Dolphin

INTERCONNECT SOLUTIONS

PCI Express High Speed Networking

A complete solution for demanding Network Applications

Table of Contents

Introduction

Maximizing application performance is a combination of processing, communication, and software. PCIe Networks combine all three elements. A Local Area Network (LAN), PCIe Networks connect Processors, I/O devices, FPGAs and GPUs into an intelligent network. They connect devices through flexible cabling or fixed backplanes. The main goal of PCIe networking is to eliminate system communication bottlenecks, allowing applications to reach their potential. To accomplish this, PCIe Networks deliver the lowest latency possible, combined with high data rates.

Dolphin's PCIe networking solution consists of standardized computer networking hardware and software. Our standard form factor boards and switches reduce time to market, enabling customers to rapidly develop PCIe networking solutions for data centers and embedded systems. Dolphin's software ensures reuse of existing applications but with better response times and data accessibility. Our SuperSocket™ and IPoPCIe software enables quick application deployment with improved overall performance and without application modification. Application tuning is available with our low level SISCI shared memory API that delivers maximum performance.

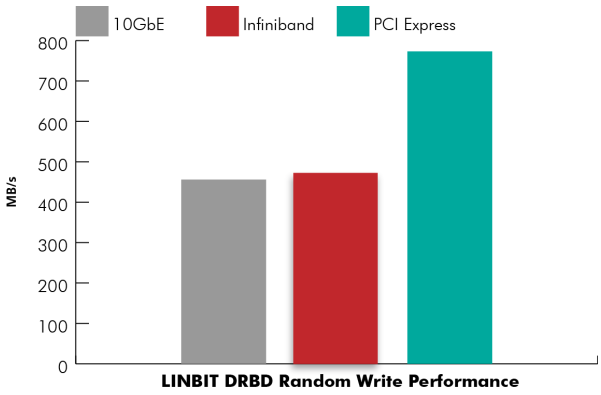
The PCI Express standard continues to improve performance, while maintaining a low cost infrastructure. The current PCI Express road map extends speeds to 128 Gbps and 256 Gbps while still maintaining backward compatibility. Standard PCIe commercial components are used by Dolphin as a road map to high performance hardware. Dolphin's solution exploits the PCI Express infrastructure to deliver next generation systems with maxim application performance. Our easy to implement and deploy solution gives customers the choice of changing or not changing their existing applications, but still taking advantage of PCI Express performance.

Why PCI Express ?

Performance

PCI Express solutions deliver outstanding performance compared to other interconnects in latency and throughput. When compared to standard 10 Gigabit Ethernet, PCI Express latency is 1/10 the measured latency. This lower latency is achieved without special tuning or complex optimization schemes.

In addition, Dolphin takes advantage of PCI Express high throughput. Our current solution utilizes Gen 2 and Gen 3 PCIe with 40 /64 Gbps data rates. Dolphin's software infrastructure allows customers to easily upgrade to next generation PCI Express with doubling bandwidth. No software changes are required. These products still maintain the low latency characteristic of PCI Express. The investment in low latency high performance Dolphin products yields dividends today and into the future.

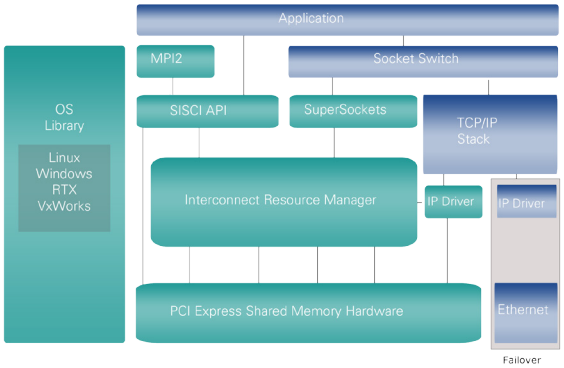


LINBIT DRBD Random Write Performance

Eliminate Software Bottlenecks

Dolphin's PCIe Software is aimed at performance critical applications. Advanced performance improving software, such as the SuperSockets API, remove traditional network bottlenecks. Sockets, IP, and custom applications utilize the low latency PIO and DMA operations within PCI Express to improve performance and reducing system overhead.

Dolphin's SuperSockets software delivers latencies around 1 µs and throughput at 54 Gbps. Other software components include an optimized TCP/IP driver for IP applications and the SISCI shared memory API. The SISCI API offers further optimization by using remote memory and multicast/ reflective memory operations. Customers benefit from even lower latencies in the range of 0.54 µs latency with higher throughput of over 6500 MBps.



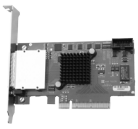
Key Applications

- Financial Trading Applications
- High Availability Systems
- Real Time Simulators
- Databases and Clustered Databases
- Network File Systems
- High Speed Storage
- Video Information Distribution
- Virtual Reality Systems
- Range and Telemetry Systems
- Medical Equipment
- Distributed Sensor-to-Processor Systems
- High Speed Video Systems
- Distributed Shared Memory Systems

Robust Features

- Lowest host to host latency and low jitter with 0.54µs for fast connections and data movement
- DMA capabilities to move large amounts of data between nodes with low system overhead and low latency. Application to application transfers exceeding 6500 MBps throughput.
- Management software to enable and disable connection and fail over to other connections
- Direct access to local and remote memory, hardware based uni- and multicast capabilities
- Set up and manage PCIe peer to peer device transfers
- High speed sockets and TCP/IP application support
- Ease installation and plug and play migration using standard network interfaces

High Performance Hardware

	Low profile PCI Express Gen 2 and Gen 3 adapter cards provide high data rates over standard cabling systems. These network interface cards are used in standard servers and PCs deployed in high performance low latency applications. Most cards incorporate standard iPass connectors. Yet Dolphin also offers a MiniSAS-HD cabled card. The host adapters feature transparent and non-transparent bridging (NTB) operation, along with clock isolation.
	XMC Adapters bring PCIe data rates and advanced connection features to embedded computers supporting standard XMC slots, VPX, VME or cPCI carrier boards. PCIe adapters expand the capabilities of embedded systems by enabling very low latency, high throughput cabled expansion. Standard PCs can easily be connected to embedded systems.
	PCI Express Gen 3 switch boxes scale out PCIe Networks. Both transparent and non-transparent devices link to a PCIe switch, increasing both I/O and processing capacity. These low latency switches scale systems while maintaining high throughput.

PCIe Software

Reflective Memory / Multicast

Dolphin's reflective memory or multicast solution reinterprets traditional reflective memory offerings. Traditional Reflective Memory solutions, which have been on the market for many years, implement a slow ring based topology. Dolphin's reflective memory solution uses a modern high speed switched architecture that delivers lower latency and higher throughput.

Dolphin's PCIe switched architecture employs multicast as a key element of our reflective memory solution. A single bus write transaction is sent to multiple remote targets or in PCI Express technical terms - multicast capability enables a single Transaction Layer Packet (TLP) to be forwarded to multiple destinations. PCI Express multicast results in a lower latency and higher bandwidth reflective memory solution. Dolphin benchmarks show end-to-end latencies as low as 0.99µs and over 5300 MBps dataflow at the application level. These performance levels solve many real time, distributed computing requirements.

Dolphin combines PCI Express multicast with our SISCI (Software Infrastructure for Shared-memory Cluster Interconnect) API. This combination allows customers to easily implement applications that directly access and utilize PCI Express multicast. Applications can be built without the need to write device drivers or spend time studying PCI Express chipset specifications.

Another main difference in Dolphin's reflective memory solution is the use of cacheable main system memory to store data. Cacheable main memory provides a significant performance and cost benefit. Remote interrupts or polling signal the arrival of data from a remote node. Polling is very fast since the memory segments are normal cacheable main memory and consume no memory bandwidth. The CPU polls for changes in its local cache. When new data arrives from the remote node, the I/O system automatically invalidates the cache and the new value is cached.

In addition to processors, FPGAs and GPUs applications can implement this reflective memory mechanism. The SISCI API can configure and enable GPUs, FPGAs etc (any PCIe master device) to send data directly to reflective memory, avoiding the need to first store the data in local memory. Data is written directly from a FPGA to multiple end points for processing or data movement. FPGAs can also receive data from multiple end points.

Reflective memory solutions are known for their simplicity, just read and write into a shared distributed memory. Our high-performance network increases simplicity with easy installation and setup. The SISCI Developers Kit includes tools to speed development and setup of your reflective memory system. Once setup, your application simply reads and writes to remote memory.

Features

- High-performance, ultra low-latency switched 64 Gbps and 40 Gbps data rate interconnect
- Gen 3 performance up to 5300 MBps data throughput
- Gen 2 performance up to 2886 MBps data throughput
- FPGA support
- Hardware based multicast
- Configurable shared memory regions
- Fiber-Optic and copper cabling support
- Scalable switched architecture
- SISCI developers kit
- Built in CRC, 8b/10b encoding
- PCI Expresshost adapters
- Expandable switch solutions

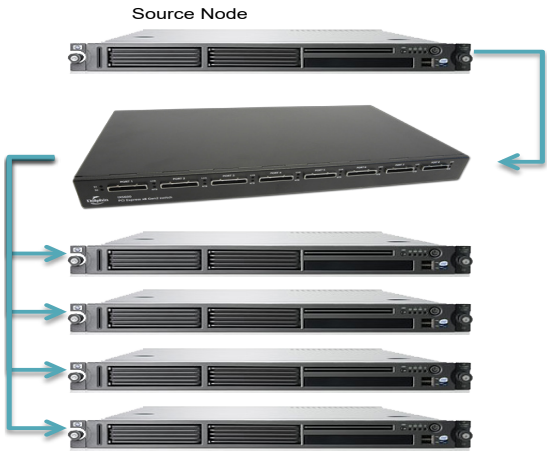


figure 1: Reflective memory diagram

Why use Dolphin Reflective Memory?

A main reason to use Dolphin's solution is the use of system memory. PCIe reflective memory creates a reflective memory address space between nodes based on system memory. This offers significantly higher performance at a much lower cost than other reflective memory solutions, while still supporting both low cost copper and long distance fiber cabling.

The combination of PCIe and system main memory offers significantly faster access to local data than other reflective memory solutions. These other solutions use expensive device memory, limiting system configuration options and increasing cost. Main memory solutions benefit from CPU caching and very high performance internal memory buses. In traditional reflective memory systems ,device memory is non-cacheable. Memory access is very expensive as the CPU must fetch the data from the card through the I/O system.

PCIe reflective memory functionality is implemented in hardware. Topology sizes range from 2 node systems without a switch to multi-switch large system configurations. For larger systems, switches are cascaded and connected to host with host adapter card.

For high throughput, the Dolphin PXH adapters come with a x8 PCIe link. This 64 Gbps link benefits customer applications that need exceptional link bandwidth. For scalablity, the IXH systems support up to 56 reflective memory nodes. The PXH family currently supports up to 12 nodes.

The implementation of a fully hardware based memory mapped data transmission system eliminates reliance on any operating system service or kernel driver functionality. Our extensive software library makes configuration and setup easy.

Reflective Memory Transfers

Reflective memory can be accessed in 3 different ways.

- CPU can do direct PIO/Store operations to a reflective memory address using a pointer or memcopy() function. The access is write posted and will typically be completed in one CPU instruction.
- PCIe DMA controllers can move data to a reflected memory address.
- Locally attached PCIe device, FPGA, GPU etc can send data to a reflective memory address. Data does not need to go to local memory first.

Reflective Memory Performance

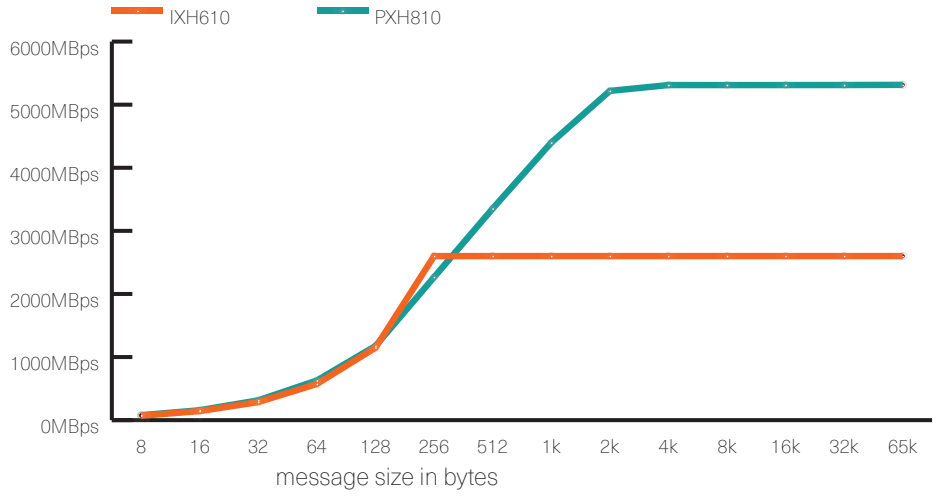
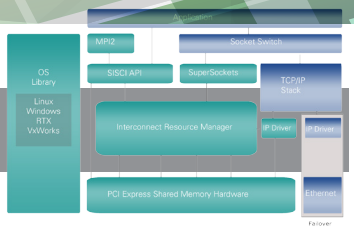


figure 2: Reflective Memory Throughput



PCIe Software

SISCI API Developer Kit

Dolphin's Software Infrastructure Shared-Memory Cluster Interconnect (SISCI) API makes developing PCI Express Network applications faster and easier. The SISCI API is a well established API for shared memory environments. In PCI Express multiprocessing architectures, the SISCI API enables PCIe based applications to use distributed resources such as CPUs, I/O, and memory. The resulting applications feature reduced system latency and increased data throughput.

For processor to processor communication, PCI Express supports both CPU driven programmed IO (PIO) and Direct Memory Access (DMA) as transports through non-transparent bridges (NTB). Dolphin's SISCI API utilizes these components in creating a development and runtime environment for systems seeking maximum performance. This very deterministic environment featuring low latency and low jitter is ideal for traditional high performance applications like real time simulators, reflective memory applications, high availability servers with fast fail-over, and high speed trading applications.

The SISCI API supports data transfers between applications and processes running in an SMP environment as well as between independent servers. SISCI's capabilities include managing and triggering of application specific local and remote interrupts, along with catching and managing events generated by the underlying PCIe system (such as a cable being unplugged). The SISCI API makes extensive use of the "resource" concept. Resources are items such as virtual devices, memory segments, and DMA queues.

The API removes the need to understand and manage low level PCIe chip registers. At the application level, developers utilize these resources without sacrificing performance. Programming features include allocating memory segments, mapping local and remote memory segments into addressable program space, and data management and transfer with DMA. The SISCI API improves overall system performance and availability with advanced caching techniques, data checking for data transfer errors, and data error correction.

Features

- Shared memory API
- PCI Express Peer to Peer support
- Replicated/reflective memory support
- Distributed shared memory and DMA support
- Low latency messaging API
- Interrupt management
- Direct memory reads and writes
- Windows, RTX, VxWorks, and Linux support
- Supports data transfers between all supported OS and platforms.
- Caching and error checking support
- Events and callbacks
- Example code available

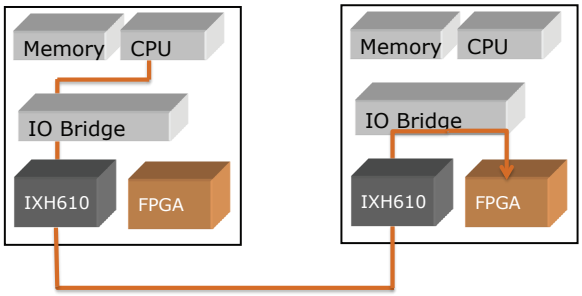


figure 3: Device to device transfers

Why use SISCI?

The SISCI software and underlying drivers simplify the process of building shared memory based applications. For PCIe based application development, the API utilizes PCI Express non-transparent bridging to maximum application performance. The shared memory API drivers allocate memory segments on the local node and make this memory available to other nodes. The local node then connects to memory segments on remote nodes.

Once available, a memory segment is accessed in two ways, either mapped into the address space of your process and accessed as a normal memory access, e.g. via pointer operations, or use the DMA engine in the PCIe chipset to transfer data. Figure 4 illustrates both data transfer options.

Mapping the remote address space and using PIO may be appropriate for control messages and data transfers up to e.g. 1k bytes, since the processor moves the data with very low latency. PIO optimizes small write transfers by requiring no memory lock down, data may already exist in the CPU cache, and the actual transfer is just a single CPU instruction – a write posted store instruction. A DMA implementations saves CPU cycles for larger transfers, enabling overlapped data transfers and computations. DMA has a higher setup cost so latencies usually increase slightly because of the time required to lock down memory and setup the DMA engine and interrupt completion time. However, more data transfers joined and sent together to the PCIe switch in order amortizes the overhead.

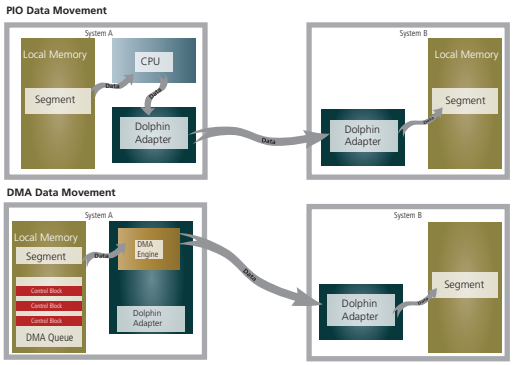


figure 4: SISCI data movement model

The built in resource management enables multiple concurrent SISCI programs and other users of the PCIe network to coexist and operate independent of each other. The SISCI API is available in user space and a similar API is available in kernel space.

SISCI Performance

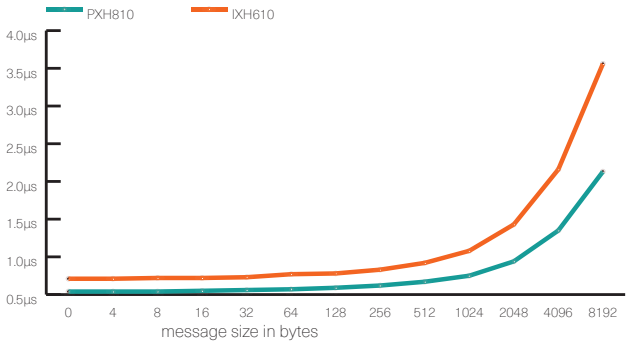


figure 5 - PXH810 and IXH610 Latency

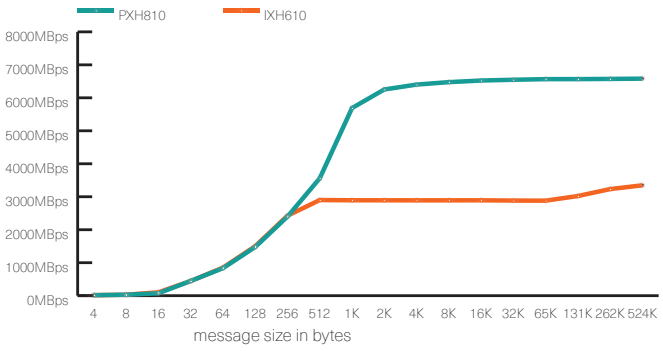
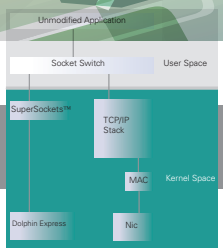


figure 6 - SISCI PIO/DMA Throughput

The SISCI API provides applications direct access to the low latency messaging enabled by PCI Express. Dolphin SISCI benchmarks show latencies as low as 0.54µs. The chart on Figure 5 show the latency at various message sizes. The SISCI API enables high throughput applications. This high performance API takes advantage of the PCI Express hardware performance to deliver over 6500 MBps for Gen 3 and 3500 MBps for Gen 2 of real application data throughput. Figure 6 shows the throughput at various message sizes using Dolphin IXH and PXH host adapters



PCIe Software

SuperSockets™

PCI Express can replace local Ethernet networks with a high speed low latency network. SuperSockets is a unique implementation of the Berkeley Sockets API. With SuperSockets , network applications transparently capitalize on the PCIe transport to achieve performance gains.

Dolphin PCIe hardware and the SuperSockets software support the most demanding sockets based applications with an ultra-low latency, high-bandwidth, low overhead, and highly available platform. New and existing Linux and Windows applications require no modification to be deployed on Dolphin's high performance platform.

Traditional implementations of TCP sockets require two major CPU consuming tasks: data copy between application buffers and NIC buffers along with TCP transport handling (segmentation, reassembly, check summing, timers, acknowledgments, etc). These operations turn into performance bottlenecks as I/O interconnect speeds increase. SuperSockets eliminates the protocol stack bottlenecks, delivering superior latency performance. Our ultra-low latency remote memory access mechanism is based on a combination of PIO (Programmed IO) for short transfers and DMA (Direct Memory Access) for longer transfers, allowing both control and data messages to experience performance improvements.

SuperSockets is unique in its support for PIO. PIO has clear advantages for short messages, such as control messages for simulations systems. Transfers complete through a single CPU store operation that moves data from CPU registers into remote system memory. In most case, SuperSockets data transfers complete before alternative technologies start their RDMA transfer.

In addition to PIO, SuperSockets implements a high speed loopback device for accelerating local system sockets communication. This reduces local sockets latency to a minimum. For SMP systems, loopback performance is increased 10 times.

SuperSockets comes with built in high availability, providing instantaneous switching during system or network errors. If the PCI Express® network fails, socket communication transfers to the regular network stack. The Linux version supports an instant fail- over and fail-forward mechanism between the PCIe and regular network.

Features

- Windows and Linux support
- Full support for socket inheritance/duplication
- Includes local loopback socket acceleration up to 10 times faster than standard Linux and Windows
- No OS patches or application modifications required
- Easy to install with no application modifications
- Linux to Windows connectivity available soon

Linux Specific Features

- TCP, UDP, and UDP multicast support
- Supports both user space and kernel space applications
- Compliant with Linux Kernel Socket library and Berkeley Sockets
- Transparent fail-over to Ethernet if high speed connection fails and falls back when problem is corrected

Windows Specific Features

- TCP support , UDP and UDP multicast being implemented
- Supports user space applications
- Compliant with WinSock2 API
- Fail-over to Ethernet if high speed connection is not available at start-up

How Does SuperSockets™ Work?

In order to divert socket communication, without touching the application, the sockets API functions must be intercepted. This is done differently in Windows and Linux environments.

Dolphin SuperSockets on Linux differs from regular sockets only in the address family. SuperSockets implement an AF_INET compliant socket transport called AF_SSOCK. The Linux LD_PRELOAD functionality is used to preload the standard sockets library with a special SuperSockets library that intercepts the socket() call and replaces the AF_INET address family with AF_SSOCK. All other sockets calls follow the usual code path. Target addresses within the Dolphin Express Network are accelerated by the SuperSockets module.

For Windows applications or services, a Layered Service Provider(LPS) module is installed and automatically configured. The LSP accelerates socket transfers initiated by AF_INET or AF_INET6, SOCK_STREAM endpoints. The SuperSockets stack provides a proxy application called dis_ssocks_run.exe that enables specific programs to use the PCI Express path. By default, the LSP is a pass-through module for all applications: the network traffic passes through the NDIS stack.

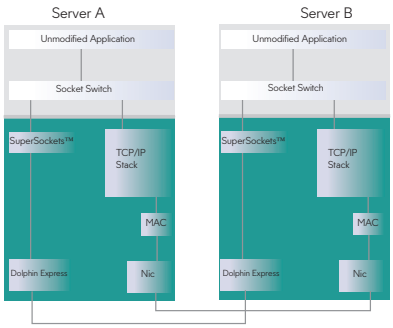


figure 8: SuperSockets™ vs. Ethernet data model

The network acceleration over PCI Express occurs when the interconnect topology is fully functional, the client and server programs are launched under the proxy application's control and both sides use the standard Winsock2 API calls. At runtime, a native socket is created and used for initial connection establishment. Therefore all connections are subject to typical network administrative policies.

The supported transfer modes are blocking, non-blocking, overlapped, asynchronous window and network events. The Service Provider balances the CPU consumption based on the traffic pattern. Dedicated operating system performance counters are additionally provided.

SuperSockets™ Performance

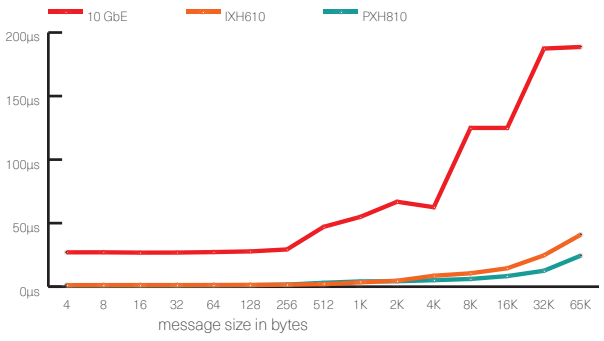


figure 9: SuperSockets latency

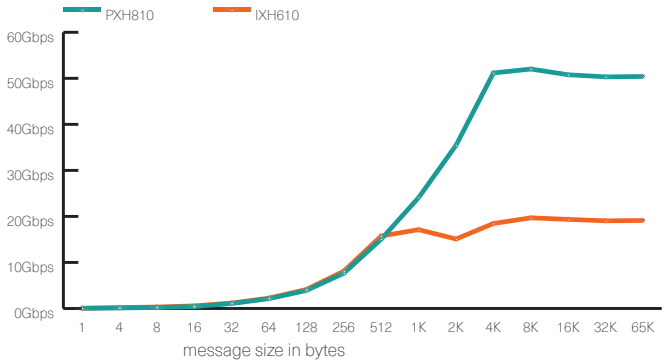


figure 10: SuperSockets™ data throughput

SuperSockets is optimized for high throughput, low latency communication by reducing system resource and interrupt usage in data transfers. The latency chart above shows performance results using PCI Express vs 10 Gigabit Ethernet. The socket ping-pong test shows the half RTT (Round Trip Time). The minimum latency for Dolphin SuperSockets is under 1 microseconds. SuperSockets also delivers high throughput with over 53 Gb/s of data throughput with our Gen3 PXH810 product.

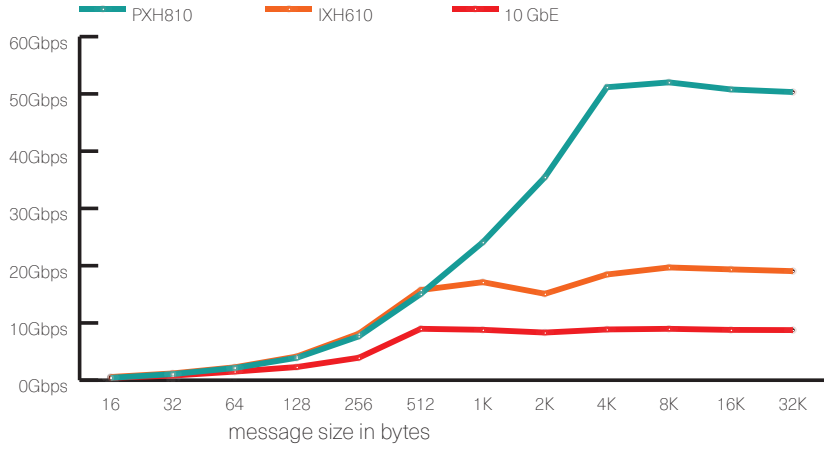
PCIe Software

IP over PCIe

Dolphins performance optimized TCP IP driver for PCIe (IPoPCIe) provides a fast and transparent way for any networked applications to dramatically improve network throughput.

The software is highly optimized to reduce system load (e.g. system interrupts) and uses both PIO and RDMA operations to implement the most efficient transfers for all message sizes. The major benefits are plug and play and much higher bandwidth and lower latency than network technologies like 40 GbE.

At the hardware level, the TCP/IP driver provides a very low latency connection. The operating system networking protocols typical introduce a significant delay for safe networking (required for non reliable networks like Ethernet). The IPoPCIe driver implements these networking protocols increasing latency. User space applications seeking the lowest possible network latency should utilize the Dolphin SuperSockets technology. The IPoPCIe driver will typically provide 5-6 times better throughput than 10G Ethernet.



Features

- All networked, users space and kernel space applications are supported.
- 100% compliant with Linux Socket library, Berkeley Socket API and Windows WinSock2.
- No OS patches or application modifications required. Just install and run
- Routing between networks
- ARP support
- Both TCP and UDP supported. (UDP multicast/broadcast is not supported yet using Linux, but SuperSockets for Linux supports UDP multicast)
- Supports hot-pluggable links for high availability operation
- Easy to install

IPoPCIe Uses

The optimized TCP/IP driver is recommended for applications like

Windows

- Microsoft Hyper-V live migration
- Network file sharing (map network drive)
- Applications that require UDP (not support by SuperSockets yet).

Linux:

- General networking
- NFS
- Cluster file systems not supported by SuperSockets
- iSCSI

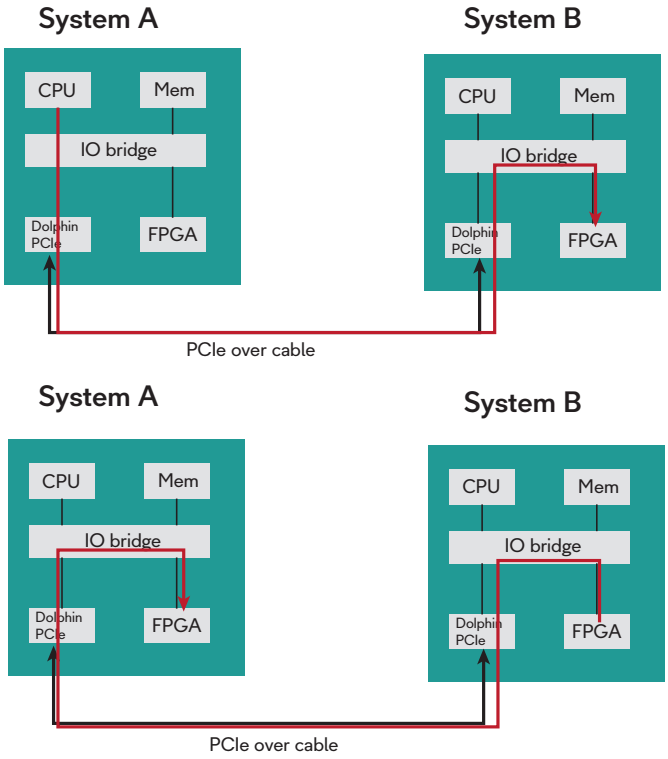
PCIe Software

FPGAs and GPUs

Remote Peer-to-Peer

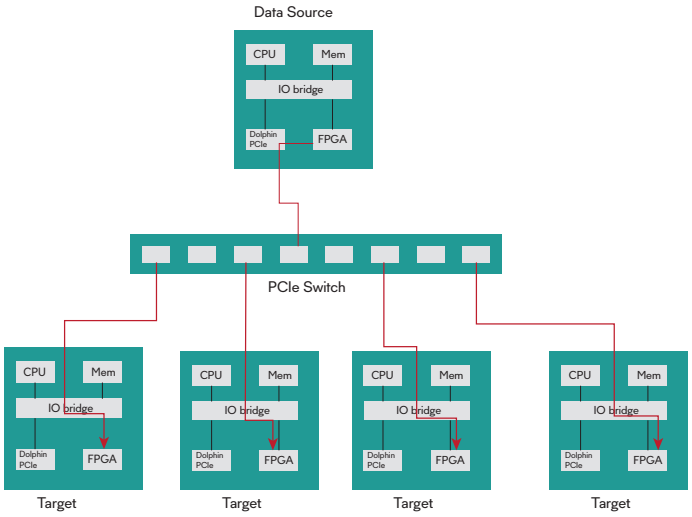
PCIe peer-to-peer communication (P2P) is a part of the PCI Express specification and enables regular PCIe devices to establish direct data transfers without the use of main memory as a temporary storage or the CPU for data movement. This significantly reduces communication lateny. PCIe Networks expand on this cabablity by enabling remote systems to establish P2P communication. Intel Phi, GPUs, FPGA, specialized data grabbers can exploit remote P2P communication to reduce latency and communication overhead.

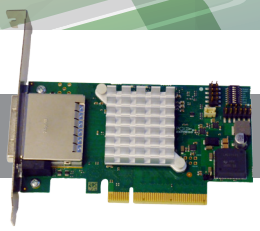
The SISCI API specification supports this functionality and provides a simplified way to setup and manage remote peer-to-peer transfers. SISCI software enables applications to use PIO or DMA operations to move data directly to and from local or remote PCIe devices.



Multicast with FPGAs

FPGAs can take advantage of PCIe multicast. It is also possible to combine P2P communications with PCIe multicast or reflective memory to transparently and instantly deliver data to multiple devices. SISCI software combined with an IX and PX cluster using the IXS600 switch installed creates a PCIe multicast environment. With PCIe multicast and PCIe peer to peer transfers a FPGA can send data to multiple targets using a single posted write transaction. This is ideal for systems needing a deterministic way to transfer FPGA data with ultra-low latency.





PCIe Hardware

PXH Gen3 PCIe Host Adapters

The PXH81x Gen3 PCIe Host Adapters are high performance network interface cards for distributed processor subsystems and I/O expansion applications. The host adapters extend PCI Express over cables to external systems. Based on an Avago Gen3 PCIe switch architecture, the PXH81x host adapter includes advanced features for non-transparent bridging (NTB) and clock isolation.

For high performance application developers, the PXH81x host adapter combines 64 Gbps performance with less than one microsecond latency, significantly improving overall inter-system communication. Network applications benefit from the high throughput and low latency. The PXH810 performs Remote Direct Memory Access (RDMA), with both DMA and Programmed IO (PIO) transfers, effectively supporting both large and small data packets. DMA transfers result in efficient larger packet transfers and processor off-load. PIO transfers optimize small packet transfers at the lowest latency. The combination of DMA and PIO creates a highly efficient data transfer system.

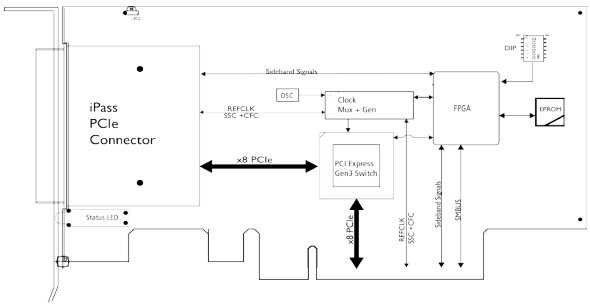
Dolphin's software suite takes advantage of PCI Express' DMA and PIO data transfer scheme. Delivering a complete deployment environment for customized and standardized applications. The suite includes a Shared-Memory Cluster Interconnect (SISCI) API as well as a IPoPCIe driver and SuperSockets software. The SISCI API is a robust and powerful shared memory programming environment. The optimized IPoPCIe driver and SuperSockets software remove traditional networking bottlenecks, allowing standard IP and sockets applications to take advantage of the high performance PCI Express interconnect without modification. The overall framework is designed for rapid development of inter-processor communication systems.

With the implementation of clock isolation, the PXH81x's signal quality is excellent. By isolating the system clock and transmitting an extremely low jitter high quality clock to downstream devices, the PXH81x offers users high signal quality and increased cable distances. The PXH81x supports x8 PCIe copper cables up to 5 meters and fiber optic cables up to 100 meters. The improved signal quality for extended copper distances and fiber optic support makes the PXH81x ideal for simulation systems. The PXH81x is also used in applications such as test and measurement equipment, medical equipment, and storage subsystem seeking high performance and data quality.

The PXH810 is targeted for network applications and includes a license for the networking software. The PXH812 is designed for I/O expansion application and supports being a host or target adapter.

Features

- PCI Express Gen3 compliant - 8.0 Gbps per lane
- Link compliant with Gen1, Gen2, and Gen3 PCIe
- PCIe iPass Connectors
- One x8 PCIe port
- RDMA support through PIO and DMA
- Copper connection up to 5 meters, Fiber-optic cable connection up to 100 meters
- Clock isolation support, CFC or SSC on cable
- Transparent host and target operations along with non-transparent bridging to cabled PCIe systems
- Low Profile PCIe form factor
- EEPROM for custom system configuration
- Link status LEDs through face plate



Cluster and I/O connections

The PXH810 adapter delivers extremely high throughput. This is demonstrated by using the SISCI API. The PXH810 PCIe hardware performance delivers over 6500 MBps of real application data throughput for high performance communication. Figure 13 shows the throughput at various message sizes using the PXH810 Gen 3 host adapters.

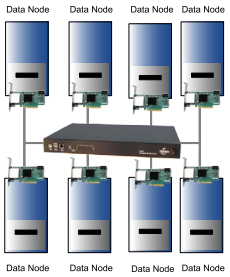


figure 12: Eight node PXH810 cluster

Performance

When used for inter-host connections, the PXH810 adapter is capable of node to node connections or connections through a IXS600 Switch as shown in figure 12. Each connection supports 64 Gbps with latencies as low as 0.54 microseconds. Designed for x8 PCIe Systems, the PXH810 supports any system with a standard x8 or x16 PCIe slot. Gen3 operation requires a system that supports Gen3 PCI Express.

The PXH812 functions as a high quality transparent connection to remote PCIe I/O subsystems. The PXH812 is specially designed for higher signal quality and support for spread spectrum clocking. It can be used as both a host and target adapter in I/O expansions.

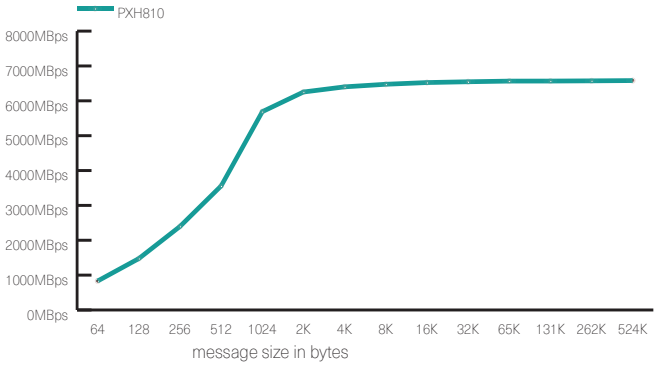
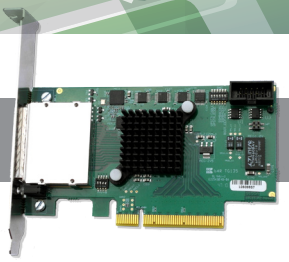


figure 13: PXH810 Throughput

Specifications

Link Speeds	64 Gbps
Application Performance	0.54 microsecond latency (application to application) 6595 MBps throughput
Active Components	PLX PCI Express Switch
PCI Express	Base Specification 3
Topologies	Switch/ point to point
Cable Connections	x8 iPass copper cable, fiber optic cable support Supports x8 to x4 transition cables
Power Consumption	12 watts
Mechanical Dimensions	Low profile - 68.90 mm (2.731 inches) x 120 mm (6.600 inches)
Operating Environment	Operating Temperature: 0°C -55°C Relative Humidity: 5% -95% non-condensing
Dolphin Software	SuperSockets Berkley Sockets API Microsoft WinSock2/LSP support IPoPCIe driver SISCI API

Usage Modes	Host/Target Non-transparent bridging
Regulatory	CE Mark FCC Class A UL94V-0 compliant RoHS Compliant
Configuration	Dip-switch configurable
Mounting Plates	Full height plate installed Half height plate included with board
Operating Systems supported	Windows, RTX, Linux, VxWorks
Product Codes	PXH810 Host NTB Adapter PXH812 Transparent only Host/ Target Adapter



PCIe Hardware

IXH Gen2 PCIe Host Adapters

The IXH610 Gen2 PCIe Host Adapter is a high performance cabled interface to external processor subsystems or I/O subsystems. Based on Gen2 PCI Express bridging architecture, the IXH610 host adapter includes advanced features such as non-transparent bridging (NTB) and clock isolation.

For high performance application developers, the IXH610 host adapter combines 40 Gbps performance with less than one microsecond latency, significantly improving overall inter-system communication. Connecting remote I/O subsystems in transparent mode requires no special drivers, so deployment is fast and easy. Inter-processor communication also benefits from the high throughput and low latency.

The IXH610 performs both Direct Memory Access (DMA) and Programmed IO (PIO) transfers, effectively supporting both large and small data packets. DMA transfers result in efficient larger packet transfers and processor off-load. PIO transfers optimize small packet transfers at the lowest latency. The combination of DMA and PIO creates a highly potent data transfer system.

Dolphin's software suite takes advantage of this data transfer scheme. Delivering a complete deployment environment for customized and standardized applications. The Shared-Memory Cluster Interconnect (SISCI) API is a robust and powerful shared memory programming environment. The optimized TCP/IP driver and SuperSocketssoftware remove traditional networking bottlenecks. IP and sockets applications take advantage of the high performance PCI Express interconnect without modification. The overall framework is designed to meet all the demands for rapid development of inter-processor communication systems.

With the implementation of clock isolation, the IXH610's signal quality is excellent. By isolating the system clock and transmitting an extremely low jitter high quality clock to downstream devices, the IXH610 offers users high signal quality and increased cable distances. Signal quality is essential for applications such as test and measurement equipment, medical equipment, and storage subsystem seeking high performance and data quality.

Features

- PCI Express 2.1 compliant - 5.0 Gbps per lane
- x8 PCI Express port - 40 Gbps
- Link compliant with Gen1 and Gen2 PCI Express
- Support Gen1, Gen2, and Gen3 PCIe Slots
- RDMA support through PIO and DMA
- PCI Express External Cabling Specification
- PCI Express x8 iPass Connectors
- Copper cables - up to 7 meters
- Clock isolation support
- Transparent bridging to cabled I/O devices
- Non-transparent bridging to cabled PCIe systems
- Low Profile PCIe form factor
- EEPROM for custom system configuration
- Link and status LEDs through face plate

Cluster connections

When used for clustered connections, the IXH610 adapter is capable of node to node connections or connections through a IXS600 Switch as shown in figure 14. Adding industrial systems is done by connecting to the IXH620 XMC adapter. Each connection supports 40 Gbps with latencies as low as 0.74 microseconds. Designed for x8 PCIe Systems, the IXH610 supports any system with a standard x8 or x16 PCIe slot. The IXH631 uses MiniSAS HD connectors to create up to a 5 node cluster without a switch.

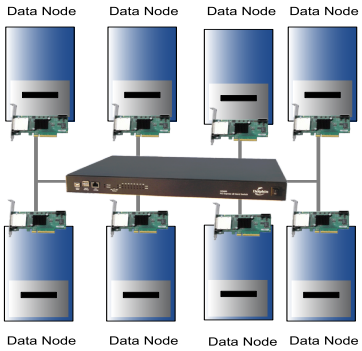


figure 14: Eight node PCI Express Cluster

Remote I/O Connections

The IXH610 functions as a high quality transparent connection to remote PCIe I/O subsystems. These subsystems include test equipment, I/O expansion systems, specialized equipment, and storage systems. The IXH610 is specially designed for higher signal quality and support for spread spectrum clocking. The IXH611 is used as a target adapter in I/O expansion applications.

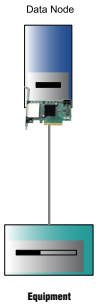


figure 15: I/O expansion with PCI Express

Specifications

Link Speeds	40 Gbps
Application Performance	0.74 ms latency (application to application) 3,500 MBps Throughput
PCI Express	Base Specification 2.1
Topologies	Point to point, Switched
Cable Connections	x8 iPass copper cable support Supports x8 to x4 transition cables
Power Consumption	7 watts
Mechanical Dimensions	PCI Express Card Electromechanical Specification 2.0
Operating Environment	Operating Temperature: -10°C -60°C Relative Humidity: 5% -95% non-condensing
Dolphin Software	SuperSockets Berkeley Sockets API Microsoft WinSock2/LSP support SISCI API IPoPCIe

User Configuration Modes	Transparent Host Target adapter Non-transparent Host(NTB)
Regulatory	CE Mark EN 55022,EN 55024-A1&A2, EN 61000-6-2 FCC Class A UL94V-0 compliant RoHS Compliant
Operating Systems supported	Windows Linux VxWorks RTX
Product Codes	IXH610 - Host / NTB Adapter IXH611 - Host/Target Adapter IXH614 - 104Mhz Overclocked Host Adapter IXH631 MiniSAS HD Host Adapter

PCIe Hardware

IXH Gen2 PCIe XMC Host Adapter

As performance needs increase in embedded systems, high performance embedded system designers are implementing PCI Express for inter-system communication. The IXH620 XMC cable adapter enables high speed PCI Express inter-system connections to external systems including servers, single board computers, and I/O subsystems.

The IXH620's 40 Gbps throughput and sub one microsecond latency deliver superior performance and low latency to systems supporting standard XMC slots or XMC VPX, VME or cPCI carrier boards. The IXH620's x8 cable connection uses an iPass™connector as a standardized connection method. The XMC adapter supports either upstream or downstream target configurations. To connect remote I/O or processing, the IXH620 implemented IDT's transparent or non-transparent bridging (NTB) functions. Used in transparent mode, standard PCI Express devices and drivers require no modifications. In NTB mode, the adapter facilitates inter-processor communication through Programmed IO (PIO) or Remote Direct Memory Access (RDMA).

The IXH620 implements a system clock isolation system for excellent signal quality. By isolating the system clock and transmitting an extremely low jitter high quality clock to downstream devices, IXH620 users benefit from improved signal quality, reliability, and cable distances.

The IXH620 comes with Dolphin's comprehensive software suite that reduces time to market for customer applications. The Shared-Memory Cluster Interconnect (SISCI) API is a robust and powerful programming environment for easy development of shared memory applications. Shared memory applications benefit from the 0.74 microsecond inter-system latency and more than 3500 Megabytes/s throughput. The optimized TCP/IP driver and SuperSockets software remove traditional networking bottlenecks. IP and Sockets applications can take advantage of the high performance PCI Express interconnect. Sockets applications experience 1.25 microsecond latency and 23 Gigabit/second user payload throughput.

These powerful features make XMC adapter an ideal interconnect for applications such as military and industrial systems that seek high performance and flexibility.

Features

- PCI Express 2.1 compliant - 5.0 Gbps per lane
- x8 PCIe port - 40 Gbps
- Link compliant with Gen1 PCI Express
- VITA 42.0-2005, ANSI/VITA 42.3-2006 compliant
- RDMA support through PIO and DMA
- PCI Express External Cabling Specification
- PCIe x8 iPass Connectors
- Copper cable connections up to 7 meters copper connections
- Clock isolation support
- Transparent bridging to cabled I/O devices
- Non-transparent bridging to cabled PCI Express systems
- Short XMC form factor
- EEPROM for custom system configuration
- XMC P15 connector
- Link and status LEDs through front panel

Host to host connections

The IXH620 connects single board computers or systems running a distributed processing environment. Industrial or military customers requiring redundancy or increased compute resources use the IXH620 by itself or in conjunction with the IXS600 switch. Figure 16 illustrates connecting two single board computers with the IXH620. Fast data transfers are enabled through Dolphin's shared memory mechanism.

figure 16: two node connection with XMC adapter

Remote I/O subsystems

To extend the I/O capabilities of an XMC enabled system, the IXH620 supports host adapter or target adapter configurations. Industrial and Military customers requiring increased I/O bandwidth for graphics, processing, or data collection can add IXH620 enabled XMC carrier cards and chassis. Figure 17 illustrates connecting a standard server to and XMC enabled chassis to attach additional XMC cards for test system or increased functionality.

figure 17 connect embedded equipment to hosts

Specifications

Link Speeds	40 Gbps
Application Performance	0.74 microsecond latency (application to application)
Specifications	PCI Express Base Specification 2.1 VITA 42.0-2005, ANSI/VITA 42.3-2006
Topologies	Point to point, Switched
Cable Connections	x8 iPass copper cable support Supports x8 to x4 transition cables
Power Consumption	7 watts
Mechanical Dimensions	XMC Short form factor
Operating Environment	Operating Temperature: -10°C -60°C Relative Humidity: 5% -95% non-condensing
Dolphin Software	SuperSockets Berkeley Sockets API Microsoft WinSock2/LSP support SISCI API IPoPCIe

User Configuration Modes	Transparent/non-transparent(NTB)
Regulatory	CE Mark EN 55022,EN 55024-A1&A2, EN 61000-6-2 FCC Class A UL94V-0 compliant RoHS Compliant
Operating Systems supported	Windows Linux VxWorks RTX
Product Codes	IXH620 Host /Target / NTB adapter

PCIe Hardware

IXS Gen3 PCIe Switch

PCI Express provides low latency, highly efficient switching for high performance applications. The IXS600 Gen3 PCI Express switch delivers a powerful, flexible, Gen3 switching solution. This powerful switch enables I/O scaling and inter-processor communication by combining transparent and non-transparent bridging capabilities with Dolphin's software and clustering technology. IXS600 users can connect multiple PCI Express devices or create a highly efficient compute cluster with PCs, servers, or SBCs with XMC sites.

The IXS600 is the switching element of Dolphin's product line. This eight port, 1U cluster switch delivers 64 Gbps of non-blocking bandwidth per port at ultra-low latencies. Each x8 PCI Express port delivers maximum bandwidth to each device while maintaining backwards compatibility with Gen1 and Gen2 components. As with other Dolphin products, the IXS600 utilizes standard iPassconnectors to string components via copper or fiber-optic cabling. IXS600 customer can link multiple standardized PCI Express products such as PXI chassis, storage, and I/O expansion units.

For Non Transparent Bridging (NTB) or clustering applications, the IXS600 integrates with Dolphin's PCI Express Host Adapters or XMC Adapter. The total NTB solution solves many of the problems related to PCI Express operations, such as power sequencing and standard software. Dolphin customers avoid the severe power-on requirements normally associated with PCI Express. Hosts, cables, and the switch can be hot-swapped and power cycled in any sequence for real plug and play.

The IXS600 switch can also be partitioned into several virtually independent partitions, e.g. mixing NTB and Transparent functionality on separate ports.

The IXS600 switch supports Dolphin's comprehensive software suite. The Shared-Memory Cluster Interconnect (SISCI) API is a robust and powerful programming environment for easy development of shared memory applications. The optimized TCP/IP driver and SuperSocketssoftware remove traditional networking bottlenecks. IP and Sockets applications can take advantage of the high performance PCI Express interconnect. This comprehensive solution is ideal for real-time, technical and high performance computing, cloud computing, and enterprise business applications.

The IXS600 switch is compatible with all Dolphin Gen2 and Gen3 PCI Express products.

Features

- PCI Express 3.0 compliant -8.0 Gbps per lane
- Eight PCI Express Gen3 x8 ports
- PCI Express x8 iPass Connectors
- Auto-training to lower lane widths
- Supports x4 lanes with a transition cable
- Link compliant with Gen1 and Gen2 PCI Express
- Transparent and Non Transparent support
- PCI Express External Cabling Specification
- Fiber-optic and copper cable support
- Hot Plug PCI Express cabling support in NTB mode
- Built in management processor
- Boot configuration and user data support for cluster and system configuration
- 19 Inch 1U rack mountable chassis
- Front and rear operational status and alert LEDs
- Redundant Fans

Cluster connections

The IXS600 is a key element in Dolphin's inter-processor connectivity strategy. Industrial, military, or enterprise customer can create diverse multiprocessing configurations. The hybrid configuration illustrated shows single board computers and data nodes connected through the switch. The IXH610 host adapter and IXH620 XMC adapter are used to connect to different compute nodes.

Increasing the number of I/O components in a system is accomplished by using the IXS600 with PCIe I/O expansion boxes. Figure 18 illustrates the IXS600 connecting 7 additional I/O expansion boxes. These boxes can accommodate components such as sensors, graphics, coprocessors, video capture cards, and other I/O devices.

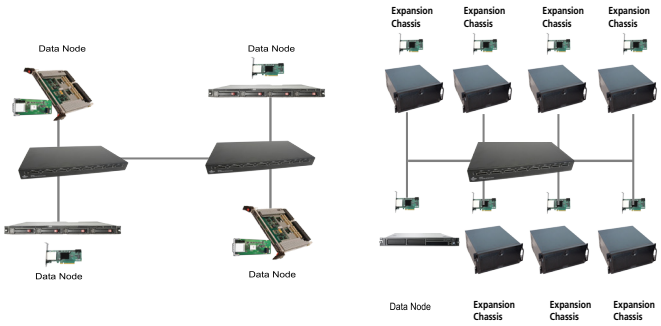


figure 18: I/O and hybrid configuration

Scalability

Scalability is achieved by connecting multiple switches. Multiple IXS600 switches are used for larger reflective memory applications or creating larger processing clusters.

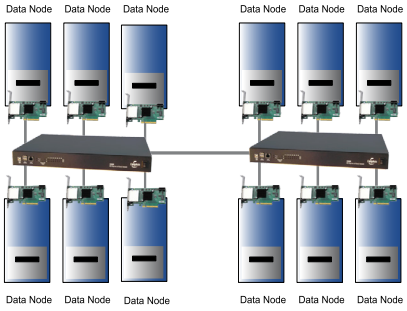


figure19: Scalability of switches

Specifications

Link Speeds	64 Gbps per port
Application Performance	6595 MBps application data rate 200ns port latency
Ports	8 - x8 non-blocking 64 Gbps ports
Cable connections	x8 iPass copper cables Fiber-Optic cables-
Management	Ethernet management port Operational status and alert LEDs
Power	Auto-sensing power supply 110 - 240 V AC 50-60Hz Power consumption (including power to the optional fiber transceiver cables) : Max 80 Watts
Mechanical	1U, 19 inch rackmountable chassis 440mm (W) x 300mm (D) x 45mm(H) Redundant Fans

User Configuration Modes	Transparent/non-transparent(NTB)
Operating Environment	Operating Temperature: 0°C -55°C Relative Humidity: 5% -95% non-condensing
Regulatory	CE Mark EN 55022,EN 55024-A1&A2, EN 61000-6-2 FCC Class A UL94V-0 compliant RoHS
Product Codes	IXS600 Gen3 100 Mhz IXS604 Gen3 104 Mhz

Product Comparison Charts

Software Comparison

Feature	SISCI	SuperSockets	IPoPCIe Driver
Linux Platform Support	x86, x86_64 ARM 64	x86, x86_64 ARM 64	x86 ARM 64
Window Platform Support	x86, x86_64	x86, x86_64	x86, x86_64
RTX Platform Support	x86, x86_64	No	No
VxWorks	x86_64	No	No
Latency	0.54 µs Linux PXH810	0.98 µs Linux	5.6 µs Linux
Max Bandwidth	6,500 MBps Linux	54 Gbps	54 Gbps
Reflective Memory Support	Yes	N/A	N/A
Address based Multicast support	Yes	N/A	N/A
TCP Support	N/A	Yes	Yes
UDP Support	N/A	Linux -Yes Windows - No	Yes
UDP Multicast Support	N/A	Linux -Yes Windows - under development	Linux -under development Windows - Yes
Application modifications required	Yes	No	No
Cross O/S and Platform data transfer support	All	Under development	Under development
Peer to Peer Transfers support	Yes	N/A	N/A
Accelerated Loopback Support	Yes	Yes	No
Open Source code available	Yes	No	No
Licensing of software available	Yes	Yes	Yes
Chipset support	PLX IDT Intel NTB	PLX IDT Intel NTB	PLX IDT Intel NTB

Hardware Comparison

Feature	PXH810	PXH812	IXH610	IXH611	IXH620	IXH614
Use	NTB/Host/ Target	Host/Target	NTB/Host	Host/Target	NTB/Host/Target	NTB/Host/Target
Form factor	Short/ Low profile PCIe	Short/ Low profile PCIe	Short/ Low profile PCIe	Short/ Low profile PCIe	Short XMC	Short/ Low profile PCIe
PCIe max connection speed	x8 Gen3	x8 Gen3	x8 Gen2	x8 Gen2	x8 Gen2	x8 Gen2
NTB	Yes	No	Yes	No	Yes	Yes
Transparent Target	Yes	Yes	No	Yes	Yes	Yes
Transparent Host	Yes	Yes	Yes	Yes	Yes	Yes
Clocking	100Mhz - 104Mhz	100MHz	100MHz	100MHz	100MHz	104MHz
Max cable length	5 Meter Copper	5 Meter Copper	7 Meter	7 Meter	7 Meter	7 Meter
Fiber support	100 Meter	100 Meter	No	No	No	No
IXS600 Switch support	Yes	Yes	Yes	Yes	Yes	Requires IXS604
Scalability NTB, general use	8 - 32 nodes	N/A	20 nodes	N/A	20 nodes	20 nodes
Scalability Reflective memory functionality only	8 - 32 nodes	N/A	56 nodes	N/A	56 nodes	56 nodes
Transparent scalability	256 bus numbers	256 bus numbers	256 bus numbers	256 bus numbers	256 bus numbers	256 bus numbers
Maximum reflective memory segment size	2 GB	N/A	2 GB	N/A	2 GB	2 GB
Number of reflective memory segments	64	N/A	4	N/A	4	4
Total reflective memory size	64 x 2 GB	N/A	4 x 2 GB	N/A	4 x 2 GB	4 X 2 GB
Max PIO Performance	5312 MBps	N/A	2950 MBps	N/A	2950 MBps	2950 MBps
Max DMA Performance	6695 MBps	N/A	3500 MBps	N/A	3500 MBps	3500 MBps
Scipp latency back to back	0.54 µs	N/A	0.74 µs	N/A	0.74 µs	0.74 µs
Scipp latency switch	0.71 µs	N/A	0.9 µs	N/A	0.9 µs	0.9 µs

Dolphin Interconnect Solutions has been a global provider of ultra-low latency, high-bandwidth computer interconnect solutions for high speed real-time systems, clustered databases, general networking, web services and industrial applications for more than 20 years. For more information, please visit www.dolphinics.com.

Visit us on LinkedIn and Facebook:

- https://www.linkedin.com/company/dolphin-interconnect-solutions
- https://www.facebook.com/dolphinInterconnect