Fermi National Accelerator Laboratory

FERMILAB-Conf-89/132

Exabyte Helical Scan Devices at Fermilab *

Penelope Constanta-Fanourakis, Ken Kaczar, Gene Oleynik,
Don Petravick, Margaret Votava, and Vicky White
Online and Data Acquisition Software Groups
Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510 U.S.A.

George Hockney
Theoretical Physics Group
Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510 U.S.A.

Steve Bracker
University of Toronto
Toronto, Ontario, Canada M5S 1A7

Jussara M. de Miranda
Centro Brasileiro de Pesquisas Fisicas
22290 Rio de Janeiro, RJ, Brazil

May 1989

* Presented at "Real Time Computer Applications in Nuclear, Particle, and Plasma Physics," Williamsburg, Virginia, May 16-19, 1989.

Exabyte Helical Scan Devices at Fermilab(*)

Penelope Constanta-Fanourakis, Ken Kaczar, Gene Oleynik, Don Petravick,
Margaret Votava, Vicky White
Online and Data Acquisition Software Groups
Fermi National Accelerator Laboratory

George Hockney
Theoretical Physics Group
Fermi National Accelerator Laboratory

Steve Bracker
University of Toronto, Toronto, Ontario

Jussara M. de Miranda
Centro Brasileiro de Pesquisas Fisicas

I. INTRODUCTION

Exabyte 8mm helical scan storage devices are in use at Fermilab in a number of applications. These devices have the functionality of magnetic tape, but use media which is much more economical and much more dense than conventional 9 track tape.

II. HELICAL SCAN TAPE SYSTEMS

Helical scan recorders write very narrow tracks at an acute angle to the edge of tape in a diagonal pattern on the tape. This technology results in very high tracks per inch which, when combined with a high linear flux density result in extremely high areal density and data storage capacity. A single 8mm tape cassette, comparable in size to an audio cassette, may hold over two gigabytes of data, and costs under seven dollars.

The device in use at Fermilab is the Exabyte EXB-8200 CTS (Cartridge Tape System). The Exabyte drive uses the Small Computer System Interface (SCSI) bus for command and data transmission. Two SCSI Interface alternatives are available, standard single ended, and differential SCSI Interface. The single ended interface is most used at Fermilab. If a SCSI port is not available on a system then an adaptor board from that system's particular bus to SCSI is required. An Exabyte system (drive plus adaptor) targeted for either VME, Q-bus or Unibus application costs below $6,000 with space requirements (including the power supply) no greater than a large shoe box.

III. USES FOR HELICAL SCAN TAPE SYSTEMS

8mm tape technology offers smaller size (easily stored and shipped), economy, larger capacity (fewer tape mounts) than 9 track tape. Additionally, this technology retains the most desirable characteristic of 9 track tape, the ability to be read on a number of computers. Unlike the international standard 9track tape, it has the disadvantage however of a sole supplier.

During the next accelerator run experiments will collect a large amount of data. Although the amount of data written by an individual Experiment at Fermilab varies, during the next fixed target run, a typical experiment is expected to write enough data to fill 6000 9 track, 6250 b.p.i. tapes, and one experiment intends to take enough data to fill 200,000 such tapes.

Analysis of all these raw data will generate a large number of Data Summary Tapes (DSTs), which may be needed in replicates, since these are sometimes analyzed at an experimenter's home institution.

The Theoretical Physics Group at Fermilab performs Lattice Gauge Theory calculations on four-dimensional lattices and generates enormous amounts of data to represent gauge configurations and quark propagators. Since these configuration and propagator objects are used many times by subsequent analysis programs (not always in the order they are created) a large data-storage system is needed. A typical run will involve 50 configurations with 5 propagators created from each configuration. Each of these objects is on the order of 10 Megabytes, so an Exabyte tape file system has the size needed to do the job.

(*)sponsored by DOE Contract No.DE-AC02-76CH03000

Computing systems are acquiring significant amounts of magnetic disk space. These systems need to be backed up frequently, preferably with minimal tape mounts and minimal operator intervention.

IV. INITIAL EVALUATION OF EXABYTES

Because this technology appeared attractive, a series of evaluations were conducted [1]. The Exabyte evaluation included testing of several Q-bus and Unibus adaptor boards and the Ciprico Rimfire 3503 VME adaptor board.

Evaluation results are summarized below, and are independent of any particular adaptor tested.

- o Sustained Transfer Rate 240 kbytes/sec
  - Time to reach Write Sustained Rate at beginning of tape --> 32 sec
  - Time to reach Read Sustained Rate at beginning of tape --> 5.5 sec
  - Transfer rate for large blocks showed small decrease, when 3 drives were daisy chained.

- o Pausing while writing:
  - Pausing for less 15 sec --> no time delays
  - Pausing for more 15 sec but less than 60 sec --> a delay of 1.2 sec for writing 256 kbytes
  - Pausing for more 60 sec --> a delay of 4.6 sec for writing 256 kbytes

- o Pausing while reading:
  - Pausing for less 60 sec --> no time delays
  - Pausing for more 60 sec --> a delay of 3.5 sec for reading 256 kbytes

- o Tape Capacity --> 2040 Mbytes
  - Capacity lost to pausing for more than 60 sec, --> 23 Kbytes/pause
  - Capacity lost to tape marks,
    --> 20 Kbytes/tape mark if the drive writes short file marks
    --> 50 Kbytes/tape mark if the drive writes long file marks.

- o Using VMS BACKUP, the transfer rate is --> 50 Kbytes/sec.

- o Tape interchangability was not a problem amongst drives and adaptors obtained from various vendors.

Five vendors supplied evaluation adaptor boards that map DIGITAL's Tape Mass Storage Control Protocol (TMSCP) onto SCSI commands:-

DILOG	(Q-bus)
Eakins TD Systems adaptor	(Q-bus, Unibus)
System Industries	(Q-bus)
Summus TD Systems adaptor	(Q-bus, Unibus)
TD Systems	(Q-bus)

Additionally, tests were performed on:-

AVIV (modified DEC standard TS11 driver) (Q-bus)
Transitional Technology Inc. (TTI supplied driver) (Q-bus, Unibus)
Ciprico Rimfire 3503 (with Fermilab software) (VME)

While some problems were detected in both drives and adaptors, the results were promising enough that these devices were brought into use. At Fermilab there are currently about 60 Exabyte drives in use. An active effort to understand the remaining problems and limitations remains underway.

V. USAGE IN THE FERMILAB LATTICE SUPERCOMPUTER PROJECT

The Advanced Computer Program Multiple Array Processor System (ACPMAPS) [2] performs Lattice Gauge theory computations. Its I/O subsystem uses VME, and is currently operating using a single Rimfire 3510 VME to SCSI adaptor with 4 Exabyte drives. The final configuration will consist of 8 adaptor boards supporting 32 Exabyte drives.

The system uses a custom file system, where some files span several 8mm tapes. The objects stored within the file system are large enough that the 64 gigabyte capacity of the eventual Exabyte system is required.

The custom file system and software driver for the tape system is written in the C language. The software driver for the tape system is divided into system independent software, which is transparent to the underlying hardware configuration, and system dependent software, which requires a knowledge of the hardware configuration.

VI. USAGE IN VME BASED DATA ACQUISITION SYSTEM

The Rimfire/Exabyte combination is appropriate for logging data in Data Acquisition systems. Up to 7 Exabytes can be attached to a single Rimfire adaptor, although the SCSI bus bandwidth is consumed by four drives. Multiple adaptors may be used concurrently in a single VME crate thus enabling access to many Exabytes by processor(s) in the same crate. The Rimfire adaptor board may be driven from multiple processes (tasks). Figure 1 illustrates such a data acquisition system.

The PAN-DA data acquisition system contains a complete set of routines to drive the Exabyte using the Ciprico Rimfire 3510 adaptor board [3]. The driver package is an adaptation and extension of the software developed for ACPMAPS. Although it was designed to run on Motorola 68000 series processor boards running a pSOS kernel operating system, the package can be easily modified for use in other environments.

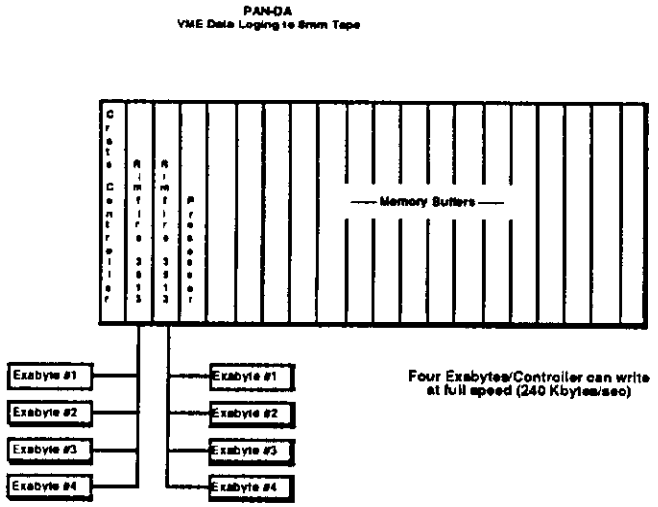

Figure 1.

In addition to supporting the Exabyte SCSI command set, the routines in the RIMFIRE package provide for the initialization of the Rimfire board and Exabyte devices, Rimfire self diagnostics, Exabyte self diagnostics, and internal (package) diagnostics. Exabyte command issuance and completion notification are handled independently in the package. Routines are supplied which poll for command completion. However, the Rimfire board can be set up by the package to generate VME interrupts on command completion. In this configuration, the user must supply an interrupt service routine.

The package provides a tracing facility [4] for pSOS based systems which, when enabled, records a time-stamped history of user initiated transactions with the Rimfire adaptor.

A restriction of the RIMFIRE software package is that there is no protection against multiple processes accessing the same Exabyte. The user is responsible for ensuring that commands are issued to a given Exabyte via a single process or for developing his own locking mechanism but different processors can access different Exabytes on the same Ciprico adaptor simultaneously.

Experiment 791 has developed an alternate set of FORTRAN routines which drive the Rimfire/Exabyte system from ACP 68020 processor nodes running under ACP system software. This experiment plans to log to 32 Exabytes concurrently (distributed amongst 8 Rimfire adaptors in 4 VME crates) during the next fixed target run.

This experiment would not be run if constrained to 9 track tape technology, for 200,000 9 track tapes would have to be written. The requirement for parallel data logging (induced by the 8 Mbyte/second data rate) would result in a tape mount every 22 seconds!

VII. PROBLEMS OBSERVED IN A VME ENVIRONMENT

In a single VME crate 24 Exabytes have been continually exercised simultaneously on 6 Rimfire 3513 adaptors (4 drives on each adaptor). All tests were performed using revision level 4$24 of the Exabyte MX board prom. The following problems were observed:

- o About a dozen occurrences (very rare) of data corruption in Exabyte/Ciprico/VME tape system over a period of a month were recorded. In all cases, errors are confined to a small segment of the record. All failures begin with the insertion of an extra longword into the data. Following longwords are then shifted forward by one place. Finally, a longword is dropped, restoring proper order for the rest of the record.

- o Persistent software hangs (SCSI reconnect timeouts) have been observed around logical end of tape (LEOT) and physical end of tape (PEOT).

- o Incorrect status is occasionally returned near PEOT. The Exabyte will sometimes return normal status for all writes past PEOT. However, it correctly returns PEOT if a filemark is written.

- o Seemingly spurious medium and hardware errors have been seen during writes.

- o Exabytes automatically go into unbuffered mode when the LEOT threshold is crossed; therefore, data transfer speeds drop dramatically. If the current buffered mode is read with a MODE SENSE operation the Exabyte will return buffered mode even though it is writing in unbuffered mode.

VIII. GENERAL USAGE UNDER VAX/VMS

These devices are used as general purpose tape drives for VAX/VMS at many locations at Fermilab. The Exabyte often shares the SCSI bus and adaptor with a disk drive to form a low cost set of peripherals for VAX, microVAX, or VAXstation computers.

The Exabyte drives are convenient for system backup, and often a complete disk backup will fit on a single 8mm tape. VAXstations are being equipped with these drives to efficiently support analysis of Data Summary Tapes.

IX. TAPECOPY FACILITY

Since 8mm tape drives are a good match to many small, powerful computers a demand has arisen to both convert 9 track tapes to 8mm cartridges and to replicate 8mm tapes for distribution to dispersed experiment collaborators.

To meet this demand, the Fermilab Computing Department has implemented a copy facility. This facility has been operational since early December 1988. It is implemented using a microVAX II and a VAXserver 3200 computer. Each system's hardware includes six Exabyte drives, two TD Systems VIKING Q-bus to SCSI adaptor boards and two STC2925 9 track tape drives. Several experiments have availed themselves of this service which compacts up to ten 6250 bpi 9 track tapes onto a single 8mm tape or replicates (up to four times) a single 8mm tape. While this system has been troublesome, at times 110 9 track tapes a day have been copied.

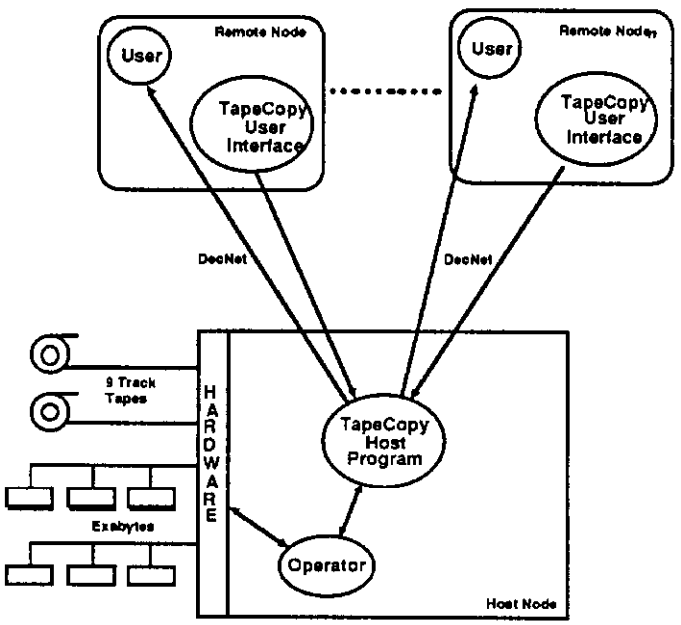

Figure 2. Tapecopy Facility Software

The TAPECOPY software consists of two parts (refer to Figure 2):

- o The user interface which accepts data and submits jobs over DECnet to the host system that has all the appropriate hardware and software to perform the copy operation [5]. Although the software fully supports direct user submission of copying jobs, the current installation of the Tapecopy Facility software requires that the user supplies the facility operators with a written request for the tape copy job. It is the responsibility of the operators to enter the data to the user interface part of the facility and respond to all tape mount requests. If requested, a log file of the job is sent to the user upon job completion (successful or not).

- o The part that performs the actual copy of one media to the other [6]. There are two types of copy that the facility can perform:

  - File copying, for file structured tapes, using the VMS COPY command. Since many 9 track tapes will fit on an 8mm tape volume, the use of VMS copy ensures that the new file headers are correctly constructed.

  - Image copying, for non file structured 9 track tapes and for 8mm tape replication. All data, without any interpretation, is copied to 8mm tape.

X. PROBLEMS OBSERVED UNDER VMS

A number of hardware and software problems have been encountered on many systems. One of the computers in the tapecopy facility serves as a test stand for pursuing these problems, since the stress of simultaneous copy jobs conjures up many sorts of errors.

The problems under VMS are hard to categorize, because of the complexity of VMS. Both a class driver (TUDRIVER) and a port driver (PUDRIVER) are invoked during an I/O operation. Accessing a device that uses the TMSCP protocol is shown in Figure 3.

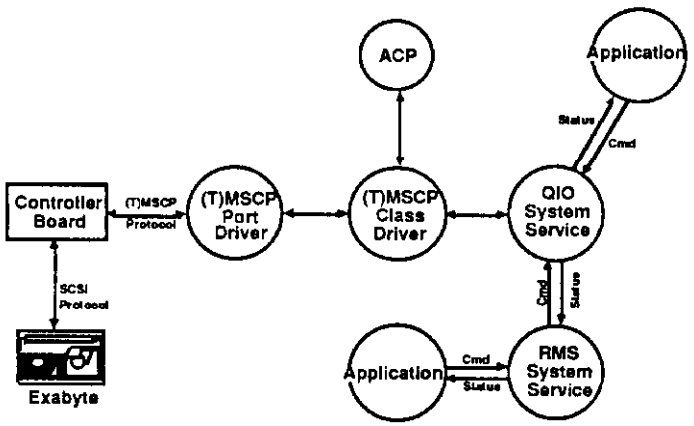

Figure 3. VMS (T)MSCP Protocol

In general, I/O operations may fail to complete, or the tape device may enter Mount Verify state. The tape Ancillary Control Program has been seen to loop at high software priority, hanging the system. Other sorts of hangs, thought to relate to the order of devices on the Q-bus, prevent halting the system by means of the HALT button or the console. The Tape position lost error is occasionally reported after a tape rewind operation. Buffers of data must be word aligned.

XI. DIAGNOSTIC TOOLS

Because of the promise of this technology, we have implemented diagnostics to reproduce and understand the remaining problems with these devices. We are in contact

with the suppliers, and notify them as specific problems are identified.

Under VME, an Exabyte exercising program that runs on a Fermilab ACP computer system has been written. This program performs operations on many Exabytes concurrently in an environment similar to what they face in a high-rate data acquisition system. It writes, reads and checks up to 16 Exabytes attached to 4 adaptors. More than 4 adaptors can be exersized simultaneously by running multiple copies of the program. Generally, tests are run overnight, exercising 24 Exabytes at a time; even very rare errors can soon be characterized in this manner.

The RIMFIRE package contains a run time diagnostic facility, including real-time tracing of operations on the adaptor.

A diagnostic program has been written for the Ciprico Tapemaster controller board for 9 track reel tape drives. This program runs on the Motorola MVME133A VME Monoboard computer. There are future plans to modify this program for the Rimfire adaptor board, using the RIMFIRE package. The diagnostic tests the internal hardware of a tape controller, the controller/tape-drive interface, and to a lesser degree, the performance of the tape drive. Diagnostics include writing variable length records onto tape, reading them back, and checking for I/O errors. This test cycle can also be set to repeat until a Write-Read-Compare error is encountered.

As noted, problems under VMS are harder to characterize exactly. The TD Systems VIKING Q-Bus to SCSI adaptors on the Tapecopy test stand have terminal lines which allow the interrogation of a diagnostic monitor. The VIKING board will, at times, insert SCSI specific information in the VAX error log.

A reading of the VMS source fiche has uncovered a trace buffer implemented within the MSCP port driver, PUDRIVER. This buffer, activated with an undocumented SYSGEN parameter, records MSCP commands packets passed between the VAX and the VIKING adaptor. Use of this facility has uncovered at least one software problem within VMS.

XII. CONCLUSIONS

8mm helical scan devices are in use at Fermilab. Two experiments plan to record raw data on these devices, and many to use them for DSTs.

The 8mm technology is relatively new and not all bugs have been shaken out of the drive and the supporting adaptors. Over the past year, when the product was first introduced, significant progress has been made in identifying and correcting hardware and firmware problems.

We anticipate Exabyte enhancements which record at twice the density of the current drives. Since the tape speed will be the same, the data recording rate will double.

XIII. REFERENCES

[1] P. Constanta-Fanourakis, M. Votava, "Evaluation results for helical scan devices", Fermilab Computing Department Note PN-360.

[2] M. Fischler, G. Hockney, et al., "The Fermilab Lattice Supercomputer Project", Fermilab-Conf-89/37

[3] G. Oleynik, M. Votava, "RIMFIRE Driver Routine Package Programmer's Guide for the PANDA System", Fermilab Computing Department Note PN-375.

[4] P. Heinicke, D. Berg, B. MacKinnon, T. Nicinski, G. Oleynik, "SCG68K Software Product, Programmer's Guide", Fermilab Computing Department Note PN-369.

[5] P. Constanta-Fanourakis, M. Votava, "Exabyte Tape Copy Facility User's Guide", Fermilab Computing Department Note PN-372.

[6] P. Constanta-Fanourakis, M. Votava, "Exabyte Tape Copy Facility Operator's Guide", Fermilab Computing Department Note PN-371.