Measuring Reliability in SSD Storage Systems

Uncorrectable Bit Errors
Temperature Model

Steven Hetzler, IBM Fellow

Reliability in Storage Systems

- **We care about all forms of data loss**
- **Device loss is one type**
- **Sector loss (uncorrectable bit errors or UBER) is another**
  - Also called non-recoverable read errors (NRRE)
- **We will concentrate on sector loss events here**
- **Measure the raw bit error rate from flash devices**
  - Need to see the ber behavior to understand NRRE

Outline

- Bit Error Rate tests
- Flash bit error rate vs. PE cycles and data age surfaces
- Flash empirical model
- Temperature tests
- Error Uniformity
- Things that go bump in the drive

Measuring NAND Bit Error Rates

Measure ber(PE cycles,data age,reads,temperature)

- Measure raw bit error rate in an SSD
  - Cycle quickly to various PE counts, age in real time
  - Can compute NRRE from ECC specs and ber
- Use SSDs with a host-managed interface (HMI)
  - Physical block access – host manages wear leveling
  - Host control of read, write, erase, raw read (no ECC correction)
  - All accesses over std. interface (e.g. SATA)
  - SSD purchased retail, firmware updated for HMI
  - Temperature controlled oven with temperature capture
  - All writes are full erase-block stripes, in page-sequential order
  - Test state info stored in a library (and limited on drive)

Simplified ber Test Flow

How flash was tested

- Select set of erase block stripes

  Erase stripe

    Wait e-w dwell

  Write stripe

    PE++

    If (PE mod 10 != 0) { wait w-e dwell; goto Erase stripe }

    Wait w-r dwell

  Read stripe

  - If (PE < PElimit) goto Erase stripe
  - Wait age_read
  - Read stripe
  - If (age > age_limit) done;
  - goto Wait age_read

We can pause the test here to add idle gaps to measure read disturb

Analyzed Behavior

Things we measured

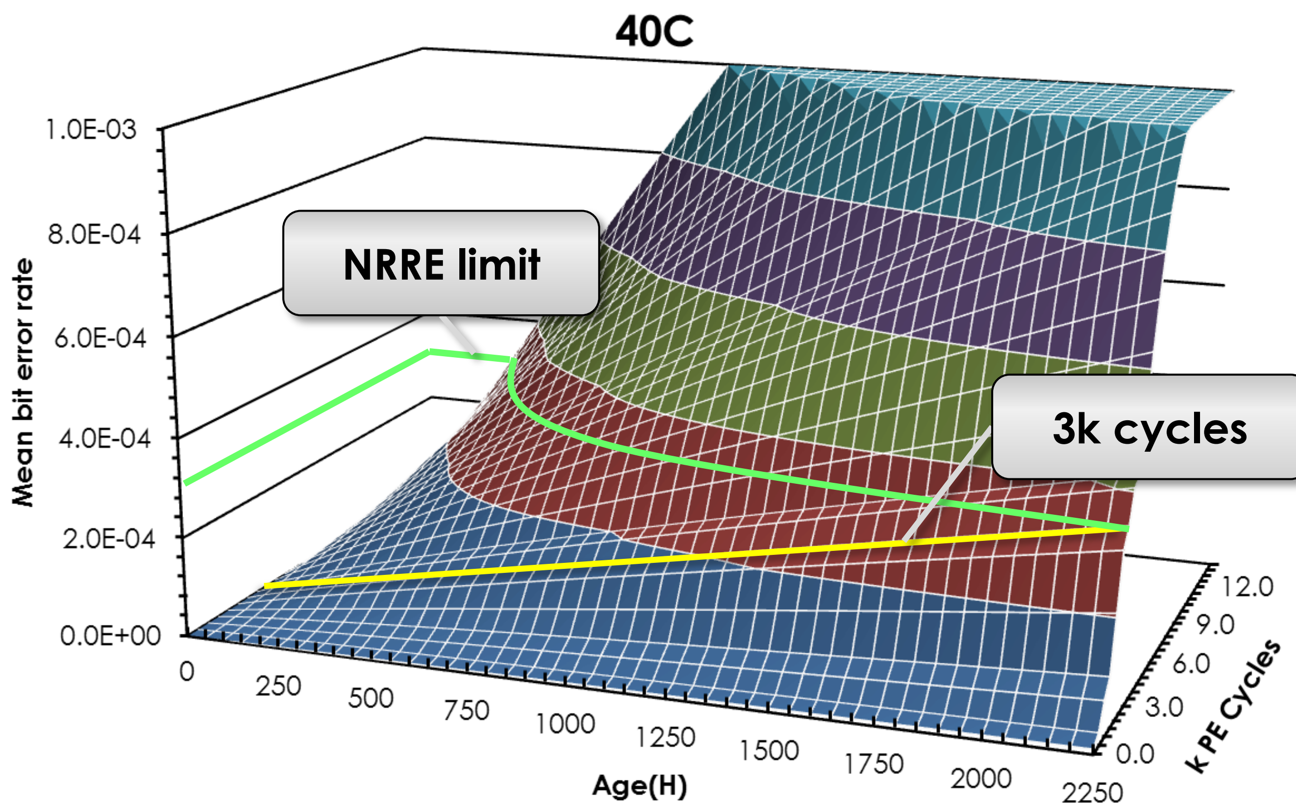
- PE Cycling
  - Look at error counts at short data ages while cycling
  - Max cycles out to 8x device spec
- Aging
  - Look at error counts while not cycling
  - Read at a constant rate – typically 10 reads/H
- Read Disturb
  - Limited to looking at effects of reads within the same stripe
  - Can separate from aging by changing the read rate
    - We set rate to 0 for 24 H at 2 different times
- Temperature
  - Looked at range of temperatures from 30C to 100C
    - Mostly <=70C to avoid device damage
  - Cycled, read and aged ALL at target temperature

Exemplary Surface 3xnm 40C

- ber limit corresponding to NRRE limit computed by inverting binomial
  - Sector bits = 4096+195
  - BCH 15 code – correct 15 error bits
  - 1e15 = 1- binom.dist(15,4291,ber,TRUE) (Excel) ➔ ber = 3.4e-4

3xnm C-SSD
3k cycle spec
1 year retain
70C
10^{15} bit NRRE int.

Note: device does not meet spec!
Will cross NRRE limit at 2,275 H =3 months



NRRE Specs and Data Loss

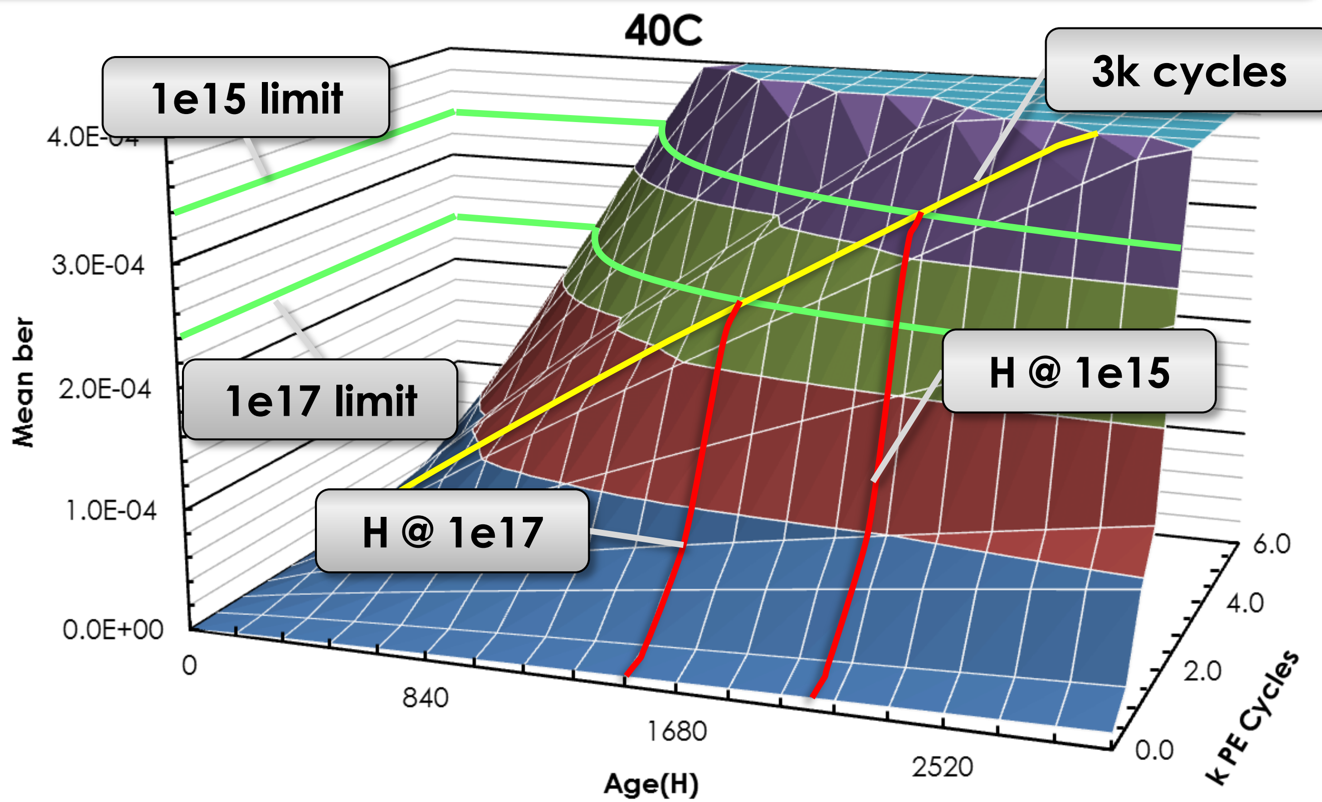
SSDs running at spec are at high risk of data loss

- A consumer SSD operating at an NRRE interval of 1e15
  - Probability sector loss per operation = 4.2e-12 (512B sectors)
- Assume 10,000 4kB IOPS (on the slow side)
  - 1.3e12 sector ops per year
- Mean years between sector loss = 0.1 (oops…)
  - MTTDL = 850 hours
  - Note: a consumer HDD does 100 4kB IOPS at 1e14 NRRE inteval
    - MTTDL 8.5kHours (1 year)
- Would be better if NRRE was 1e17 for C SSDs
  - 10 Y/sector loss, or 85,000H MTBF

Impacts of NRRE Limits

- JEDEC NRRE interval for consumer SSD is 1e15
- My suggested value is 1e17
- How much would this impact retention?

- 1e15:
  - 2,275H @ 3k PE
- 1e17:
  - 1,525H @ 3k PE
- Net is loss of 1/3 retention
  - (Which was already out of spec)



NAND ber Surface Equation

The empirical ber model from all the devices tested

- The empirical bit error equation looks like this:

A = data age

C = program-erase cycles

R = reads since last write

a is a constant scale factor

Power law in age, reads

$$\textbf{bit\_errors\_stripe} = hA^{k}R^{g} + a/(1+(b/C)^{d}) \qquad \textbf{(eqn. 1)}$$

Log-logistic-curve in PE cycles

- No one said this was going to be easy…

Reliability Examination

Device Specifications

- NRRE interval spec is 1e15
  - Equivalent probability of sector loss = 4.2e-12
- Retention specification of 1 year at 3,000 PE cycles at 70C

Measured Results

- NRRE limit reached at 3 months at 3,000 PE cycles at 40C
- Prob Sector Loss at 1 year projected as = 7.7e-5 (NRRE interval 5.6e7)
  - Missing retention limit by 4x results in a probability of sector loss of 1e7x!

How can the devices be this far off of spec?

- Perhaps vendors aware, rely on FTL to relocate
  - Can't easily test the device to confirm
- Perhaps they don't know because they do accelerated temp testing??

FLASH TEMPERATURE TESTING AND MODELING

Spoiler: It's not Arrhenius

Commonly Assumed T Model

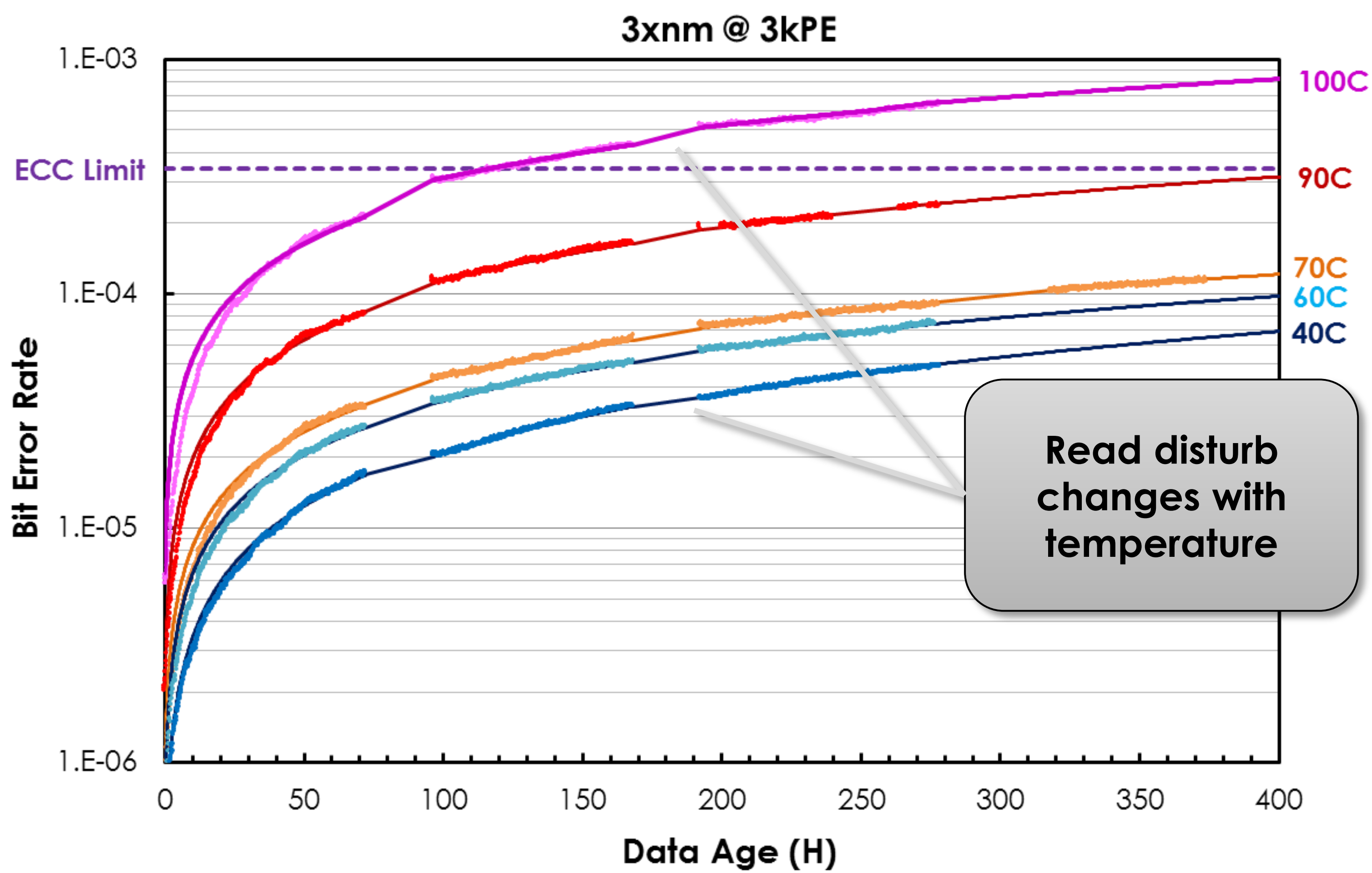
The Arrhenius model is widely assumed to be accurate for NAND

- Arrhenius model acceleration factor: $a_f = exp\left(\frac{E_A}{k}\left(\frac{1}{T1} - \frac{1}{T2}\right)\right)$
- Arrhenius model assumes a single activation energy
  - In a flash cell, failure is the value of the bit changing
  - In an SSD, failure is a sector loss event
    - Note that Arrhenius doesn't know what the ECC correction power is, thus failure rate must not depend on the error count chosen
- Charge de-trapping widely modeled as E_A = 1.1eV
  - Assuming Arrhenius accelerated test at 125C, acceleration factor to 55C is 936x
  - Thus a target 1Y retention would take only 9 Hours
- However, other error mechanisms are known to exist
  - Stress induced leakage current
    - Behaves like negative activation energy - so not Arrhenius
- We need to validate the temperature model
  - Measure time to a given ber
  - Differences between gate level and ber level measurements
  - Not clear we expect NRRE to be Arrhenius even if the gates are!

Measured Temp Behavior of ber

- The curves scale with temperature as:

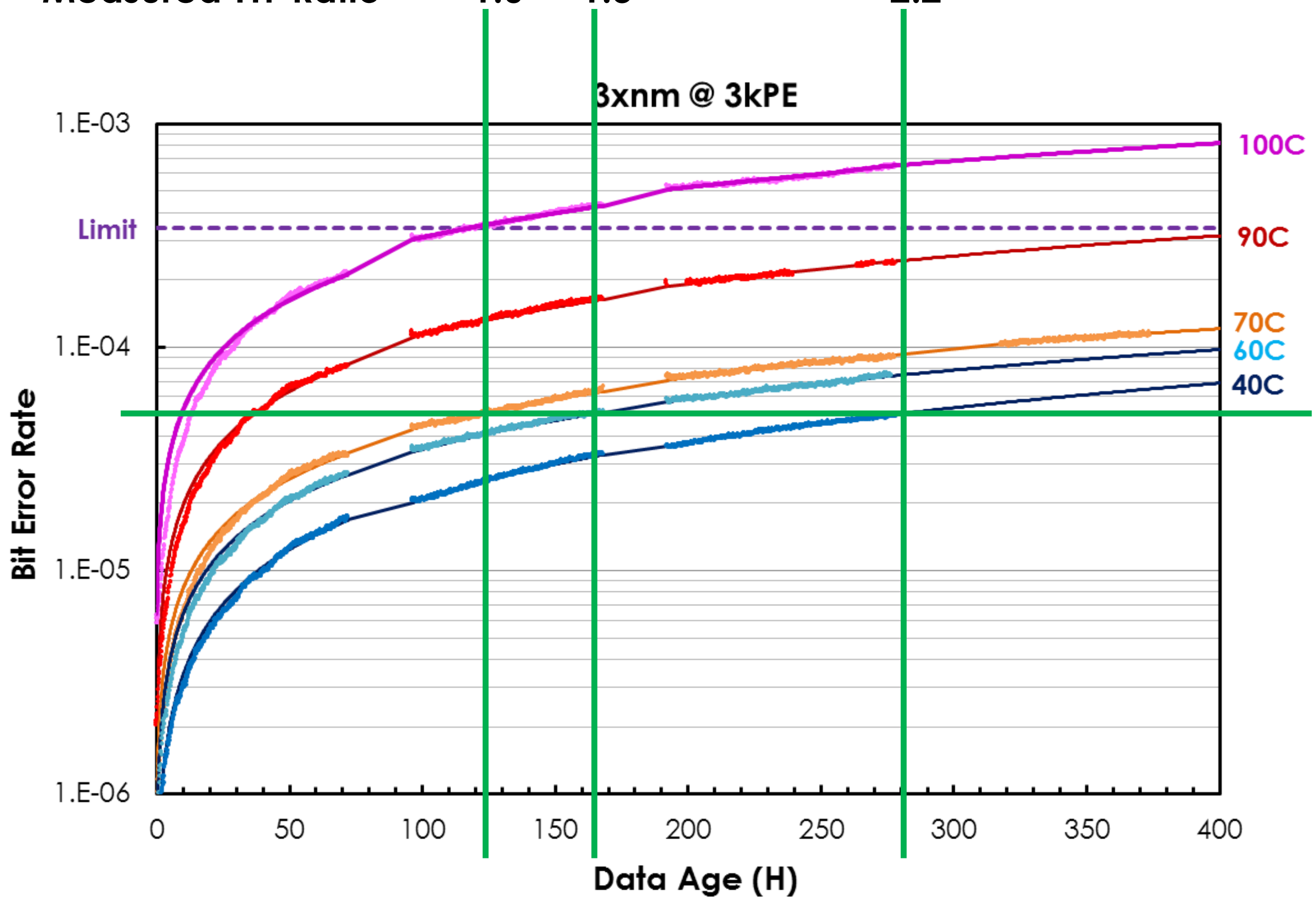
  ber(T1) = m * ber(T2) (log plot so vertical offset is a multiplier)



Time to bit error rate Data

- **Time to ber**
  - Choose 5e-5 ber
  - Arrhenius 1.1eV vastly over estimates the acceleration factor between 70C and 40C

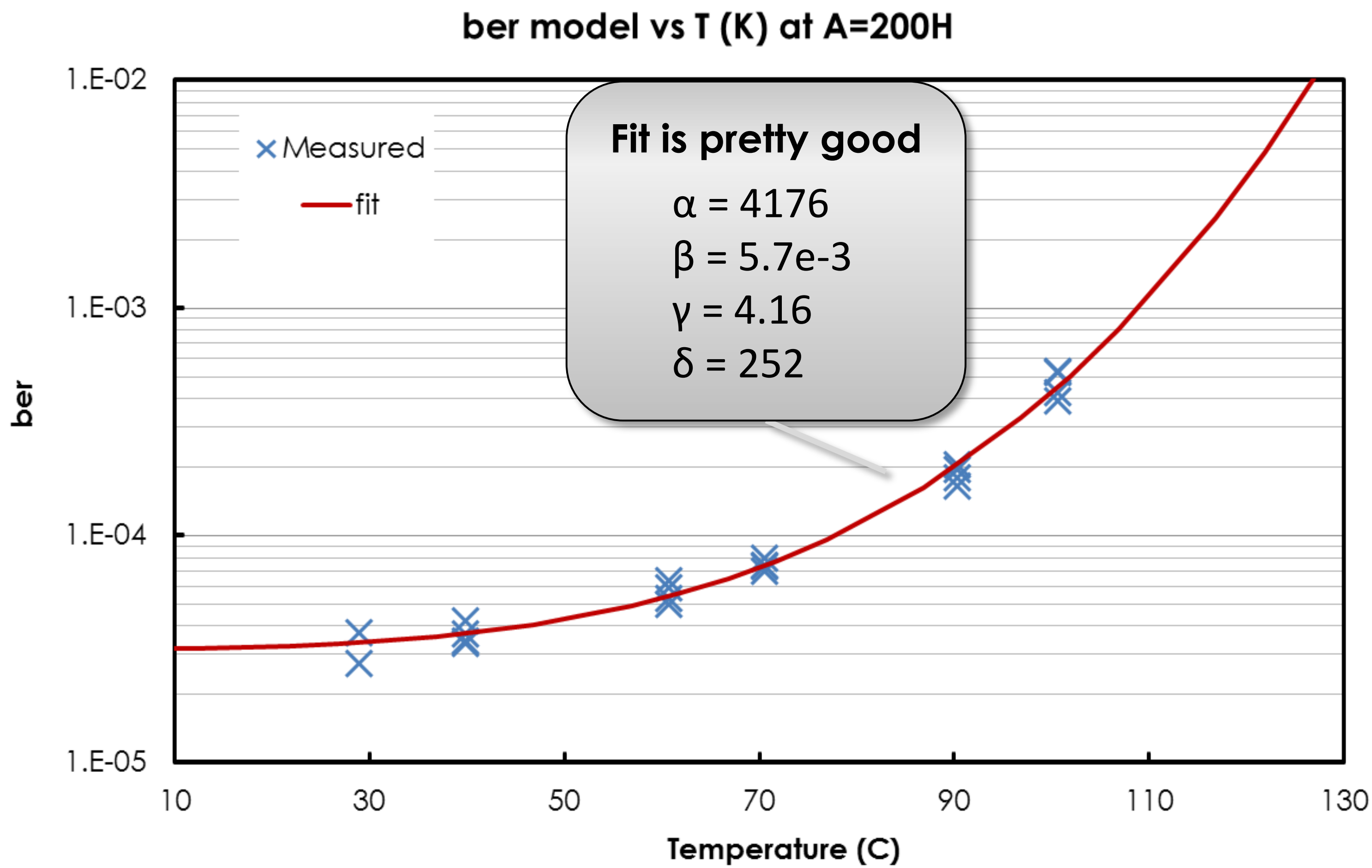
Arrhenius 1.1eV Ratio	**1.0**	**3.0**	**38.0**
Measured TTF Ratio	**1.0**	**1.3**	**2.2**



ber vs. Temperature

- **Observed ber vs T at 3,000 PE cycles, 200 Hours and 10 reads/hour**

$$\text{ber}(T) = \alpha e^{(\beta(T-\delta))^{\gamma}} \qquad \text{(eqn. 2)}$$



ber is Super Exponential in T

The ber temperature model

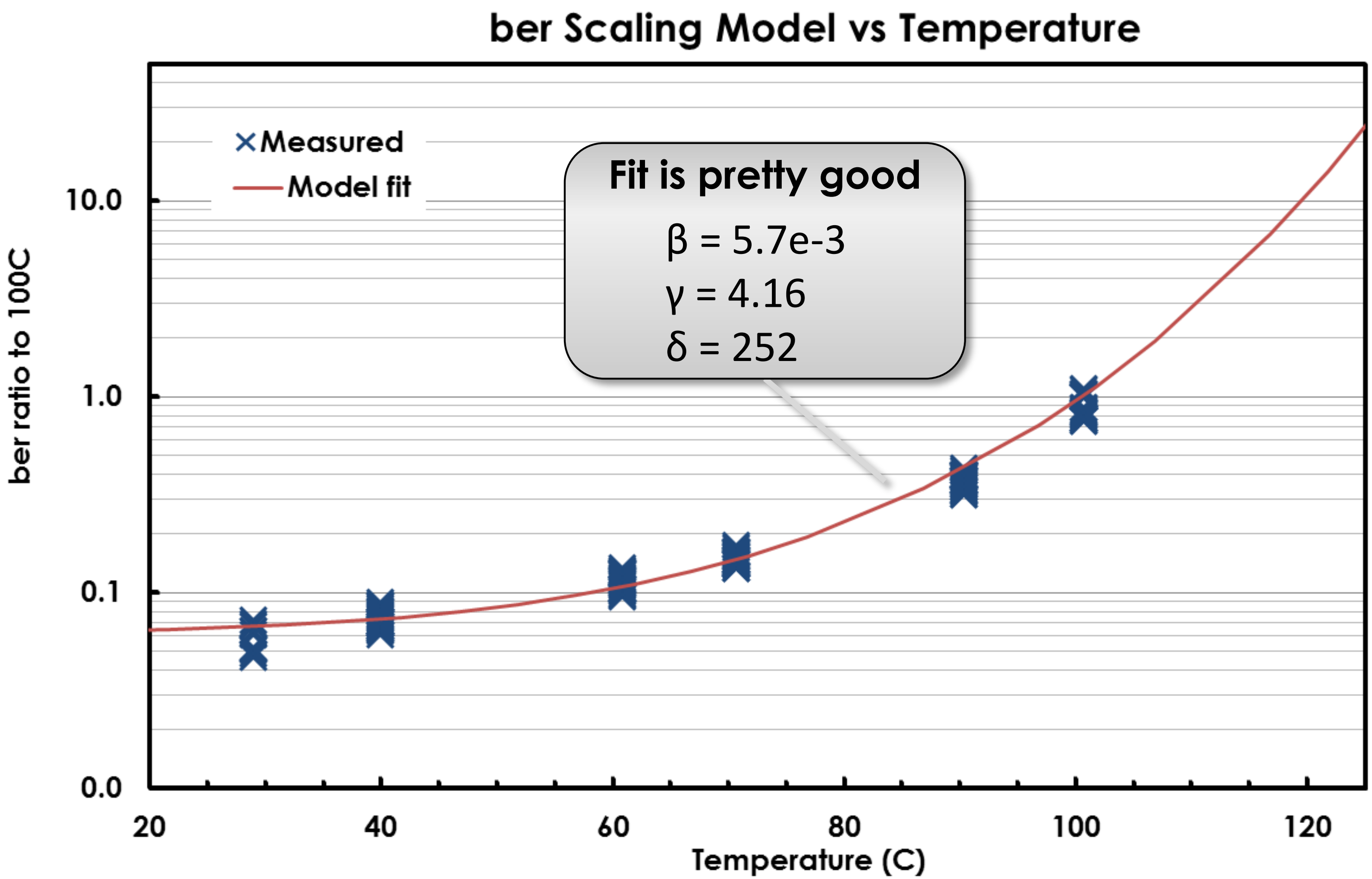
- Observed ber vs T in this range:

$$\mathbf{ber}(\mathbf{T}) = \alpha e^{\left(\beta(\mathbf{T}-\delta)\right)^{\gamma}} \qquad \textbf{(eqn. 2)}$$

- Measure ratio of ber at different temps
  - Each ratio point at constant age, read rate
- Start with a sample set of data ages (30-275H)
  - Compute average ber at that age from measurements
  - Compute ratio of each ber to 100C value
  - Using multiple devices, 8 data ages each, 6 temperatures each

Temperature Scaling Law

$$\frac{\text{ber}(\mathbf{T2})}{\text{ber}(\mathbf{T1})} = e^{\left(\beta^{\gamma}\left((\mathbf{T2}-\delta)^{\gamma}-(\mathbf{T1}-\delta)^{\gamma}\right)\right)} \quad \text{(eqn. 3)}$$



Comparison with Arrhenius

Maybe it's Arrh-ain't-ius?

- The Arrhenius model computes a ratio of times to failure at different temperatures
  - Failure here is a particular bit error rate
  - If Arrhenius, can't depend on the ECC bits chosen
    - So should be independent of the error rate chosen

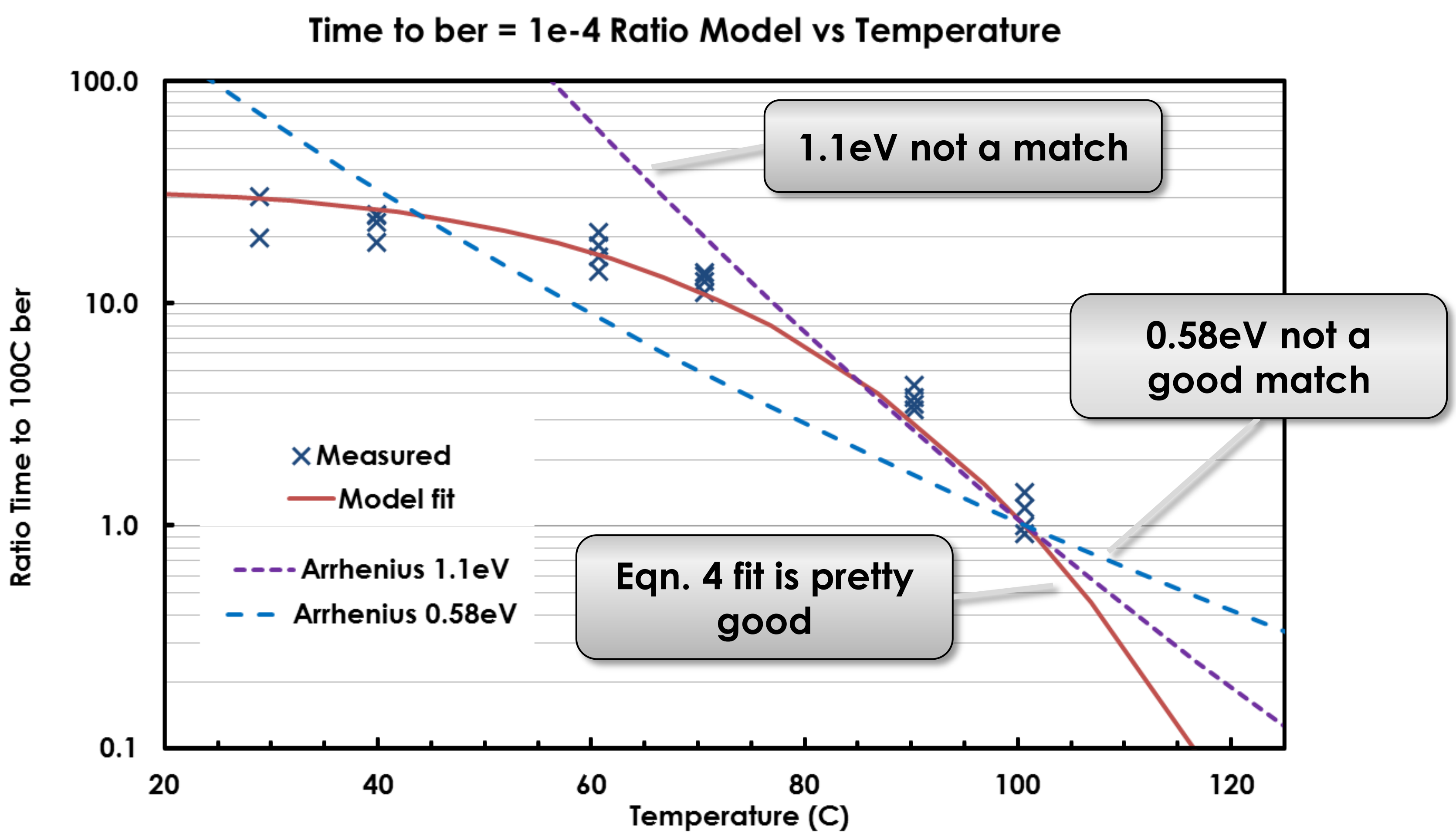
- We can invert the ber equation when reading at constant rate to get the acceleration factor:

$$a_f = \left(\frac{ber(T2)}{ber(T1)}\right)^{1/(k+g)} = \left(e^{\left(\beta^{\gamma}\left((T2-\delta)^{\gamma}-(T1-\delta)^{\gamma}\right)\right)}\right)^{1/(k+g)} \qquad \textbf{(eqn. 4)}$$

Which isn't Arrhenius at all: $a_f = exp\left(\frac{E_A}{k}\left(\frac{1}{T1} - \frac{1}{T2}\right)\right)$

Time to ber Acceleration Behavior

- **Arrhenius @ 1.1eV is not a good match to the data**
- **Arrhenius @ 0.58ev is best Arrhenius fit**
  - A broken watch fit – it's correct at 2 points



Temperature Results

The behavior is not consistent with a 1.1eV activation energy

- Time to failure acceleration factors
- Data is not Arrhenius!

T Hi (C)	T Lo (C)	Model	1.1 eV	0.58 eV
100	40	13x	702x	31x

What might explain the situation

- Point of failure depends on set of worst-case bits from a population
- Tests done on small population at gate level may not show effect
- Extrapolation from short duration tests subject to slope change
  - ber(A) is not exponential

Temperature acceleration models should always be validated

- Best to test full device, operating and aging at temp (as in the field)
- With a testable device, you can measure the total behavior (HMI)
  - An interface which bypasses FTL and ECC allows tests to be done in the field
- An unverified model isn't suitable for accelerated testing

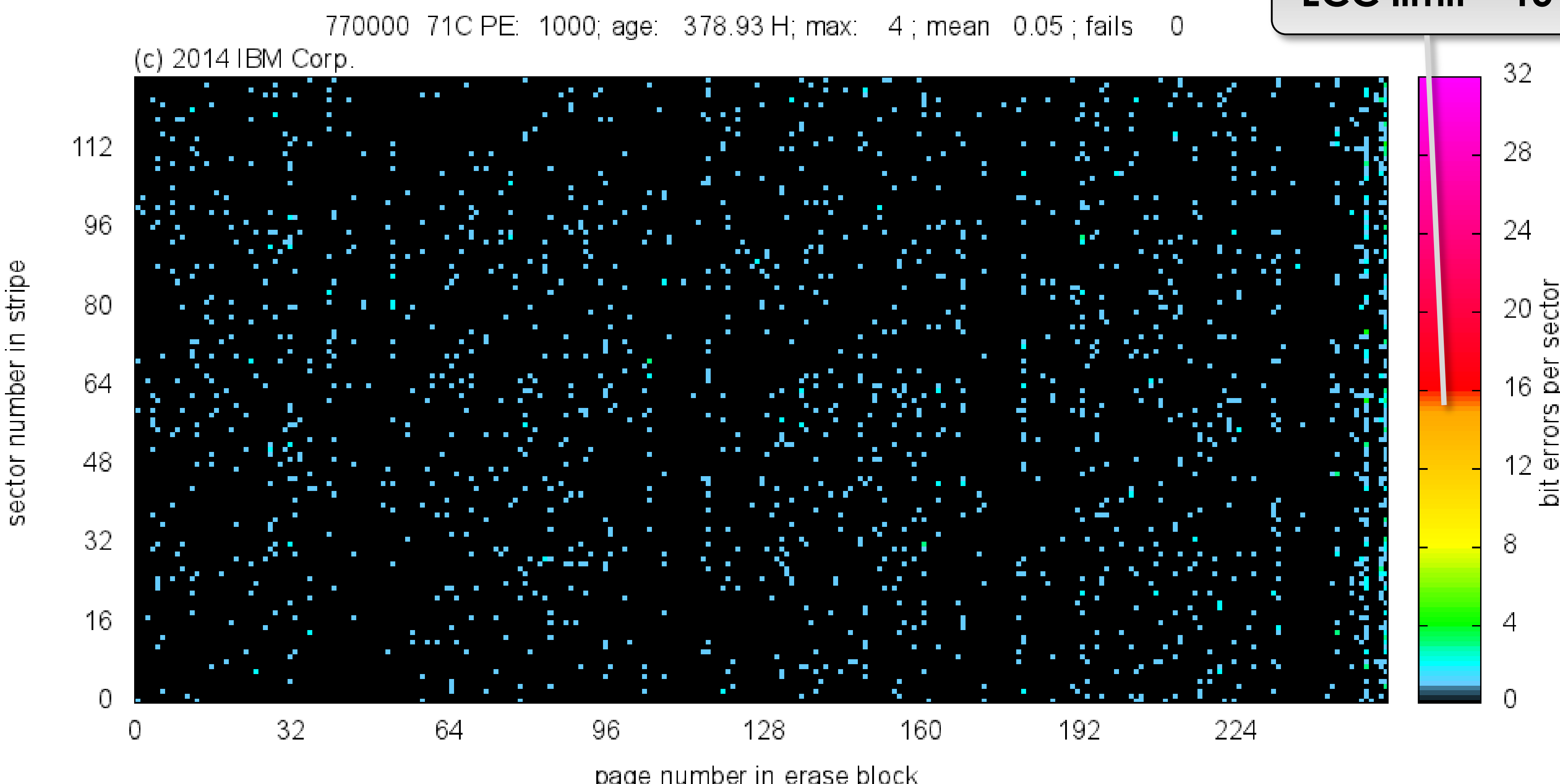
ERROR UNIFORMITY

And you thought it couldn't get more confusing

Location Dependence

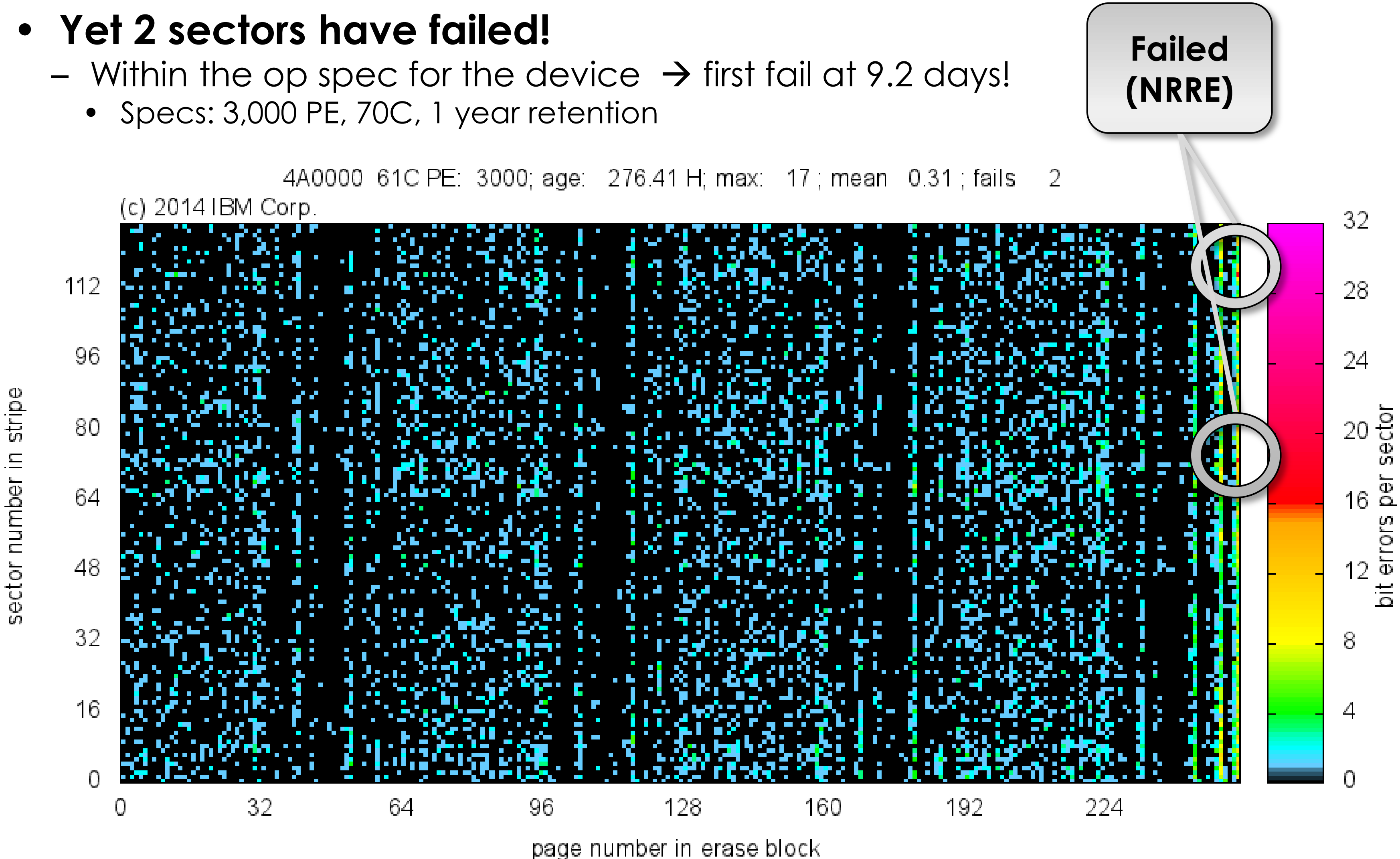
- **Sector error count bitmap (3xnm)**
  - "If you see red, the sector is dead"
  - Error patterns aren't that random looking…

ECC limit = 15



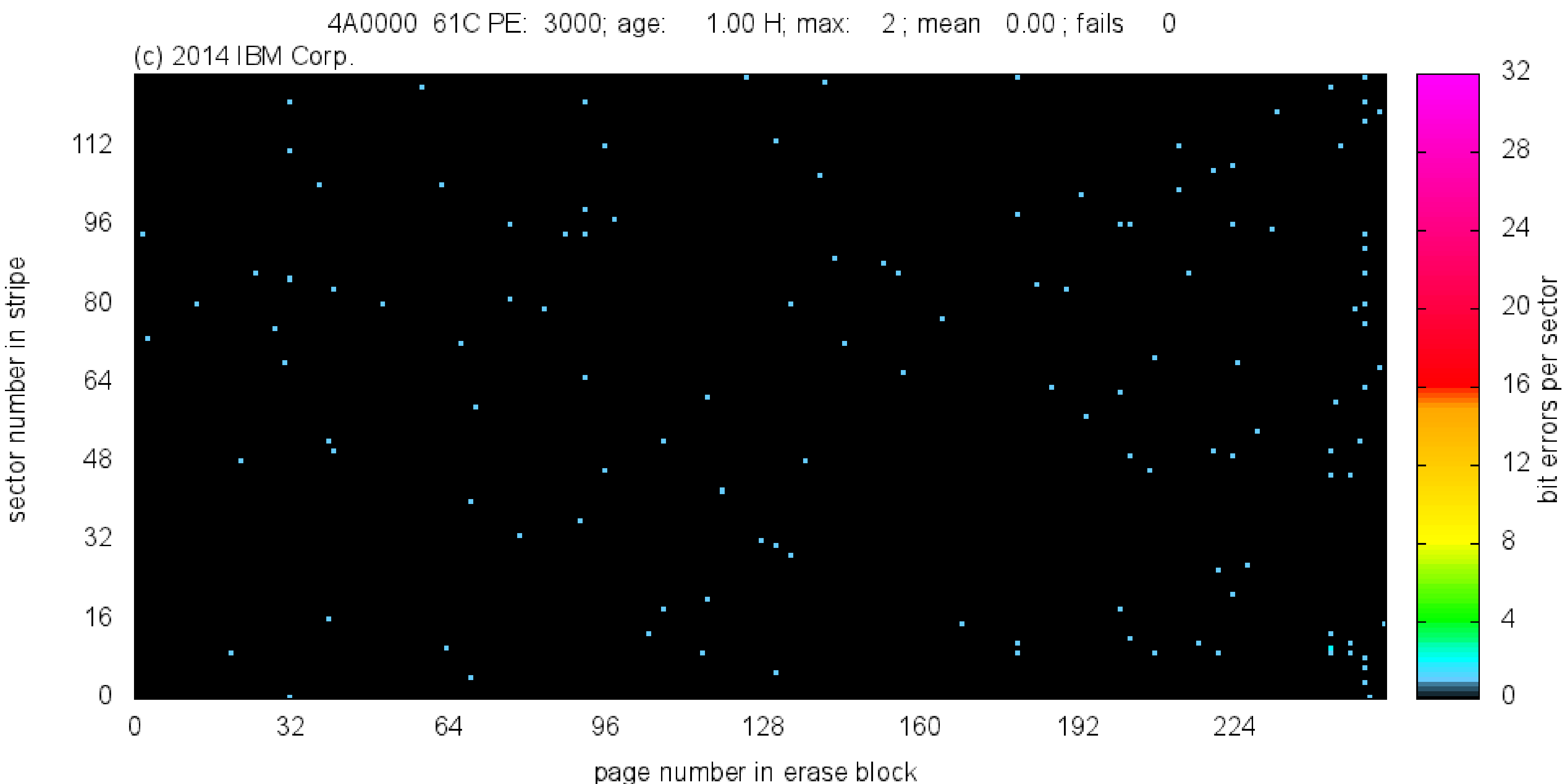
3xnm, @ 60C, 3K PE, 276Hours

- **The mean is in spec**
  - 0.31 errors/4096bits = 7.6e-5 ber
- **Yet 2 sectors have failed!**
  - Within the op spec for the device → first fail at 9.2 days!
    - Specs: 3,000 PE, 70C, 1 year retention



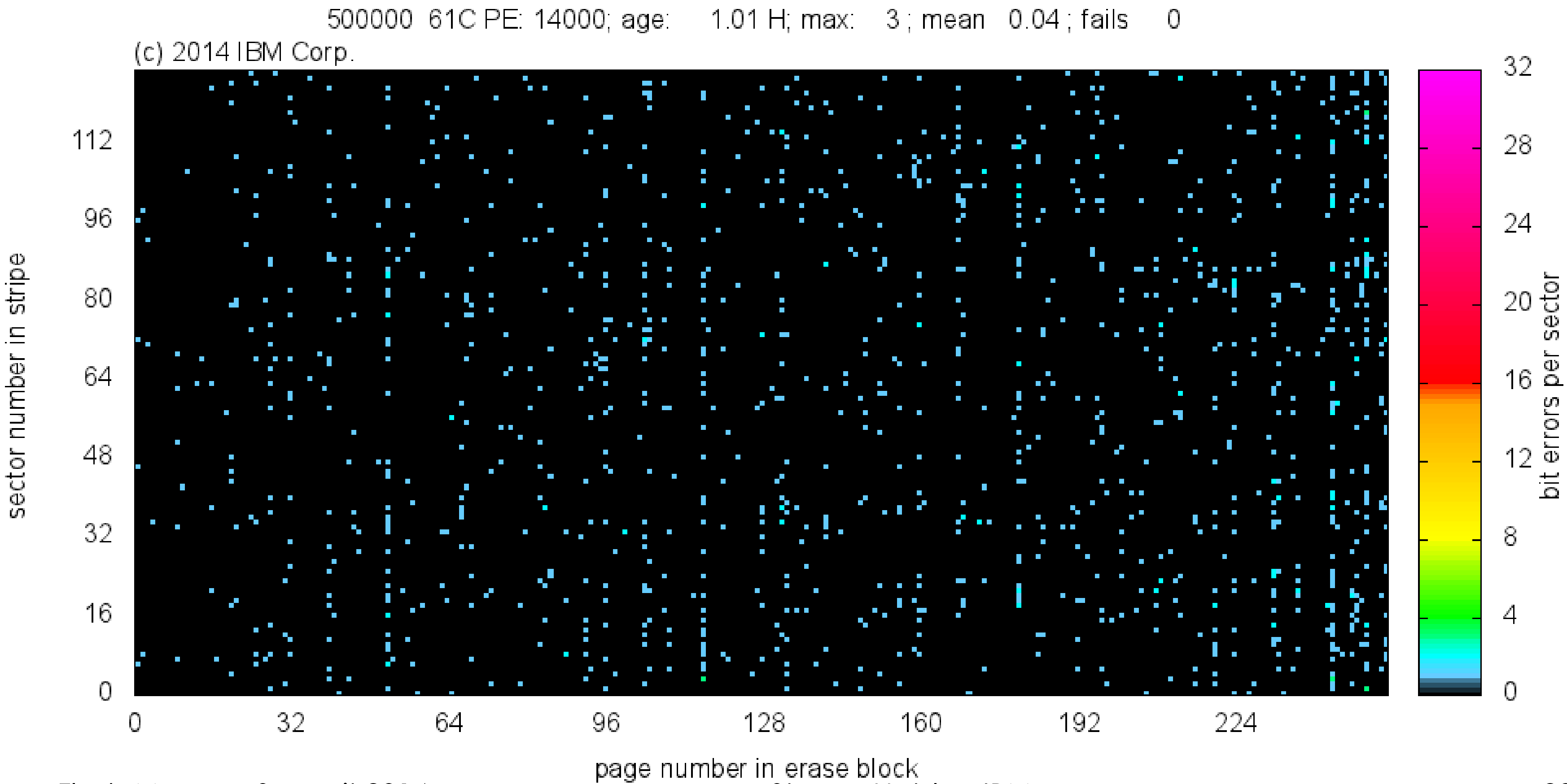
Error Patterns vs. Age

- **Time evolution of aging – operating within spec**
  - How quickly they forget…
    - (The bits, that is)



An Extreme Case

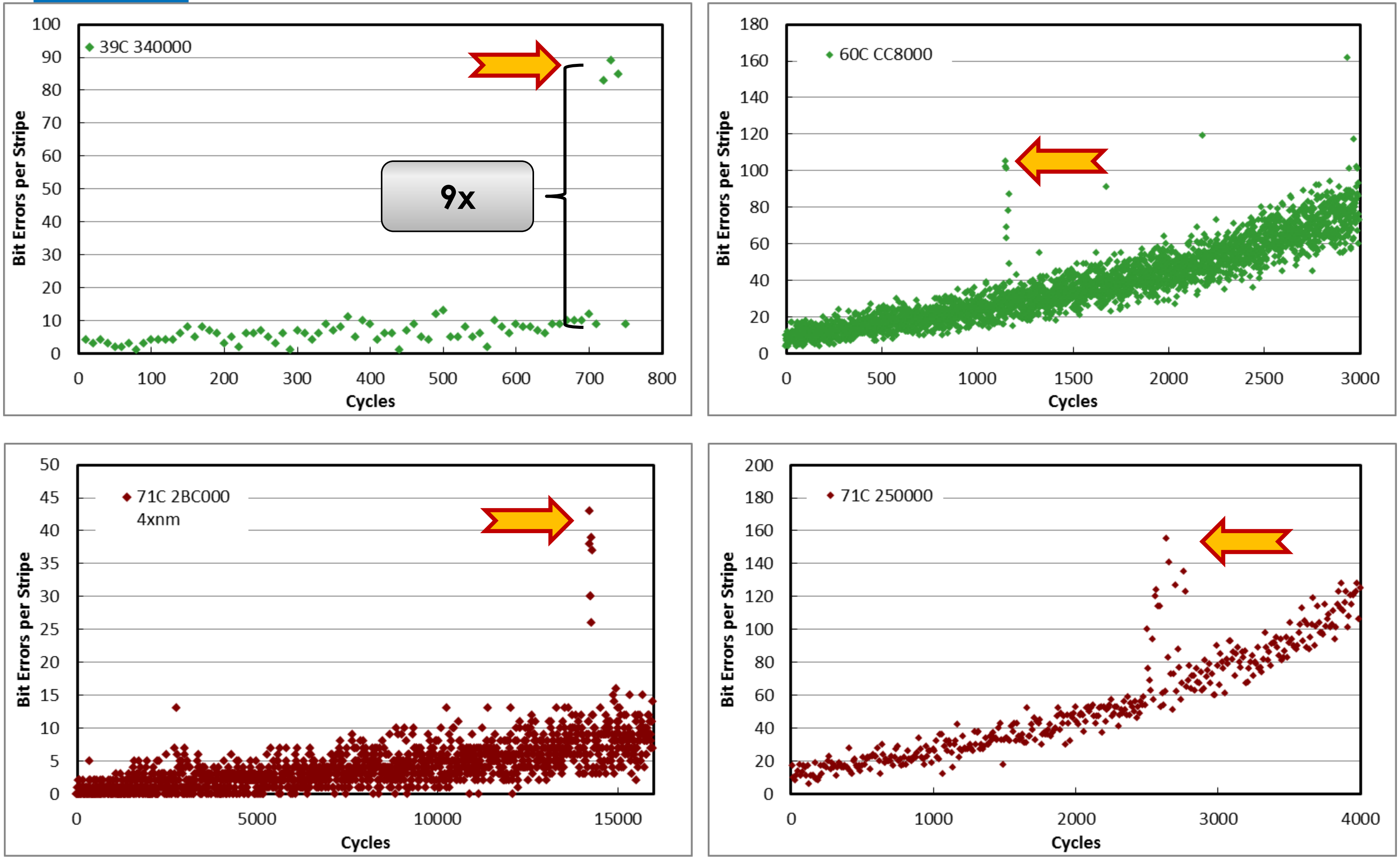
- **You could push endurance farther with stronger ECC**
  - Max 101 errors/sector → 2.5% ber



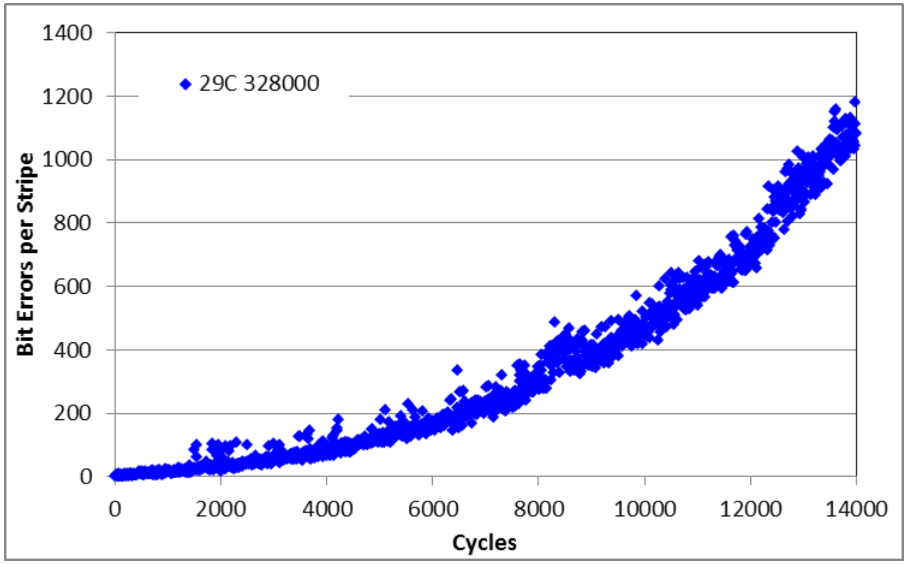
THINGS THAT GO BUMP IN THE DRIVE: ERROR FOUNTAINS

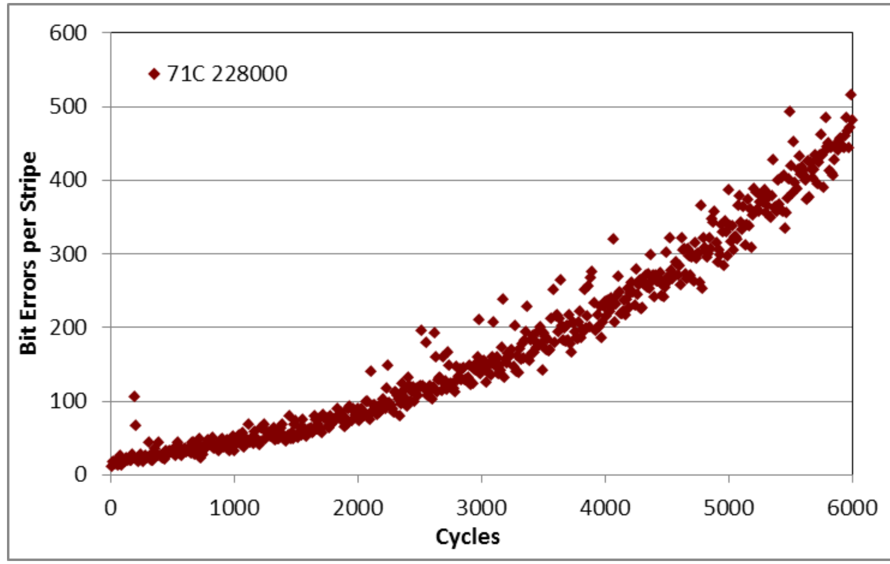
Where SSDs wrong a few writes

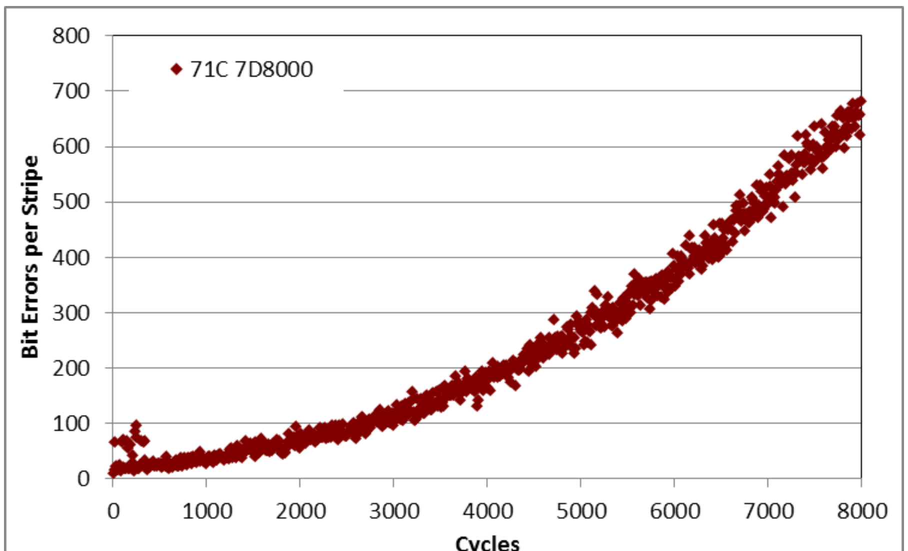
Error Fountain Taxonomy - Isolated

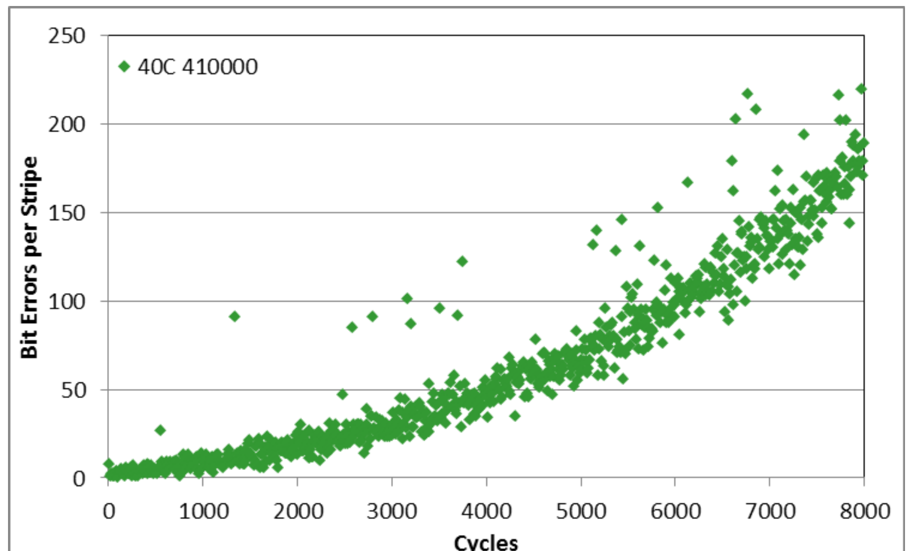


Error Fountain Taxonomy - Noisy

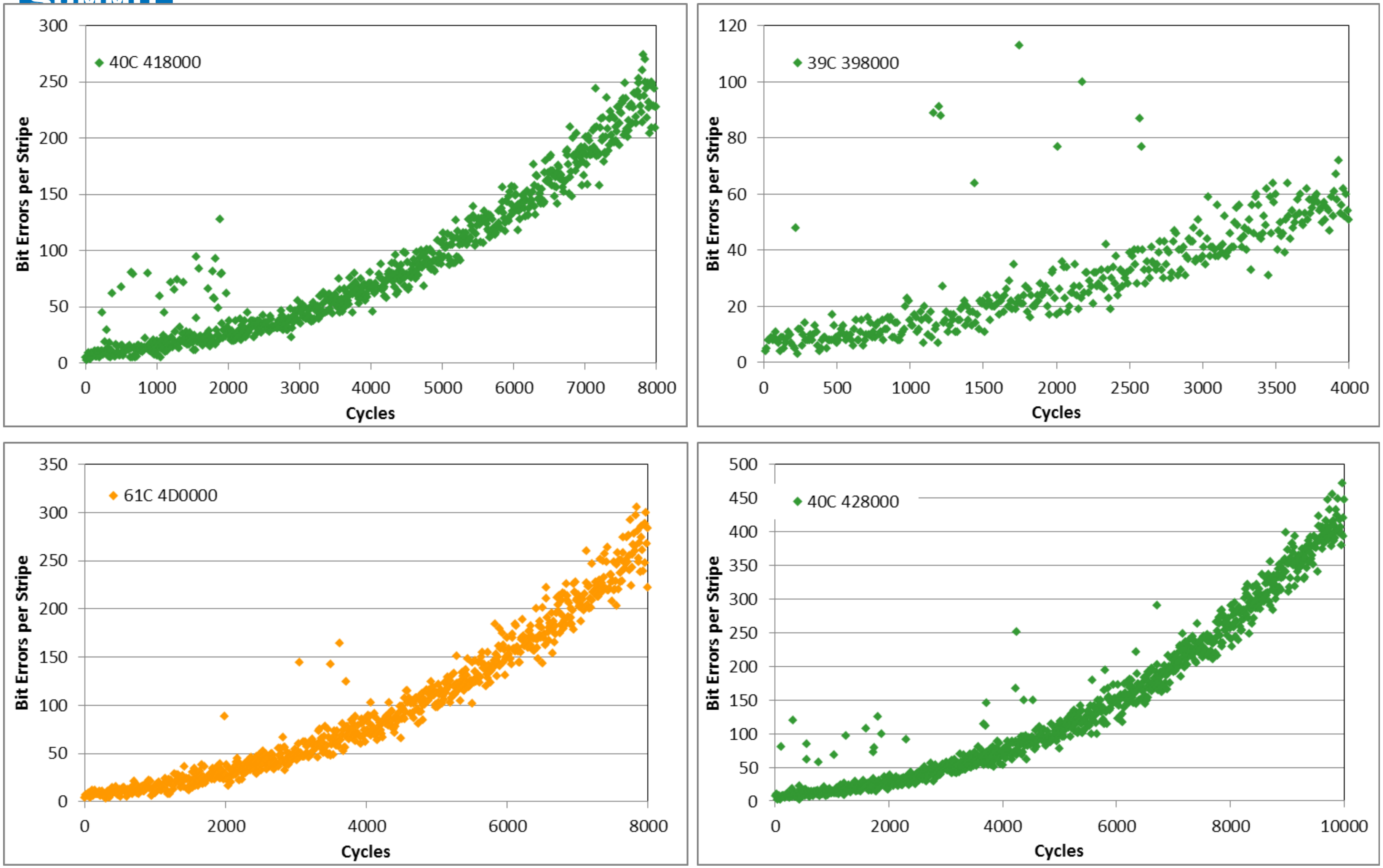




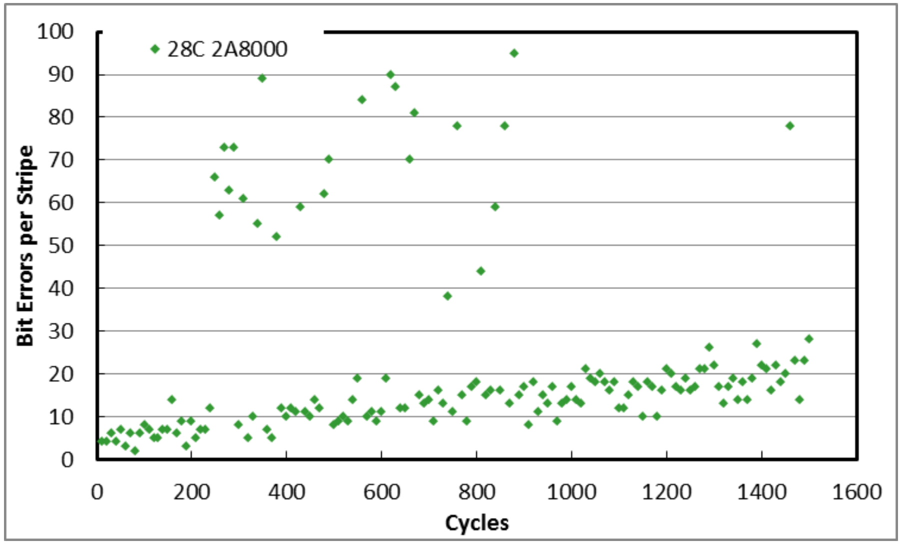


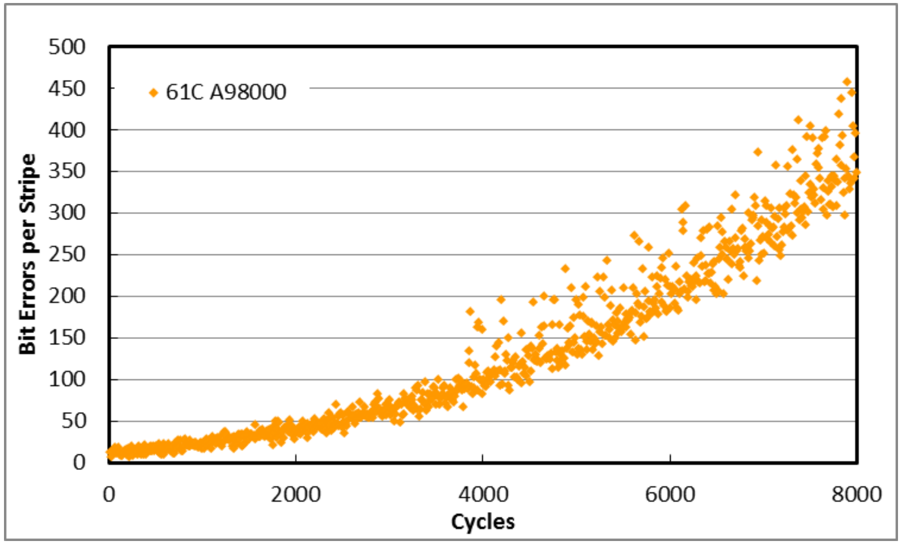


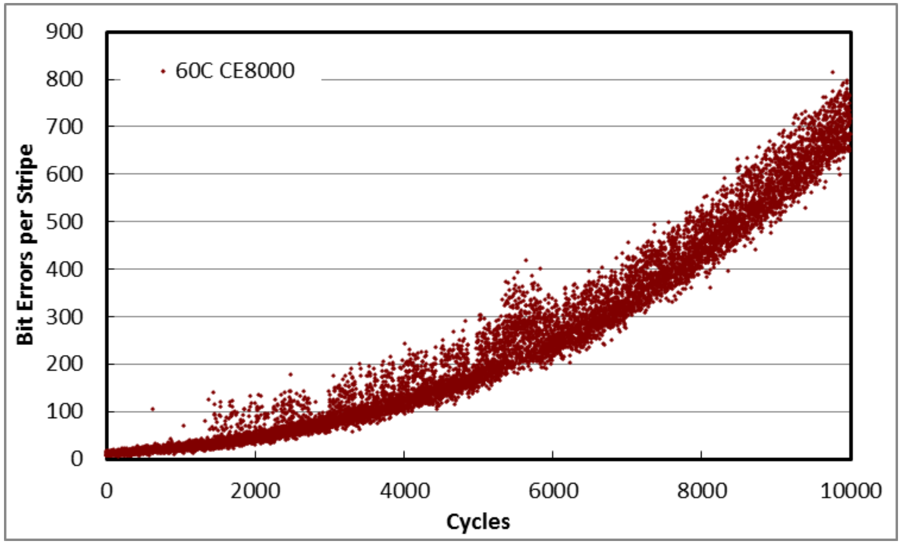
Error Fountain Taxonomy - Bursty

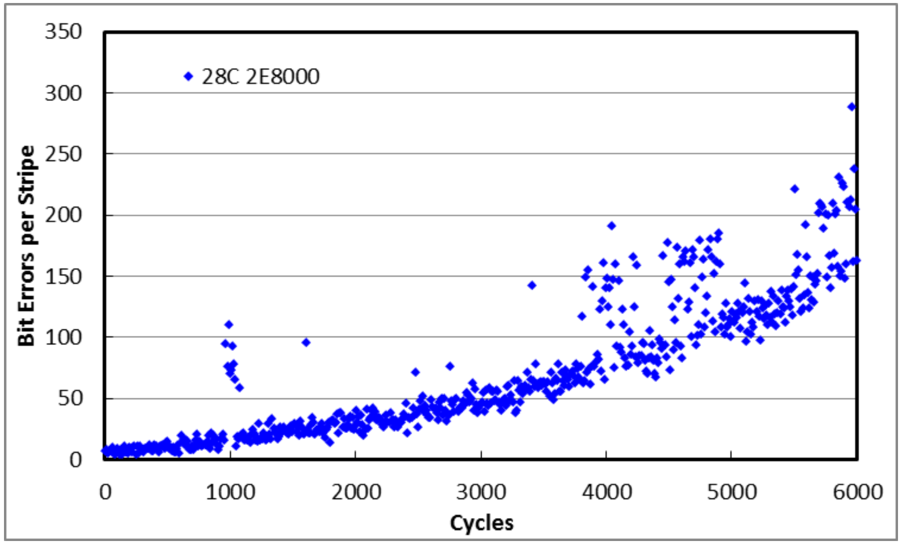


Error Fountain Taxonomy – Bellagio?



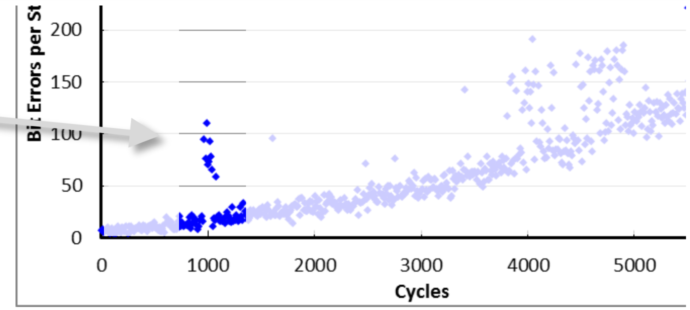


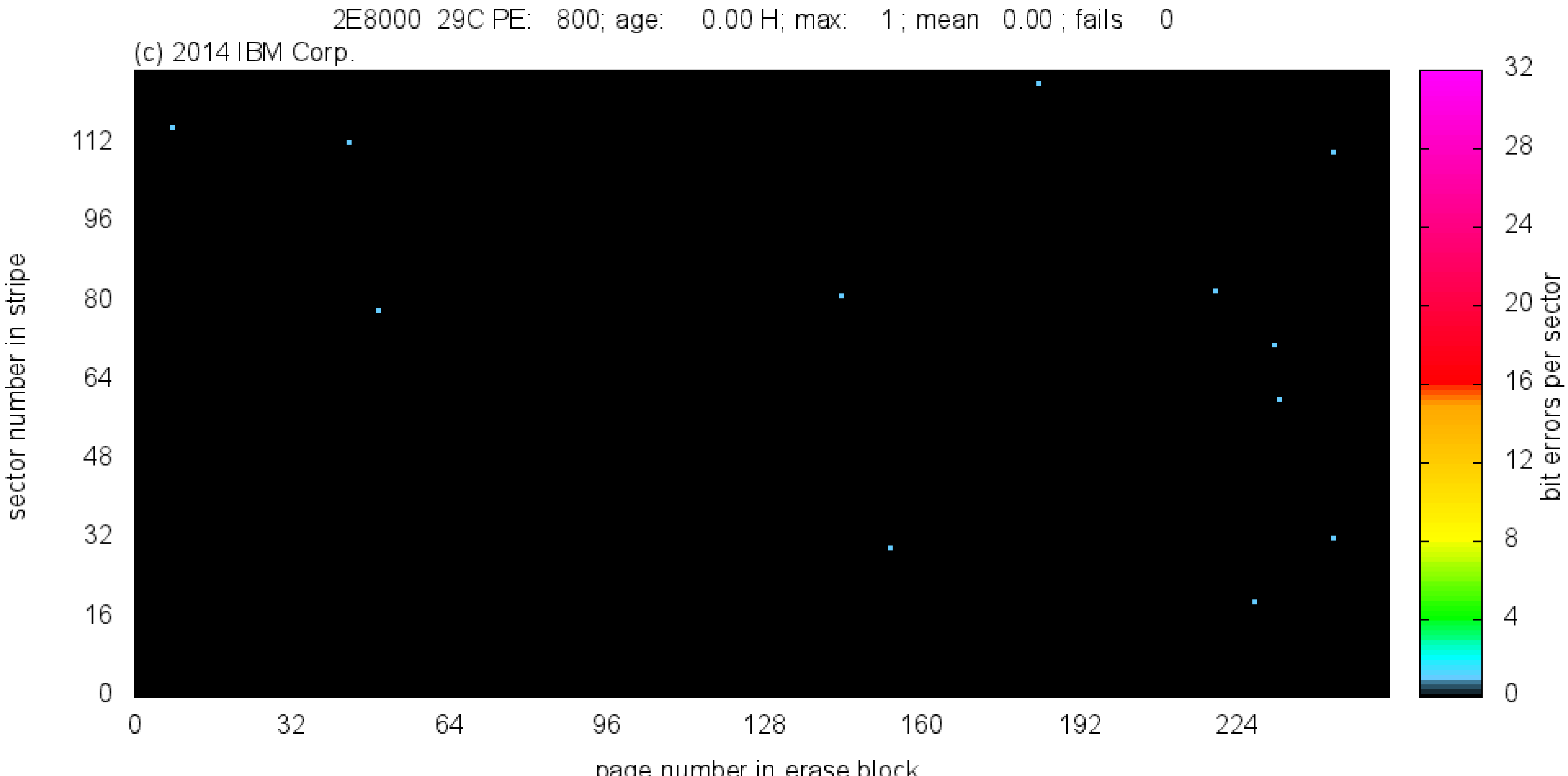




Error Fountains- Bellagio Behavior 1

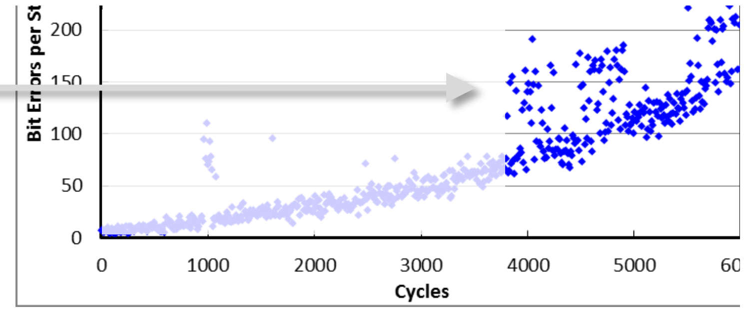
- **Let's look at PE 1,000**
  - See what's going on

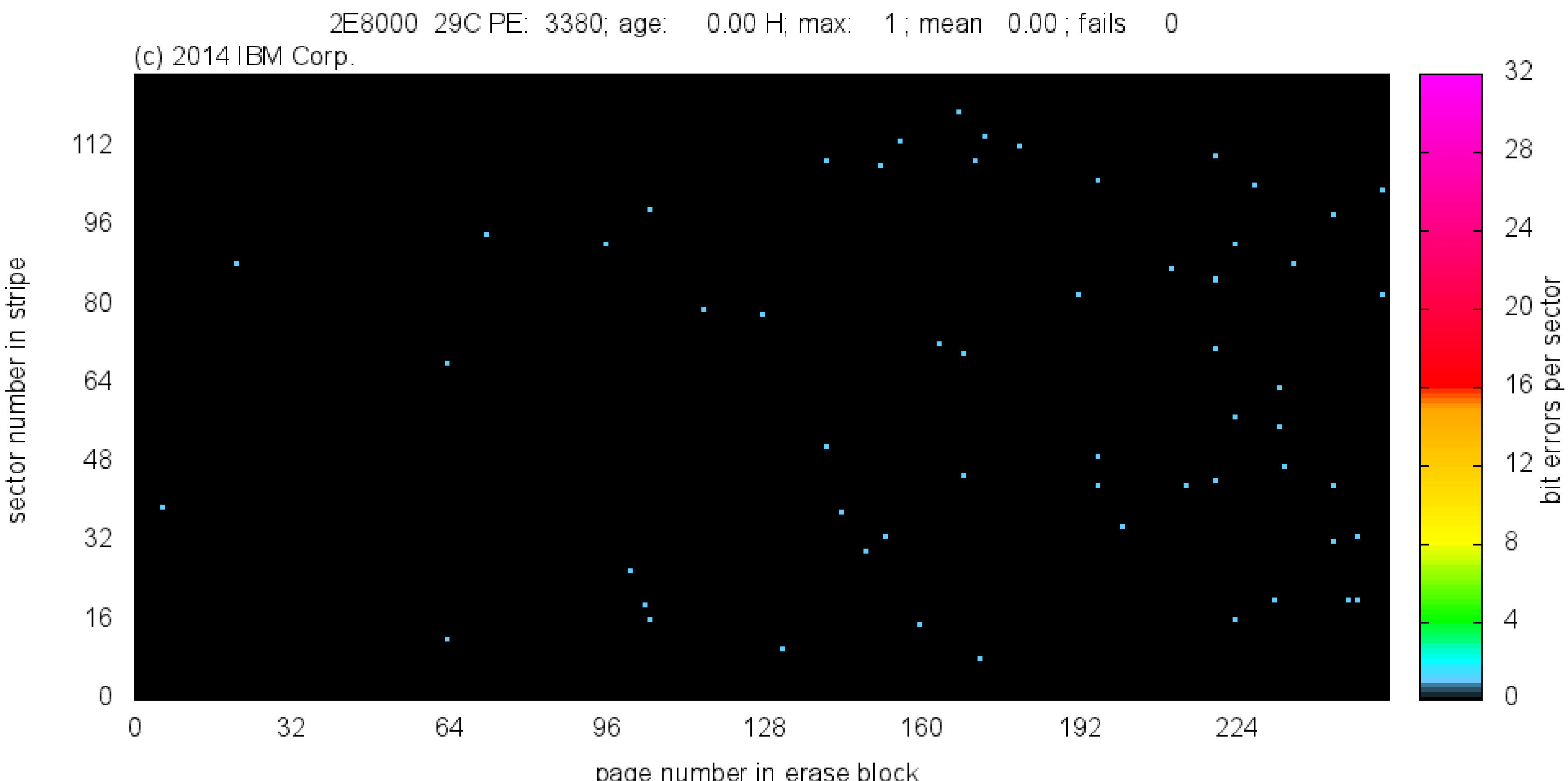




Error Fountains- Bellagio Behavior 2

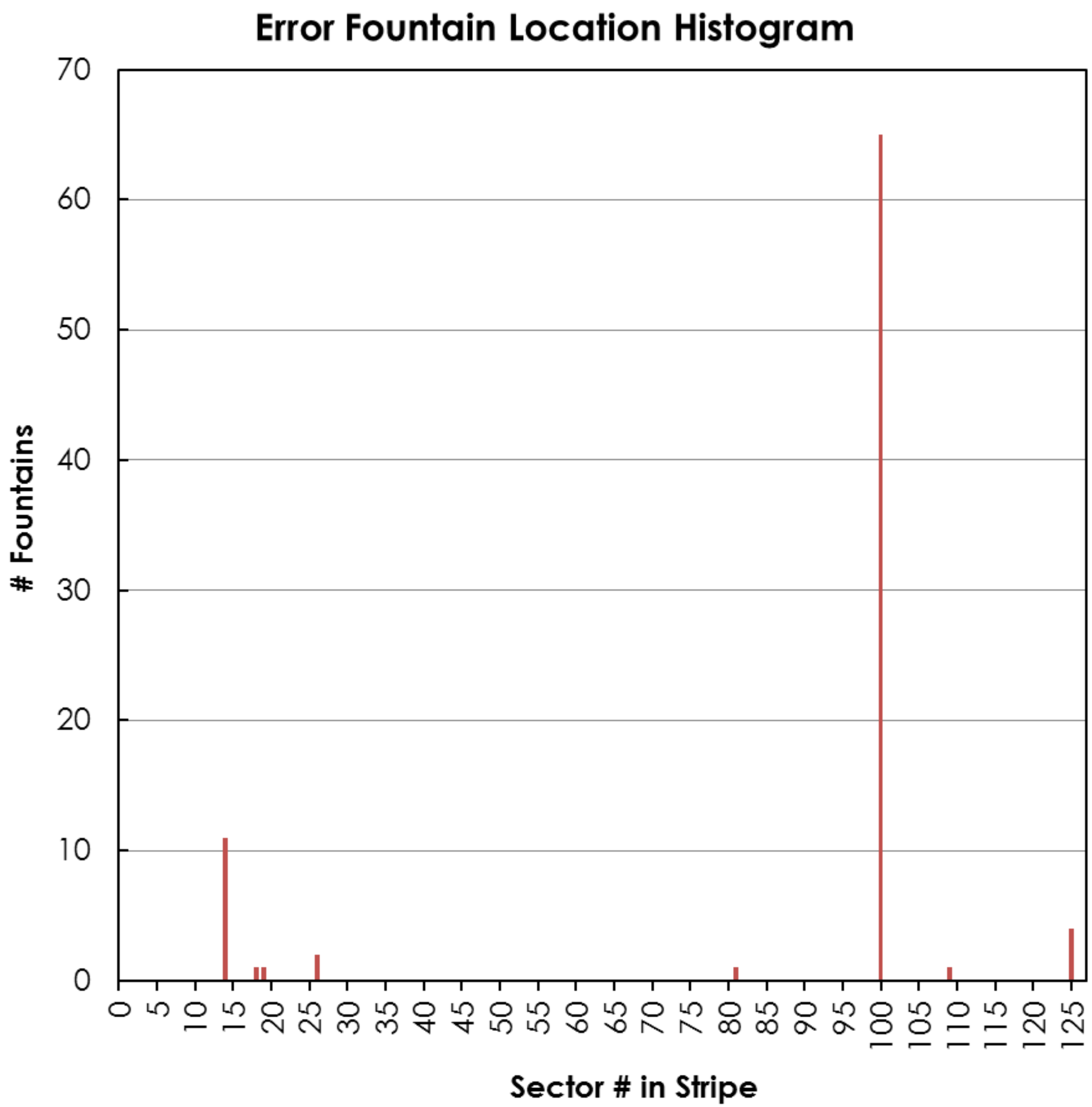
- **Let's look at PE 4,000**
  - Watch the party
  - Did you see the double?





Error Fountain Summary

- **Errors fire along most sectors of a given page offset in an erase block**
  - Not always cleared on a subsequent PE cycle
- **Many tend to be repeat offenders**
  - Note: this data is only reading every 10th write, so we don't have full data
- **I wrote an automated detection algorithm**
  - Results shown at right for the prior data sets



Summary

- **We have a testing methodology for SSDs**
  - Utilizes an SSD with a host-managed interface
- **Have a model for bit error rate (Age, Cycles, Reads)**

$$\mathbf{ber} = \mathbf{h}A^{k}R^{g} + \frac{a}{\left(1+\left(b/C\right)^{d}\right)} \qquad \text{(eqn. 1)}$$

  - A: age, R: reads since written, C: cycle count
- **Have a model for the temperature dependence**

$$\mathbf{ber}(T) = \alpha e^{\left(\beta(T-\delta)\right)^{\gamma}} \qquad \text{@ fixed age, cycles, reads} \qquad \text{(eqn. 2)}$$

$$a_f = \left(\frac{\mathbf{ber}(T2)}{\mathbf{ber}(T1)}\right)^{1/(k+g)} = \left(e^{\left(\beta^{\gamma}\left((T2-\delta)^{\gamma}-(T1-\delta)^{\gamma}\right)\right)}\right)^{1/(k+g)} \qquad \text{(eqn. 4)}$$

  - Not Arrhenius – but allows temperature acceleration
- **Saw some new effects**
  - The errors are not very uniformly distributed
  - Error fountains
- **There's more we didn't get to**
  - To learn more, visit: DrHetzler.com