Reliability HotWire
The eMagazine for the Reliability Professional
Reliability HotWire
Issue 21, November 2002

Hot Topics

The Bathtub Curve and Product Failure Behavior
Part One - The Bathtub Curve, Infant Mortality and Burn-in

by Dennis J. Wilkins
Retired Hewlett-Packard Senior Reliability Specialist, currently a ReliaSoft Reliability Field Consultant
This paper is adapted with permission from work done while at Hewlett-Packard.

Reliability specialists often describe the lifetime of a population of products using a graphical representation called the bathtub curve. The bathtub curve consists of three periods: an infant mortality period with a decreasing failure rate followed by a normal life period (also known as "useful life") with a low, relatively constant failure rate and concluding with a wear-out period that exhibits an increasing failure rate. This article provides an overview of how infant mortality, normal life failures and wear-out modes combine to create the overall product failure distributions. It describes methods to reduce failures at each stage of product life and shows how burn-in, when appropriate, can significantly reduce operational failure rate by screening out infant mortality failures. The material will be presented in two parts. Part One (presented in this issue) introduces the bathtub curve and covers infant mortality and burn-in. Part Two (presented in next month's *HotWire*) will address the remaining two periods of the bathtub curve: normal life failures and end of life wear-out.

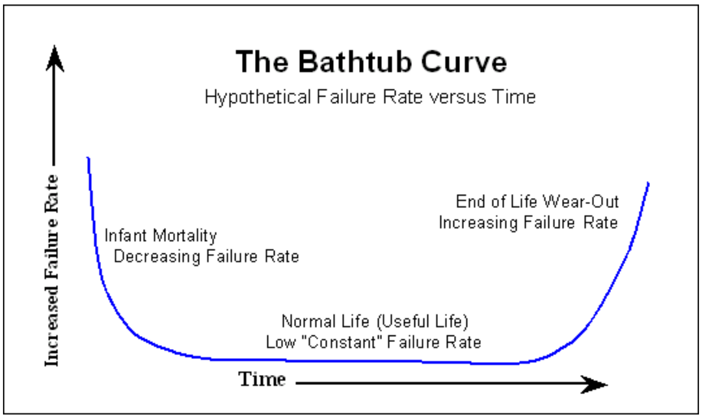

Figure 1: The Bathtub Curve

The bathtub curve, displayed in Figure 1 above, does *not* depict the failure rate of a single item, but describes the relative failure rate of an entire population of products over time. Some individual units will fail relatively early (infant mortality failures), others (we hope most) will last until wear-out, and some will fail during the relatively long period typically called normal life. Failures during infant mortality are *highly undesirable* and are always caused by defects and blunders: material defects, design blunders, errors in assembly, etc. Normal life failures are normally considered to be random cases of "stress exceeding strength." However, as we'll see, many failures often considered normal life failures are actually infant mortality failures. Wear-out is a fact of life due to fatigue or depletion of materials (such as lubrication depletion in bearings). A product's useful life is limited by its shortest-lived component. A product manufacturer must assure that all specified materials are adequate to function through the intended product life.

Note that the bathtub curve is typically used as a visual model to illustrate the three key periods of product failure and not calibrated to depict a graph of the expected behavior for a particular product family. It is rare to have enough short-term and long-term failure information to actually model a population of products with a calibrated bathtub curve.

Also note that the actual time periods for these three characteristic failure distributions can vary greatly. Infant mortality does not mean "products that fail within 90 days" or any other defined time period. Infant mortality is the time over which the failure rate of a product is decreasing, and may last for years. Conversely, wear-out will not always happen long after the expected product life. It is a period when the failure rate is increasing, and has been observed in products after just a few months of use. This, of course, is a disaster from a warranty standpoint!

We are interested in the characteristics illustrated by the entire bathtub curve. The infant mortality period is a time when the failure rate is dropping, but is undesirable because a significant number of failures occur in a short time, causing early customer dissatisfaction and warranty expense. Theoretically, the failures during normal life occur at random but with a relatively constant rate when measured over a long period of time. Because these failures may incur warranty expense or create service support costs, we want the bottom of the bathtub to be as low as possible. And we don't want any wear-out failures to occur during the expected useful lifetime of the product.

Infant Mortality What Causes It and What to Do About It?
From a customer satisfaction viewpoint, infant mortalities are unacceptable. They cause "dead-on-arrival" products and undermine customer confidence. They are caused by defects designed into or built into a product. Therefore, to avoid infant mortalities, the product manufacturer must determine methods to eliminate the defects. Appropriate specifications, adequate design tolerance and sufficient component derating

can help, and should always be used, but even the best design intent can fail to cover all possible interactions of components in operation. In addition to the best design approaches, stress testing should be started at the earliest development phases and used to evaluate design weaknesses and uncover specific assembly and materials problems. Tests like these are called HALT (Highly Accelerated Life Test) or HAST (Highly Accelerated Stress Test) and should be applied, with increasing stress levels as needed, until failures are precipitated. The failures should be investigated and design improvements should be made to improve product robustness. Such an approach can help to eliminate design and material defects that would otherwise show up with product failures in the field.

After manufacturing of a product begins, a stress test can still be valuable. There are two distinct uses for stress testing in production. One purpose (often called HASA, Highly Accelerated Stress Audit) is to identify defects caused by assembly or material variations that can lead to failure and to take action to remove the root causes of these defects. The other purpose (often called burn-in) is to use stress tests as an ongoing 100% screen to weed out defects in a product where the root causes cannot be eliminated.

The first approach, eliminating root causes, is generally the best approach and can significantly reduce infant mortalities. It is usually most cost-effective to run 100% stress screens only for early production, then reduce the screen to an audit (or entirely eliminate it) as root causes are identified, the process/design is corrected and significant problems are removed. Unfortunately, some companies put 100% burn-in processes in place and keep using them, addressing the symptoms rather than identifying the root causes. They just keep scrapping and/or reworking the same defects over and over. For most products, this is not effective from a cost standpoint or from a reliability improvement standpoint.

There is a class of products where ongoing 100% burn-in has proven to be effective. This is with technology that is "state-of-the-art," such as leading edge semiconductor chips. There are bulk defects in silicon and minute fabrication variances that cannot be designed out with the current state of technology. These defects can cause some parts to fail very early relative to the majority of the population. Burn-in can be an effective way to screen out these weak parts. This will be addressed later in this article.

A Quantitative Look at Infant Mortality Failures Using the Weibull Distribution
The Weibull distribution is a very flexible life distribution model that can be used to characterize failure distributions in all three phases of the bathtub curve. The basic Weibull distribution has two parameters, a shape parameter, often termed beta (

β), and a scale parameter, often termed eta (η). The scale parameter, eta, determines when, in time, a given portion of the population will fail, *i.e.* 63.2%. The shape parameter, beta, is the key feature of the Weibull distribution that enables it to be applied to any phase of the bathtub curve. A beta less than 1 models a failure rate that decreases with time, as in the infant mortality period. A beta equal to 1 models a constant failure rate, as in the normal life period. And a beta greater than 1 models an increasing failure rate, as during wear-out. There are several ways to view this distribution, including probability plots, survival plots and failure rate versus time plots. The bathtub curve is a failure rate vs. time plot.

Typical infant mortality distributions for state-of-the-art semiconductor chips follow a Weibull model with a beta in the range of 0.2 to 0.6. If such a distribution is viewed in terms of failure rate versus time, it looks like the plot in Figure 2.

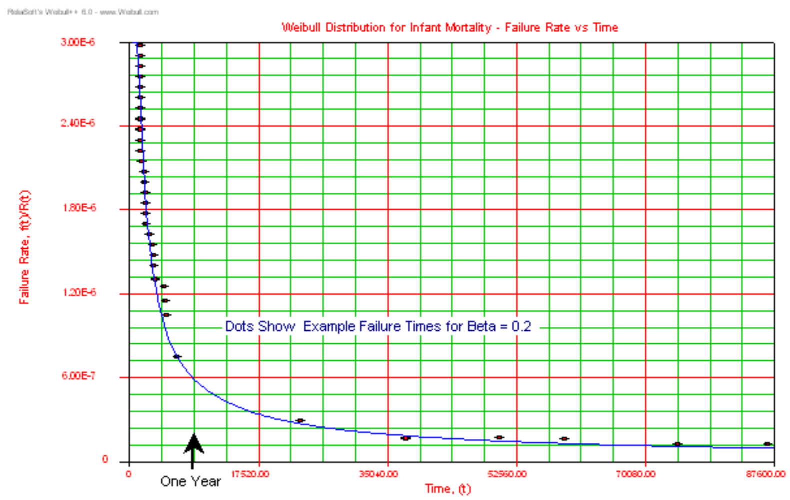

Figure 2: Infant Mortality Curve - Failure Rate vs. Time

This plot shows ten years (87,600 hours) of time on the x-axis with failure rate on the y-axis. It looks a lot like the infant mortality and normal life portions of the bathtub curve in Figure 1, but this curve models only infant mortality (decreasing failure rate). Dots on this plot represent failure times typical of an infant mortality with Weibull beta = 0.2. As you can see, there are 27 failures before one year, and only 6 failures from one to ten years. People observing this curve, and the failure points plotted, could not be blamed for thinking it represents both infant mortality failures (in the first year or so), and normal life failures after that. But these are only infant mortality failures - all the way out to ten years!

This plot shows the distribution for a beta value typical of complex, high-density integrated circuits (VLSI or Very Large Scale Integrated circuits). Parts such as CPUs, interface controller and video processing chips often exhibit this kind of failure distribution over time. A look at this plot shows that if you could run these parts for the equivalent of three years and discard the failed parts, the reliability of the surviving parts would be much higher out to ten years. In fact, until a wear-out mode occurs, the reliability would continue to improve over time. If there are mechanisms that can produce normal life failures (theoretically a constant failure rate) mixed in with the defects that cause the infant mortalities shown above, burn-in can still provide significant improvement as long as the constant failure rate is relatively low.

Burn-In for Leading Edge Technologies
To see how burn-in can improve the reliability of high tech parts, we'll use a chart that looks somewhat like the failure rate vs time curve in Figure 2, but is more useful. This is a survival plot that directly shows how many units from a population have survived to a given time. Figure 3 is a plot for a typical VLSI process with a small "weak" sub-population (defective parts that will fail as infant mortalities) and a larger sub-population of parts that will fail randomly at a very low rate over the normal operating life. The x-axis scale is in years of use (zero to 100 years!) and the y-axis is percent of parts still operating to spec (starting at 100% and dropping to 50%).

Figure 3 shows that, of the failures that occur in the first 20 years (about 4%), most failures occur in the first year or so, just like we observed in the infant mortality example above. Because there is a low level, constant failure rate, this plot shows failures continuing for a hundred years. Of course, there could be a wear-out mode that comes into play before a hundred years has elapsed, but no wear-out distribution is considered here. Electronic components, unlike mechanical assemblies, rarely have wear-out mechanisms that are significant before many decades of operation.

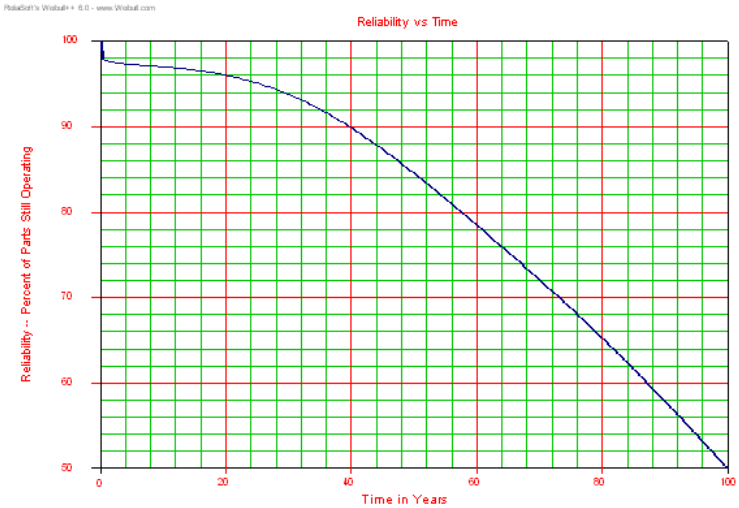

Figure 3: Mixed Infant Mortality and Normal Life Survival Plot

We're not really interested in the failures much beyond ten years, so let's look at this same model for only the first ten years. In Figure 4, we have included sample failure points from the simulation model used to create the plot. These enable us to view which population (infant mortality or normal life) the failure came from.

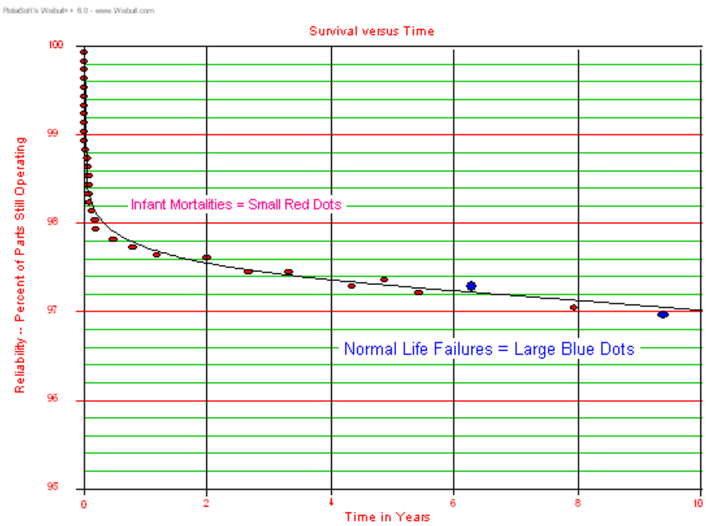

Figure 4: Mixed Infant Mortality and Normal Life Failures

We see that the plot in Figure 4 looks like the early life and normal life portions of the bathtub curve, and in fact includes both distributions. We see that over 2% of the units fail in the first year, but it takes ten years for 3% to fail. In actuality, there are still "infant" mortalities occurring well beyond ten years in this model, but at an ever-decreasing rate. In fact, in the ten year span of this model there would be very few normal life failures. Only two failures (~5% of all failures) in this example (large blue dots) come from the normal life failure population. About 95% of the failures plotted above (small red dots) are infant mortality failures! This is what the integrated circuits (IC) industry has observed with complex solid-state devices. Even after ten years of operation the primary failure cause for ICs is still infant mortality. In other words, failures are still driven primarily by defects.

In such cases, burn-in can help. In the plot above you can see that if you could get three years of operation on this part before you shipped it, you would have screened out over 80% (2% divided by 3%) of the parts that would fail in ten years. So if we were to come up with a method to effectively "age" the parts the equivalent of three years and eliminate most of the infant mortalities, the remaining parts would be more reliable than the original population. Of course, the parts that go through the three-year "burn-in" would have to last an additional ten years in the field, for a total of thirteen years. Let's see what this looks like in Figure 5.

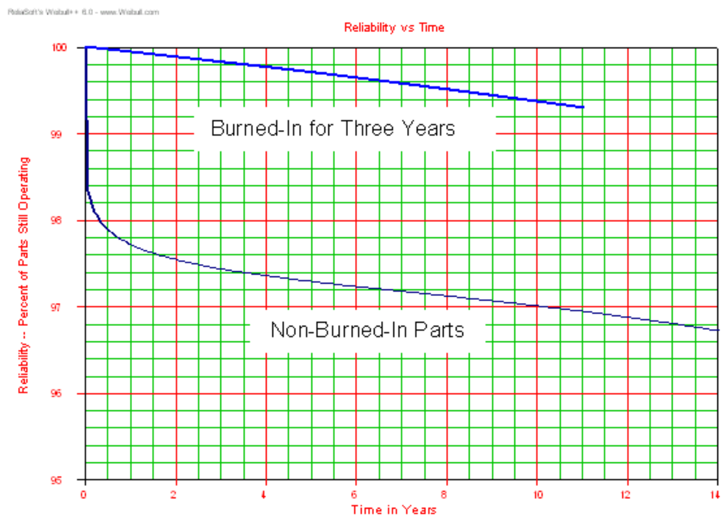

Figure 5: Comparison of Failures from Raw and Burned-in Parts

Above, we see fourteen years of failure distribution for the original parts (not burned-in) and eleven years of expected failure distribution for parts that received three years of burn-in. In this example, the total cumulative failures between three years and thirteen years for the original parts (or from zero to ten years for burned-in parts) is about 0.6%. Without burn-in, the first ten years would have had about 3% cumulative failures. This is about a five times reduction in cumulative failures by using burn-in, or in terms of a change, we would have about 2% fewer cumulative failures in ten years with burn-in if a dominant infant mortality failure mode exists. Note that in the first year or two, the relative improvement in reliability is even greater. At two years, only about 0.1% failures are expected after burn-in but almost 2% without burn-in; a ratio of almost 25:1!

In reality, manufacturers don't have two to three years to spend on burn-in. They need an accelerated stress test. In the IC industry there are usually two stresses that are used to accelerate the effective time of burn-in: temperature and voltage. Increased temperature (relative to normal operating temperatures) can provide an acceleration of tens of times (10x to 30x is typical). Increased voltages (relative to normal operating levels) can provide even higher acceleration factors on many types of ICs. Combined acceleration factors in the range of 1000:1, or more, are typical for many IC burn-in processes. Therefore, burn-in times of tens of hours can provide effective operating times of one to five years, significantly reducing the proportion of parts with infant mortality defects.

What if we try burn-in on a product with no dominant infant mortality problems? The survival plot for an assembly with a 1% per year "constant" failure rate (normal life period) is shown below in Figure 6.

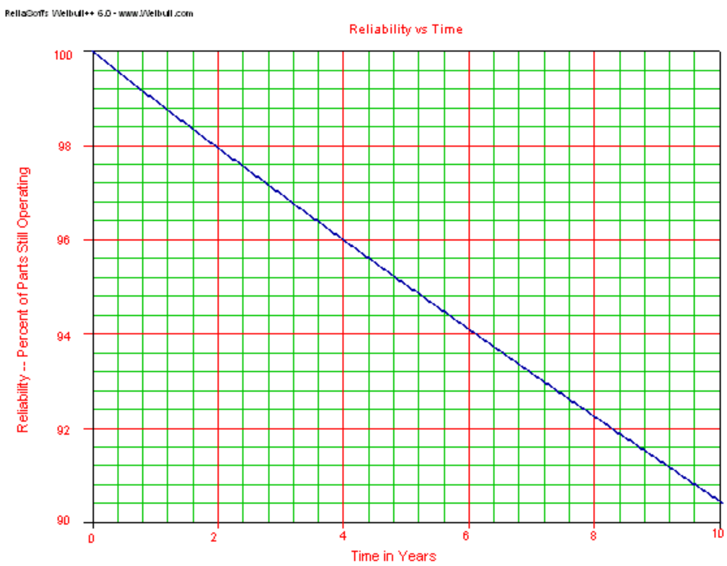

Figure 6: Survival Plot for Constant Failure Rate

It's pretty easy to see that burn-in for two years would find ~2% failures, but operation for an additional two years would find another ~2%. At ten years, we would have found about 10%. Note, the line is not really a straight line because a constant failure rate (equivalent to the normal life part of the bathtub) acts on the remaining population and the remaining population is decreasing as units fail. Looking at the same burn-in conditions as in the last example, if we were to provide three years of operation on these parts and then use them for an additional ten years, what results would we have? The cumulative failures of the units that passed this screen would be very close to 9.5%. Without burn-in, the cumulative failures in ten years would be the same, about 9.5%. There is no advantage to burn-in with a constant (normal life) failure rate.

It should be obvious that burn-in of an assembly that is failing due to a wear-out failure mode (failure rate increasing with time) will actually yield assemblies that are worse than units that did not go through burn-in. This is simply because the probability of failure is increasing for every hour the parts run. Adding operating time simply increases the possibility of a failure in any future period of time!

Conclusion

In this issue, Part One, we have introduced the concept of the bathtub curve and discussed issues related to the first period, infant mortality, as well as the practices, such as burn-in, that are used to address failures of this type. As this article demonstrates, although burn-in practices are not usually a practical economic method of reducing infant mortality failures, burn-in has proven effective for state-of-the-art semiconductors where root cause defects cannot be eliminated. For most products, stress testing, such as HALT/HAST should be used during design and early production phases to precipitate failures, followed by analysis of the resulting failures and corrective action through redesign to eliminate the root causes. In Part Two (presented in next month's *HotWire*), we will examine the final two periods of the bathtub curve: normal life failures and end of life wear-out.