MASTER

A LARGE MASS STORAGE FACILITY*

Arnold M. Peskin, BNL

INTRODUCTION

In May 1978 the Chairman of the BNL Applied Mathematics Department organized a group to study the questions surrounding the proposed acquisition of a large mass storage device for the Laboratory's central science computing facility (CSCF). This action was based on some optimistic forecasts for funding of such a procurement in the near future. The study group was to perform several functions:

1. Confirm the programmatic justification for this device.
2. Survey the spectrum of current and future vendor products that may meet the need.
3. Draft a set of specifications for the system.
4. Discuss possible alternatives to alleviate the problem in the event that procurement is delayed or denied because of budgetary limitations.

The group (see Table 4) consisted of persons with backgrounds representing a broad range of viewpoints in computing: hardware and software, development and maintenance, user and purveyor of the data processing service.

This paper is the final report of the group and serves as the report to the BNL Computer Policy Advisory Committee.

THE CURRENT COMPUTING ENVIRONMENT

The CSCF is the primary facility for meeting the computational needs of the multitude of scientific programs at BNL. It operates in an "open shop" environment. During periods of peak activity, the CSCF has recently begun to show signs of system saturation. Steps are being taken to augment the configuration in stages to meet the growing needs.

The computer configuration of the CSCF is shown in Fig. 1. The triplex of major mainframes in the facility is a Control Data Cyber 76 computer supported by two Control Data 6600 computers. Each 6600 is connected to a common 1.5 million words of extended capacity storage. The facility has the usual complement of unit record peripherals, 7- and 9-track tape drives, and a number of disk storage drives. Figure 1 also indicates local communications controllers and additional interface equipment. Near-term anticipated improvements include increasing the communications capabilities and expanding the disk capacity toward a realistic upper limit of about one billion 60-bit words.¹

PROGRAMMATIC JUSTIFICATION

BNL has been requesting funds for several years for an on-line large mass storage (LMS). The basis for the request was the con-
Supported in part by the United States Department of Energy.

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

Fig. 1. C3CF equipment configuration.

viction that a device was needed which would have very large capacity (such as manually mounted magnetic tape) and would be almost as fast in access as disk storage. It was envisioned that a 10^{12} -bit storage with an access time on the order of seconds would significantly enhance conventional system performance and provide a new functional capability to the users as well.

Table 1 summarizes a part of the programmatic storage requirements that would be met by the LMS. The energy and environment

Table 1
Mass Storage Facility Program Requirements

	10^9 Bits	
	Now	In 5 yr
User allocated storage		
Energy and environment	45	100
Nuclear engineering	5	10
High energy	105	2000
Other programs	5	10
System allocated storage		
Dynamic	72	150
Archival	500	1000
Total	732	3270

and nuclear engineering entries for current needs are based on actual programs and data bases. For the high energy physics segment of users, estimates had to be made, based on certain assump-

tions. It was noted that during a typical active week physics experiments transmit ~500 tapes or 5×10^{10} bits of information to the CSCF for temporary storage. It was estimated that a similar amount of vastly reduced processed data retained for a much longer period of time would occupy a similar amount of storage space on an LMS. Including an additional five billion bits for frequently used program libraries and utilities, the high energy physics total needs already exceed 10^{11} bits.

The computers' operating systems would use space on the proposed device to excellent advantage. A dynamic allocation area can augment present rotating mass storage facilities, and an archival area can reduce user tape traffic. Both disk and tape activities have been identified as major bottlenecks to system throughput. Consequently the equivalent of about 20 additional 844 disk drives and 5000 library tapes is the estimate for system allocated usage of the LMS, largely for user files.

Other BNL programs are also expected to utilize the LMS, with smaller requirements, although storage estimates for five years hence tend to be uncertain. It is likely that in time the requirements of ISABELLE will dominate. Several estimates have been made of the computational and data requirements for ISABELLE. The BNL study group estimate is, conservatively, two trillion bits, assuming that current patterns of usage will be continued.

VENDOR SURVEY

The Advisory Committee identified four products as likely to apply to the mass store programmatic needs. Systems manufactured by IBM, Control Data (CDC), Calcomp, and SDC are listed in Table 2 along with some of their characteristics. A fifth option, that of adding 100 more disk drives, is included for comparison. Note that the capacity column lists the capacity of a system responsive to the 10^{12} -bit requirement and is not the minimum module of the product.

The relative merits of these systems were discussed at a Mass Storage Workshop held in October 1977, sponsored jointly by the National Center for Atmospheric Research and the Department of Energy, and attended by prospective customers from the two agencies, vendors, and other researchers; the Proceedings provide an excellent review of existing storage hardware and vendor developments.² An IEEE Special Issue similarly highlights research in storage media and techniques.³

The SDC system, known as Terabit or TBM (originally developed and marketed by Ampex), uses video tape recording techniques and provides hardware and software interfaces to certain CDC, IBM, and DEC computer systems.

The Calcomp advanced tape library (ATL) consists of a tape handling "robot," tape storage vault, and rudimentary interfaces to IBM 360 and 370 systems. It uses standard computer tape, which can be an attractive feature. This is the lowest cost and most modular option, but it is also the most primitive and probably

Table 2

Vendors

	Storage medium	Capacity (10 ¹² bits)	Access (sec)	Transfer rate (Mbits/sec)	Host support	Expanded up to (10 ¹² bits)
Ampex/ SDC TBM	video tape	1.00	17 av	2.7	7600 360/370 DEC hw & sw	2.91
Calcomp ATL	std. comp. tape	1.00	15-20	**	360/370	4
CDC 38500	*	1.024	8.3	3.0	360/370 cyber/ NOS	22
IBM 3850	*	1.35	15	3.0	360/370	3.8
100 844-41 equiv.	mag disk	0.14	55 x10 ⁻⁸	6.45	all vendors	none

* Special magnetic tape cartridge.

** Tape drive and channel speed (6 Mbits for 6000).

would involve the greatest amount of development for installation and integration into a computing center.

The CDC and IBM LMS systems are quite similar. They use essentially the same wide-tape cartridge recording techniques, they exhibit similar performance, and both are designed for the IBM market (although CDC is beginning to support its own device on one of its operating systems and may soon have support for its other one.)

As a guide for procurement, the BNL working group has prepared a draft specification for LMS.

If procurement of a major LMS such as that described by this specification does not materialize, it may be possible to meet some of the needs partially by much more modest investment. An alternative system might include the Calcomp ATL (which handles conventional magnetic tape reels), existing tape drives, some aspects of magnetic tape management software already in use,^{4,5} and other software yet to be developed. It may, after considerable system development effort, exhibit enough of the characteristics of an ideal system to be a worthwhile endeavor, should a more complete procurement prove unattainable. A similar approach has been proposed by Lawrence Berkeley Laboratory.²

MASS STORE RESEARCH AND DEVELOPMENT

Some of the future possibilities and trends indicated by the cited surveys are presented in Table 3, which lists the type of

Table 3
Future Recording Media

Medium	Years	Developer	Bit density (Mbits/in.)	Implied speed (Mbytes)	Comments
Photographic	1960's	IBM	1.3	0.25- 1.5	Nonerasable
Magnetic	1970's	IBM	2.5	1 -10	Erasable
NRZI		CDC	(625x400		
PI		Calcomp	tracks)		
Video		Ampex			
Magnetic	1980's	Bell	1*	2.5 -10	Read & written in
bubbles		IBM			shift loop
					Cost a big problem
Optic &	1980's	Ampex	100	10	Erasable
Magneto optic		Honeywell			Lasers write onto
		IBM			thin film
		PI			Utilize thermomag-
					netic coercivity
					change effect
					(except for PI)
Optical	1980's	Kodak	600	>>10	Marriage of con-
Video		RCA			sumer & digital
Disk		Xerox			technology
					Nonerasable

* At present (3 microns/bit eventually).

storage medium, its probable decade of appearance, and the organizations that have made significant contributions to its development. The expected performance is characterized by the bit density and recording/readback speed parameters. To provide historical perspective, the IBM photographic chip store of the early 1960's is included.

Table 3 serves to highlight the more promising techniques expected in mass store development in the next few years. It appears that optical physics will play an important role in devices of the future. Coherent light magneto optic effect systems appear to be promising for erasable storage applications. Video recording methods are being developed for relatively cheap permanent storage applications. Holography methods have great advantages because of their inherent redundancy, but development has been slow and

several larger companies have reduced their interest recently. The mass store field has clearly attracted a number of large corporations, which should influence the rate of development. Much of this development is currently aided by various federal agencies seeking more cost-effective ways to store satellite-derived mapping data, with 10^{15} -bit requirements.

THE FUTURE COMPUTING ENVIRONMENT

An LMS facility is a long-term investment. One must ensure that it is flexible enough to fit into whatever the future computational environment at BNL is likely to be. One fundamental question is the future mix of larger central computing facilities, in view of recent trends in mini- and midi-computers and the amazing growth of microprocessors and personalized computing. Many persuasive arguments still favor a strong central facility. Certain economies of scale that first brought about large facilities may be expected to hold, especially in the sharing of larger peripherals such as LMS. The need for increased reliability implies a redundant configuration that distributed single machines are unlikely to provide. The human resources concentrated at a central facility are at least as important as the hardware located there. Overall systems maintenance, library support, and computer operations problems increase when the computing capacity is dispersed.

What is a reasonable configuration for meeting the growth of such programs as ISABELLE? One possible approach which includes the idea of large machines in a central complex is shown in Fig. 2.

Fig. 2. Local network topology.

Within DOE, computer mainframes have recently been classified by their computational power. The CDC 7600 and the IBM 370-168 and 370-195 are called Class IV machines. The CRAY 1 is Class VI;

the 6600 is Class III; and the VAX 11/780 is perhaps almost Class III. If current BNL computing practices continue, it is estimated that needs will increase tenfold during the next 7 years, with ISABELLE being responsible for the largest part of the growth. In a model with concentrated computing power, this growth suggests at least one Class VI machine. The demand for a decrease in the rate and duration of service interruption would be best met by providing two similar main machines and parallel data paths. Supporting the main worker machines would be an array of Class III machines closely coupled by a high speed network. One of these would be an interactive time-sharing processor for such uses as program development, editing, and graphics. Another would be a data management device, to which the LMS might be attached. Others might service links to compatible data acquisition processors, which would be attached to the various experiments in the ISABELLE intersections and presumably managed by ISABELLE operations. The requisite local networking technology implied by this complex is being developed at DOE laboratories, among which BNL is an active participant, and elsewhere.

Several important questions related to mass storage were raised at this Workshop, for example, whether it is feasible and desirable to capture the "raw" physics data in the experimental halls and insert it directly into an LMS. Also, the data storage needs in most experiments fall fairly obviously into two categories, archival and dynamic. The primary input data tend to be computed once or twice, and then to be retained or "archived." Most of the reduced data are used somewhat more often, in summary and various physics interpretation procedures. As they are often wanted on rather short notice, they are relatively dynamic. Each type of storage

Table 4

LMS Working Group:

B. Arbeit	A. Peskin
S. Aronson	D.A. Ravenhall
C. Benkovitz	A. Scott
J. Niederer	N. Schumburg
M. Losquadro	

Programmatic Justification
Subcommittee:

D.A. Ravenhall
S. Aronson
C. Benkovitz

Vendor Subcommittee:

N. Schumburg
B. Arbeit
M. Losquadro
A. Scott

Specifications Subcommittee:

A. Peskin
B. Arbeit

can be best handled by a somewhat different set of data management methods. This distinction applies also to a large part of other BNL data storage needs. Perhaps two classes of mass storage strategies, rather than a single larger one, might be a better approach. These and related options will be studied in the coming months.

REFERENCES

1. ADP Long-Range Plan, 1979-1984, BNL, March 1978.
2. DOE/NCAR Mass Storage Workshop, National Center for Atmospheric Research, Boulder, CO, May 1978.
3. Special Issue on Large Capacity Digital Storage Systems, Proc. IEEE 65, No. 8 (1975).
4. S. Heller and L. Linton, Externals on Archiving, AMD Doc. File No. 20.5.2, BNL, Jan. 1978.
5. D. Stampf and R. Horwitz, SUPEROX Users' Guides, AMD Internal Memorandum, BNL, June 1978.