Qualification, Reliability, and Extrapolation Under Uncertainty

Module 8 : Test and Qualification | Lecture 2

Asif Khan
Associate Professor
School of Electrical and Computer Engineering
School of Materials Science and Engineering
Georgia Institute of Technology

Where we left off, and what changes today

1

Last lecture — Test

Billions of measurements. Immediate, complete labels. One defective part in ten thousand. The difficulty was imbalance, and the answer was screening plus a cost function.

2

This lecture — Qualification

Tens of units. Weeks of stress. Often zero failures. The difficulty is scarcity, and the answer is structure, priors, and honest uncertainty.

Qualification is an extrapolation problem with almost no data.

- **Interpolation forgives a wrong model.** Extrapolation does not. Four decades of extrapolation in time makes the model choice, not the data, the dominant term.

SEGMENT 1

What qualification actually is

Standards, sample sizes, acceleration models, and lifetime distributions.

What qualification means

- **A qualification is a fixed programme of stress tests** that a product must pass before it can be sold. It is agreed between supplier and customer, usually by reference to a published standard.
- **It certifies the product and the process together.** A change to either one — a new package, a new fab, a shrink — triggers a requalification.
- **It is a gate, not a measurement.** The output is a pass or a fail against a stated criterion, not an estimate of lifetime with error bars.

That last point is the opening for this lecture. A gate answers yes or no. The engineering question is how much margin there is, and by how much the answer could be wrong.

A TYPICAL QUALIFICATION PROGRAMME

High-temperature operating life
1000 h at 125 °C, biased

Temperature–humidity bias
1000 h at 85 °C / 85 % RH

Unbiased highly accelerated stress
96 h at 130 °C / 85 % RH

Temperature cycling
1000 cycles, −55 °C to 150 °C

Early-life failure rate
48 h burn-in on a large sample

Electrostatic discharge
Human body model, charged device model

Latch-up and solderability
Per the relevant standard

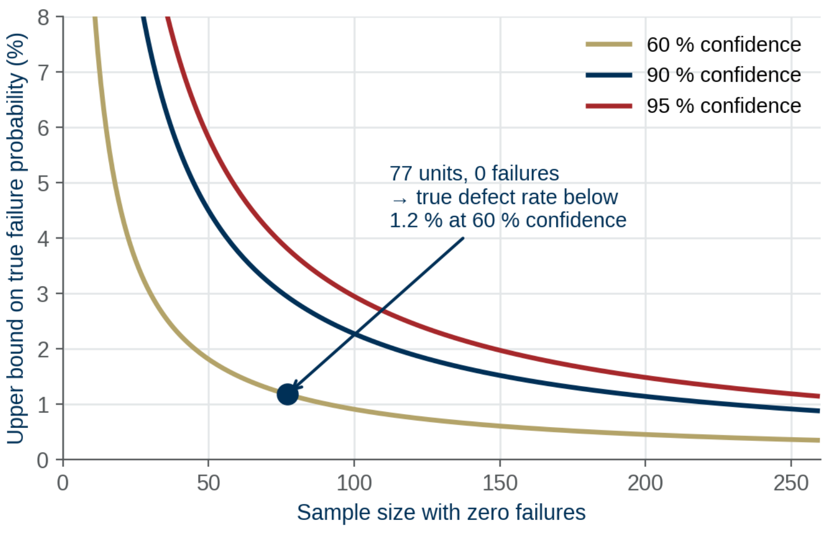
The standards landscape

JESD47 JEDEC	Stress-test-driven qualification of integrated circuits. Defines the test suite, the conditions, and the sample sizes for commercial parts.
JEP122 JEDEC	Failure mechanisms and models for semiconductor devices. This is where the acceleration models and their activation energies are written down.
AEC-Q100 Automotive Electronics Council	Stress qualification for integrated circuits in automotive use. Adds grades by operating temperature and is substantially harder than JESD47.
JESD22 series JEDEC	The individual test method definitions that the suite documents refer to.

Read JEP122 if you read only one. It is the physics document, and every acceleration model in the rest of this lecture appears in it.

The statistics of a real qualification

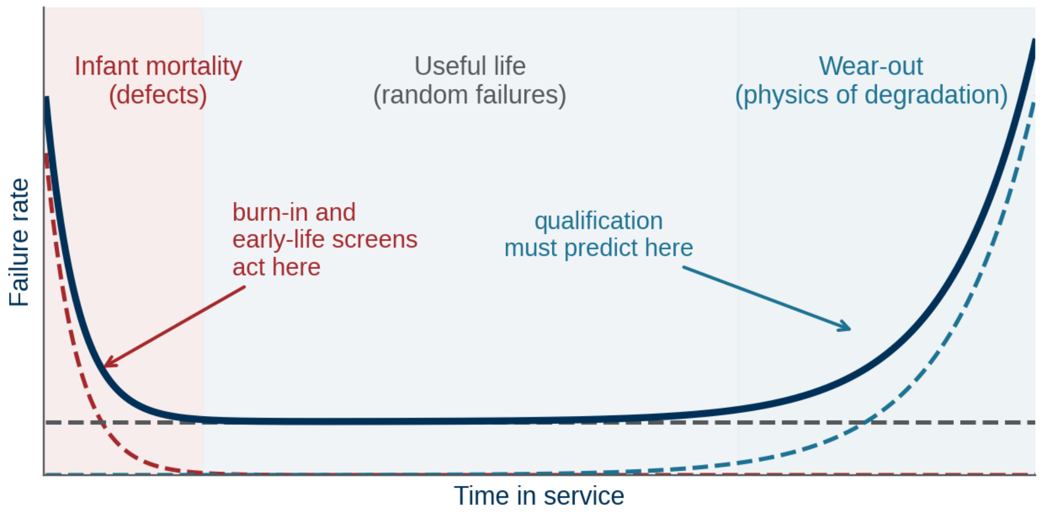
A common criterion: three lots, 77 units per lot, zero failures allowed.



- **Where 77 comes from.** It is the smallest sample that demonstrates a defect rate below about 1 % at 60 % confidence when nothing fails.
- **What zero failures proves.** An upper bound on the failure probability, at a stated confidence. Nothing more.
- **What it does not prove.** It gives no estimate of the lifetime distribution, no shape parameter, and no margin.

The bound scales as 1/n. Proving 100 DPPM directly by this method would need tens of thousands of units.

The bathtub curve, and which part each activity owns



- **Qualification targets wear-out.** It asks whether the rising part of the curve stays beyond the product's intended life.

THREE REGIMES, THREE CAUSES

Infant mortality

Manufacturing defects. Falls with time because the weak parts die off. This is what last lecture's screens attack.

Useful life

Roughly constant rate. Caused by events rather than by wear: radiation, transients, handling.

Wear-out

Rises with time. Caused by physical degradation of the material: bond breaking, oxide damage, atomic transport.

Acceleration: the only way this is possible at all

- **The problem.** You need a ten-year answer and you have a few weeks.
- **The move.** Run the part at a stress higher than it will ever see in use. Degradation that would take years happens in hours.
- **The acceleration factor** is the ratio of life at use conditions to life at stress conditions. Multiply your test hours by it to get equivalent field hours.

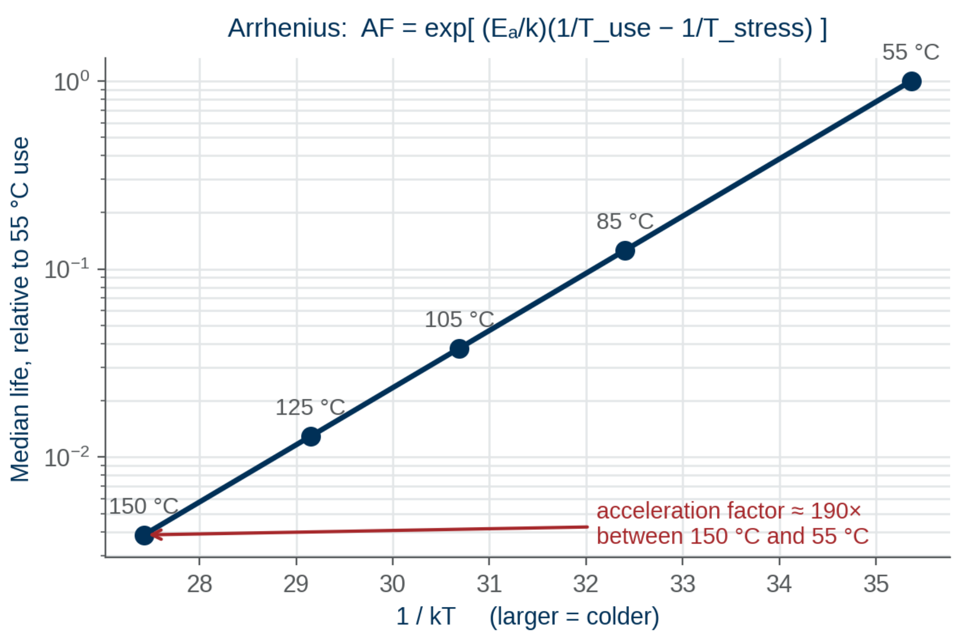
The assumption that makes it work

The stress changes the rate of the failure mechanism and does not change the mechanism itself.

If a new mechanism activates at the elevated stress, the extrapolation is invalid, and the data will usually look fine anyway.

Stress variable	Model	Mechanisms it governs
Temperature	Arrhenius	Most chemical and diffusion-driven processes
Current density	Black's equation	Electromigration in interconnect
Voltage / field	Power law or exponential	Time-dependent dielectric breakdown, BTI
Humidity	Peck's model	Corrosion, package moisture ingress
Temperature swing	Coffin–Manson	Solder and package fatigue under cycling
Several at once	Eyring	The general multi-stress form

Arrhenius: temperature acceleration



$AF = \exp[(E_a / k) \cdot (1/T\_use - 1/T\_stress)]$

Temperatures in kelvin. k is Boltzmann's constant, 8.617×10^{-5} eV/K.

- **E_a is an activation energy,** in electronvolts. It is a property of the failure mechanism and it is the whole model.
- **Typical values.** 0.3 eV for some interface effects, 0.7 eV for many oxide and diffusion processes, above 1 eV for some metallurgical reactions.
- **The sensitivity is severe.** Between 55 °C and 150 °C, E_a of 0.7 eV gives about 190×. E_a of 0.5 eV gives about 35×. A 0.2 eV disagreement moves the answer by more than five times.
- **Consequence.** Whoever chooses E_a chooses the answer. That choice deserves an uncertainty, and it rarely gets one.

Black's equation and the general Eyring form

BLACK'S EQUATION — electromigration

$$MTTF = A \cdot J^{-n} \cdot \exp(E_a / kT)$$

J is current density. n is typically near 2. Metal atoms are pushed along a conductor by the momentum of the electron flow, and a void eventually opens the line.

Two stresses appear together: current density and temperature. Joule heating couples them, so raising current also raises T.

EYRING — the general multi-stress form

$$Life = A \cdot T^{a} \cdot \exp(E_a/kT) \cdot \exp(S \cdot (B + C/kT))$$

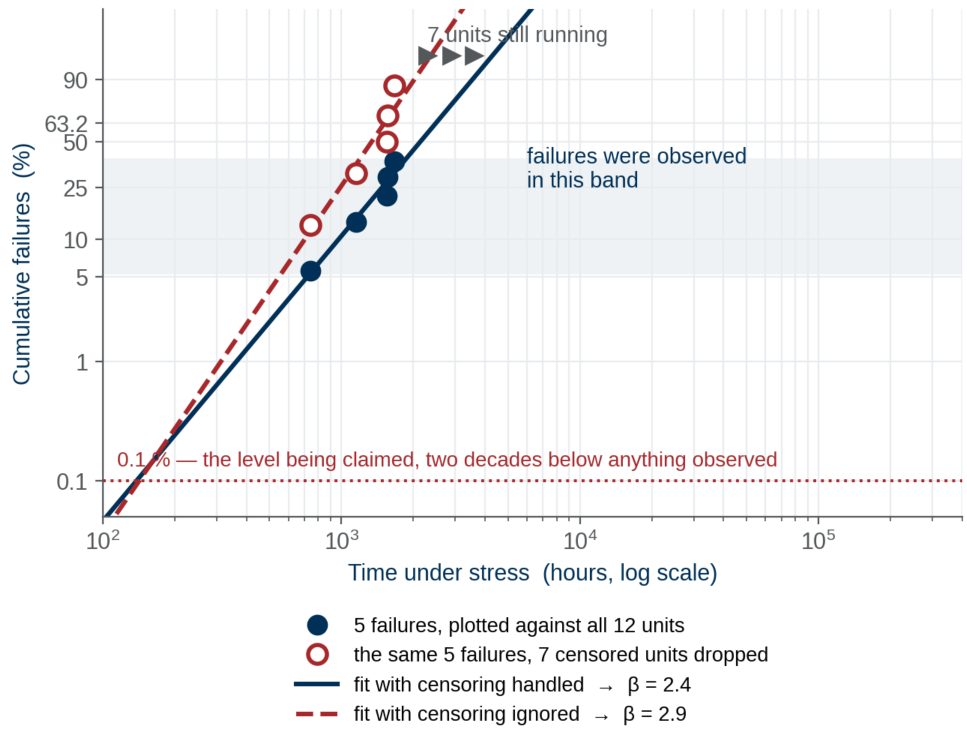
S is a second stress: voltage, humidity, or mechanical load. The cross term C/kT lets the effect of S depend on temperature.

Arrhenius, Black and Peck are all special cases of this form.

What these models have in common, and why it matters for this course

- Each is a small parametric form with two to four constants, derived from a physical picture of one mechanism.
- The constants are fitted from very few points, and they enter the answer through an exponential.
- The extrapolation is far outside the fitted range, so parameter uncertainty is amplified rather than averaged away.

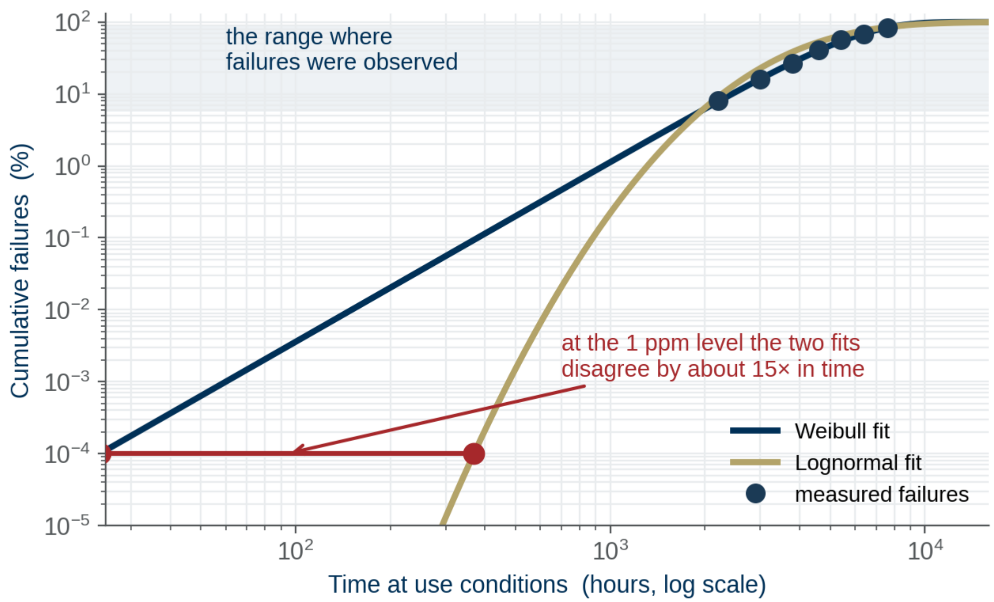
Lifetime distributions and censoring



- **Weibull.** $F(t) = 1 - \exp[-(t/\eta)^\beta]$. β is the shape and it carries the physics: $\beta < 1$ means infant mortality, $\beta = 1$ means constant rate, $\beta > 1$ means wear-out.
- **Censoring.** A unit that has not failed when the test stops is right-censored. You know it survived past that time and nothing more.
- **Handle censored units, do not discard them.** Dropping them biases the fit. The maximum likelihood function has a term for survivors and it must be used.

The shaded band is the range of failure fractions actually observed. The claim sits two decades below it. Acceleration buys time; it does not buy low-probability information.

The model choice dominates the extrapolation



- **Weibull and lognormal are both standard,** and for most datasets both fit the observed range acceptably.
- **They differ in the lower tail.** Weibull has a power-law tail. Lognormal decays much faster. The disagreement grows as you go to lower failure probabilities.
- **The customer cares about the lower tail.** The question is when the first parts per million fail, and that is exactly where the two models disagree most.

Two models that fit the same data differ by roughly 15× in the predicted time to 1 ppm failures. Choosing between them by goodness of fit inside the data is not possible.

Burn-in and early-life failure rate

- **Burn-in attacks the left of the bathtub curve.** Hold parts at elevated voltage and temperature so the weak ones fail on your equipment.
- **ELFR testing measures the early-life rate.** A large sample is stressed for a short time, and the failures are counted.
- **Burn-in is expensive.** Ovens, boards, sockets, floor space, handling, and the risk of damaging good parts by the stress itself.

Policy	Cost	Escape reduction
Burn in every part	Highest	Highest
Burn in a sample per lot	Low	Monitors only; removes nothing
Burn in flagged parts only	Moderate	Close to full, if the model is good
No burn-in, rely on screening	Lowest	Depends entirely on the screen

The same cost calculation as last lecture

Burn-in cost per part versus the escapes it removes. The industry answer has moved from burn-in on everything, to burn-in on a sample, to burn-in on the parts a model flags as risky.

Adaptive burn-in is where Lecture 1 and Lecture 2 meet. The risk model is built from test data. The decision it drives is a reliability decision.

SEGMENT 2

Where AI enters

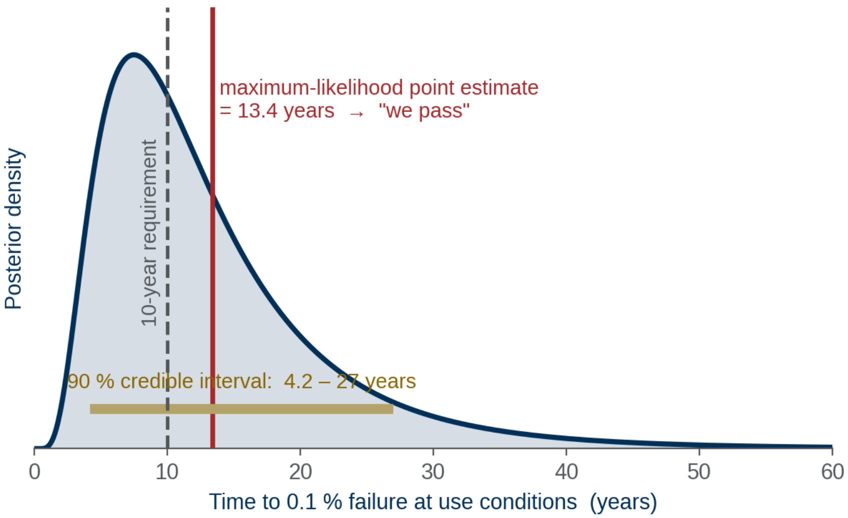
Bayesian inference, active experiment design, and physics-informed surrogates.

Bayesian inference with censored, small-sample data

- **The likelihood already handles censoring.** Failures contribute a density term. Survivors contribute a survival term. Nothing needs to be invented.
- **The prior carries the physics.** You have decades of published values for activation energies and Weibull shapes for known mechanisms. That knowledge belongs in the model.
- **The posterior is the deliverable.** A distribution over lifetime at use conditions, rather than a single number.
- **Small samples are the natural case.** With ten units, the maximum likelihood estimate is unstable. The posterior stays well defined and reports the instability as width.

$$L(\theta) = \prod\_\text{failures}\ f(t_i \mid \theta) \cdot \prod\_\text{survivors}\ [\,1 - F(t_j \mid \theta)\,]$$

Posterior ∝ likelihood × prior. Sample it, and propagate to any quantity you need.



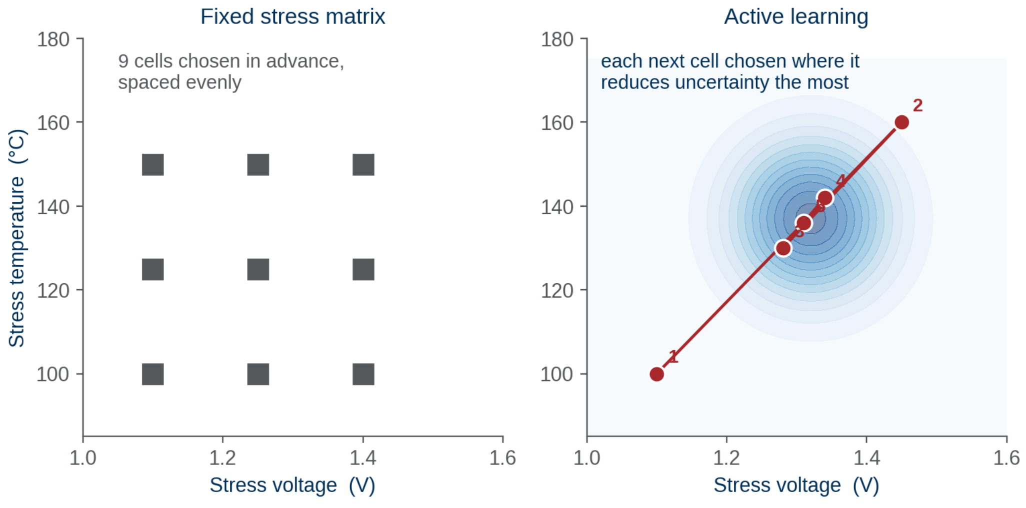
What a prior does here, in practice

Parameter	Where the prior comes from	Effect on the answer
Weibull shape β	Published values for the mechanism; previous generations of the same product	Prevents an absurd β from a handful of points
Activation energy E_a	JEP122 and the literature for the mechanism	Turns a fixed assumed value into a stated uncertainty
Voltage exponent n	Physics of the breakdown model	Constrains the extrapolation in voltage
Lot-to-lot variation	Historical qualification data across lots	Separates unit variation from lot variation

A fixed assumed activation energy is already a prior. It is a prior with zero variance, and nobody writes it down as an assumption.

- **Hierarchical models are the natural extension:** share strength across lots, across stress conditions, and across product generations.

Active learning: choose the stress conditions



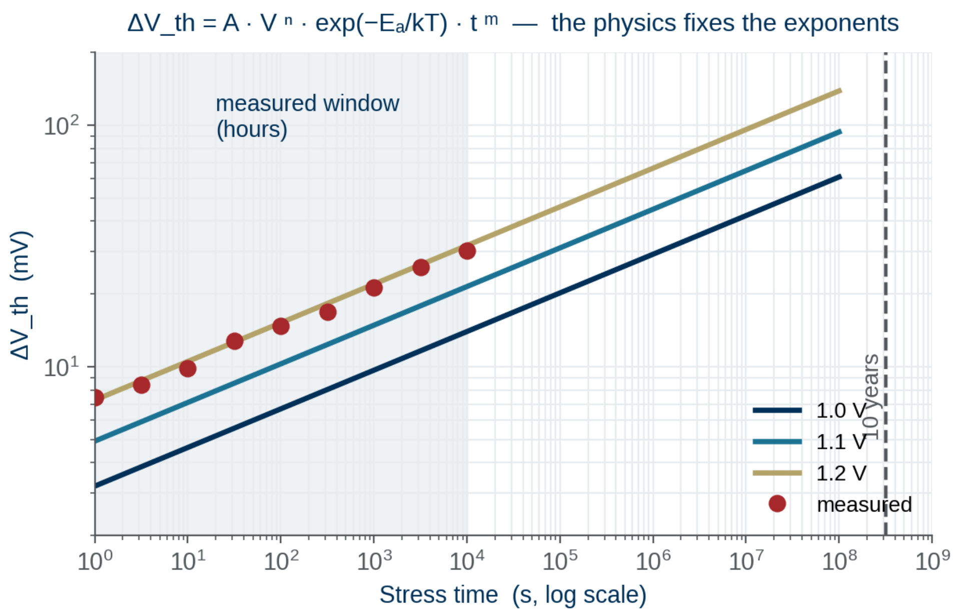
- **The fixed matrix spends units evenly,** including at conditions that turn out to be uninformative.
- **Active learning picks the next cell** where it most reduces uncertainty in the quantity you actually care about.

This is Bayesian optimal experiment design. The objective is information about the use-condition lifetime, not coverage of the stress space.

THE LOOP

1. **Fit**
   Update the posterior over model parameters with everything measured so far.
2. **Score**
   For each candidate stress condition, compute the expected reduction in uncertainty about life at use conditions.
3. **Choose**
   Run the condition with the best score, subject to the constraint that it must not activate a new mechanism.
4. **Repeat**
   Stop when the credible interval is narrow enough, or the units run out.

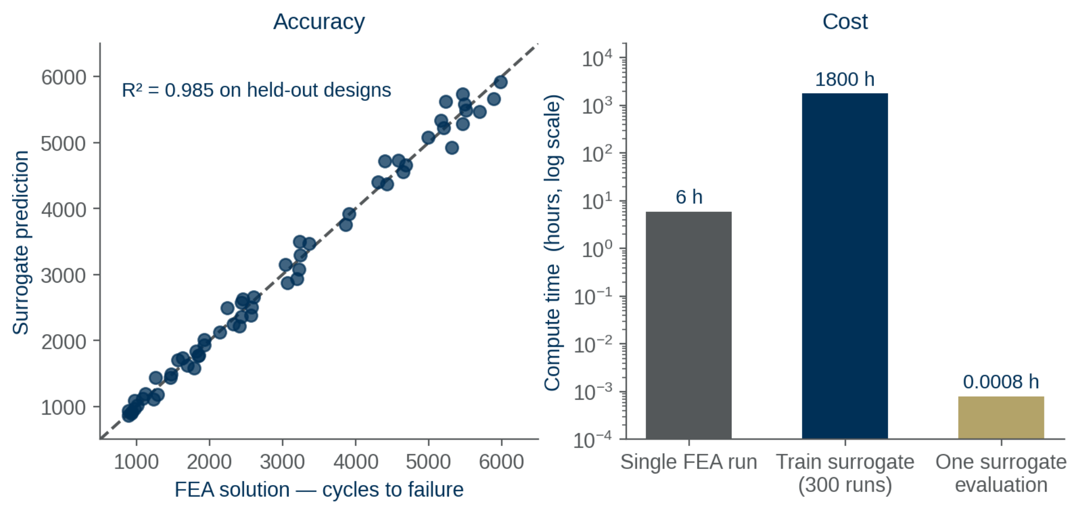
Physics-informed surrogates for degradation



- **The mechanisms have known functional forms.** Bias temperature instability follows a power law in time with an Arrhenius temperature term and a strong voltage dependence.
- **Fitting a free-form model to this data would fail.** There are too few points, and a flexible model extrapolates arbitrarily.
- **The physics-informed approach constrains the model** to obey the known form, and uses learning for what the form leaves open: coefficients, workload dependence, recovery, and device-to-device spread.

Mechanism	What it degrades
BTI	Threshold voltage drifts, circuits slow down
TDDB	Gate dielectric breaks down, hard short
Hot-carrier injection	Interface damage, drive current falls
Electromigration	Interconnect voids, line goes open

Virtual qualification: surrogates for package simulation



- **Package failures are mechanical.** Solder joints crack, underfill delaminates, and dies warp, because materials with different expansion coefficients are bonded together and cycled in temperature.
- **Finite element analysis predicts this well,** and one run takes hours. A design study needs thousands of runs.

THE SURROGATE ARGUMENT

Run FEA a few hundred times over the design space. Train a model to reproduce the output. Then explore with the surrogate.

The training cost is paid once. After that, evaluation is effectively free, and design space exploration becomes possible.

Callback

This is the same machinery as the thermal co-design homework. There the expensive function was a thermal solve and Bayesian optimisation searched the design space. Here the expensive function is a thermomechanical solve and the objective is cycles to failure.

The method transfers unchanged. Only the objective is new.

SEGMENT 3

A worked extrapolation, and how it goes wrong

Claiming ten years from a few weeks of data.

The setup

1. **The requirement**
   Under 1 % cumulative failures at ten years, at use conditions.

2. **The experiment**
   30 units. Elevated voltage and temperature. Stress until the equivalent of one year at use conditions.

3. **The observation**
   A handful of failures, all early, following a clean power law.

4. **The analysis**
   Fit a Weibull. Convert to use conditions with an Arrhenius factor. Extrapolate to ten years.

5. **The result**
   About 1 % failures at ten years. The part passes. The report is one page.

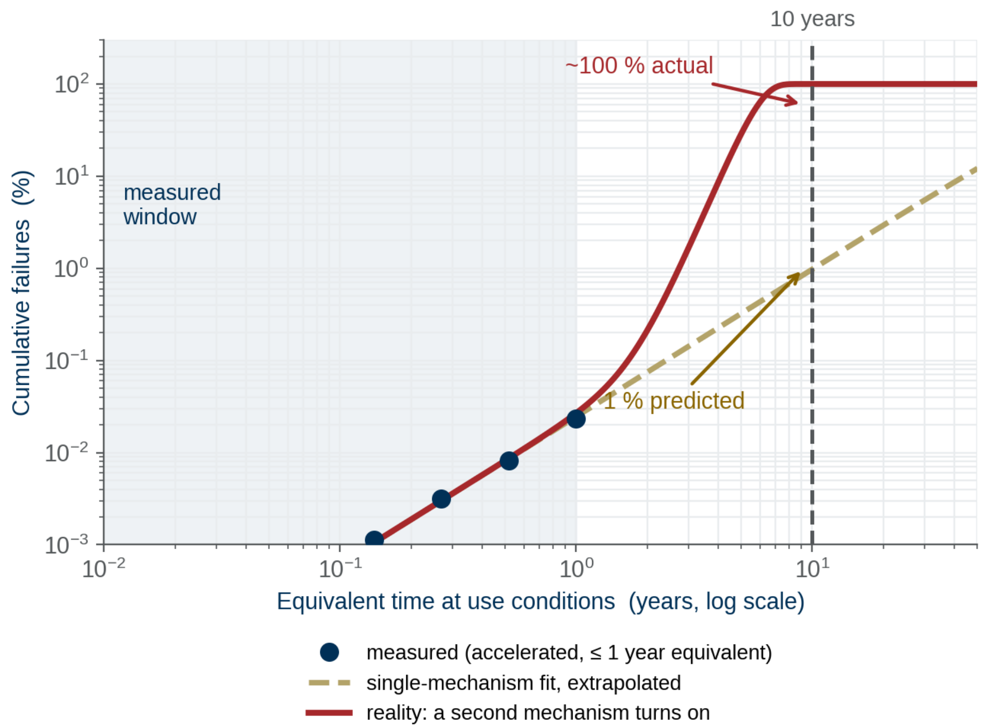
EVERY STEP IS STANDARD PRACTICE

Nothing above is careless. The sample size follows the standard. The distribution is the usual one. The acceleration model is the one the standard names for this mechanism. The fit is good.

The report will be accepted, and the conclusion may still be wrong by two orders of magnitude.

The next slide shows why.

What the data could not tell you



- **A second mechanism exists.** It has a much larger shape parameter, so it contributes nothing measurable early and then rises steeply.
- **Inside the measured window it is invisible.** The data is consistent with one mechanism, and the fit quality gives no warning.
- **Beyond the window it dominates.** At ten years the single-mechanism fit predicts about 1 %. The truth is close to 100 %.
- **The residuals looked fine the whole time.** Goodness of fit inside the data says nothing about the extrapolation.

The error is not statistical. It is structural. No amount of additional data inside the same window fixes it.

What actually defends against this

1. **Test at more than one stress level, and check the model holds across them**
   If one mechanism governs, the fitted parameters agree across stress levels. Disagreement is the signal that something else is happening.

2. **Fit mixtures, not single distributions**
   Let the model contain two competing mechanisms with a prior on each. The posterior then reports how much room the data leaves for a hidden second mechanism.

3. **Use physical failure analysis, not only statistics**
   Cross-section the failed parts. Knowing the mechanism physically is worth more than any amount of curve fitting.

4. **Report an interval, and report what it assumes**
   A number with a credible interval and a stated model assumption is auditable. A single number is not.

5. **Instrument the field and keep watching**
   The strongest defence is measuring the population in service, which is the next slide.

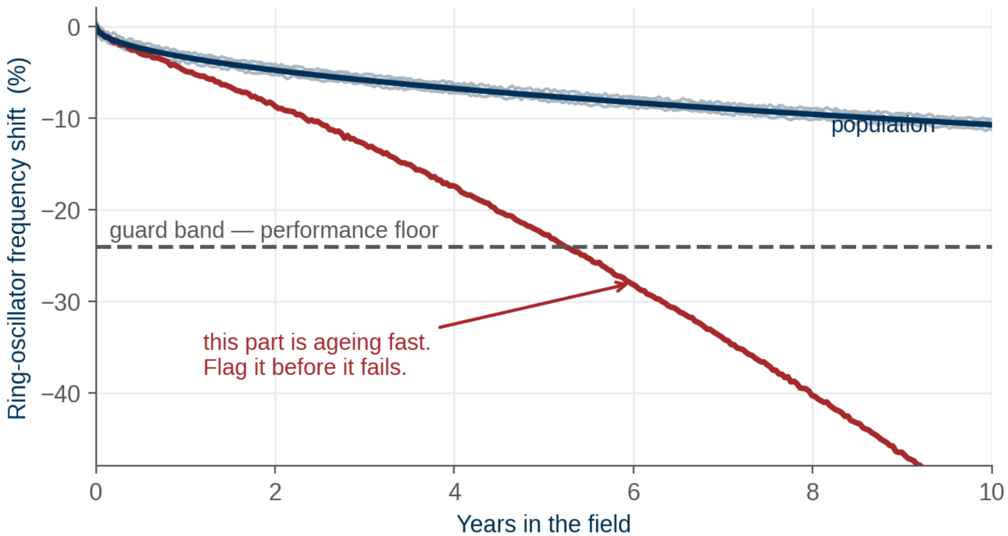
None of these are exotic. The reason they are not universal is that a qualification is scheduled as a gate to be passed, and every item on this list costs time.

SEGMENT 4

Silicon lifecycle management

Closing the loop from the field back to design.

On-die telemetry and in-field aging prediction



- **The monitors from Lecture 1 are still on the die,** and they are readable in the field. The same ring oscillators that screened the part now report how it is ageing.

WHAT THE LOOP LOOKS LIKE

1. **Measure**
   Read on-die monitors periodically over the part's service life.
2. **Compare**
   Track each part against the population and against the model's prediction.
3. **Predict**
   Estimate remaining useful life for that individual part.
4. **Act**
   Schedule maintenance, lower the clock, or retire the part before it fails.
5. **Feed back**
   Return real field degradation data to the next qualification and the next design.

The loop closes back to design

Design	Test	Qualify	Field	Next design
margin, guard bands, monitor placement	screen, adaptive routing	stress, models, intervals	telemetry, aging prediction	real degradation data as input

Each stage in this module produces data that the previous stage can learn from. The digital twin stops being a forward simulation and becomes a loop that is corrected by measurement.

- **Design margin is currently set by worst-case assumptions.** Field data replaces those assumptions with distributions, which usually means less margin and more performance.
- **Qualification is currently a gate.** Continuous measurement makes it an estimate that gets better over the product's life.

Summary

1. **Qualification is extrapolation from almost no data.**
   A few dozen units, a few weeks, often zero failures, and a ten-year claim at the end.

2. **Zero failures in 77 units is an upper bound at a stated confidence.**
   It contains no lifetime distribution, no shape parameter, and no margin.

3. **Acceleration models are small parametric forms whose constants enter exponentially.**
   A 0.2 eV disagreement in activation energy moves the answer by more than five times.

4. **The model choice dominates the extrapolation, and the data cannot settle it.**
   Weibull and lognormal fit the same data and differ by about 15× at the 1 ppm level.

5. **Bayesian methods and physics-informed models do not create data.**
   They make assumptions explicit, use prior knowledge efficiently, and report an honest interval.

6. **Field telemetry turns a one-time gate into a continuous measurement.**
   The population in service becomes the experiment, at real conditions, at real scale.

Two lectures, one argument

In both lectures the mathematics was standard. The hard part was writing down what the answer is worth, and how wrong it could be.

1

Lecture 1 asked for a cost function

Screening is easy to do and hard to justify. The escape–overkill curve says what is possible. Only an explicit price on each error says where to operate.

2

Lecture 2 asked for an interval

Fitting a lifetime model is easy. Stating how much of the answer came from data and how much came from assumption is the real work.

Both are habits rather than techniques, and both transfer to any problem where a model output drives a decision with an asymmetric cost.