The Dolphin SCI Interconnect

White Paper

February 1996

Dolphin

INTERCONNECT SOLUTIONS

Dolphin Interconnect Solutions, AS
Olaf Helsets Vei 6
0621 Oslo, Norway
E-mail: info@dolphinics.no
Phone: +47 22 62 70 00
Fax: +47 22 62 71 80

Dolphin Interconnect Solutions, Inc
3625 East Thousand Oaks Blvd
Suite 50
Westlake Village, CA 91362
E-mail: info@dolphinics.no
Phone: (805) 371-9493
Fax: (805) 371-9785

Dolphin Interconnect Solutions, Inc
959 Concord Street
Framingham, MA 01701-4682
E-mail: info@dolphinics.no
Phone: (508) 875-3030
Fax: (508) 875-1517

1. Multiprocessor Architecture Overview

The economics of the microprocessor and volume manufactured PC units are changing the rules for how multiprocessor computer systems will be built in the future. The signs are already here driven by market demands such as:

- *Scalability* - The ability of a system to grow with increasing customer needs for higher performance.
- *Low cost* - Customers demand low entry level costs and want to expand their multiprocessor systems using inexpensive building blocks.
- *Reliability and availability* - Reduced computer down time and continuous availability of mission-critical data are becoming increasingly important.

In order to meet these market demands, computer makers need to employ systems in different architectures depending on scalability, cost and reliability/availability requirements. Independent of which architecture is used, existing applications must run unmodified and an operating system must be readily available.

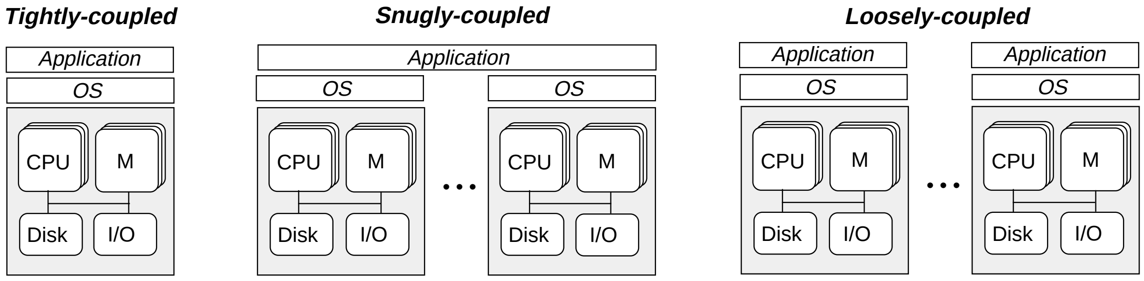

Figure 1—Three different multiprocessor architectures

Figure 1 illustrates three multiprocessor systems architectures. A **tightly-coupled** system is characterized by multiple processors sharing memory and I/O across a system backplane bus. A tightly-coupled system is the basis for *shared memory* symmetric multiprocessor (SMP) servers. In an SMP server, the operating system and the application can execute on any CPU and the server is under control of one operating system. SMP is the dominant architecture for commercial multiprocessor servers. An SMP system is widely recognized for its ease of application programming and manageability, and it is well suited for on-line transaction processing (OLTP), decision support systems (DSS) and general server applications. Scalability of an SMP system is constrained by the capacity of the backplane bus that connects processors and memory. As the system is expanded by adding more processors, the backplane bus becomes a bottleneck. Another scalability limit for SMP is that processors continue to get faster with improved chip technology, but the backplane performance is limited by physics (i.e. the speed of light). The result is that the number of processors you can lash together on a single bus gets smaller over time. In addition, the performance of a large multiprocessor system depends to a reasonable degree on how well the operating system scales. Another drawback of an SMP system is that it does not tolerate failures of processors within the system, a major problem for applications that require a high degree of data availability.

The counter architecture to a tightly-coupled system is a loosely-coupled system. A **loosely-coupled** system is characterized by each node having its own local copy of the operating system. Each node is typically an SMP system, thus a loosely-coupled configuration is also referred to as a cluster of SMP systems. Applications running on each node communicate with each other using explicit *message passing* protocols. Loosely-coupled systems can be configured as shared disk or shared nothing. A *shared disk* configuration is used in systems where applications need high bandwidth and low latency access to stored data, typically in a database server. A *shared nothing* configuration uses message passing protocols for both data transfers and process to process communication and has been a favored architecture for massively parallel (MPP) systems manufacturers. Although a shared nothing system has great scalability for some applications, it is known to be difficult to program and it is difficult to achieve performance scalability on a wide set of applications. One of the great benefits of a clustered SMP system is that it overcomes the single point of failure problem of a single SMP system. With the appropriate cluster-aware application, a node within an SMP cluster can fail, yet the application remains operational. These applications and SMP clusters are used in high availability database server systems where reducing the amount of system downtime is very important.

A **snugly-coupled** configuration combines the benefits of shared memory with the reliability benefits of clustered SMP systems. In a snugly-coupled configuration, each node runs a copy of the operating system whereas the application is distributed across the nodes. Since each node has a copy of the OS, node failures do not halt the system and management software can cope with the application interruption. At the same time, the application continues to see shared memory, thus applications do not need to be re-written to run on this configuration. This configuration is also called shared memory cluster (SMC). An additional benefit of an SMC system is that it can use standard operating systems with no additional scaling requirements beyond what is needed within a node. Table 1 lists key differences between these architectures.

Table 1—Key difference between multiprocessor architectures

	Tightly-coupled	**Snugly-coupled**	**Loosely-coupled**
Programming Model	Shared Memory	Shared Memory	Message Passing
Multiprocessor Architecture	Shared Memory	Shared Memory	Shared Disk Shared Nothing
Pros	Easy to program Easy to manage	Easy to program Improved availability	Good scalability High availability
Cons	Limited scalability Single point of failure	Harder to manage Application must be tuned	Difficult to program Limited set of applications
Commonly used terminology	Symmetric multiprocessor system (SMP)	Shared memory cluster (SMC)	Cluster of SMPs
Dolphin SCI support	✓	✓	✓

Key to all of these multiprocessor architectures are at least three technologies. When building a multiprocessor system it is desirable to use an inexpensive *building block* that can be replicated many times. Such a building block can be a microprocessor or a motherboard with many microprocessors. One example of an attractive building block is the Intel Pentium Pro motherboard that contains up to 4 processors, memory and I/O. In addition, a *scalable operating system and/or application* is needed so that the added number of processors translates into improved system performance. The third key technology is a *scalable interconnect* that enables combination of the building blocks into a cost-effective and powerful system.

Dolphin provides interconnect products that enable integration of systems using any of the above mentioned architectures. The **Dolphin SCI** product line has been designed to maximize flexibility by supporting tightly-coupled, snugly-coupled and loosely-coupled architectures. This product line consists of *chip sets*, plug-in *adapter cards*, and *fabric* consisting of switches and copper/fiber-optic cables. In addition, driver and cluster management *software* help system integrators to integrate a system quickly.

2. Introducing Dolphin SCI

2.1 SCI short facts

Dolphin is basing its product line on the newly adopted ANSI/IEEE 1596-1992 Scalable Coherent Interface (SCI) standard.

- The SCI standard defines a point to point interface and a set of packet protocols. The SCI protocols use small packets with a 16-byte header and data sizes of 16, 64 and 256 bytes. Each packet is protected by a 16-bit CRC code.

- The standard defines 1 Gigabit/sec serial fiber-optic links and 1 GByte/sec parallel copper links. An SCI interface has two unidirectional links that operates concurrently.

- The SCI protocols support shared memory by encapsulating bus requests and responses into SCI request and response packets. Packet-based handshake protocols guarantee reliable data delivery. A set of cache coherence protocols are defined to maintain cache coherence in a shared memory system.

- Message passing is supported by a compatible subset of the SCI protocols. This protocol subset does not invoke SCI cache coherence protocols.

- SCI use 64 bits addressing and the most significant 16 bits are used for addressing up to 64K nodes.

2.2 Dolphin SCI features and benefits

The Dolphin SCI product line has been designed to support different multiprocessor system architectures and applications. In order to achieve this flexibility, the products support the two main programming models: shared memory and message passing.

2.2.1 Shared memory

- *Shared memory* - Dolphin SCI products support shared memory protocols so that all processors in a system see one global shared memory. In most systems the memory is distributed on the nodes within the system and SCI links these nodes into a *global distributed shared memory* architecture.

- *Cache coherence* - All high performance microprocessors use caches to reduce the average memory access time and improve overall processor performance. In a multiprocessor system, multiple processors with caches lead to a cache consistency problem. Buses have solved this problem using cache coherence protocols such as bus snooping to ensure that data located in the caches do not become stale. SCI interfaces overcome the scalability problem of bus snooping protocols and use distributed *directory-based cache coherence* protocols to maintain cache coherence.

- *Atomic operations* - Processes and process threads in a multiprocessor system need to synchronize access to shared resources. Buses use bus locks which enable an atomic read-modify-write operation to implement locks such as test&set. SCI interfaces support a rich set of *lock primitives* including test&set, compare&swap and fetch&add.

2.2.2 Message passing

- *Direct memory access (DMA)* - For high throughput and low overhead data transfers, an SCI interface uses a DMA controller. The DMA controller can transfer data directly from user memory of one node into the user memory of another node, with no need for intermediate buffering.
- *Protected shared memory* - The performance limitation of parallel applications is often linked to the latency of communication for small messages. Most networking interfaces have been optimized for high throughput at the expense of low latency. SCI takes a different approach by allowing small data transfers (~500 bytes) to be sent between processes with exceptionally low latency. A message passing library can use the underlying shared memory protocols which enable *load/store operations* to reliably send data. The shared memory interface can also be used directly by an application programmer. Shared memory segments have access protection to prevent message passing libraries or user programs to erroneously access illegal memory addresses. When an SCI interface is used in message passing mode, the latency is not limited by the overhead of system calls but by the hardware latency of routing a processor store through the adapter card to memory of another node (~1-2µs).
- *Atomic operations* - The SCI interfaces enable software to generate a read-modify-write operation on remote memory through SCI lock transactions.
- *Interrupts* - Interrupts are generated on remote nodes through on-board registers or a conditional atomic operation.

2.2.3 Higher bandwidth

Dolphin SCI products use single-chip SCI interfaces for node to node communication. An SCI interface has two unidirectional parallel links that connect nodes in various topologies. Because all signals are unidirectional, the performance of a link is not limited by reverse handshake delay, but is limited by speed of light and advances in semiconductor technology.

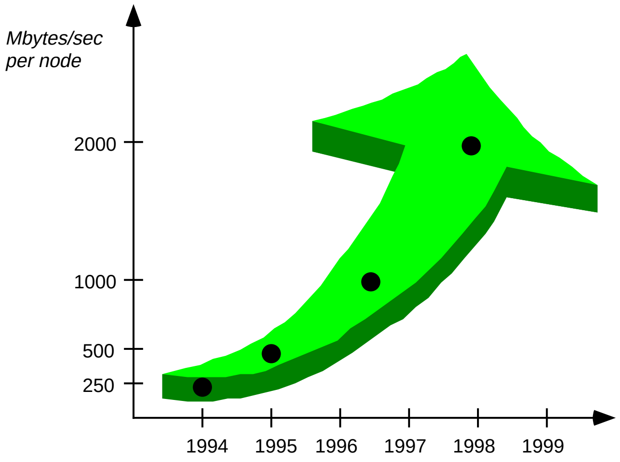

Figure 2—Duplex bandwidth per node (CMOS)

Figure 2 illustrates the current performance and the expected performance of SCI link interface chips. Each SCI interface has one receive and transmit link that operates concurrently, thus a 500 Mbytes/sec link bandwidth means a total receive/transmit bandwidth of 1000 Mbytes/sec per node. Total aggregate bandwidth in a system depends on the connection topology (ring or switch) but could be as high as the sum of all node interface bandwidths.

2.2.4 Lower latency

Dolphin SCI products have been designed to link nodes into a dedicated multiprocessor system. Multiprocessor applications require low latency and high througput in order to scale well and to deliver high performance.

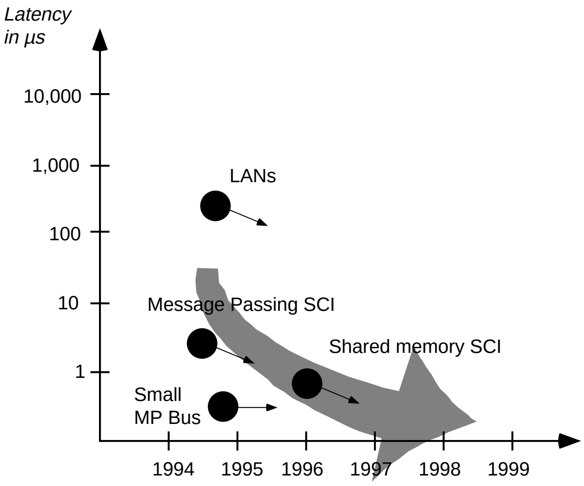

Figure 3—User level base latency comparisons

Figure 3 illustrates the latency of communication for SCI, a small MP bus and local area networks. The SCI protocols have been designed using "RISC" principles using few transaction types and packet sizes so that link chips and bridge chips can process accesses with low latency. In addition, the SCI chips use pipelining to achieve high throughput operation.

- *Shared memory latency* - SCI used as a backplane bus replacement for an SMP server has latencies which are comparable to those of a multiprocessor backplane bus. In a shared bus architecture, all accesses to the bus are serialized causing long latencies when there is contention on the bus. The point to point architecture of SCI allows accesses to be overlapped reducing the time a processor must wait for data to be returned from memory.

- *Message passing latency* - The key to achieving low message passing latency is to enable user-level applications to bypass time consuming system calls to transmit data from one node to the other. The SCI interfaces enable reliable data transfers between memories with little or no software overhead resulting in dramatically lower latency. LAN protocols have been designed to connect a very large number of nodes and assume unreliable hardware protocols, thus the overhead of making a reliable data transfer is much higher. Performance benchmarks have demonstrated that Dolphin SCI products have message passing latencies 50-100 times lower than LANs for small messages.

2.2.5 Improved scalability

Current multiprocessor backplane buses can connect up to a few tens of cards. However, higher processor frequencies are forcing manufacturers to reduce the number of bus slots to improve bus performance and to deliver more bandwidth to each processor. Thus as processor performance increases, the bus becomes less and less scalable as illustrated in Figure 4 .

In a bus architecture, all processors share the total available bandwidth. When a processor is added to the bus, the bandwidth available to each processor is reduced. The scalability of a bus is therefore limited by the bandwidth and latency of a shared bus architecture. When a processor is added using SCI, each processor node adds bandwidth to the system, thus the bandwidth available to each processor is constant. In addition, SCI supports concurrent accesses which reduce latency in a loaded system and improve overall system scalability. As a result, Dolphin SCI products will enable systems that can scale from a few processors to several hundred processors in shared memory or shared nothing architectures.

The size of a commercial multiprocessor system will be limited by the scalability of applications and operating systems. However, Dolphin SCI products make a quantum leap in scalability compared to backplane bus technology.

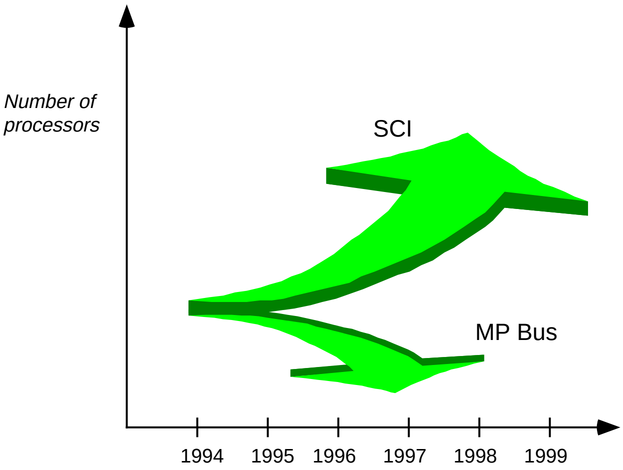

Figure 4—SCI overcomes bus scalability limitations

2.2.6 Low cost

- *Singe-chip interface* - Dolphin has implemented SCI interfaces in single-chip link solutions that can be manufactured in high volume. The link interfaces do not need external transceivers to connect to cables nor are external buffer chips needed. Thus improved reliability and reduced cost is achieved.
- *High volume* - Dolphin's link interface chips are designed using CMOS technology and can be produced in high volume.
- *Standard packaging* -Dolphin's custom link designs have been optimized for high volume manufacturing using a small die size and industry-standard packaging.
- *Low power* - CMOS implementations of link interface chips and controller ASICs mean low power consumption and no special cooling requirements.
- *Standard cables* - Adapter cards use standard copper or fiber-optic cables.

2.2.7 Reliability, Availability, Serviceability

- *Reliability* - The SCI protocols have been designed to guarantee data delivery through a set of handshake packet protocols. However, hardware failures, such as unplugged or broken cables and noise, do occur. SCI interfaces have built-in error detection and error logging functions so that software can determine where an error occurred and what type of error it was. In addition, the interface hardware supports error containments so that failing nodes do not cause failures in operating nodes.
- *Availability* - SCI interfaces support redundant links and switches for systems that require high availability operation. Multiple adapter cards can be used in each node to achieve high performance and failover capabilities. Driver software supports multiple cards per node and various mechanisms for detecting failing nodes.

- *Serviceability* - SCI interfaces have been designed as plug-in modules that can be replaced in the field. The interfaces also support on-line replacement of nodes or plug-in cards.

3. Scalable Pentium Pro SMP Servers

3.1 Today's SMP Servers

- *PC servers* - PC servers typically have up to 4 Pentium processors connected on a multiprocessor bus using a multiprocessor PC chip set. A PC server has low cost and runs shrinkwrapped operating systems, but does not scale beyond 4 processors due to the scalability limitation of the multiprocessor bus and the operating system.
- *Enterprise servers* - Enterprise servers have up to 32 processors connected on a large high performance backplane bus running a scalable proprietary version of the Unix operating system. Enterprise servers scale better than PC servers, but have much higher cost.

3.2 Improved and cost-effective scalability

Dolphin SCI products have been designed to improve the scalability of PC servers and to reduce the cost of enterprise servers.

- Dolphin SCI improves scalability allowing systems to scale from 4 to 64 processors or more.
- Dolphin SCI lowers entry-cost and enables users to "pay-as-you-go".
- Dolphin SCI preserves investments in applications.
- Dolphin SCI uses high-volume Pentium Pro building blocks.

3.3 The Pentium Pro SHVS building block

With the introduction of the Pentium Pro microprocessor, Intel has defined an architecture for multiprocessor servers known as standard high-volume servers (SHVS). An SHV server is a system built on the same production line as a desktop computer but including the more sophisticated features required by a server implementation. An SHV server consists of up to 4 Pentium Pro processors, 1-4 Gbyte of memory, 1-2 PCI buses and disks as illustrated in Figure 5 .

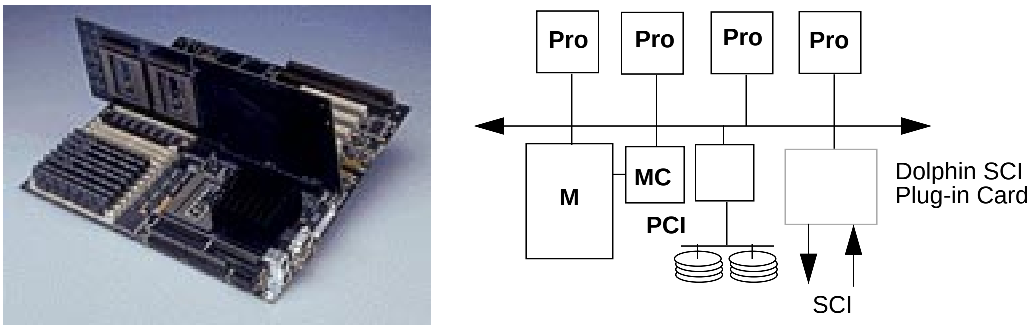

Figure 5—Intel Pentium Pro SHVS and Dolphin SCI interface

Intel has defined a cluster port interface on the Pentium Pro multiprocessor bus which can be used to connect SHVS nodes into a high performance cluster. Dolphin’s plug-in adapter card is compatible with the cluster port and the Pentium Pro bus protocols.

3.4 Leveraging SHVS in a larger system

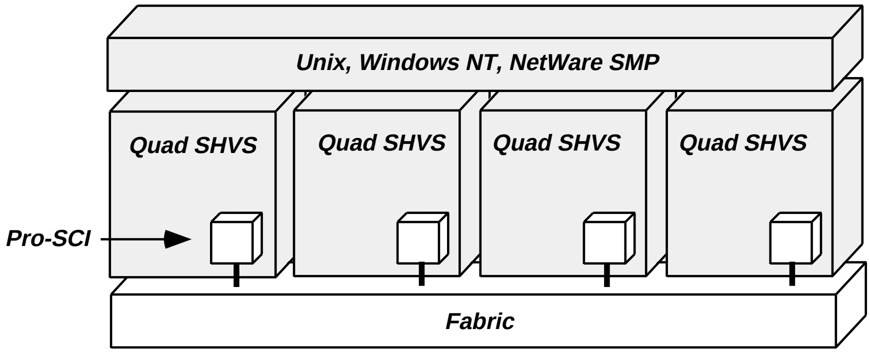

Figure 6—A 16-processor SMP server using Intel SHVS and Dolphin Pro-SCI

Figure 6 illustrates a tightly-coupled SMP server where multiple quad (4 processors) SHVS nodes are connected using Dolphin SCI. The Pro-SCI adapter card makes it appear to applications and the operating system as if they are running on a single SHVS node, when in fact they are running across multiple nodes. The speed and protocols of SCI make a virtual bus, linking the Pentium Pro buses together as if they were one large bus. Processor requests are mapped into small SCI packets which are processed at very high speed through the interconnect fabric. All nodes can communicate simultaneously and at any time, reducing latency and maintaining high performance.

3.4.1 Dolphin Pro-SCI features

Transparent operation:

- An SCI memory controller gives each processor a global distributed shared memory view. A processor does not notice any differences whether accessing memory on the local SHVS or a different SHVS, except for the added latency. Pro-SCI supports up to 64 Gbytes of global shared memory.

- An SCI cache controller guarantees cache coherence between cached data on each SHV node. Accesses to memory on another SHVS node are cached by the Pro-SCI interface. The SCI cache controller maintains cache coherence between the Pro-SCI caches and the Pentium Pro processor caches.

High performance:

- A 1000 Mbytes/sec duplex link interface guarantees low latency accesses between SHVS memories (1-2 µs) and high bandwidth. SHVS nodes can be linked in a ring, dual ring or switch topology using the high speed point to point links.

Scalability:

- The Pro-SCI interface can address up to 12K SHVS nodes. Practical system sizes will be limited by application and operating system scalability. Initial systems are expected to be in the 8-64 processors range.

Compatibility:

- An Intel-defined connector on the SHVS motherboard allow SHVS nodes to be added to the system by plugging in a Pro-SCI card. The Pro-SCI adapter card is an EISA form factor card that conforms to this connector specification.

- The Pro-SCI interface is compliant with the ANSI/IEEE 1596-1992 Scalable Coherent Interface standard.

- The Pro-SCI interface is compliant with Intel’s multiprocessor specification (MPS) and multicomputer-specification (MCS).

Reliability:

- Pro-SCI has been designed to maintain the same level of reliability as a backplane bus. Memory and tags on the Pro-SCI adapter card are protected by error correction codes. SCI packets transmitted on cables are protected by a 16-bit CCITT CRC code. Links use differential signals and shielded twisted pair copper cables.

Availability:

- Pro-SCI supports a set of hardware functions to improve availability compared to a backplane bus. Operating systems and applications can use these functions to reduce the effect of software, hardware or transmission failures.

- The interface supports access protection on control and status registers as well as on memory address ranges which can be reached from other nodes. This prevents erroneous software or hardware failures from corrupting register contents and operating system or application data structures. Pro-SCI can also exclude specific nodes from accessing one node.

- Pro-SCI supports dual links for redundant ring configurations and switches.

4. Cluster of SMP servers

4.1 Today's SMP clusters

- *Database server clusters* - A cluster of SMP servers has become a popular platform for high availability databases. When multiple servers are connected in a cluster, failover software enables continuous operation even if one node within the cluster fails. Some databases can also operate in a parallel mode to improve system performance in addition to supporting failover modes. Most high availability database clusters consist of two nodes, but there is a need to be able to scale the system to many nodes. Current database clusters use Ethernet for cluster communication.

- *High performance compute clusters* - The performance of SMP systems, either as a workstation or a server, makes them attractive for computation-intensive applications. In order to solve large and time consuming problems, multiple SMP systems can be connected in a cluster to expand the total processing capability. This is an attractive alternative to dedicated massively parallel systems with respect to both performance and cost. Compute clusters can consist of many nodes, although most clusters consist of 4-16 nodes. Current compute server clusters use Ethernet or FDDI for cluster communication.

4.2 Improving performance and scalability

- Dolphin SCI has the best *combination of low latency and high bandwidth* to support both fine grained and coarse grained parallel applications.

- Dolphin SCI can support many nodes in a cluster to overcome the latency and bandwidth limitations of local area networks.

- Dolphin SCI supports fiber and copper links and allows nodes to be separated over long distances.

- Dolphin SCI supports redundant configurations for high availability requirements.

4.3 Parallel database cluster

Parallel database server clusters are becoming increasingly popular. A user can install one SMP server and run a database on that system. As the database grows and the number of clients increases, the user can install a second SMP system and a parallel database module. The dual node configuration with the parallel database software allows the database to deliver up to twice the capacity *and* high availability. The database software has been designed to detect failures in one node and switch over to the other node in case of failures with no loss of data availability for the clients.

Figure 7 illustrates a dual host parallel database (PDB) cluster. Each node runs an instance of the parallel database software and an operating system. A lock manager synchronizes accesses to the database. This example also shows a shared disk configuration where each host has direct access to the disk system. A typical database used in this configuration is Oracle Parallel Server (OPS). The performance of the lock manager which synchronizes database accesses is sensitive to the latency of communication between the servers. By using Dolphin SCI, latency and throughput can be significantly improved compared to LAN implementations.

Figure 7 shows a configuration where Dolphin's PCI-SCI adapter card is used to link the nodes within the cluster. In order to support the high availability requirements, two cards can be used per host so that a broken cable or adapter does not cause loss of data availability for the clients.

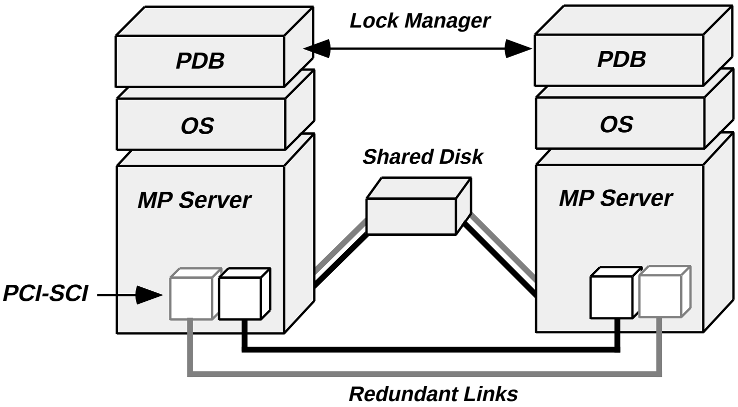

Figure 7—High availability database server with PCI-SCI

4.3.1 Dolphin PCI-SCI features

Message passing support:

- The PCI-SCI adapter card driver has an interface optimized for low latency and high throughput operation. The driver supports both memory to memory DMA and low latency programmed I/O.
- The adapter card and driver support atomic operations and remote interrupt mechanisms.
- The driver supports both a standard IP interface as well as a high performance light weight message passing protocol.

High performance:

- The PCI-SCI adapter uses a 400 MBytes/sec duplex link interface. The card allows small messages to be communicated between user processes with 4µs latency .

Scalability:

- The PCI-SCI adapter card can address up to 64K nodes. Initial systems will be in the 2-16 nodes range.
- The PCI-SCI adapter supports switch configurations including redundant switches.

Compatibility:

- The PCI-SCI card is compatible with PCI specification revision 2.1.
- The PCI-SCI card is compatible with the ANSI/IEEE 1596-1992 Scalable Coherent Interface (SCI) standard.

Reliability:

- SCI packets transmitted on cables are protected by a 16-bit CCITT CRC code. The PCI-SCI adapter supports parity on PCI bus. Links use differential signals and shielded twisted pair copper cables or standard fiber-optic cables.
- The PCI-SCI card supports guaranteed data delivery of requests and responses. Extensive error detection and error logging are available.

Availability:

- The interface supports access protection on control and status registers as well as on memory address ranges which can be reached from other nodes. This prevents erroneous software or hardware failures from corrupting register contents and operating system or application datastructures. PCI-SCI can also exclude specific nodes from accessing one node.
- In the case of hardware errors, the PCI-SCI card detects the error and can enable software to retry a transmission.
- PCI-SCI supports dual cards per host for redundant link configurations.

5. Shared Memory Cluster

5.1 Operating system approaches

- *Standard clusters* - Independence of systems provides flexibility and availability, but users don't get shared memory. Applications must use explicit message passing or shared disks to make use of multiple nodes.

- *Monolithic kernel* - Although this is the traditional solution, it requires a major system software effort, which only vendors of proprietary operating systems have undertaken to date. In the near future, commodity operating systems will not scale beyond a small number of processors.

- *Shared memory cluster* - Combines the availability benefits of standard clusters with the shared memory benefits of monolithic kernels. SMC uses add-in software to provide a single system image to applications even though it looks like a shared nothing cluster at the base operating system level.

5.2 SMC - improving scalability and availability

The limited scalability of an SMP backplane bus is solved by using Dolphin Pro-SCI adapter card. However, in order to take advantage of the larger system, the operating system must provide performance that scales with the number of processors. Traditionally, providers of enterprise servers have developed scalable monolithic Unix kernels that support proprietary server products. Since these operating systems are not readily available and are not viewed as industry standards, integrators of open servers find their systems limited in scalability by the UnixWare and Windows NT operating systems.

In order to overcome this problem, Dolphin is providing shared memory cluster (SMC) software modules that allow a system to scale using standard off-the-shelf versions of UnixWare and Windows NT.

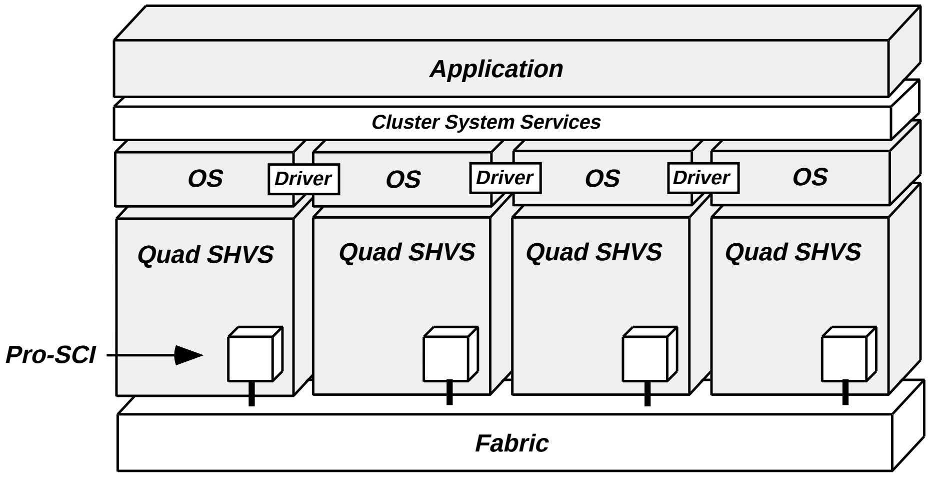

Figure 8—A 16 processors shared memory cluster

Figure 8 illustrates a shared memory cluster system. This is a snugly-coupled architecture where each node in the system runs a copy of the operating system, but where the application is distributed across the nodes. SMC software modules provide the ability for the application to treat the cluster as if it were a single large SMP:

- *Cluster System Services (CSS)*, a dynamically linked library that provides the illusion of a single system image to the applications.
- *Shared Memory Cluster Driver (SMC Driver)*, a loadable device driver that manages the SCI hardware and links together memories of the nodes into a large shared memory.

These SMC modules permit many multiprocessing applications, such as Oracle™, to take full advantage of the cluster without any modifications. Furthermore, this snugly-coupled architecture dramatically reduces the impact of a node failure compared to a conventional SMP that cannot tolerate a processor failure.

5.3 Dolphin SCI SMC features

Low cost:

- SMC systems use standard volume manufactured Pentium Pro microprocessors and SHVS systems as well as standard UnixWare and Windows NT operating systems.

High performance:

- Applications running on an SMC system use coherent shared memory for high performance.
- Low latency, low-overhead signalling and message-passing between nodes support rapid interprocess communication.

Higher availability:

- Independent operating systems make it feasible to add or remove nodes without rebooting the entire system.
- Failing hardware components will be detected, reported and automatically removed for the configuration.

Portability and compatibility:

- Applications will run as-is. Since the kernel is not changed and the hardware supports coherent shared memory, 100% compatibility is assured.
- The SMC software is portable so that the solution can be applied to a variety of platforms without high porting cost.

Improved application scalability:

- The base operating system only needs to scale to the number of processors in a node.
- Easy to add nodes for incremental scalability.

Ease of administration:

- Tools are provided so most administration operations are performed on the whole system.
- Global visibility of files makes database administration easier.
- Configuration is simple, with automatic reconfiguration omitting failed components.