Magnetic Tape Media:

Proven Method for the Retention & Recovery of Data

John Bordynuik Inc
4536 Portage Rd
Niagara Falls, Ontario Canada L2E 6A6
905-354-7222
www.johnbordynuik.com

Overview

- The Challenge – “Store Data Forever, Once”
- About Us
- Magnetic Tape Theory
- Data Validation
- Case Studies
- Recommendations

The Challenge

To archive “massive” amounts of data:

- with the least risk of loss over the next 50 years.
- using proven media and technology.
- while preserving the integrity of data.
- with the ability to use and access the data in the future.

About Us

JBI's 7 and 9 Track Tape Recovery Operation

About Us

- We are data recovery experts.
- We will be a publicly traded domestic US corporation in Nov 2008.
- We have an engineering unit in Cambridge, Massachusetts.
- We have dozens of custom tape drives to read legacy magnetic tape media.
- We do not use off-the-shelf components or drives.
- We recover data from any computer-output media including paper, film, fiche, optical and magnetic.

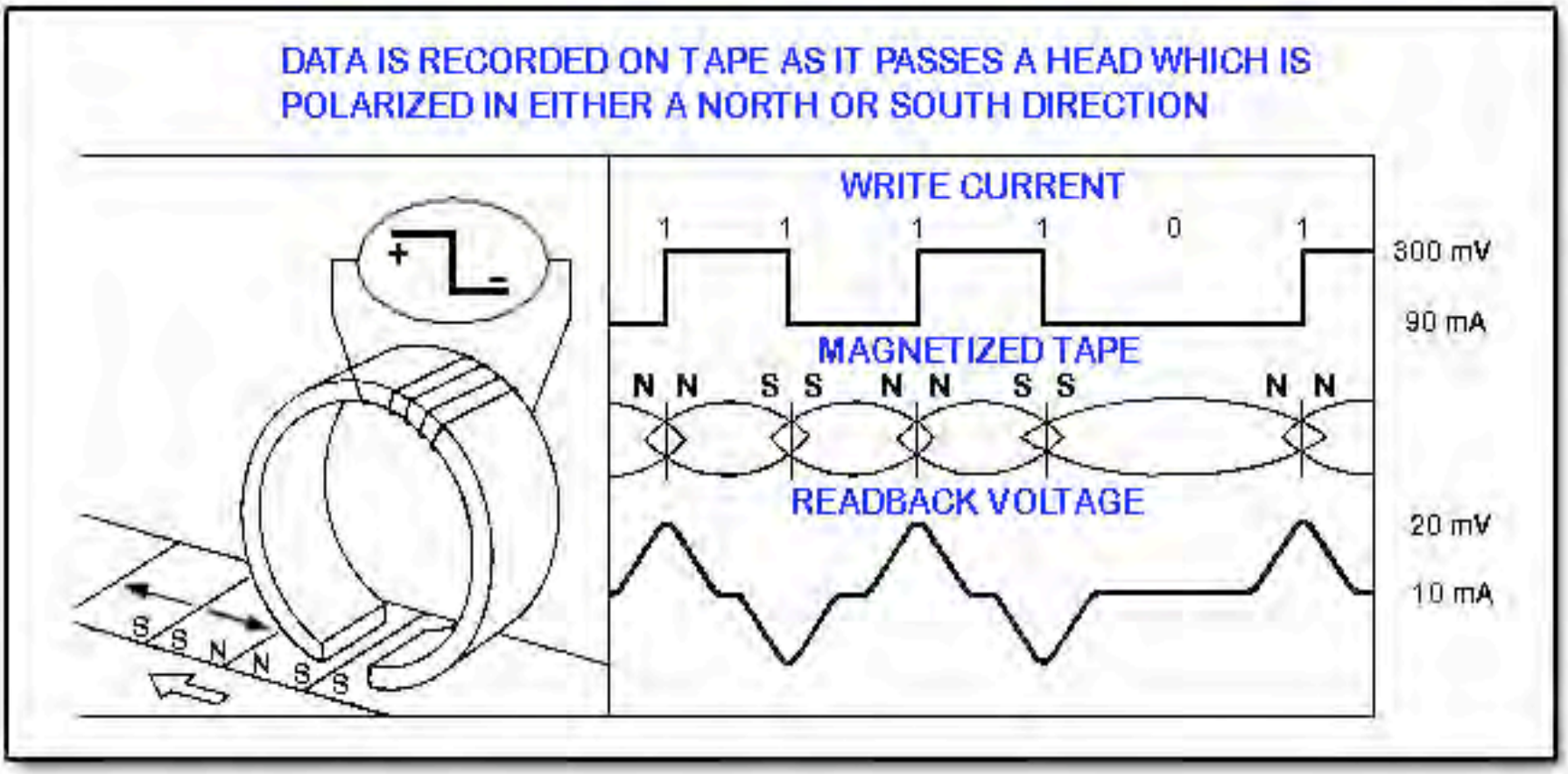
Our Experience

- We design high-volume data recovery hardware for all computer output media.
- We've read thousands of computer-generated media from the 1950's to present.
- We recover the most sensitive and valuable data on the planet.
- We have 20 years of experience.
- We are sole-sourced by MIT, NASA, many educational institutions, Fortune-100, and Fortune-100 founders.

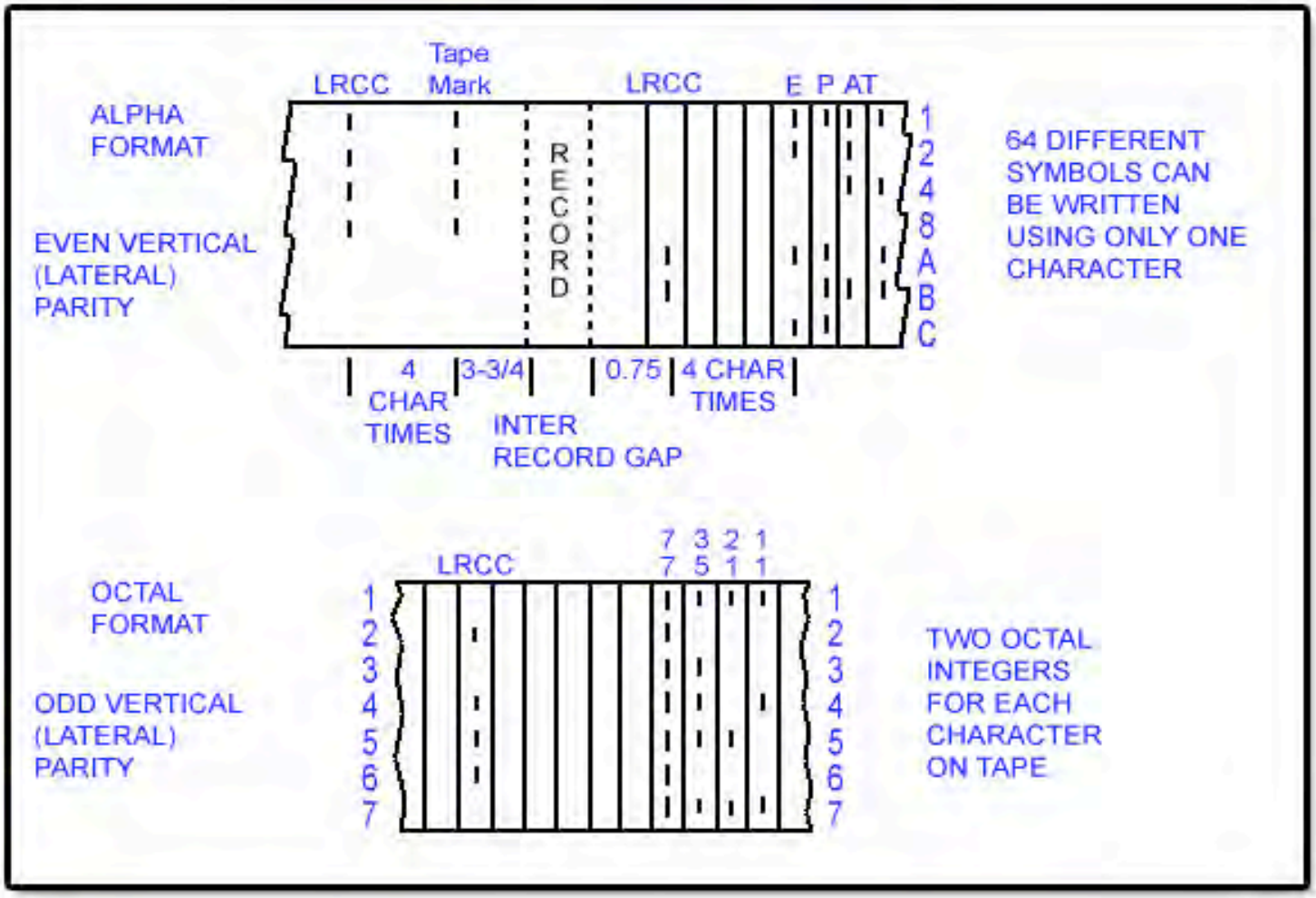
Theory Overview

- Data Recording Process
- Data Validation
- Tape Records

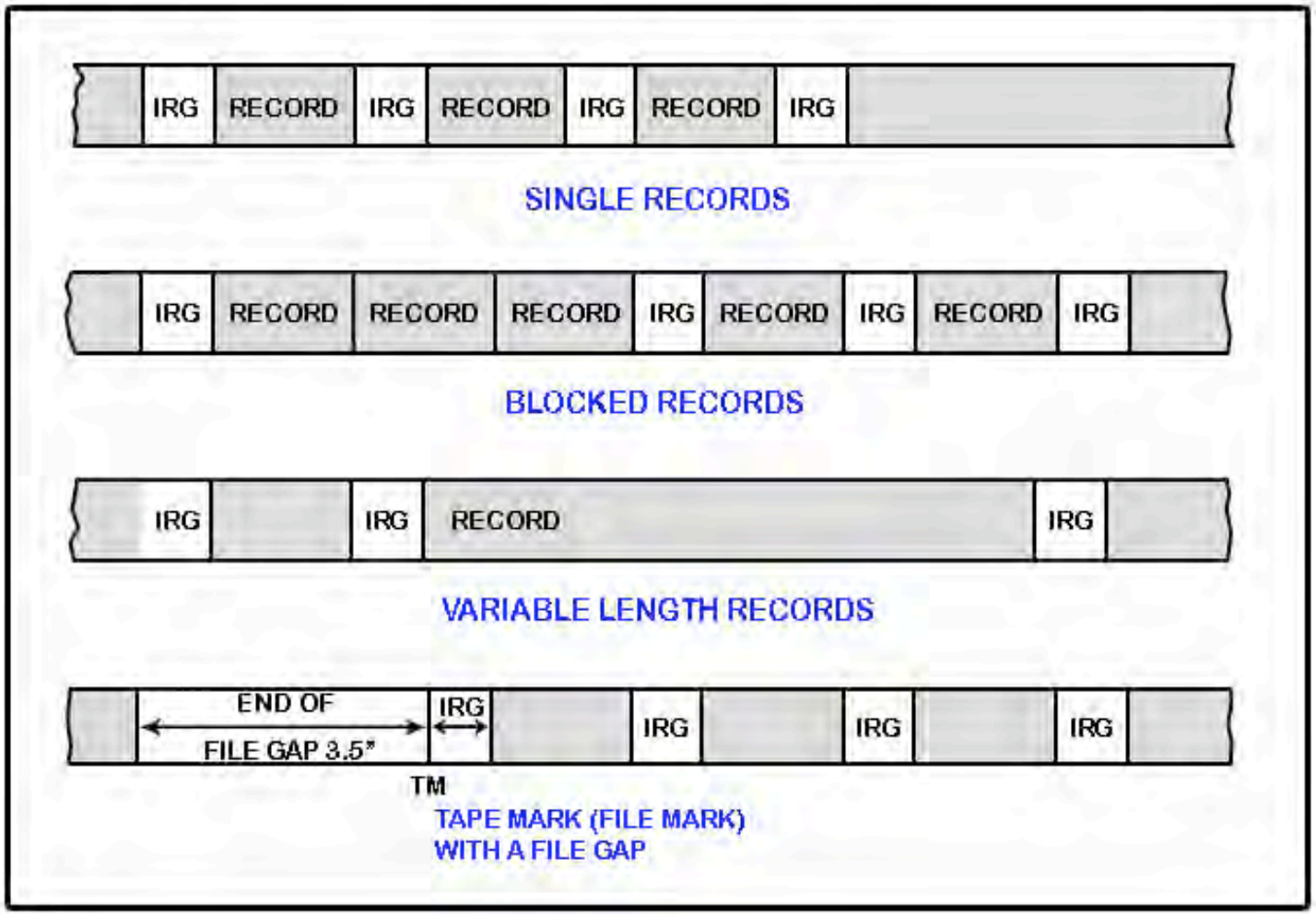
The Data Recording Process



7 Track NRZI Triple Density Format



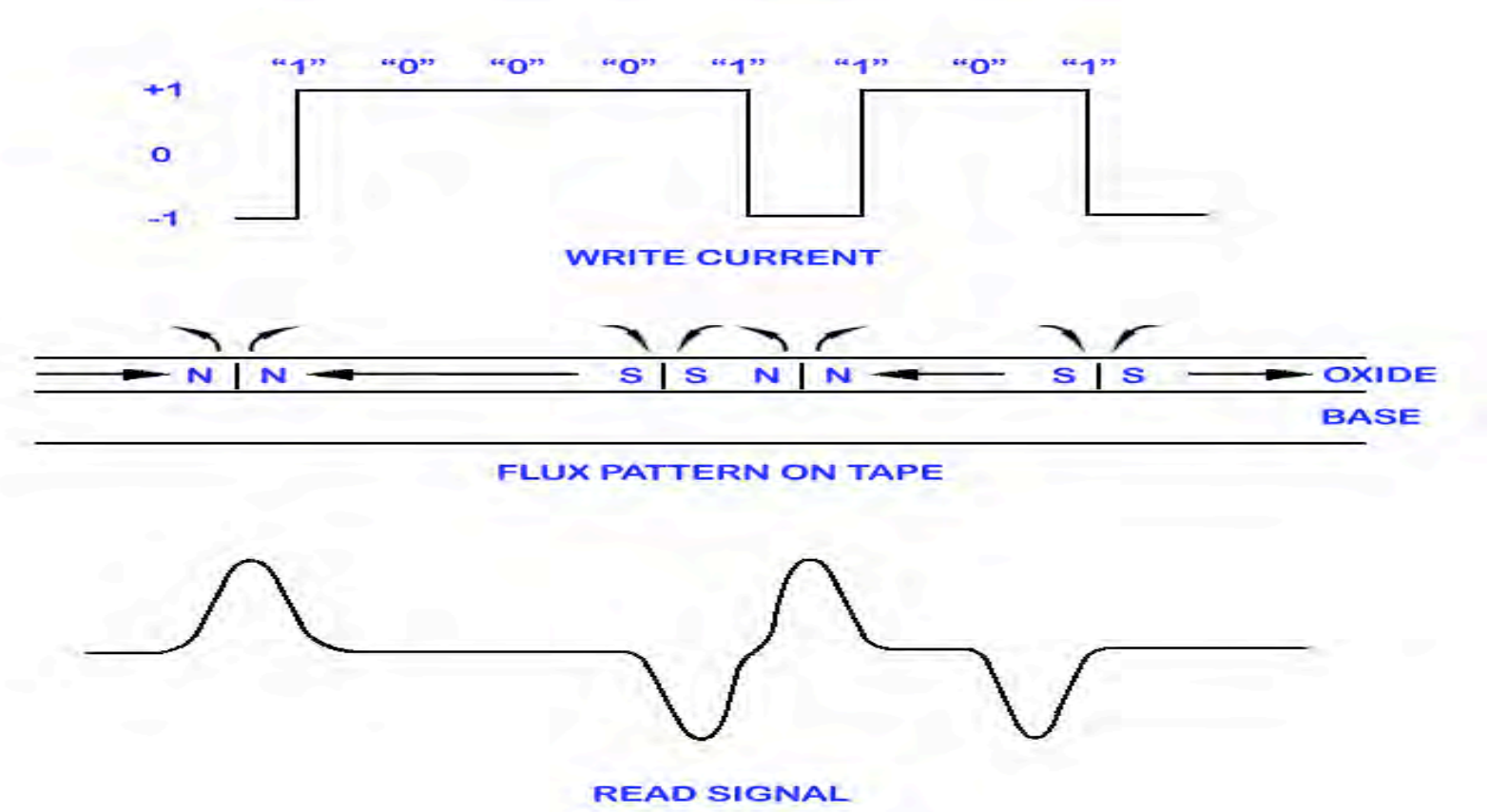
Record Types



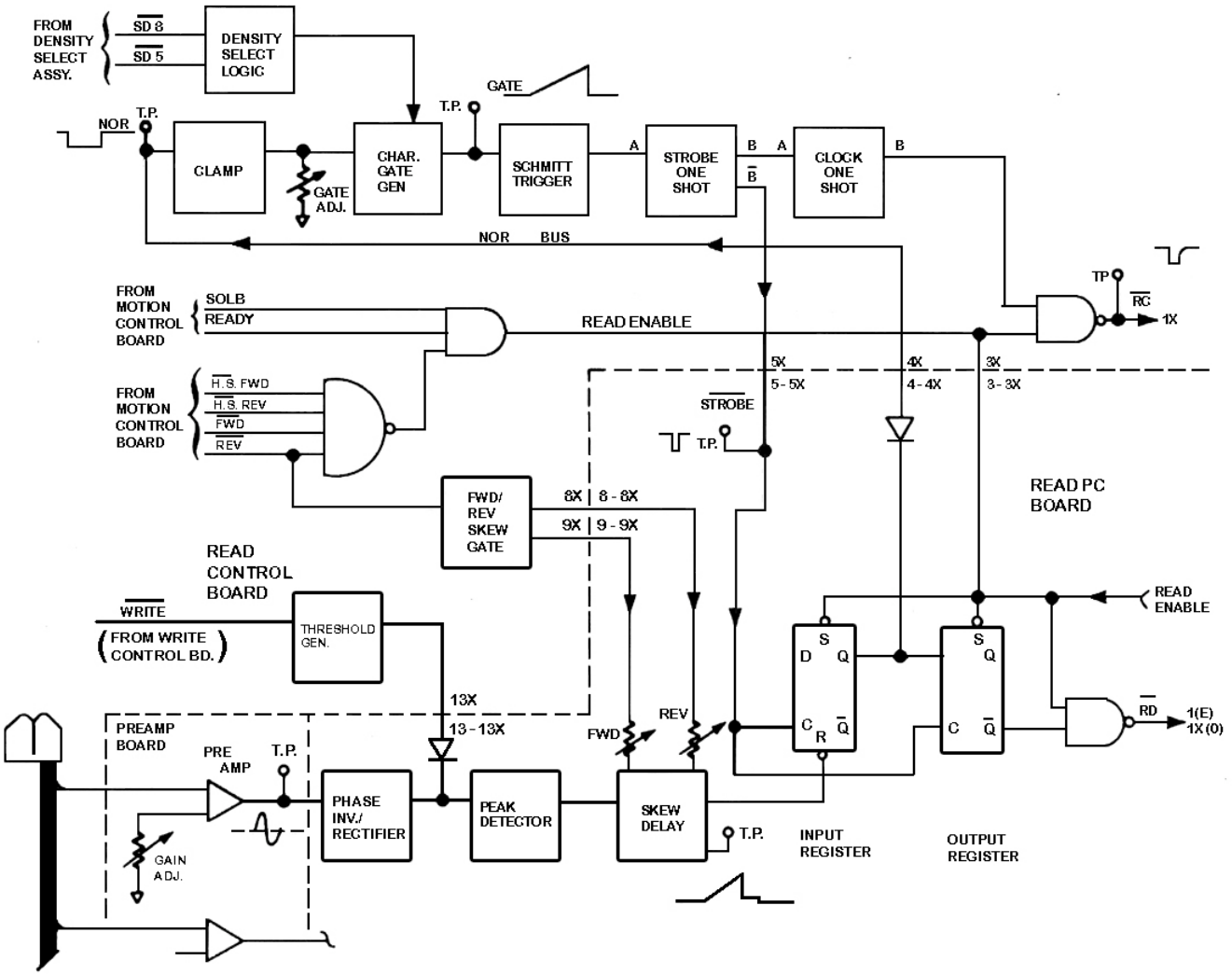
Technology: JBI 7-9 Track Tape Drive

Proven Choice: Magnetic Media

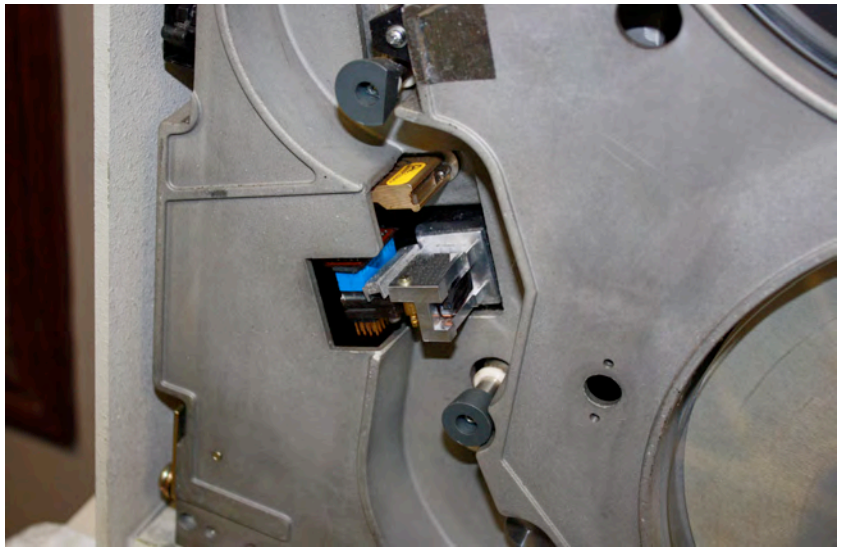
- Physical properties do not change with new tape technologies.
- Flux reversals provide bit-level mechanical checksums.



Original Equipment – No Bit Validation



Technology: New Process

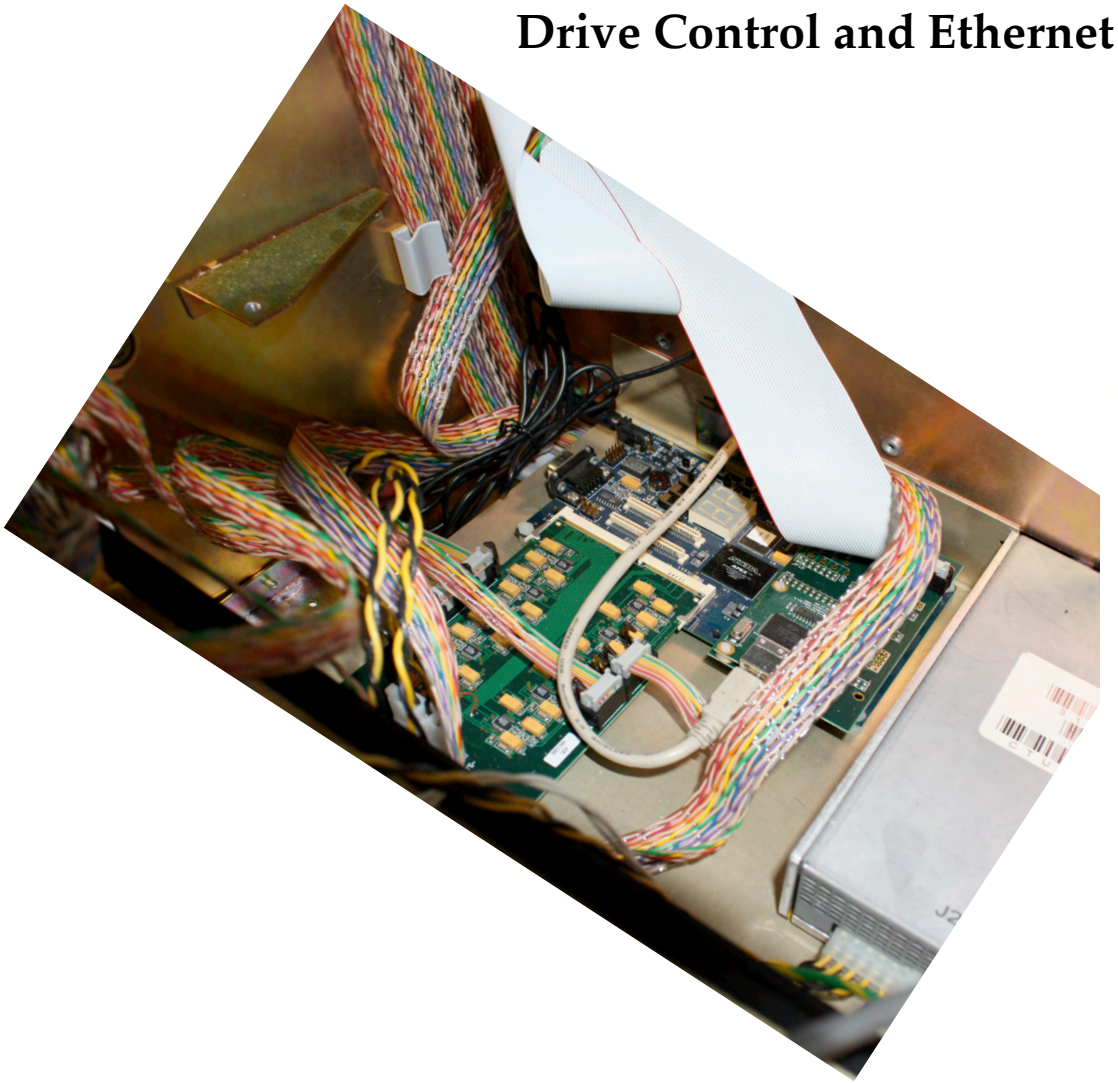


MR Head – speed independent

Custom bias and amplifiers tuned for weak bits

Technology: New Process

ADC Board, Altera Stratix,
Drive Control and Ethernet Interface



Altera's
"Cray-on-a-chip"

Original Drive vs. Custom Recovery Drive

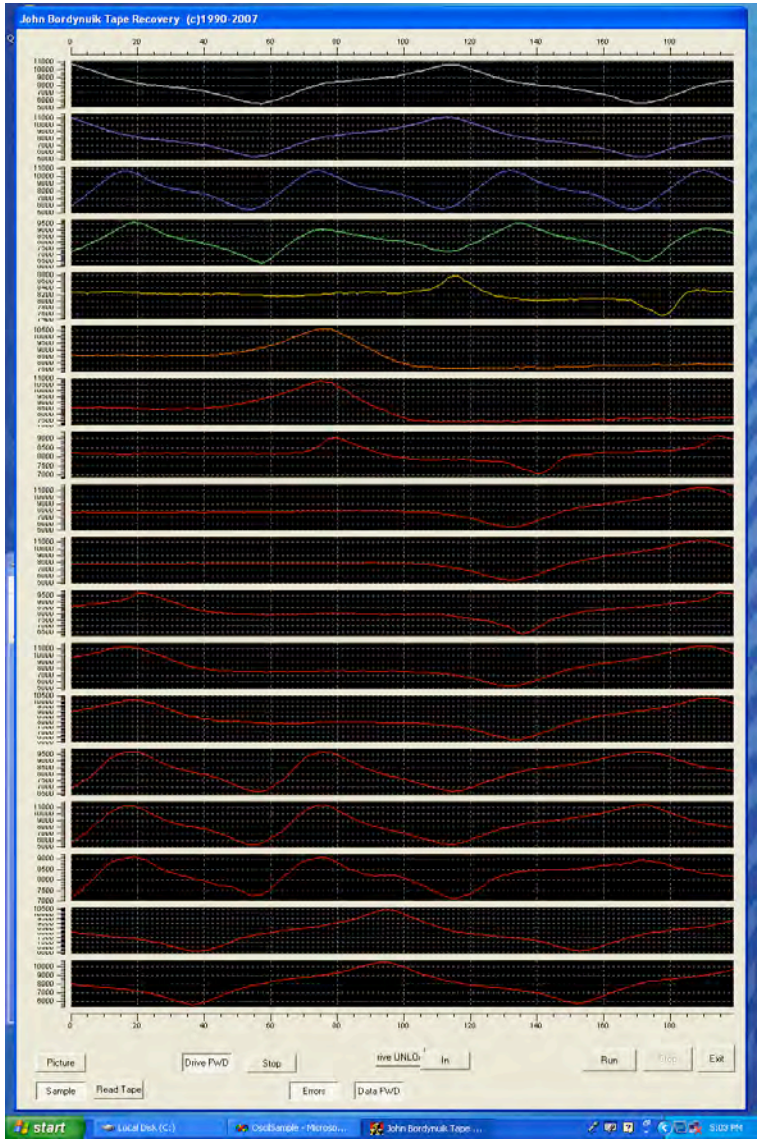
Description	Original Equipment	Custom
Tape Head	Induction – 7/9 track	MR – 36 track, incredibly sensitive
Tape Speed	110 IPS. Speed dependent	5-40 IPS. Speed independent
Bit Detection	Peak Detection or Window Comparison	ADC with analysis in FPGA
Tape Skew	+/- 100 micro inches- static	+/- 6000 micro inches – fully dynamic
Mechanical Validation	None	Every flux reversal validated in hardware
Data Validation	Poor – many drives are programmed to ignore errors	Bit validation, parity, LRC, then deciphering. Intelligent Drive
Alignment	None	36 heads reading tape
Stretch	Fatal	Automatic recovery for stretched tape or bit crowding
Automatic Gain Control	Yes	None – Maximum Gain

Diagnostics

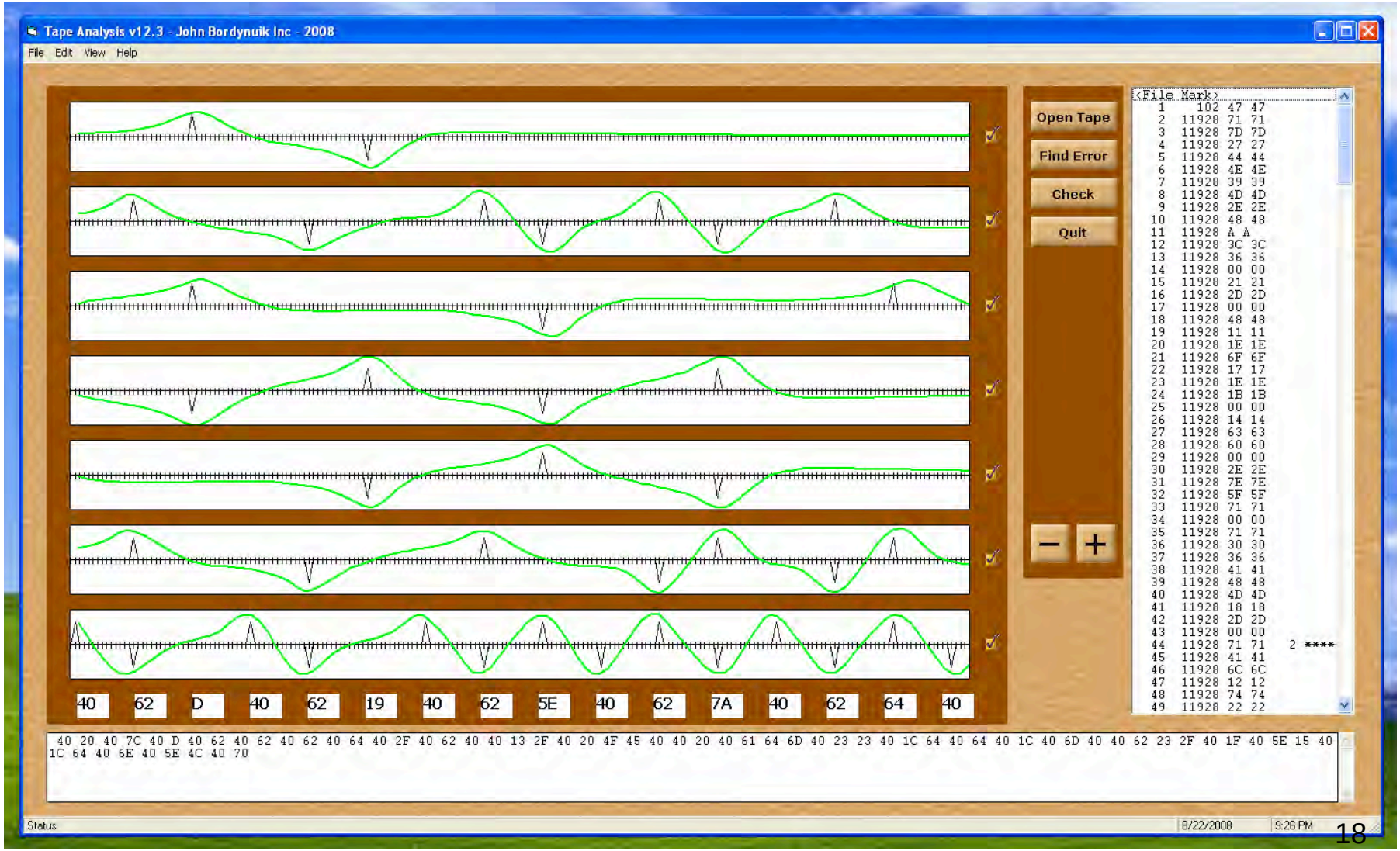
This is a screen shot from our diagnostics workstation displaying data from a 7-track tape.

The 18 waveforms are representative of the analog data amplified from our 36 track MR head. This allows us to recover data from tapes severely misaligned or tapes that momentarily shifted in the tape path.

A single track on a 7 track tape is wide enough to span across 4 to 5 MR heads.



Analysis: Mechanical Bit Validation



NASA Case Study

NASA Case Study - Output Media

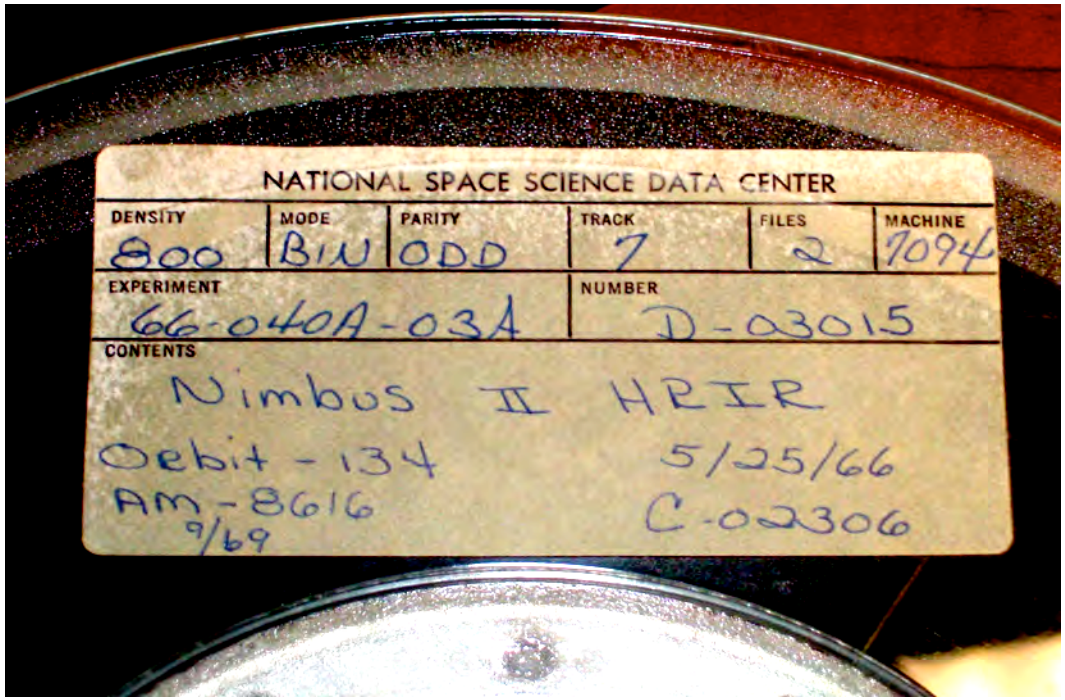
- Computer output media from the 50's through the 80's

- Tapes were stored in ideal conditions.

- Data was archived on high-use tapes.

Nimbus II Recovery Project

- 7-Track tapes written in 1966
- HRIR Sensor data from Nimbus II Satellite
- Completed recovery April 2008



Enterprise Computing 1966

IBM 7094 System

Data Analysis

- 36 bit computer, 6-bit “bytes”
- Data was stored in multiple bit widths
- Little documentation/ no source code
- HRIR (temperature data) for Earth

Deciphering the Data

Organizing:

- Fields and Headers are non-standard
  - No Filenames
  - No Operating System

Deciphering the Data

Visualizing:

- 5MB of highly valuable data…

Deciphering the Data

Visualizing:

- Africa and Madagascar temperatures…

MIT Case Study

MIT Case Study

- 13,000 7 & 9 track tapes
from the 60's
through 90's

- Tapes were stored in poor conditions

What we have learned….

- 50+ years of data
- Thousands of backups
- ARPANET to present

Nothing has changed in 50 years…

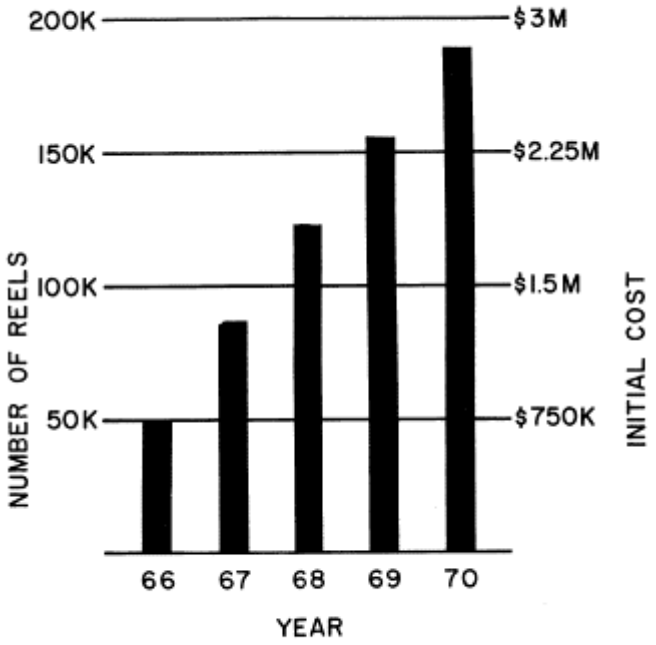
THE GSFC SCIENTIFIC DATA STORAGE PROBLEM

BY

RICHARD G. HOLMES
WILLIAM B. POLAND
ALLAN M. DEMMERLE

JANUARY 1967

NASA
GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND



PROJECTED STORAGE REQUIREMENT FOR DIGITAL DATA (35K REELS PER YEAR AT $15.00 PER REEL)

FIG 2

Nothing has changed in 50 years…

BASIC DATACOMPUTER USE

V. R. Pratt

M.I.T.

5/9/77

BRIEF DESCRIPTION

The Datacomputer (DC) is a slow computer in 575 Technology Square, at the CCA (Computer Corporation of America) node of the ARPA network. Its interesting feature is that it has 3 terabits (3,000, 000,000,000 bits) of memory. It serves as an archive device for the entire ARPA network. The purpose of this note is to tell you just enough about it so you can use it to store files you don't use regularly but don't want to dump onto magnetic tape. This document presents a simplified model of what is going on in DC, to avoid first-encounter confusion; for more sophisticated use, see the more complete manual residing on .INFO.;DFTP ORDER. Hopefully DC will alleviate the present space crunch on the AI computer. I have been using DC myself for the past five months and have found it an excellent way of uncluttering my directory.

Options Available for Long-Term Data Storage

Technology	Risk Level	History
Magnetic Tape	**Low** - quality of transport, tape and storage are all factors.	Excellent Preservation – new technologies have the same magnetic properties. Lifetime > 50 years.
New Optical Media	**High** - unproven, significant historical problems	Poor - chemical changes and physical damage causing irreversible data loss.
Hard Drives	**Medium** - Exceptionally dense data on media is not easily accessible	In 20 years, retrofitting drives written today would be costly, slow and risky.
New Disruptive Technologies	**High** - unproven, lifetime unknown	Could prove fatal for large data sets.

General Findings – Our Experience…

- Data migrated using original equipment to higher density tapes were poor.
- Original equipment was unreliable/unable to read data from legacy tapes.
- 15% greater problems found when tapes were stored in less than ideal conditions.
- 15% greater problems encountered when data was archived on well-used tapes for long-term storage .
- Actual lifetime of magnetic tape media is still unknown. Our equipment was capable of accurately detecting weak bits when 10% of the original amplitude was present.
- **We successfully read tapes that were written in 1952.**
- We have re-read tape archives that were migrated using original equipment because data integrity was found to be poor.

Recommendations

- Use only enterprise-grade tape drives.
- Archive long-term data on new tapes or rarely used tapes.
- Store tapes in climate-controlled facilities.
- Avoid unproven “shiny” new disruptive technologies.
- Do not attempt to migrate data from old tapes using original drives.
- **Archive source code on tapes with sensor data.**
- Keep a backup of a system tape with applications for emulation in the future.
- The data you archive now may be infinitely valuable in the future, treat it like gold.

Retain Source Code with Data

- It's an intellectual tool that describes "What is" and "How to" of data.
- Unlike written documentation, it is completely unambiguous and computationally effective.
- Terminology in documentation evolves but precisely expressed functions in source code do not.
- Our data and data structures are becoming far more complex every year making decoding more and more difficult.

Risks and Reasons for Recovery

- Many of our clients are recovering data from old tapes to seek solutions to present day problems.
- Litigation
- 1960's NASA Earth Science data stored on 10,000 tapes can be retrieved and analyzed on a common desktop computer today. Enterprise computing technology in 1966 was 8 kilobytes of RAM and no disk drive.
- As computers and algorithms evolve, sensor data will be reprocessed in the future.
- National Asset Amalgamation – reduction in business unit overhead relating to asset tracking.