A Digital Video Disk Array Primer

By Charles F. McConathy

This paper reviews data storage technology and reveals some of the mysteries associated with video disk arrays as related to digital video editing and audio applications. Subjects such as embedded servo, thermal recalibration, mode page settings, head technology, UltraSCSI, Fibre Channel, and 10,000 RPM disk drives are discussed.

With the advent of desktop digital video and audio editing applications, the need for high-capacity storage systems capable of streaming data without interruption has steadily increased. Typically, such systems employ computer disk drives, which are becoming as important to video production as videotape recorders (VTRs). Although tape transports represent sophisticated technology, disk drives are even more advanced, involving extremely small parts hermetically sealed in hand-sized assemblies. Video disk arrays become essential to nonlinear editing systems and digital audio workstations, and are increasingly finding their way into other audio and video recording applications.

Storage vendors have responded to the demand for audio/video disk drives with a multitude of advanced technologies, including embedded servo, magneto-resistive (MR) heads, partial response maximum likelihood (PRML), redundant array of inexpensive drives (RAID), UltraSCSI, Fibre Channel Arbitrated-Loop, FireWire, fast/wide small computer systems interface (SCSI) host adapters, as well as improved connectors, data cables, and terminators. This article is written to unveil some of the mysteries of storage systems and to help users decide which of the above features will be of value.

Presented at the 31st SMPTE Advanced Motion Imaging Conference (paper no. 31-20), in New York, N.Y. February 6-8, 1997. Charles F. McConathy is president of ProMax Systems, Inc., Irvine, CA 92618. An unedited version of this paper appears in *The Age of Compression: Nonlinear Editing, Digital Broadcasting, and Other Wonders*, SMPTE, 1997.
Copyright © 1998 by SMPTE.

UltraSCSI Drives

SCSI hard disk drives are built using various numbers of magnetic coated disks that currently rotate at 5,400, 7,200, and soon 10,000 RPM. Each disk has two recording surfaces. The disks are logically divided into concentric circles called tracks. A set of tracks at a given position on the disks is known as a cylinder. The number of cylinders is the same as the number of tracks across the disks. Tracks are divided into varying numbers of sectors. The number of sectors varies according to the position of the track located on the disk. This technology is known as zone bit recording (ZBR).

Data are written to one track of a given surface at a time. If more space is needed, a head switch takes place and data then are written to the next surface. Head switches continue down the cylinder until the last track in the cylinder is filled. If additional space is needed, the heads are stepped to a new cylinder and the process of head switching and track stepping continues until the file is completed. ZBR technology has greatly increased the ability to store data on a given set of disks. Previous to ZBR, all tracks had the same number of sectors, which wasted storage space.

Using ZBR, the outside tracks contain more sectors than the inside tracks since there is more physical space on the outside tracks. The transfer rate is greater from the outside since a greater amount of data is available from a single rotation of the disks. The first partition created uses the outside tracks and moves to inside tracks as the drive or drives are fully partitioned. It is a good idea to use a naming convention for partitions so that one knows in which order they were created for performance reasons. Since the outside partition performs best, inside partitions can be used for less intensive functions such as copying files from a working partition, which will quickly defragment them.

To further increase disk capacities and transfer rates, manufacturers of disk drives have begun to employ MR heads. MR heads allow higher areal density. Areal density, which has been doubling every 18 months, is simply the number of tracks of data per inch (TPI) times the bits of data per inch (BPI). Higher areal densities increase the chance of read/write errors. Try to picture ten disks mounted on a common spindle that are spinning about 120 MPH with 20 tiny MR heads hovering a few microns over 3,711 finely spaced tracks. Even the slightest thermal change can cause the heads to be off track, creating read/write errors, which in turn can cause the drive(s) to spin extra times in a retry process.

If the storage system does not recover within a couple of retries, the disk drive's buffer is likely to empty, and dropped audio or video frames will most likely take place. To compensate for thermal changes, some disk drives take time out to recalibrate (TCAL). Simply put, the read/write heads are moved to find the outer and innermost tracks, and the stepper is then calibrated to match thermal changes accordingly. TCAL was a major issue for drive makers to overcome before data storage systems could be made to stream uninterrupted data at rates needed for audio/video applications.

Disk drives that employ embedded servo technology never take time out to TCAL. Information is embedded across the tracks of all the disks so that the read/write heads know their exact location by continuously compensating for temperature changes.

Picking disk drives that use embedded servo technology is key to building disk arrays that can stream audio/video data without interruptions. In addition to staying on track, disk drives with embedded servo tend to be more reliable.

PRML

PRML helps disk drives to handle higher areal densities. As areal density increases, the possibility of interference from surrounding tracks also increases. PRML helps to filter wanted data with minimum encoding resulting in increased data transfer rates.

5,400, 7,200 Vs. 10,200 RPM Spindle Disk Drives

Disk drives that rotate at 7,200 or 10,000 RPM move data under the heads more quickly than slower spinning drives. Key factors in the faster RPM drives are higher transfer rates and lower latency. Some audio/video applications create multiple files, causing the heads to seek frequently, which results in latency. Drives spinning at 5,400 RPM with suitable transfer rates work fine with most single-stream audio/video applications. Some digital editing software offers the user the option to lay audio and video on different drives. In this case, latency is not as much an issue since two sets of heads will be used. To accomplish this, audio and video must be on different drives, not different partitions of the same drive.

A problem with the faster RPM drives is the heat they produce due to the extra power required. Later versions of the Barracuda 7200 RPM do not run as hot as early versions. A drive that runs hot not only shortens its useful life, but cannot stay on track which often results in dropped video frames.

The environment in the chamber of a disk drive is highly controlled. Makers of disk drives are careful to seal the drive, using inert material and precise amounts of humidity. The inert material is used to keep gases from forming and condensation from turning into "acid rain," which can create bad blocks on the disks. Too much humidity in the drive chamber can cause head stiction, and too little can result in destructive head landings.

One other factor that heat introduces is lowering relative humidity in the drive's chamber. The read/write heads are much like lightweight sail planes that are designed to fly a few microns above the disks. During a seek in low humidity, they tend to go into a yaw position more easily. Thicker air helps them to get on track, eliminating errors and retries. Adequate air management inside computers and cabinets is essential to keep hard disk drives within temperature specifications, allowing them to stay on track and ensuring their reliability. Hot disk drives simply cannot produce reliable streams of data needed for audio/video applications.

Optimizing Mode Pages for Audio/Video Applications

Mode Page settings are one of the deeper mysteries of disk drives and are little understood by most users. Mode Pages are parameters that are stored in memory on the controller of the drive. Factory default Mode Pages are generally set for short random-access transfers used by database applications. With proper Mode Page settings you can improve the ability of a disk drive or a disk array to stream long sequential audio/video data by as much as 40%.

Unless a user is experienced, it is better to have a trained person change Mode Page settings. Applications such as Remus Control (from Adaptec's Trillium Division), which runs on the Macintosh, make changing the parameters relatively easy. Mode Page changes made using a Macintosh are effective on other platforms such as Windows 95 or newer technology systems. These changes do not affect data, and there is no need to reformat. The changes are effective immediately and remain stored in memory on the drive until changed by software.

Some SCSI utilities will alter Mode Page settings should a drive be reformatted. These changes may or may not be suitable for streaming audio/video data. It is a good idea to run a benchmark program, such as the Remus Performance Application, to prove the new settings. Do this before you make any changes and again after each set of changes. Note the changes so you can return the values just in case they do not improve performance. In general, Mode Pages should be set as follows:

Mode Page 1

Set Read and Write Retry counts to 2. In this case the drive will try twice and then move on. The error might only represent a couple of pixels. It is better to miss two pixels rather than Retry 19 times, which could result in several frames being dropped. Disable AWRE (Automatic Write Reallocation Enabled), ARRE (Automatic Read Reallocation Enabled), and RC (Read Continuous). Enable PER (Post Error Recovery) and TB (Transfer Block).

Mode Page 2

Buffer Full and Buffer Empty Ratios should be set to 0 and 0, or 0 and 4, in that order. The factory settings for these are often higher like 127 over 127. By lowering these settings, you can generally improve Reads, which can improve video playback as much as 20%. If after running a benchmark application the Read rates do not improve with 0 and 0, then move the numbers to 0 and 4, or 4 and 12, or maybe something like 42 and 94. Experimenting with these two values can prove to be beneficial.

Mode Page 8

WCE should be enable or set. Set Minimum Prefetch to 0. Set Maximum Prefetch to 255, and the Ceiling to 65,535. Set Cache Segments to 1.

Standard, Wide, or UltraWide SCSI drives—Which is Best?

Eight-bit, narrow or fast SCSI-2 drives burst at 10 Mbytes/sec. A single narrow drive can sustain about 4 Mbytes/sec. Sixteen-bit, wide SCSI-2 drives burst at 20 Mbytes/sec. A single wide drive can sustain about 7 Mbytes/sec. Ultrawide SCSI-3 drives burst at 40 Mbytes/sec. A single ultrawide drive can sustain about 12 Mbytes/sec. Wide SCSI offers higher transfer rates, so it is the best interface for high-end video applications. A pair of ultrawide SCSI drives connected to an ultrawide SCSI host adapter and striped as RAID-0 can sustain as much as 28 Mbytes/sec.

As a rule of thumb, generally under QuickTime due to overhead, a user will need about double the sustained transfer rate to maintain a given video capture and playback rate. For example, in Adobe Premiere, if you set to maintain 4.5 Mbytes/sec, you will need a disk array that can sustain 9 Mbytes/sec without audio. If audio is added at 16-bit stereo at 44.1 MHz, the need for data will increase to about 11 Mbytes/sec. With Media 100 xs to produce 2:1 compression digital video (30 frames/sec x 300 Kbytes/frame = 9 Mbytes/sec), a pair of ultrawide drives striped as RAID-0 that can maintain at least 16 Mbytes/sec will be required. As applications that employ dual streams of video and/or digital video effects are developed, the need for higher data rates will steadily increase.

UltraSCSI is Here

UltraSCSI disk drives are readily available. UltraSCSI, wide-40 drives burst at 40 Mbytes/sec and will sustain about 50% more than wide-20 drives. The reason for the higher sustained rate is the employment of read/write technology with internal transfer rates as high as 120 Mbits/sec. The internal transfer rates of wide-20 drives topped out around 80 Mbits/sec. A single UltraSCSI, wide-40 drive can sustain about 12 Mbytes/sec. A pair of UltraSCSI, wide-40 disk drives striped as RAID-0 will sustain as high as 28 Mbytes/sec. In order to get UltraSCSI transfer rates, the disk drives must be attached to an UltraSCSI host adapter, such as the PowerDomain 2940UW by Adaptec. A dual channel, single chip (no bridge chip), UltraSCSI host adapter will be available soon. These new single chip cards will not have the overhead that the current three chip host adapters have since they do not require two SCSI chips and a bridge chip to accomplish dual channels of SCSI.

SCSI Host Adapters

Fast/wide SCSI host adapters are essential to eliminate the bottleneck of the native SCSI ports of Power Macintosh computers. These cards have one internal and one external 16-bit, wide SCSI, 68-pin connector, and a single internal, 50-pin, fast SCSI-2 connector. The cards also autoterminate and are capable of using fast SCSI-2, wide SCSI-2, and UltraSCSI on the same bus.

For video applications it is better to create a SCSI bus with the same type of drives, that is, all fast or all wide. It is best not to mix tape backup and scanners on a fast/wide SCSI bus with high-performance video disk arrays. If fast SCSI-2, 50-pin drives are used for audio, they need to be connected to a different SCSI host adapter or to the native SCSI port. It is important to note that 8-bit drives use 50-pin connectors and wide 16-bit SCSI drives utilize 68-pin. If for some reason it is necessary to daisy chain a 50-pin drive with 68-pin drive, the 68-pin drive must be placed first in the chain. This way the 68-pin drive gets full benefit of wide performance provided it is connected to a fast/wide SCSI host adapter. In this case termination can be an issue since the terminator on the 50-pin drive only terminates 50 of the 68 lines to the wide drive(s).

How Many Drives Can Be Connected?

Seven fast-10 drives and 14 wide-20 drives can be connected to the same SCSI bus. SCSI IDs available for fast SCSI are "0" to "7." SCSI IDs available for wide SCSI drives are "0" to "15." In each case the host requires one SCSI ID, which is generally set at "7." One thing to note is that each SCSI port or host adapter allows a new set of identical SCSI IDs. Therefore there is no conflict with native SCSI ports and host adapters. With UltraSCSI wide-40 drives, the maximum drives on a single bus is eight.

Daisy Chaining Standard and Wide Drives

Some people ask, is it possible to connect a wide 68-pin SCSI drive to a standard 25 or 50-pin SCSI port? The answer is yes. A 68-pin wide external drive can be connected to a DB25 SCSI port by using an HD68 to DB25 cable. In this case the drive will perform at standard or fast SCSI speeds. HD68 to C50 cables are also available so that wide external drives can be daisy chained with standard 50-pin SCSI systems as well. In this case it is best to place the wide drive last and use a 68-pin active terminator so that all lines are terminated.

Cables, Connectors, and SCSI Terminators

Since audio/video demands steady streams of data, it is imperative that SCSI cable and termination rules be followed as closely as possible. Although the SCSI specifications state that standard 50-pin SCSI buses should not exceed 20 ft, wide-20 SCSI, 10 ft, and UltraSCSI wide-40, 5 ft, it is best to keep a SCSI bus as short as possible. UltraSCSI wide-40 must be kept to 5 ft or shorter to work reliably. Wide-20 SCSI works well out to about 8 ft and standard 50-pin SCSI to 12 ft.

A good rule to follow is to have the least amount of connectors and cable types possible in the chain and to use active termination. This helps to cut reflections and noise on the bus, which can be a source of errors resulting in dropped video frames. Impedance matching of cables helps as well. Another good rule to follow is to use the same brand of SCSI cables throughout the chain as far as possible. Transitions between connectors and different kinds of cables cause reflections, which can result in corrupted data, Type 11 system errors, and system hangs.

Sixty-eight-pin round system data cables used with wide SCSI drives are available with 105Ω impedance. Flat ribbon cable insulated with Tempflex Teflon for use inside of drive enclosures are available with 95Ω impedance. This is the best impedance match to be found, and, when implemented in a SCSI bus, it drops SCSI termination issues to a minimum. Some round data cables run as high as 130Ω and flat ribbon cables as low as 70Ω, a poor match often causing reflection problems resulting in errors, as well as a source of directory corruption.

The use of thumb screws instead of latches creates a more stable SCSI bus. Thumb screws pull connectors tight. This lessens the chance of noise and reduces the chance of errors. Also, latches are easy to knock loose and thus can be a source of errors.

Data cables and video cables that run near monitors can be a source of dropped frames and loss of quality. All cables should be properly shielded and routed to stay clear of magnetic fields. Serial cables that control video decks should also be shielded and have been known to cause system hangs.

Multiple Drive Enclosures

Enclosures that house multiple SCSI drives often use a printed circuit board as a backplane to connect the drives. This backplane shortens the SCSI bus and if designed properly is an excellent way to cleanly connect multiple drives, especially UltraSCSI.

One thing to be mindful of is that multiple drive systems often use canisters to mount the drives. These cannisters most likely will have a short ribbon cable that connects the drive in the cannister to a connector that mates with the backplane. The use of several of these short SCSI stubs on one bus can be a source of data errors. If the cannisters are not properly designed and the drive enclosures do not have sufficient air flow, problems due to heat could also be encountered. Other problems with enclosures with backplanes include the fact that they may not handle UltraSCSI wide drives unless they have been specifically engineered to handle the higher transfer rates.

Which RAID Level to Use for Video

Because video needs all the bandwidth it can get, most desktop video storage systems use a software-created RAID Level 0 disk array. The arrays are generally made up of striped pairs of like drives with as many as four or five pairs on a single chain. If additional drives are needed, then use a dual-channel or a second single-channel SCSI host adapter. RAID Level 1 or mirroring is not practical since Write times are slow and require twice the disk space. RAID Level 3 or 5, if implemented with a RAID hardware controller, can yield data rates suitable for certain types of video editing.

Fibre Channel Arbitrated-Loop

Fibre Channel Arbitrated-Loop (FC-AL) will be a strong contender for high-end storage systems for use with multiple nonlinear editing workstations, adding the ability of each station to share data stored on a local or remote FC-AL disk farm. FC-AL offers enough bandwidth to produce 300 Kbytes/frame of video between multiple workstations spread over a wide area. Fibre Channel has no ID or termination issues like SCSI to contend with, which will make it easy to install. Copper cable can have runs as long as 30 m between FC-AL devices while fiber cable can be kilometers. FC-AL should become more cost effective approaching the price of SCSI. Cost of FC-AL host adapters, hubs, and cables will drop quickly once FC-AL starts to hit critical mass.

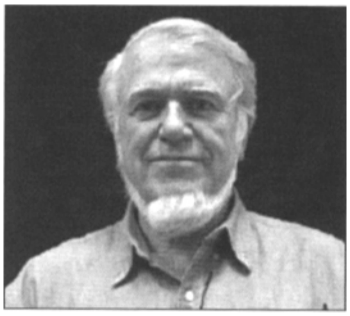
FireWire (P-1394)

FireWire peripherals and host adapters have begun to appear, such as the Sony VX-1000 digital camera. FireWire is an excellent interface for various peripherals such as digital cameras, DVD, CD-ROM, and video decks, etc. The cost to implement FireWire is low, and it offers guaranteed bandwidth for video applications. Many have asked why not FireWire disk drives? I don't think the drive industry will build FireWire drives until FireWire ports appear on computers. Meanwhile data that comes into a computer via PCI FireWire can be stored on SCSI drives. Some think that FireWire could replace a number of ports on current computers such as desktop bus and enhanced integrated drive electronics (EIDE). FireWire is easy to connect since it too does not have IDs or termination issues like SCSI.

Conclusion

For reliability and high rates of uninterrupted data, pick a hard disk array that does not get hot and employs embedded servo technology. Never try to create a disk array with mismatched pairs of hard disk drives. Use impedance-matched, high-quality shielded data cables and active termination. Keep the SCSI bus as short as possible. Be aware that video and SCSI data cables can be affected by magnetic fields that surround large video monitors. Make sure the Mode Pages have been set to stream data for audio/video.

THE AUTHOR



Charles F. McConathy is president of ProMax Systems, Inc., a vendor of Macintosh-based turn-key digital video editing workstations using solutions from Apple, Media 100, Scitex, Truevision, Adaptec, Fast Electronics, Adobe, Seagate, Sony, and others. McConathy was the founder of MicroNet Technology, a noted vendor of high-performance SCSI data storage systems for Macintosh, SGI, Sun, IBM, and compatible workstations. His past experience includes technical support management, product management, director of marketing, and more than 30 years of pre-press and advertising production. His background has helped him become an expert in overcoming performance bottlenecks in the configuration of computer systems for video, desktop imaging, and graphic design.

He has been a speaker at SMPTE, Seybold, Conceppts, Color Connections, Macintosh Summit, and other conferences in the U.S. and Tokyo.