Towards Improved Tape Storage and Retrieval Response Time

John J. Gniewek
IBM Corporation
9000 S. Rita Road
Tucson, Arizona, 85744
gniewekj@us.ibm.com
Tel: +1-520-799-2390
Fax: +1-520-799-4138

Introduction

Magnetic tape has maintained a prominent place in the computer data storage hierarchy for in excess of forty years. As a result of continuing technology advances, the improvements made in both areal and volumetric recording density have assured that magnetic tape remains the lowest cost storage medium for most computer data storage applications. Relative to the impressive advances in reduction of disk storage costs, tape storage has maintained a 10X to 100X advantage in lower cost per Gigabyte. For applications such as back-up, disaster recovery, and passive archive of very infrequently accessed files, this lower cost, as well as the removability and transportability of the tape storage media, has ensured a continuing role for magnetic tape. The frequent projections of the demise of magnetic tape as a result of disk storage price-performance improvements have ignored the enablement of new applications that require significantly more storage capacity and hence, provide the driving force to maintain the 40+ year storage hierarchy paradigm. Several of these new applications place increased demands on the ability to retrieve data objects quickly and require that the tape storage serve as an active member of the hierarchy, not merely a passive member, as in applications such as disaster recovery. Until recently however, not much effort had been made to improve the retrieval response time of tape. In this metric, tape storage has continued to be disadvantaged by three to four orders of magnitude relative to disk storage.

Concomitant with the very significant tape technology advancements in the last several years, there has been, and continues to be a proliferation of new types of recording devices and robot systems. In order to provide a means for judging suitability of particular types of devices for these new emerging tape applications, it would be desirable to be able to develop a figure of merit for comparing different recording systems, each with widely different component characteristics. The analyses presented here has been of value to design engineers and can be expected to be of value to application systems engineers and system integrators.

Figure of Merit--What is Appropriate?

The diversity of applications precludes the possibility of any one figure of merit representing <u>all</u> requirements equally well. However, we can start by identifying individual favorable attributes and proceed to integrate the individual parameters into a composite figure of merit. This composite figure of merit is developed as a result of applying the analyses to meet the requirements of several specific applications. The approach to constructing appropriate figures of merit is developed in three stages. In Figure 1, a parameter called the ‘effective data rate’, (EDR), is first presented without system capacity or cost considerations. In Figure 2, drive cost, but not total storage capacity, supplements the EDR parameter. Finally, in Figures 3 and 4, total system cost is factored in and is presented in units of (MB^2/second/$) as a proposed comparative figure of merit. At this last stage, in addition to the device hardware variables, two new application variables, aggregate data rate, and library capacity are introduced.

Efforts to make improvements in retrieval response time for tape storage systems must be considered at the system level involving hardware, software, and application data organization approaches. Hardware factors include items such as robot cycle time, drive load and unload times, tape search and rewind times, drive data rate, recording density (linear and track), cartridge capacity, etc. In addition to these more obvious factors, items such as track format and cartridge design can also provide a means for improving retrieval response time. An example of two different types of drive/cartridge designs illustrated the advantage a midpoint load cartridge design could provide in improving response time for a serpentine longitudinal recording format (1). For certain types of retrieval patterns, this design has the effect of doubling the search speed. The purely mechanical aspects of improving response time by simply going faster runs counter to wanting to reduce costs since without concomitant increases in recording density or mass reductions, faster implies larger, more expensive motors. Thus, in addition to the mechanical improvements, there is room for significant improvements via I/O scheduling algorithms (2), enhancing locality of reference by data organization during the writing operation, and utilizing special attributes of the recording technology to enable partitioning of the cartridge recording format such that a high capacity cartridge can have multiple partitions, all located at the logical beginning of tape.

A) Effective Data Rate--Enhancing Locality of Reference

Smart software and intelligent data organization with some prior knowledge of the expected data retrieval pattern can be expected to have a profound effect on the retrieval response times. This data organization is generally referred to as improving the “locality of reference”. The benefits result from a higher ‘hit rate’ on a mounted cartridge with a data rate more closely representing the device streaming rate. The ‘read’ performance with various intelligent I/O scheduling algorithms and with knowledge of the drive characteristics has been analyzed and reported at this conference by Hillyer (3) for the 3570 Magstar MP device. Sometimes, however, the read retrieval benefits are not without cost to the ‘write’ operation. In general, in order to improve the locality of reference, an increased number of tape mounts is required during the data entry operations such that common data types may be collocated on separate cartridge classes. An example from a commercial application of where this would apply is described and used to develop one relevant composite figure of merit parameter. Other examples from scientific/technical applications can be envisioned and could include the appropriate collocation of spatial/temporal data.

For a number of commercial financial applications involving monthly billing cycles, such as check and credit card statement processing, very large numbers of small objects must be processed expeditiously. As this industry evolves to include images, the data capacity requirements increase dramatically. As a result of the storage cost advantage of tape, these high total capacity requirements make it highly desirable to be able to employ tape storage for a portion of the processing requirements. Thus, this application may be defined as: 1) a streaming data ingestion (e.g. from an image capture check sorter, or an instrumentation satellite), 2) intermediate data labeling and processing, and 3) storage to tape in several versions; a) back-up for the original information until first-order processing is complete, b) chronological entry for ad-hoc retrievals for some period of time, and c) collocated data organization for anticipated future processing. In the case of financial records, this would correspond to end-of-the-month statement processing. It is the high capacity requirement that provides the driving force for wanting to use tape. For low capacity applications, all the intermediate processing would employ disk storage.

The relevant figure of merit for this type of application is thus defined as “effective data rate” (EDR) and corresponds to the total amount of data transferred versus clock time. Clock time includes not only the data transfer time, but also all the other nonproductive times including load, search, robot moves, etc. Thus, the EDR figure of merit encompasses more than just the native device data rate. EDR thereby serves as a means of integrating several independent parameters such that a comparison can be made between types of devices with widely different characteristics. Since the effective data rate for any given device is higher for larger object sizes (a greater percentage of the time is spent in the data transfer mode), by evaluating EDR as a function of object size, the interests of different types of applications are considered.

1) Defining the EDR Parameter

Using the commercial application as a template, in order to minimize processing time during the end-of-the-month processing cycle, it is necessary to provide collocation of the ingested data into G groups (a group is a set of cartridges that contains data only for that group). The ingest rate is defined as IR (MB/sec). The analysis requires that the disk buffer size and the number of tape storage devices be balanced to provide the most economical solution. Obviously, if G devices were provided, there would be no need to swap cartridges during the loading operation other than for full cartridges. However, this is not the most economical solution. Rather, we seek to configure the system such that the effective data rate for the write operation balances the ingest rate. An adjustable parameter is the size of the object that will be transferred each time the cartridge is mounted. Thus, EDR as a function of object size written becomes a relevant parameter and serves to estimate the number of drives required given the constraint of a certain disk buffer capacity. The EDR thus serves as an integrated figure of merit and allows direct comparison between devices with diverse characteristics such as, for example,: a) a device with high data rate and high capacity (but long search and rewind times due to long tape length) with b) a device with moderate data rate and capacity, but short search and rewind times. For illustrative purposes, several hypothetical devices are constructed and compared for EDR values. This application corresponds to a scheduled write operation and no queuing is involved. Higher numbers for EDR is in the direction of goodness.

2) Quantifying EDR

The sequence of operations used to develop an expression for EDR is shown in Table 1.

Table 1
Sequence of Operations for Defining EDR

Operation	Parameter
Robot Get Cartridge	
Cartridge Exchange	AS
Load to Drive	LD
Search	C/2KV*
Read	O/D
Rewind	C/2KV*
Unload	ULD

* The search and rewind times are expressed in terms of cartridge capacity, C, search velocity, V, and recording density, K (MB/M). O is object size in MB. All other terms are in self-consistent units to express EDR in MB/sec. (1).

The expression for the effective data rate, EDR is then given as:

$$EDR = \frac{O}{\left[AS + LD + \left(\frac{C}{K \bullet V}\right) + \frac{O}{D} + ULD\right]}$$

In Figure 1, EDR is expressed in (MB/sec) and represents a situation corresponding to a series of APPEND operations that include multiple cartridge mounts. It would approximate the measured elapsed time data rate under conditions where the appended file size is small relative to the cartridge capacity and the cartridge is filled by the multiple write operations. It would not apply for low capacity cartridges where, for example, an 800 MB object is written to an 800 MB cartridge. In this case, a full cartridge write would result from a single cartridge mount and no search and rewind operations would be invoked.

The EDR figure of merit defined in this manner may not represent a specific tape application but it does represent 'goodness' of the device when considered as a measure of throughput. It becomes more relevant as advances in tape recording technology provide higher capacity cartridges. The parameters chosen for several hypothetical devices are shown in Table 2. Results for EDR as a function of object size for these devices are shown in Figure 1. The results presented in Figure 1 represent the base performance of the different devices without regard to cost. In Figure 2, EDR for each device is divided by the device cost so that the figure of merit is represented in units of Bytes/second/$. In both Figures 1 and 2, higher numbers are better. These results provide a figure of merit which values effective data rate without valuing cartridge capacity. For a given technology, higher capacity cartridges result in lower EDR as a result of longer search and rewind times. A means to incorporate capacity and data rate into an integrated system level figure of merit is developed in a subsequent section.

Table 2a
Hypothetical Device Parameters

Device ID

	A	B	C	D	E	F
Parameter						
AS (sec)	8	8	8	8	8	8
LD (sec)	20	40	5	10	15	5
ULD (sec)	10	15	5	5	15	5
D (MB/sec)	10	5	2	3	12	7
V (M/sec)	5	4	10*	1	4	10*
K (MB/M)	35	60	35	80	120	35
C (MB)	10,000	40,000	5,000	20,000	50,000	5,000
DC ($)	30,000	10,000	8,000	8,000	80,000	10,000
CC ($)	50	100	50	50	100	50

Table 2b
Hypothetical Device Parameters

Device ID

	G	H	J	K	M	N
Parameter						
AS (sec)	8	8	8	8	8	8
LD (sec)	20	20	5	5	5	5
ULD (sec)	10	10	5	5	5	5
D (MB/sec)	10	10	7	7	15	15
V (M/sec)	5	5	10*	10*	15	15
K (MB/M)	35	70	35	70	150	150
C (MB)	20,000	20,000	10,000	10,000	30,000	100,000
DC ($)	30,000	30,000	10,000	10,000	30,000	30,000
CC ($)	100	50	100	50	100	150

*Effective Search velocity for midpoint load two-reel cartridge design.
DC and CC are hypothetical costs of drive and cartridge respectively.

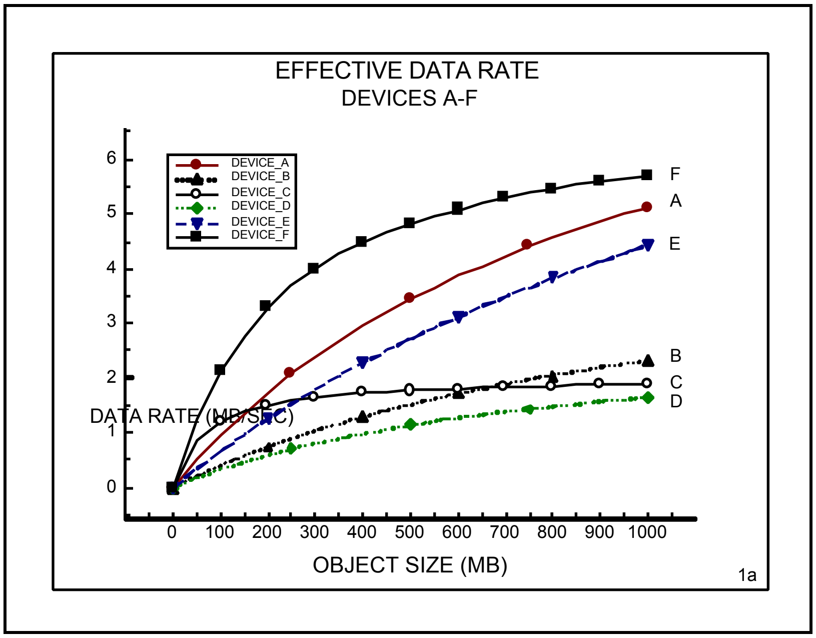

Figure 1a. Effective Data Rate (EDR), (MB/second), as a function of object size transferred in a cartridge mount cycle. Devices A-F. This metric corresponds to random retrievals from full cartridges or to separate sequential write 'Append' operations. See text.

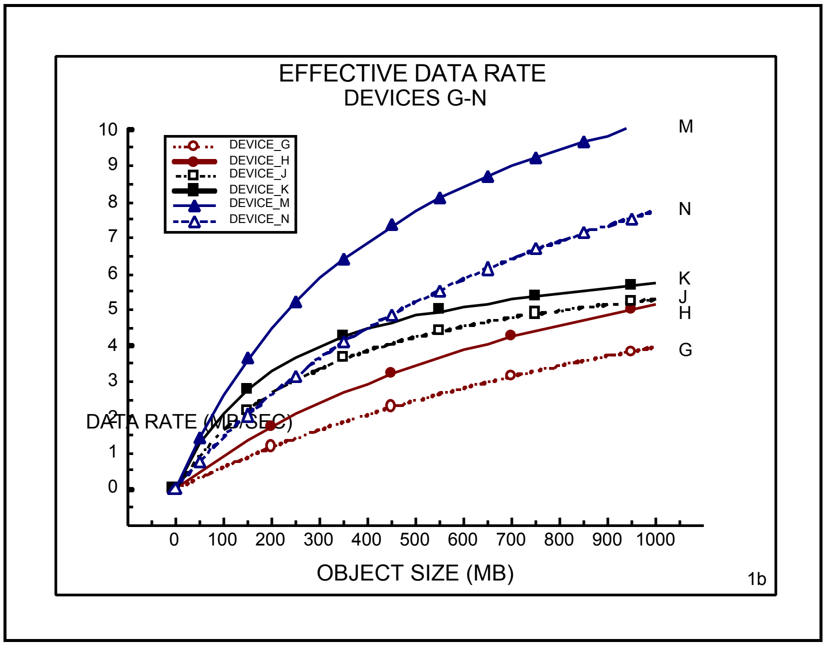

Figure 1b. Effective Data Rate (EDR), (MB/second), as a function of object size transferred in a cartridge mount cycle. Devices G-N. This metric corresponds to random retrievals from full cartridges or to separate sequential write 'Append' operations. See text.

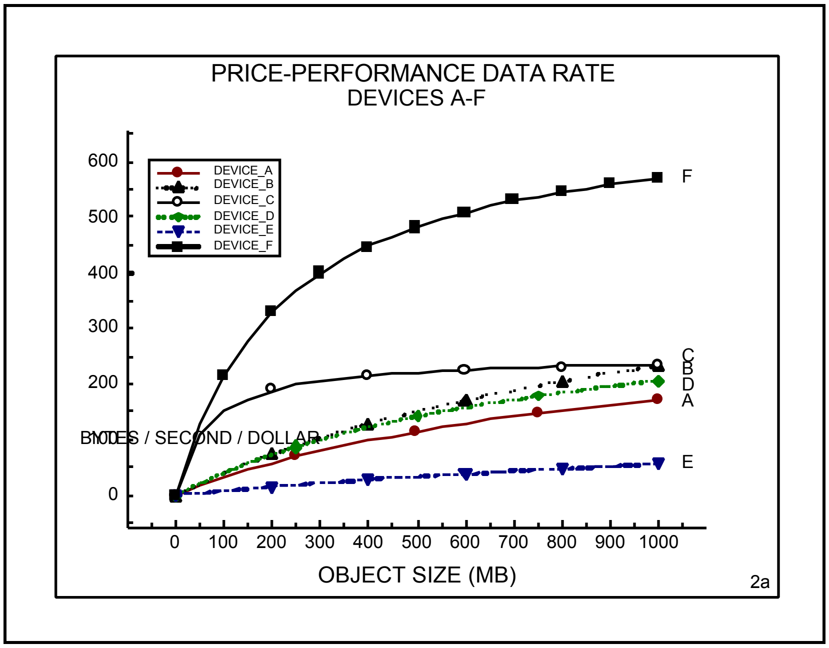

Figure 2a. Price-Performance effective data rate, (MB/second/$). Devices A-F. Calculated as EDR/drive cost per characteristics listed in Table 2. See text.

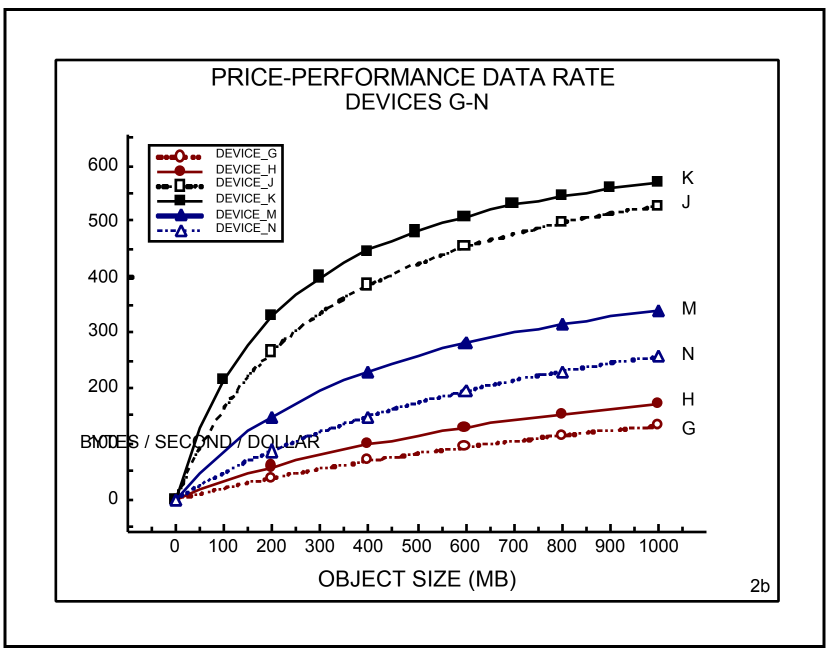

Figure 2b. Price-Performance effective data rate, (MB/second/$). Devices G-N. Calculated as EDR/drive cost per characteristics listed in Table 2. See text.

3) EDR-Discussion of Results

The six devices hypothesized in Table 2a are composites of characteristics similar to those existing in several different current technology devices. From the table of numbers alone, because of the widely disparate individual parameters, it would be difficult to rank the devices in order of effective data rate over the range of object sizes considered. Some non-intuitive results are, however, apparent from Figure 1a which corresponds to these six devices. Device C, which has the lowest native data rate (2 MB/sec), provides a higher effective data rate than all other devices other than Device F for the following object size ranges: Device A (10 MB/sec.), up to approximately 150 MB object size; Device B (5 MB/sec.), up to approximately 600 MB; Device D (3 MB/sec), up to > 1000 MB; Device E (12 MB/sec.), up to approximately 250 MB. When the hypothetical device costs are considered, the results are presented in Figure 2a in units of Bytes/second/$. Devices F and C are then ranked one and two for all object sizes up to 1 GB.

In Table 2b, Devices G and H were constructed such that the cartridge capacity for both is double the capacity listed for Device A. However, this was achieved by different means. Device H doubled capacity by doubling the recording density (and keeping tape length the same), while Device G doubled capacity by increasing tape length at the same recording density as Device A. The improved performance of H compared to G is evident in Figure 1b. Similarly, Devices J and K double the capacity of Device C in an analogous manner. Devices M and N assume a further improvement in recording density representative of what is expected to appear in devices in the near future. These data are presented in Figures 1b and 2b. Note the scale change in Figure 1b relative to 1a. The areal density improvements assumed in Devices M and N are expected reasonable extensions of current technology, but are still far short of the 10 year goals projected by a recent NSIC tape storage study group (4).

B) Cartridge Capacity Considered

The manner by which cartridge capacity may be incorporated into an appropriate figure of merit is not unambiguous. Value can be judged by several different criteria. For a single user desktop application, probably the most important criteria is the ability to provide a single cartridge file backup. Drive cost would also be high on the priority list, but because of the limited number of cartridges likely to be in use, media cost would not be heavily weighted.

For larger users, particularly on a multi-user network sharing common tape storage, tape automation is considered to be essential. The analysis developed here addresses the needs of medium and large size storage libraries where the storage system is composed of drives, media, and automation. A system is defined by specifying: A) the required library capacity (LCAP), and B) a required aggregate data rate (ADR). Device and media costs are assumed per the values in Table 2. Automation costs are introduced in a simplified manner by assuming a fixed cost per slot (in the examples given here, a value of $100/slot is used). The value of higher capacity cartridges is then reflected in the figure of merit as a result of requiring fewer slots and hence lower automation costs. For a given technology recording density, higher capacity cartridges will however, reduce the EDR as a result of longer search and rewind times. Thus, if 'goodness' is considered to be high effective data rate, high library capacity, and low system cost, the figure of merit chosen to represent this composite is: ((EDR) x (LCAP))/System cost. System cost is defined as the sum of media cost (number of cartridges for a capacity of LCAP times the cartridge cost), plus the drive cost (number of drives required to meet the ADR using EDR as the individual device data rate times the individual drive cost), plus the automation cost (simplified here as $100 per slot times the number of slots required). In this analysis, an adjustment factor is provided to allow for different cartridge sizes and the number of cartridges that may fit in a given wall space of the automated library.

The results are presented in Figures 3 and 4 for library capacities of 1 TB and 10 TB respectively. The units are MB^2/sec./$. This results from ((MB/sec.) x (MB))/$. The units do not have any direct functional application and should thus be considered strictly as a figure of merit providing appropriate weighting for data rate, capacity, cost, and application conditions, i.e. library capacity, aggregate data rate, and object size transferred. The step function in the graphs results when the number of drives required to achieve the desired aggregate data rate decrements by one as a result of the higher EDR at larger object sizes. For this example, a value of 20 MB/second was used for the desired aggregate data rate. The difference in the ranking of device types at 1 TB and 10 TB libraries reflects the weight given to higher capacity cartridges at larger library sizes and the amortization of system cost over a greater number of cartridges. Note the scale change between Figures 3 and 4.

The concepts developed in this analysis have been simplified compared to what would be required for a specific application system configuration analysis. Refinements, such as allowance for variable cost of storage slot, alternative definitions for the effective data rate, and treatment of costs as reflective of total cost of ownership, could be introduced. Also, there may be other specific constraints, such as maximum object retrieval response time, storage space, reliability factors, etc., which could provide additional weighting factors in a comparative evaluation. The analysis given here is sufficient to guide product development engineers as to the relative importance of various individual attributes towards a competitive design point and may be useful to systems and application engineers charged with evaluating many diverse potential storage solutions.

Tape Track Format Design

The development of high track density serpentine linear recording formats has provided additional opportunities to improve the retrieval response time from magnetic tape storage devices. In formats of this type, the time required to 'jump' across tracks is small compared to the time required to search down the length of a track. These attributes, when coupled to intelligent request reordering algorithms, can improve the response time required for retrieving multiple objects randomly located within a given cartridge (2,3).

The manner by which the serpentine track format is written varies among the different types of devices. In some cases, the tracks are written by a 'shingling' process where the width of the track left on tape is less than the write head width as a result of partially overwriting the previous track when the write head is indexed to a new position. In other designs employing precise track-following head positioning servo systems, each track is independently written without overlapping the neighboring tracks. This format design is thus amenable to logically partitioning the cartridge capacity into multiple partitions, each partition located at the logical beginning of tape and each partition capable of being individually rewritten while maintaining the data integrity of the neighboring partitions. The number of partitions and the capacity of each is a function of the total cartridge capacity and the number of tracks that are written concurrently.

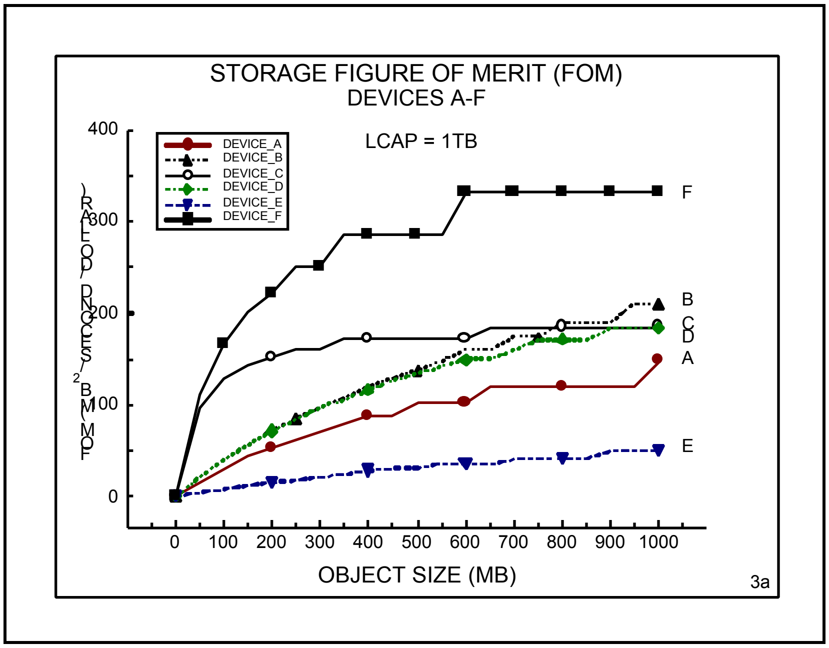

Figure 3a. Storage Figure of Merit in units of (MB^2/second/$). Devices A-F. Library capacity is 1.0 TB. Aggregate effective data rate is 20 MB/second. See text.

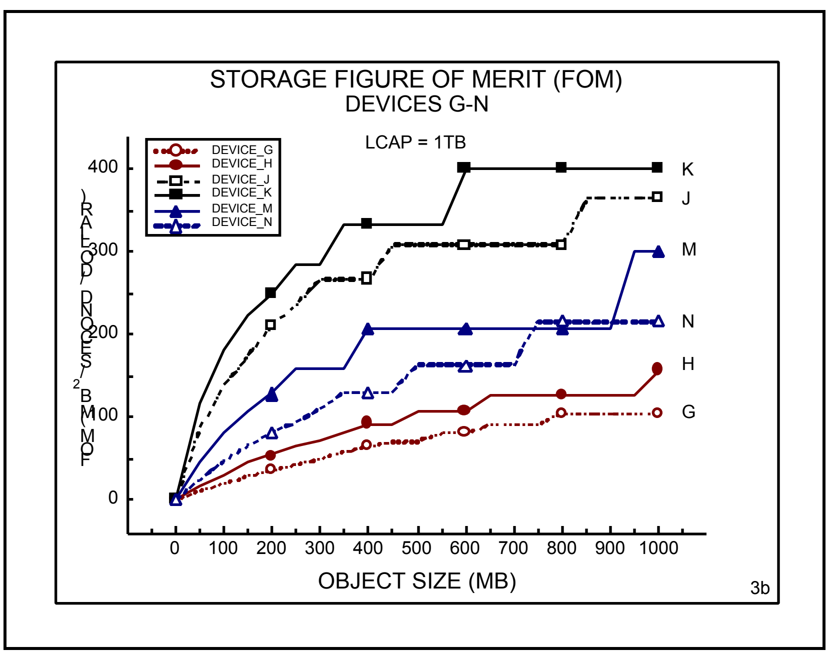

Figure 3b. Storage Figure of Merit in units of (MB^2/second/$). Devices G-N. Library capacity is 1.0 TB. Aggregate effective data rate is 20 MB/second. See text.

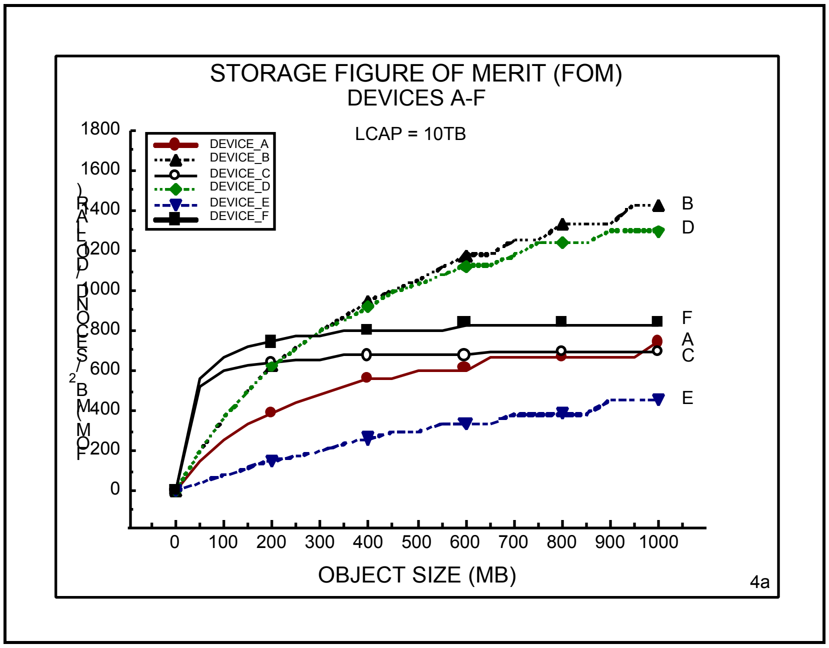

Figure 4a. Storage Figure of Merit in units of (MB^2/second/$). Devices A-F. Library capacity is 10.0 TB. Aggregate effective data rate is 20 MB/second. See text.

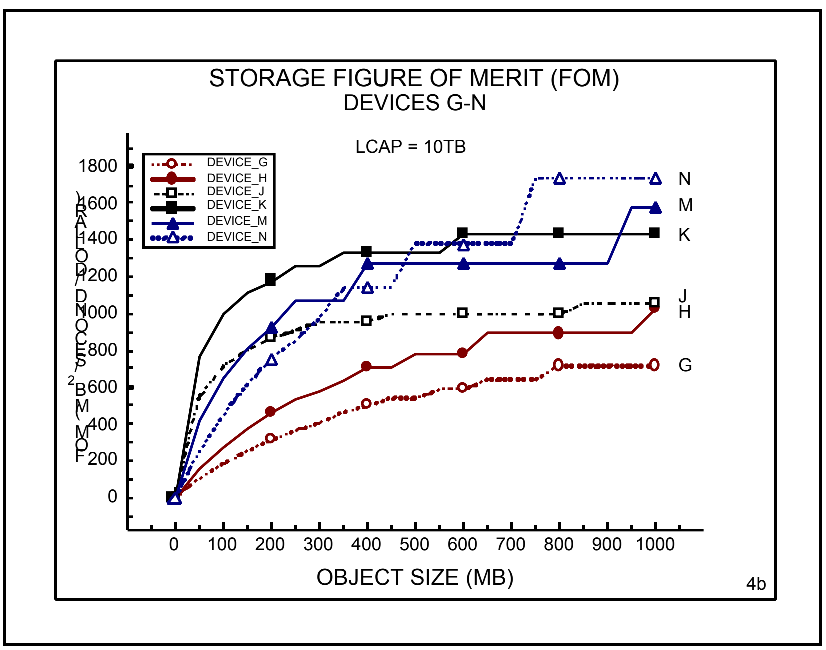

Figure 4b. Storage Figure of Merit in units of (MB^2/second/$). Devices G-N. Library capacity is 10.0 TB. Aggregate effective data rate is 20 MB/second. See text.

Consider a cartridge with 128 tape tracks and a capacity of 5000 MB, written 4 tracks at a time, and with a midpoint load. In this case, a total of 32 partitions, each of approximately 155 MB capacity and each beginning at the logical beginning of tape (LBOT) would result. If the application required data set sizes in this range, the response time to first byte of data would be improved as a result of eliminating the search time. Technology factors that enable this capability include track-following head positioning servo systems and precise dimension thin film recording head fabrication processes. A schematic of the track layout illustrating this tape format is shown in Figure 5. Data management software enhancements are required to exploit these features.

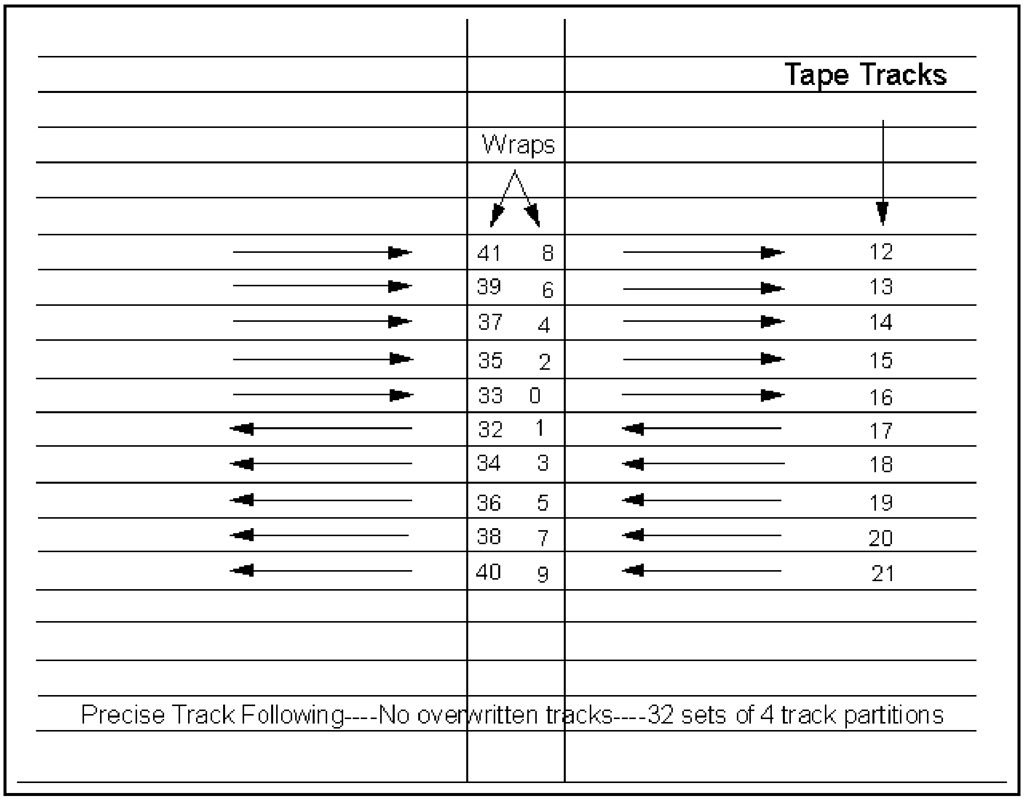

Figure 5. Schematic of track layout for a track-following servo, midpoint load tape device illustrating capability of multiple logical partitions, all beginning at the logical beginning of tape. Sequentially written tracks do not overlap or overwrite previously written tracks. Illustrated is a portion of the track sequence for one head element of a track format with four concurrently written tracks. 'Tape Tracks' shows track location on tape. There are a total of 128 data tracks on tape. The sequence of written tracks proceeds from 0 to 1 to 2....etc. Middle region is the midpoint load region. The right half is written first whereupon the left half begins writing at wrap # 32. Any full wrap, for example (4,5), may then be overwritten while maintaining integrity of the neighboring tracks.

Integrated System Solutions

Current open system storage systems incorporate, via software, direct access to tape storage in a manner that appears as expanded disk storage, albeit with longer response time. Recently several system solutions that feature 'virtual' tape drive capability for the mainframe market segment have been introduced. Common to both approaches is the use of a disk buffer that results in greatly improved performance in the write to tape mode and in the ability to get full utilization of the cartridge capacity. Retrieval of data is improved to the extent that the ‘hit rate’ to the disk cache is significant. Thus, the aggregate performance will be very much application dependent. In situations where the library capacity is large and there is a random retrieval pattern, real tape devices with fast response time are required. Under high load conditions, queuing delays become significant. The response time as a function of various application conditions and types of devices has previously been presented (1).

Tape Storage Trends

The natural operating domain for tape storage is in systems requiring large storage capacity. What is considered large is, however, a moving target. The lower cost per MB of tape storage must translate into significant actual dollars to accommodate the total costs of a storage hierarchy. Continued decreases in $/MB for disk storage and the introduction of new technologies such as high density removable floppy disk storage, and increased density recordable optical disk products are putting pressure on tape storage devices at the individual desktop user market segment. For tape storage to maintain a presence in any individual market segment, tape storage costs should maintain at least one order of magnitude advantage relative to competing technologies that have more favorable response time attributes. This is expected to be maintained for midsize and large storage capacities, but is doubtful for low-end individual user segments (4). Based on the NSIC study, improvements in tape storage volumetric density can be expected to advance in an evolutionary manner by 10-20X over the next 10 years. Some of these technology advances will be used to improve tape storage and retrieval response time.

Conclusions

The analyses presented here have made an approach to developing a comparative figure of merit that allows quantitative comparison of devices and systems with widely different characteristics. This was prompted by the arrival of new applications for tape storage which must be analyzed in a manner that requires greater sophistication than the historical metrics of only cartridge capacity and device data rate. Use of such comparisons can be expected to lead to further product enhancements that will provide improved response time from tape systems.

1. Effective data rate is proposed as a measurement parameter, in general, more useful than native device data rate. It gives appropriate weighting to other drive attributes and is reflective of the throughput that can be expected. Examples were given that illustrated higher effective data rates for lower native data rate devices.

2. The key tape device technology factor that would allow improved response time (at constant cartridge capacity) is the recording density, K, given as MB per meter length of tape. From basic recording physics constraints, this can be expected to be achieved predominantly by higher track densities. Most recent developments have concentrated on increasing capacity via thinner, longer length tape. Response time is degraded relative to what could be achieved from density improvements.

3. A system level figure of merit that integrates cartridge capacity, cost, and application conditions, as well as EDR, is proposed as a means of quantitatively comparing widely different system component characteristics. The analysis is of value in guiding development engineering priorities. Enhancements and refinements to the methodology can be expected.

4. Intelligent software is required to take advantage of device characteristics that could improve response time from current technology devices. This could include I/O scheduling algorithms and use of multiple logical volume partitioning within a high capacity cartridge. Continuing future increases to cartridge capacity can be expected to increase the demand for such features.

References

1) J. Gniewek, “Evolving Requirements for Magnetic Tape Data Storage Systems”, NASA Conference Publication 3340. Fifth NASA Goddard Conference on Mass Storage Systems and Technologies, pp. 477-491, September, 1996.

2) B.K. Hillyer and A. Silberschatz, “Random I/O Scheduling in Online Tertiary Storage Systems, Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, Montreal, Canada, June 3-6, pp. 195-204.

3) B.K. Hillyer, “Scheduling Noncontiguous Tape Retrievals”, Joint NASA and IEEE Mass Storage Conference, March 23-26, 1998.

4) National Storage Industry Consortium (NSIC) Tape Storage Workshop, September, 1997. Boulder, Colorado. Final report due 1Q98.