STORAGE DEVELOPER CONFERENCE

SDC 21

BY Developers FOR Developers

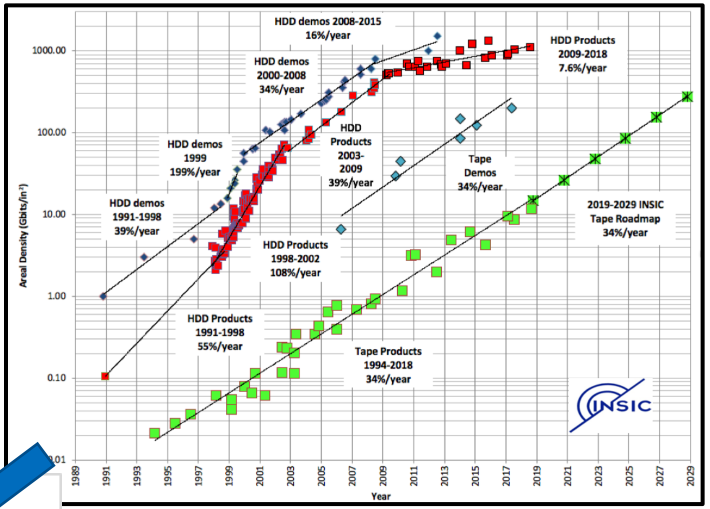
Virtual Conference
September 28-29, 2021

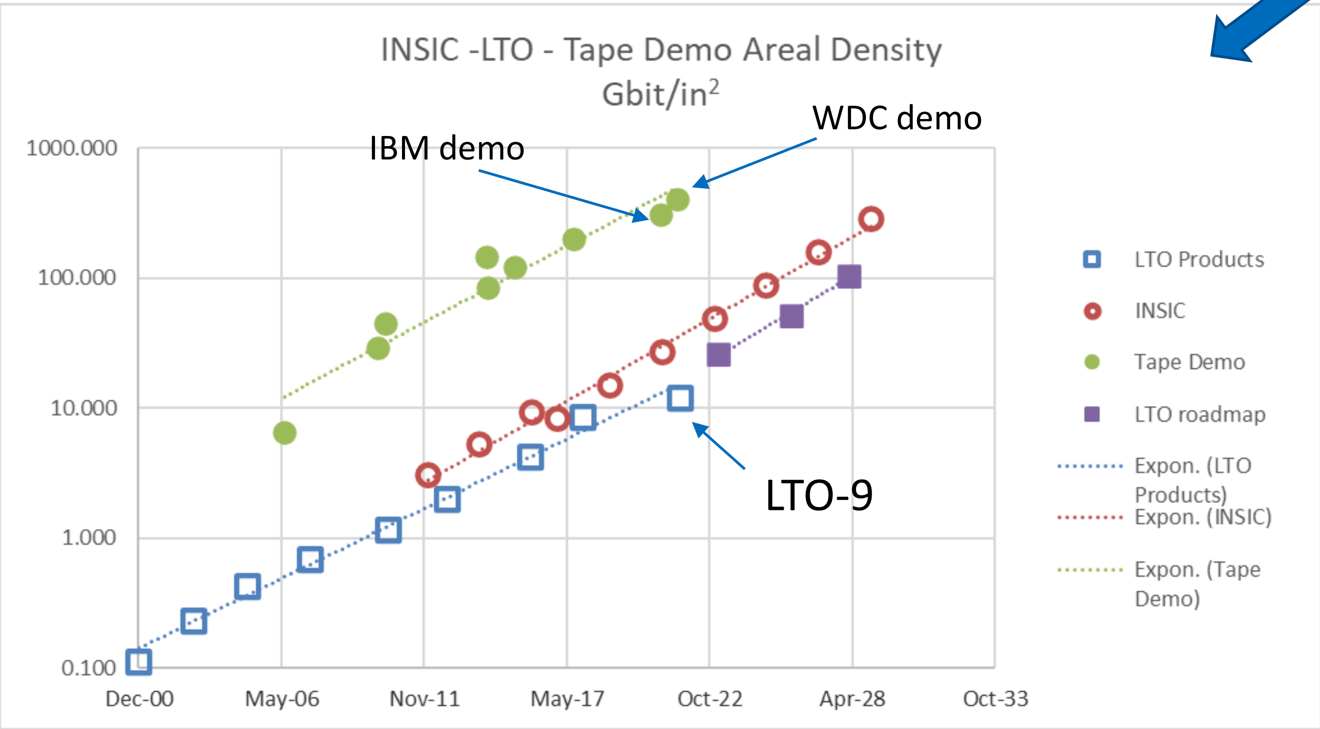
LTO-9 Technology and Two-Dimensional Erasure-Coded Long-Term Archival Storage with RAIL Architecture

Presented by Turguy Goker Quantum Corporation

Tape Is NOT Dead:

Has a new home at Hyperscale Data Centers





https://www.insic.org/insic-application-systems-and-technology-roadmap/

- ❑ LTO-9 has Volumetric Density advantage over HDD
  - ❑ 18TB Exos X18 HDD 1022Gbpi^2 while LTO-9 uses only 12Gbpi^2
  - ❑ LTO-9 Track Pitch is ~30X bigger than similar capacity HDD
  - ❑ 12.5m^2 magnetic tape area for data + format
  - ❑ Total volume of Cartridge is 60% of the 3.5” FF HDD volume
- ❑ Tape is Green and Low Cost
- ❑ Tape has Air Gap ideal for Ransomware

INSIC 2012 Prediction and LTO-9 Technology

- Bit Aspect Ratio (BAR) is key for Capacities establishes Magnetic Recording and Tracking technology requirements; goal is to push Tape BAR towards HDD, so smaller is better.
- LTO-9 achieves the TPI and BAR targets by employing a new Cartridge Optimization Algorithm with one time calibration enabling precise Shingle Recording with tape over environmental conditions

Key Cap Technology Metrics	INSIC prediction	LTO-9
TPI/ Track Pitch (nm)	22498/1129	21897/1160
AD	16.67	12
BAR	33	25
Tape Length (m)	1202	997
Head Technology	TMR	TMR
UBER / ECC Technology	10^{20}	10^{20}
TDS control	Substrate, Data Bands	Cartridge Optimization

LTO-9 followed INSIC BAR predictions in line
Typical 18TB HDD BAR is 5; same capacity LTO-9 BAR is 25

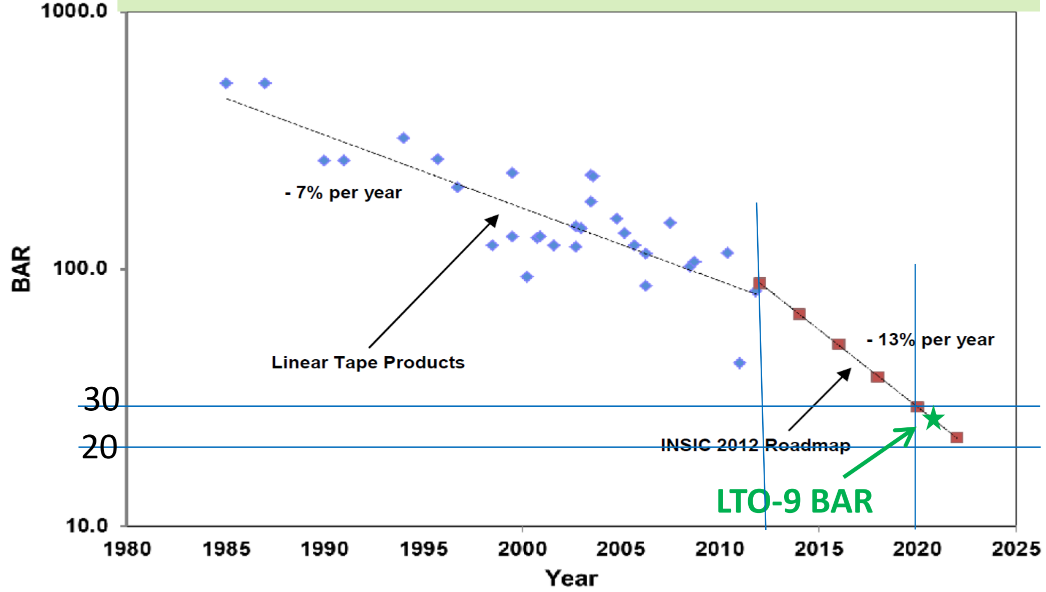
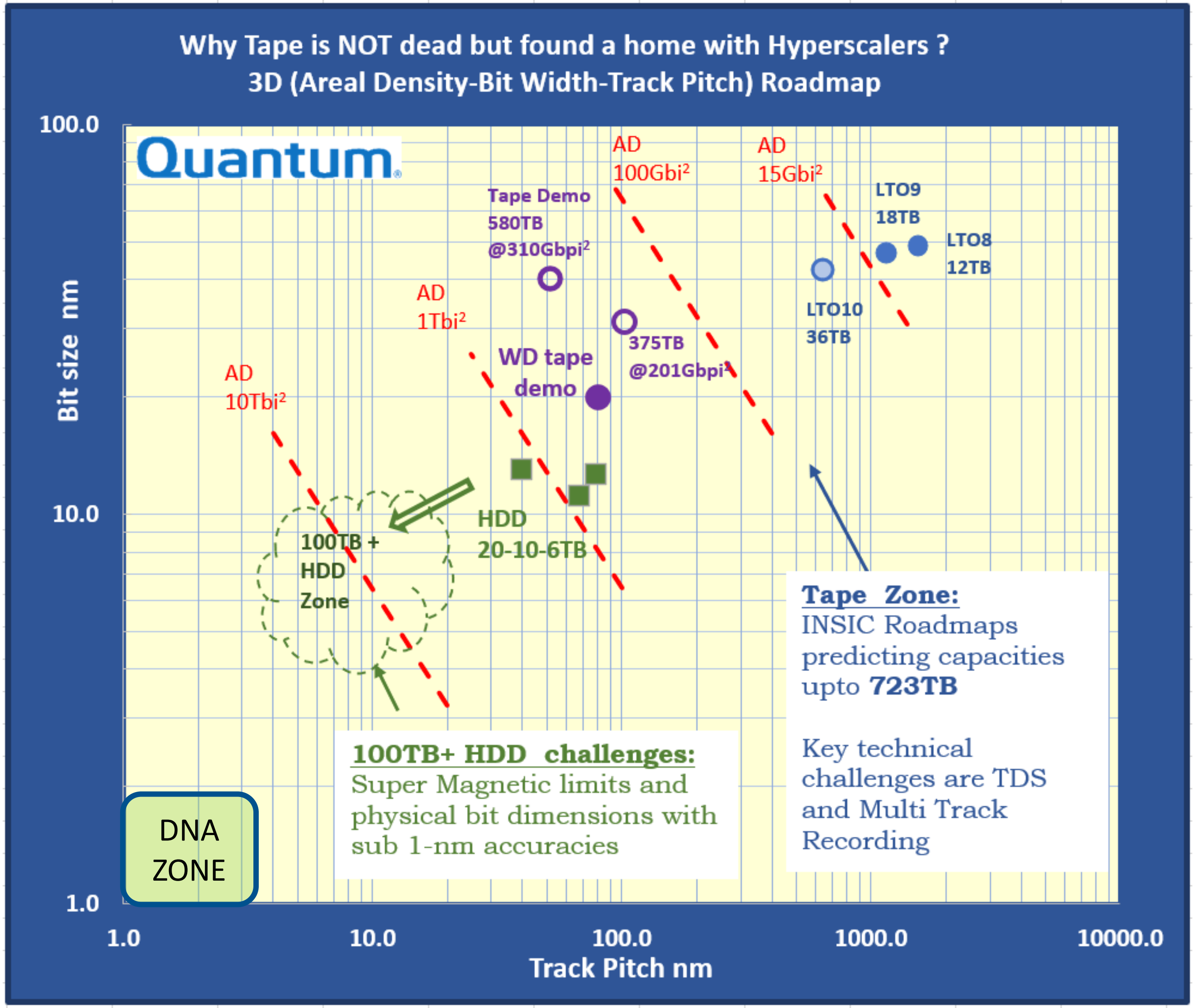


Figure 6: Tape Bit Aspect Ratio (BAR) Trend.

Areal Density Technology Deep Dive



- Dimensions below a few nm, DNA technology may be the solution, but it has its own challenges
- Some of the key HDD challenges for 100TB+:
  - How to manage Bit dimensions below 10 nm
  - SNR and BER at scale
- LTO Challenges:
  - TDS
  - Track Trimming
  - New thinner stable substrate technologies
  - Environmental Controls

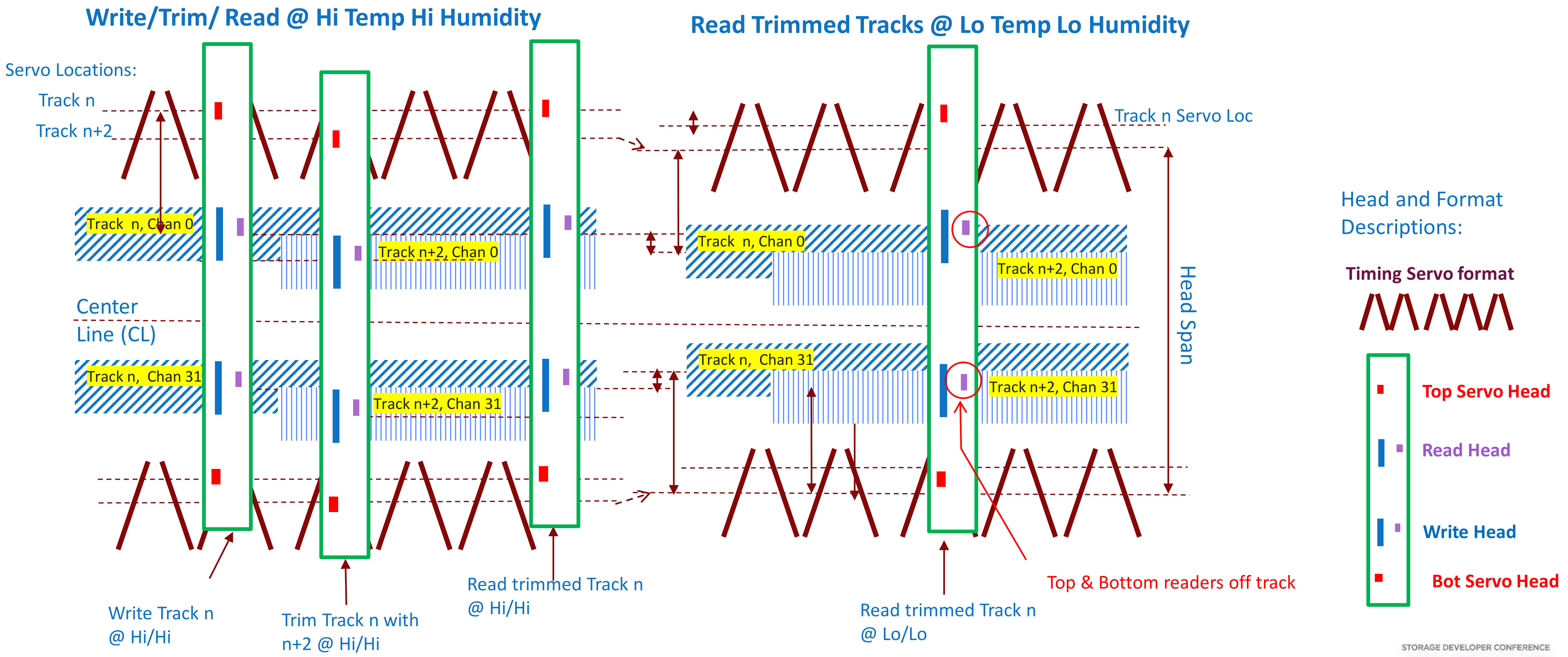
STORAGE DEVELOPER CONFERENCE SDC 21

TDS and Compensation Techniques

- Tape **T**ransverse **D**imensional **S**tability (TDS) is the allowed change in tape width in ppm due to changes in temperature, humidity, ageing and tape tension as well as in-cartridge creep due to pack stresses.

- TDS compensation methods developed over covered by various patents:
  - Tension to control media width to compensate mainly the Humidity part of TDS
  - ECC; best used to recover the errors at edge tracks using C1-C2-C1-C2 type iteration
  - Reduced head span best done by increasing Servo Bands from 4 currently to 8 or even 12
  - Azimuth head to compensate for reading TDS compromised tracks at different conditions
  - Shearing head by Piezo
  - Stretching head potentially by heating or other means
  - Heads with different reader sizes; narrow readers at the edges and wider ones in the center
  - Substrate Technologies; Aramid offers 4X reduction compared to PEN/PET but High Cost
  - Adjacent Write Technology, writers are trimming each other in a Matrix Head configuration

Simple TDS Illustration



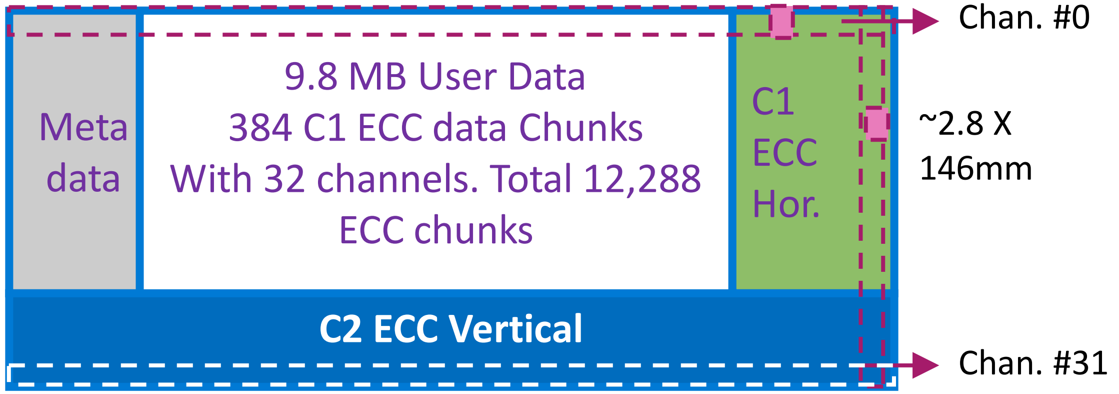
LTO-9 Numbers

	Units	LTO-9	LTO-8
Nominal Capacity	TB	18	12
Application Design Capacity **	TB	17.4	NA
Areal Density	Gb/in2	11.9	8.5
Tape Length - recordable	m	997	922
Number of Wraps	#	280	208
Size of a Data Set	MB	9.805	5.031
UBER *	#	10^20	10^19
Number of 9's	#	16	15
Data per Wrap (Ideal)	GB	66.957	59.688
# Max Data Sets Per Wrap (Ideal)	#	6828	11864
4- Partition Nominal Size	TB	4.500	2.885

* UBER for LTO-9 is based on Quantum's ECC and channel model

** Application design Capacity (ADC) is best for Constant Capacity Applications where rewrites are critical for performance

LTO-9 Format & Data Durability; the UBER Story



- **Magnetics 101:** 2X increase in Linear Density cost 6 dB SNR and 2X increase in Track Density cost 3dB SNR
- **Optimized BaFe particle media**, new TMR heads and new powerful C2 ECC altogether supported 41% increase in Areal density
- Novel **Iterative C1-C2-C1-C2 ECC decode** algorithm unique to Tape Format improved data recovery robustness in Read Error Recovery modes
- QTM's ECC and Channel Models estimate **LTO-9 UBER with new C2 ECC to be 10X better than LTO-8** with error rates greater than 10^{20}
- Using new larger Data Set size of 9.8MB, 10^{20} BER translates to **16 9's durability** assuming errors are random
- For Correlated Errors and full cartridge error cases, QTM promotes its **2D ECC protection** or a combination of replication and ECC
- New C2 ECC also has additional benefits as a **TDS compensation** since even with loss of edge channels to remaining C2 parities are still same as LTO-8
- New C2 ECC is robust to Shock events and off-track conditions during Read Mode where **up 18 mm of data loss can be tolerated before read error**

STORAGE DEVELOPER CONFERENCE SDC 21

LTO-9 Capacity Efficient Partition Format

- Capacity Efficient Partition Format
- LTO-8 and earlier gens used 2 guard wraps (FWD and REV) for isolation resulting in capacity loss
- In LTO-8 4 partition format resulted in 2.885 TB partition vs. 3 TB due to use of guard wraps
- LTO-9 format offers guard less partitions where Servo Bands are used for natural Guarding for 2/4 partitions
- In a 4-partition equal size format, users can write 4.5TB nominal and even more if Variable Capacity is used

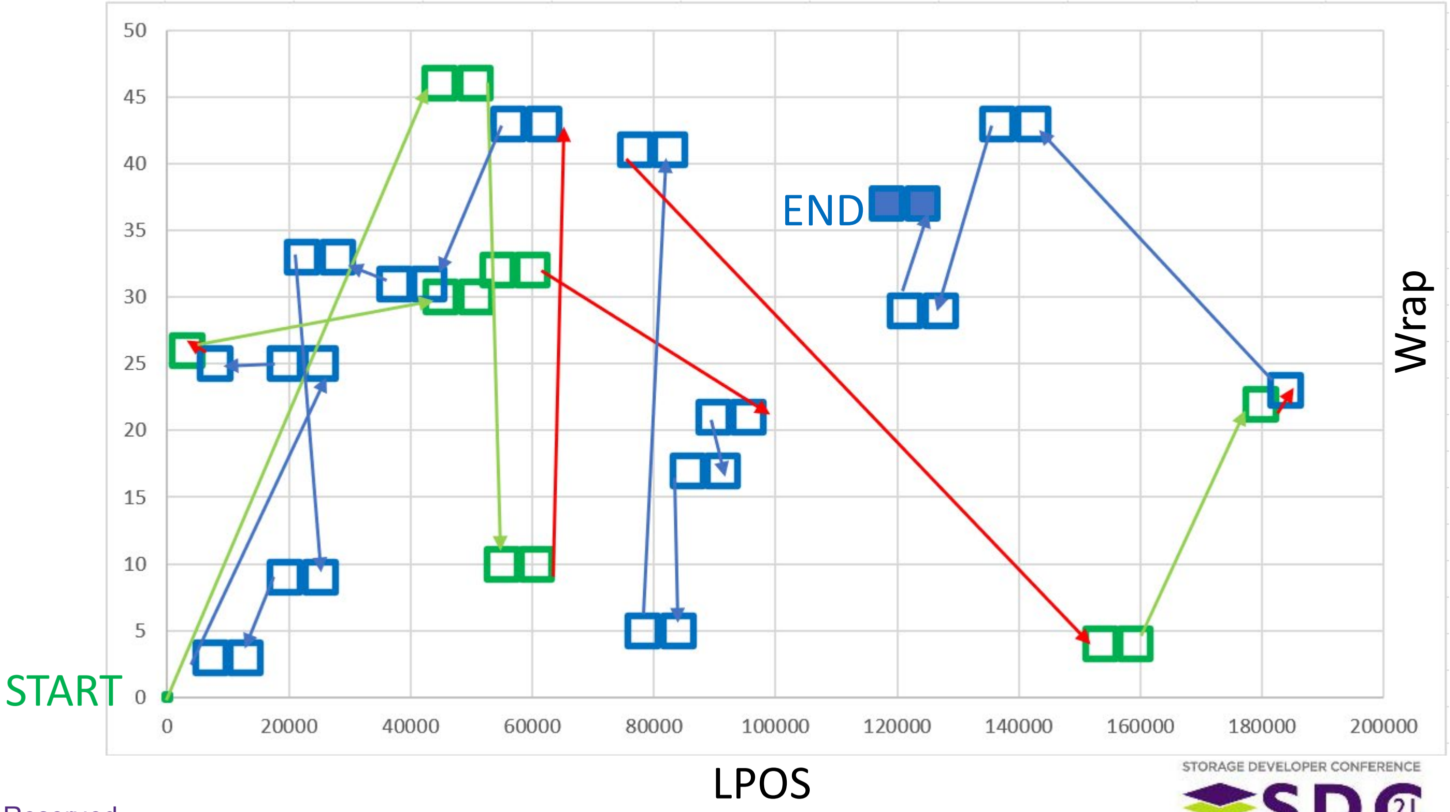
LTO-9 Drive –Data Durability Optimization Algorithms

- Preparing Cartridge for Storage:
  - Unique unload process by managing tape packing at Cartridge Reel to deal with Pack Stresses
  - Increases Unload times up to 4 minutes based on last operation of media
  - Crucial for data durability with thinner and longer media due to pack stresses when stored
- One Time Cartridge Optimization
  - 21,896 TPI with thinner, longer tape over environmental conditions requires precision track trimming to compensate for TDS providing 16 NINES user data durability over the recommended environmental conditions.
  - One time Media Optimization algorithm performed on first load creates a referenced calibration meta data for each cartridge enabling accurately trimming tracks to compensate for TDS. Any drive can use this reference to write or append tracks to that cartridge resulting in highly durable user data format with log-term media durability.
  - Recommendation is to perform first load in the location of deployment, which should be in a stable environment that meets the recommended environmental specification.
  - Media optimization averages 40 minutes per first load of a cartridge to a tape drive. Although most media optimizations will complete within 60 minutes some media optimizations may take up to 2 hours.

LTO-9 FH Drive Only–New RAO Algorithm

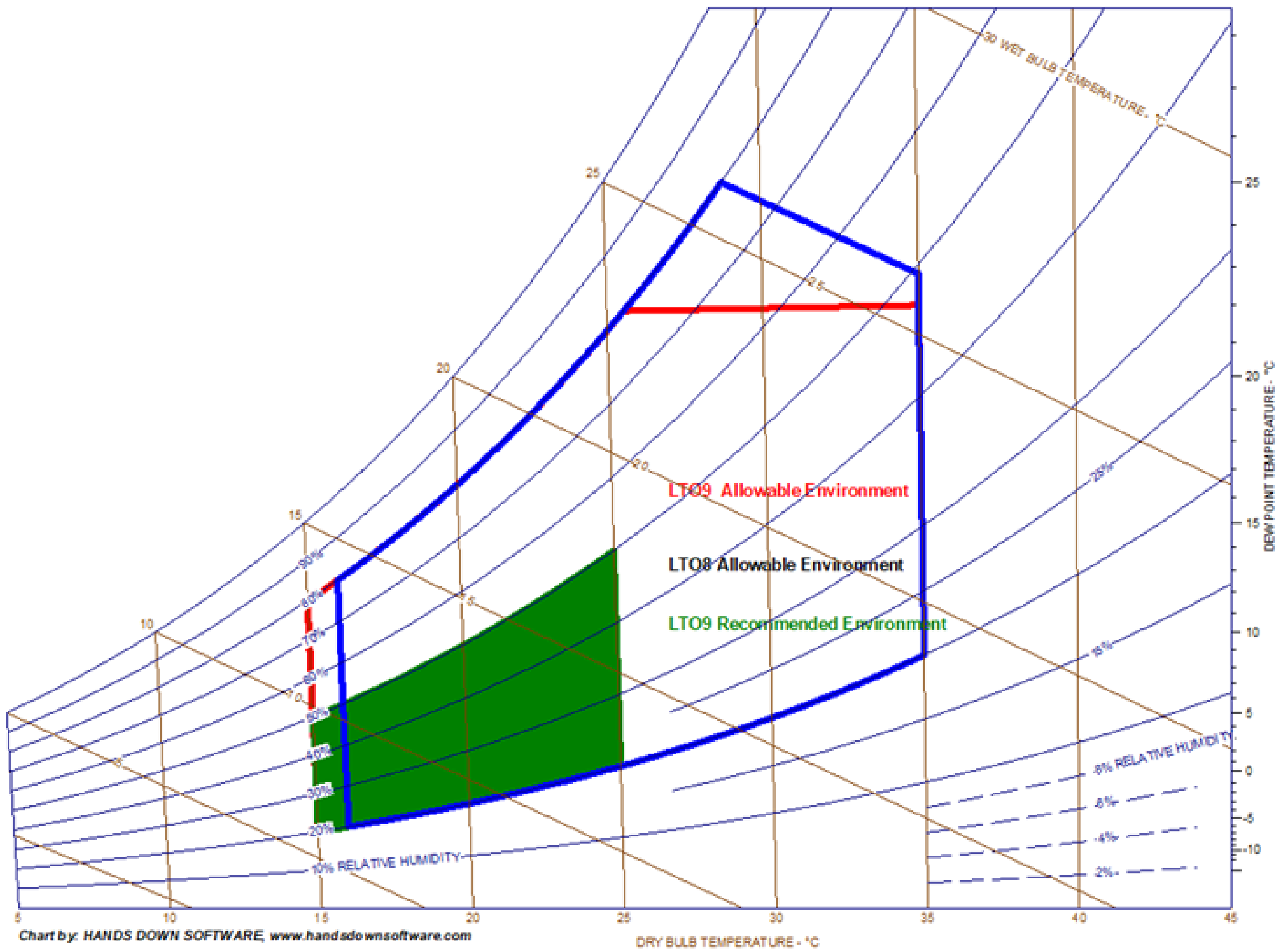
- Drive based RAO algorithm sample result (20 files 2GB each) showing the way drive decides which files to read in physical order for minimum time.
- Send Drive the list of Logical Block Addresses for records to be read using GRAO command
- Drive will use this command and the list of records with Logical Block Addresses to order them in a physical order based on internal Re-Ordered algorithm. It will read CM to know where each records starts and finishes.
- Now Host will send the next command RRAO to the Drive and Drive will now send back the list of records put in a physical order back to the Host
- Host now will use this order to read the records from the drive using established SCSI read commands.
- QTM has also developed a Host Based RAO which can be programmed for different optimization goals managed all by the Host

	w/RAO	wo/RAO
Read Time, min	11	18
Tape Motion, m	5328	8201



STORAGE DEVELOPER CONFERENCE SDC 21

LTO-9 Environmental Requirements-New



- Blue region as shown in plot is the LTO-8 allowable range with the 26°C Wet Bulb limit
- Because of TDS with high track densities and thinner media, the Wet Bulb temperature has been replaced by the 22°C Dew Point
- However, Recommended Range is unchanged
- Now LTO-9 drives come with environmental sensors

STORAGE DEVELOPER CONFERENCE SDC 21

Overview: Locally Repairable 2-Dimensional Erasure-Coded Tapes

- Quantum's patented technology plus pending ones
- Ability to read Erasure Coded files using single tape but with protection of multiple tapes
- 11 9's Durability most Tape and Drive correlated errors using Per Tape ECC
- >15 9's Durability using iterative decoding with 2-Dimentianal ECC
- Low Erasure Code overhead; less than 60% compared to 100% for 2 Copy and 200% 3 Copy
- Performance & Recovery
  - Repairs are done locally within single tape for most common Tape and Drive errors
  - For loss of tapes or severe errors, iterative decoding with Multiple tapes are used
  - Reads bypasses most drive recoveries using local erasure codes, extending tape and head life
  - Works with RAIL

Typical Tape and Drive Errors Affecting NINES

Typical Errors	Estim. 9's for errors & potential protection methods		
	% Error	**Nines**	**Protection**
Random or small burst errors	$> 10^{19}$	> 11	C1-C2 ECC
Defects: new, acceleration post write and scratches	< 0.05%	< 4	PT & 2D
Edge quality and packing problems, bad CM			PT & 2D
Local worn sections due to stop starts, back hitches			PT & 2D
Lost, damaged, misplaced or the human factor	< 0.1%	< 3	2D
Pin drops, tape stiction, cut tape, tension control	<0.001%	< 5	2D
Typical correlated Drive-Tape read errors	3-4%	< 2	PT & 2D

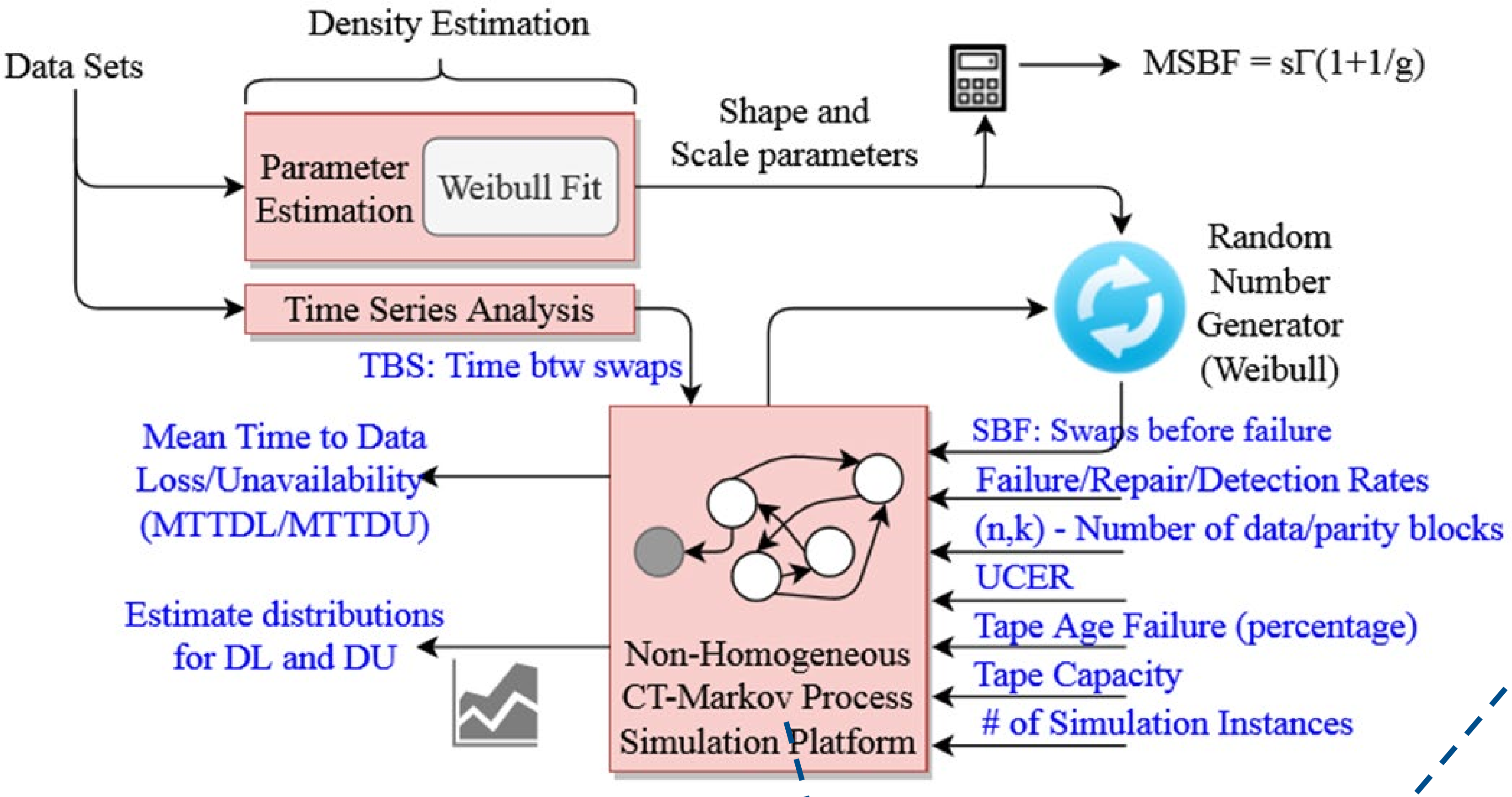
PT : 1st Dimension of 2D ECC which is Per Tape; most efficient parity to use requiring the same tape
2D: 2^{nd} dimension using multiple Tapes; used only for rare cases or to achieve higher transfer rates
C1-C2: LTO's internal ECC format

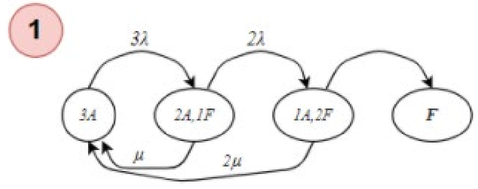
Err Rate	NINES
3.0%	1.6
1.0%	2
0.5%	2.4
0.1%	3

$$NINES = log10(\frac{1}{Prob\ of\ Err})$$

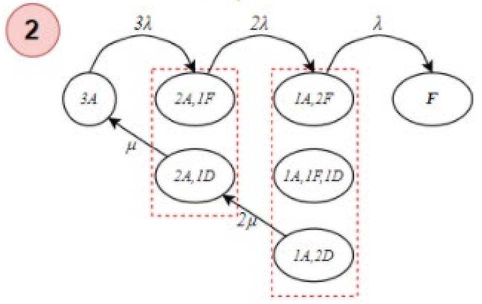
% Error numbers are estimates based on Quantum's experience with LTO media and drives

Quantum’s Data Durability and Availability Computer Model for RAIL System using ECC, Copy, or Combination Protection Policy



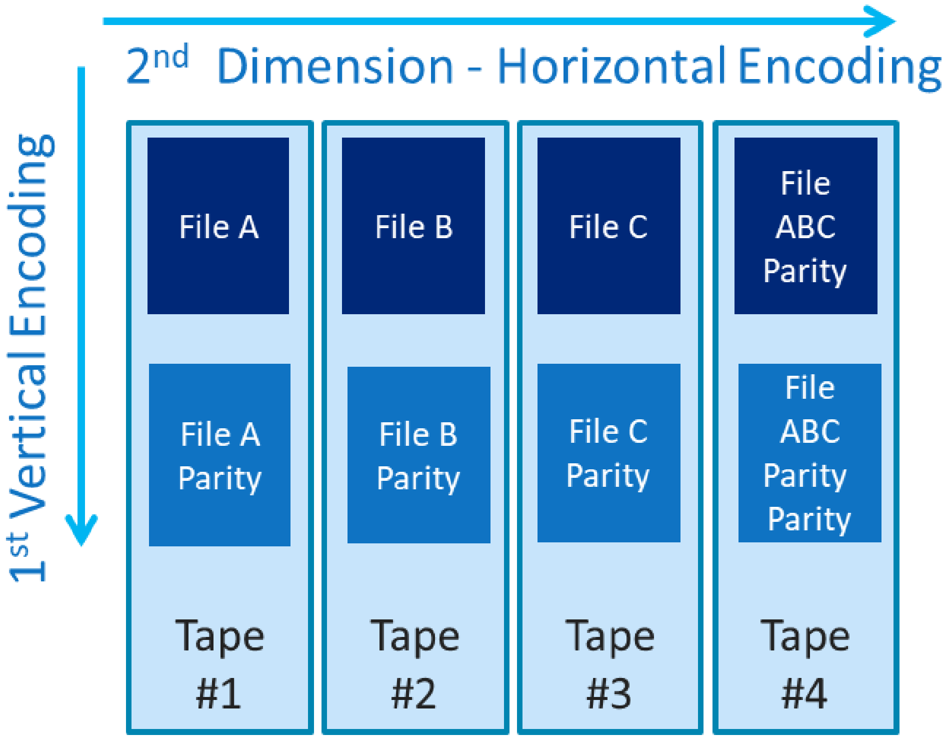


Typical Markov States Available & Fail



Cold Storage Markov States Per Quantum Model: Available, Detected & Fail

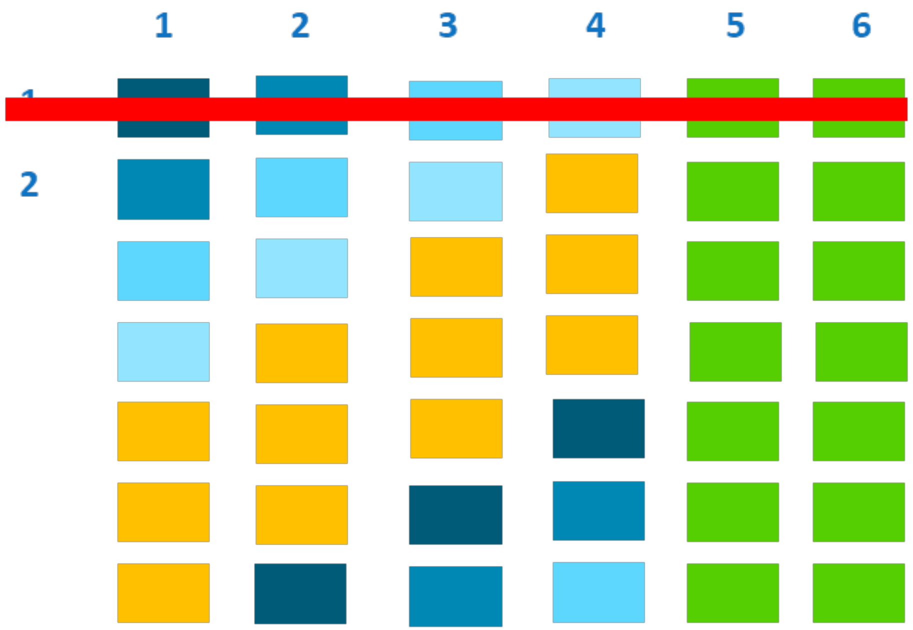
Basic Concept for 2-D ECC



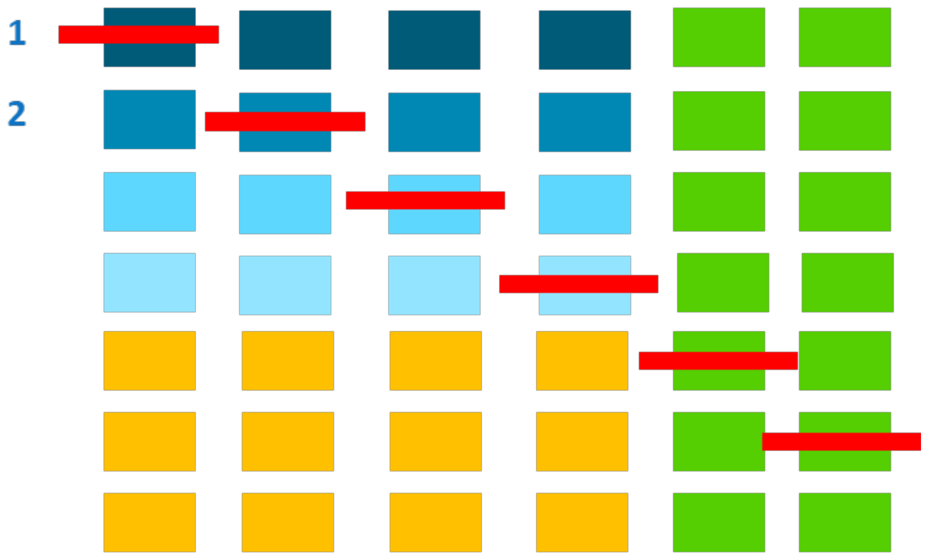
- Per Tape is 1st Dimension
- Multi Tape is 2nd Dimension
- Synchronization during Write is not required
- Read any user file using a single tape but with local protection; this is called “Per Tape EC”
- In case a single tape self-repair is not sufficient, then run global repair using parities across tapes
- Iterative Column - Row - Column type Decoding
- With Per Tape EC, drive skips over errors
- Error recovery on the fly without stops and back hitches also feasible

Spreading Across Tapes for Randomization of Correlated Errors

- Randomize Erasure Coded chunks across tapes using rotation for correlated errors such as tracking at BOT or EOT; for high Areal Density and high TPI LTO’s



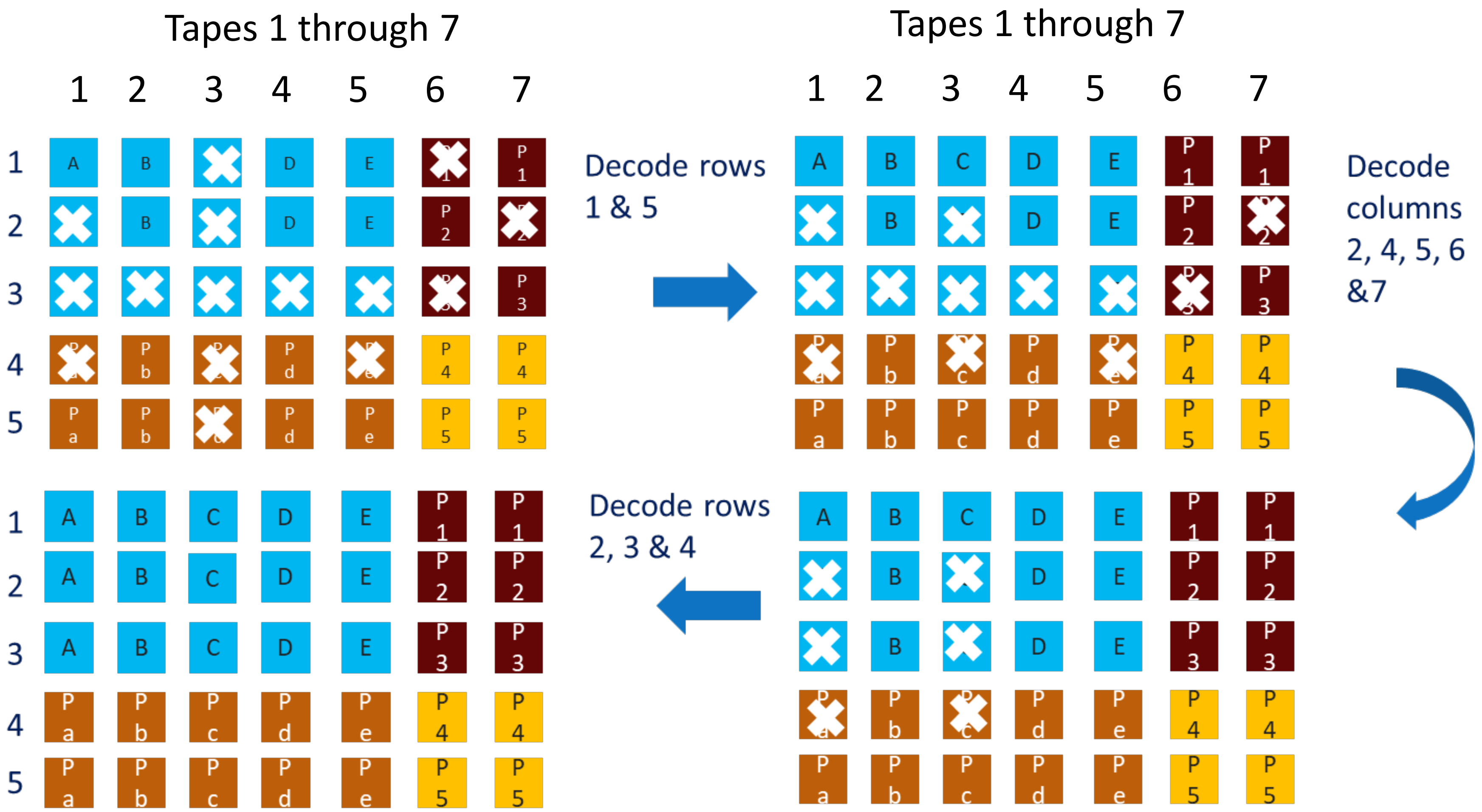
BOT Error without Rotation



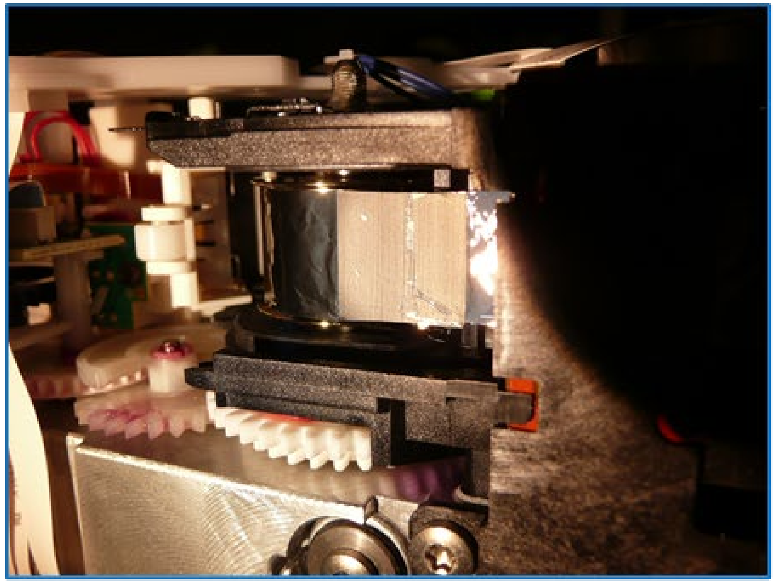
BOT Error with Rotation

2-Dimensional Row & Column Iterative Decode Example

“White X’s” are erasure / error locations on tapes; Multi Tape 7/2 & Per Tape 5/2 EC Policy
Note in this example bot EC policies can correct up to 2 erasures



Per Tape EC Tape Demo



- Tape was erasure coded using Per Tape EC
- Tapes were intentionally damaged; left tape's magnetic section including servo totally removed; right tape was cut , 1-m tape removed and spliced back
- Data was recovered at near max performance using PER Tape EC with 10% overhead

Please take a moment to rate this session.

Your feedback is important to us.

Back Up Slides

Tape: Benefits, Errors, and the Role of 2-Dimensional EC with RAIL

Typical Tape and Drive Errors

Tape C1-C2 ECC format deals mostly with Random Errors or small burst ones

Most real-world errors are Not Random but correlated

Once data is written to tape error free and stored @ Archival Conditions, data is good for 30 years

Data on Tape is safe until you want to load it to a Drive; this is where 2-D ECC comes in

Multi Tape ECC Protection:

- **Human Errors**
- **Lost or Stolen Tapes**
- **Tape Cut / Shred**
- **Tape Stretch due to Tension Problems**
- **Accidental Erasures**
- **Pin Drops**
- **Load problems**

Per Tape ECC Protection:

- **Media & Drive Issues**
- **Stagger Wraps**
- **Tracking problems**
- **BOT, EOT and Directional Errors**
- **Local errors due to media wear**
- **Edge damage**

Nines Modelling Example Using Sample Policies with Comparison to Copy Methods

SUMMARY					
Policy	Copy		Per Tape EC		2-DIM EC
Copy	No	Yes	No	Yes	No
Long-period tape scrub & repair on demand			Yes	Yes	Yes
# of Total Tapes	1	2	1	2	12
# of Total Data Tapes	1	1	1	1	9
Storage Overhead	0%	100%	15%	130%	53%
Errors Corrected excluding BER	No	All	Some	All	All
Full Tape Damage robustness	No	Yes	No	Yes	Yes
Per Tape Damage robustness	No	Yes	Yes	Yes	Yes
9's for BER corrected by Tape Format	11	11	11	11	11
Single Tape 9's excluding full tape errors	NA	4	11.0	11.0	11.0
Nines based on for all errors	< 5 *	< 10*	< 5*	< 13	> 15

** Human related errors such as lost and misplaced tape are excluded*

2-Dimensional Erasure-Coded RAIL with LTO

- Redundant Array of Independent Libraries
- RAIL and 2-D ECC
- High Availability 9’s for robotic errors
- High performance for Random Restores