High Performance RAIT

Jim Hughes *STK fellow*
Jacques Debiez *R&D mgr.*
Charles Milligan *R&D mgr.*

RAIT : only RAID for Tape?

- Increased Performance
  - Up to 10x faster
  - Backup / Archive faster
  - Restore faster
- Increased Reliability and Availability
  - Drive failures non-disruptive
  - Damaged/ lost tapes non-disruptive
  - Media errors non-disruptive
- Increased… but

  with scalable Performance & Reliability
- Virtual Tape Device for ‘compatible fit’
  - Transparent mounts of an array of tape drives
  - Transparent access to an array of tape volumes

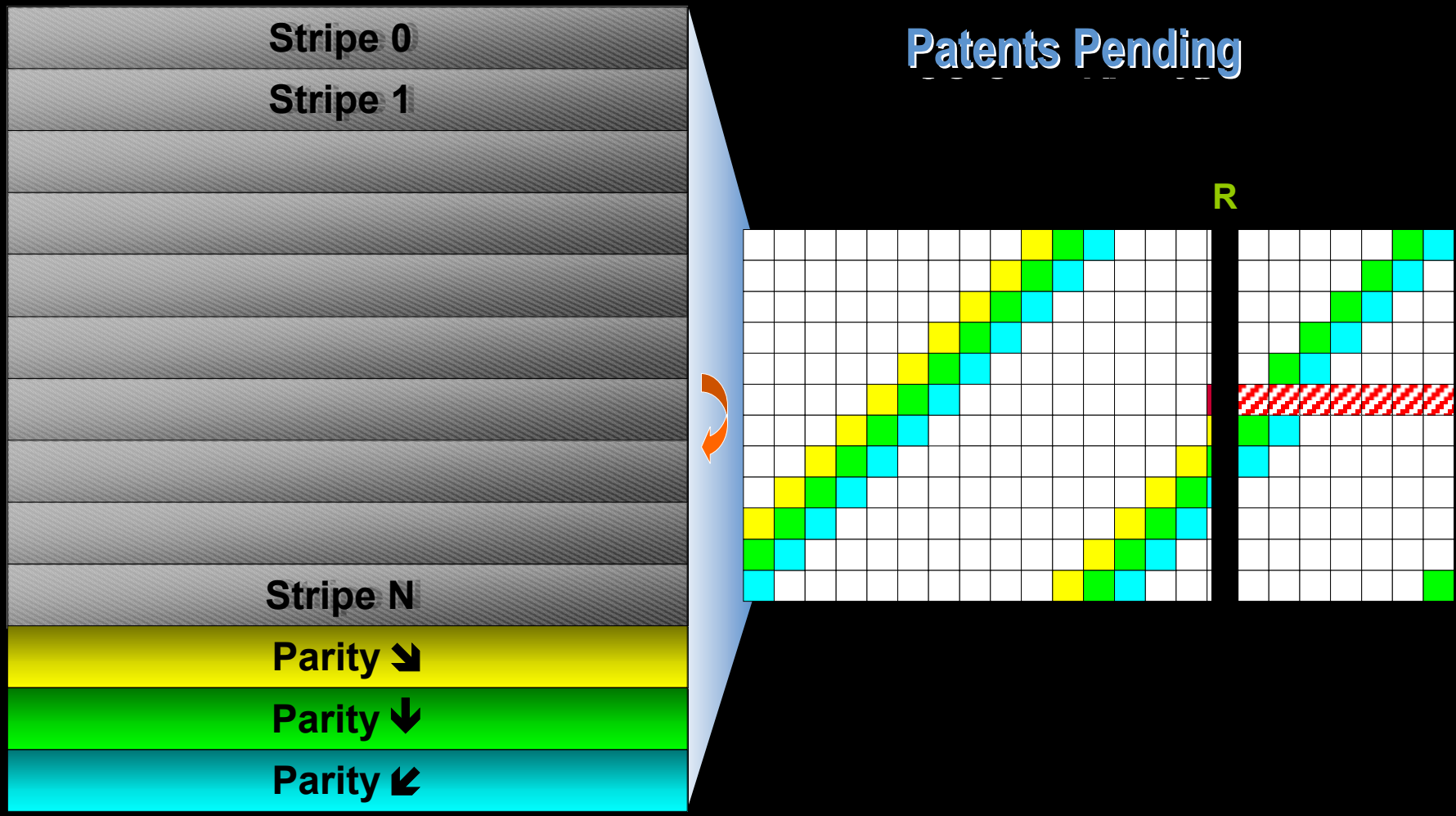
Implementation Basics: RAID for Tapes

- Requirements
  - High Throughput
  - Robustness / physical
  - Robustness / system
  - Self-defining array
    - RAIT 'Volume'
  - RAIT Virtualization
- Solutions
  - RAIT '0'
    - Striping
  - Media Errors
    - Drive internal ECC 'on'
  - RAIT '5+'
    - Parity redundancy + ✕ + ↻
  - Specific Tape Format
    - Information on each tape
  - One Volume / One Drive
    - 'Hidden' RAITape
    - 'Hidden' RAILibrary

RAIT Volumes Primer

- n+p sets Quality of Service
  - n number of data stripes
  - p number of parity stripes
- Performance
  - ≈ n x unit drive throughput
- Reliability / Non-disruptive operation
  - Complete loss of p tapes / stripes / drives over n+p
    - When Reading & Writing
    - Unique cross parity system — patent pending
    - Unique self definition mechanism — patent pending
      - dynamic volume 'striping width' & 'length'
    - 8+2 improves failure rate per 10^9
    - 8+2 simply better than 8+8 mirroring
- Compressibility
  - Parity rotation ensures length averaging over n+p tapes set

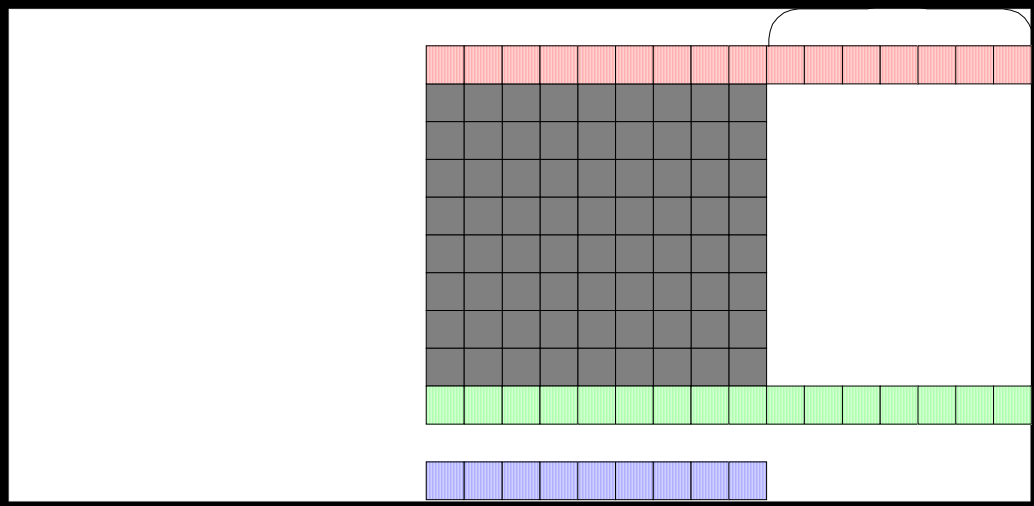
Adaptive Cross-Parity for RAIT



Parity Coverage / Write / 8.3

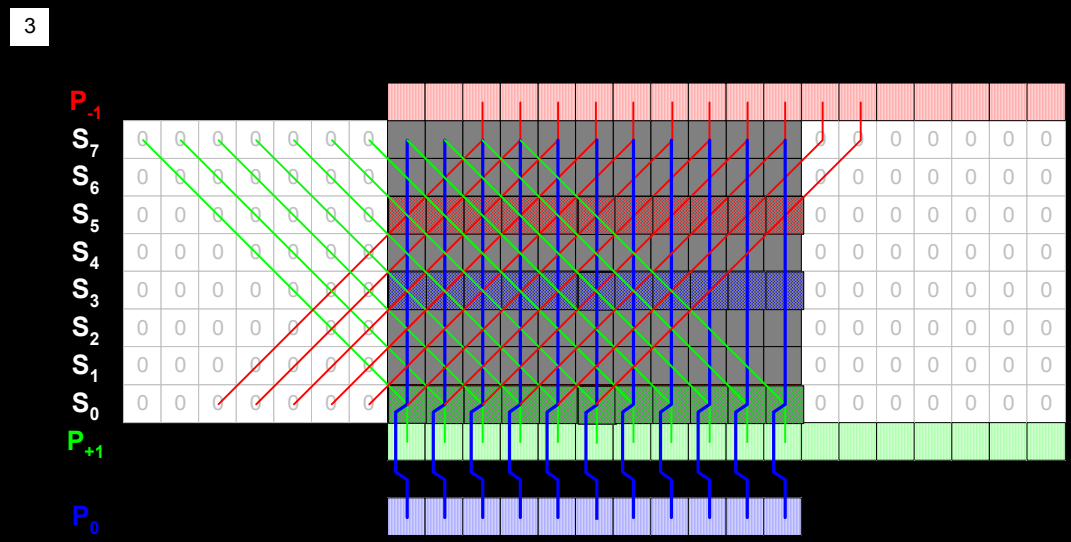
Overall Constant Length

9



Data Reconstruction / 8.3

Iterative process



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

Overall Architecture

- Virtualization : RAIT hidden to the Client Host
  - RAIT Volumes
    - one RAIT volume = one single entity (mount…)
  - High Performance / High Reliability RAIT Drives
    - access / R / W on Virtual RAIT Volumes
  - RAIT Automation Management
    - Automatically handles Virtual RAIT Volumes
  - RAIT Library / RAIL (one ACS)
    - Virtual RAIT Volumes Storage
- Separate RAIT Administration
  - RAIT Administration Interface
    - manages “RAIT specific” operations / monitoring

Architecture important features

- Virtual to Physical Drives Mapping
  - Optimize availability of physical drives to RAIT ‘drives pool’
  - Dynamic mapping of physical drives
    - takes any available drives within resource pool

    ‘optimized spares’
- RAIT Volume Information stored in a specific database
  - faster RAIT management wrt. self-defining information on tapes
- RAIT Volumes managed in ‘pools’
  - RAIT volume creation sets Quality of Service: n,p
  - “OnLine” pool in ACS / implicit n,p
  - “Repair” pool in ACS /
  - “Shelf Storage” Volumes still managed ! import /export function
- Offline Tools:
  - RAIT volumes accessible out of RAIT system
- RAILibraries: tolerance to Library failure

Conclusion

- Who needs/wants RAIT?
  - Supercomputers owners
    - ASCI Case : presently, CPU time avg. < archiving time
    - Non replicated archive : too expensive to mirror
  - "Commercial" Short timeslot archive / backup & restore
    - one hour for one TB just 4 RAIT drives
    - Non-disruptive operation / reliability / data availability
    - mirror for disaster recovery site RAIT51
  - RAIT-1, Simple Mirroring
    - Also possible using Volume virtualization