CDF Note 5385
D0 Note 3771
JP Note 0034

Tape Technology for Run II

The Second Report of the Run II Serial Media Working Group

D. Amidei, J. Bakken, D. Benjamin, B. Blumenfeld,
K. Chadwick, M. Diesburg, S. Fuess, K. Genser,
S. Lammel, L. Lueking, J. Patrick, D. Petravick,
D. Schamberger, H. Schellman, R. Thies, A.Walters

Abstract: The Joint Run II Serial Media Working Group recommends that the main tape technology for Run II should be **Exabyte Mammoth-2**, with **SONY AIT-2** to be considered as a backup.

1. Introduction

The Serial Media Working group is a Joint Project of CDF, D0, and CD to make a consensus proposal for the magnetic tape technology to be used for the Run II data-sets. The two basic performance parameters for the problem were laid out in the report from the Physics Analysis Resources Working Group (CDF-4100, D0-3197) [1].

- the total tape bandwidth for all activities at full luminosity of 2E32 is of order 100-200 MB/s at each experiment
- the total capacity for 2 fb-1 is 1 PByte per experiment, preferably with the full volume storable in the tape robot of each experiment

The total tape solution must also satisfy obvious practical goals: the cost/GB should be minimized, the capacity of a fixed volume robot should be maximized, offsite tape use must be supported, and there should be an upgrade path for both bandwidth and capacity which is fiscally and technically achievable.

The First Report of the Serial Media Working Group, CDF-4560/D0-3435, March1998 [2] established an overall strategy. A survey comparing data center vs. commodity solutions found interesting general trends. Data center drives had higher bandwidth, but were also an order of magnitude more expensive than commodity drives. Unit media costs were comparable between the two, with data center tapes having higher capacity. Most notably, several commodity technologies were found to have near term development paths that would meet our bandwidth and capacity specs while creating a cost savings at the level of millions of dollars. Inexpensive commodity drives would make bandwidth upgrades cheap, and provide a unified solution for both data archive and export to universities. Further detail can be found in the First Report, or in a representative re-comparison in Appendix I of this note. The decision of the group was to commit to a general purpose “multimedia” tape robot from EMASS/ADIC which is capable of supporting many technologies, and move on to a next phase of waiting out and deciding between the evolving commodity solutions. This Second Report outlines this work and a final conclusion.

2. Four Possible Technologies

The commodity technologies whose development roadmaps were roughly matched to our performance, cost, and schedule constraints were Quantum DLT, SONY Advanced Intelligent Tape Drive (AIT) , Exabyte Mammoth, and IBM/HP/Seagate Linear Tape-Open (LTO).

The DLT7000 is a linear serpentine device with 5 MB/s, and 35 GB/tape. It is well known due to a big market-base in back up applications. A "Super-DLT" of 10MB/s and 100 GB/tape has been promised for some time. Experience with the use of DLT as an active data medium at KTeV has been disappointing with respect to robustness and reliability, and this was verified in independent tests by the SMWG [5,6]. This fact, plus the schedule uncertainty of Super-DLT, lead to the early disqualification of this technology. At time of this writing, Super-DLT is still unavailable.

LTO is a new linear scan product developed in a formidable alliance of IBM, HP, and Seagate. A device capable of 18 MB/s and 200GB/tape was expected in late 1999, but development has dragged on. At time of this writing, evaluation units are expected in Fall 2000. This device is too late for Run 2a.

Our candidates then boil down to AIT and Mammoth, which are both 8mm helical scan devices using metal evaporated tape. A great advantage in the examination of these two options is that (unlike LTO) the "first generation" was available for testing, and (unlike DLT) the subsequent generations were expected to be mechanically equivalent. This meant that we could conduct extensive reliability tests on Generation 1 and plausibly extrapolate them forward to Generation 2, for which there would be only a brief testing opportunity before the need of a decision.
The road maps for the devices are as below. For both devices Generation-N is planned to be backwards compatible with tapes written by Generation N-1, enabling smooth evolution.

SONY Advanced Intelligent Tape (AIT)

- AIT-1 25GB 3MB/s
- AIT-2 50GB 6MB/s shipping since Jan 99
$3000/drive, $95/tape
- AIT-3 100GB 12MB/s "2001"

EXABYTE Mammoth

- M-1 20GB 3MB/s
- M-2 60GB 12MB/s shipping since Jan 2000
$4000/drive, $80/tape
- M-3 120GB 18MB/s "2001"

The prices listed above are representative, further detail on this below. Our investigation of these technologies included vendor site visits, meetings with Sony and Exabyte representatives at Fermilab, and an extensive test program at Fermilab.

3. Examination of Mammoth-1 and AIT-1 and a Preliminary Conclusion

Most of our testing was done on the first generation devices, AIT-1 and M-1. A first set of manual bench tests were performed by S. Lammel in late 1998. One of each drive was set up to read/write new tapes with 10 passes/tape, trying various file sizes, and checking for data corruption. Another of each drive was set up to read the same tape over and over. A second set of tests, supervised by K. Genser in Spring 99, used the ADIC robot to perform a large number of typical operations such as mount/write/rewind/skip/read/dismount on both devices.

Lammel's tests are documented in CDF-4856/D0-3590 [3] and CDF-4858/D0-3589 [4], Genser's results are documented in Joint Project Note JP0033 [5]. The two sets of tests gave somewhat repeatable results:

- After a large number of passes, AIT-1 develops both hard and soft R/W errors traceable to "dirty" tape or drive. A soft error causes multiple rewrites, degrading the total tape capacity, a hard error stalls the drive. An internal mechanism that is supposed to monitor the head cleanliness did not pick up the problem (i.e. the "cleaning light" did not come on). This problem cropped up for 3 drives after ~3700, 5100, and 12,000 passes respectively. In Genser's test, one AIT-1 drive was still running without problem after ~18,000 passes.

  In Lammel's tests there were also a significant number of failures which seemed to be associated with a bad batch of media. The tapes were returned to SONY for analysis, which ultimately revealed media damage by the drive.

- For M-1, Lammel observed tape cleanliness problems after ~1000 passes, but Exabyte subsequently made an engineering change which fixed the problem. In Genser's tests an M-1 drive was still running after 20,800 mounts.

The overwhelming first conclusion is that the reliability of AIT and Mammoth is far better than our 8mm experience in Run I: compare the survival of drives over multiple thousands of passes to the average of lifetime of 450 passes with Exabyte 8500. Assuming that next generation would have the same mechanical and stress related performance as the first generation, we conclude, based on these tests, that *either* technology would be suitable and robust enough to employ in Run II, although Mammoth has an edge in bandwidth and reliability.

The weak point in this argument is the assumption that reliability can be extrapolated to the next generation. Our final conclusion has been waiting for tests on AIT-2 and M-2.

4. Examination of Mammoth-2 and AIT-2

M-2 was late coming to market and still somewhat "green". In addition to the capacity and bandwidth improvements, M-2 has a new feature that the tape leader is a cleaning tape, and when the drive detects it is dirty, it spaces to the leader to clean itself.

We have worked on a succession of M-2 evaluation units since Jan 2000. In the most recent tests, with hardware received in May 2000, 3 drives have been through 1400, 1200, and 60 R/W passes in desktop units. Tape capacity is verified to be 60 GB. At the 0.2% level there are problems writing filemarks which occur each time at different locations on the tape. The device detects the error and rewrites. Some problems at end of tape and one problem with a wrong measure of remaining tape seem to be bugs introduced in an update of the microcode. In the test batch of 15 tapes, there was 1 pathological tape which would stall any drive, this has been sent to Exabyte for analysis.

Although AIT-2 was available since Jan 99, the 50GB media was extremely scarce, and has only lately become available in quantity. Three drives have been put through approximately 500 R/W passes. Tape capacity is verified to be 50 GB. Two drives had soft errors writing filemarks, similar to those of M-2, however they occurred at repeatable tape locations, suggesting media defects. One drive had a hard error, hanging after 487 passes with same error message about filemark write. Another drive had a hard media error and hang after 470 passes. There are not large statistics here, but the impression is that we are seeing low level media problems, and a sensitivity to this problem which varies from drive to drive. This is reminiscent of the media related problems seen with AIT-1.

5. Cost Considerations

Although we have focussed on the technical merits of the drives, there is still the bottom line. In terms of street price, the media cost is $1.90/GB for AIT and $1.33/GB for Mammoth. The bandwidth cost is $500/MB/s for AIT and $333/MB/s for Mammoth.

Multiplying the media costs by large volumes shows a huge price advantage for Mammoth. In a competitive Request-for-Quote on the sale of 1 Pbyte, 2 out of 5 vendors were able to push the AIT media costs into a competitive range, but only on the volume sale. University collaborators buying university-sized volumes would still find a price disadvantage with respect to AIT. There is a modest cost advantage for Mammoth on the large bandwidth driven central tape plant.

Finally, we note that the best quotes on media put the cost at ~ **$1.1/GByte**, which is an extremely cost effective outcome, and (we feel) validates the SMG strategy of waiting out the commodity market.

6. Conclusion

After extensive study and comparison, we conclude that Mammoth-II is a technically superior and cost effective tape system and we recommend that it be our first choice for the tape storage technology in Run II.

SONY AIT-II is runner-up in performance and cost, and would be an acceptable contingency solution.

References

1. L. Buckley-Geer, et al., “Joint D0/CDF/CD Run II Data Management Needs Assessment Report, CDF 4100 and D0 3197, March, 1997.

2. J. Bakken, et al., “Serial Media Working Group First Run II Report”, CDF 4560 and D0 3435, March, 1998.

3. S. Lammel, “Exabyte Mammoth Tape Drive Stress Tests”, CDF 4856 and D0 3590, Jan, 1999.

4. S. Lammel, “SONY AIT-1 Tape Drive Stress Tests”, CDF 4858 and D0 3589, Jan, 1999.

5. S. Lammel, “Quantum DLT Tape Drive Stress Tests”, CDF 4858 and D0 3591, Jan, 1999.

6. K. Genser et al., “Results from 1999 Serial Media Working Group (SMWG) Tape Drive Tests”, http://cddocs.fnal.gov/cfdocs/productsDB/docdetail.cfm?DocNum=JP0033.

Appendix I: A Re-Comparison of Commodity and Data Center Options   L. Lueking, May 99

Ampex is a representative data center technology, with high throughput drives and extremely high capacity tapes. The drives are roughly 20 times as expensive as Mammoth-II for less than a factor of 2 in bandwidth. The cartridges are large, difficult to handle, and offer no advantage in total capacity relative to total shelf space available in the robot. In a direct cost comparison, Mammoth-II saves $400,000 *per experiment.*

Table I. Drive streaming throughput, media and AML2 capacities. Mammoth II numbers assume 230m tapes (60GB).

	Ampex	Mammoth I	Mammoth II
Drive (MBps)	20	3	12
Media capacity (GB)	279	18	60
AML2 tower slots	1152	5185	5185
AML2 tower capacity (TB)	321	93	311

Table II. Quantity, unit price, and total costs for drives to support 200 MBps (streaming) and media to store 1 PB of data. Mammoth II assumes 230m tapes.

	AMPEX			Mammoth I			Mammoth II		
	Number	@	Total	Number	@	Total	Number	@	Total
Drive	10	100k	1,000k	70	2.5k	175k	20	4k	80k
Media	3584	250	896k	55,555	80	4.45M	16,666	80	1.33M
Total			$1,896k			$4,625k			$1,413

Appendix II: SMWG Activities

1/19/99 Quantum DLT Site visit

1/20/99 Writeups of previous round of tests of DLT, Mammoth-1,
AIT-1. (S. Lammel, CDF-4856, 4857, 4858)

2/17/99 Begin new program of tape drive tests with Mammoth-1,
AIT-1, and DLT on ADIC robot. K. Genser supervising.

2/23/99 Exabyte Site visit (DA,SL,MD,KG)

3/5/99 List of performance and reliability questions sent to
Exabyte (re Mammoth-2) and Sony (re AIT-2)

3/12/99 Exabyte team @ FNAL, inc Jorgen Frandsen, Dir. Eng.

3/31/99 Receive replies to questions from Exabyte

4/21/99 Exabyte transmits Mammoth-2 product schedule
("General availability 9/28/99")

5/5/99 Report from EMASS test program
Reappraisal of Ampex option (L. Lueking)
Sense of committee: Mammoth "best", but keep options open

5/11/99 Initiate P.O. for 3 AIT-2. Media limit! 2 50G tape/drive

8/25/99 Rumor: 50G AIT-2 media has technical problems. A summer long
focussed program at Sony fails to improve yields

8/29/99 Second Exabyte Site Visit
Mammoth-2 test drives delivered but general availability is delayed

Fall 99 Many iterations with buggy Mammoth-2 microcode

Comdex 99: Super DLT and LTO, expected Q4-99, are delayed

1/5/00 Exabyte Ships First "Customer" Mammoth-2's

1/12/00 Meet with ADIC representative. Learn that their 8mm install
base is > 15,000 AIT, almost no Mammoth

1/20/00 Rumor: Sony is adding a second supplier for AIT-2 tapes
Will “eliminate backlog” by April

1/23/00 R/W full Mammoth-2 tape at FNAL

2/21/00 ADIC and SONY reps at FNAL to discuss AIT-2 media problem

2/25/00 Mammoth-2 testing at FNAL plagued by microcode problems

3/22/00 Analyse needs for CDF commissioning run. CDF commits to AIT-2 for commissioning run.

4/7/00 K. Kondo (CDF) meets with high level SONY officials in Japan to discuss AIT-2 media problems. (Thanks Kuni!)

4/19/00 Follow-up meeting with SONY media engineers at Fermilab. Extensive discussion somewhat resolves concern about AIT-2 media.

4/20/00 After many firmware updates, M-2 running reliably on bench at FNAL

5/3/00 "Informal" price quotes suggest M-2 advantage of $600,000 per Pbyte.

5/23/00 ADIC balks on support for Mammoth-2 in the robot

6/5/00 Begin joint discussions with FNAL CD, FNAL business office and ADIC regarding support for Mammoth

6/20/00 Begin automated testing of AIT-2 and M-2 in the D0 robot

7/5/00 ADIC agrees to install/support M-2 by 1 October 2000

7/6/00 Results on AIT-2 and M-2 tests

7/12/00 Consensus decision: First choice is Mammoth 2.