A Conceptual Model for "Inherent Reliability" for Nuclear W SAND2009-4720C

Rene L. Bierbaum, Sandia National Laboratories

Key Words: bathtub curve, inherent reliability one-shot devices, test and evaluation

SUMMARY & CONCLUSIONS

Many people, when thinking about different stages of a particular device's life vis-à-vis defectiveness, use the notion of the "bathtub curve" as a model. However this model is not fully applicable for nuclear weapons (and one-shot devices in general). A new model will be proposed that includes two regimes: birth defect dominated and time-dependent dominated. A short discussion of why a bathtub curve might mistakenly be inferred is included. Finally, the relationship between inherent and estimated reliability will be described in the context of this model.

1 NUCLEAR WEAPON RELIABILITY

Nuclear weapons are generally considered as one-shot devices, even though they are partially composed of subsystems capable of multiple operations. This is because their operational time (on the order of seconds) is very short compared to the time they spend in dormant storage (decades). Hence performance is quantified in terms of failure probability – what is the probability that a weapon will fail to achieve the specified nuclear output if functioned? This probability can be thought of in a very general sense as the percentage of failures one would expect to observe in a given number of weapons operated. This is contrasted with the usual characterization of performance for continuously operating devices in terms of failure rate (the number of observed failures divided by the operating time). Failure probability, as used for nuclear weapons, is unit-less and in particular does not include the unit of time, as does failure rate. Failure rate for continuously operating systems can be used to calculate a failure probability by integrating over some time period. This in fact gives meaningful information about a population. However trying to use failure probability for one-shot devices to calculate a failure rate is much less useful since there is no "operating time" per se associated with them. For the purpose of evaluation, they either work or they don't when selected.

2 THE BATHTUB CURVE MODEL

Many people, when thinking about different stages of a particular device's life vis-à-vis defectiveness, use the notion of the "bathtub curve" as a model. This model in fact is used quite pervasively to think about product defectiveness. As will be discussed below, this model is not very apt for nuclear weapons (and one-shot devices in general) but it is useful to review it first and then describe its shortcomings with respect to weapons.

A useful discussion of the bathtub curve model is provided below by Wilkins [1], and Figure 1 is extracted from the same article:

"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."

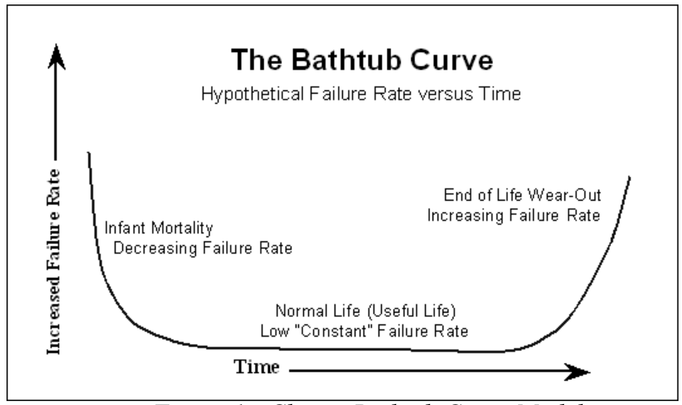

Figure 1 - Classic Bathtub Curve Model

Note that the bathtub curve model is specifically intended to depict the behavior of continuously operating systems – the y-axis is failure rate not failure probability and the x-axis is operating time, not "calendar time" as one would use for nuclear weapons in the stockpile. As a depiction of failure rate, the bathtub curve only applies to those items that are "alive" or working at a given time. That is, if the failure rate is λ at time t, then given that a unit is alive and operating at time t the probability of failing or "dying" in the next time period Δt is approximated by $\lambda \cdot \Delta t$. Therefore, a constant failure rate implies an increasing failure probability function, since the probability of having failed by a certain time is an accumulation of the chances of having failed at any of the previous times.

In addition, the terms "infant mortality" and "wear-out" refer to failures that are experienced during (and as a result of) operation and thus neither of these concepts are germane to one-shot devices that spend most of their lives in dormant storage. Finally, systems that are continuously operating allow for (nearly) immediate detection and removal of defective units when they occur. This is most assuredly not the case for weapons, where defects may be present in the stockpile for long periods of time degrading reliability but not being known. Stated another way, weapons with birth defects do not get removed from the stockpile unless we do sampling and testing to find them and then take action, whereas continuously operating systems with infant mortality defects remove themselves from the population by failing during operation early on. We can never infer from nuclear weapon testing when a defect has occurred – only that it occurred sometime between the present test and any prior applicable test. In short, while the bathtub curve is a commonly used and easily understood model, it is not particularly applicable to nuclear weapons. As a side note, if a continuously operating system was not operated continuously or was stored for a while before being used, the resulting failure rate plotted as a function of calendar time would not look like the bathtub curve either.

3 A NEW MODEL FOR WEAPON RELIABILITY

A more suitable model for one-shot device reliability is proposed in this section. This model has served to lay the foundation for a new test and evaluation program for nuclear weapons that is tailored for the unique stages of a weapon's life. The new model has two regimes. While to a modest degree there is an analogy to the bathtub curve model, the new model is indeed different.

Understanding this model requires one to differentiate between inherent reliability (which "is what it is" but is never known) and estimated reliability (the estimate of inherent reliability based upon knowledge gained through experience, testing, and analysis). In all of the examples shown below, inherent reliability (or failure probability) is shown as a dashed line and estimated reliability (or failure probability) is shown as a dotted line. For clarity, failure probability is used in each of the following figures. The y-axis is labeled as "Average System Failure Probability" and represents a calculation of the average failure probability for a population of weapons that share the same design (referred to as a "weapon system"). The x-axis represents "calendar time" and would generally be in units of years, with the entire length of the x-axis representing the total lifetime of a weapon system.

The model of inherent reliability that will be used is shown in Figure 2.

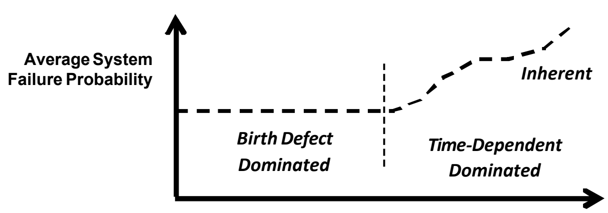

Figure 2 - Notional Model of Inherent Reliability

There are two discrete regimes identified in the figure: Birth Defect Dominated and Time-Dependent Dominated.

3.1 Birth Defect Dominated

The first regime begins at production, and it is characterized by the presence of defects that are in the weapons when they enter the stockpile. These defects may be due to design errors, production or assembly problems, or material flaws that were not detected during product acceptance testing. Thus these defects are in the stockpile from Day 1 and will remain there unless they are fixed – and of course, fixing them requires them to be found first. That, in turn, relies on sampling and testing. This initial regime will be called "Birth Defect Dominated". With regard to failure rate in the first regime the assumption is that, given that a unit is working (reliable) at time t, it remains reliable for the foreseeable future. Note that a failure rate at time t only applies to units that are alive or reliable at time t, and does not apply to those units with birth defects and are not reliable at time t. This effectively means that the failure rate for reliable units is, for all practical purposes, zero in the first regime.

3.2 Time-Dependent Dominated

The second regime of weapon life begins when there is any onset of time-dependent defects affecting reliability. Obviously one does not know when this transition point occurs. In the past, there have been time-dependent problems observed through testing that have affected reliability very soon after production. In general though, weapon systems have not experienced many time-dependent failures during their lifetime. To some extent, onset of this regime is little more than a hypothesis at this point because there have been few time-dependent issues that have arisen that affect performance (ability to get nuclear output when functioned). On the other hand, investigations have revealed changes in either materials or parameters which indicate that time-dependent behavior is afoot, even if it doesn't yet affect performance.

Note that this regime is not the same as "wear-out", as used in the bathtub curve model. Defects are not due to long-term use and wear, but rather arise during dormant storage as materials change with age. Material properties may eventually shift to the extent that performance is impacted. As with birth defects, this will not be detectable without on-going evaluation.

The new test and evaluation program is intended to capitalize on this model. First, it is recognized that there must be a deliberate, carefully-planned effort made early in the life of a weapon to detect the unknown birth defects that are present at Day 1. Without such an effort, these defects will remain unknown. Eventually though, there is a point of diminishing returns. Smaller and smaller birth defects will continue to be detected with further testing, but their impact is so small that the costs outweigh the benefits. At this point, the major concern becomes time-dependent behavior that may arise to jeopardize performance. Thus the focus shifts from a broad search for unknown birth defects across the entire state space of conditions to monitoring specific high-risk issues (parameters with small margin, known or suspected trends, material vulnerabilities, etc.).

4 CHANGES IN INHERENT RELIABILITY

Although as noted earlier, inherent reliability "is what it is", there are things that can cause it to go up or down. Figure 3 shows the ways in which inherent reliability can change. As before, the y-axis is average system failure probability and the x-axis is weapon system lifetime in years.

Example A shows that the inherent failure probability can decrease if an existing problem is corrected. This change is shown as gradual slope downward, where the transition between failure probability values extends over the duration of the repair operation on the population of weapons. Example B, which unfortunately does occur, shows a case where an action results in the introduction of a new problem which increases failure probability; this could occur during the course of adding of a new capability, for example. As with Example A, the change in failure probability is gradual and occurs over the duration of the incorporation into the stockpile. The final graph (C) shows an example of a time-dependent issue that afflicts increasingly large segments of the stockpile over time. These changes can be due to stockpile aging or aggregated exposure time to a particular accelerating environment such as the field storage environment.

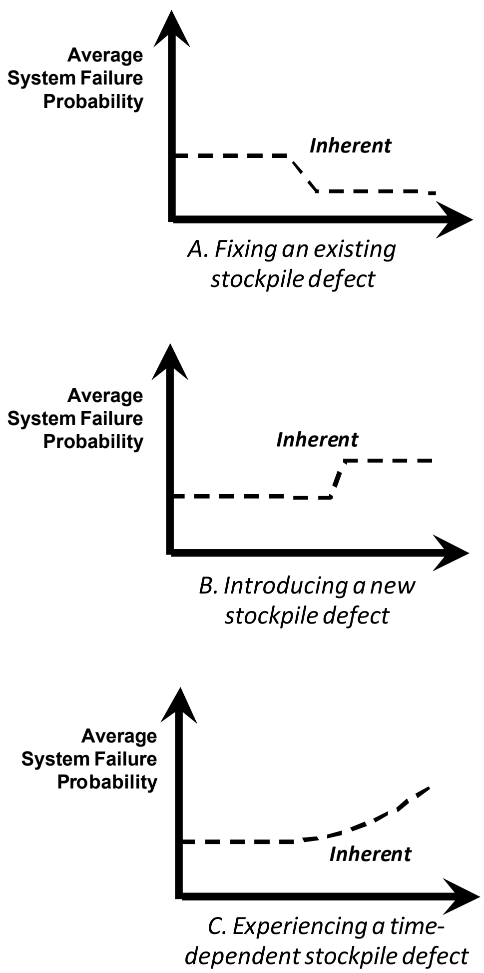

Figure 3 - How Inherent Reliability Can Change

Note that there are many ways in which a time-dependent problem can manifest itself, depending upon (1) the speed at which the mechanism evolves and (2) the degree of homogeneity in how it evolves (i.e., does the change take place in all units of the stockpile at an equal rate, or does it evolve at different rates for different units – or perhaps not at all for some units?). This is shown by the multiple curves in Figure 4 that illustrate different types of behavior. They are offset on the failure probability axis for clarity.

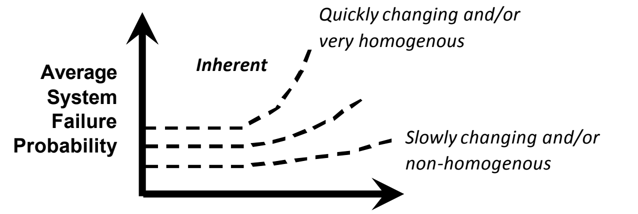

Figure 4 - Examples of Time-Dependent Failure Probabilities

5 RELIABILITY ESTIMATION AND THE NEW MODEL

A reliability estimate is an attempt to quantify the inherent reliability based upon all available knowledge. This estimate is used to make key operational and programmatic decisions. Hence it is also helpful to review some examples of the relationship between inherent and estimated reliability during the course of a weapon's life. These are shown notionally in Figure 5.

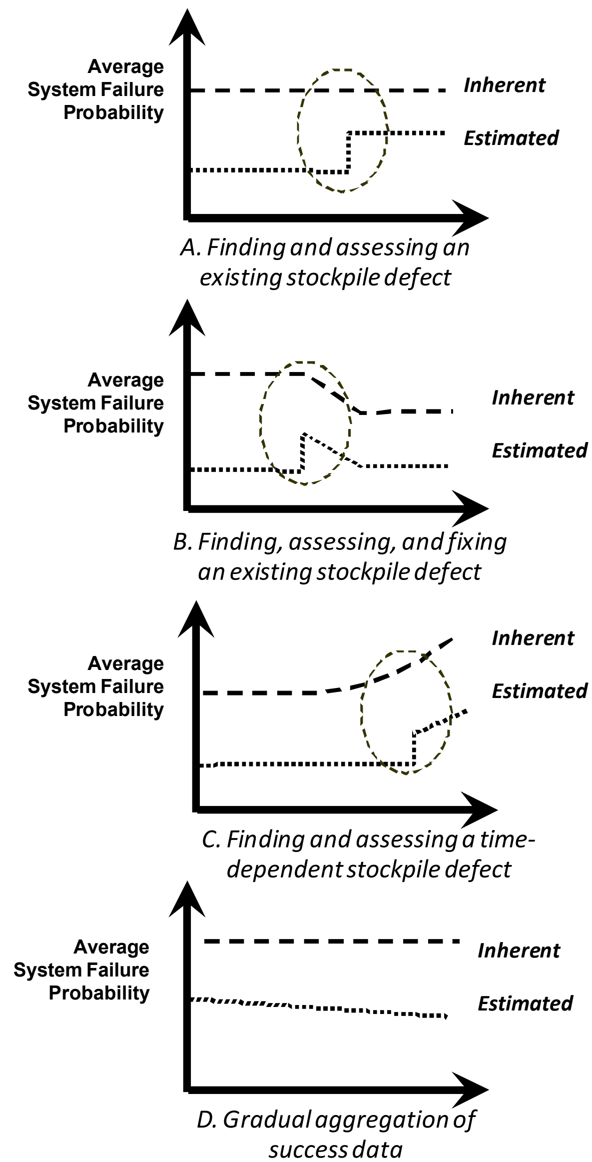

Figure 5 - Relationship of Inherent and Estimated Reliability

Note that for each example in Figure 5, the inherent failure probability is shown as being larger than the estimated failure probability. However this will not necessarily always be the case – estimated failure probability could also be larger than the inherent failure probability or the lines could cross during the course of a corrective action. In practice, one sometimes finds that the initial calculated impact due to detecting a defect is based upon limited data. Thus a failure probability estimate may initially be large and then decrease over time if more success data are collected. This may result in estimated failure probability being higher than inherent failure probability. On the other hand, undetected defects may result in the estimated failure probability being lower than the inherent failure probability.

The goal of course is to have estimated reliability converge over time to inherent reliability, and every test observation must be considered carefully to understand its impact on the estimate. Issues such as sample quantities, sampling philosophy, and test quality and diversity do not directly enter into calculation of the estimate. However these issues are quite critical in that they affect the ability to detect previously unknown defectiveness. Generally undetected defects cannot be credibly assessed and included in the estimate.

In summary, estimated failure probability changes when (1) defects are found through sampling and testing indicating a higher failure probability than currently estimated, (2) success data are aggregated over time indicating that a lower failure probability than currently estimated, and (3) defects are being fixed. Here as well, sample testing allows one to make better estimates as test results are aggregated.

The lifetime profile for a particular weapon type can be quite complex since there are so many combinations of changing inherent reliability and changing estimated reliability. A notional example is shown in Figure 6; again, inherent failure probability is shown as being larger than the estimate for clarity, but these lines can be in inverse position or cross multiple times during the life of a weapon.

As indicated earlier, the intent is to have the estimated reliability line eventually coincide over time with the inherent reliability, based upon the knowledge gained by sampling and testing (both detection of unknown defects as well as aggregation of success data). Unknown defects in the stockpile are the greatest risk that these lines will not coincide, and the point of the sample testing and analysis programs is to reduce the gap between the two failure probability lines to as small as possible (given risk and cost considerations). Note that success data has an important role here too in helping the inherent and estimated reliability converge, although emphasis tends to be on tests that reveal defects.

6 INCORRECT INFERENCE OF THE BATHTUB CURVE

Historical evaluation results have often been used to infer that nuclear weapons follow the bathtub curve model. This is not an unreasonable inference to make, given the data, but it can be explained more properly by the new conceptual model for inherent reliability.

There are two major artifacts that cause people to assume that a bathtub curve model is appropriate for nuclear weapons. The first is that the defect detection rate (i.e., the number of defects detected per unit time) for a particular weapon system is generally higher early in its lifetime than later. [As a side note, it should be pointed out that defect detection rate for nuclear weapons is unrelated to reliability impact. Some defects are determined to have a small failure probability impact and some a large one, and thus defect detection rate does not have a direct relationship to reliability.] The second artifact, which does relate to reliability, is that the impacts of newly identified defects for a particular weapon tend to be higher earlier in stockpile life.

In each of these cases, use of the bathtub curve model to explain the results is flawed. For the first case, it is vital to note that the quantity of tests performed over a weapon's lifetime is heavily front-loaded by intent. Because more units are being tested, the expectation is that more birth defects will be found. This is not due to a changing product as the bathtub curve implies (i.e., a changing failure rate over time) – instead, it is due to the profile of the test program, where testing more units gives a higher likelihood of finding birth defects present in the stockpile from Day 1. In short, it is incorrect to infer that the product is changing based upon this pattern alone. If there were no evaluation program to find the defects (or no fixes once the defects were found), then the inherent reliability curve would be flat until the onset of aging problems. Similarly, if we were to test fewer units, it would take longer to find these defects and the hump of the curve would move to the right.

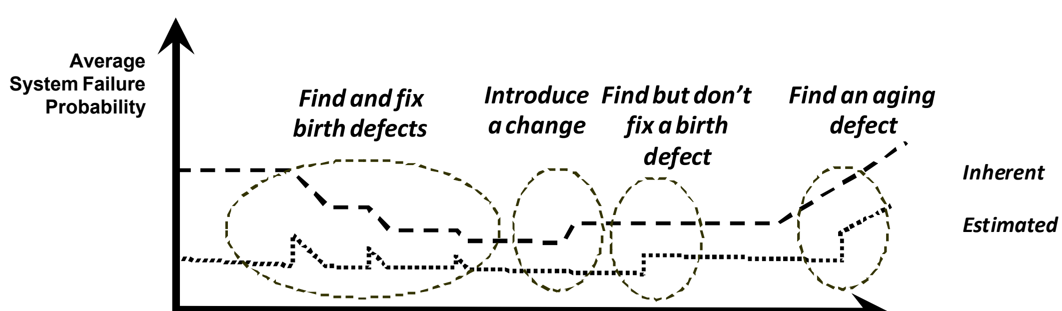

Figure 6 - Comparison of Inherent and Estimated Reliability

For the second case above, there is also a better technical explanation for this pattern. Larger defects, in general, take fewer tests to detect than smaller defects. For example, on average it takes 100 tests to find a 1% defect (assuming it is detectable in the test) whereas it takes 1000 tests on average to find a 0.1% defect. Thus we would expect a pattern of decreasing defect impact for newly identified defects as time goes by and more tests are aggregated. Note that the overall defect detection rate (new defects as well as already identified defects) may not decrease over time at all, if action is not taken to remove identified defects from the stockpile.

One other pertinent historical observation is that very large birth defects have sometimes been discovered relatively late in a weapon's life, which seems to contradict the new model. However, analysis of historical defects suggests that these cases have been due to the following causes:

1. The test program that was being done was completely incapable of detecting a particular birth defect. Once a new test regime was introduced later in life that allowed for detection, the defect was then discovered.
2. The defect was manifested under a narrow range of conditions that were examined very infrequently in the test program. Thus the time to detection was greatly extended.
3. The birth defect, while severe, only affected a subpopulation.
4. A new birth defect was introduced through an upgrade in capability conducted later in the life of a weapon system.

7 CONCLUSION

In summary, the bathtub curve loosely appears to explain the test history for nuclear weapons but is an incorrect model to assume. The test history instead is chiefly a manifestation of the way that tests are spread over time as well as the simple statistical principle that larger defects will take fewer tests to detect than smaller defects. The traditional bathtub curve presents a failure rate as a function of operating time, t, and is meant to characterize the chances of a unit "dying" or becoming unreliable in the immediate near future from time t given that the unit was, in fact, "alive" or reliably operating at time t. As such, it does not address the observations resulting from sample testing that reveals underlying "deaths", all of which would have occurred in the past.

This new model may have applicability for other one-shot devices and systems that share similar characteristics. Key consequences of this are as follows:

1. It will almost certainly require explicit action to achieve convergence of the estimated reliability with the inherent reliability, through sampling, testing, and analysis.
2. Interpretation of defect detection history for one-shot devices must be done carefully with consideration of the underlying evaluation program. Unsurprisingly, more tests will typically yield more defects, so some attempt to normalize with respect to test quantities may be helpful.
3. One must generally choose to take action to improve reliability for one-shot devices when defects are found; this is different for the case of continuously operating systems where failed units essentially remove themselves from the population. As noted before, this makes it essential to conduct testing on samples to identify defects present in the stockpile such that they can be addressed if appropriate. This continues to be important as the systems age and potentially enter the regime of time-dependent change.

ACKNOWLEDGMENTS

The author is grateful to Floyd Spencer, Janet Sjulin, Robert Monson, Kathleen Diegert, and Rena Zurn for their insightful comments on this paper.

REFERENCES

1. Wilkins, Dennis J., "The Bathtub Curve and Product Failure Behavior: Part One - The Bathtub Curve, Infant Mortality and Burn-in", *Reliability HotWire*, Issue 21, November 2002.

BIOGRAPHY

Rene L. Bierbaum
Sandia National Laboratories
P.O. Box 969, MS9007
Livermore, CA 94551-0969 USA

e-mail: rlbierb@sandia.gov

Rene Bierbaum received a B.S. in Electrical Engineering from the University of Nebraska in 1983 and an M.S. in Electrical Engineering from Stanford University in 1984. She is a Distinguished Member of Technical Staff at Sandia National Laboratories in Livermore, CA. She is currently affiliated with the Reliability and Electrical Systems Department at Sandia, working as a reliability analyst for the nuclear weapons program.