Carnegie Mellon University

Gotta have HeART

Improving storage efficiency by exploiting disk-reliability heterogeneity

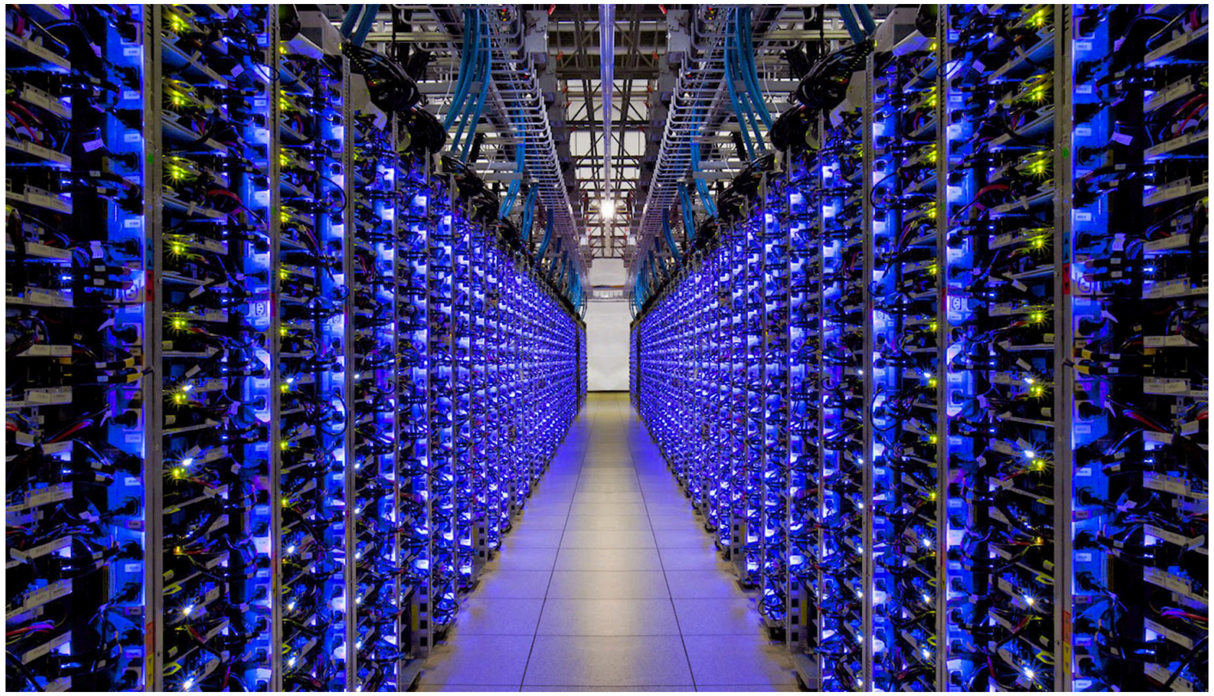
Saurabh Kadekodi

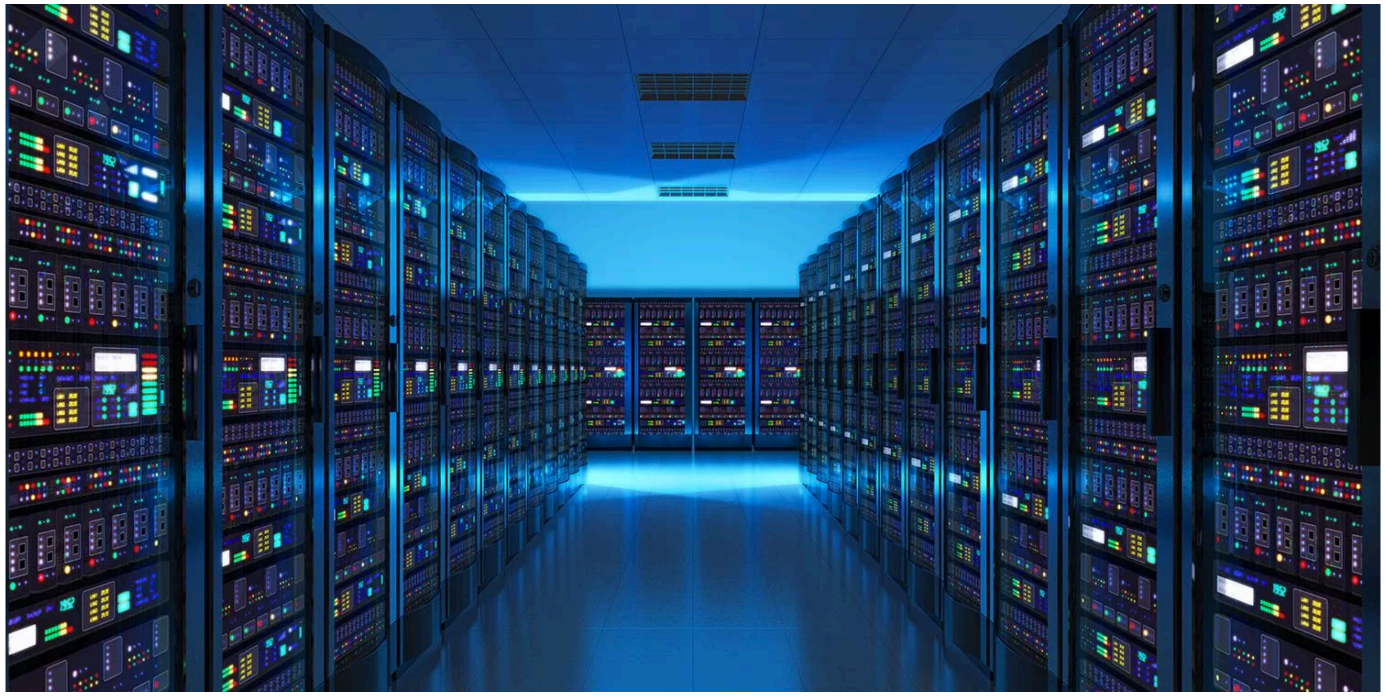
Rashmi Vinayak

Greg Ganger

Cluster storage systems

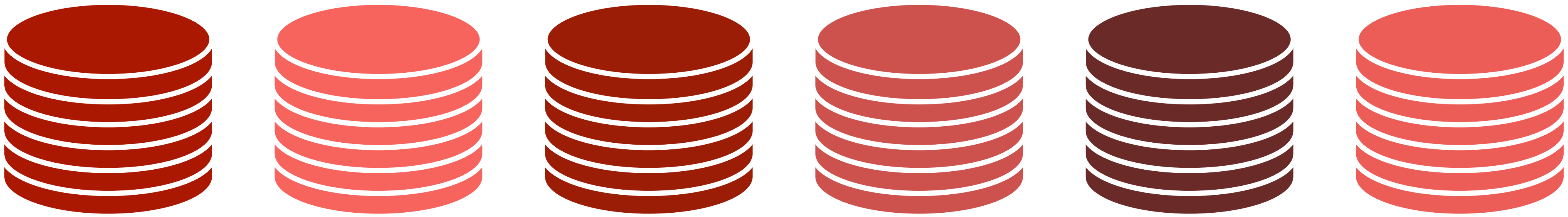
- Storage subsystem of distributed systems





- Thousands to millions of disks in primary storage tier
- Built incrementally according to demand

Reliability heterogeneity in disks



- Disk fleet has heterogeneous collection of disks
- **Different in reliability**
  - Manufacturing differences across makes/models
  - Different vibration / temperature experiences
  - I/O churn

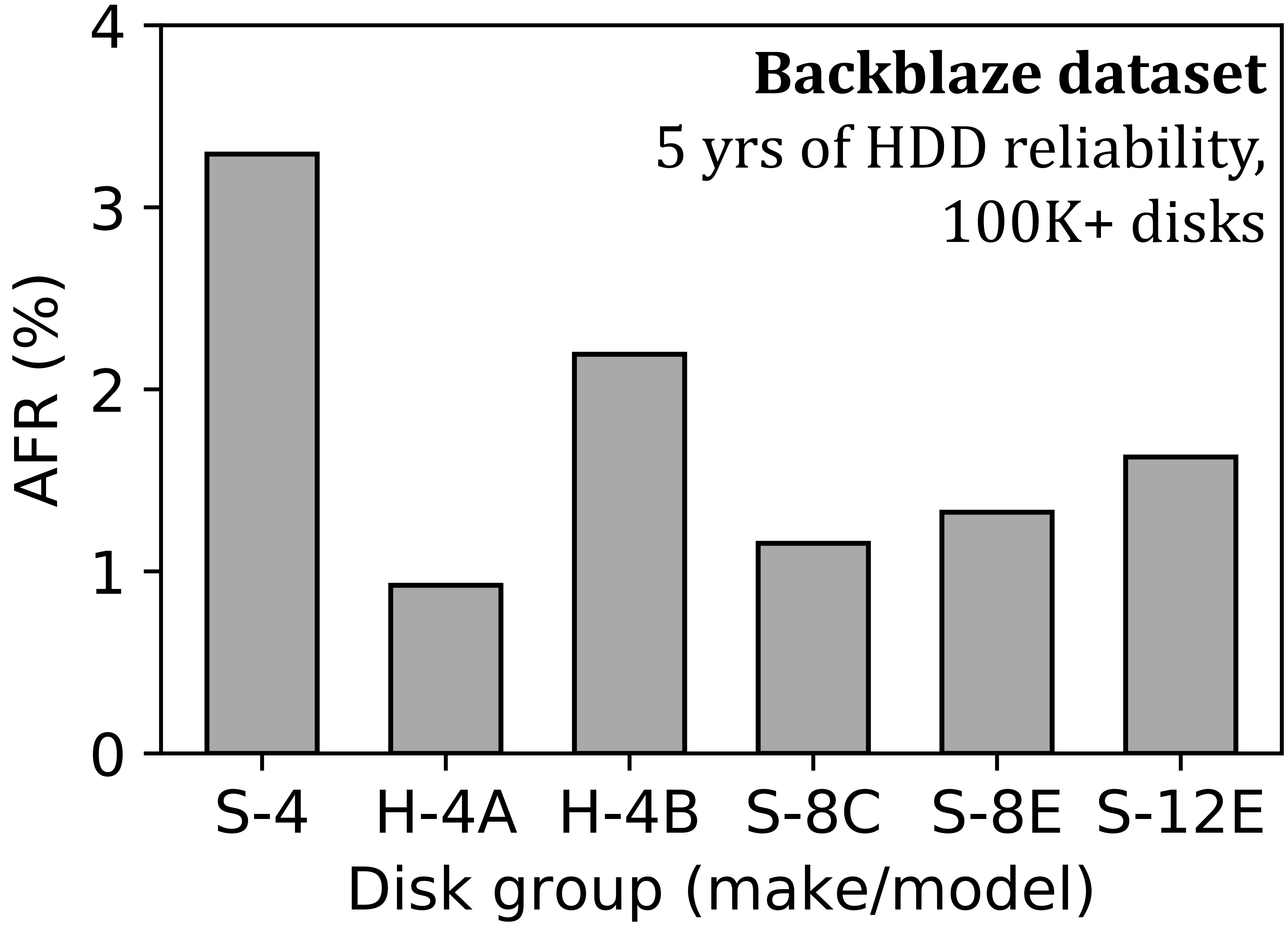
Overview of exploiting reliability heterogeneity

- Data redundancy typically same across disk fleet
  - E.g., 3-replication: 3 copies of data on independent devices
- Disks from same storage tier vary a lot in failure rates
  - E.g., HDDs from different makes/models fail differently
- **Explicitly consider reliability heterogeneity in deciding redundancy**
- **HeART**: **He**terogeneity **A**ware **R**edundancy **T**uner
  - Tailors redundancy to disk failure rate heterogeneity
  - A safe, accurate and online framework
  - Reduces storage overhead, and thus cost
- **HeART could save 11-33% disk space** on a production dataset

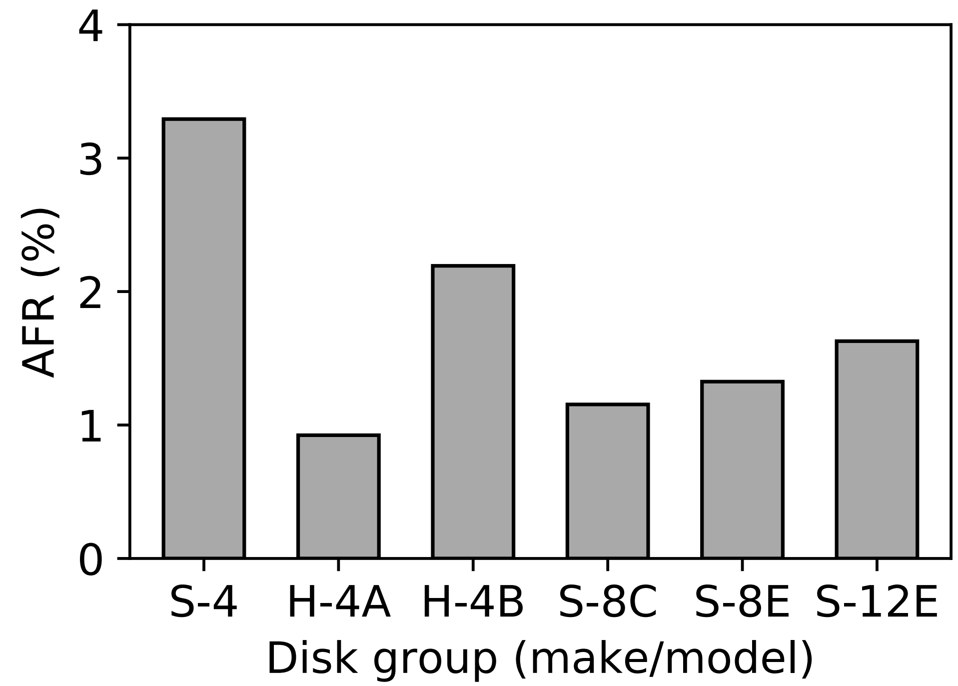
Cluster storage system reliability

- Failures common in today's cluster storage systems
  - Disk failures measured as annualized failure rates (AFR)
  - AFR ⟶ expected % of disk failures in a year
- Popular fault tolerance mechanism ⟶ redundancy
  - Full data replication
  - Erasure coding
- Redundancy configurations ignore disk AFR differences

Reliability heterogeneity



Reliability heterogeneity

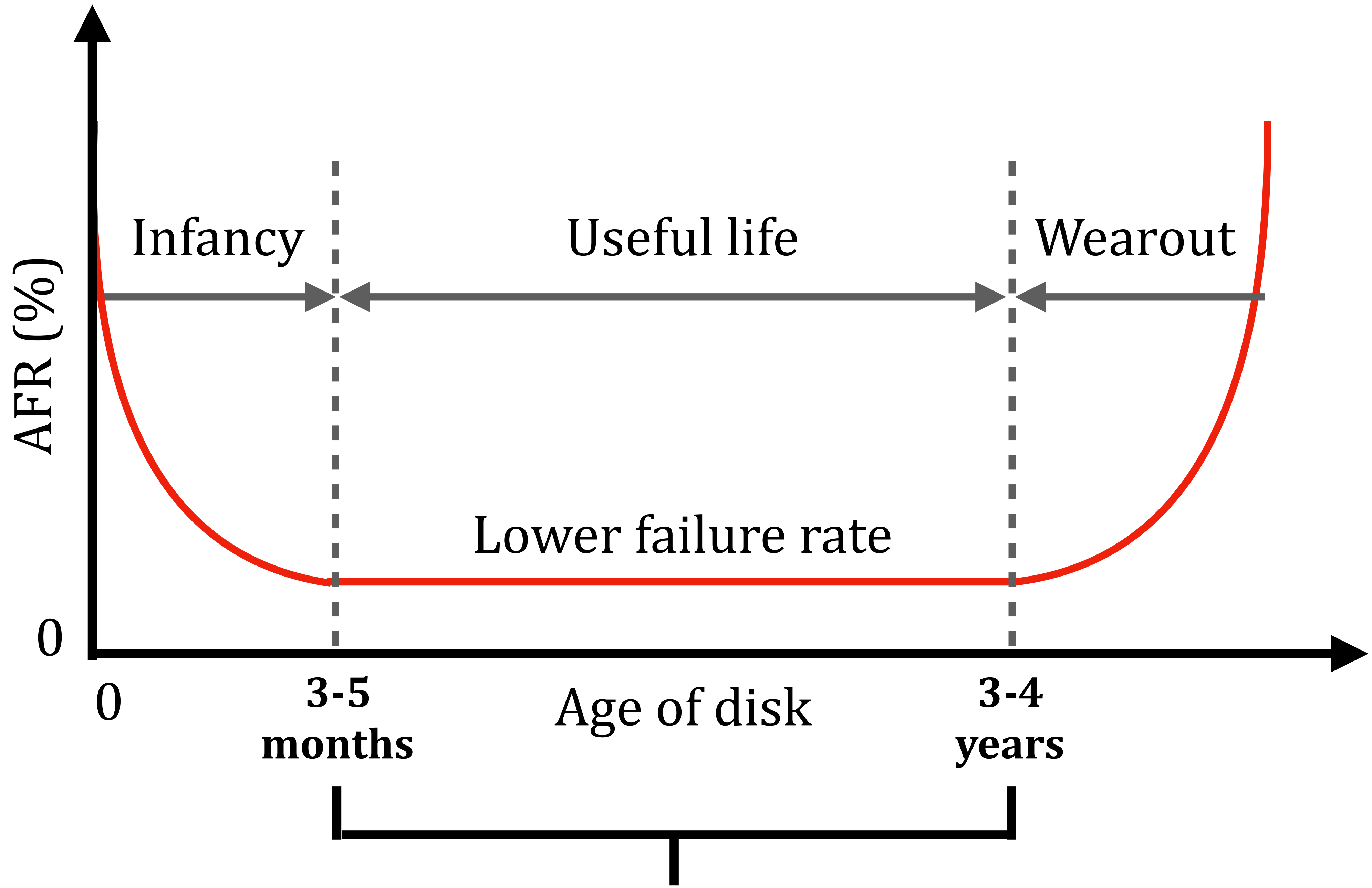


- HDD failure rates vary a lot in the field
  - Also shown by Schroeder et al. for SSDs in FAST 2016
- No single redundancy scheme is good enough for all disks
  - Conservative redundancy ⟶ overprotection for strong disk types
  - Lower redundancy ⟶ subset of disks risk data loss

Exploiting reliability heterogeneity

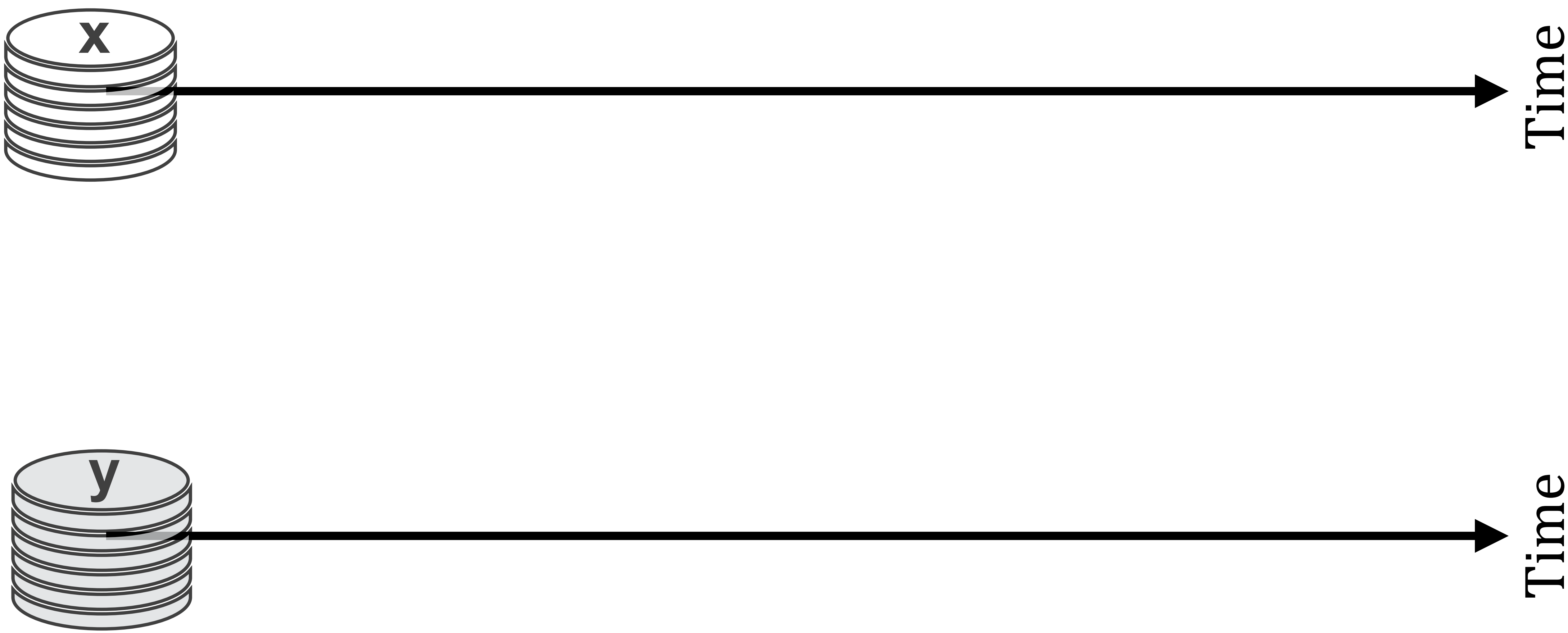
- Redundancy decisions informed by AFR differences
- **Challenges**
  1. Has to be **monitored in the field**
  2. Disk failure rate **varies over its lifetime**
- Redundancy tailoring mechanism needs to be:
  - **Safe**: prevent under-redundancy from causing data loss
  - **Accurate**: identify different reliability phases correctly
  - **Online**: benefits only realizable during disk's low failure rate

The bathtub curve (each disk group)

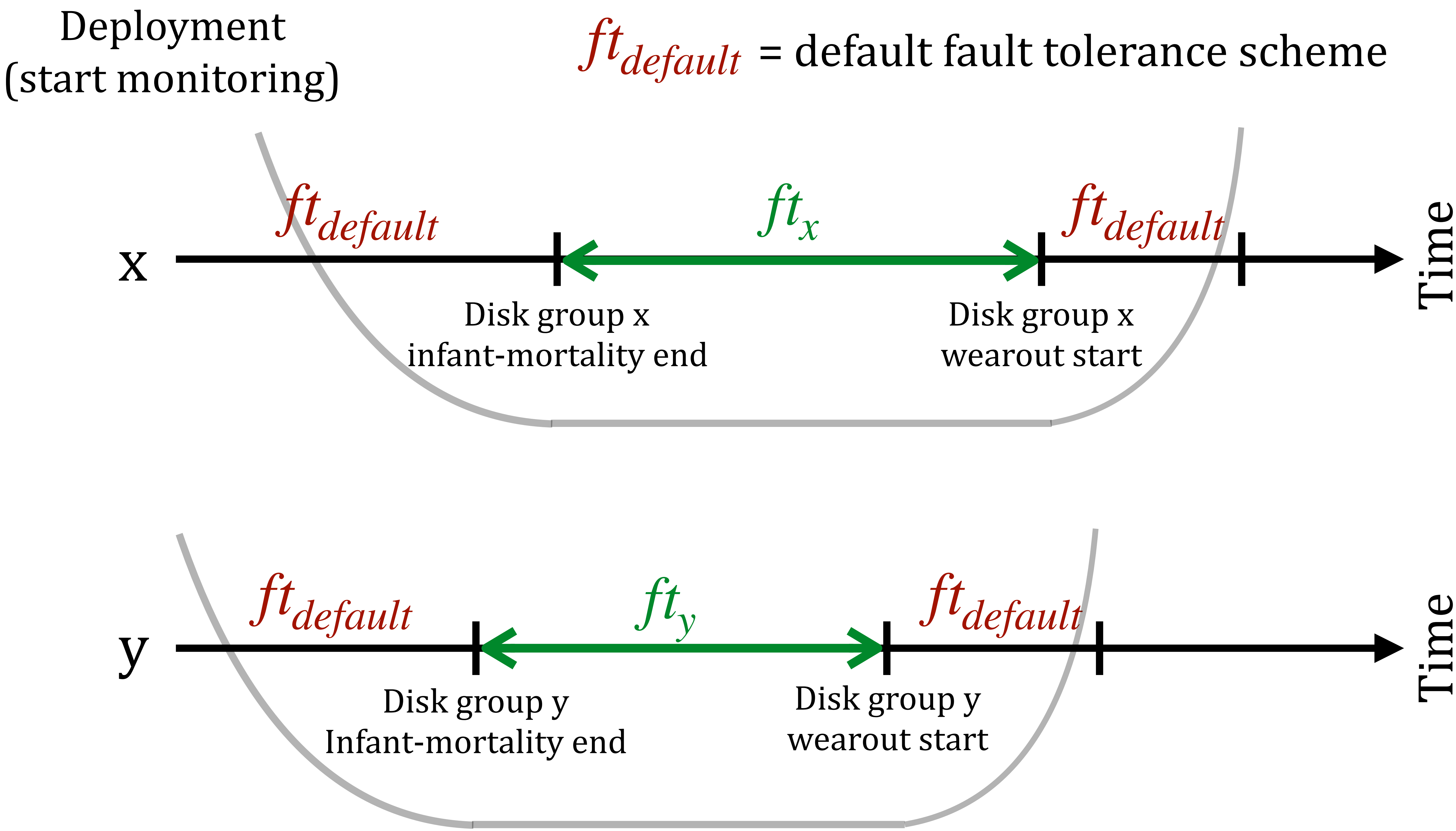


lower AFR → **lower redundancy** → **lower storage cost**

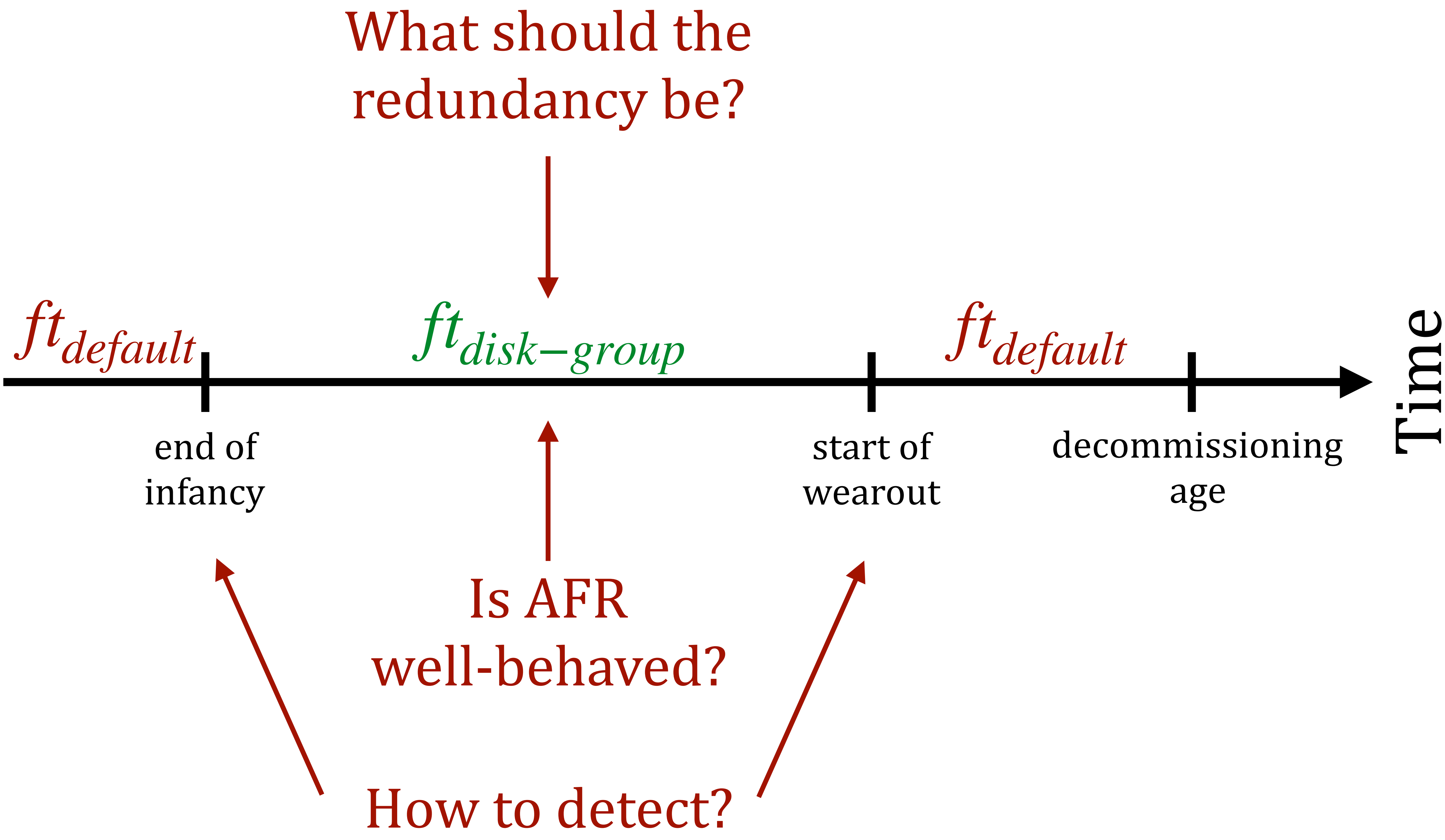
Two disk groups over time



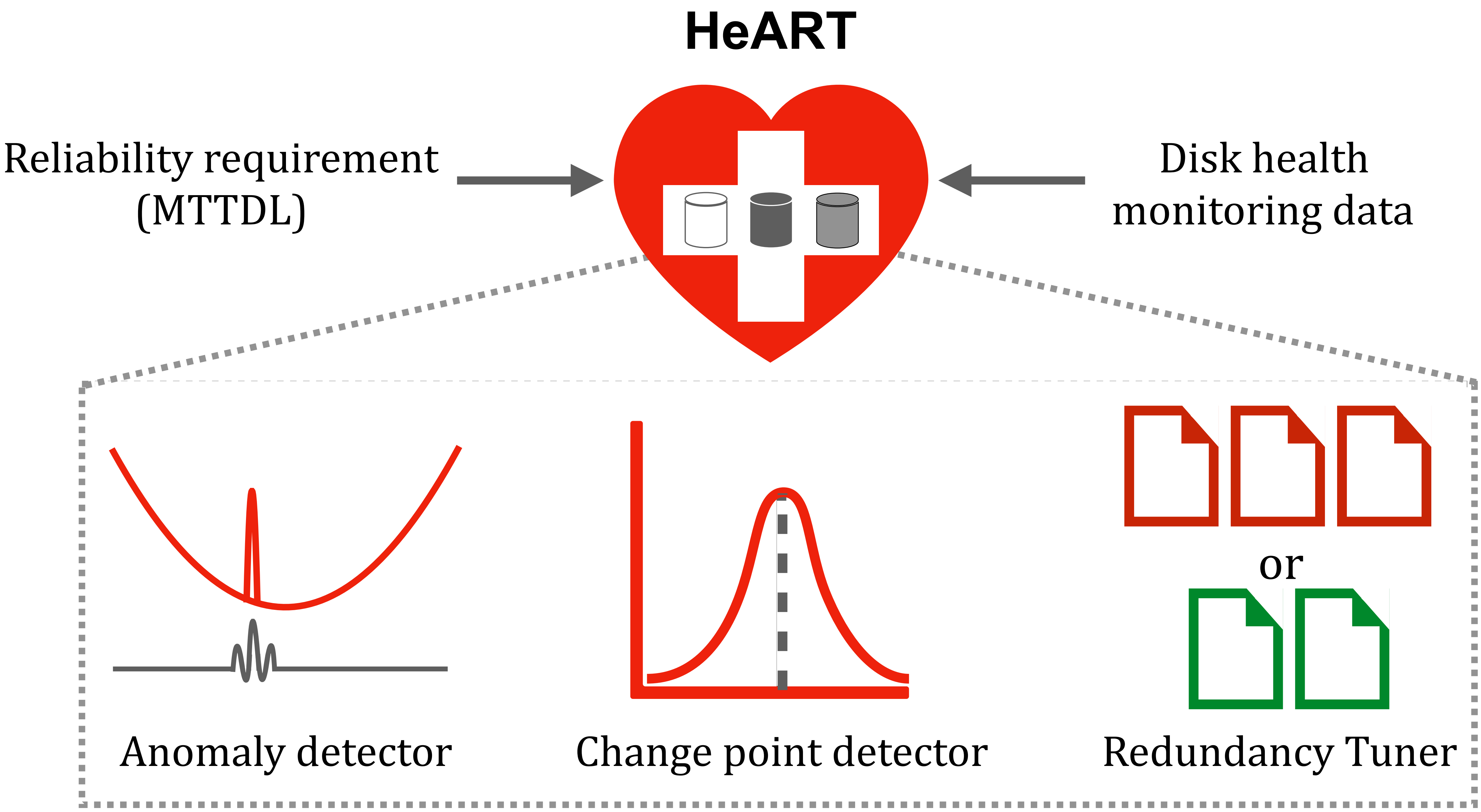
Two disk groups over time



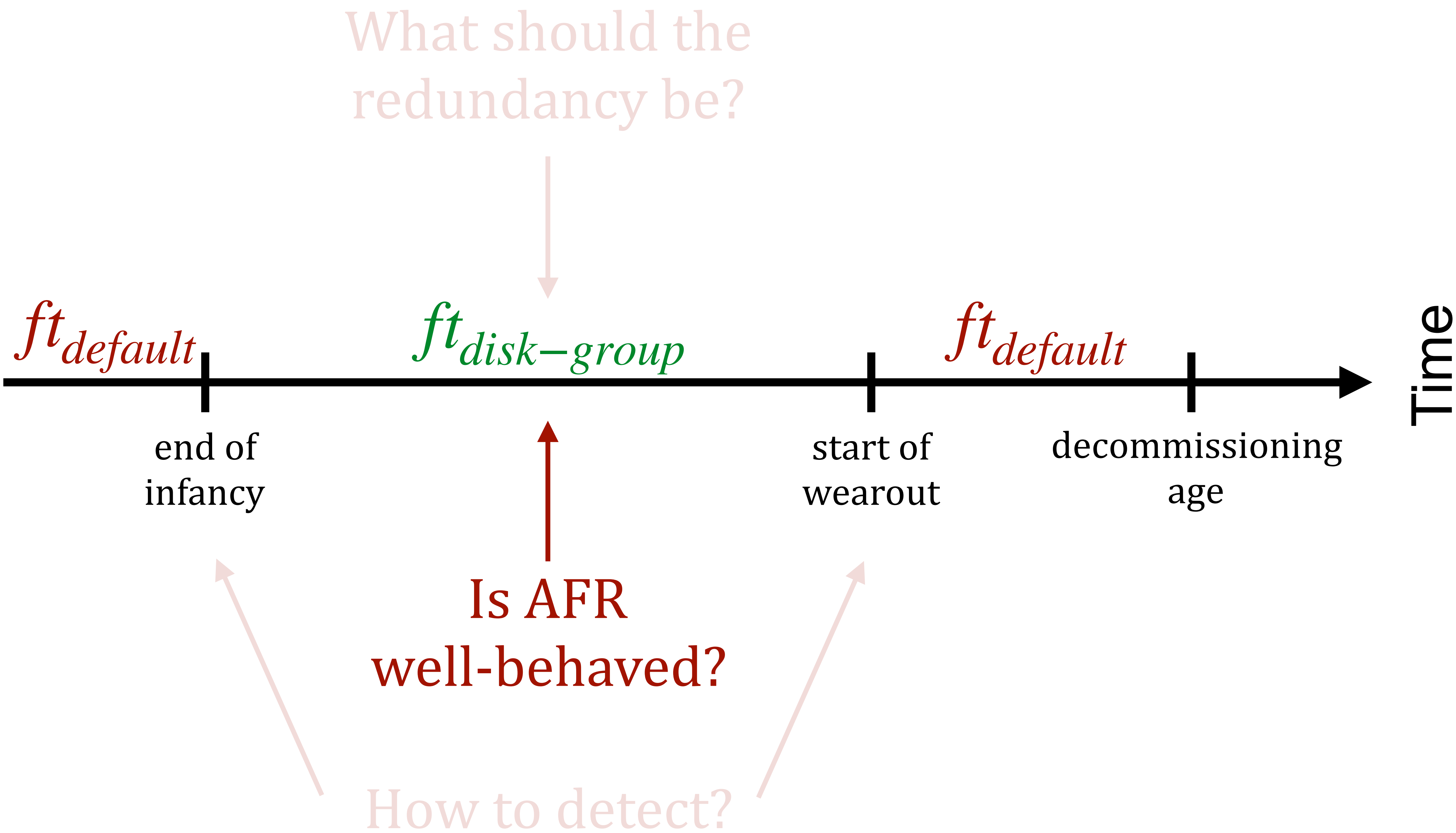
Disk-group reliability timeline



Heteretogeneity-Aware Redundancy Tuner

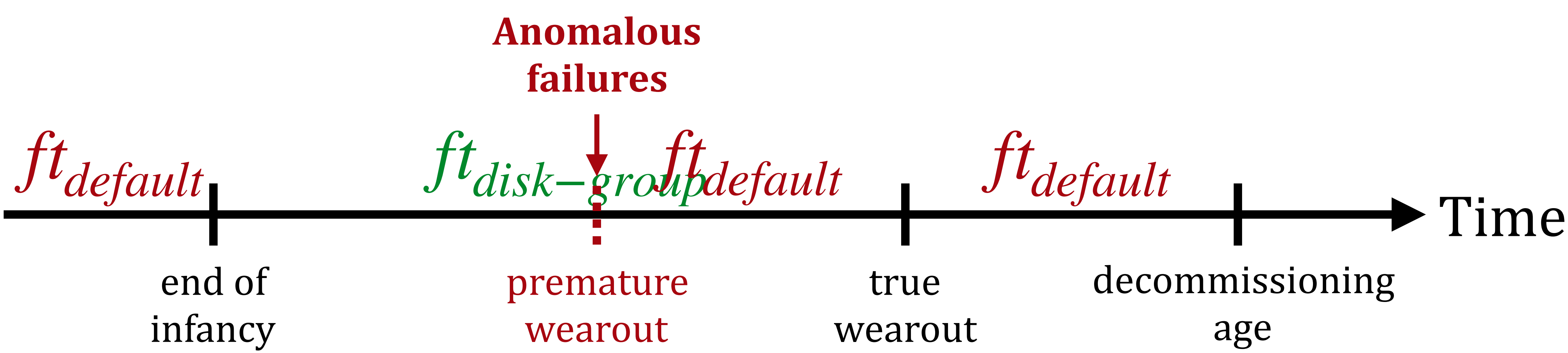


Disk-group reliability timeline

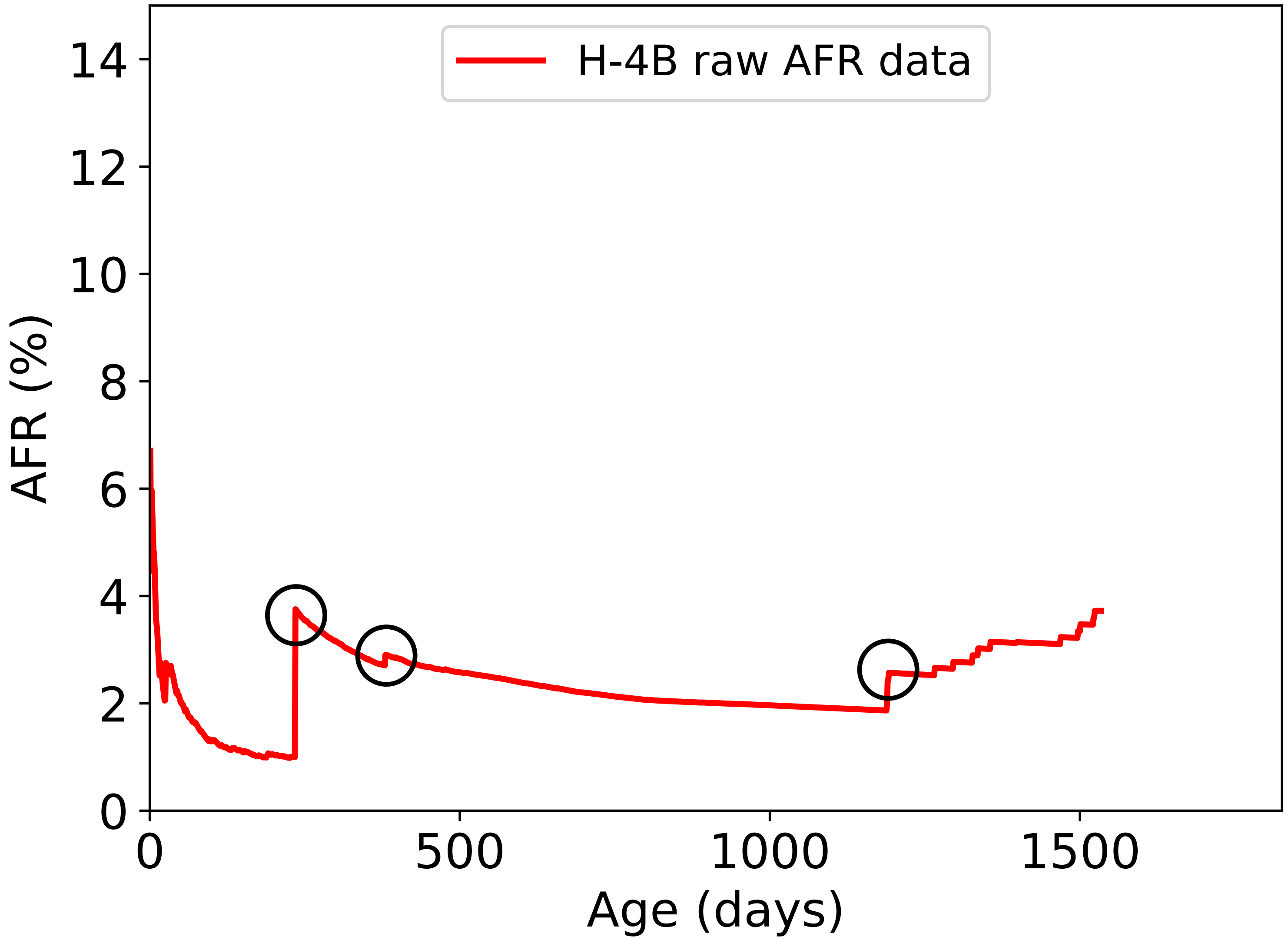


AFR in useful life: stability & anomalies

- Useful life AFR is typically stable, within reasonable bounds
- External factors can cause simultaneous bulk failures
  - Rack power failure, accidents, human error, etc.
- "Anomalies" appear like (premature) wearout
  - Benefits proportional to length of useful life
  - Bulk failures may not reflect true HDD failure rate



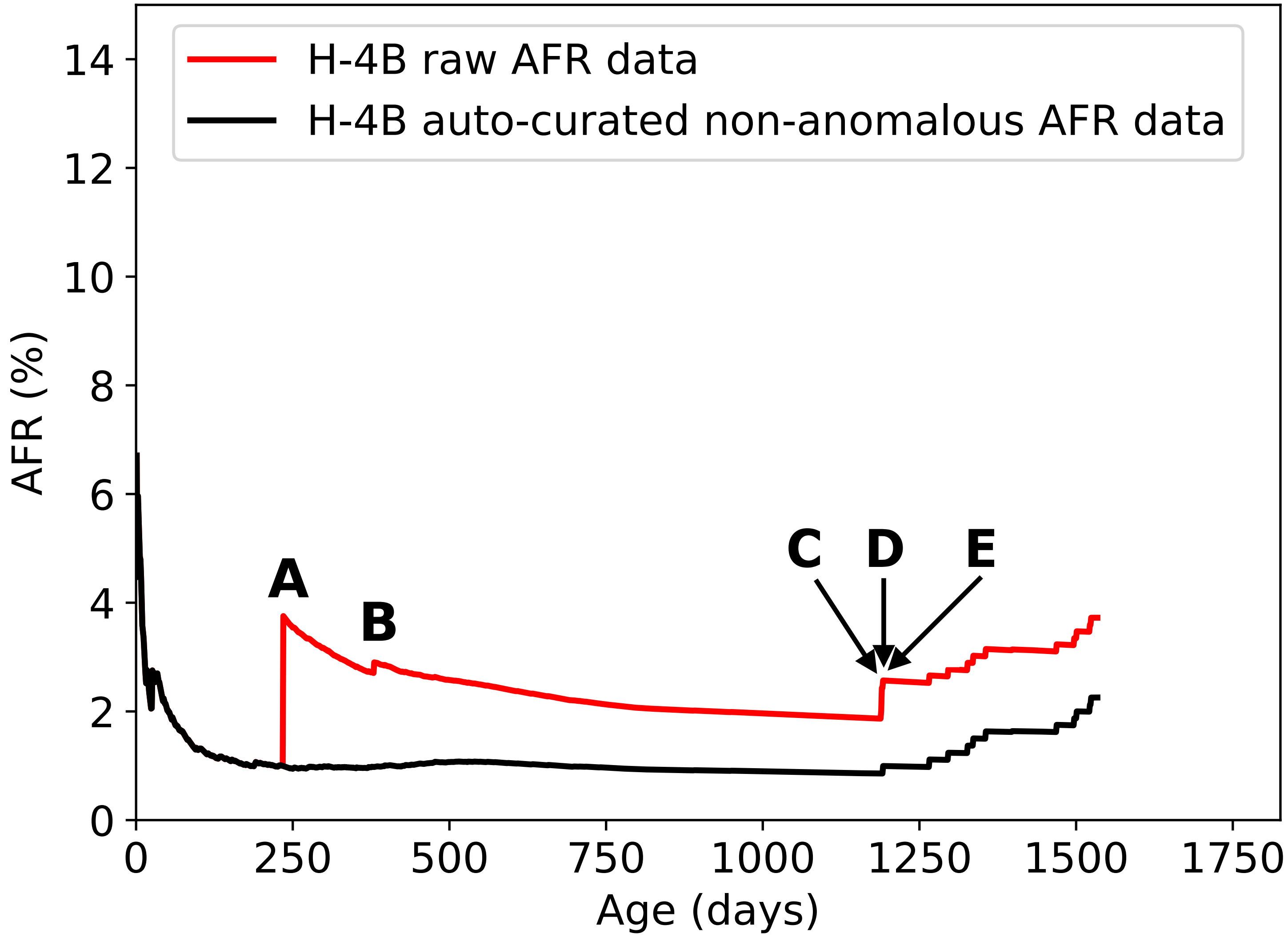
AFR in useful life: stability & anomalies



Spikes due to **simultaneous** bulk failures

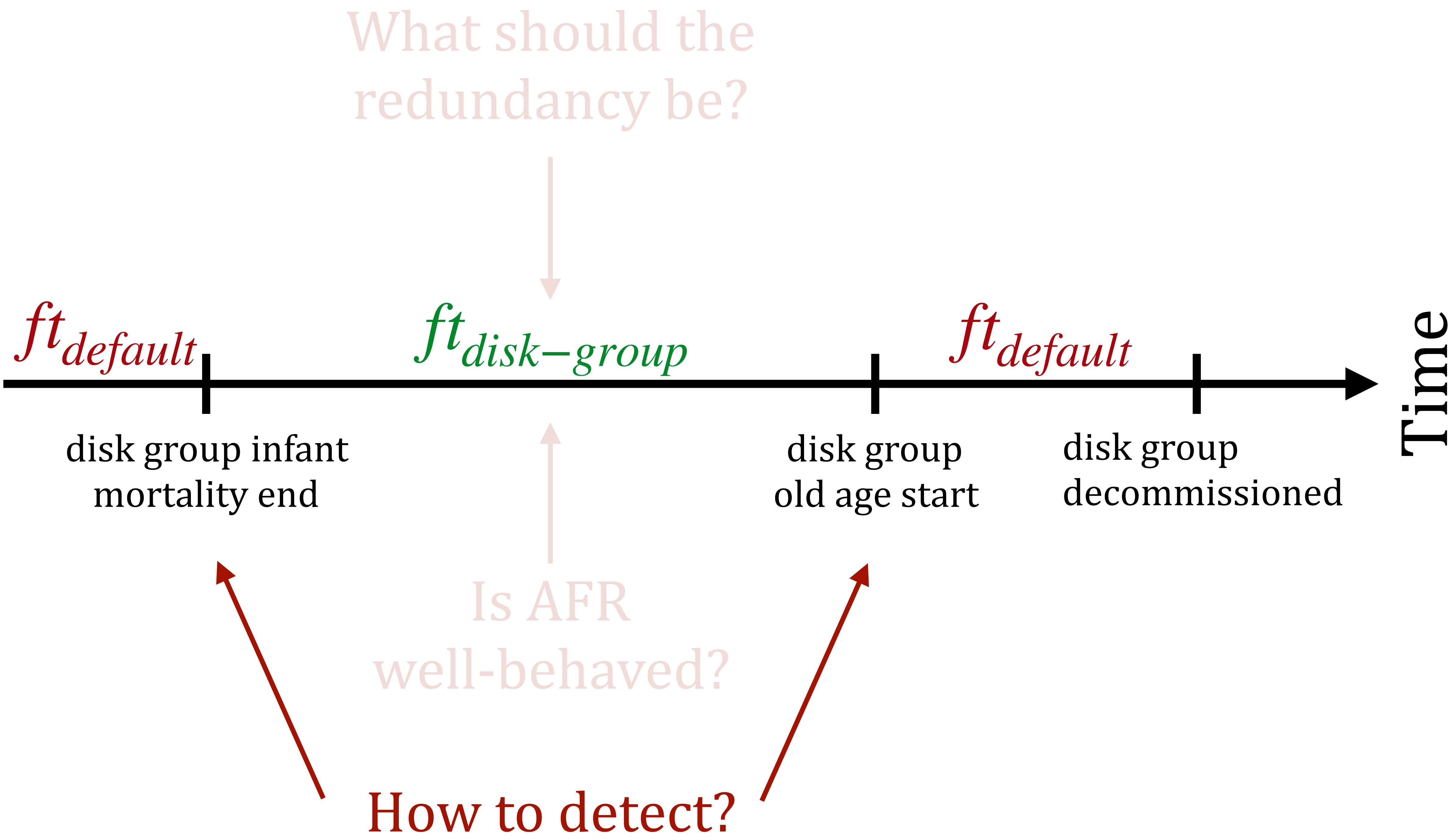
Need for anomaly detector

AFR in useful life: filtering anomalies

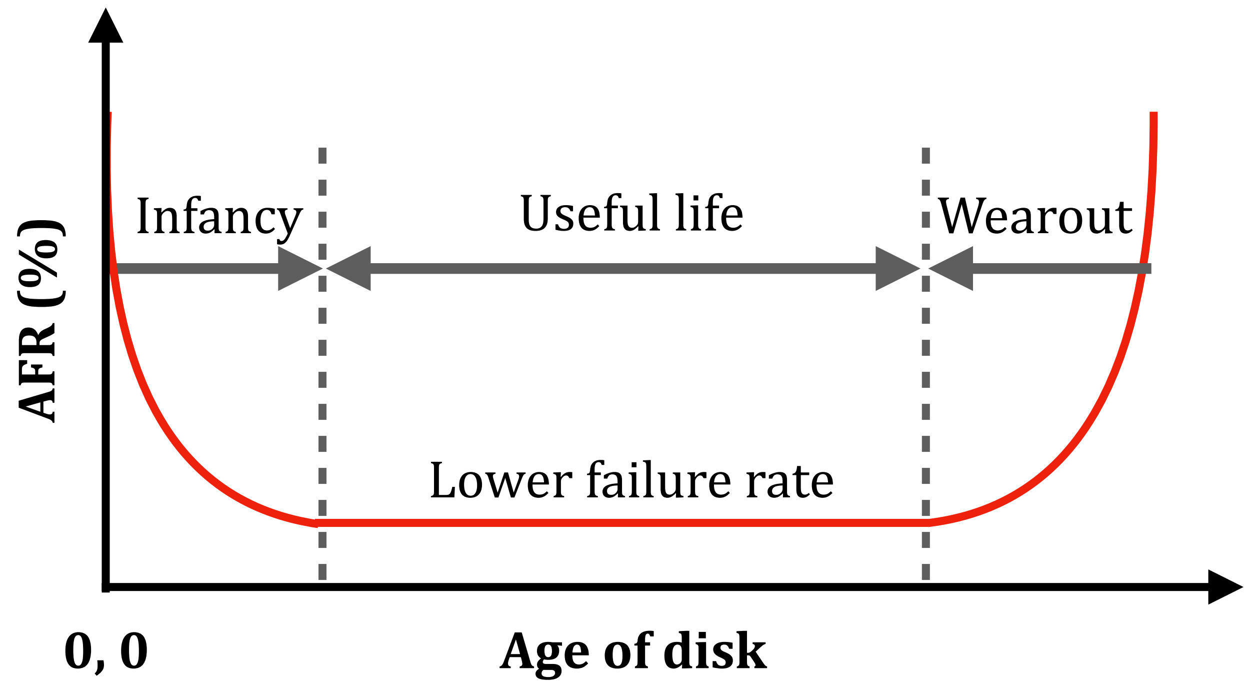


A and B: **300+ disks failed simultaneously**

Disk-group reliability timeline



Change point detection



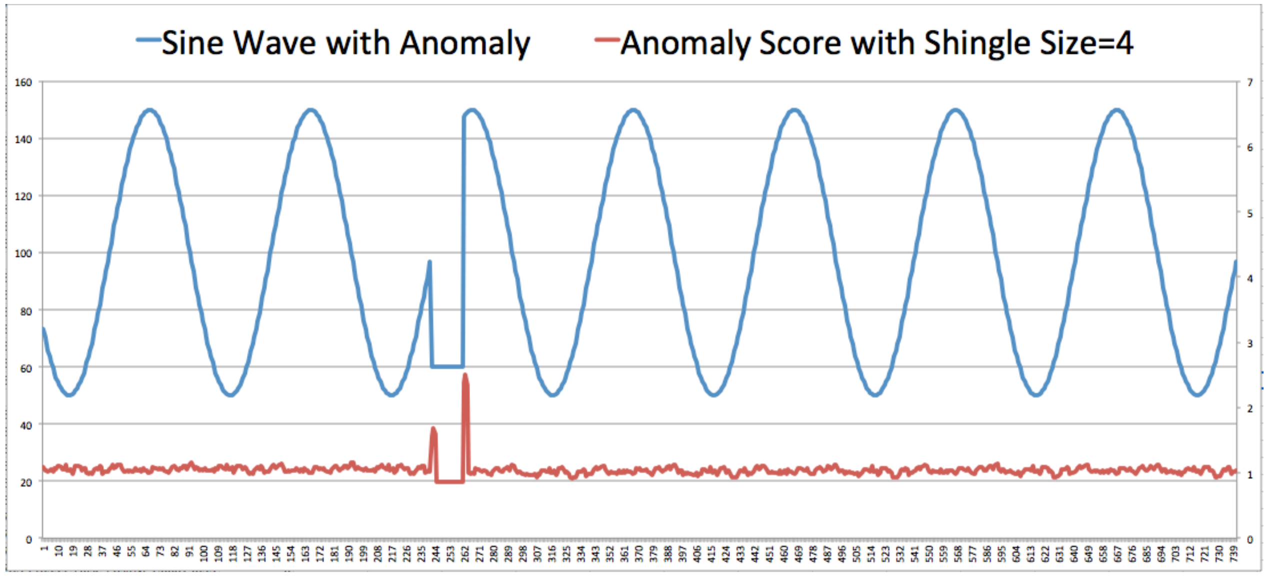
- Reliability target can be missed if:
  - Hasty declaration of end of infancy
  - Delayed declaration of onset of wearout
- Tradeoff between extracting benefits and safety
- Use online change point detectors to identify change points

Leveraging existing algorithms

- **Online anomaly detection (RRCF) [3]:**

Original signal

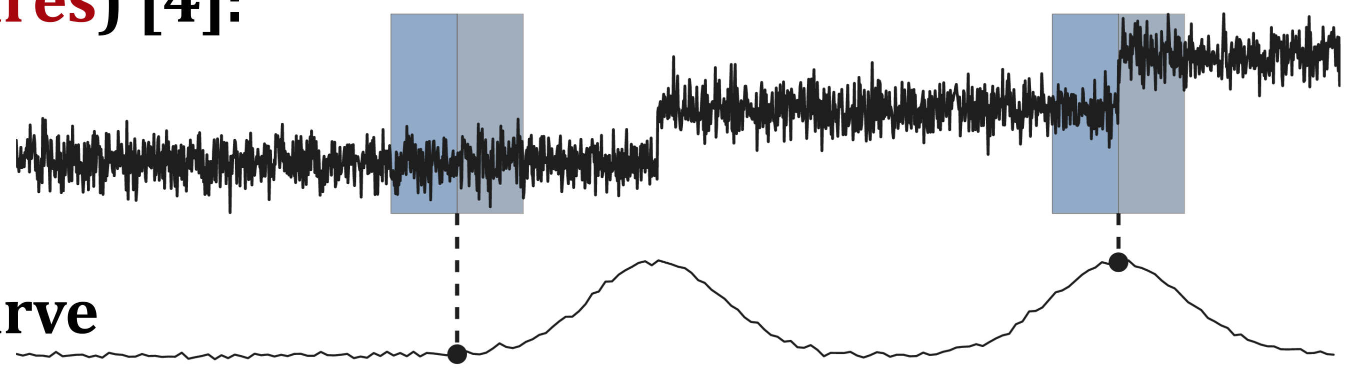
Discrepancy curve



- **Online change point detection (Ruptures) [4]:**

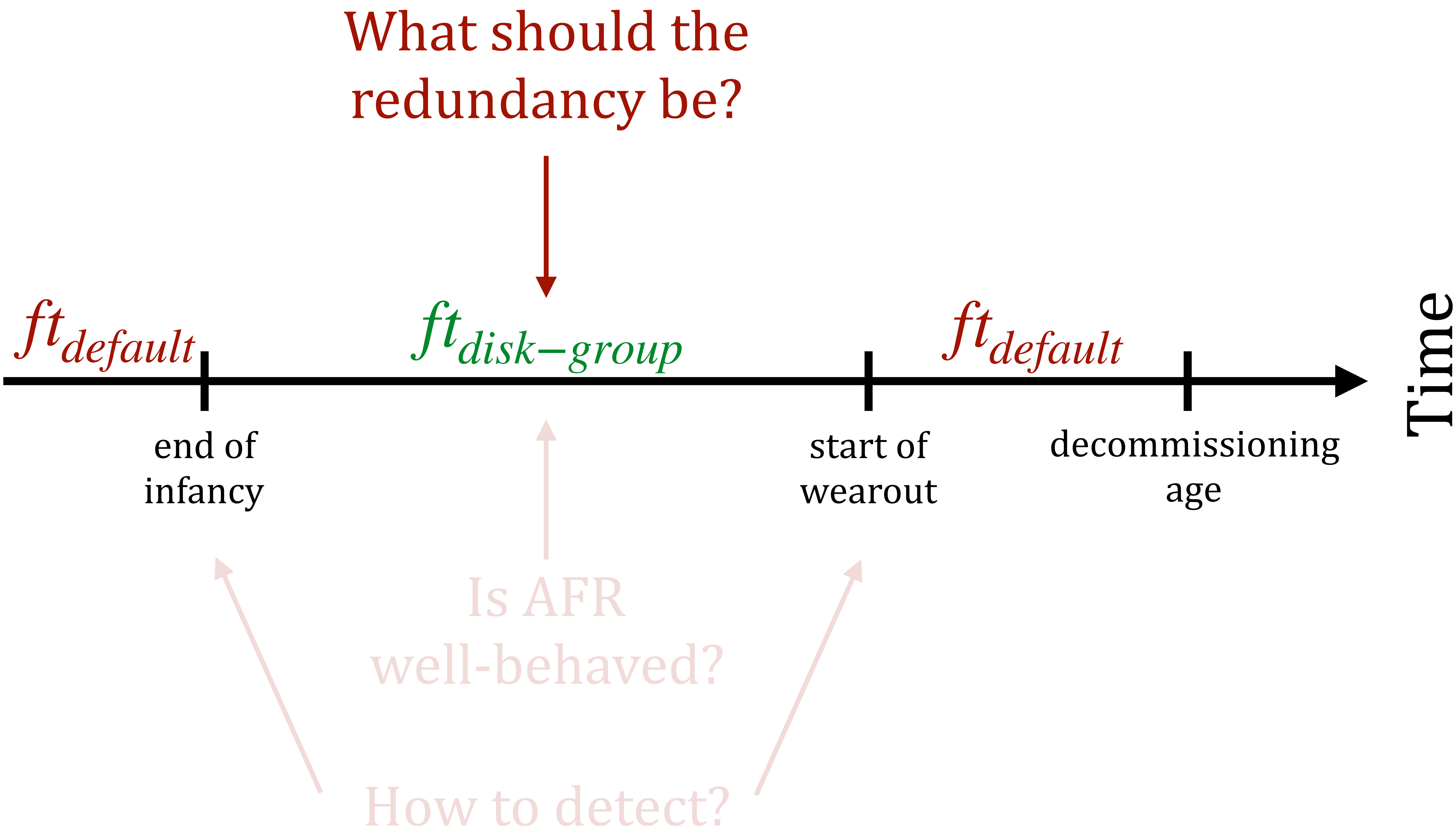
Original signal

Discrepancy curve



- Ruptures compares discrepancy in adjacent **sliding windows**
- Window length is one month by default
  - weekly or fortnightly AFRs are too jumpy
  - monthly AFRs balance reaction time with accuracy of AFR

Disk-group reliability timeline



The Backblaze dataset

- 100K+ HDDs belonging to Backblaze: a backup company
  - Daily reliability statistics from mid 2013 - mid 2018
  - Open sourced
  - 6 drive makes/models with significant number of disks to test:

Disk Grp	Num Drives	Num Failed	Age so far (yrs)
S-4	37015	9539	5
H-4A	8715	3939	4.77
H-4B	15048	1276	4.2
S-8C	9885	186	1.99
S-8E	14395	162	1.2
S-12E	21581	148	0.64

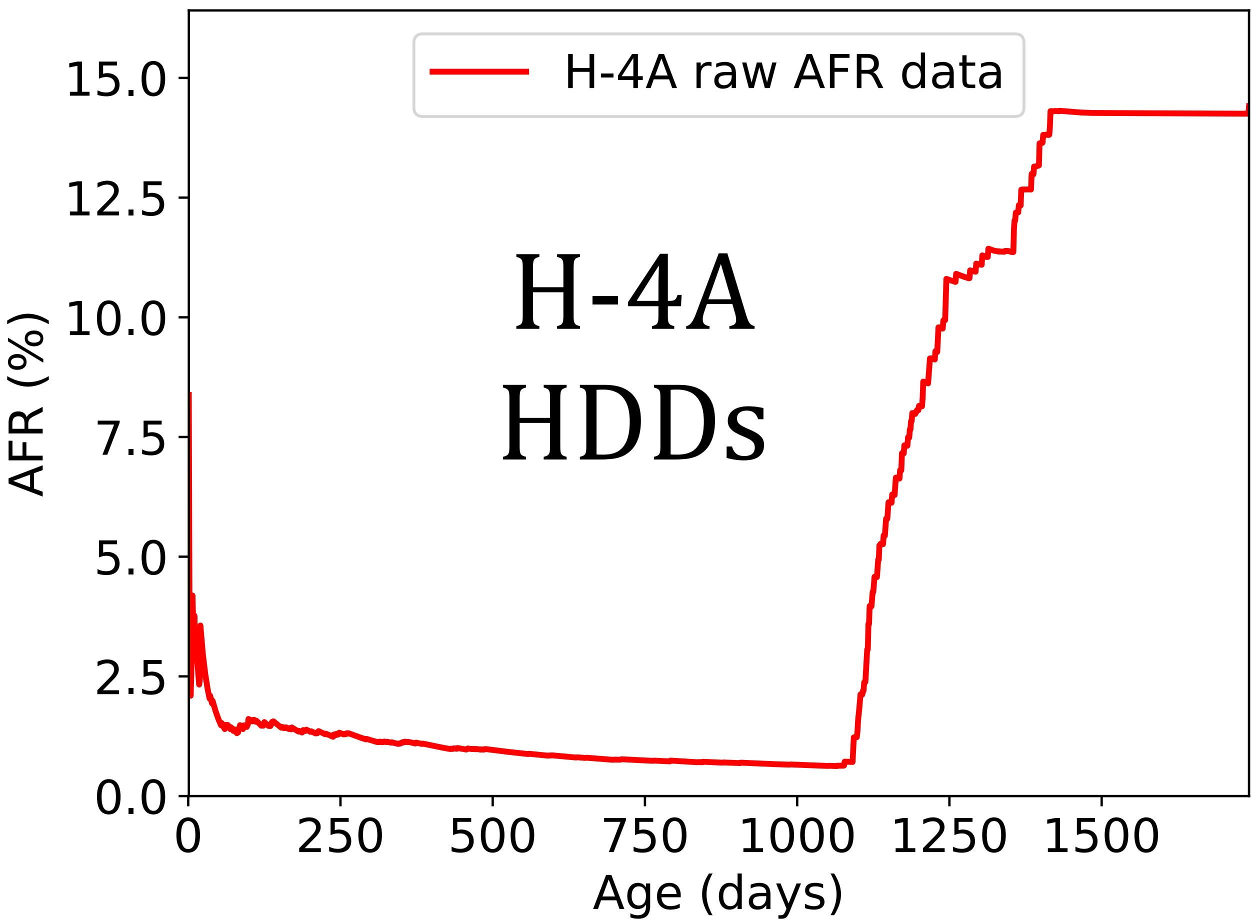
Methodology to evaluate cost reductions

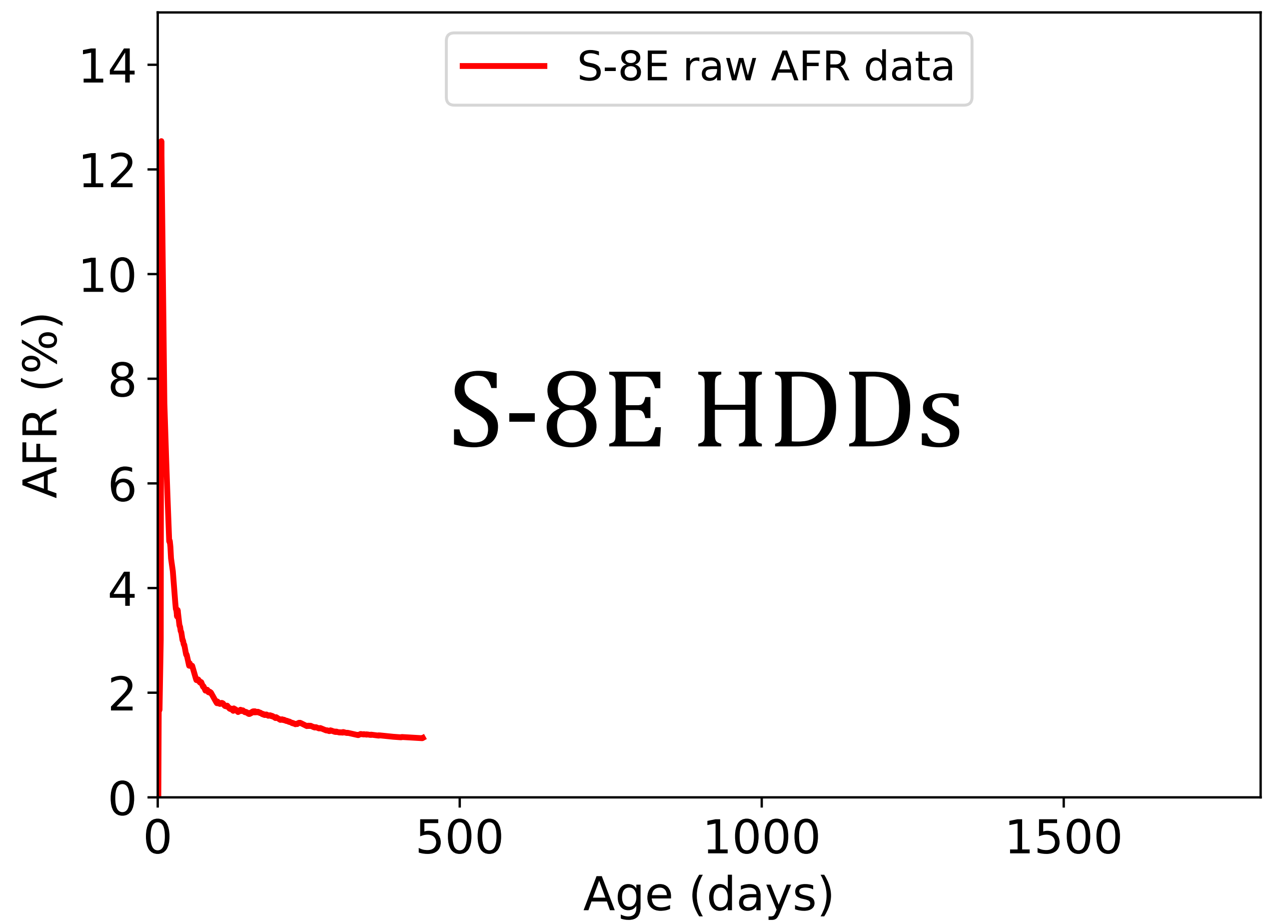
- Need target reliability (MTTDL)
  - Higher the MTTDL, the more resilient the data
- Reliability target decided by disk group with highest AFR
  - Currently $ft_{default}$ on all disks provided acceptable MTTDL
  - MTTDL is only higher for disks with lower AFR
- Find cheaper $ft_{disk-group}$ that meets MTTDL
  - Cheaper redundancy $\longrightarrow$ lower storage cost
- $ft_{disk-group}$ is decided with the following constraints:
  - Tolerate at least as many failures as $ft_{default}$
  - Have an upper bound on stripe width

Evaluation on Backblaze dataset

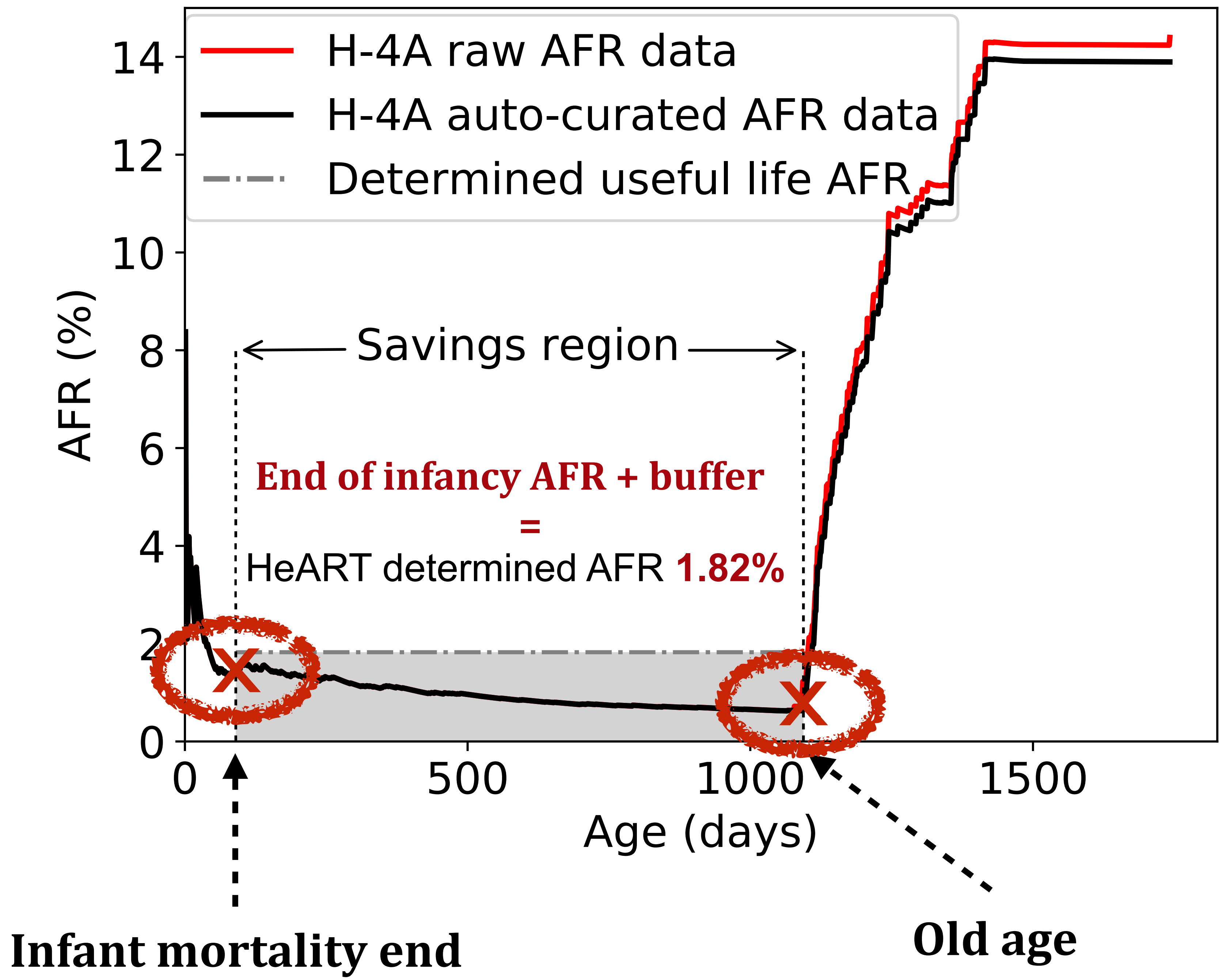
- S-4 disks have the highest AFR (4.01%) in Backblaze
  - Reliability target is MTTDL of $ft_{default}$ on S-4 HDDs
- Upper bound on stripe width = 2x $ft_{default}$
- $ft_{default}$ options evaluated:
  - 3-replication
  - 6-of-9 erasure code
  - 10-of-14 erasure code

Disk group AFR

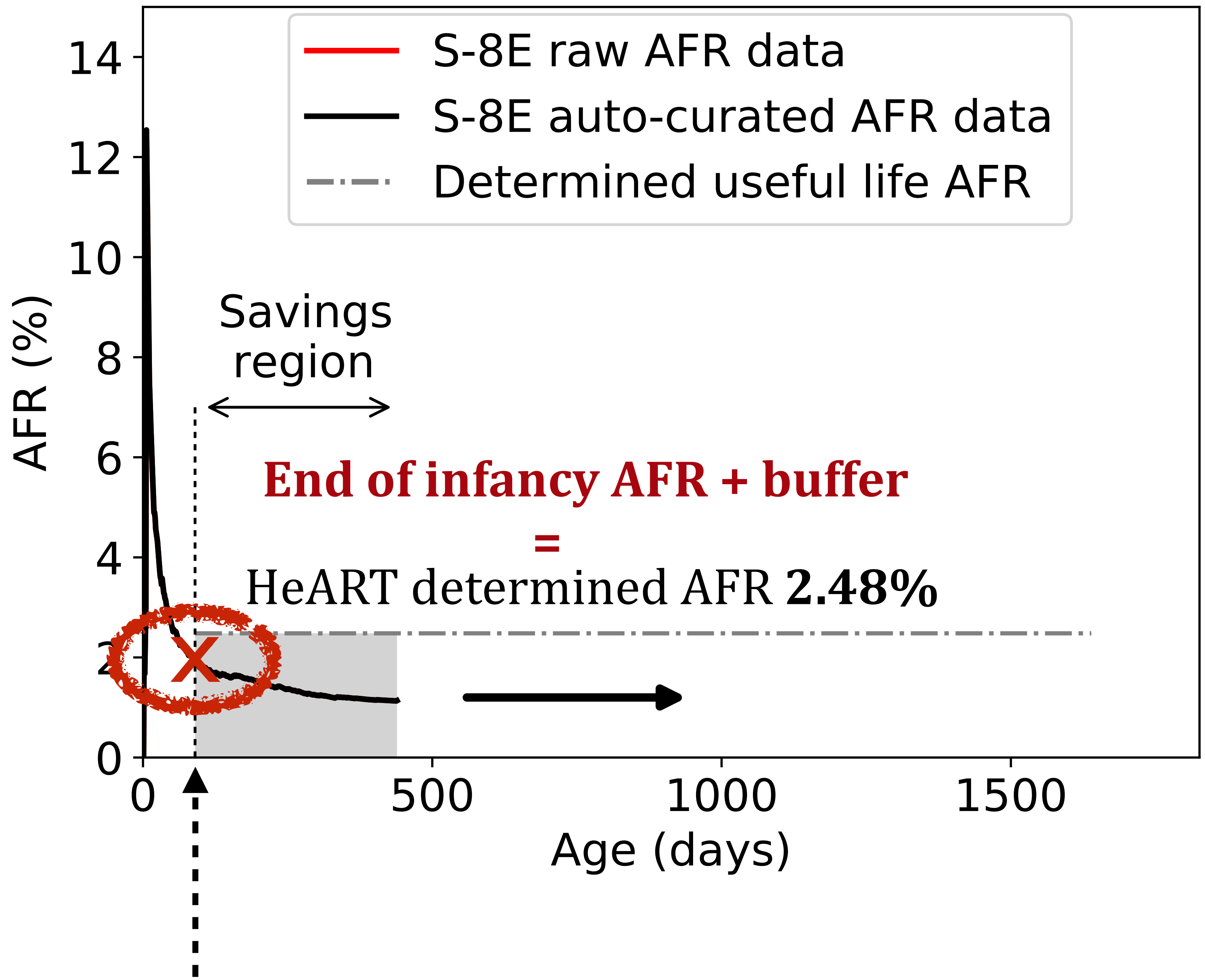




HeART in action: **H-4A HDDs**



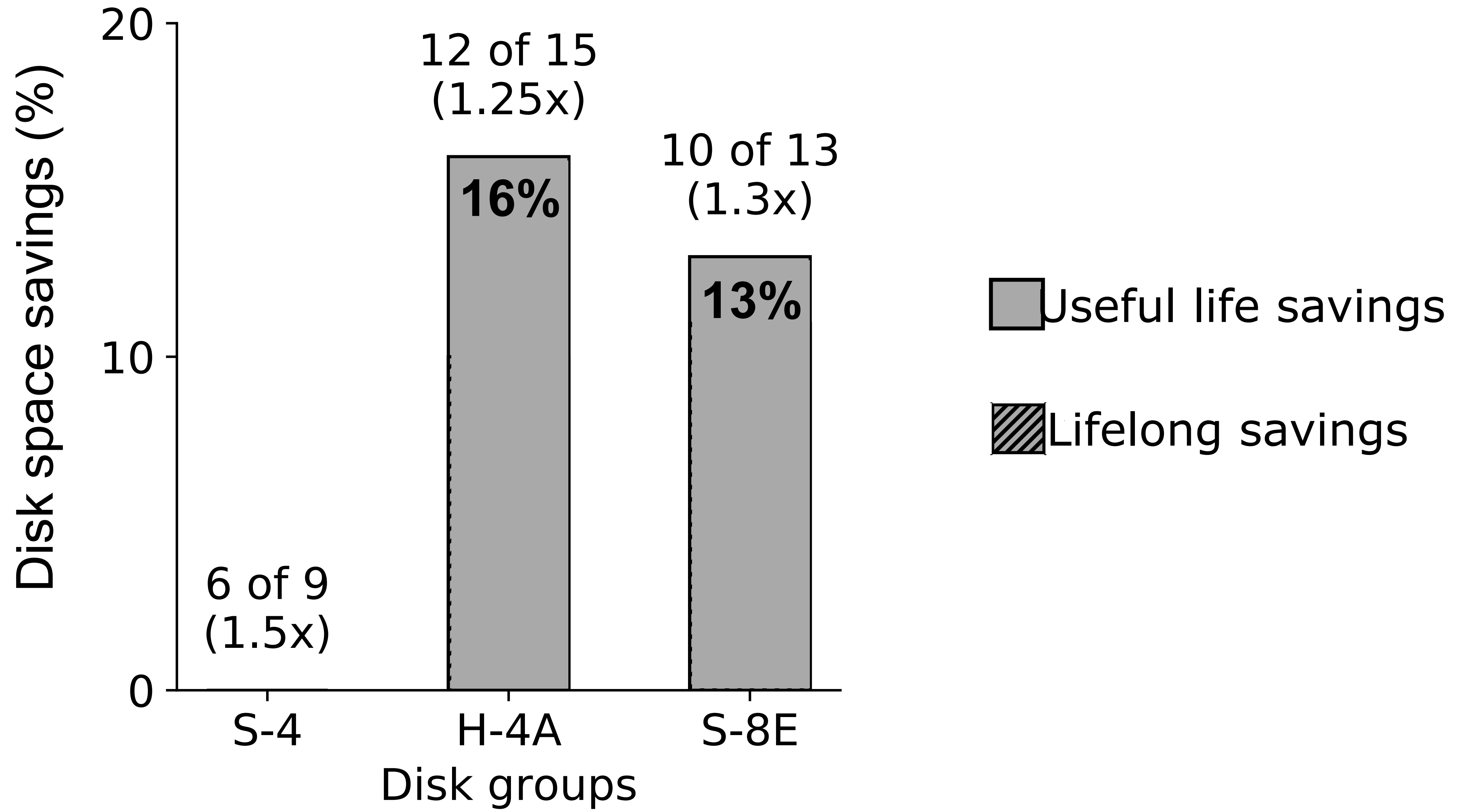
HeART in action: **S-8E HDDs**



Only infant mortality end detected

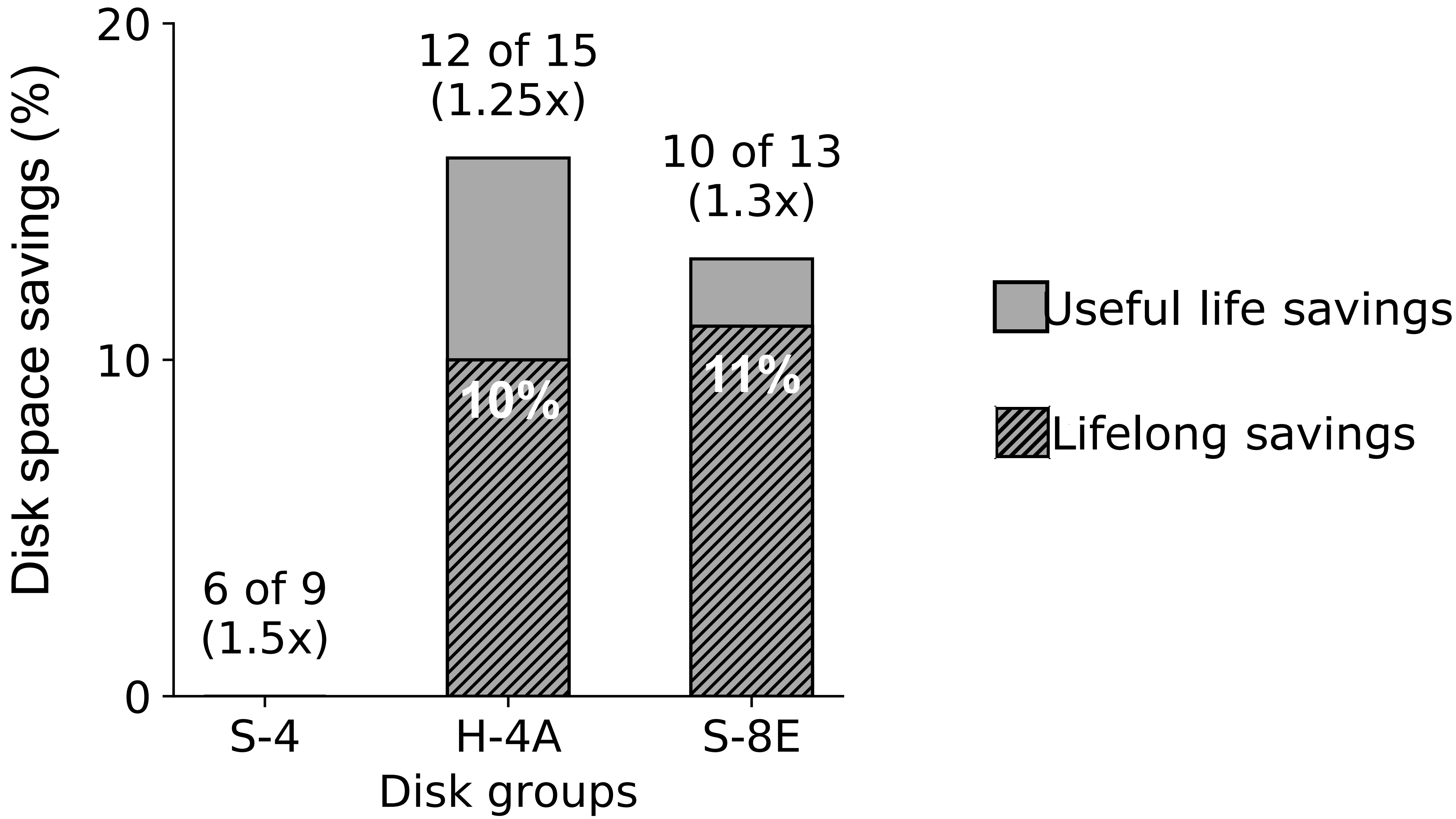
Storage savings: **6-of-9**

- Small storage overhead of only 1.5x



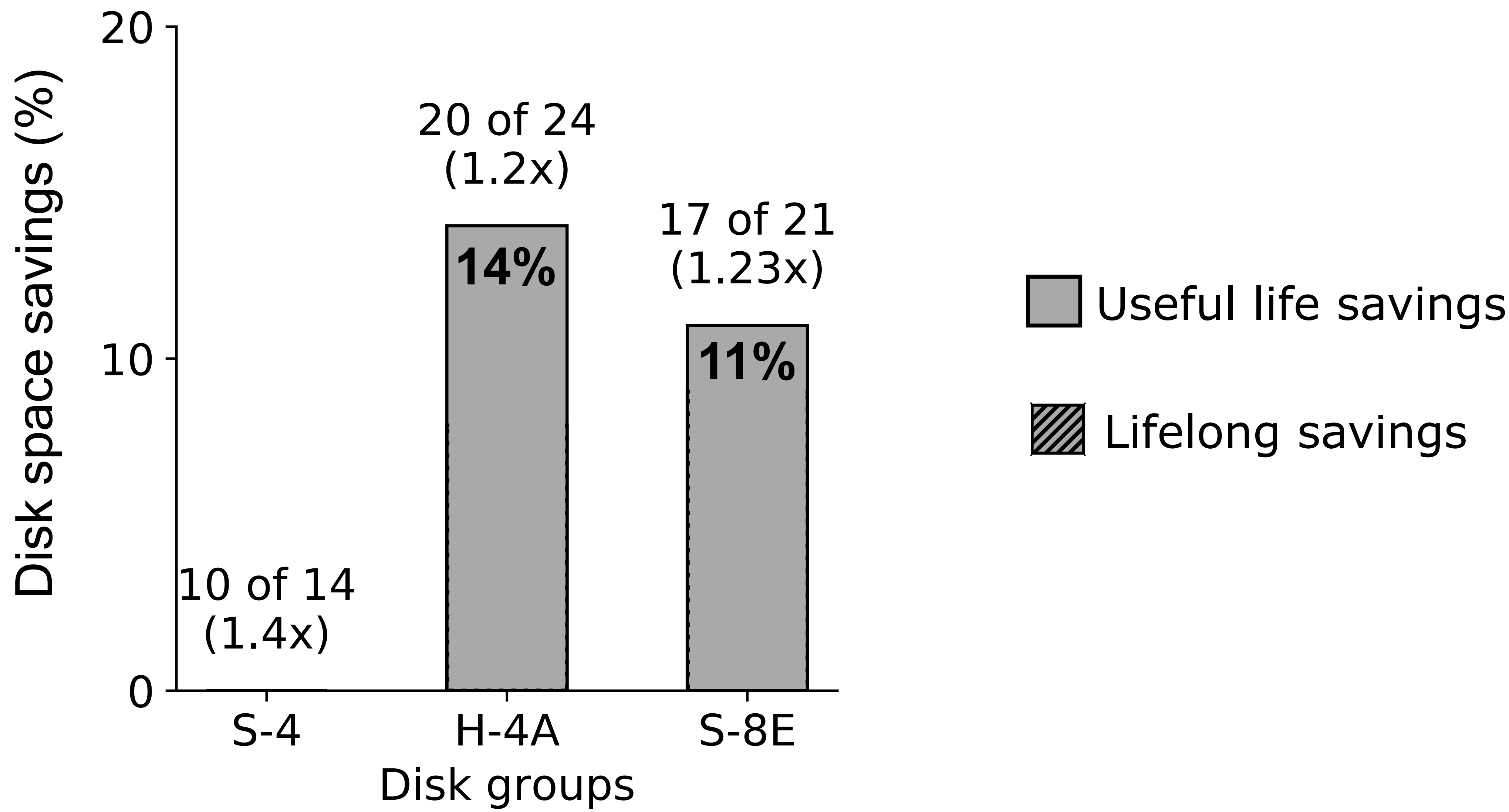
Storage savings: **6-of-9**

- Small storage overhead of only 1.5x



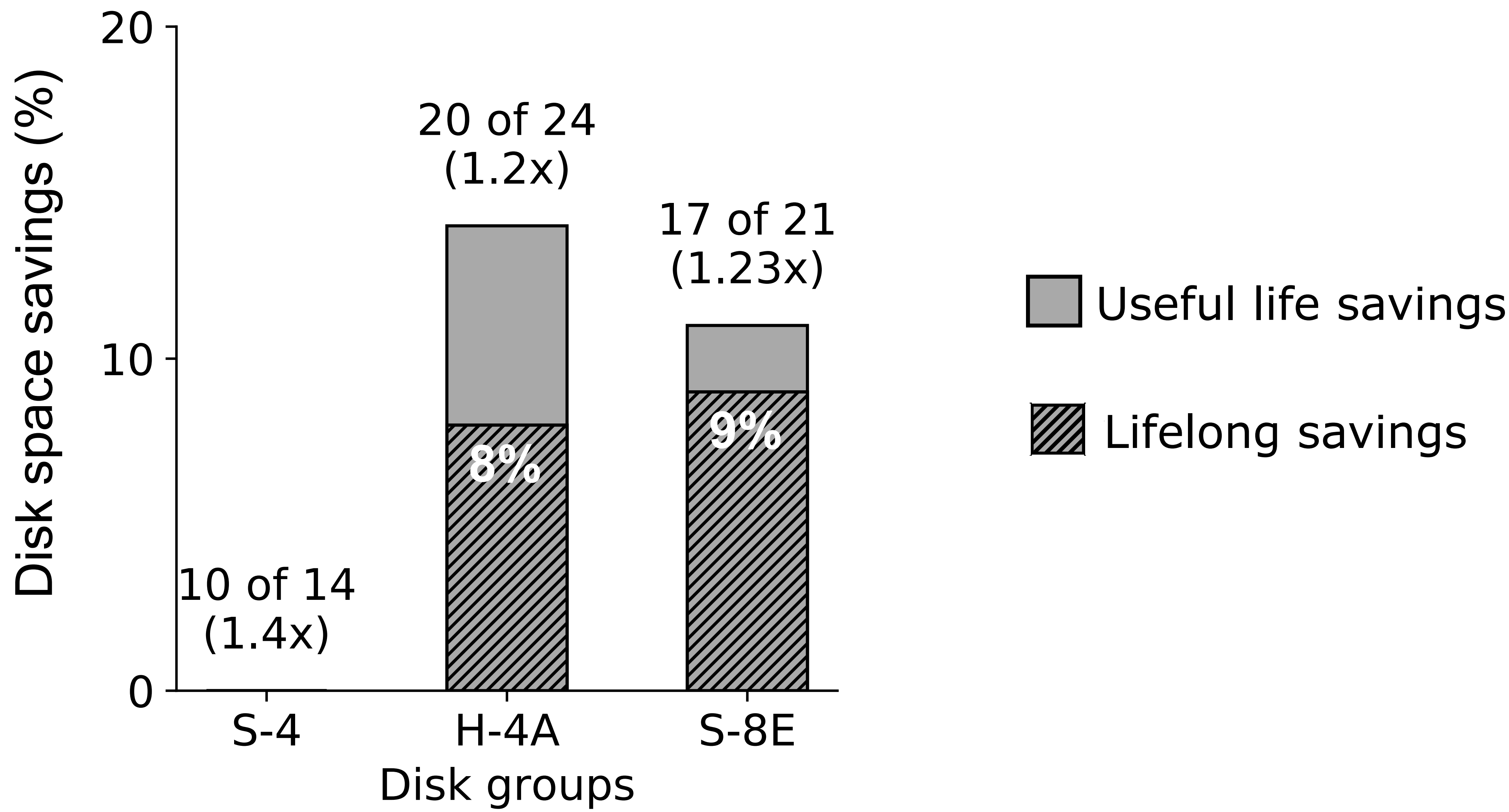
Storage savings: **10-of-14**

- Even smaller storage overhead of only 1.4x



Storage savings: **10-of-14**

- Even smaller storage overhead of only 1.4x

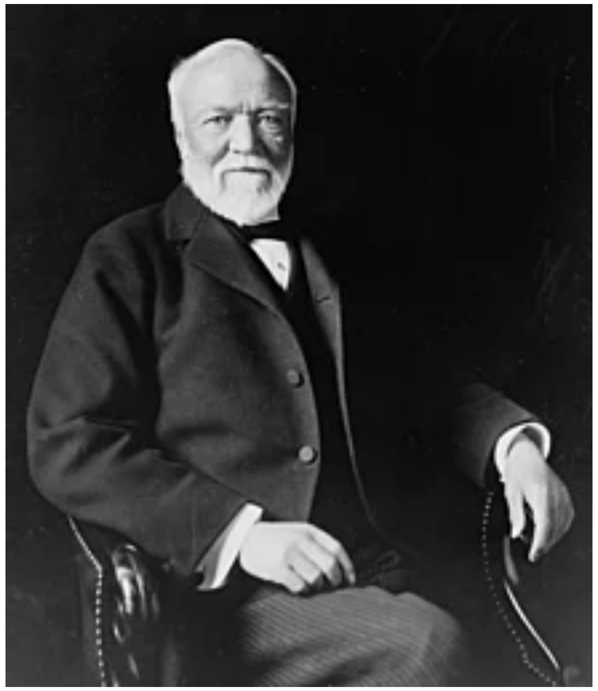


Conclusion

- Exploiting reliability heterogeneity reduces storage cost
- 11-33% space savings observed on production dataset
- **HeART**: an online heterogeneity-aware redundancy tuner
  - actively engages with disk bathtub curves
  - built-in online anomaly and change point detector
  - suggests cheap redundancy schemes that meet reliability

Thank you!

Questions?



“My heart is in the work”

“My work is in the HeART”

References

1. Xia, Mingyuan, et al. "**A Tale of Two Erasure Codes in HDFS.**" *FAST*. 2015.
2. Sathiamoorthy, Maheswaran, et al. "**Xoring elephants: Novel erasure codes for big data.**" *VLDB.* 2013.
3. Guha, Sudipto, et al. "**Robust random cut forest based anomaly detection on streams.**" *ICML*. 2016.
4. Truong, Charles, Laurent Oudre, and Nicolas Vayatis. "**ruptures: change point detection in Python.**" *arXiv:1801.00826.* 2018.
5. Rashmi, K. V., et al. **"A hitchhiker's guide to fast and efficient data reconstruction in erasure-coded data centers."** *ACM SIGCOMM Computer Communication Review.* 2015.