Description and use of Electron Beam Accessed Memory systems

by Donald O. Smith†

EBAM systems having capacities in the range of 256 Mbytes, access time of less than 100 μsec, and 2 Kbyte read transfer times of 424 μsec, are currently under development. The unique technical features of these systems are described and a brief discussion of potential applications in computer systems is given. Future EBAM systems of much greater capacity are predicted.

1. Introduction

Electron beam accessed memory (EBAM) systems suitable for use as auxiliary storage in general purpose digital computers have been under active development for the last ten years. This activity has been based on fundamental advances in electron optics and memory materials which occurred in the early 60s. The advance in optics is to use two stages of deflection in order to access 100 to 1,000 times more resolvable spots that can be accessed using a single stage of deflection.[1] The advance in memory materials was due to the discovery that storage or removal of positive charge in the oxide of a MOS capacitor can be accomplished by simultaneous application of an electron beam and a charge or discharge bias across the oxide.[2] The present cycle of EBAM development started in the early 70s with the subsequent demonstration of practical methods of reading out the presence or absence of stored charge.[3,4,5,6] Operation of the memory target is described below.

EBAM development has now reached the stage that commercial memories with capacities in the 256 Mbyte range can be expected in the early 80s. A description of the overall features of these systems and their application in computer systems is described below. In conclusion, a brief survey of potential future developments in EBAM technology is given.

2. System uses of EBAM

2.1 Block diagram of an EBAM system

Figure 1 is a block diagram of the principal function elements in an EBAM system connected to a larger host system. Three major distinct elements of the EBAM are

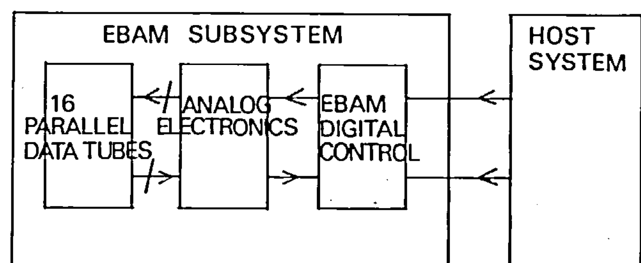

Fig. 1 Block diagram of an EBAM subsystem connected to a host system.

identified, namely:
(i) 16 parallel data tubes;
(ii) analog electronics directly interfacing with the tubes; and
(iii) digital control electronics which manages the host system.

2.2 EBAM parameters which determine performance and system use

The principal EBAM parameters which determine performance and system use are storage capacity, and service time; the latter being determined by access time, read and write data rates and data format. These parameters for a Control Data EBAM development system (D3 system) are given in Table I.

The capacity of a single D3 tube is 128 M bits and the D3 EBAM system consists of 16 such tubes for a total system capacity of 256 M bytes. Data is formatted in 1,024 bit blocks per tube corresponding to 2048 bytes per system. Access time, measured from the time that a request for service arrives at the EBAM subsystem from the host system (see Fig. 1) until the start of data transmission from or to the EBAM is 80μs for read and 58μs for write.

†Control Data Corporation, Lexington, Mass 02173.

Table I D3 EBAM parameters which determine performance and use in a computer system

Capacity and Format	
M bytes/system	256
M bits/tube	128
Data tubes/system	16
Bytes/block	2048
Access Time (μs)	
Read	80
Write	58
Data Transfer	
Block Transfer Time (μs/block)	
Read	424
Write	1351

Only about half of this time is required for deflection of the electron beam, the rest of the time being used mainly to read and error correct the first line of data.

The read and write block transfer times given in Table I include time for reading preamble and error correction bits plus time for interline delays and is the result of 16 tubes operating in parallel at 4 M bits/sec per tube. The per tube write rate is one-third the read rate resulting in a system write rate of $^8/_3$ M bytes/sec. The data format organises data into blocks of 1,024 data bits per tube corresponding at the system level to 2,048 bytes. At the system read rate of 8 M bytes/sec the block read time would then be 256μs. However, interline delays, and the time required to read preamble bits and error correction bits, increases the block read time to 424μs. Since the maximum write time is one-third the maximum read time, the block write time is very nearly $3\times445=1{,}335$μs (actually 1,351μs due to the time required to read and write the preamble).

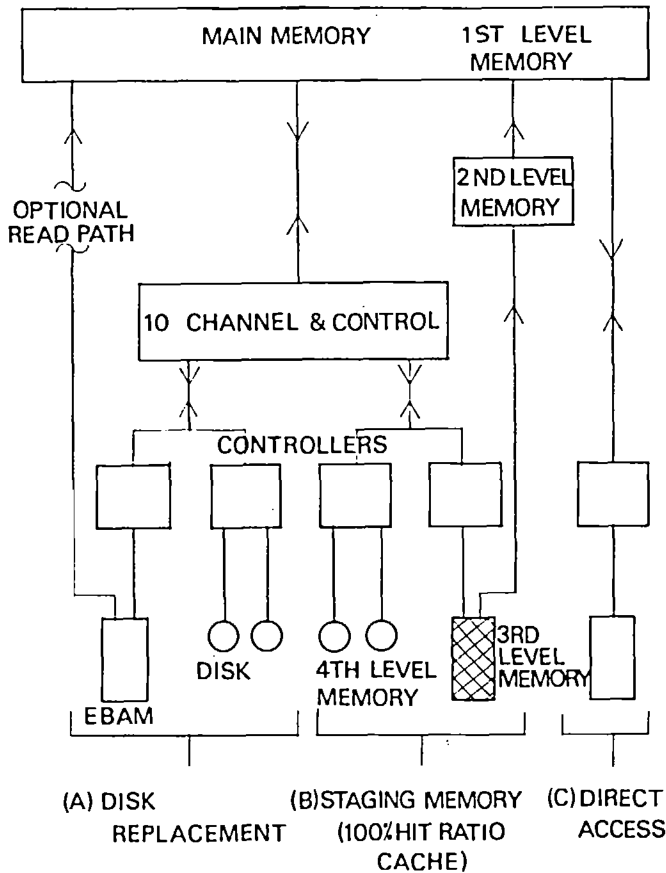

Fig. 2 EBAM used as (a) disk replacement, (b) staging memory, or (c) direct access memory.

2.3 Functional system uses of EBAM

The performance parameters given in Table I suggest that from a system point of view EBAM be thought of as a nearly zero-access-time high-data-rate disk. Hence the simplest initial system use would be as a disk replacement. More sophisticated uses as a cache for a disk, as a staging memory, and as a direct access memory are also possible. These functional applications are summarised schematically in Figs. 2 and 3 and briefly described below.

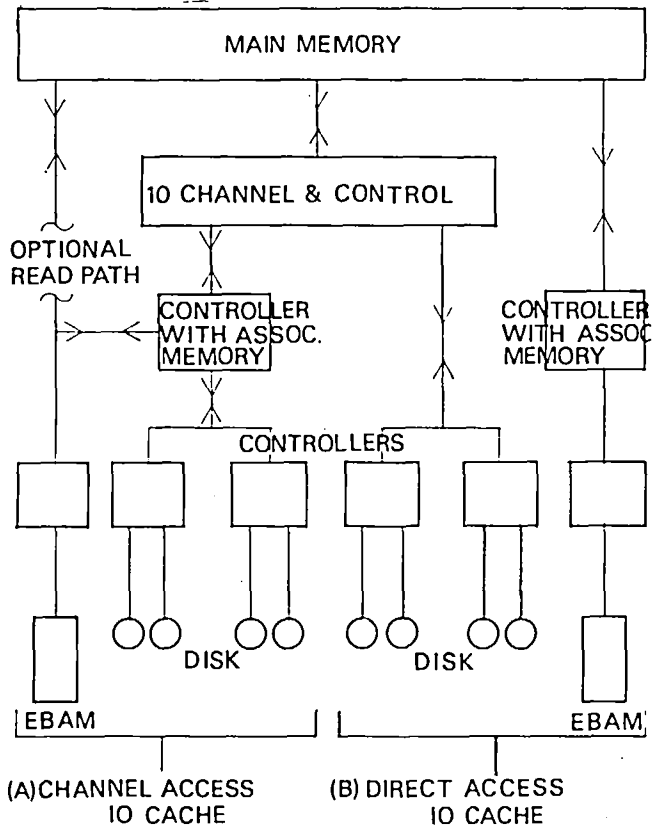

Fig. 3 EBAM used as a cache for disk (I/O cache): (a) channel access, and (b) direct access.

Disk replacement. The simplest use for EBAM is as a direct replacement for disk in Fig. 2a. However, in most of the software operating systems presently used the time to process a request to I/O can easily mask much of the reduction in hardware access and service time to EBAM. In an exemplary system which manages I/O at the channel level the software overhead is 5 ms for sequential files and 23 ms for random files.

In such a circumstance there is clearly no advantage in using an EBAM. However by moving the I/O control to the CPU these times can be reduced. Typically the software overhead times would be reduced by a factor of 10 to 100. Furthermore it is possible to design software operating systems which reduce the random file overhead time to the symmetrical time. Hence steps can be taken which will reduce I/O software overhead time to low enough values that advantage can be taken of the fast access time of EBAM.

Cache for disk (I/O cache). EBAM used as a cache for disk is shown schematically in Fig. 3. The cache function is depicted by the presence of associative memories which check an incoming address for residence in the cache EBAM. Cache use of EBAM is graphically indicated by single cross-hatching.

Channel access I/O cache. In this configuration (Fig. 3a) requests for information go through the I/O channel and if the desired information is in EBAM access to disk is avoided and the resulting service time is reduced. If the EBAM read data rate exceeds the I/O channel rate, then an optional read path direct to the main memory might be provided. Cache management could be either at the CPU level or at the I/O channel level depending on the amount of intelligence built into the channel. Of course software overhead time must be reduced to be comparable to EBAM service time.

Direct access I/O cache. Better service time can be obtained by providing direct access to the EBAM cache as shown in Fig. 3b. In this case channel delays are completely avoided if the desired information is found in EBAM. Cache management in this case would be at the CPU level.

Staging memory. Use of EBAM as a staging memory in a four-level hierarchy is shown in Fig. 2b. At the start of a job all of the information required to service this job is transferred from disk to EBAM. During execution EBAM then serves as a 100% hit ratio cache for the disk (indicated graphically by double cross-hatching); the second level memory then operates as a traditional cache in front of EBAM.

Direct access from main memory. A conceptually simply way of using EBAM is as a direct extension of main memory as shown in Fig. 2c. In this configuration all channel delays are avoided and maximum advantage is taken of the fast access time of EBAM. Hardware and software implementation of this configuration is also relatively simple.

2.4 Performance improvement of host systems using EBAM

Simulation of queuing model studies of host system performance with the inclusion of EBAM in the various configurations described above are needed to provide detailed data for system design. However, some qualitative observations are possible as to the type of host systems which will be benefited and the impact on operating system software corresponding to the EBAM configurations described in Table II. The first uses of EBAM are expected to be in general purpose systems as disk replacement (Fig. 2a) or as channel access I/O cache (Fig. 3a), mainly because of the minimal impact on software.

Direct access, either as an I/O cache or main memory extension, will take the greatest advantage of the fast access and service time of EBAM but are not likely to be among the first applications due to the substantial impacts on operating system software. The first application of direct access I/O cache is likely to be in large scientific machines.

Use of EBAM as a staging memory (Fig. 2b) is particularly attractive for large multi-user interactive systems and may become one of the earliest applications of EBAM.

3. Basic elements of an EBAM tube

3.1 Memory target

The memory target is a MOS structure shown schematically in Fig. 4. Positive charge is stored or removed from a silicon dioxide film under the combined action of the electron beam and pulsed bias fields applied across the oxide. Built-in ten volt avalanche junctions permit the application of both polarities of voltage across the oxide.

Table II Applications of EBAM in different types of host systems and the corresponding impact on the software operating system necessary to introduce EBAM.

		Type of host system	*Impact on operating system software*
Disk replacement		General purpose	Small
IO cache	Channel access	General purpose	Very small
	Direct access	Large scientific	Large
Staging memory		Multiuser interactive	Moderate
Direct access from main memory		General purpose; Large scientific	Moderate

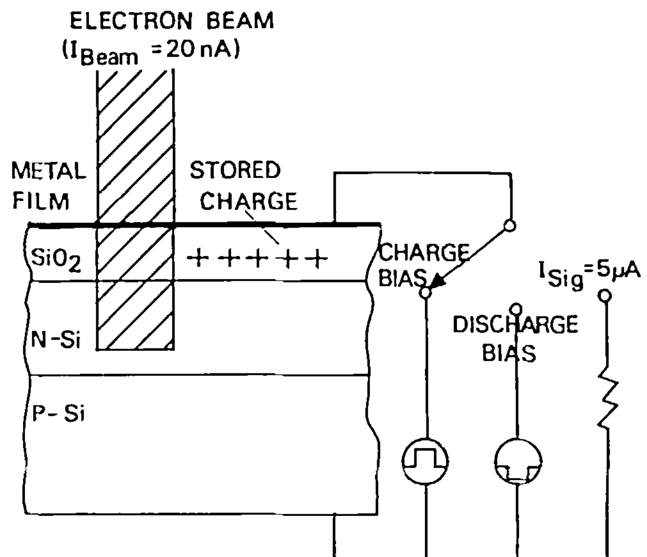

Fig. 4 Memory target construction and operation

Read-out employs the same electron beam used for writing by utilising the electron-hole pairs which are generated by this beam in the n-type silicon just below the oxide film. Stored charge in the oxide modulates the silicon surface potential so as to repel the beam-generated minority carriers (holes) which are consequently collected by the underlying reverse-biased np junction. In regions of oxide in which there is no stored charge beam-generated holes are not collected by the junction but instead diffuse to the Si-SiO_2 interface and interact with interface states to recombine with electrons. The combined process of electron-hole pair generation and subsequent modulation of hole collection by the stored charge results in a signal output current several hundred times greater than the input beam current. In effect a low-noise transistor amplifier is located at each of the stored bits.

The storage and read-out described above does not require any fine structure in the plane of the memory target and only a small number of leads must be brought out from the target to accomplish writing and sensing. These are the basic reasons for the extremely high bit density (~80 Mbits/inch²) and low cost of the D3 memory target.

It is of importance to note that charge storage in the SiO_2 layer of the target is highly non-volatile. Physics experiments have demonstrated storage for several years and the total possible time has not been studied. Although the maximum practical storage time for real bits is not yet known, recovery after several hours has been demonstrated.

Information is phase encoded on the target as shown in Fig. 5 in which a bit is represented by a pair of cells in either the sequence charge-discharge (CD) or discharge-charge (DC). In turn a cell is irradiated by a beam having a spot size somewhat less than a cell. In practice approximately four spots are required per bit.

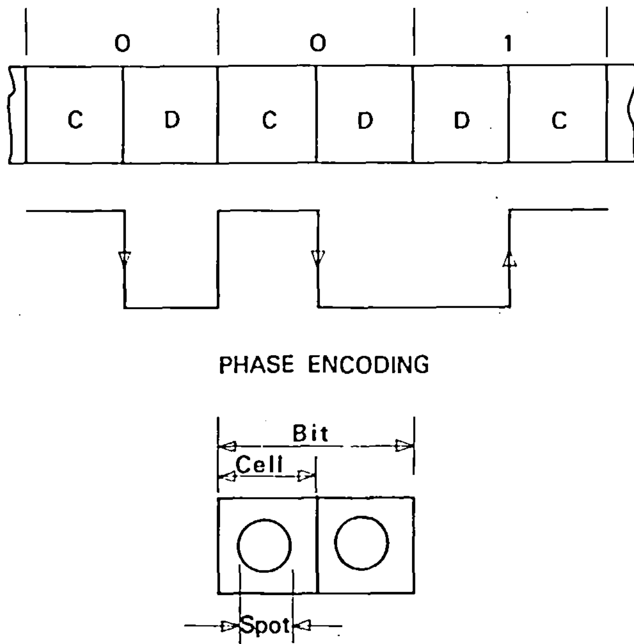

Fig. 5 Method of recording information on the target and terminology defining a spot, cell, and bit.

3.2 Double-deflection electron optics

Using presently existing engineering technology, the number of resolvable spots with sufficient current to operate the memory target of Fig. 4 and which can be accessed by a single deflector is of the order of 16 to 32 million. Since approximately 4 spots/bit are required single deflection optics can yield memory tubes containing up to 4 to 8 Mbits. Attempts to increase the capacity beyond that point encounter two major problems, namely: (1) the deflected spot size increases due to deflection aberrations; and (2) the required accuracy of deflection electronics exceeds present technology.

Both of the above problems can be overcome by the introduction of double-deflection optics as shown in Fig. 6. In this tube electrons are first deflected by a coarse deflector to select one lenslet of an array of lenses containing a total of 1,024 lenslets. The array lens is followed by an array of fine deflectors containing 1.024 deflectors in one-to-one correspondence with the lenslets. In the tube shown in Fig. 6 the final spot is 1.2μm in diameter and there are 128K bits in the field of each lenslet and associated fine deflector. Then since there are 1.024 lenslets the total tube capacity is 128M bits.

Double-deflection tubes have been built and tested at performance levels near the requirements for D3 memory system operation. A program of systematic improvement to reach the requirements of product performance with high yield and reliability is in progress.

4. Multitube EBAM subsystem

4.1 Special data management requirements

Two properties of the MOS EBAM target lead to requirements for special data management in an EBAM system, namely: (1) the target 'fatigues' and hence has a limited lifetime when irradicated with an electron beam; and (2) there is no threshold at which charge is stored or removed from the target but instead the target charges or discharges continuously depending on bias and irradiation.

At the system level the above two target characteristics lead to the problems of target fatigue and

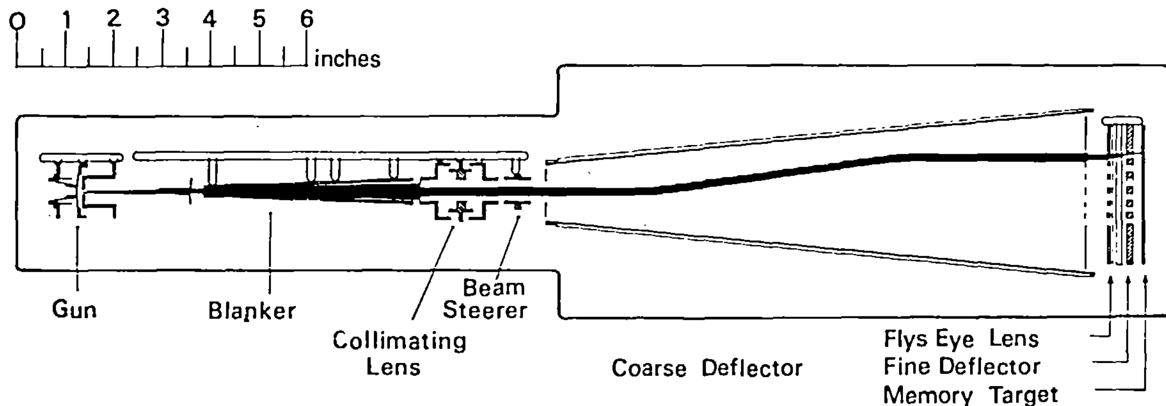

Fig. 6 Double-deflection EBAM tube.

disturbance of data. Methods of data management which overcome these problems are discussed below.

Fatigue. For presently used targets, the fatigue lifetime is measured to be of the order of 0.5 coulomb/cm^2. The lifetime t for one cell repetitively accessed is then given by $t=aq/i$, where a=spot area, q=fatigue charge dosage density, and i=spot current. For the D3 tube $a=(\pi/4)(1.2)^2\times10^{-8}$ cm^2 and $i\times20$ ns so that $t\simeq0.3$ sec, which is not a practical lifetime.

Practical times are achieved by using a data management algorithm which ensures that repetitive or even weighted accesses sufficient to cause fatigue cannot occur at any particular location before occurring at every other location, i.e., on the average the target is accessed uniformly. Assuming for the moment that such algorithms exist, then the target lifetime is extended by multiplying the spot lifetime by the number of spots. Assuming 4 spots/bit and 20% error correction bits then gives the number of spots as $4\times128M/.8=640M$. The lifetime is then $0.3=640M$ sec=6 yrs.

The uniform usage algorithm presently in use is a simple permute algorithm described in Fig. 7. In this case after every tenth write request to EBAM data is written into an empty block from the adjacent block in a list of addresses ordered from 1 to N. In this way the empty block is gradually permuted through the memory until the initial starting location is reached. At this point all physical data locations have been shifted by one in the list. Hence, in order to map a logical address into a physical location, two numbers are required, namely: (1) the position of the empty block, and (2) the number of times (modulo N) that the empty block has cycled through the physical addresses.

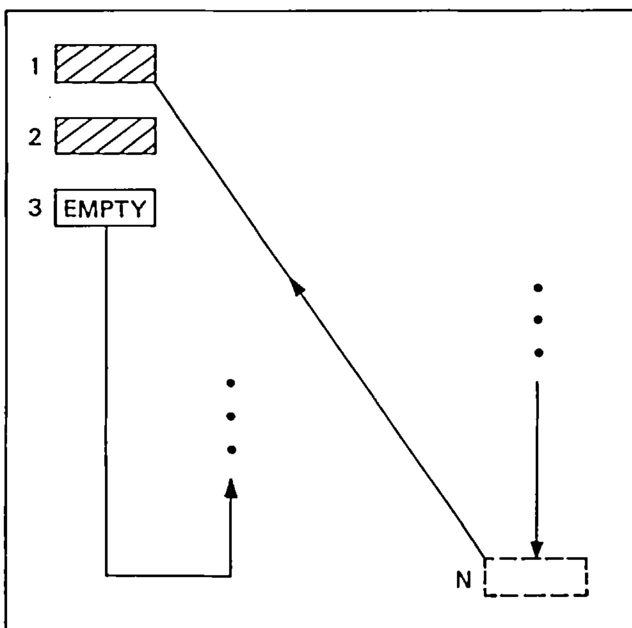

Fig. 7 Permutation algorithm used in experimental EBAM system. The physical address is determined by: (1) the position of the empty block, and (2) the number of times (modulo N) that the empty block has cycled through the memory addresses.

Note that according to the above algorithm the condition to permute does not depend on the particular addresses of the previous ten write operations. Hence a necessary condition that this algorithm be guaranteed to work is that repetitive addressing to a particular block can only result in a small fraction of the fatigue dosage before the empty block arrives to permute the data. For a memory tube with a capacity as large as the D3 tube (128 Mbits) it is necessary to supply an external first-in-last-out block buffer in order to ensure this condition.

The simple permute algorithm described above could be made to fail by a 'test' program which counted write operations and permuted a logical address in the opposite direction at just the right time to keep addressing the same physical location. Such an eventuality can be overcome by introducing random features into the permute algorithm. As an example, permute might occur after 5 to 20 writes, the exact number depending on a random number generator.

Disturb. Since any level of irradiation will change the charge state of the target data can be disturbed by the beam used to read data (self disturb), by beam tails (near-neighbour disturb), and by scattered electrons (far-field disturb). Separate data management methods are used to handle each of these effects.

Self disturb, which can be described as partially destructive read, requires that after some specified number of reads the data be restored. For the D3 tube restore is required after between three to five read operations. A reads counter memory could be provided to manage restore.

Near-neighbour disturb is reduced to a problem of disturbing neighbouring blocks and is handled in two ways, namely: (1) a gap is left between blocks thus reducing the disturbing effect of beam tails, and (2) a block buffer is provided to reduce the number of possible block accesses to a particular block before permutation of neighbouring blocks occurs. This is an example of an additional use of permute, namely to refresh data. The size of the block buffer must be chosen such that permutation of neighbouring blocks occurs before disturb of these blocks can occur. Clearly the disturb buffer and permute buffer are the same buffer with the size determined by whichever requirement is the larger.

Far-field disturb is a small but significant effect which occurs due to scattered electrons. Although studies of this effect are not yet complete, it is expected that the permute operation will also serve to refresh the far-field.

4.2 Error correction

Since several EBAM applications resemble disk applications, it is natural to consider disk performance as a measure for the error rate of EBAM. The uncorrected error rate of a modern disk is typically 10^{-12} errors/bit. In order to determine if this is an acceptable rate consider the interrupt time between errors assuming the disk to be 50% busy and a transfer rate of 5 Mbits/sec. Then the interrupt time is $2\times10^{12}/5\times10^{6}=2\times10^{5}$ sec$\simeq$100 hrs. However, the use of error correction and multiple re-reads serves to increase the interrupt time due to uncorrectable errors by a factor of at least 30, i.e., to 3,000 hrs, with a corresponding error rate of 3×10^{-14} errors/bit. The time between undetected errors is a factor of 10 to 100 times longer,

 continued from page 9

i.e., in the range of 1 to 10 years.

The performance of disk described above sets stringent conditions for comparable EBAM performance. Several factors make the problem harder:

(i) the parallel data rate from 16 tubes each operating at 4 Mbits/sec is 16 times higher than for a single disk head, hence requiring EBAM to operate at 2×10^{-15} errors/bit to match disk performance;

(ii) in the first EBAM products a higher defect density on the storage surface is expected for EBAM than for disk; and

(iii) only one re-read can be counted on in the case of an uncorrected EBAM error.

Although the error correction scheme for the D3 system has not yet been fully decided on, it is instructive to consider the error correcting used on a previous experimental EBAM system (D2 system). In this system a product correction code was used as illustrated in Fig. 8 and which consisted of parallel Hamming correction and serial burst correction. In Fig. 8 one line from each of 16 data tubes plus six Hamming tubes are shown. As these lines are read out in parallel Hamming SEC-DED is performed in real time. For the defect count achievable with present targets the incidence of double Hamming errors is high enough to require the addition of serial burst correction for each separate line as shown in Fig. 8.

The predicted performance of the above error correction scheme would not quite meet the required error rate. However, subsequent data processing which takes account of cross correlation between the two codes would yield the desired error rate. The principle drawback of the product code is the addition of the six Hamming tubes. It is probable that more efficient codes can be found which will reduce the tube count.

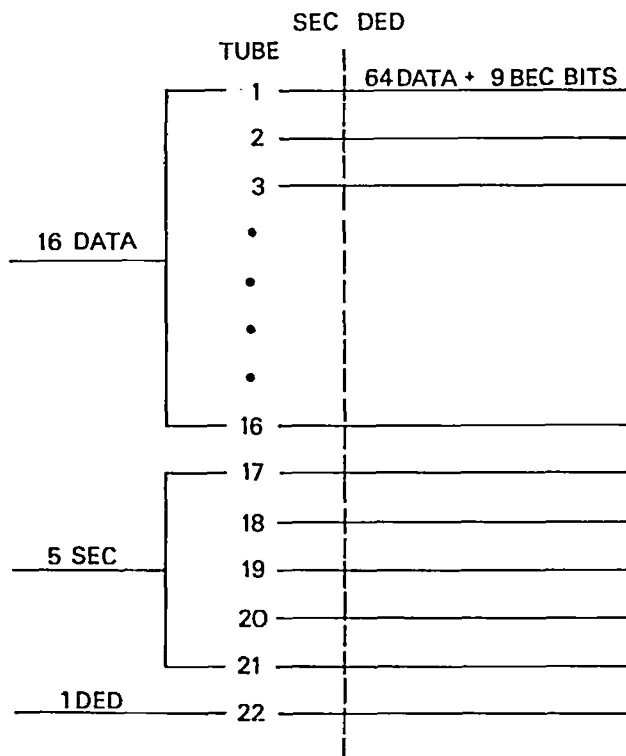

Fig. 8 Error correction used in D2 experimental EBAM system.

5. Reliability and packaging

5.1 Reliability

An EBAM system contains technical elements which are qualitatively different from the usual ones found in computing systems. It is worthwhile to examine these new elements from the standpoint of system reliability. The new reliability-risk areas are: (1) high voltage; (2) precision analog electronics; (3) target defects; (4) array lens defects; (5) cathode life; and (6) target life.

Among these items cathode and target life will affect long term life but are unlikely to affect operating reliability and hence do not properly belong on the list. Their inclusion here is in the interest of explicitly stating the reliability implications of these unusual technical features of EBAM.

High voltage arcs are the principal reliability problem associated with the high voltage used in EBAM systems (−10kV and ≃ −5kV for the cathode and lenses, respectively). Although proper design can make arcing an infrequent event, conservatively it should be assumed that complete elimination of arcs will not occur. Hence the system should be designed such that arcing cannot damage equipment or cause loss of data.

In the event of a major arc which causes system shutdown automatic recovery should occur. Smaller arcs (micro arcs) which do not cause shutdown should also be absorbed by automatic system response without loss of data. The above general system response to arcs have been developed in previous experimental EBAM systems and will be incorporated into the D3 system.

The use of precision analog electronic circuits in EBAM represents a technology which has been consciously designed out of computing systems whenever possible. A notable exception of course occurs in the case of magnetic disks and tapes, which are based on mechanical analog motion. In the case of EBAM, an important simplification is obtained by requiring only relative precision instead of absolute precision whenever possible. For example, the fact that the storage area of the target is unstructured at the bit level allows the use of less than perfectly linear deflection and ramp electronics. A second technique which reduces the need for absolute analog precision is the use of frequent (≃1 sec intervals) automatic calibration of the high voltage and deflection electronics against some common reference such as a zener diode or fiducial marks on a memory target (see above).

By the application of the above general principles and by sound engineering design there is every expectation that the analog electronics required for reliable EBAM operation can be achieved.

Target and array lens defects are conveniently considered together. At this stage of EBAM development it is expected that production yields will be unacceptably low if perfect targets and array lenses are required. Hence system design is proceeding on the assumption that targets with defects which cause errors and array lenses with unusable lenslets will be used in constructing tubes. The problem then becomes one of designing an error correction system which will deliver the required error rate given some specified incidence of target and lens defects. Since the actual relationship between defects and target and lens yield has not yet been fully established, initial EBAM system will be

conservatively designed with respect to error correction capability.

5.2 Packaging

Tube packaging presents unique packaging problem in an EBAM system. Vibration and stray AC or DC magnetic fields are the most critical parameters which must be controlled. For installation in a ground based computer room simple shock mounting of the tube cabinet provides sufficient vibration control. Preliminary studies show that using specially designed tubes, operation in certain military environments should also be possible.

Magnetic shielding requires the use of three concentric magnetic cylinders, the inner two of μ-metal and the outside one of soft iron. The total volume required for tubes plus shielding varies roughly as the third power of the tube diameter which for the D3 tube is six inches. Assuming the error correction scheme of Fig. 8, the corresponding minimum space for 22 tubes has dimensions and volume of $\sim 3 \times 3 \times 6 = 54$ ft^3. Improved D3 tube design could reduce the diameter to four inches with a corresponding space requirement of $\sim 2 \times 2.5 \times 3 = 15$ ft^3.

6. Future developments

6.1 Near and intermediate term

The near and intermediate term future of EBAM depends on tubes of the general design of the D3 tube (Fig. 6). Tubes of this type can reach capacities of up to 1 Bbit, corresponding to a spot size of $\sim 0.5\mu$m. Access time should improve from the present 80 μs down to 5 to 10 μs. Average read transfer rate can be improved by increasing the number of reads between restore (see above), and by increasing the maximum read rate from 4 to 8 Mbits/tube. Simultaneously the write rate would increase from 4/3 to 4 Mbits/tube.

6.2 Long range

The most dramatic long range development which would occur in EBAM technology would be an increase in bit packing density from the D3 level of ~ 10 Mbits/cm^2 (1.2μm diameter spots) to the theoretical limit of ~ 10 Bbits/cm^2 (0.053mm diameter spots). This theoretical capacity limit is based on considerations of the basic physical limits imposed by the MOS target, electron emitters, and electron optics. With respect to cathodes it is assumed that thermal cathodes are replaced by field emitters. The target and optics improvements would be based on evolutionary development of present technology.

Equally dramatic reduction in tube packaging can also be anticipated. By using an array of selectable micro field emitters[7] the need for a coarse deflector can be eliminated resulting in a tube which is only one or two inches long. Hence 4 Bbits could be packaged in a tube one inch in diameter and two inches long. By selecting more than one cathode at a time high parallel data rates could also be obtained. For example, simultaneously selecting eight cathodes would result in a byte rate of up to 8 Mbytes/tube. The per bit storage cost for such a tube should be less than for disk storage and since EBAM storage is non-volatile disks could be replaced by these miniature tubes. Of course this would have a major impact on the computer industry.

7. References

[1] Newberry, S. P., "Problems of Microspace Information Storage", Proc. Fourth Symposium on Electron Beam Technology, 1962.

[2] Peck, D. S., and Schmid, E. R., "Effects of Radiation on Transistors in the First Telstar Satellite", *Nature*, **199**, p.741, 1963.

[3] Huber, E. E., Cohen, M. S., and Smith, D. O., "Electron Beam Detection of Charge Store in MOS Capacitors", *Appl. Phys. Letters*, **16**, p.147, 1970.

[4] Ellis, G.W., Possin, G.E., and Wilson, R. H., "Diode Detection of Information Stored in Electron Beam Addressed MOS Structure", *Appl. Phys. Letters*, **24**, p.419, 1974.

[5] Hughes, W. C., *et al*, "A Semiconductor Non-volatile Electron Beam Accessed Mass Memory", *Proc. IEEE*, **63**, p.1230, 1975.

[6] Speliotis, D. E., Smith, D. O., Harte, K. J., and Arntz, F. O., "Electron Beam Memories", ELECTRO/76 Boston, Mass., May 1976.

[7] Spindt, C. A., Brodie, I., Humphrey, L., and Westerberg, E. R., "Physical properties of thin-film field emission cathodes with molybdenum cones", *J. Appl. Phys.*, p.5248, 1976.