A Proposal for a Digital Tape Format to Complement Disc-Based Editing

By Alec Cawley

Editing techniques in film and television are changing rapidly. Nonlinear, disc-based systems already enjoy a substantial installed base addressing off-line, news quality, and full post-production requirements. Undoubtedly the role of the VTR in editing and post-production is diminishing, being superseded by more flexible and cost-effective nonlinear systems. This trend suggests a tapeless future, and proposals already exist for disc-based camcorders and magneto-optical (M/O) technology archiving. Technologically exciting, many such schemes capitalize on the inflexibilities and occasional unreliability built into videotape formats by the need to edit on tape.

Starting from the viewpoint that all editing will be performed in a nontape environment, this paper analyzes the true expectations and requirements of a tape format. Concentrating only on tape for acquisition frees the format in many ways. Constant speed and data rate may be parameters of the past. Flying erase heads, full-width erase heads, and frame-accurate servos could be technology of the past. A cruder variable-speed recorder may suffice. A new format capable of handling compressed and full CCIR 601 quality could enjoy wide acceptance.

In today's linear versus nonlinear climate, the debate of tape versus disc continues to dominate the headlines. The debate often ignores the fundamental requirements of television production techniques. Arguably, the weakness of today's videotape formats is that they do not reflect the fact that today's and tomorrow's editing is disc-based.

This paper represents an idealistic view of the role of tape acquisition in future television production. It argues that in today's disc-based editing environment, tape has a clear future, provided that tape editing requirements are eliminated.

Videotape is central to every facet of TV production. Cassettes in various formats fill shelves and vaults and litter the desks of broadcasters, editors, and production houses everywhere. For the past 40 years, every TV professional has used tape, in one way or another, many times every working day. In recent years, domestic VCRs have spread tape, often in untidy heaps, from the professional world into homes and offices everywhere.

The uses of videotape are many. Tape is the basic medium for long-term storage of TV pictures, whether raw footage or edited programs. It is increasingly the medium upon which images are first captured, usually from camera (though film still retains a large minority of adherents). It is the primary method for transport of programs, whether from office to office or from continent to continent, and is also the most widely used medium for editing moving pictures.

Now, just when it seemed that tape was spreading into every corner of life except the bathroom, something has invaded the very citadel of tape usage, the Edit Suite. The interloper is nonlinear, or disc-based, editing. In a nonlinear editing system, all pictures are stored on discs — usually fixed, but sometimes exchangeable disc packs. Discs can be randomly accessed, so editing on disc is simply a matter of playing the frames off the disc in the desired order. This contrasts with tape, which has to be shuttled laboriously backwards and forwards to access different scenes stored on the same tape.

Nonlinear Editing

Nonlinear editing systems are now widely available. Early systems were off-line, where a low-quality, disc-based system is used to create an Edit Decision List (EDL), which then has to be reexecuted (conformed) on a full-quality system, which might be either tape or disc-based. The advantage of such a system is that the lengthy decision-making process is carried out on relatively inexpensive, reduced-quality equipment. The expensive, full-bandwidth on-line equipment is then needed only for a relatively short time for the final conform. Increasing numbers of on-line nonlinear editing systems are now available, operating at full uncompressed CCIR 601 quality. These can be used either for the whole of an edit, or to conform EDLs from off-line nonlinear systems, and perhaps to perform some fine tuning on the final edit. In either case, no tape editing is needed.

Is tape, therefore, dead — or, at least, dying? Some people have suggested that disc-based systems will totally supersede tape, and the familiar tape cartridge will disappear from our offices, to be replaced by the disc pack and the central disc-based Video Server. This can easily be shown to be seriously overstated, simply by counting the number of tape cartridges in any TV facility. Disc-based systems will always be significantly more expensive than tape-based systems. The two technologies use the same underlying magnetic storage mechanism, but discs must achieve a significantly higher packing density because a disc needs random access to the whole of its magnetic storage area in order to achieve the benefits already described. Tape, able to roll its magnetic media up out of the way, is able to take a much more relaxed attitude toward storage density and, therefore, to use much cheaper components. Most

Revised version of a paper entitled "Digital Acquisition Without Compromise — The Missing Link," presented at the 1995 SMPTE Advanced Television and Electronic Imaging Conference in San Francisco (paper no. 29-26) on February 11, 1995.
Alec Cawley is with Quantel Ltd., Newbury, Berkshire, England RG 13 2NE.
An unedited version of this paper appears in *Proceedings of the Advanced Television and Electronic Imaging Conference: New Foundations for Video Technology*, SMPTE, 1995.

discs also have an expensive overhead of drive motors, arms, actuators, etc., which are fixed to the drive or data pack and therefore add to the cost of the cartridge, as opposed to tape, where such components form part of the VCR and so need to be bought only once. (There are exceptions to this, but they tend to be at the lower performance end of the spectrum and therefore offer restricted, or no video capabilities.)

It is therefore inconceivable that disc systems will, within the foreseeable future, be used for holding completed programs on the shelf awaiting a repeat. It is improbable that bulk program matter (drama, light entertainment, etc.) will be held on disc for the weeks between shoot and broadcast. Future news and sports material may well be held on disc for the short time between event and first transmission, but it will almost certainly be transferred to tape for reference and archive purposes. Disc-based cameras are under development, but these have limited applications. There are already conflicts over the cost of keeping raw camera footage available on tape; if the media become 20 to 100 times more expensive, this proves to be totally uneconomical. This means that the camera disc packs would need to be routinely copied off to tape, in order to keep costs down. If you are going to copy to tape, why not record on tape in the first place?

Tape Today

Tape, then, is not dead. Tape editing, however, if not dying, is at least wounded. Nonlinear systems are already dominant for short, high-intensity edits such as commercials and action sequences. In the near future, it is likely that the storage capacity and price of nonlinear systems will expand to the point where it becomes economical to perform more mundane edits in a nonlinear system. This is significant because the requirements of tape-based editing have driven all modern tape formats, whether analog or digital, coded or decoded. Central to this function is the ability to perform an Insert Edit; that is, to replace a number of frames on a tape with an identical number of new frames (usually copies from another tape) without disturbing the frames on either side of the edit. This is quite a strong constraint, which has probably added significantly to the cost of both tape and VCRs, and which has precluded (as shown in a later section) some valuable savings.

Tape Tomorrow

It is therefore time to consider the possibility of a new tape format that is optimized for all the uses to which videotape is used, with the exception of tape editing, and with additional features optimizing it for nonlinear editing.

It should be clear that any new professional tape format must be digital. The problems of analog tapes, with generation loss and drifting setups, are too recent history (or too current a misery) to think of returning to them if it can be avoided. The same logic says that, if the new format uses compression, it must be capable of doing so in a manner that offers negligible picture degradation. A digitally perfect recording mode, based on CCIR 601, is highly desirable.

The first thing to throw out is the fixed tape frame size. In order for a tape system to do an Insert Edit, a fixed frame size is mandatory because the replacement frames may not occupy more space than those they overwrite (or they will overrun those after the edit), nor significantly less (or there will be a gap between the end of the insert and succeeding frames). Once we have dropped this requirement, however, we can allow different frames to occupy different areas. This immediately opens the door to data compression, in either "lossy" or "lossless" forms. There is much room for discussion of different forms of compression, but this is rather a technical matter that is discussed in a later section. First, the operational consequences of variable frame sizes will be discussed.

One consequence of variable frame sizes is that when we record on tape, we do not know how much we are overwriting. Therefore, we effectively endanger all material beyond the point to which we record. At first sight this may seem a rather dramatic and undesirable consequence. However, if we examine how tape is used in real applications outside the edit suite, it will be seen that this is actually how tape is usually used. In a camera shoot, for example, the cameraman may rewind and overwrite a failed take, but he never tries to recycle the bad take much earlier up the tape; the risk of overrunning and destroying valuable material is too great. Similarly, archivists append reference clips onto tapes for the library but do not try to fit new material over earlier footage. A distribution tape may contain several episodes of a drama, several edits of a commercial, or several "stings" for an evening's broadcasting, but these would certainly have been recorded in sequential order.

A tape in the new format will therefore contain a number of different fragments of video, each an indivisible entity recorded at a single pass. These might be individual takes from a camera shoot, segments from a news or magazine program, or whole programs. The term "clip" refers to any of these items. The recorder that created these clips might be a camcorder, a nonlinear editing system, or a simple dubbing system. There is no necessity for a frame-accurate join between adjacent clips. Indeed, it will probably make life considerably easier for the recording system if we demand that a short section of run-in (displayed as black, title, or clock) separate each pair of clips. The making of accurate splices is the province of the nonlinear editing system.

Any tape format needs to be able to record and play in exact real time. In addition, most VCRs are capable of playing at a range of speeds, both faster and slower than real time. Recording has hitherto been regarded as necessarily a real-time affair, and for a camera source, this is obviously still true. Recording camera sources, however, represents a minority of recording operations. Most tapes are produced by editors or simply by dubbing other tapes.

This paper discusses a tape format intended for nonlinear editing. Generally speaking, the discs used in a nonlinear system can record continuous data significantly faster than real time, because they have to have spare capacity to handle the random seeks produced during editing. This is particularly so if there is some mechanism whereby the disc can control the flow of data being recorded from tape, slowing it down if the disc falls behind and speeding it up as it catches up. This process can be eased by a relatively modest and inexpensive video buffer a few frames long, such as is routinely fitted to modern digital video systems.

Before a video sequence can be edited on a nonlinear editing system, the image must be copied onto disc (a process sometimes misleadingly called digitizing). If we can do this faster than real time, there are obvious time savings in loading up the editor, and there are similar savings after the edit is playing out the edited clip (which we have already reviewed on the editing system, so that there is no need to watch it again) to tape. Alternatively, if the tape system is capable of running very slowly, or even in a start/stop mode, a nonlinear editing system could use "spare" bandwidth to record or play out to tape as a background task, at the same time as it is being used for a different editing job.

It is obvious that for simple tape copying (an activity upon which a surprising amount of time and effort is spent), the ability to copy at double speed, for instance, would be a boon. It is ironic that domestic dual-deck VCRs are available with double-speed copying, while this facility is not available to the TV professional.

It is therefore proposed that the new format should be designed so that both recording and playing can be performed at a range of speeds, both slower and faster than real time. As well as the traditional video speed input and output ports, whether analog or digital, we need a new port, the dubbing port. This port carries digital video, but instead of carrying it always at the lock-step synchronized video speed, it allows for variable speed. It also carries extra signal lines, by means of which two devices (a recorder and a player) can send requests to each other to speed up or slow down.

Having obtained our new format, how do we find material on it? We do not always want to play a tape from the start, particularly if (as is usually the case) it has a number of different items on it. There are two traditional methods: scan-by-eye and time code. Scan-by-eye is done simply by putting the tape into a fast shuttle mode and watching the action fly past. This requires that the tape drive be able to read enough information off the tape to produce a recognizable picture while shuttling the tape at high speed. If the shuttle speed is to be high, it places a demanding, and thus expensive, constraint upon the system. It will also probably compromise the way in which data is put onto the tape, which we would like to avoid if possible. If the shuttle speed is low, too much time is wasted while the tape spools by, searching for a clip we want. If a tape contains much similar material (e.g., successive takes from the same scene, or in the same location), it is difficult to decide when we have reached the segment we want.

The alternative is time code, which allows us (usually with the aid of a computer of some sort) to go to the exact frame we want. Time code, however, is not without its disadvantages. Without detailing all the problems involved, examples are tape striping, discontinuous time code, drop-frame versus non-drop-frame, and midnight wraparound. Furthermore, the proposed format has already removed the possibility of time code by allowing (even demanding) intersegment gaps. Traditional time code assumes that a tape consists of a continuous sequence of frames; now we have a discontinuity between different segments on tape.

Even when we have a time code to shuttle to, how did we get it? Nearly always, from a piece of paper that travels with the tape. These pieces of paper vary from smartly printed labels to scrawls on the back of old memos. They may be stuck to the tape (in which case they cannot be read once the tape is inside the VCR) or stuck to nothing (in which case they get lost). If the tape is being used for a compilation, or being reused, the papers become obsolete (or out of date).

Even when we have the right piece of paper, how did the information get onto it? Someone read it, from a location clock or by watching the tape noting significant events, and wrote or typed it for us. We then read it and type it back into another system. Once we gain a certain amount of experience, we stop typing it in incorrectly and speeding to the wrong end of the tape. This is a system crying out for improvement.

What we want is some memory inside the tape cartridge, which can be accessed (both read and written) without having to scan the tape. It does not need to be very large, because we will typically have one entry per clip, and a fairly busy tape might well contain a few tens of clips. This memory will store a clip directory. The minimum data for each entry is a clip title, in a format suited to humans (i.e., text), and a clip start position, in a format that allows the VCR to find the start of the clip (e.g., inches of tape). Essentially, this is the same information as is scrawled on a clapper board on a prepared shoot. Unfortunately, clapper boards are not usually available in news or sports operations. Even when available, the information has to be manually copied (onto the piece of paper we discussed earlier) before editing can start. Furthermore, the clip directory can be edited after an individual recording has ended; for example, to title a dramatic news event or to annotate a take that went wrong but might be retrievable.

With a built-in clip directory, as soon as we put a tape into a VCR, it can present us with a list of clips. Finding a clip is then just a matter of scrolling through the list of clips until we find the one we want, and pressing a seek button to scan to it. While seeking, or playing through the gap between clips, the VCR can display the information from the clip directory (such the current clips, the clip it is seeking, and a countdown clock).

Once we have put a memory for the clip titles into the cassette, it is easy to think of other information that could go into it: clip duration; name of cameraman or editor, director, sound; type of camera or editor; date and place of shoot; copyright information; tape library number; date copy was made, etc. One can imagine other improvements to the organization of a shoot. For example, a production assistant's handheld time-code unit could become a remote titler (connected via an infrared link), titling clips as they are recorded without distracting the cameraman from his job. Of course, when the rushes are transferred into a nonlinear editor, the title and ancillary information can be transferred with them (via the dubbing link discussed earlier), so that the raw rushes are fully titled for the editor when he wants them. Similarly, the editor titles the clip as he records it back to tape. The cart machine then reads the title from the same memory, with no possibility of getting the wrong cartridge, entry point, or duration.

The technology used to implement this built-in memory need not concern

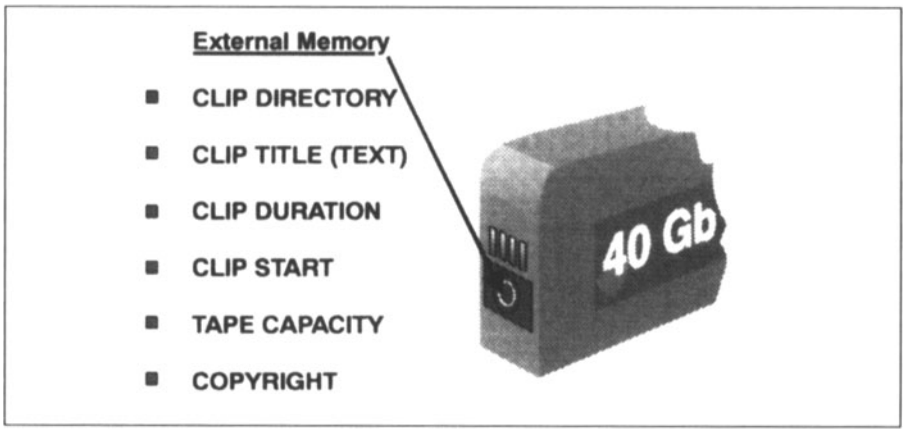

Figure 1. Clip memory.

us at this point; there are a number of possibilities, such as Flash EPROM and Magnetic Stripe (Fig. 1). The memory must be read/write (like the tape it indexes), be capable of being read or written in about one second (perhaps while the tape is threaded or unthreaded), contain at least a few thousands of bytes of data, and be very reliable.

A VCR will be able to read the directory of a tape almost instantly. Of course, we don't always have a VCR on hand when we want to see what is on a tape. Paper labels will therefore always exist. However, it is much easier to go from electronic memory to paper than vice versa. It is not implausible to think that VCRs might well be provided with inexpensive label printers to print out the directory of the cassette as it is recorded.

By specifying that the tape is aimed solely at nonlinear editing systems, we have released the tape from the need to do frame-accurate inserts. By adding the clip directory, we have removed the need to view the tape while shuttling at high speed. We have added the desire to be able to play perfectly (i.e., not in a reduced-quality shuttle mode) and record at speeds higher and lower than real time. Note that this is a desire only; the format should not lock it out, but it is by no means necessary that all transports implement it. A camera, for example, need only implement real-time recording, while a dubbing recorder would implement copy as fast as possible, with no need for a slow-speed mode.

Format

The preceding argument suggests that a streaming tape format, such as is used for computer tape drives, might well be suitable. In a streaming tape, the tape transport does not attempt to stay exactly at the correct position in the tape. Instead, it makes use of a random access memory (RAM) buffer, which contains enough memory to buffer, and thus conceal, the time it takes to start or stop the drive. When replaying, the tape partly fills the buffer and attempts to keep it full to a pre-determined level. If the buffer fills above this level, the tape is slowed or even stopped until the buffered data has been consumed. Similarly, on recording, data is accumulated into this buffer until a worthwhile amount has been recorded, when the tape is started. The tape is then speeded or slowed in order to keep the buffer from overflowing, while still maintaining the speed necessary for accurate data flow. If the VCR gets ahead of itself, by overfilling the buffer while playing or emptying it while recording, it may even have to stop, back up, and start again; however, this is innocuous because the buffer is designed to be large enough to conceal this from the user. Some digital VCRs already contain elements of such a system, but they are still constrained by the requirement for frame-accurate editing.

Variable Data Rate

Variable-speed recording is a necessity if we are to properly exploit data compression, since different scenes carry different amounts of information, and under any reasonable compression scheme will compress to different sizes. A simple graphic or frame from an animated cartoon will usually compress very efficiently, because it contains large areas of flat color. A busy camera scene (sports footage tends to be among the worst) requires much more information to reproduce. If we shoehorn the data into a fixed size, we either grant excessive bandwidth to the cartoon or get low-quality sports pictures. The former is chosen, for obvious reasons, but it raises tape costs for lower bandwidth recordings.

By adopting variable-speed recording, we allow the tape speed to be varied in order to record whatever amount of data is produced by our compression system at our selected level of compression. If the recording contains little information, the data will be reduced and the tape slowed. Conversely, as the picture gets busier, the tape speeds up.

Compression

From the preceding discussion, it will be obvious that this format should include data compression. The quality

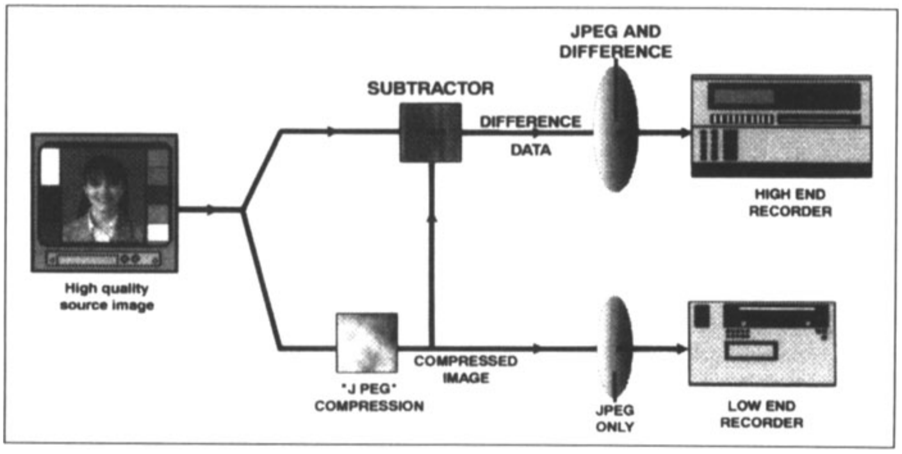

Figure 2. Recording.

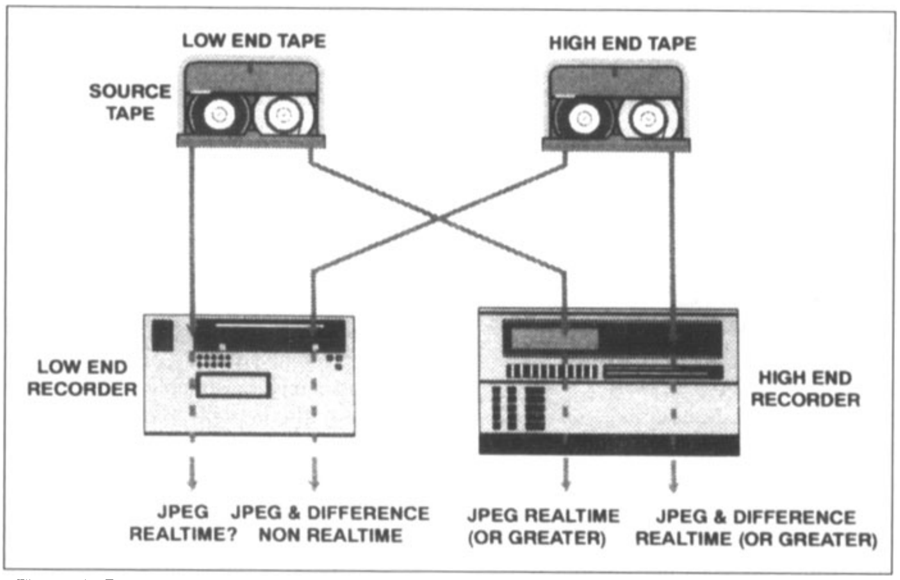

Figure 3. Playout.

of modern compression schemes is too high for the savings they produce to be ignored, especially in less demanding roles such as news gathering. On the other hand, the capability for true CCIR 601 recording is essential for any system used for storing high-quality post-production material, because even small compression artifacts are unacceptable for top-quality post-production.

The solution is to pick up a proposal from the JPEG standards, which may not as yet have been used in moving picture systems. This is to achieve lossless compression by combining a lossy compression (such as the JPEG system, although others, such as fractal or wavelet, might be used) with error-correction information, which fixes the errors produced by the lossy system. Such a system usually produces some data reduction over the original, although not nearly as much as the baseline JPEG system. Experience shows that the best available lossless compression for busy pictures is about 1.8 or 2.0 to 1. The baseline compressed data and the error correction information would be stored on tape in separate, variable-size fields, and the error correction information would be optional.

A lower performance system would not need to record the error correction information and could ignore it on replay. This would be particularly suitable for cameras, because experience has shown that a reasonable level of data compression is quite acceptable on pictures that consist only of a single real-world image. Pictures containing multiple images and digitally generated graphics (such as those produced by editing systems) tend to produce compression artifacts and would need the error-correction information. Figure 2 shows high and low-end recorders capturing the same image.

Audio could be handled in the same way. The number of audio tracks needed for a recording is always at least one more than the number available. If audio is allocated fixed space, more audio tracks need additional tape and bandwidth, even if they are not used. If audio is added as further optional fields in the recording format, the number of audio tracks can be varied as needed. Cameras will probably record only one stereo pair, while a distribution tape could carry many stereo tracks, including music, effects, and speech in several languages. A player need not be capable of playing all audio tracks, provided that it can be instructed to pick out one or two desired tracks from the set available on the tape. (The functions allocated to the tracks will of course be available in the clip directory.)

Dubbing and Multitracking

The specification demands that the tape should, in theory at least, be capable of recording at double speed for dubbing purposes, although not all transports need implement this facility. Instead of using this facility for doubling dubbing speed, why not put two (or even more) video fields into the data fields on tape? This would allow video to have the same kind of multitrack ability as is taken for granted in the audio world and would be a boon for multicamera shooting, cutting recording costs while simultaneously simplifying the edit. As stated earlier, a moderate level of compression is usually acceptable for raw camera footage without error correction. If the transport has the bandwidth to support real-time lossless recording with an effect compression of 2.0:1, it can support five channels at 10:1 compression. Such tapes might not be playable on low-end replay systems (low-end equipment must be inexpensive), but low-end recording could be played on high-end players, which would be attached to the off-line editing system. Examples are shown in Fig. 3.

As one of the premises to the discussion, all editing will be done by nonlinear, disc-based, systems. However, as mentioned earlier, there is bound to be a need for a non-disc-based dubbing system. It is easy to see that the capabilities of such a system can be quite restricted. It need only copy whole clips, which it appends to the end of the record tape. It copies all information from the clip directory (so copyright information always travels with the pictures — a significant gain). While the system must copy all information on the tape, it needs to display only the compressed image without error correction and one or two audio tracks (if, indeed, any video display is needed at all). Clips to copy can be selected by a simple menu presentation of the directory.

Conclusion

Although this paper cannot accurately predict the viability of such a proposal, in terms of the exact design parameter of a tape transport and the physical medium, there is enough evidence to suggest that technology exists to economically implement such a proposal. Whether or not simply adding features to one of today's high-end tape streamer systems will suffice is yet to be seen, but undoubtedly if the assumptions of past practice are thrown away, the industry may be able to create an acquisition format that is both cheaper and better than its predecessors.