

Redundant Array of Independent Tape Rait

Mike Fisher

Storage Technology Corporation

280 N High St, Ste 850

Columbus OH 43215

Phone: +1-614-227-5852 FAX: +1-614-227-5888

E-mail: michael_fisher@storagetek.com

**Presented at the THIC Meeting at the Naval Surface Warfare
Center Carderock, 9500 MacArthur Blvd**

West Bethesda MD 20817-5700

October 4, 2000

The Premier Advanced Recording Technology Forum

THIC Inc.



STORAGETEK

INFORMATION *made* POWERFUL

Redundant Array of Independent Tape Rait

Jim Hughes unable to present

Mike Fisher

Storage Technology Corporation

Michael_Fisher@StorageTek.com

Presented at the THIC Meeting at the Naval Surface

Warfare Center Carderock

October 4, 2000

ASCI RAIT prototype

- *Accelerate the creation of a virtual tape drive*
 - *Commercial / Reliable / Fault Tolerant / High performance*
- Initial customer ASCI Pathforward:
- Initial Target performance:
 - 1GB/s archival bandwidth
- Demonstrate IBM's HPSS compatibility
- Future COTS feature development
 - based on StorageTek's COTS Access Controller SN6000

Implementation Basics

■ Requirements

- High Throughput
- Robustness / *intra*
- Robustness / *inter*
- Virtualization

■ Solutions

- ⇒ RAIT '0'
 - ⇒ *Striping*
- ⇒ Media Errors
 - ⇒ *Drive internal ECC 'on'*
- ⇒ RAIT '5+'
 - ⇒ *Parity redundancy + X + ()*
- ⇒ One Volume
 - ⇒ *'Hidden' RAITape*
 - ⇒ *'Hidden' RAILibrary*

RAIT Volumes

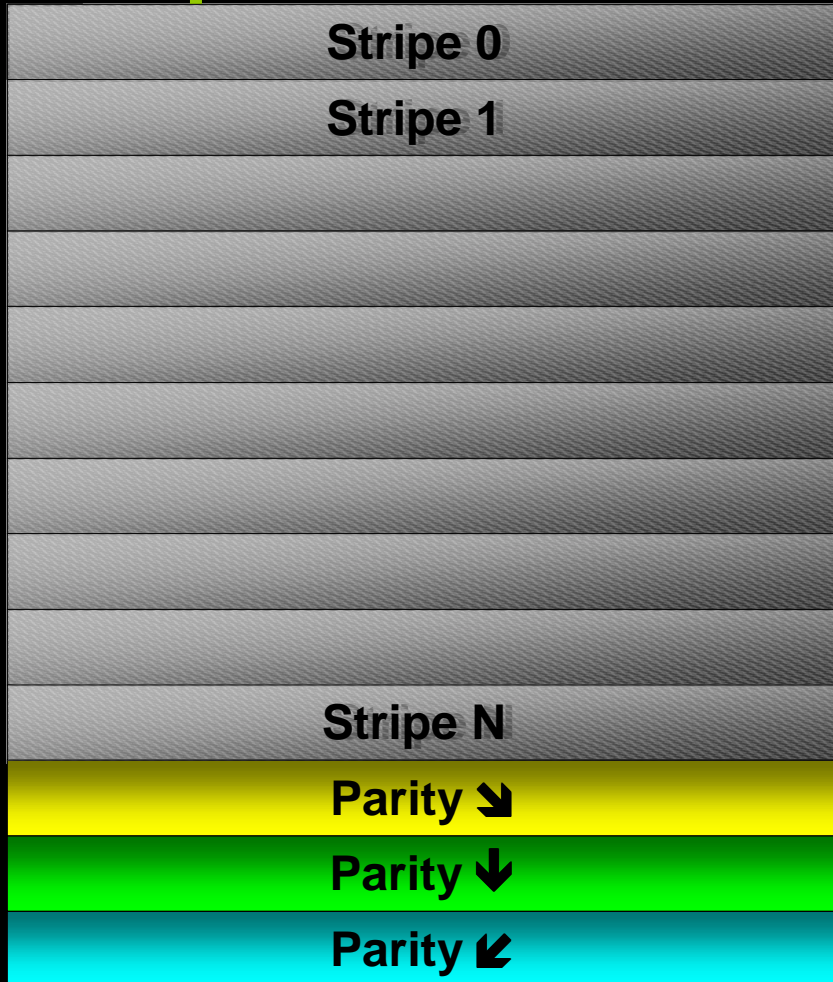
- **n+p sets Quality of Service**
 - n: number of data stripes, p: number of parity stripes
 - Defined at scratch mount time
- **Reliability**
 - Complete loss of **p tapes** / stripes / drives over n+p
 - ◆ When Reading & Writing
 - ◆ Unique cross parity system
 - ◆ Unique self definition mechanism
 - dynamic volume 'width' & 'length'
 - ◆ 8+2 better than 8+8 mirroring
- **Performance**
 - \approx n x unit drive throughput
- **Compressibility**
 - ◆ Parity rotation ensures averaging over n+p tapes set

What Parity Is Required

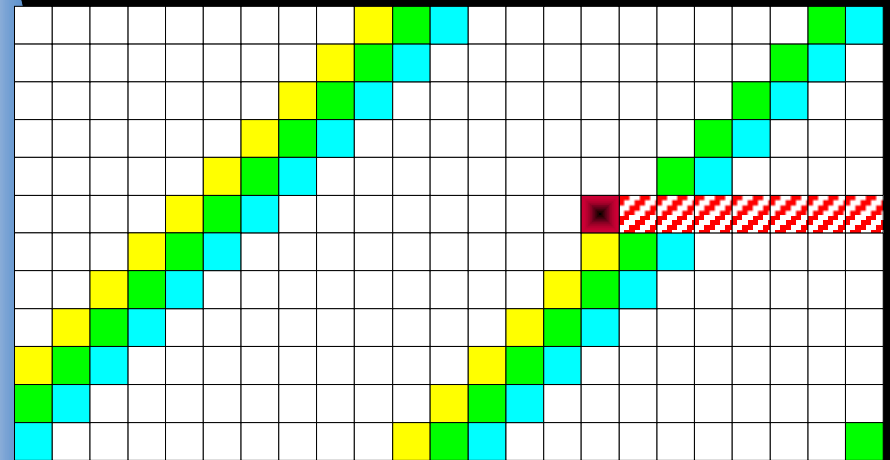
- Number of 160 GB Virtual-Volume Reads before Data Loss

Protection Scheme	Total # of tapes	Number of Missing Tapes		
		0 missing	1 missing	2 missing
None	8	10^2	1	1
1 parity	8 + 1	10^9	10^2	1
2 parity	8 + 2	10^{16}	10^9	10^2
3 parity	8 + 3	10^{23}	10^{16}	10^9
4 parity	8 + 4	10^{30}	10^{23}	10^{16}
Mirroring	8 + 8	10^{10}	10^3	15

Adaptive Cross-Parity for RAIT

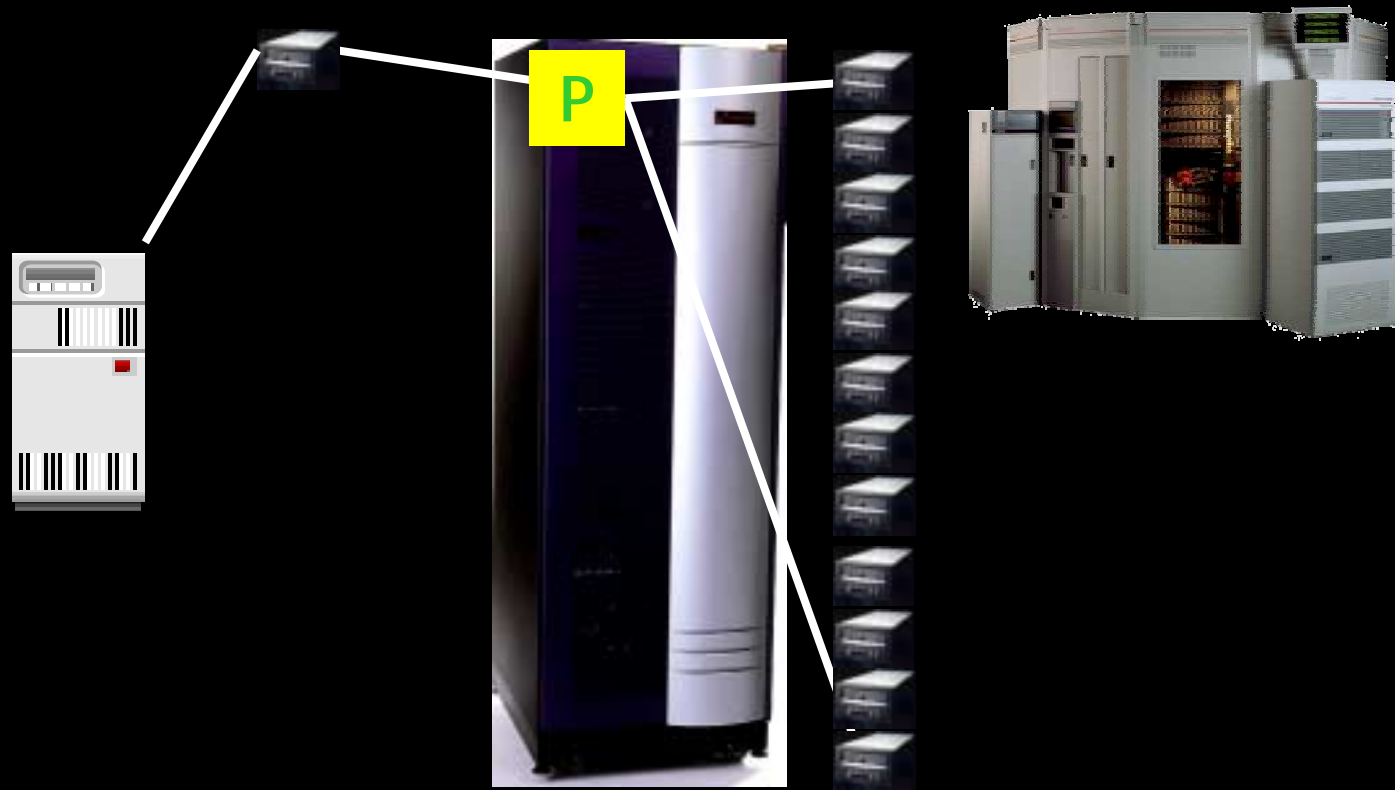


Patent Pending



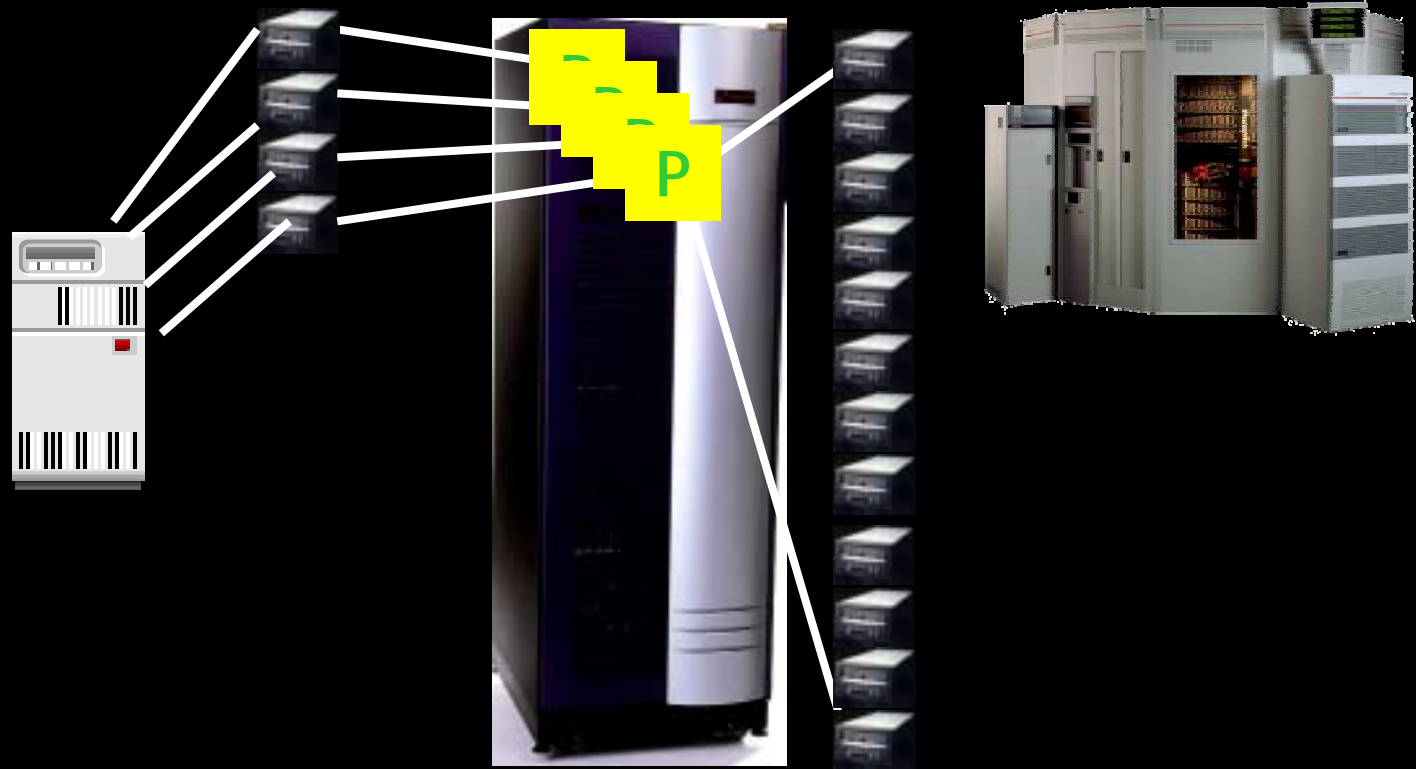
Drive compression still effective

80MBytes/second
0.29 TBytes/hour



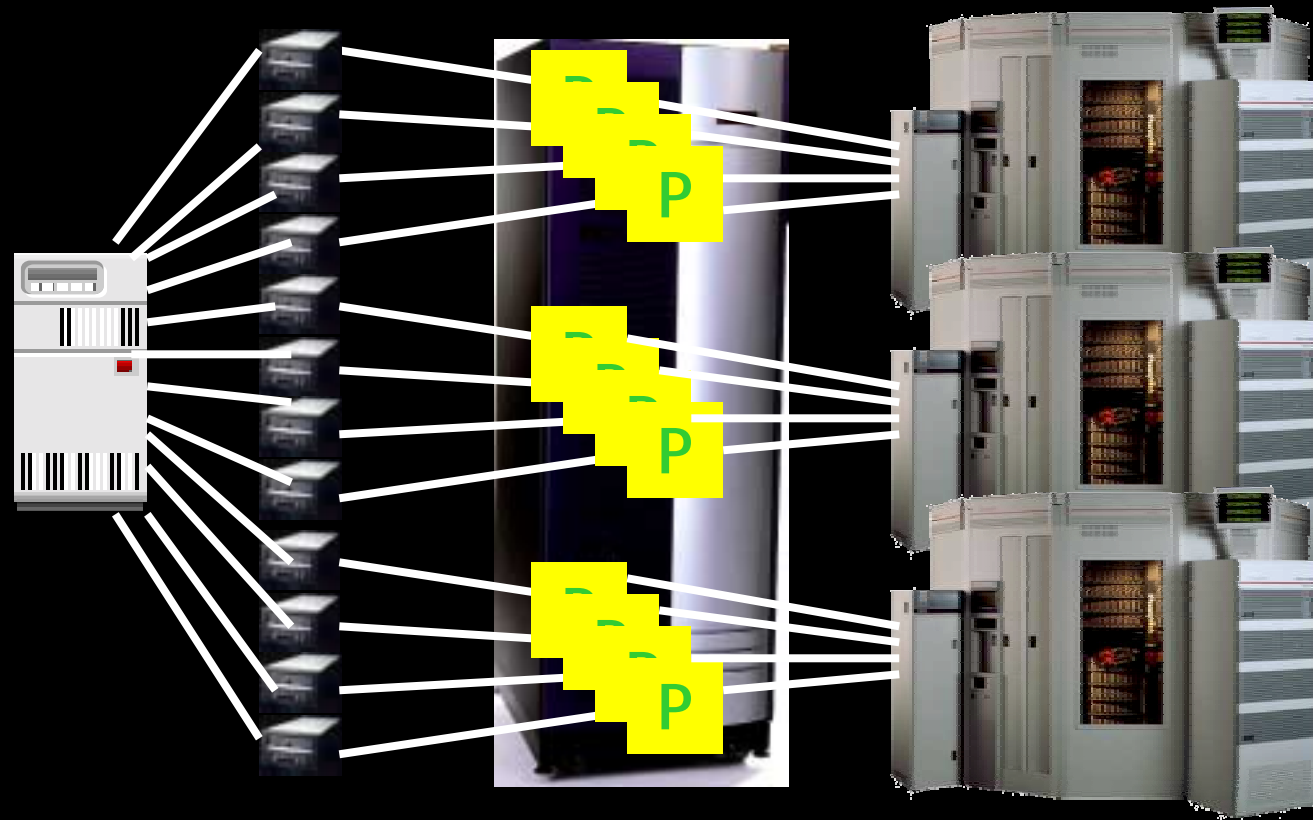
■ Without compression

320MBytes/s
1.15 TBytes/hour



■ Without compression

960 MBytes/s
3.46 TBytes/hour



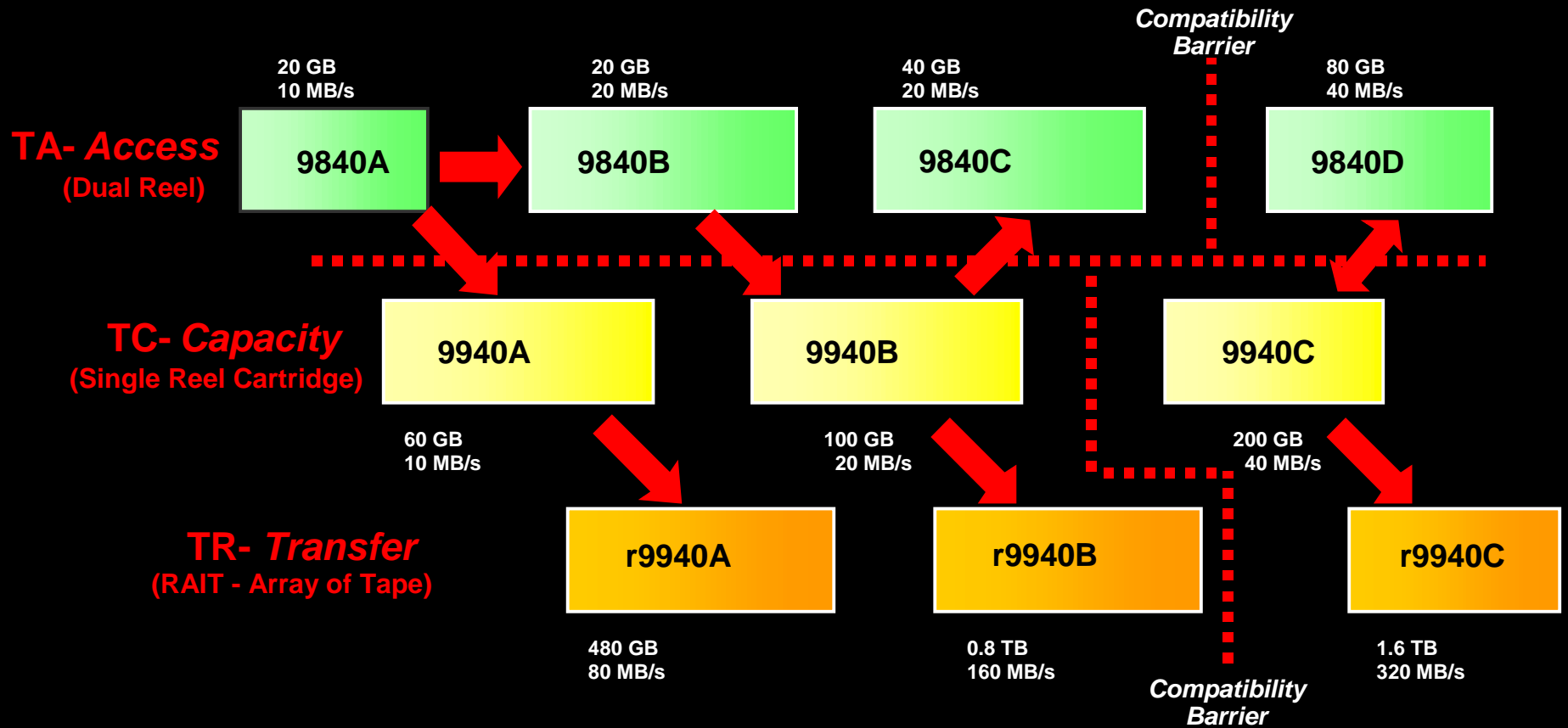
■ Without compression

OffLine Utilities

- **Real Offline: ejected RAIT set**
 - Site not equipped with RAIT
 - Data Mining with reduced equipment
 - Repairs
- **RAIT Offline: RAIT volumes in silos**
 - Repairs

Prototype vs Production

- **STK is committed to COTS RAIT in 2001.**
- **Additional features**
 - **More drive types. DLT (and compatible), 9940, Magstar, etc.**
 - ◆ **We are looking at LTO and LOTS.**
 - **Drives shared as pool**
 - ◆ **Singles and Raits allocated as needed**
 - **Up to 16 flexible Rait controllers per chassis**
 - ◆ **Integrated Switch, Single pool**
 - **Support other applications**
 - ◆ **Veritas, Reel, LSCi, etc...**



RAIT

■ Who needs RAIT?

- Scientific community
 - ASCI Case : job CPU avg. = 6 hours < result storage
- Commercial database users
 - ◆ **Short timeslot archive / backup / restore!**
 - “one hour for one TB” Case

■ Non Replicated Archive

- A cartridge fell out of the truck

■ Efficiency of SN6k Architecture:

- COTS Technology upgrade path