LA-UR-04-0496

Approved for public release; distribution is unlimited.

Title: Physics package confidence: "one" vs. "1.0"

Author(s): D. H. Sharp
T. C. Wallstrom
M. M. Wood-Schultz

Submitted to: Proceeding of the 2003 NEDPC Conference

Los Alamos
NATIONAL LABORATORY

Physics Package Confidence: "ONE" vs. "1.0" (U)

D.H. Sharp,* T.C. Wallstrom, M.M. Wood-Schultz,† LANL**

*T-13 and X-DO; ** T-13; †X-2

The "reliability" of the nuclear explosive package of a stockpiled nuclear weapon has historically been stated to be "ONE," with the intent to convey very high confidence that a device that was properly constructed and that had been properly handled would perform as expected on receipt of the appropriate arming, fusing, and firing signals. In this paper, we report on the results of recent work clarifying the basis for assertions of confidence when applied to high consequence systems in the context of Quantified Metrics and Uncertainty (QMU). Previous work on QMU has used a conservative approximation that assigns a confidence of "ONE" or "Not ONE" for nuclear weapons. We extend QMU to a fully probabilistic setting, in which confidence in performance can be assigned a probability between zero and one. We use this more general formulation to examine the assumptions underlying the more conservative model. The potential for this work to support a quantitative evaluation of physics reliability "beyond ONE" is also discussed. (U)

Introduction

The problem of establishing the level of confidence that can be assigned to assessments of the performance, safety and reliability of nuclear weapons in an evolving stockpile is a central question in the nuclear weapons program. Various sources of uncertainty affect assessments of weapons behavior, and a fundamental problem is to estimate confidence in weapon performance and reliability in the light of these uncertainties.

The approach to confidence that has been taken historically is based on conservative bounds on uncertainty, grounded in nuclear test results and supplemented by scientific judgment. This has led to a binary assessment of confidence as "ONE" (we have high enough confidence that the weapon will work properly to allow certification) or "Not ONE" (we do not have sufficient confidence for certification). A confidence assessment of "ONE" is thus not an assertion that the probability of some event (successful operation of a nuclear weapon) is 1.0. It is rather a statement that the balance of evidence is sufficient to support certification.

A traditional approach to assigning confidence would determine the probability of successful performance, a number p between 0 and 1. This approach is more rigorous, to the extent that the assignment of probabilities is based on actual experimental data, and, in principle, it gives more information, allowing for "gradation" of levels of confidence. However, it requires knowledge of the probability distributions characterizing different aspects of a weapon's performance, and thus places more stringent demands on data and/or scientific judgment.

Scientific judgment has always been required to bring closure to the scientific process and to render certification decisions tractable. For example, it was always necessary to extend the experimental determination of weapons performance from a limited set of tested conditions to the full set of potential deployment conditions. Very often, a stockpiled device differed in significant ways from the devices tested for its development, so stockpile performance was inferred from test results for a "similar" but not identical device. In such cases, the adequacy of the data and processes used for these extensions is ultimately a matter of judgment, and it was and is a function of technical management to evaluate and mitigate the associated risks.

The questions that must be answered reliably to maintain confidence in the stockpile include assessments of weapons' behavior in circumstances where (1) aging, engineering flaws or manufacturing defects result in stockpile devices that fail to meet original specifications or (2) nuclear design flaws, apparent or suspected, come to light. In addition, certification of new designs and/or new applications of existing devices could be desired if deemed necessary for national security. Under a comprehensive test ban (CTB), the design laboratories must attempt to answer this range of questions without further nuclear tests. It seems likely that this can be done with the requisite confidence for some questions but not for others. To push the envelope of what can be reliably certified without nuclear testing as far as possible requires advances in predictive science across a broad front including experimental, modeling and simulation capabilities.

We believe that we state the obvious in saying that the most crucial role of any certification methodology is to clarify the choices and judgments made in deciding whether or not to certify a device at all. The determination that a device fully meets the weapon system military characteristics (MCs), and the predictions of the range of its performance are qualitatively different products of the certification process. Thus while reliability and performance are both important, they are not the same thing for a nuclear weapon system.

Because of the potential consequences of a weapon failure, a policy of strict conservatism is usually adopted. This is implemented in part by requiring that a certification be based on persuasive evidence that the device *will work*, as well as on the *absence* of significant evidence that it might not work. Note that conservatism too requires judgment in its application. Both too much conservatism and too little have their costs.

The method of Quantified Margins and Uncertainty — QMU — has been introduced to provide a systematic and explicit framework for explaining the scientific basis for confidence in assessments of the performance, safety and reliability of nuclear weapons. QMU is a logical framework built on salient characteristics of a weapon's performance, each of which is known as a metric, derived from an analysis of experimental data and computer simulations. Requirements for robust operation for the metrics are termed "gates," and the margins of interest to QMU are the amounts by which the metrics exceeds the requirements. Gates and margins are well suited to inform the binary decision of whether or not to certify a weapon. That aspect of certification is addressed here.

A key feature of QMU is the integration of the uncertainties into both the requirements for the metrics -- the gates -- and for the metric values for real-world systems. Consequently, QMU is sufficiently flexible that it provides a very general framework for the analysis of confidence. The simplest formulation of QMU, here called "interval QMU," typically uses uniform probability distributions over a finite range of metric values and leads to the ONE/Not ONE description of confidence. A fully probabilistic formulation of QMU, which permits probability distributions of arbitrary form, can be used to produce more general probabilistic estimates when the necessary data is available. (See, for example, McLenithan, 2001 NEDPC.) Interval QMU is of course a special case of fully probabilistic QMU. We stress that QMU is a framework for analysis, and does not itself supply the science necessary for the evaluation of gates and metrics. However, in calling for these quantities, and especially their uncertainties, QMU engenders a scientific program.

In the next section, we discuss the rationale for QMU in greater detail. In Section 3, we briefly describe the basic ideas of QMU. In Section 4 we present the fully probabilistic formulation of QMU. Interval QMU is discussed in Section 5. Both sections emphasize the assumptions and approximations underlying the two versions of QMU. We discuss the idea of a confidence interval in Section 6. We carry the analysis further in Section 7 by reviewing the explicit and implicit assumptions that are made in the choice of specific simple probability distributions. In Section 8, we touch on some issues that would arise in applying classical statistical reliability theory to the nuclear explosives package. Our conclusions are summarized in the final section.

2. Why QMU?

Ideally, a full description of the behavior of a nuclear weapon would be contained in the solution of the physics equations used to model its performance. However, nuclear weapons—perhaps the most subtle and complex of all man-made objects—are particularly difficult to simulate. Many interesting nonlinear phenomena are operative. The interactions occur on an extremely wide range of length scales and both fundamental material properties and the properties of manufactured "parts" affect device operation. Multiscale phenomena can be sometimes be handled by developing a model for the average or aggregate effects of short length scale interactions and using it to explicitly include the effects of the sub-grid scale phenomena in macroscopic calculations. However, it is not possible to ensure that all of the phenomena emerging from multiple non-linear interacting phenomena have been anticipated, so a general simulation of nuclear weapons has been intractable in practice. Because of their unique physical regime of operation, many of the physical models and data used in other areas of science and technology are not immediately applicable to nuclear weapons.

The equations and the detailed device description on which a weapons simulation is based are for these reasons incomplete and inaccurate. In addition, solution errors for coupled non-linear equations are inevitable. The net result is that computed solutions are rarely adequate for stockpile stewardship needs by themselves. This kind of situation—needing a simulation capability which is by itself inadequate for the ultimate application—is undoubtedly not unique to nuclear weapons. The only sure solution is to turn to full scale testing to supplement and confirm (or not) the simulation results, but multiple obstacles have always precluded the exhaustive—and in some cases even adequate—testing of nuclear weapons, and the scarcity of detailed data on nuclear weapons operation is a pervasive problem.

The cost of a nuclear test kept the number of tests low relative to the development needs, including experiments to directly demonstrate repeatability of performance or parameter sensitivities. The physics regime inside an operating nuclear weapon was itself a major impediment to direct or detailed measurements. National policies did of course play a major role in limiting progress in understanding nuclear weapons: first in the transition from atmospheric to underground testing under the Limited Test-Ban Treaty (LTBT), then to yield-limited tests under the Threshold Test-Ban Treaty (TTBT; not ratified, but implemented by executive order), and now a functional CTB.

As opportunities to obtain directly applicable data have diminished, the nuclear weapons community has increasingly compensated for this lack of data by the use of "normalized simulations" in which adjustments of free parameters and/or simulation "knobs" are used to bring calculated results into agreement with experimental measurements. The adequacy of such models for untested configurations, that is, their usefulness for prediction, cannot be determined directly. Because the model predictions provide the basis for decisions concerning both the safety and reliability of the nation's nuclear deterrent, some form of quality assurance is needed for the modeling and decision processes. We believe that the balancing of requirements and uncertainty that occurs in a valid QMU analysis provides appropriate quality assurance for this application.

QMU is far from foolproof. Some acceptance of the sufficiency of the supporting analysis is a fact of life in any decision making process (excluding so-called "gaming"). The thorough application of QMU to an inappropriate or inadequate set of metrics will not produce the desired results. Its successful application depends on correctly judging the sufficiency of the metrics used. The equivalent statement is true for all other decision paradigms. The value of QMU or an equivalent methodology lies in its mandate to clearly and completely identify all pertinent assumptions and in the increased likelihood that requirements will be dealt with consistently and scientifically.

3. Fundamentals of QMU

The quantity of information necessary to completely specify the detailed spatially and time resolved state of the many physical variables important to the operation of a complex system is enormous. Even when such information is available, for example as it could be from a simulation, it is neither feasible nor useful to attempt to draw meaning from any significant fraction of these individual values. In practice, of course, one integrates some or all of this high-dimensional set of information into a relatively small number of understandable physical quantities. In QMU, a set of such quantities -- termed "metrics" -- are used as indicators of the robustness of the system under analysis.

QMU metrics were originally conceived as key characteristics of the device at important junctures in its operating history. We adopt a more general definition, in which a metric for a physical system can be any quantity that depends on the physical characteristics and state of the system and/or its operation. The set of metrics that it is useful to consider in a QMU analysis will depend on the system and decision in question. For example, the metrics for determining the acceptability of a new high explosive will not be the same as the metrics for determining the minimum isotopic enrichment of a fissile material.

A set of metrics might include one that summarizes a "make-or-break" characteristic, such as the peak Pu compression in a primary or an aspect of device performance; a quantity that characterizes a critical juncture in time, such as just pre-boost; or a calculated or measured quantity that has been highly predictive of operational success, such as explosion alpha in a primary.

Metrics are not generally independent. For example, a poor value for one metric may imply, or be strongly correlated with, a poor value for another metric. And while metrics frequently describe a weapon at different stages in its time evolution, one might choose to describe some stages with several metrics. Metrics which characterize non-physical, idealized states, such as a peak condition obtained without yield, are also useful. Figure 1 illustrates the main ideas discussed below.

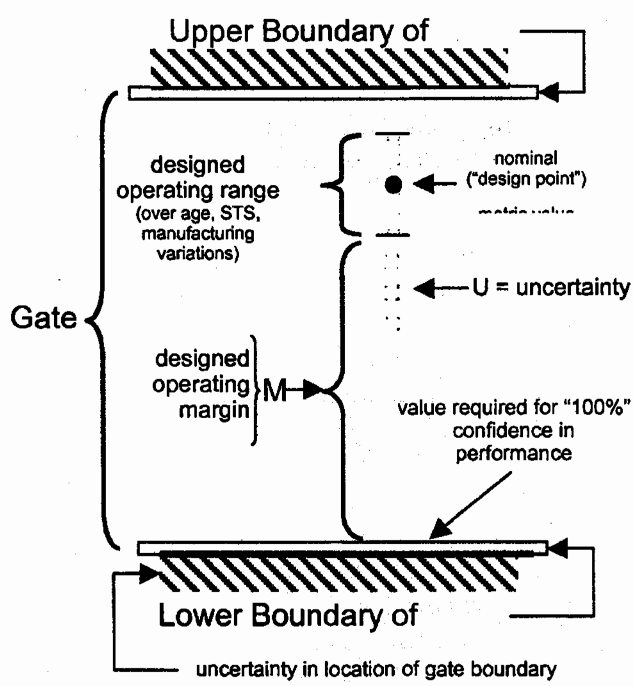

Figure 1. The main features of a performance gate are illustrated.

A corollary to the definition of a metric as a characteristic important to successful operation is the idea that certain ranges of values for each metric represent successful operation and others do not. We define a QMU gate as the range of metric values that is believed to be necessary for robust operation. The conservatism intrinsic in this definition is not essential to the QMU concept, but is considered appropriate for nuclear weapons applications as discussed above. The modifiers "necessary" and "robust" function to exclude any explicit indicator of *risk*: the sufficiency of any gate or set of gates is neither assumed nor implied by the QMU formalism (necessary, but not sufficient), and regimes of incipient failure (not robust) are unacceptable. We note that a gate can be one or two-sided, depending on whether it represents a threshold or a requirement that the metric fall between both lower and upper bounds. For simplicity, we assume that the gate has a single, lower bound; extensions to two-sided gates are straightforward.

When used as the basis for decision-making, gates are interpreted as a set of requirements that are sufficient to ensure that the device *will work*. This is an approximation because restricting our focus to a limited set of metrics creates the risk of neglecting other phenomena that affect device performance. It is generally not known that the converse assumption holds, that is, that the device will definitely fail if any metric is outside of its performance gate. However, in order to ensure that our confidence errs on the side of conservatism, the device is assumed to fail if the metrics fall outside the performance gate.

As discussed in the Introduction, sufficiency is not guaranteed. This is especially obvious in weapons because our understanding of the detailed processes occurring in their operation is incomplete. The point of that previous discussion is important enough that it bears repeating: the question of sufficiency of a set of gates lies outside the purview of QMU itself.

The value of the metric corresponding to the operation of a weapon under nominal conditions is called the "design point". A weapon is designed to operate properly over a range of conditions (the STS), and the value of the metric may vary as a result. This set of values, which by construction contains the design point, is termed the "operating range (OR)."

In order for a weapon to be reliable, the OR must fit "comfortably" within the performance gate. Margin in a metric is simply the "clearance" between the performance gate and the OR. For example, a two-sided gate has two margins—upper and lower. We define the margin in terms of the boundary of the OR, because we want to evaluate our confidence relative to the extreme values of the device operating range.

If the device were completely understood and accurately modeled, any positive value for a margin would be sufficient to ensure successful performance at a gate. In reality, of course, there are uncertainties in both the location of the OR and the performance gate. We describe some of these uncertainties below. The margin must be sufficiently large compared to the uncertainties to ensure confidence in performance. This is of course the familiar idea of a designed "safety margin" for performance— sometimes referred to as the "design cushion." This idea is captured in QMU by defining a confidence ratio, *CR* = *M/U*, and requiring *M/U>1*, where *M* and *U* are numbers characterizing the margin and uncertainty. This is discussed further below.

Interval and full QMU differ in the amount of information and level of detail they require for the useful description of the uncertainties. We discuss these formalisms next.

4. Full QMU

In full QMU, the uncertainties in the OR and performance gate boundaries are characterized by full probability distributions. Since the gates are not independent, we should really introduce a multivariate distribution for the entire set of gates. However, we will restrict ourselves here to the case of a single gate.

For a one-sided lower gate, the margin *M* is defined as $x_o - x_g$, the difference between the lower boundary of the OR, x_o, and the location of the lower performance gate, x_g. We assume that the distributions of x_o and x_g are independent. This assumption is almost surely not strictly true, especially as concerns the mean of the distribution. However, it may be a sufficiently good approximation when calculating the difference $x_o - x_g$. Once we have assigned distributions to x_o and x_g, we need only convolve these distributions to calculate the distribution of the margin. (McLenithan, NEDPEC 2001.) We illustrate this procedure with two simple examples.

Gaussian Case

Assuming x_o and x_g are independent, the density of *M* is given by:

$$f_M(x) = \int_{-\infty}^{\infty} f_g(z-x) f_o(z)\, dz$$

where f is the density function of the corresponding random variable (*M*, x_o, or x_g).

If x_g is normal with mean and variance μ_g and σ_g^2, and similarly for x_o, then *M* is normal with mean $\mu_o - \mu_g$ and variance $\sigma_M^2 = \sigma_o^2 + \sigma_g^2$, by elementary properties of the Gaussian. The fact that the squares of the uncertainties add, rather than the uncertainties themselves, leads to a smaller uncertainty for *M*, and reflects the fact that uncertainties in the gate and the operating range will tend to cancel, to the degree they are independent.

If $\sigma_g \to 0$ and $\sigma_o \to 0$, then $\sigma_M \to 0$, and f_M becomes a Dirac δ-function centered at $\mu_o - \mu_g$. We thereby recover a "deterministic" expression for *M*.

Rectangular Case

In this case, $f_g(x) = 1/(2h)$, if $-h < x - \mu_g < h$ and zero elsewhere, and analogously for f_o, with *k* replacing *h* and μ_o replacing μ_g. We assume that *h<k*; if *h>k* then the following result is valid if *h* and *k* are interchanged. Performing the convolution, we obtain a trapezoidal distribution function for *M*:

$$f_M(x) = \begin{cases} 1/(2k) & |\xi| < k-h \\ ((k+h)-|\xi|)/4hk & k-h < |\xi| < k+h \\ 0 & \text{elsewhere} \end{cases}$$

where $\xi = x - (\mu_o - \mu_g)$. If *h*=*k* the distribution is triangular. As $h \to 0$, corresponding to zero uncertainty in either the gate or the OR, the distribution becomes rectangular. If one quantifies the uncertainties by the root-mean-square error, one finds, as in the Gaussian case, that the uncertainties are subadditive, because of error cancellation.

5. Interval QMU

In interval QMU, the uncertainty in the lower boundary of the operating range is characterized by a single number, U_o, and the uncertainty of the performance gate is characterized by a second number U_g. The interpretation of these uncertainties is that we are "nearly certain" that the true value does not differ from the mean value *x* by an amount greater than *U*. In other words, we are asserting that, to a good approximation, the support of the distribution is $[x-U, x+U]$. (The support of a function is the set on which it is nonzero.)

The interpretation of the uncertainty in *M* as the support of its distribution implies that when adding (or subtracting) random variables, we should add their uncertainties. If x_o and x_g are supported in intervals $\mu_o \pm U_o$ and $\mu_g \pm U_g$, then *M* will be supported in the interval $(\mu_o - \mu_g) \pm (U_o + U_g)$. This is a conservative, worst-case, assumption, which assumes that there will be no error cancellation, or equivalently, that the errors are completely correlated.

To treat Gaussian distributions, which have support on the entire real line (infinite support), within the framework of Interval QMU we must approximate the Gaussian by a distribution having compact support. We do this by using the uncertainty as given by full QMU to define an effective uncertainty for Interval QMU. For example, we could define the support of the distribution by the limits of the 90% confidence interval for Gaussian distributions. For the case of a rectangular distribution, the uncertainties in x_g and x_o are *h* and *k*, and the uncertainty in *M* is *h*+*k*. In any case, the uncertainties add, so Interval QMU will lead to higher uncertainty estimates than Full QMU.

6. Confidence Ratio

In principle the margin, considered as a random variable, contains all statistical information about whether the design point will fall within the performance gate. We write M_α for the αth quantile of *M*, i.e., the value of *M* that will be exceeded $100(1-\alpha)$ % of the time. By adjusting the value of α, one can impose any desired degree of confidence on the system, simply by requiring that $M_\alpha > 0$.

In practice, however, one is often not sure about the shape of the distribution, particularly in the region of the tails. For this reason, the more conservative method of ensuring confidence via a Confidence Ratio (CR) is used. Using $\bar{M}$ for the mean value of the margin, and letting U be some estimate of the magnitude of the central region of the distribution, the confidence ratio (CR) can be defined as $\bar{M}/U$. If the confidence ratio is large, then we have high confidence that the uncertainty is less than the mean margin, so that the actual margin will very probably be greater than zero.

As an alternative, therefore, to requiring that M_α be greater than 0, for some small α, we could require that CR be greater than γ, for some $\gamma > 1$. As α gets smaller, or γ gets larger, we would gain more confidence. The advantage of using γ is that it depends mainly on the central portion of the distribution, whereas α is sensitive to small changes in the tails.

A concrete definition applicable for the case of Full QMU is $U_\alpha = \bar{M} - M_\alpha$, and $CR_\alpha = \bar{M}/U_\alpha$. If we require that $CR_{0.05} > 1$, then we are requiring a 95% probability that M is greater than zero. If we require that $CR_{0.05} > \gamma$, where γ is greater than one, then we are requiring a correspondingly higher degree of confidence that M is greater than zero, although we cannot determine the probability exactly without making potentially questionable assumptions about the probability distribution of M. Qualitatively, however, we can say that if the tail falls off quickly, then a value of γ that is only a little greater than one should lead to high confidence; if it falls off slowly, then it is prudent to require a larger value of γ.

For the case of Interval QMU we take U to be the half-width of the support, and $\bar{M}$ to be the midpoint of the support. If we interpret the distribution literally, then the probability that $M<0$ is zero if the $CR>1$. However, it is still a good idea to insist that $CR>1$, because the assumption of zero probability outside the support of the distribution is an idealization that may not be valid in practice.

The preceding discussion provides some insight into the meaning of the numerical value of the CR, in that different values of the CR reflect different assumptions about properties of the relevant probability distributions. We emphasize that confidence in successful device operation requires successful performance at each gate. Lack of independence of the basic quantities defining the various gates precludes defining a single confidence ratio for overall system performance by any simple recipe. Instead we adopt a "weakest-link-in-the-chain" strategy, and take a CR<1 at any gate as a signal of possible system failure (reliability Not ONE), and the requirement CR>1 *for every* gate as an indicator of robust performance (reliability = ONE). This is consistent with the conservative ONE/Not ONE binary assignment of confidence that has been used historically.

7. Determining the Probability Distribution for Full QMU

As stated above, the successful use of full QMU at a specific gate requires knowledge of the probability density functions (PDFs) for the limiting metric value and for the gate itself. The determination of the PDFs is therefore extremely important, but for nuclear weapons, it is typically extremely complex and difficult. Detailed uncertainty analyses that may lead to PDFs have been undertaken in several important weapons areas, such as pit radiography. Each of these endeavors is both a significant and unique undertaking. However, some simple observations based on the defining requirements and the overall context of predictive capability for nuclear weapons follow fairly directly.

As a rule, the determination of the operating range and the location of the boundaries of performance gates will use both experimental data for real devices as well as simulation results for these devices. These determinations will typically reflect uncertainties in the measured data and in the simulations. Errors in the latter have several sources: (i) input error, including database errors; (ii) solution errors, which arise from deficiencies in numerical methods; and (iii) modeling errors (errors in the physics equations). Unfortunately, the various kinds of errors are often intertwined. So while in practice, simulations inevitably reflect *experimental* errors in input data (e.g., EOS), utilizing experimental data often requires model-based inference, and so reflects modeling error as well as the measurement errors themselves.

The most important part of any uncertainty analysis is a careful study of the potential sources of error based on a scientific understanding of the operation of the experimental measuring instrument or the simulation code. This is particularly important when systematic errors are the issue, as is frequently the case in both experiments and simulations, because systematic errors are difficult, if not impossible, to estimate using statistical methods alone.

A sufficiently large sample of data points enables a good model for the distribution through fitting to a familiar parametric distribution or to an empirical distribution. If measured data is not available, one alternative is to generate simulated data for a "closely" related problem and then evaluate its applicability to the problem of interest. In many cases, however, few applicable data -- real or synthetic -- will be available, and an estimated for the PDF can only be based on largely qualitative insights.

We mention two potentially useful techniques: Maximum Entropy, and optimal estimates of the mean. If The Maximum Entropy method requires knowledge of the moments or range of the distribution, which may be available on theoretical grounds in certain applications. When such information is available, the Maximum Entropy distribution may be obtained by maximizing the continuous entropysubject to the constraints. The Maximum Entropy distribution can be used as a starting point, or prior, for Bayesian Inference, when measurement data is also available.

The second technique infers the distribution from a given method for estimating the mean. Common examples are: as an average of all measurements, as a weighted average in which the "outliers" are given reduced weight, as the average of the extreme observations, and as the median. For a specific problem, an analyst may use physics reasoning to supports the choice of one method of estimating the mean over another. Gauss showed that if one assumes that the chosen form corresponds to an unbiased estimate, then the form of the distribution is actually largely determined by this choice. This means that any method for estimating the mean entails an implicit assumption about the form of the distribution. The first three methods, for example, correspond to a

Gaussian distribution, a fat-tailed distribution, such as a *t* distribution, and a rectangular distribution (thin-tailed). It is invariably easier to find intuitive justification for a method of estimating the mean than to produce a direct intuitive argument for the distribution itself, and it is fortunate that this window into the form of the distribution can be so directly exploited.

Except in cases where the probability of satisfying the gate requirement is low, the probability of interest is determined by integration over one or more tails of the PDFs. The tail of the distribution is most affected by inaccuracies in either the form or the distribution parameters, e.g. the variance of the PDF. The confidence ratio is crucial in full QMU to provide compensation for these inaccuracies. The aspect of a PDF most important to the effects of the tails is its "kurtosis," which characterizes the degree to which the tails are thin or thick. Although the above methods do not always yield quantitative answers, they should often be able to tell us qualitatively whether the kurtosis is large or small, and this may be more than adequate for setting the confidence ratio. If no practical approach can be convincingly justified, then full QMU cannot provide the basis for a high confidence decision concerning the problem being considered.

We place considerable stress on the need for rigor and conservatism in the determination of probability distributions for use in QMU. The reason is that conclusions can be very sensitive to the choice of distributions (or its parameters); an incorrect choice can lead to disastrously wrong answers.

8. QMU and Reliability

Classical statistical reliability theory (Barlow & Proshan) deals with two main concepts: (i) "Reliability" — the probability that a device will perform properly for the period of time intended under the operating conditions encountered and (ii) "Pointwise availability" — the probability that a device will operate to specifications at a specific time. These concepts are used to describe both component reliability and the overall reliability of a system consisting of multiple components. An associated idea that plays a prominent role is the probability of component or device failure, and also the rate of change of these probabilities of failure.

Here we discuss these ideas in a QMU framework. The metric at a gate, which usually refers to a process rather than to the state of a piece of hardware, is the analogue of a "component". Pointwise availability thus corresponds to the probability of successful performance at a gate — this is $P(m>0)$. The probability of failure $P(m<0)=1-P(m>0)$. Given the preceding discussion of the role of the tails of the distribution in full QMU, it is evident that the probability of failure requires knowledge of $P(m)$ where it is least likely to be known accurately.

Some time dependence of $P(m)$ can be expected due to aging effects, both expected, such as aging of gas, and unexpected where chemical and/or environmental effects occurring during a device's tenure in the nuclear stockpile result in out-of-specification conditions.

In many applications, the probability of device failure as a function of time is obtained by direct observation of the frequency of occurrence of component or system failure. This is not a feasible approach for a nuclear explosive package. Instead, the surveillance program seeks to observe the frequency of out-of-specification conditions, but relating the conditions to the probability of failure requires a significant model-based inference.

Any estimate of system reliability depends first and foremost on an understanding of system failure modes, which is obviously a very important thing to study no matter how one intends to use the results.

Inferring the reliability of a complex system has always been a key part of the field; an early example is the Moore-Shannon theorem relating the reliability of a multi-component relay circuit to that of its individual components. Using system reliability results in the QMU framework in principle requires the determination of the conditional probabilities relating metrics at gates N and N-1. In nuclear weapons, this information is not readily available. Sufficient data exist that it may be available from observations in robust regimes of operation, but not in the problematic failure, or near-failure, regimes. The most direct alternative would be to use a model to *derive* the required conditional probabilities between gates. (The dilemma is that the regime where QMU is most needed as a safety net for conclusions drawn from inaccurate simulations is precisely where the weapon model falsely predicts acceptable behavior.) An approach often used in analyzing relay circuits is to impose a binary on/off state for relay operations. Such a "lumped" description avoids dependence on a detailed physical model of a relay. In a weapon, however, the specific value of metrics at different gates will usually be important. Because these values depend crucially on the behavior of the system upstream to the gate in question, a lumped model of the dynamics is not useful, except perhaps for order of magnitude estimates.

It is worthwhile to ask how accurately one needs to know the PDFs that come into play in reliability theory in order to produce estimates that are sufficient for certification purposes. This question deserves careful study. We make two simple observations here.

First, concerning the determination of the needed PDFs, the best hope is to concentrate on the "central" region of the distribution, corresponding to a regime of robust operation. This aligns with the refurbishment strategy of maintaining the device in a condition where the existing performance data were collected. It also illustrates the reasonableness of the traditional use of interval PDFs in certification.

Second, some very worthwhile aspects of reliability analysis do not depend sensitively on details of the probability distributions. As commented above, a first step in the analysis of system reliability is usually a careful analysis of failure modes and their effects. Likewise, powerful results in reliability theory sometimes follow from a knowledge of the mean of the distribution and general assumptions, such as monotonicity of the probability of failure, that are frequently applicable. Explorations of such approaches in analyzing the reliability of the nuclear explosives package appear worthwhile; these methods typically lead to reliability bounds.

Progress in implementing the approaches mentioned in the above two paragraphs should clarify the scientific basis for the ONE/Not ONE characterization of reliability. They can be expected to confirm that this characterization of reliability is based on some hard facts of life. Going "beyond ONE" to a numerical estimate of reliability based on knowledge of full probability distributions in conjunction with QMU would place great demands on our ability to characterize uncertainties. In view of this, it is inevitable that there would be pressure to adopt "short cuts" by simply assuming the forms of PDFs or using PDFs that are not based on some but inadequate supporting data. The response to such pressure would make or break nuclear certification. No analysis that is based on speculation or that neglects significant possibilities can lead to genuine confidence, but instead will frequently lead to over-confidence or under-confidence, both of which carry severe costs.

9. Summary and Comments

We have presented two ways of formulating QMU that we have termed "Interval", and "Full." The difference between the two formalisms stems solely from the restriction of the PDFs representing the probability distributions for the values of a metric for a given device configuration and for the limits of a gate to those with non-zero values for a single finite interval in Interval QMU. The use of either formulation requires approximations and scientific judgment. However, fully probabilistic analyses require knowledge of uncertainties that as a rule can be expected to be difficult to obtain.

"ONE/Not ONE" estimates of reliability are consistent with historical practice, and both are a response to the very limited availability of detailed uncertainty information. Extending this class of estimates to current and future stockpile questions is itself a significant challenge.

The US nuclear weapons community is just beginning to explore where QMU can be useful. The emphasis in QMU on characterization of uncertainties is essential for determining which stockpile questions can be answered with confidence using the data and simulation capabilities that are available at any given time. Improvements in predictive science may allow improved estimates of uncertainties, and potentially may also allow them to be reduced. This would affect the scope of questions that can be dealt with successfully, but significant limitations are expected to remain. Identifying these limits is an important task of the nuclear weapons program.

We submit that when carefully applied, QMU can improve our basis for assessing the reliability of stockpile decisions.

Acknowledgments

References

R. E. Barlow and F. Proschan, Mathematical Theory of Reliability, (SIAM, Philadelphia), 1996.

K. McLenithan, Proceedings of the 2001 NEDPC Conference. [in press]

D. H. Sharp and M. M. Wood-Schultz, QMU and Nuclear Weapons Certification: What's under the Hood?, Los Alamos Science, Number 28, 2003, pp. 47-53.