	7
<b>Section 4: Typical Storage Applications for SiByte™</b> .....	8
SAN Switch .....	8
iSCSI Gateway .....	9
InfiniBand® Line Card .....	9
IP Storage and NAS-SAN Bridge .....	10
<b>Section 5: Summary</b> .....	12
SiByte Processor Family .....	12

---

## LIST OF FIGURES

Figure 1: SiByte™ BCM1250 Block Diagram ..... 4

Figure 2: Network Storage and Server Appliances ..... 5

Figure 3: Typical SAN Switch Implementation ..... 8

Figure 4: Artis Microsystems' CPCI-A7200 CompactPCI Single-Board..... 10

Figure 5: Artis Microsystems' CPCI-A7200 Block Diagram ..... 11

## Section 1: Introduction

The rate at which increasingly powerful network storage systems are implemented is accelerating. Data proliferation, Internet and e-mail traffic with media rich attachments, greater sharing of files among heterogeneous network environments, more powerful PCs, and high speed Internet access have created the need for faster, more powerful storage solutions in capacities ranging from Gigabytes to Terabytes. With the typical enterprise's data storage needs doubling every year, IT managers are constantly adding more storage servers and disks and seeking ways to increase both the capacity and throughput of their networked storage. As a result, IT managers are turning to storage area networks (SAN), network attached storage (NAS), and other direct attached storage solutions.

Today's enterprise often features various storage solutions acquired at different times from different vendors, and even recently installed solutions may not have enough input/output (I/O) bandwidth and processing power to keep pace with accelerating capacity and throughput requirements. Growing demand for greater performance, intelligent functionality, scalability and management have created a need for very high performance processors optimized to address the specialized needs of network storage.

## Section 2: Network Storage Design Challenges

Storage system designers and architects face a number of critical technical challenges today. Four issues of particular concern to both designers and customers are:

- Processing performance
- Flexibility
- I/O and memory bottlenecks
- Fixed power and space constraints

Each of these issues is discussed in more detail throughout this section.

---

### PROCESSING PERFORMANCE

Storage systems must be able to handle more complex packet processing than traditional networking systems. Storage protocol mediation requires the termination and regeneration of each protocol; furthermore, packets must be reformatted from one protocol to the other (e.g. IP, SCSI). In more advanced storage applications, the actual implementation and processing of the upper layers of the protocol (SCSI) may have to be performed at the processor level, making storage protocol mediation a highly compute-intensive task.

Two recent developments in storage networking, IP storage, and storage virtualization, offer the potential for significant improvement, cost reduction or easier management, but also add to the requirement for greater processing power. IP Storage, such as iSCSI (Internet Small Computer Systems Interface) enables SCSI traffic to be transported over standard Ethernet networks using the Internet Protocol (IP) rather than using separate Fibre Channel (FC) networks. IP storage solutions simplify the management and deployment of networked storage by leveraging the installed base of Transmission Control Protocol/Internet Protocol (TCP/IP) networks, but the protocol mediation involved in iSCSI requires very heavy-duty packet processing. Storage virtualization separates the logical view of storage from the underlying physical devices, enabling it to be treated as a shared pool through online re-provisioning. This enables volume management and added security functionality but adds to the specialized processing requirements of a network storage system.

Another consideration is the need for high performance processing in both the data plane (in-line) and the control plane (exception path). System designers need to evaluate not only the speed, or frequency, of the processor solution, but also the headroom and intelligence on the chip to support features such as TCP checksum and deep packet examination and manipulation.

Furthermore, since storage applications are especially sensitive to latency considerations, integration at the silicon level becomes key. In a multi-chip solution, there is added latency each time a task gets passed to another chip. Higher levels of on-chip integration can speed up overall system communication, thus reducing response time and improving overall processing performance.

---

## **FLEXIBILITY**

While new trends such as iSCSI and Fibre Channel IP (FCIP) are exciting, these specifications are still evolving and are not yet finalized. To support new protocols and emerging standards in a constantly evolving industry, storage equipment vendors require processor solutions that can be programmed and updated easily using widely available software architectures and tools. In this way, system vendors can perform software-based field upgrades to boost performance and deliver new features, thereby maximizing the time-in-market of their system solutions.

---

## **I/O AND MEMORY BOTTLENECKS**

With processor clock speeds reaching frequencies of 1 GHz and higher, system designers are facing a new problem: these high-performance processors need equally fast access to high-speed I/O and memory to sustain wire-speed performance. For example, the peripheral component interconnect (PCI) bus is widely used today as the primary I/O interface for chip-to-chip interconnect within a system. However, the PCI bus bandwidth becomes quickly saturated when very high data rates must be sustained, especially with added peripherals such as Gigabit Ethernet PCI cards. Processor solutions that can integrate high-speed network I/O interfaces on-chip become very attractive.

Storage processing is also bound by memory considerations. Not only is support for large external memory (DRAM) required, but there is also a need for high-speed access to memory in order to minimize latency and improve response time for data requests. Next-generation systems will need to address memory bandwidth to keep pace with advances in system I/O and overall processing performance.

In short, to support next-generation storage processing requirements, new storage solutions with very powerful processors and high I/O and memory bandwidth are needed to deliver ever-greater performance.

---

## **FIXED POWER AND SPACE CONSTRAINTS**

In the data center, there are severe space and power constraints. System architects developing next-generation systems must balance system size and power consumption with performance considerations. Processor solutions that optimize power and performance while maintaining a high level of integration are key success factors.

## Section 3: Multi-Processor for Network Storage Designs

A new generation of high-performance, low-power, highly integrated processors from Broadcom offer a promising alternative to address the needs of next-generation storage systems. These processors provide advantages over both general purpose processors and traditional network processors for storage system applications such as IP-based SAN gateways, SAN routers, SAN and NAS systems, and iSCSI cards. Broadcom’s SiByte™ family of processors delivers the industry’s highest performance MIPS-based™ multiprocessor solutions while achieving the highest level of integration and lowest power consumption. At a fraction of the size and power of alternative solutions, the SiByte processors enable superior control processing as well as fast path (data) processing for high throughput applications. The tightly integrated solution achieves new levels of performance and flexibility for storage system design while enabling development of more compact, lower power systems. As a result, the SiByte processor family can help to meet the conflicting challenges of today’s storage system design.

The first member of the family, the BCM1250, features two 64-bit central processing units (CPUs) scalable from 600 MHz to 1 GHz, or up to 128 gigabits per second (Gbps) bus bandwidth, a high-speed memory subsystem delivering 50 Gbps memory bandwidth, and up to 30 Gbps I/O bandwidth tightly integrated onto a 60-million transistor silicon chip. Our 256-bit internal bus, the ZBbus, provides an ultra-high speed link between major blocks of the processor (see Figure 1). For smaller systems, a family of single CPU derivative-core processors, the BCM1125 and BCM1125H, provides original equipment manufacturers (OEMs) with a powerful alternative.

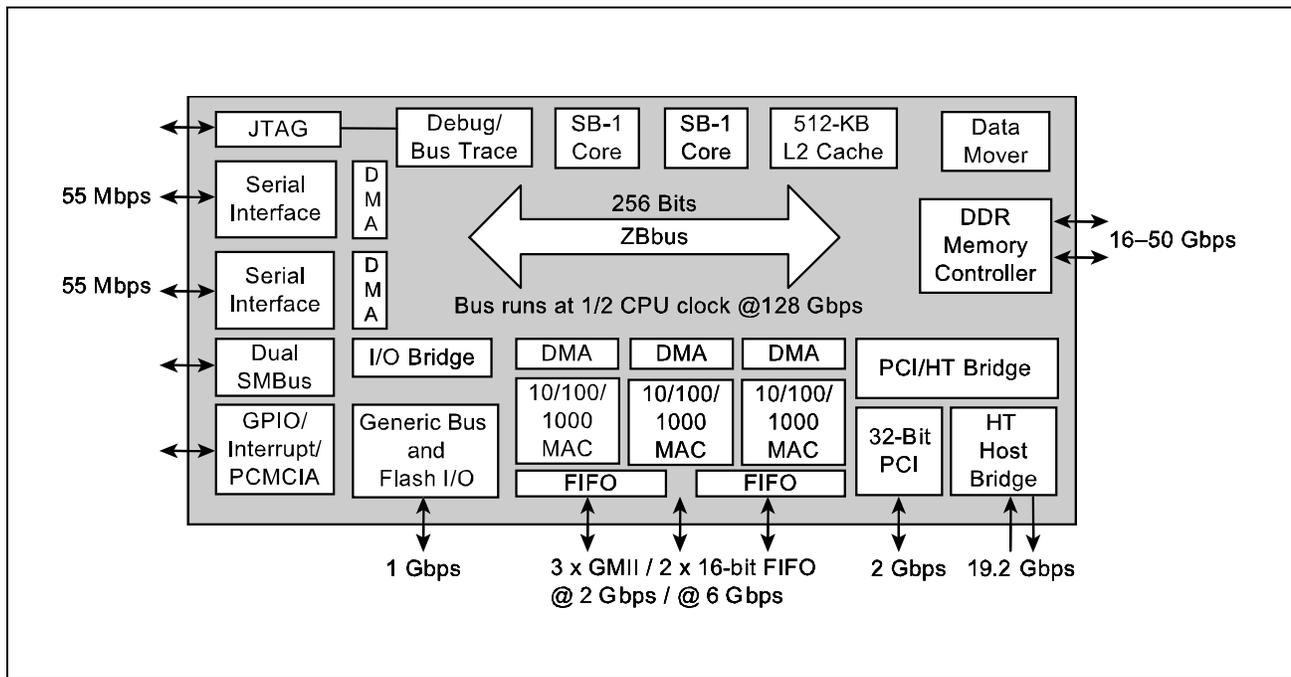


Figure 1: SiByte™ BCM1250 Block Diagram

The highly integrated multi-processor for networking and communications applications integrates two 64-bit MIPS® cores, scalable from 600 MHz to 1 GHz, with multiple I/O options to support high data throughput, including three Gigabit Ethernet Media Access Controllers (MACs) and a HyperTransport™ interface.



Flexible enough to handle both wirespeed in-line processing as well as exception path processing requirements, the BCM1125 and BCM1250 are well suited for SAN, NAS, and iSCSI storage applications. The flexible architecture combined with a MIPS-based processor, supported by widely available software tools, enables system designers to optimize overall system performance, power, programmability, and cost. Figure 2 shows a few of the network storage/server appliances in which the SiByte processors can achieve new levels of performance while helping system designers meet power, space, and cost budgets.

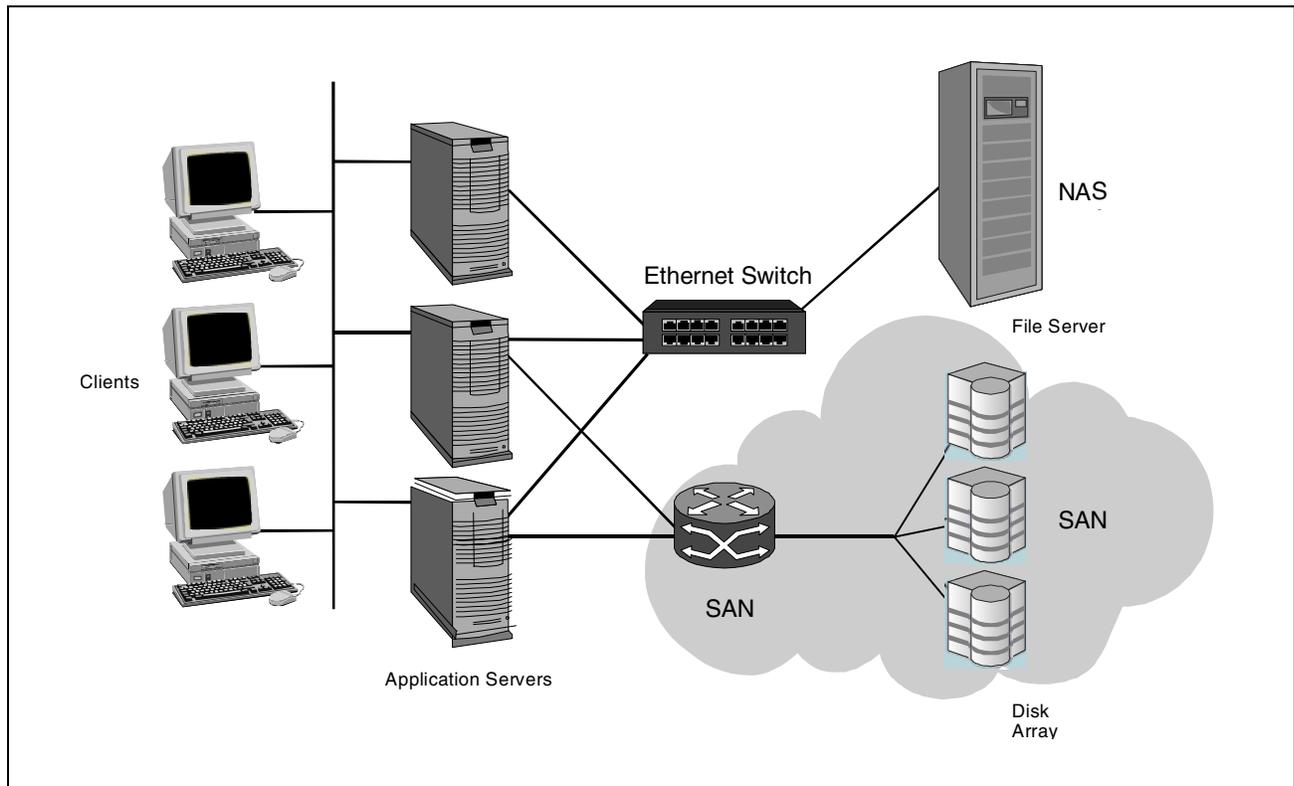


Figure 2: Network Storage and Server Appliances

## UNPARALLELED PROCESSING PERFORMANCE

The BCM1250 tightly integrates two 64-bit MIPS CPU cores, each scalable from 600 MHz to 1 GHz. The SiByte processor cores are high performance implementations of the standard MIPS64 Instruction Set Architecture (ISA), and incorporate the MIPS-3D and MIPS-MDMX Application Specific Extensions. Each core supports a four-issue enhanced skew pipeline and can issue up to two memory and two ALU (Integer, Floating Point, MDMS, or MIPS-3D) computational instructions per cycle. To minimize software development efforts, a comprehensive software development kit based on MIPS ISA tools and software supports the SiByte processors and provides high programming flexibility. The SiByte cores include a 4-way associative 32 KB instruction cache, and a 4-way associative 32 KB data cache with two accesses per cycle. The chip also provides hardware acceleration for TCP/IP offloading, such as IP header size and TCP/IP checksum.

---

## ZBBUS

At the heart of the SiByte processor is a high-speed split-transaction multiprocessor bus that connects all major blocks of the processor. This 256-bit bus runs at half the CPU core frequency to provide ultra-fast data transfer at greater than 100 Gbps for a 1 GHz CPU speed. The bus implements the standard modified/exclusive/shared/invalid (MESI) protocol to ensure coherence between the two CPUs, the L2 cache, memory, and I/O agents.

---

## ADVANCED MEMORY ACCESS

To support the dual processor cores, the BCM1250 integrates a shared 512 KB L2 cache, and a Double Data Rate (DDR) memory controller that supports 16 to 50 Gbps peak memory bandwidth. This high memory bandwidth supports processing at Layer 7, which requires at least twice the packet memory bandwidth of Layers 3 and 4. The DRAM interface supports up to 2 GB of memory using 512 MB synchronous DRAM (SDRAM) dual in-line memory modules (DIMMs) and allows expansion up to 8 GB. The large shared cache ensures fast memory accesses with minimal latency.

---

## INTEGRATED I/O

The architecture of the SiByte processors is optimized for maximum I/O throughput, with three Ethernet MACs, a 32-bit/66 MHz PCI host bridge, HyperTransport Host Bridge, SMBus, GPIO, flash interface, interrupts, timers and DMA all integrated onto the single 60-million transistor device. These integrated I/O functions eliminate the need for a separate system controller, and provide flexibility in design, saving space and power. Three Gigabit Ethernet MACs provides auto-sensing 10/100/1000 BASE-T Ethernet connections for large I/O bandwidth or redundancy. In cases where Ethernet protocol is not required, the MACs can also be configured as three 8-bit or two 16-bit Packet FIFO interfaces. Two serial ports are available to use as UARTs for console ports or as asynchronous interfaces. The high speed I/O needed to support an advanced storage system is available through an interface for HyperTransport I/O fabric and a 32-bit PCI (rev 2.2) local bus.

---

## HYPERTRANSPORT FOR FASTER CHIP-TO-CHIP DATA TRANSFER

In the SiByte processor family, the HyperTransport Host Bus provides a high-speed interface for connecting co-processors, such as encryption engines, as well as peripherals or multiple SiByte processors. HyperTransport is a new high-speed, packet-based I/O bus with a peak data transfer rate of 8 Gbps providing an innovative chip-to-chip system link that can reduce data bottlenecks in computers, networking equipment, and communications devices. HyperTransport is supported by the industry group HyperTransport Consortium which has a growing list of members including Broadcom, Cisco, Nvidia, and Sun. HyperTransport provides better than an order of magnitude increase in bus transaction throughput over existing bus architectures such as PCI, PCI extended (PCI-X) and accelerated graphics port (AGP).

The HyperTransport interface logically looks like PCI and uses a PCI configuration mechanism. Broadcom's 8-bit implementation in the BCM1250 provides 9.6 Gbps bi-directional throughput and serves as a host bridge that can support up to 31 devices, such as PCI/PCI-X bridges, HyperTransport switches, south bridges, MACs, or graphics. From a system design standpoint, the HyperTransport interface provides for the same type of buses as the PCI bus and uses the same ordering rules as PCI. It also maintains backward compatibility with existing software developed for the PCI, including the ability to support memory read/write operations.

---

## ON-CHIP DEBUG FEATURES

On-chip debug, trace, and performance monitoring functions assist both hardware and software designers in debugging and tuning the system. Supported functions include reset and configuration, debug, bus trace and performance monitor, timers, and interrupts. An on-chip JTAG interface enables an external debugger to access the system control and debug functions to facilitate debugging and bringing up the system without microcode.

## Section 4: Typical Storage Applications for SiByte™

The application summaries below illustrate how the SiByte family of processors can be implemented in various applications to address the need for high speed, high capacity storage in data centers, including NAS, SAN, and IP Storage.

### SAN SWITCH

The BCM1250 is well suited as the control processor for a high-performance SAN switch. A typical SAN Switch implementation is shown in Figure 3, in which two SiByte processors are linked via a HyperTransport Bridge. The SAN switch links to an Ethernet network via the gigabit media-independent interfaces (GMII) and uses PCI as an interface to the crucial Fibre Channel connections. A Fibre Channel Host Bus Adapter (HBA) can be connected to the BCM1250 via a HyperTransport-to-PCI bridge. Figure 3 also shows the scalability of this approach. For high-performance SAN boards, the BCM1250 can be doubled-up to offer four processors, using HyperTransport to support tunneling (devices can be “daisy-chained” in-line.) One high-end scalable multi-protocol SAN switch under development will feature up to 500 ports and support 2 Gbps and 10 Gbps Fibre Channel connections.

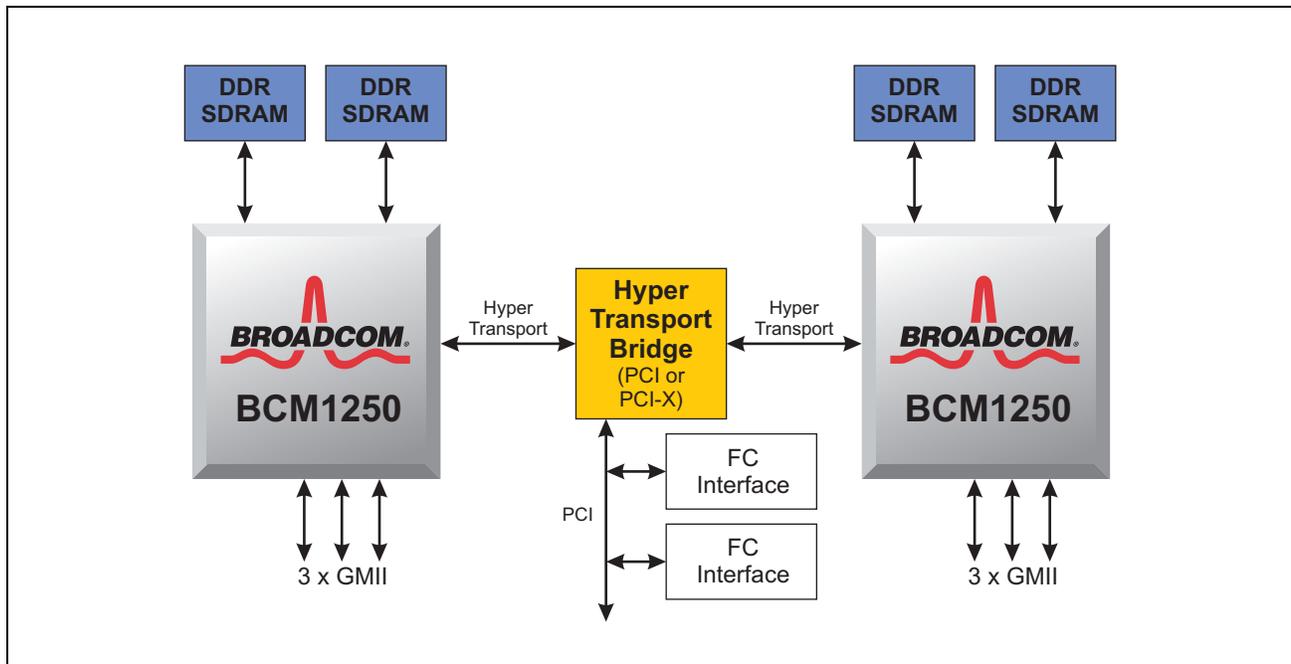


Figure 3: Typical SAN Switch Implementation

---

## ISCSI GATEWAY

Another example of an advanced storage system based on the SiByte processors is in an iSCSI Gateway. This provides iSCSI hosts and FC networked hosts with secure and trusted access to a variety of logical volumes residing on diverse storage systems within a SAN.

A multi-function high-end switch needs a processing platform that can handle multiple tasks beyond standard enterprise SAN requirements. The 1 GHz clock speed, high-speed I/O, and fast memory access of the SiByte processor can run a single execution software stack that performs all tasks without memory re-writes, eliminating the performance degradation memory re-writes can cause. In addition, this dual-processor platform provides the power to support iSCSI to FC and SCSI protocol conversion and block level virtualization. Each of these tasks is complex, and combining the two requires a combination of efficient code and high performance hardware. A single transaction involves accepting the incoming iSCSI packet, checking security, un-encoding it, un-encrypting it, checking data integrity, looking at the original command and address, making necessary changes, re-encoding, and re-encrypting it before sending the packet on its way. To do this on the fly at wire speeds TCP data integrity checksums and security checks can be offloaded onto the BCM1250 processor, thus reducing the burden of processing these tasks in software, making the code lighter and improving performance. The dual-processors also support real-time OS, which allows multiple tasks to run independently on the two processors while keeping the tasks perfectly synchronized.

---

## INFINIBAND<sup>®</sup> LINE CARD

Another advanced network storage solution that incorporates the SiByte processor is InfiniCon's InfiniO™ Shared I/O System and Clustering System, a family of InfiniBand<sup>®</sup> (IB)-enabled offerings providing seamless access to high-speed server-to-server communications and transparent integration of InfiniBand-enabled servers into Fibre Channel and/or Ethernet networks. Through this shared approach to I/O, InfiniO makes data center infrastructures less costly and complex to manage than traditional server deployment methods. By simplifying the process of adding additional capacity to new or expanding applications, data centers can respond to changing business requirements at a significantly accelerated pace.

Key to InfiniCon's value proposition is that no changes are required at the software layers (either application or O/S software), which enables non-disruptive deployment. As a result, all protocols, including those between FC, Ethernet and IB conversions, must be handled on the processor packet. Broadcom's BCM1250 was selected because its powerful processing and on-chip memory architecture prevent the InfiniO from being the bottleneck and reducing latency. Given InfiniCon's space and port density requirements, the highly integrated BCM1250 was a natural choice. Last, Broadcom's early availability of the BCM91250A evaluation boards allowed parallel efforts on both PCB and software development.

InfiniCon's shared I/O and clustering system removes the cost and complexity of traditional dedicated I/O subsystems found on today's server, and places I/O system resources into an external, shared unit that many multiple servers can use simultaneously. This means that the PCI busses, and all of the associated Gigabit Ethernet NICs and Fibre Channel HBAs that formerly were required to be purchased and installed in every server (and by extension, forcing every server to be installed/configured/managed on each of those networks) are no longer necessary. These servers can be attached with a high-speed, low-latency IB connection to the InfiniO and still realize greater amounts of connectivity, performance, availability, and function than they had before. For technical product information of InfiniO, visit InfiniCon's website, [www.infinicon.com](http://www.infinicon.com).

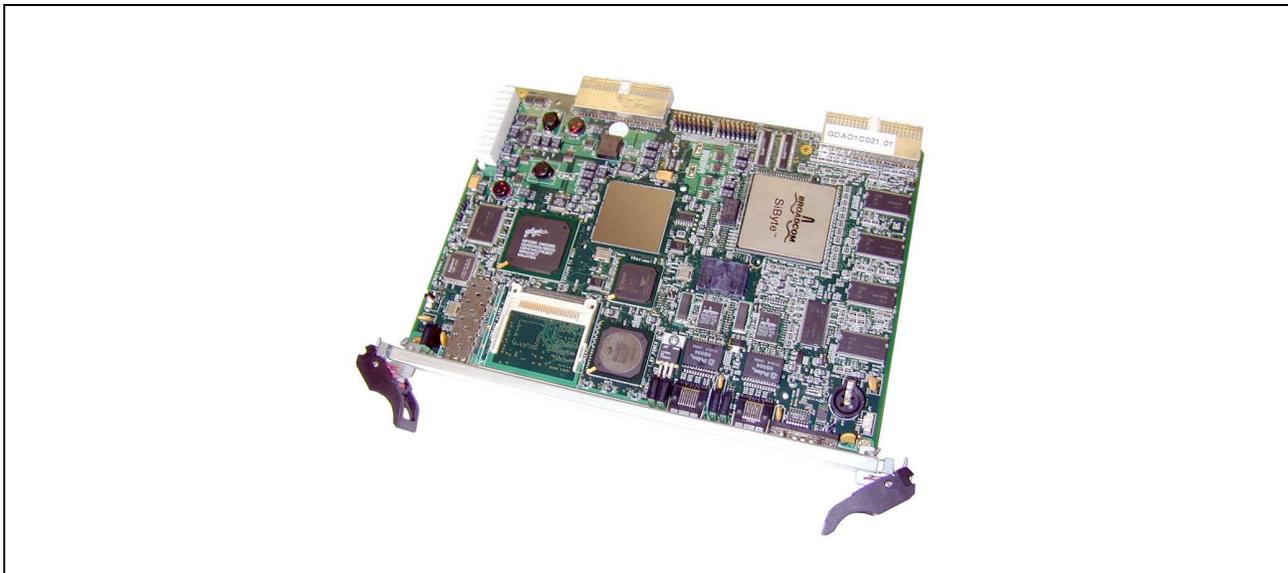
---

## IP STORAGE AND NAS-SAN BRIDGE

To address next-generation storage applications in IP Storage and NAS-SAN bridge products, Artis Microsystems Inc. has developed a single-board computer that combines the power of two 64-bit MIPS processors with the connectivity of Gigabit Ethernet and Fibre Channel interfaces. Artis Microsystems' CPCI-A7200 CompactPCI single-board computer is targeted for SAN, NAS, IP Storage, firewall, security, and VPN applications.

One of the challenges in IP storage applications is in handling the enormous overhead of processing iSCSI or iFCP protocols over TCP. The objective is to achieve low-latency without impacting performance, despite these compute-intensive protocol conversations. Artis selected Broadcom's BCM1250 dual processor for the CPCI-A7200 because it provides both the processing power and high-performance I/O required to handle these tasks with ease. Another reason for the selection of the BCM1250 is the level of integration in the system-on-chip (SoC) device. This reduced a number of the external components and allowed Artis to integrate the enormous computing power with two Gigabit Ethernet ports and one Gigabit Fibre Channel port in a 160 mm x 233 mm CompactPCI card.

The Artis CPCI-A7200 is well suited for storage applications, with scalable processing power of 600 MHz to 1 GHz in each of the two processors, combined with a Fibre Channel interface and two Gigabit Ethernet interfaces. The card is intended to extend the connectivity of corporate FC-SAN networks in multiple geographical locations connected through Ethernet/IP networks. Specifically, using protocols such as iFCP, FCIP and iSCSI, it can be used for bridging a Fibre Channel-based SAN to Ethernet/TCP/IP based NAS system. The Artis CPCI-A7200, shown in Figure 4, provides 1 Mb Content Addressable Memory; a fast lookup table device supporting ternary elements to facilitate the creation of advanced networking and communication applications. For technical product information of the CPCI-A7200, visit Artis Microsystems' website, [www.artismicro.com](http://www.artismicro.com).



**Figure 4: Artis Microsystems' CPCI-A7200 CompactPCI Single-Board**

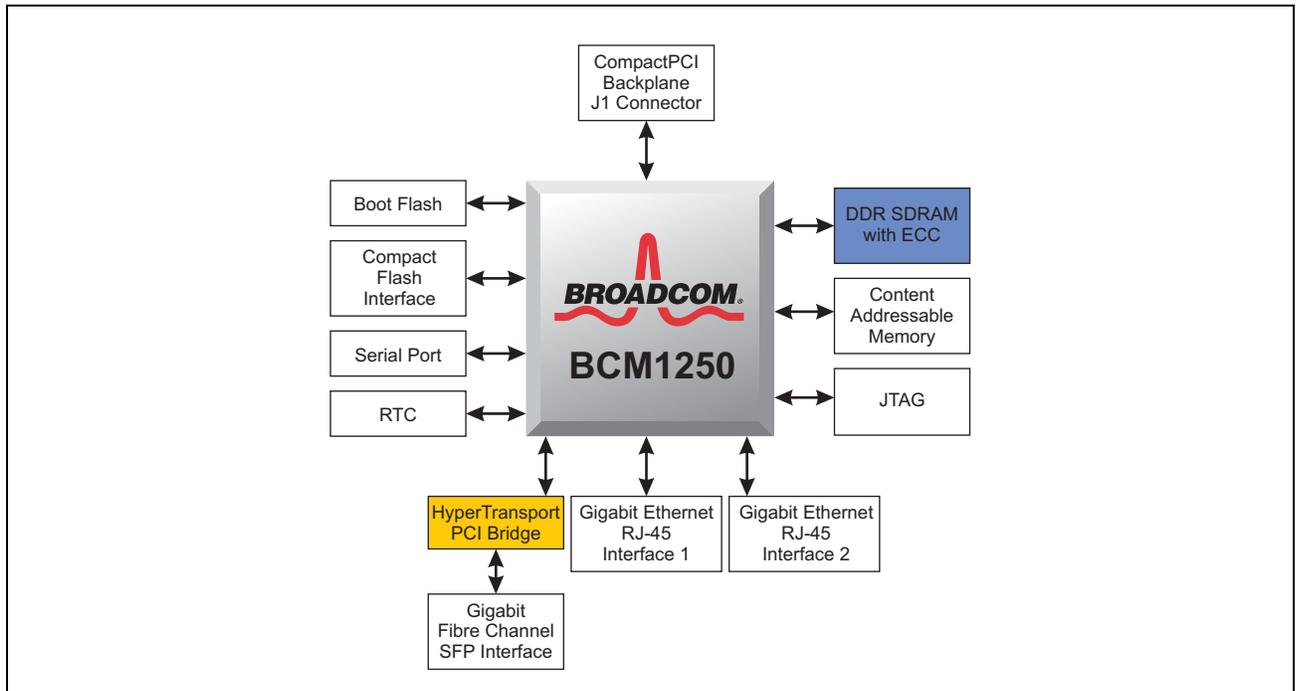


Figure 5: Artis Microsystems' CPCI-A7200 Block Diagram

## Section 5: Summary

---

### SiBYTE PROCESSOR FAMILY

In summary, the single and dual processor core members of Broadcom's SiByte family of processors offer a versatile platform that provide the power and flexibility to help storage system designers balance their requirements for performance, intelligent functionality, I/O throughput, space, and power. Planned additions to the family will provide even greater scalability and design flexibility to meet the ever-increasing demands of the industry.

---

### *Broadcom Corporation*

16215 Alton Parkway  
P.O. Box 57013  
Irvine, California 92619-7013  
Phone: 949-450-8700  
Fax: 949-450-8710

Broadcom<sup>®</sup> Corporation reserves the right to make changes without further notice to any products or data herein to improve reliability, function, or design. Information furnished by Broadcom Corporation is believed to be accurate and reliable. However, Broadcom Corporation does not assume any liability arising out of the application or use of this information, nor the application or use of any product or circuit described herein, neither does it convey any license under its patent rights nor the rights of others.