

# **Fibre Channel: The Preferred Performance Path**

## **Why people are choosing Fibre Channel**

A clear trend in realm of high-end storage interface technology is the growing adoption of Fibre Channel-Arbitrated Loop (FC-AL) technology. Discussions with customers have revealed a consistent preference for Fibre Channel as the high-end serial interface-of-choice, and while each customer has unique aspects to his requirements, some prevalent themes emerge as reasons for that preference.

It might seem that performance would be the biggest reason for preferring Fibre Channel. Certainly, its 100 MB/s bandwidth gets a lot of attention. In fact, architectural considerations are even more significant. First, we will look at where performance sits today, then the benefits of Fibre Channel architecture will be discussed. Finally, a couple of additional aspects of Fibre Channel that realize even more benefits for the user will be described.

## **Performance**

FC-AL shows big improvements over parallel SCSI for transaction processing. The significantly-reduced overhead supports much longer strings per host port. A big advantage for FC-AL is that the same port developed for transaction processing could serve equally well for high data rate applications. High data rate applications are increasingly becoming mainstream requirements, so a traditional workstation configuration is expected to be usable for video and the like. FC-AL allows the workstation manufacturer to offer one hardware platform with the right performance characteristics for both application segments. This serves both the vendor and the end user well by getting the most value from a single investment. The workstation purchaser can buy a system with FC-AL and not have to worry if he will have sufficient performance if he later wants to add a video application.

## **Architecture**

There are six fundamental aspects of FC-AL architecture that yield distinct advantages over other interface technologies:

1. Absolute addressing
2. Multiple device failure support on FC-AL
3. No protocol limit on FC-AL transmission distance
4. Any point-to-any point transfer capabilities
5. Parallel SCSI SCA compatibility on FC-AL
6. FC is a network, as well as storage, interface standard.

### **1. Absolute addressing versus no fixed address**

FC-AL offers the user the choice of two absolute addressing methods. All drives support both methods, and the subsystem developer can choose which he prefers. With FC-AL, no jumpers or switches are required on the drive. The first method is a worldwide unique address built into every FC device. The second method is an absolute address based on the physical location into which the drive is plugged. In neither case does the drive address ever change during operation. If an I/O request is queued in a peripheral it can be executed and the results returned independently of

any other device's status.

## **2. Multiple device failure support.**

FC-AL can tolerate any number of device failures. This has important implications in a several areas:

### **Low-end configurations: single port subsystems**

Fibre Channel can support fully hot-pluggable RAID on a single loop. It would be not fault tolerant if the loop broke, but it could provide complete protection against loss of data, including the extra protection of a hot spare or even RAID "6" dual level parity protection.

Since most systems today do not support dual controller and dual paths to common storage, this topology would be consistent with the majority of today's SCSI based arrays.

### **Very high availability subsystems: fault tolerance and mirroring**

With the increased dependence on computer systems, companies have invested in more sophisticated hardware to insure the highest level of availability. Disc arrays and mirrored subsystems have proven to be effective in making business data available even after a device failure. With mirrored - or shadowed - storage there is a duplicate device for every data drive. If any should fail its mate would allow operations to continue. Note that several drives could fail, as long as not both a primary and its copy became unavailable, and the system could continue to run.

A corollary technique with disc arrays is the use of multiple parity levels, which may well become more popular as drive strings get longer. This enables operation in the event of more than one device failing.

### **Non-disc devices on the same interface**

Another value to allowing multiple devices failing without interrupting the subsystem is to make practical the attachment of non-disc peripherals on the same interface as on-line disc storage. With FC-AL, no device failing has any effect on the availability of any other.

### **Partially-populated subsystems**

It is often desirable to ship a cabinet with only some of the drives actually installed. The customer can later add drives as he needs them. With FC-AL, this is simple; any number of drives can be missing. What is more, any number of drives can be added during operation as easily as a single drive can be added.

## **3. Transmission distance**

One of the intriguing features of FC-AL is that it can connect to peripherals up to 10 km away. This makes possible offering a new level of system reliability where on-line storage can be instantly mirrored at a remote site. This is particularly attractive to companies and organizations that are critically dependent on their computer systems. If the primary system should fail, such as in a building fire, the remote site would be

immediately ready to take over processing; the data on the remote storage being an always up to date copy of the primary systems.

#### **4. Non adjacent communications**

An FC-AL device is not limited in only being able to communicate with its immediate neighboring devices. This is useful when managing larger configurations. Assume that a long string of drives is attached to dual controllers over dual loops. It is common practice to isolate some subset of the devices on one controller or the other to balance or otherwise manage the workload. For example, with IPI-2, SMD or dual port parallel SCSI in mainframe configurations, it was very common to have one adapter be the primary controller for some drives and the other for the remaining. FC-AL maintains and builds on this capability.

#### **5. Parallel SCSI mechanical compatibility**

Fibre Channel disc attachment uses a 40 pin version of the parallel SCSI SCA blindmate connector. FC-AL has built-in accommodation for hot plugging, including elimination of the power surge and arcing problems that plague power supply design in arrays. Particularly attractive to developers of parallel SCSI is the fact the drive can plug directly onto the back plane with no extra electronics in the drive chassis. This reduces the drive hot plugging problem to little more than a simple mechanical housing or pair of guide rails. Moreover, Fibre Channel's use of the SCA connector means that a subsystem designed for parallel SCA can be adapted to Fibre Channel with no significant mechanical changes.

#### **Other Fibre Channel Advantages**

In addition to the performance and architectural features described above, two more benefits of Fibre Channel are worth noting: better support for the new SCSI disc array commands and network compatibility.

#### **Disc array command support**

A fundamental concept behind the definition of FC-AL is that it improve disc array design and performance. Previous sections illustrated how Fibre Channel disc attachment streamlined disc array mechanics and improved the flexibility of managing strings of drives. To improve disc array performance several new SCSI commands have been proposed, which involve the drives participating in performing RAID 5 XOR functions.

These commands were developed in large part expressly to be used on Fibre Channel and are already on FC-AL drives. The architectural differences described above have a profound effect on the usability of these array commands on the two interfaces and illustrate again the advantages of Fibre Channel.

A traditional RAID 5 Read-Modify-Write sequence involves six steps:

1. Read the old data
2. Write the new data
3. Read the old parity
4. XOR the old data and the new data
5. XOR the result of step #4 and the old parity
6. Write the result of step #5 as the new parity

This sequence involves a considerable amount of controller activity in addition to four bus transfers - one reason why some RAID 5 controllers seem to be somewhat slow doing short write operations. A pair of new XOR commands will greatly simplify the process.

Using a new SCSI command, XDWRITE, the array controller can tell the old data drive to:

1. Read the old data
2. XOR the old data and the new data
3. Send another new SCSI command, XPWRITE, to the parity drive
4. Write the new data

The XPWRITE command will cause the parity drive to:

1. Read the old parity
2. XOR the result of step #2 in the XDWRITE command with the old parity
3. Write the new parity

This has reduced the number of bus transfers from four to two and number of tasks the controller must manage from six steps to one!

In addition there are new commands for rebuilding a drive that has been replaced (REBUILD) and constructing data for a drive that is not available (BUILD).

### **Network compatibility**

It was described earlier that a workstation manufacturer could offer to his customers a valuable investment in a Fibre Channel interface as could support both transaction and high data rate applications. In fact, that interface has an even bigger advantage:

An FC-AL disc interface can also serve as a network connection. In high data rate applications especially, the network bandwidth tends to be one of the most severe limitations on performance. Fibre Channel not only offers 100 MB/s disc attachment, the same connection can connect systems in a 100 MB/s network. Fibre Channel started its existence as a definition for network attachment. FC-AL is completely consistent with the original standard and the same physical facility used to connect discs to multiple systems can support TCP/IP between those systems. Not only is this much faster than any network alternative today, it is essentially like connecting workstations via a virtual backplane and enables separate system to collaborate on multiprocessing applications to a degree that has not been possible until now because there has not been a network interface fast enough to support it.

As an example, assume a workstation has a dual port Fibre Channel interface built in. One port could be used for disc attachment the other for a network connection. For high data rate collaborations like studio video editing and image processing, a Fibre Channel network of workstations sharing 100 MB/s paths to storage and to each other provide more than enough bandwidth to have several users working on the same application and sharing data. This points out even more sharply the superior value Fibre Channel offers by giving to the user so many options for making the most of his hardware investment for even the most demanding application requirement.

Going along with Fibre Channel as a network connection is its wide acceptance as a system to box connection. Fibre Channel is both a controller-to-drive and system-to-controller interface.

### **New Fibre Channel Functionality**

Recently, Seagate has been a leader in extending the capabilities of Fibre Channel technology for even greater support for new storage architectures, including Storage Networking (SN). Two of the most key efforts include Public Loop Support (PLS) in Fibre Channel drives as well as the extension of the original Fibre Channel spec. to multi-gigabit transmission rates.

### **Public Loop Support**

PLS enables disc drives to be directly attached to the Fibre Channel fabric, which is commonly composed of switches and hubs. The technology is key to the deployment of large-scale Storage Networks because it allows disc drives to interact directly with the network without an intermediary. This can lower overall system cost and enhance system performance significantly. Seagate's new Cheetah and Barracuda disc drives all ship with onboard Public Loop Support.

### **Multi-Gigabit Fibre Channel**

Seagate is also working with other leading Fibre Channel components vendors to develop multi-gigabit Fibre Channel support. A multi-fold increase in bandwidth addresses the emerging needs of computer system OEMs to build arbitrated loops with higher bandwidth, and to build topologies capable of cascading several 1 Gbit/sec loops together via high speed switches that are interconnected by multi-gigabit links. The enhanced data rate implementation follows standards defined in the original Fibre Channel specification. Multi-gigabit Fibre Channel products will seamlessly interoperate and co-exist with today's 1 Gbit/sec products, as well as with other vendors' future multi-gigabit products. Multi-gigabit Fibre Channel products are expected to begin emerging in 2000.

### **Implementing Fibre Channel disc subsystems**

The basic elements of a Fibre Channel disc subsystem are similar to a parallel SCSI configuration: host adapter, drives, and an optional separate cabinet. The biggest difference is that, while most SCSI subsystems are connected by a traditional ribbon cable, Fibre Channel drives will only plug into a backplane, exactly like SCSI using the SCA connector. Compared to an SCA implementation of parallel SCSI, the principal change going to FC-AL is the increase flexibility in cabling.

Assuming the host adapter is a standard S-Bus or PCI bus card, it will have cable from it to the external cabinet or an internal backplane. Of course those cables will look quite a bit different from ribbon cables.

Some of the possibilities for cabling include:

1. Shielded miniature coaxial cables: 10 m
2. Twinaxial cables: 30 m

3. RG 59 COAX: 30 m
4. Multimode fibre optics: 2 km
5. Single-mode fibre optic: 10 km

A simple subsystem might look quite a bit like a parallel SCSI based system, while larger systems will probably make use of dual porting for full fault tolerant storage:

This Port Bypass Circuit is an active component designed into the FC-AL interface to accommodate the insertion, removal or absence of a device without breaking the loop. The peripheral actually controls this active component and can take itself off the loop as a result of any of several conditions:

1. It is not powered on.
  2. In a self test it has detected a problem.
  3. The host has ordered the device to take itself off the loop.
- (The host can also instruct a drive to put itself back on the loop.)

A disadvantage of the PBC is that it puts an active component on the backplane. An elegant solution to this makes the PBC's also hot pluggable by putting them on a card. In the following diagram the PBC's for each loop are accumulated on a board that can be hot plugged into the backplane. This makes a completely fault tolerant subsystem with a truly passive backplane. If a PBC should fail, the card containing the failed component can be hot swapped while the drives continue to run across the other loop. After the card has been replaced, the first loop would again be up and running, with no interruption in data availability.

Of course, the PBC daughter card can also be a controller; it need not be a separate card or dedicated to the PBC's. (Interestingly, a standard parallel SCSI 80 pin SCA connector has enough pins to support the PBC's for eight drive configuration shown above and makes an effective and economical connector for a card of PBC's.)

Many controller companies have mentioned that the way to make large configurations of multiple cabinets more flexible will be to use a hub that make loops physically star configurations. A hub is essentially a simple-minded switch that preserves the loop in the event of a cabinet or system becoming unavailable. These will not only allow for some elements on the loop to be powered down or off-line, but also add:

1. Ability to add cabinets during operations with no down time
2. Offer a convenient point for conversion to fibre optics
3. Enable multiple local loops to be connected together into a larger network

Taking advantage of the XOR commands described above will lead to products that dramatically change the economic and performance characteristics of disc arrays. Instead of needing expensive and complex parallel controllers - and thus dictating to the user the parity amortization - the array supplier can offer an array built on a loop of Fibre Channel drives that lets the user pick his parity coverage: 5 + 1, 18 + 1, or even 32 + 1. Since the XOR engines are each time a drive is added to an array, the ability of the subsystem to perform XOR's increases with the number of drives.

These, then are the essential elements of a Fibre Channel subsystem.

The whole idea behind Fibre Channel has been to take standardization several steps beyond where parallel SCSI left off. Fibre Channel makes the task of doing what users want to do with storage, simpler, more standard, and much less expensive.

By Dave Anderson and Tyson Heyn  
November, 1998

Copyright 1998, Seagate Technology, Inc. All rights reserved.