HDD Ghosts, Goblins and Failures in RAID Storage Systems

Jon G. Elerath, Ph. D.
Reliability Consulting
www.ElerathReliability.com

February 19, 2013

Agenda RAID > HDDs > RAID

- Redundant Array of Independent Disks
  - What is *RAID* storage?
  - How does *RAID* work?
  - Common configurations
- Why is RAID Needed?
- HDD "ghosts, goblins and failures" (errors and failures)
- HDD Field Reliability for RAIDed HDDs
- HDD Reliability Metric
- Realities of RAID Operations
- Impact on RAID Reliability Estimates

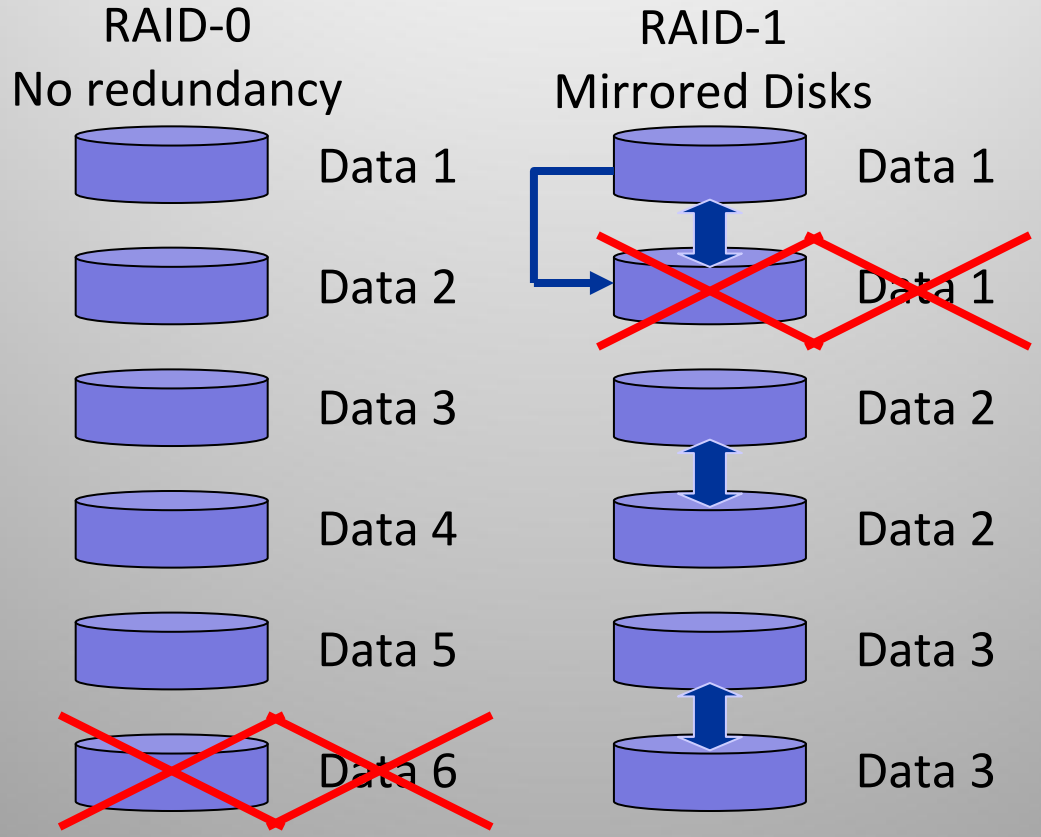
HDDs in RAID Storage Systems

- What is *RAID* storage?
  - Redundant Array of Independent Disks
- How does RAID work?
  - Data Redundancy
  - Parity: Exclusive “OR” logic
  - Single parity creates one bit for a data stripe
  - Dual parity creates two parity bits across different stripes of data
- Why is it needed?

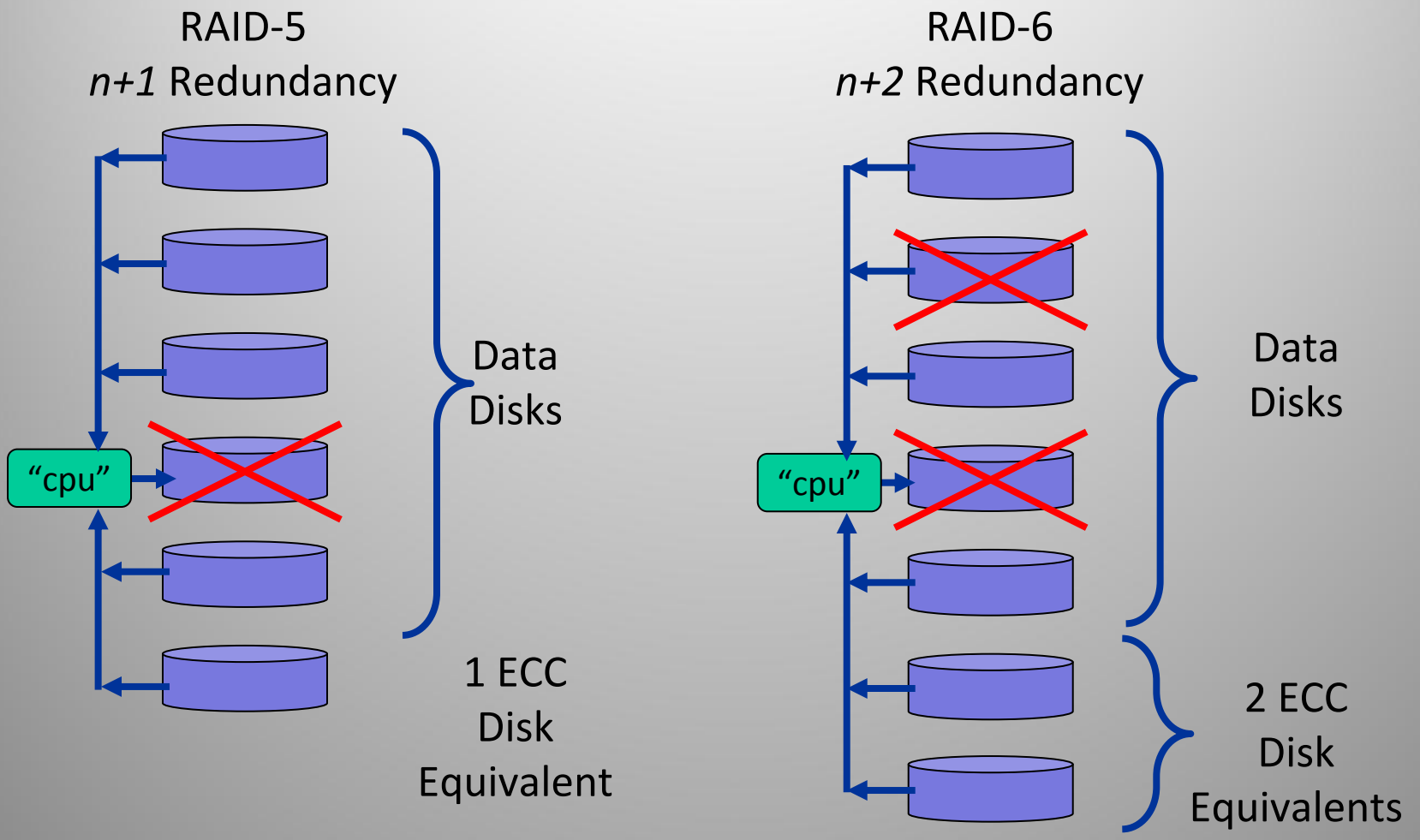
RAID Storage Systems

- Common configurations
  - RAID-0 No redundancy
    - Any failure causes data loss
  - RAID-1 Mirrored disks (*n*+*n*)
    - all data is saved on two identical HDDs
    - requires both disks to fail concurrently
  - RAID-4 and RAID-5 (*n*+1)
    - *n* data disks + 1 parity disk
    - requires 2 “failures” to lose data
  - RAID-6 (RAID-DP) (*n*+2)
    - *n* data disks + 2 parity disks
    - requires 3 “failures” to lose data

RAID Operation (Overly Simplified)



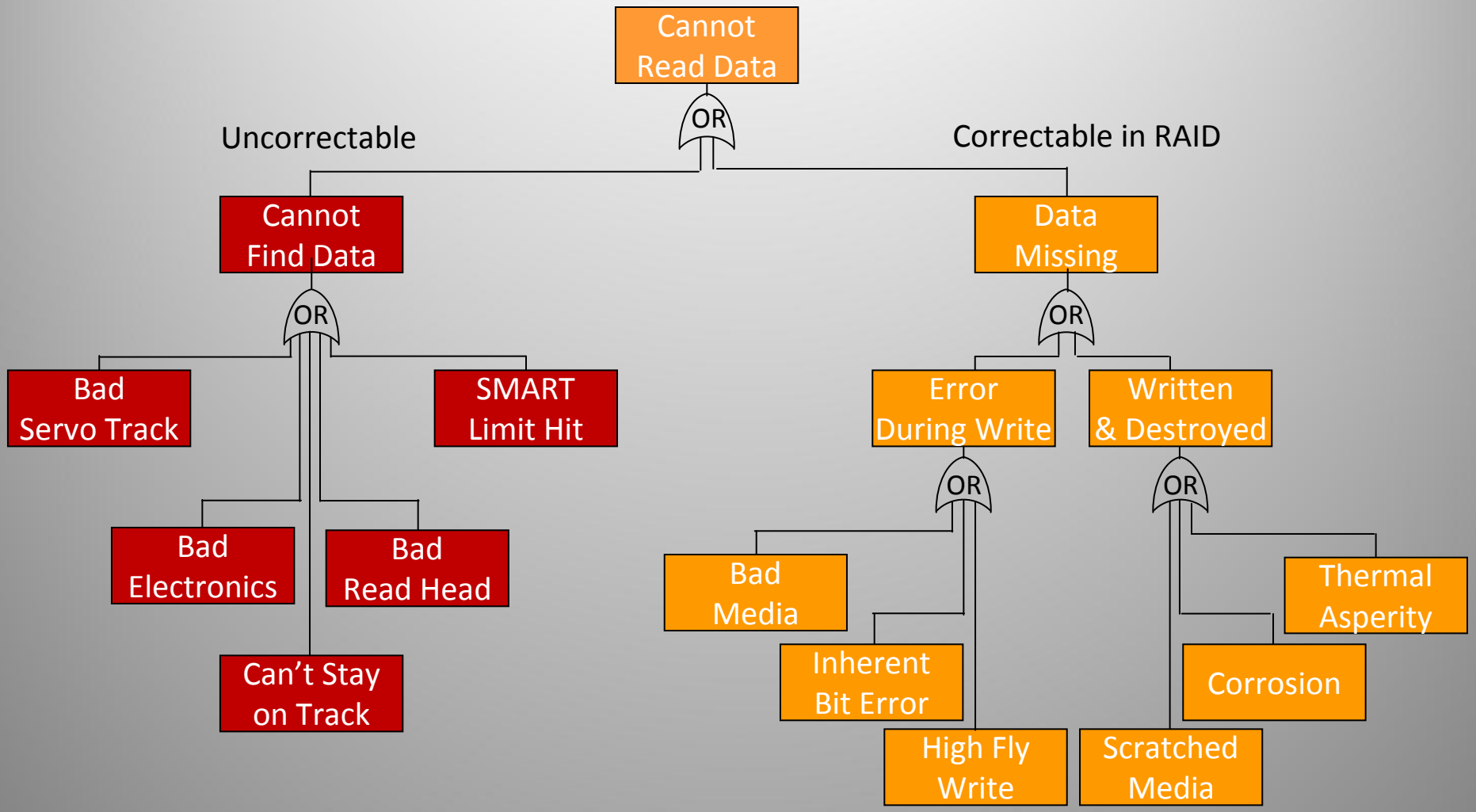
RAID Operation (Unbelievably Oversimplified)



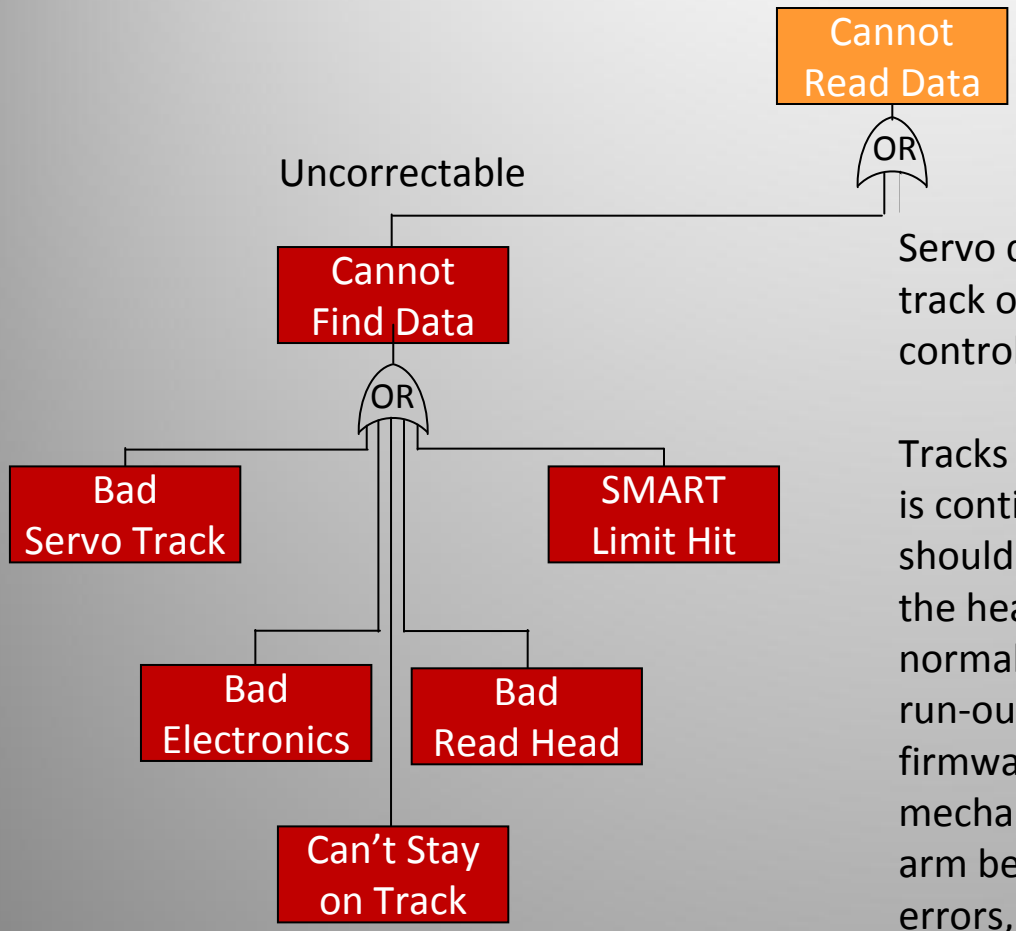
HDDs in RAID Storage Systems

- What is *RAID* storage?
  - Redundant Array of Independent Disks
- How does RAID work?
  - Data Redundancy
  - Parity: Exclusive "OR" logic
  - Single parity creates one bit for a data stripe
  - Dual parity creates two parity bits across different stripes of data
- Why is RAID needed?
  - "Disk drives suck" – A perception by Dan Warmenhoven, CEO NetApp (circa 2005)

HDD Events (Ghosts, Goblins and Failures)



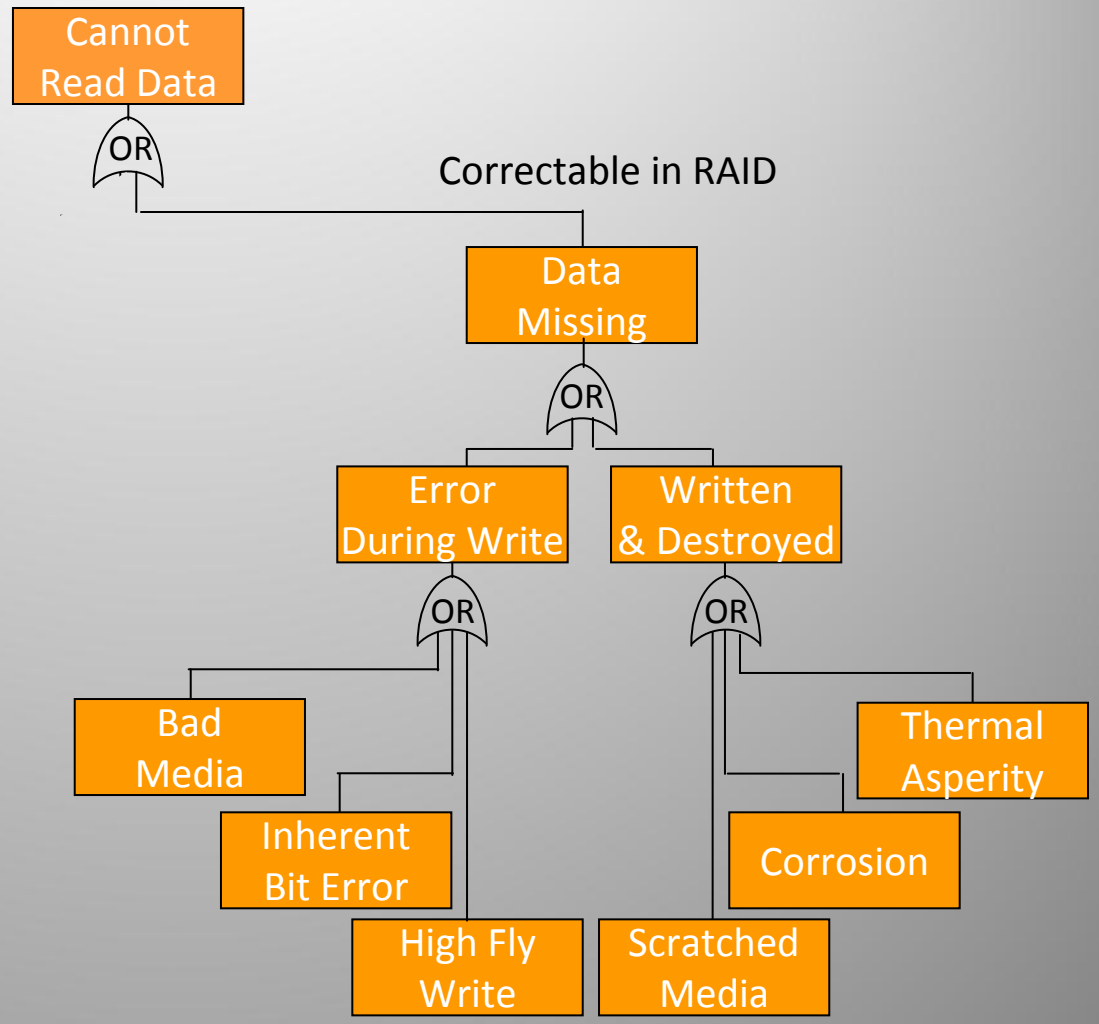
HDD Events (Ghosts, Goblins and Failures)



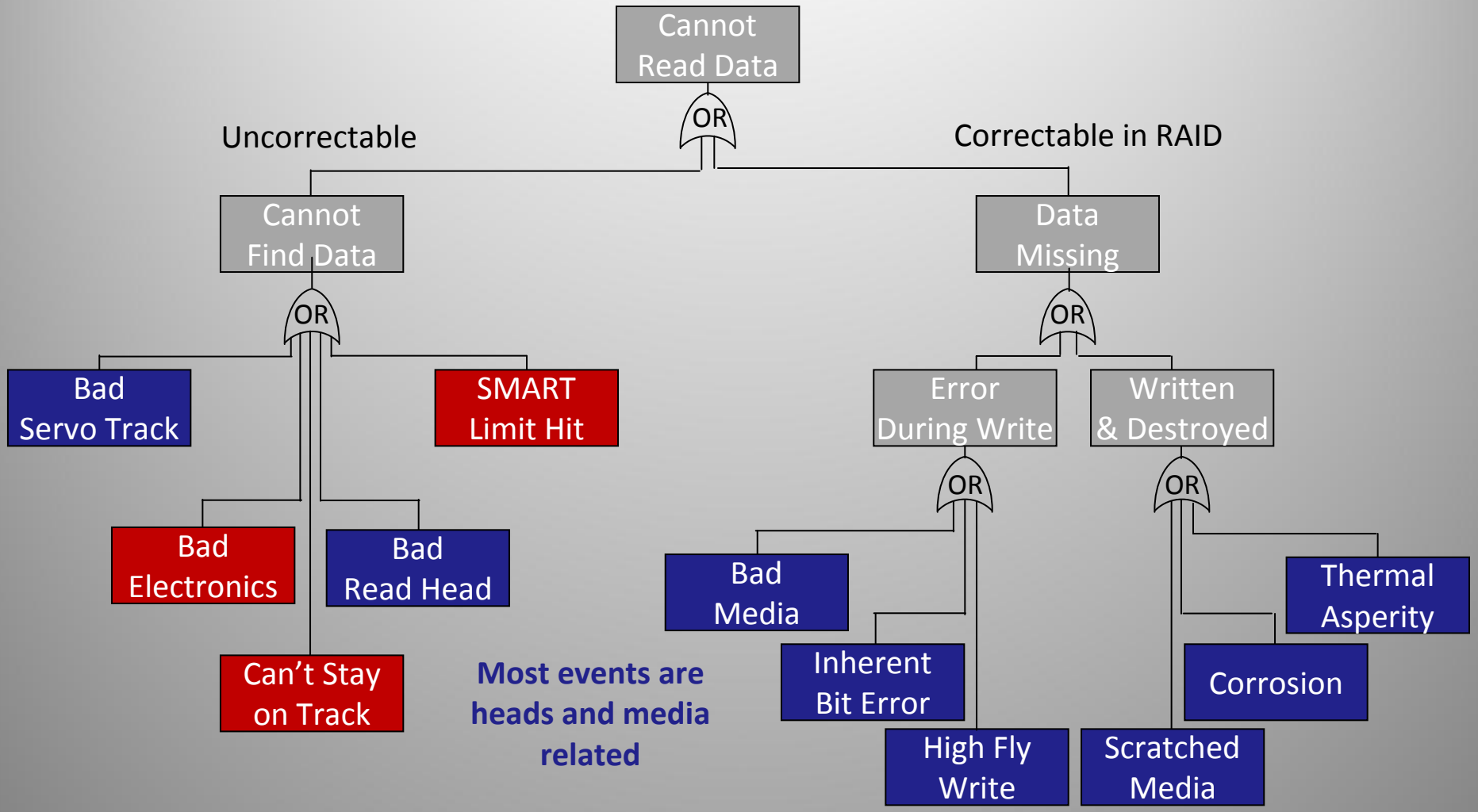
Servo data is written at regular intervals on every data track of every disk surface. The servo data is used to control the positioning of the read/write heads.

Tracks on a HDD are not perfectly circular. Head position is continuously measured and compared to where it should be. A position error signal is used to reposition the head over the track. Repeatable run-out is part of normal HDD head positioning control. Non-repeatable run-out (NRRO) cannot be corrected by the HDD firmware since it is non-repeatable. NRRO, caused by mechanical tolerances from the motor bearings, actuator arm bearings, noise, vibration and servo-loop response errors, can cause the head positioning to take too long to lock onto a track and ultimately produce an error.

HDD Events (Ghosts, Goblins and Failures)



HDD Events (Ghosts, Goblins and Failures)

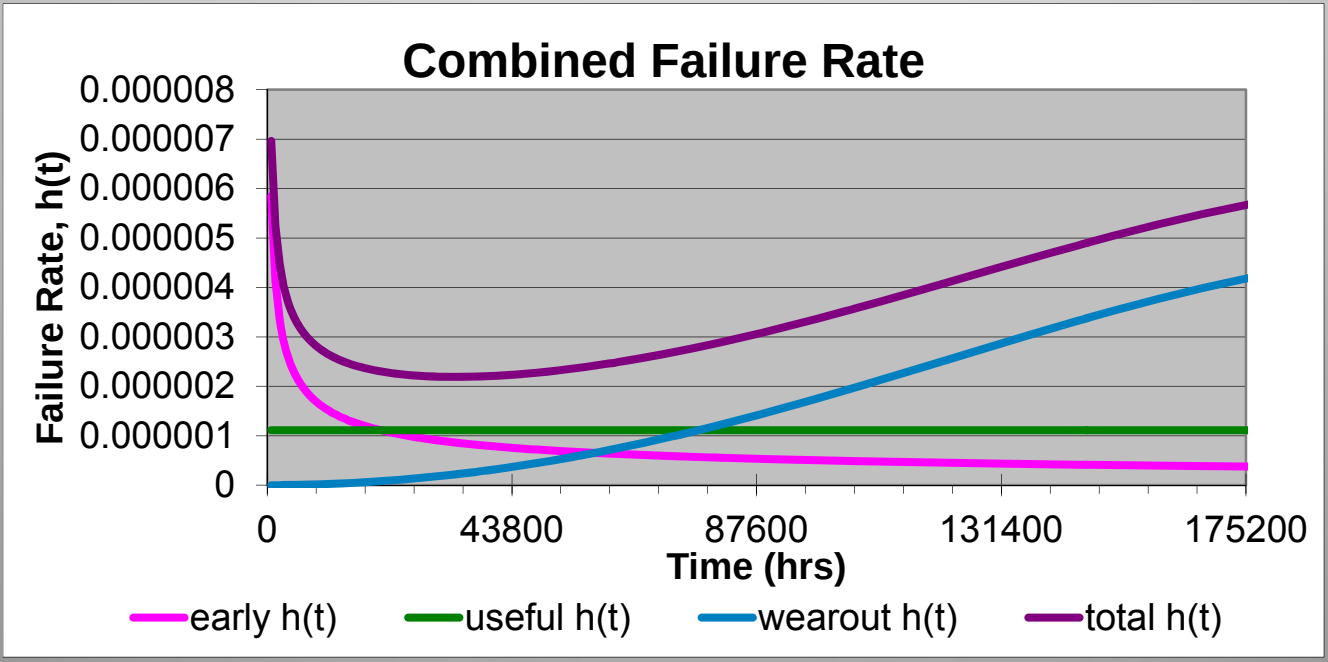


Uncorrectable Errors and Field Failures (RAID)

- Uncorrectable “errors” constitute what most people think of as failures
  - HDD “crashed”
  - Won’t spin
- Most field data is based on these ONLY
- Field data do not match vendor specifications
- HDD reliability depends on the HDD’s
  - Manufacturer
  - Product Family from the manufacturer
  - Capacity
  - Age
  - Use conditions

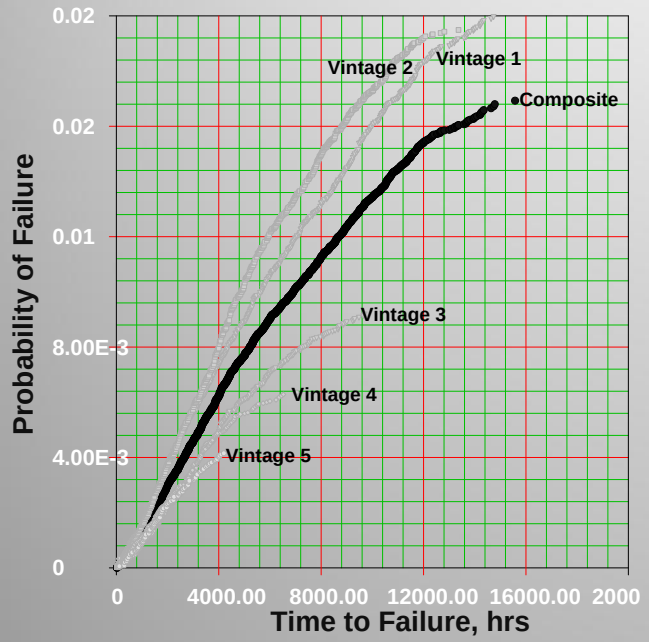
Mean Time Between Failures MTBF

- MTBF is ***NOT*** a good metric for measuring HDD reliability (Elerath, J. G. 1998. Specifying Reliability in the Disk Drive Industry. *Proc. Annual Rel. and Maint. Symposium*, IEEE. pp194-199.)
- What does MTBF *really* mean to a consumer?
- Does NOT include correctable events

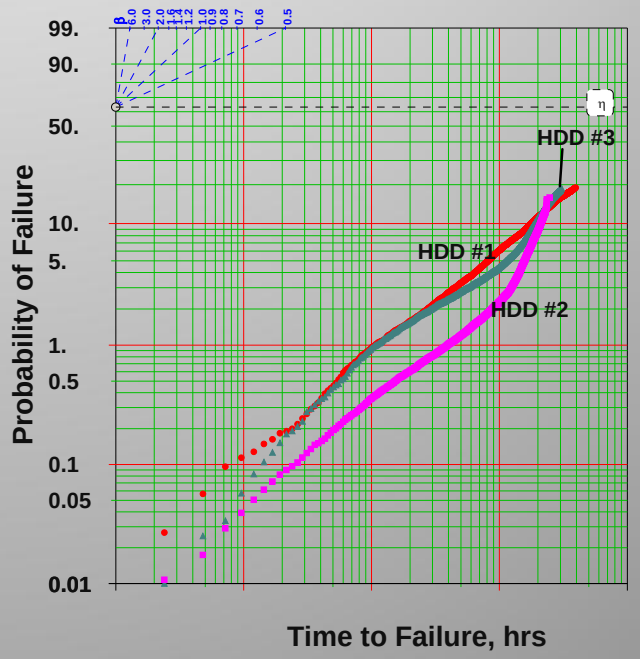


Real HDD Field Reliability Data

- Vintage affects reliability



- Slightly decreasing failure rates, then increasing rate (esp. HDD#2)



Reality of RAID Operations

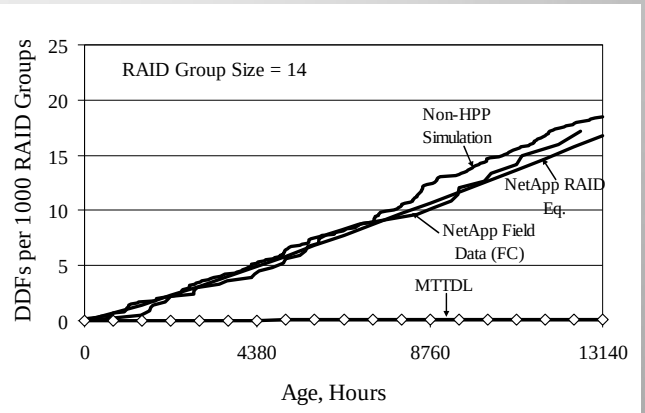
- RAID 4 & RAID 5 *(n+1)*
  - Most likely combination of failures is a unknown data corruption in a small amount of data (several sectors) followed by a concurrent, non-correctable error in a different HDD
  - During reconstruction onto a new HDD, the corrupted data cannot be recovered
- RAID 6 *(n+2)*
  - Most likely combinations are 1 data corruption, followed by two uncorrectable failures (HDD failures) simultaneously. The second failure occurs before the first can be incorporated into the RAID group and reconstructed. Far less likely than RAID-4 or RAID-5 data losses, but still measurable.

Reality of RAID Operations

- Read-after-write is too slow; not used. Checked only on READ.
- Data corruptions can occur
  - During READ operations
  - During WRITE operations
  - Whenever an HDD is spinning
  - Sometimes when the HDD is not spinning (depending on the design)
- "Small" Data Corruptions
  - All HDDs have error recovery algorithms built into the HDD
  - "Small" errors are corrected on the fly (less than one revolution)
- "Large" Data Corruptions
  - Require ECC and data from other HDDs in the RAID group
  - Disk scrubbing by the RAID controller proactively finds these "larger" errors and corrects them to prevent data loss at the RAID group level

Impact on RAID Reliability Estimates

- Mean Time To Data Loss (MTTDL)
  - Concept statistically flawed[1]
  - Doesn't include undiscovered defects
  - Ridiculously optimistic
- RAID 4/5 Reliability Estimates
  - Monte Carlo Simulations[1]
  - Simple Equation[2]
- RAID 6
  - Monte Carlo Simulations
  - Simple Equation[3]



http://raideqn.netapp.com

[1] Elerath & Pecht. A Highly Accurate Method for Assessing Reliability of Redundant Arrays of Inexpensive Disks (RAID). *Trans. on Computers*. 58, 3, March 2009.

[2] Elerath. A Simple Equation for Estimating Reliability of an N+1 Redundant Array of Independent Disks (RAID). DSN 2009.

[3] Elerath & Schindler. Beyond MTTDL: A closed-form RAID 6 reliability equation. WIP. FAST 2013.

Conclusion

- HDD head/disk interface is critical to achieving RAID reliability
- RAID designers must do background media scrubbing
- Scrub frequency is critical to RAID reliability
- Larger HDDs (2+ TB) have more opportunities to develop media defects and require long times to scrub
- Field data does not map to true defect rates