SCSI and IPI Disk Interface Directions

Introduction

One of the primary goals of Sun Microsystems® has always been to deliver the best technology for maximum performance at the most favorable price. These goals are the driving force behind technology changes at Sun™. Over the past year, SCSI-2 has displaced IPI-2 in high end deskside workstation and server products. Over the next year, SCSI-2 will replace IPI-2 even for large Data Center Servers. Why is this going to occur? What has changed that allowed SCSI to migrate from a moderate performance interface to a high performance interface like IPI? Why change the interface for high end systems even if SCSI can reach equal performance levels? This paper will explain why today's SCSI meets Sun's goals and our customer's needs better than IPI, and will explain the rational behind the direction Sun is taking.

Sun I/O Standards

Over the years, Sun has implemented three industry standard I/O architectures designed to best meet the needs of the wide range of systems. The three interfaces that have been deployed are SCSI/SCSI-2 (Small Computer System Interface), SMD (Storage Module Device), and IPI-2 (Intelligent Peripheral Interface). Currently Sun offers only SCSI-2 and IPI-2 interfaces.

The original SCSI implementation at Sun was a single ended asynchronous SCSI and was used to communicate with both embedded and non-embedded SCSI disk and 1/4" tape devices. SCSI was originally used as a low cost, moderate performance interface and was primarily used for desktop systems. With the introduction of the SPARCstation™ 1 in 1989, Sun brought the SCSI implementation into compliance with the SCSI-2 command set and implemented a new SCSI driver architecture called the Sun Common SCSI Architecture (SCSA), to facilitate the writing and addition of new target or host adapter drivers. Sun's SCSA implementation also supports synchronous data transfers. SCSI is now used to control more devices such as 1/2" tape, 8mm tape, and CD-ROM.

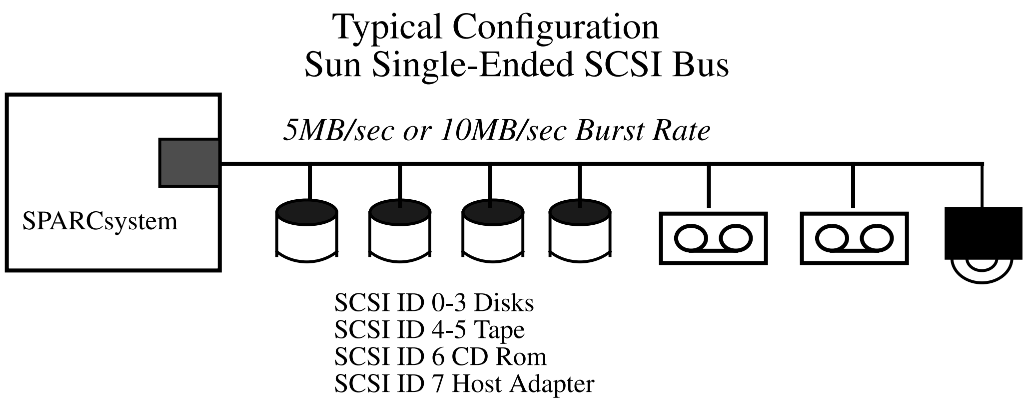
On high-end deskside workstations and servers, Sun originally used the SMD interface to support the highest performance industry standard drives available at the time. However, SMD provided no performance growth path to keep pace with the increasing performance of the systems. SMD has an architectural limitation of 3MB/sec without major changes to the electrical interface. Also, being a device level interface it did not provide the intelligence needed to optimize performance across a large number of drives. In addition, the use of dual cables (command and data) made large dual ported configurations very difficult. For these reasons SMD was replaced by IPI for Data Centers and deskside systems.

Sun SCSI Overview

The SCSI interface is an intelligent device independent interface. Sun's implementation is in compliance with the SCSI-2 ANSI standard. SCSI is a peer to peer protocol which allows devices to arbitrate for the bus. The single ended 8-bit wide interface has a maximum cable length of 6 meters and a burst synchronous transfer rate of up to 10MB/sec. From the most simplistic standpoint, SCSI-2 can be thought of as the merging of a refined SCSI hardware specification and a more tightly defined CCS 4B command set into a single specification. SCSI-2 features several interface features and optional commands. SCSI-2 does not mean Fast or Wide, although both interface implementations are covered in the specification. Also there are many commands, such as the Copy command, which are not required for compliance to the specification.

SCSI-2 - Lots of options

As noted above, in addition to the 1.5MB/sec asynchronous and 5MB/sec synchronous transfer rates, the SCSI-2 specification also defines options for 10MB/sec and 20MB/sec SCSI. These are commonly referred to as fast and wide SCSI. Fast SCSI at 10MB/sec uses new SCSI chip sets which Sun first implemented on the SPARCstation 10. The bus is still 8-bits wide but with these faster chips, the bus can accommodate 10MB/sec, 5MB/sec and asynchronous devices. Both single-ended and differential versions of Fast SCSI are available. During initialization, the host adapter determines the maximum transfer rate of each of the devices attached to the SCSI bus through a negotiation process with each target. Sun's single-ended Fast SCSI adheres to the industry standard and allows Sun customers to mix 5MB/sec devices and 10MB/sec devices on the same bus. Single ended 10MB/sec SBus host adapters are also available for workstation and server SCSI expansion.



In addition to single ended SCSI, the specification has always defined differential SCSI to increase noise immunity on the SCSI bus and allow longer cable lengths. Differential SCSI supports cable lengths of up to 25 meters compared to 6 meters for single ended SCSI. Differential SCSI uses different drivers and receivers and references the signals to each other rather than to ground. Now that SCSI performance has reached levels appropriate for Data Center applications, the differential interface will become increasingly important to provide the connectivity needed by large system configurations. Sun's differential SCSI is now available on Data Centers to provide a lower cost high performance alternative to IPI-2 for large server systems.

Fast wide SCSI requires a different chip set that allows for 16-bit wide data transfers. These chips are now available from a number of sources. These chip sets can double the burst transfer rate on the SCSI bus to 20MB/sec over a 16-bit bus, and can often alternatively support two separate 10MB/sec 8-bit busses. Due to competition from higher performance serial interface designs it is unclear if Fast Wide SCSI will be widely implemented in the industry.

Intelligent Buffering to Improve Performance

Almost all SCSI devices today have complex controllers integrated or embedded on the device. First generation implementations usually required a bridge controller board to convert SCSI commands to device specific commands such as ST-506 or ESDI. Second and third generation devices available today are all embedded designs which means that the SCSI controller is integrated right on the device. These devices also feature built in buffers that provide burst rates at synchronous SCSI speeds rather than disk speeds. For example, the 2.1-GB 5.25-inch differential SCSI drive offered by Sun has a sustained average data rate of around 4.5MB/sec. However, short data transfers are typically burst out of the buffer at about 10MB/sec. Buffering on the drive can also help to hide mechanical latencies associated with disk drives. In a multi drive system with a number of active processes against different drives, the buffer would help in the following manner. The system would send the request to the appropriate drive. The drive would then disconnect from the bus and seek to the area of the disk with the requested data. While this drive is seeking, the system can send a request to another drive and so on. The first drive to reach the requested data will start reading it into the buffer and arbitrate for the SCSI bus. If the bus is busy, the drive can hold the data in the buffer until the bus is free. Because of this, the drive will have the data ready and be able to burst the data across to the host computer as soon as the bus is available. This reduces the impact of rotational latency that would generally be experienced if the drive was not buffered. The ability to burst data at the maximum bus rate, and to overlap operations, is extremely important to multi-disk/multi-user system performance. The high burst rates means that each transfer takes less time on the bus. With the ability to overlap operations and utilize intelligent drive caching, this translates directly into maximum bus utilization and maximum IOPS.

IPI Overview

IPI is a channel architecture that has migrated from mainframe and super minicomputer systems to Unix workstation and server systems. IPI was first integrated into the Unix market by Sun with the introduction of the Sun designed ISP-80 disk controller and the 3MB/sec 1GByte IPI drive in 1989. IPI is also an ANSI standard and is divided into 3 layers. IPI-1 is the physical interface specification. IPI-2 is a device level interface that is used to communicate with the disk drives. IPI-3 is a higher level device independent interface that is used for communication from the host computer to IPI-3 peripherals. These can be devices such as very high performance tape drives, disk drives, or IPI-2 string controllers.

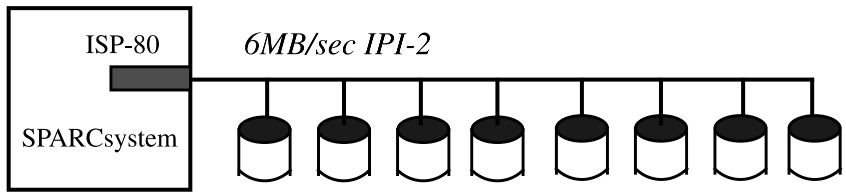
IPI was originally specified as a 16-bit wide interface with channel rates of up to 10MB/sec.The enhanced IPI-1 specification also specifies 32-bit, 64-bit wide busses. Virtually all IPI-2 drives and controllers, including the ISP-80 use the 16-bit wide data path. The next generation of industry standard controllers will support channel rates of 10MB/sec to 25MB/sec.

Originating in the mainframe world, IPI-2 and IPI-3 fully support dual porting allowing a second controller to access a disk in the event of a controller or cpu failure. In fact every IPI-2 disk drive has dual ports installed to support high availability configurations. Sun supports IPI dual porting through On-Line: IPI Dual Port software.

IPI-2 drives are typically high performance drives. The drives were originally based on the highest performance SMD disk drives. Today most IPI-2 drives are being developed around the 5.25" form factor. These drives usually share a common HDA (head disk assembly) with the highest performance SCSI drives.

Today's IPI-2 drives, unlike SCSI drives do not have on-board buffers to assist with reducing mechanical latencies. However, the ANSI IPI committee is adding drive buffering to the IPI-2 specification. Enhanced IPI-2 drives with 10MB/sec buffered burst rates are expected to become available late 1992. These buffers will be simple track buffers designed to help maximize bus utilization. However, unlike SCSI drive buffers which now are often segmentable, predictive, and self optimizing, IPI-2 will still rely on the disk controller to perform this optimization. For example, the Sun ISP-80 controller uses a technique known as RPS (rotational position sensing) to keep track of the position of the heads on all the disks that it is managing.

Each ISP-80 Can Manage up to 8 Disk Drives



By using RPS the controller can ensure that only the disk that is ready to transfer data has possession of the bus. This keeps mechanical latency to a minimum. These techniques are required to achieve good performance on multi-disk systems running applications that support a large number of users simultaneously.

Interface Directions

In mid 1991 with the introduction of the SS600MP series of systems, Sun began the transition from IPI-2 to SCSI by making SCSI the standard disk interface for deskside systems. IPI-2 was still needed for Data Center systems to provide the required level of connectivity and availability features. During the second half of 1992, Sun announced the first phase of completing this transition with the introduction of high capacity differential SCSI drive on Data Center systems. IPI-2 will continue to also be available on the SS690MP Data Center. The transition will be complete with the next generation of Data Centers.

Why Sun Is Moving Away From IPI

IPI was originally selected for some very important reasons. These include:

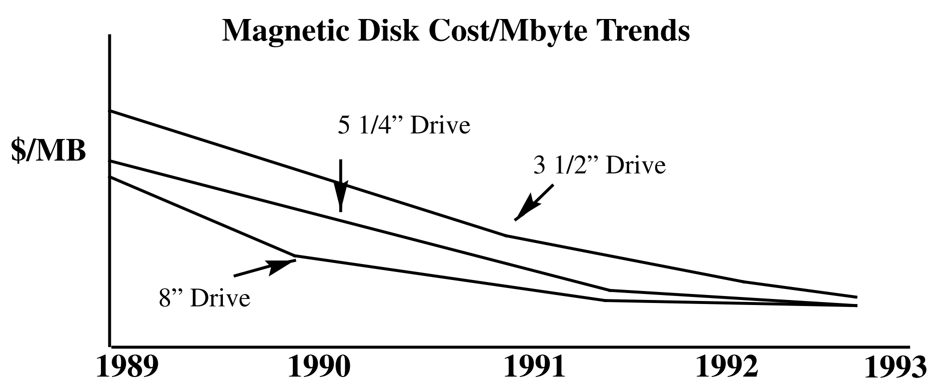
- IPI was much higher performance than SCSI.
- SMD had hit its limit, IPI was needed to balance system performance.
- IPI provided twice the connectivity of SMD or SCSI.

- IPI provided dual porting for higher data availability
- IPI provided a growth path to 25MB/sec and could also provide a very clean channel architecture for large systems with IPI-3.

There were several powerful reasons to use IPI-2 as the high performance interface when Sun introduced it in 1989. However, since then a number of things have changed in the industry which no longer make IPI a very attractive choice.

First of all the performance of SCSI has grown up to meet server requirements. In fact, the two interfaces are now roughly the same in terms of performance and capabilities. In addition, the drives themselves are now identical. Only the interfaces are different, so the performance gap between the SCSI and IPI peripherals has disappeared.

Secondly, the IPI-2 interface will not be standard on sub 5.25" drives. Sun has already announced a 1.05-GByte 3.5" disk drive with a 10MB/sec interface. Due to physical limitations, IPI-2 is not likely to migrate into standard 3.5" disk products. This means that both the availability and cost of small form factor IPI-2 devices will be severely impacted. The following chart shows the current cost/MB trend for different disk diameters.



Just as 8 inch drives have been largely obsoleted by the current generation of 5.25 inch drives it is expected that 3.5 inch drives will replace 5.25 inch drives after one or two more generations. The higher cost and limited availability of non-standard sub 5.25” drives certainly limits the useful life of IPI-2 as a main stream interface. In addition, IPI-2 or IPI-3 require more expensive controllers and only attain the same level of performance offered by today's SCSI solutions.

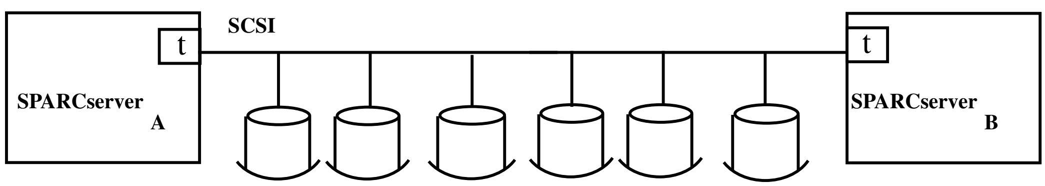
The table below shows a comparison of the three I/O interfaces used by Sun. As you can see, IPI-2 and SCSI-2 are both equally good on most of the performance parameters. IPI-2 specifies transfer rates of up to 25MB/sec. However, this can only be realized with expensive multi-parallel head drives or when enhanced IPI-2 buffered drives become available. In addition, most commercially available IPI-2 solutions available today have a maximum burst transfer rate capability of about 10MB/sec. SCSI also specifies a maximum burst transfer rate of 20MB/sec, but again, the most commonly available devices are either 5MB/sec or 10MB/sec implementations. Both SCSI and IPI-2 can provide scalable random performance which is so important for database and NFS servers. Here again SCSI now can actually perform as well as IPI-2 due to features such as intelligently managed buffers and tagged queueing.

Industry Standard Interface Comparison

Attribute	SMD	IPI-2	SCSI
Transfer Rate		✓	✓
Connectivity		✓	✓
Intelligence			✓
Availability Features	✓	✓	✓
Scalable Performance		✓	✓
Device Independence			✓
Widely supported on a variety of devices			✓
Standard on 3.5" and smaller devices			✓

One of the other recent changes in the industry is also highlighted by the above chart. Availability features have only recently begun to be implemented on SCSI. Although a limited number of dual ported SCSI drives may be produced, the most common method of providing a level of data availability comparable to IPI-2 dual port will be through multi-initiator SCSI. Just like with dual ported IPI drives, multi-initiator SCSI allows an alternate CPU to access the data if the primary CPU should fail. The difference is that with multi-initiator SCSI, both of the hosts are sharing the same data path to the drives.

Multi-initiator Cabling Scheme



Multi-initiator SCSI requires changes to the driver as well as new higher level software. The need for this type of higher availability is just now becoming a reality because SCSI is now powerful enough to provide the ideal interface solution for large Data Center configurations.

There are other big wins for the SCSI interface. One is that SCSI is a widely used interface, as opposed to IPI-2 which is disk specific. This makes programmers and users lives easier by only needing to understand a single peripheral interface. Another is that SCSI will continue to be a standard interface on 3.5″ and 2.5″ peripherals. Today we are seeing 1GByte 3.5″ disks. It won’t be very long before 2.5″ drives will have sufficient capacity to be the standard disk offering for most desktop systems. SCSI will continue to be available on these products extending the life of the interface. A third benefit of SCSI is the lower cost of the interface due to high volume.

Future I/O Interfaces

SCSI is replacing IPI on high end systems and is being be used across Sun's entire product line for disk systems. But will SCSI always be the interface of choice for all products? The answer to this question is no. Just like SMD and IPI-2 had limited life spans, SCSI will also be replaced, at least at the physical layer, for many applications. Drive manufacturers are looking at very small serial interfaces for the 2.5" and smaller diameter drives they are working on. With such an interface running at 25MB/sec or more, very low cost rings of drives could be built. Implementations such as this may be common by the middle of the decade. Other serial channels are also being investigated that promise to provide much higher bandwidth and connectivity than is possible with either SCSI or IPI. The likely protocols that will be run on these channels are SCSI and IPI-3 for I/O devices, but these same serial channels may also find other applications as well.

Sun Microsystems, Inc.
2550 Garcia Avenue
Mountain View, CA 94043
415 960-1300
FAX 415 969-9131

For U.S. Sales Office locations, call:
800 821-4643
In California:
800 821-4642

Australia: (02) 413 2666
Belgium: 32-2-759 5925
Canada: 416 477-6745
Finland: 358-0-502 27 00
France: (1) 30 67 50 00
Germany: (0) 89-46 00 8-0
Hong Kong: 852 802 4188
Italy: 039 60551
Japan: (03) 221-7021
Korea: 822-563-8700
Latin America: 415 688-9464
The Netherlands: 033 501234
New Zealand: (04) 499 2344
Nordic Countries: +46 (0) 8 623 90 00
PRC: 861-831-5568
Singapore: 224 3388
Spain: (91) 5551648
Switzerland: (1) 825 71 11
Taiwan: 2-514-0567
UK: 0276 20444

Elsewhere in the world,
call Corporate Headquarters:
415 960-1300
Intercontinental Sales: 415 688-9000