TECHNICAL MEMORANDUM

34

Review of computer mass storage systems

N. Storer

Operations Department

July 1981

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen

This document is an attempt to present the current 'state of the art' of mass storage systems (MSS). The term 'mass storage' refers to systems with a capacity in the region of about a tarabit (10^{12} bits).

Historically, there have been few of these systems. The main reason for this is that potential users have had to design and implement the software required to use the various pieces of hardware, which have from time to time become available. MSS became more widely used with the advent of the tape cartridge systems implemented by IBM (3850) and CDC (38500), each being supplied with the necessary software. Earlier equipment such as the IBM 1360 Photostore and the Ampex Terabit Memory Systems were never produced in quantity (in fact, less than 10 such systems were sold), and are no longer supported, though several are still in use. So, at present the 3850 and 38500 dominate the market.

However, some very interesting projects are under development in several laboratories, employing optical disk technology. These projects have developed as a spin-off from the commercial/entertainment world of optical disks. Optical disks will be capable of holding between 10^{10} and 10^{11} bits each. The hardware should arrive on the scene within the next few years but obviously, we will have to wait for some time after this for an integrated system to be brought forth.

In the meantime, it is possible to make better use of existing ½ inch 2400 feet tapes by means of the Calcomp Advanced Tape Librarian (ATL). Various sites have used this as the basis for an interim solution to their MSS needs and most find it a very cost effective solution. It is cheaper than the 3850/38500, as it is a mixture of new equipment (ATL) and existing equipment (tapes and drives).

Along with developments in MSS came developments in the linking of MSS devices to computer networks. Two such developments are the Network Systems Corporation Hyperchannel and CDC's Loosely Coupled Network (LCN). The Hyperchannel has been employed by various sites to connect machines and devices. One vendor, MASSTOR Systems Corportation, produce an integrated MSS based upon a MASSTOR file processor, Hyperchannel and various combinations of Calcomp ATL, IBM 3850, tapes and disks. In fact, the system is very open-ended and any storage devices available for IBM or IBM Plug Compatible equipment can be included.

Annex I of this paper is a report of an IEEE symposium on MSS.

The digest of papers which were presented at that symposium is now available from:

IEEE Computer Society
10622 Los Vaqueros Circle
Los Alamitos
California 9022
U.S.A.

ANNEX I

Report on the Fourth IEEE Symposium on Mass Storage Systems

(The Gap between MSS Products and User Requirements)

ANNEX I

4th IEEE Symposium on Mass Storage Systems, Denver, Colorado,
15 - 17 April 1980

1) This opened with a talk by Stephen Miller (SRI International), the main topic of which was the lack of standards for data movement between mass storage and main memory. The point being put forward that until such standards are defined, mass storage will not 'take-off' in the same way that telecommunications equipment was only used by a specialised few until communications protocol standards were defined. A proposal was put forward that the committee should look into the feasibility of designing such a standard.

2) Lynn Shirley (MASSTOR) gave a resumé of the state of the mass storage 'scene'. The impression gained seemed to be that nothing much has happened since the last IEEE symposium in 1976. Ampex no longer support the TBM. Several microfiche storage devices had been produced, but for special projects only. The optical disk is still not with us, 'but just around the corner'. Optical disks would have the capacity of about 10^{11} bits each (50 x 844) and would cost $200K to $300K per system (disk and drive). A general complaint was made that up to now mass storage backup and archiving systems have been left to the user to design, a full dataset management capability was required.

3) Jim Thornton (NSC) gave a talk on the NSC hyperchannel as a means of shipping data between mass storage devices and several mainframes. Lee Johnson (CDC) gave a talk on CDC's MSS. Neither had anything new to say.

4) Les Burns (OMEX - formerly Precision Instruments) gave a talk on OMEX's latest project which is based on magazines holding slides. The 4" x 4" slides each hold about 10,000 documents (3×10^9 bits/slide). There are 100 slides per magazine (this will later be upgraded to 2000 or even 6000) and 5 magazines per system, making 1.5×10^{12} bits in total (10^{14} later?). Slides of a cumbersome prototype were shown, and plans to make the system less cumbersome and more reliable were explained.

5) John Bock (MASSTOR) talked about the MASSTOR system, which is connected via links using the NSC hyperchannel to the mainframes. The system consists of a minicomputer with access to a disk and calcomp ATL (Advanced Tape Library). The disk is used for staging files from tape, which are then shipped across to the mainframe requesting the file. A large amount of buffering in the file processor goes on. The first system is due for release by the beginning of August this year.

6) Dave Dodd (STC) gave an uninteresting talk on STC's plans on the MSS market. There was a vague reference to an STC disk drive holding 1000 Mbytes to be announced soon, but he was unwilling to comment further.

7) Donald Hertzog (RCA) gave a very interesting talk on RCA's optical disk project which is in an advanced stage of development. An engineering model under test has the following characteristics:

Capacity:	2×10^{10} bits (laser diode) or 1.5×10^{11} (gas laser)	
Access:	500 msec max displacement	
	200 msec close displacement	(~10 tracks)
	80 msec min displacement	(next track)
Transfer rate:	100 Mbits/sec (2 channels)	

Systems would be available within 2 or 3 years, costing $20K to $50K. The number of channels would be upgraded to 8 (320 Mbit/sec rate). It is envisaged that a 'juke-box' type system with up to 100 disks would be produced, costing $200K to $300K. The disks themselves would be cheap at $10 to $200 depending upon the demand.

8) Leonard Laub (EXXON) gave an extremely interesting talk on their optical disk project which is nearing completion. This system uses equipment initially developed by Thomson-CSF (France) and is planned for delivery to a customer by early 1982. The features of the drive are:

Capacity:	6×10^{10} bits
Warm up start:	15 minutes
Access:	50 msec Average
	5 msec Local seek (10 tracks)
	1 msec Min. displacement (next track)
Transfer rate:	15 Mbits/sec.
Soft error rate:	1 in 10^{11}
Hard error rate:	1 in 10^{12}
Physical dimensions :	4' x 2' x 3'

The cost would be $41K per drive ($36K for the simple version B1), $34K for the storage control unit. The disks themselves cost $350 each, more than the RCA disk because they are specially pre-formatted.

9) Gerry Duggan (Harris Corporation) gave an interesting talk on MASTAR (Mass Archival Storage and Retrieval). This is based upon microfiche technology. It is Harris' response to a requirement by NASA-MARSHALL for a system with the following characteristics:

10^{13} bits with 10^{12} online
expandable up to 10^{15} bits
10 yr archival life
1 in 10^{9} bit error rate
50 Mbits/sec transfer rate
15 second access time (per 10^{12} bit group)
operational by December 1981

System characteristics:

Rotary fiche format, silver halide film
Laser spot recording
2 x 10^{9} bits/fiche
1024 fiche/carousel (2' long x 1' diameter)
50 Mbit/sec transfer rate
7 second access time (fiche to fiche)
½ second access time (within a fiche)
>10 year lifetime
Replication offline via contact printing

An interesting point is that the angular speed of the fiche, which rotates, is not constant: it is 1 rev/sec for the innermost track which holds 2.4 x 10^{7} bits, and 2 revs/sec for the outermost track which holds 7.9 x 10^{7} bits.

10) David Ullmer (EROS Data Centre) gave an interesting talk about acquisition of satellite data from the LANDSAT project. They will have a requirement for 10^{15} bits of storage by 1987, and they visualise this to be in the form of optical disks. Several important points were made:

each disk must be self-contained, hence each must have its own catalogue recorded;

each disk must have a life of at least 20 years and it must be possible to update the data on the disk at any time during those 20 years.

11) Bill Crittendon (NCAR) talked about the Ampex TBM system and the file management system written by NCAR to make effective use of this mass storage media. It is clear from the discussion that NCAR have put a lot of effort into designing this system, and keeping track of which tapes hold which data, also clearing up and compressing tapes is done automatically. Even so, the system needs at least 1 member of staff permanently working on the housekeeping side of things.

12) Bill Collins (LASL) gave a talk about CFS, the Common File System. The system has a lot in common with MASSTOR's. There is a separate computer looking after the system. It has a hierarchy of file storage: Disk-MSS-offline storage cabinets. The computer used is an IBM 370/148 which unfortunately has a very poor availability record. The MSS is IBM's 3850 which has an excellect availability record >99%. The smallest amount of data that can be transferred is the file. There is no file conversion done (transparent) and file transmissions are multiplexed. Complete directory recovery is possible in the event of a failure. Most of the active files reside on disk (5×10^{10} bits), the 3850 holds the next most regularly used files (1.5×10^{12} bits) and the offline storage system is mainly used for archival storage ($>10^{13}$ bits).

13) James Baker (LBL) gave a talk on the possibility of recording whole libraries on microfiche. The only information I could glean from this is that a normal length novel can be held in 10^7 bits.

14) Jim Boswell (Bank of America) gave a talk on BA's future needs. By 1984 they expect to have:

11000 terminals hooked onto
23 mainframes (IBM 3033's)
2000 3350 disk drives
500 6250 tape drives
50 2305 drums

The talk was not really relevant to the symposium.

15) Patrick Savage (Shell Development Corp.) talked about seismic computing done on UNIVAC 1100's at Shell. Last year they purchased 5 Calcomp ATL's (20 drives) to replace 150 tape drives on the Univacs. They found that this is an excellent medium-term solution to mass-storage problems. Of course, it also required a file management system to be designed, but this was kept fairly simple. They get through about 2800 tape mounts/day on average and 4000 per day maximum. One complaint is that the capacity of the ATL is too small. It could have been better designed to hold perhaps 10 times as much.

16) George Michael (LLL) gave a very interesting talk on the LLL proposal for a mass-storage system to replace the now phased-out IBM photostore. The proposal is for a system based on Digital-Optical techniques (optical disk) holding 10^{11} bits in 1981 upgraded to 10^{13} bits in 1983 to 10^{15} bits in 1985. The device controllers must be computer-based (intelligent) with 1 per interface channel. There must be 2 independent channels. The volume-capacity 10^9 - 10^{11} bits with a lifetime of >20 years (10^5 load passes). It must be possible to write to a

volume at any time during its lifetime. Each unit should have a minimum of 3 stations (1 read-write, 2 read only). Read-only should not be able to accidentally write. It should take less that 5 secs to mount a full volume and have it ready for access. File access, once mounted should be >1 or 0.1 sec. There should be a sustained rate of >5 Mbits/sec, a peak one of >40 Mbits/sec. There should be no impact on these rates by non-data accesses. There should be a bit error rate of 1 in 10^{11} or 10^{12}, and the device should report the error to the block in error. It should also provide the bad data if requested.

Summary

The impression gained from the meeting seemed to be that optical disks are the way forward. The Calcomp ATL appears to be a very cost-effective stop-gap measure, as optical disks are still 2-3 years off. The standardisation problem has yet to be tackled. File management systems for mass storage are in their infancy and need urgent attention. The non-rewriteability of optical disks is not necessarily a drawback, in some cases it is a positive asset.