Goal Based Testing: A Risk Informed Process

Chester Everline[1]
Jet Propulsion Laboratory/California Institute of Technology, Pasadena, CA 91109

Clayton Smith[2]
Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723

Sal Distefano[3]
Jet Propulsion Laboratory/California Institute of Technology, Pasadena, CA 91109

and

Natalie Goldin[4]
NASA/Glenn Research Center, Cleveland, OH 44135

A process for life demonstration testing is developed, which can reduce the number of resources required by conventional sampling theory while still maintaining the same degree of rigor and confidence level. This process incorporates state-of-the-art probabilistic thinking and is consistent with existing NASA guidance documentation. This view of life testing changes the paradigm of testing a system for many hours to show confidence that a system will last for the required number of years to one that focuses efforts and resources on exploring how the system can fail at end-of-life and building confidence that the failure mechanisms are understood and well mitigated.

Nomenclature

λ	=	failure rate
$\pi_0(x)$	=	prior probability density function for the random variable, x
$\pi_1(x\|E)$	=	posterior probability density function for the random variable, x, given evidence, E
A_f	=	acceleration factor for testing
E	=	evidence
$erf(x)$	=	error function of x
$f(r\|E)$	=	conditional probability density function for r, given evidence, E
$ln(x)$	=	natural logarithm of x
M	=	number of units tested
r	=	reliability
t	=	test duration
T	=	mission duration

I. Introduction

A risk informed goal based testing process was developed for NASA's Radioisotope Power Systems Program at Glenn Research Center[1] from the perspective of minimizing the resources (e.g., project budget and schedule) needed to demonstrate compliance with stakeholder expectations with respect to the lifetime of a new technology. The current paper augments that process by explaining how a suitably risk informed process can support stakeholders in establishing lifetime goals which can be demonstrated with available resources.

---

[1] System Engineer, Product and Circuit Reliability, 4800 Oak Grove Drive/MS 156-206, Pasadena CA 91109.
[2] System Engineer, Risk and Reliability, 11000 Johns Hopkins Road, Laurel, MD 20723.
[3] System Engineer, Power and Sensor Systems, 4800 Oak Grove Drive/MS 303-300, Pasadena CA 91109.
[4] Senior Systems Engineer, Systems Engineering and Architecture Division/Human Space Flight Systems Branch, 21000 Brookpark Road/MS 142-6, Cleveland OH 44135.

II. Problem Statement

The dilemma for demonstrating reliability or service life for items that must last a long time (perhaps over a decade) is that the testing is expensive, time consuming, must satisfy multiple sets of expectations, and address a variety of requirements simultaneously. When designing a testing regiment to validate a reliability requirement, programs often turn to classical statistical methods in order to determine the amount of time needed. For example, if a program requires a reliability of 0.90 at 15 years (130,000 hours) with a confidence of 90%, the number of hours needed for demonstration with no failures is 2.9 million hours using sample theory with an exponential distribution assumption.[2] A straight forward method advocated by DoD, MIL-HDBK-781[3] for reliability demonstration test planning yields test hours from 1.1 million to 1.6 million depending on risk ratios designated. Bayesian methods such as the WeiBayes Zero-Failure formula[4] yields 1.2 million hours. These hours increase tremendously if failures have to be taken into account. Again applying sample theory with an exponential distribution assumption, the number of hours needed to demonstrate 90% reliability at 15 years with 90% confidence increases from 2.9 million hours to 4.8 million hours if one failure is allowed, and increases to 6.6 million test hours if two test failures are allowed. Rather than focus exclusively on how to demonstrate compliance with reliability requirements, this augmented process addresses tailoring reliability requirements to available resources in a manner that is satisfactory to stakeholders.

III. Process Description

Our process begins with requirements for a new technology based on inputs from the various stakeholders. These requirements will encompass safety, technical performance (including lifetime), cost, and schedule.

For long duration missions with the expectation of high reliability from the new technology, the challenge is to demonstrate compliance with the technical performance requirements in a manner compliant with cost and schedule constraints. In systems engineering parlance, demonstrating compliance with technical performance requirements is the Evaluation Process in NASA's Systems Engineering Engine (Figure 1).[5-7]

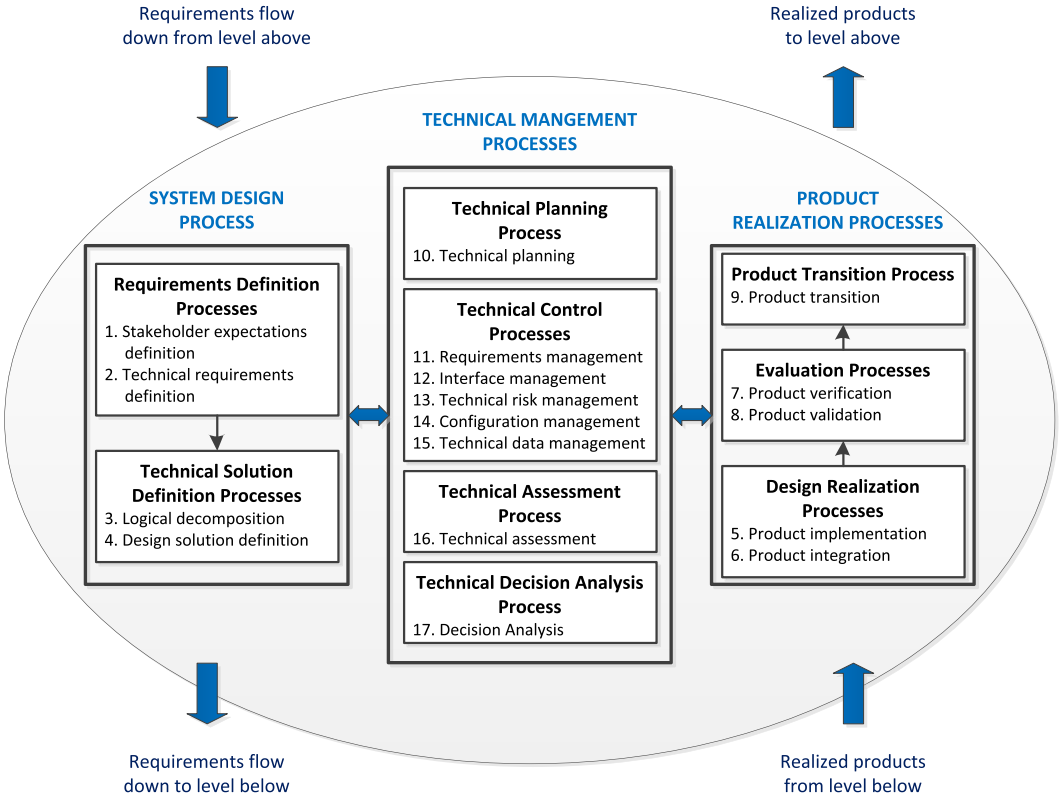

Figure 1. Systems Engineering Engine.

The *Requirements Definition Process* also merits emphasis in our methodology because unless lifetime reliability is specified properly, the *Evaluation Process* can become argumentative due to ambiguity in what constitutes an acceptable demonstration of adequate lifetime. A simple illustration of this is afforded by considering a new technology intended for a ten year robotic science mission into deep space. The stakeholders agree that 90% reliability over ten years is a good representation of their expectations. Suppose that this technology is so novel that its ability to operate in deep space is unknown, but we test it for a month under flight-like conditions. Using a standard exponential distribution to model reliability and, since the technology has no significant heritage, a Jeffreys' non-informative prior,[8] the probability the new technology will furnish 90% reliability over the ten year mission is on the order of a few percent. Most assuredly, no stakeholder, project manager, or systems engineer would accept this calculation as a demonstration that the new technology satisfies their expectations. The mission is 120 times longer than the life test, there is no insight into whether aging and wear-out will begin to degrade performance after a few years of operation (such phenomena would certainly invalidate the application of an exponential reliability model), nor was there any agreement on the type of statistics appropriate for demonstrating lifetime (e.g., Bayesian statistics, sample theory, or MIL-HDBK-781). These same issues are also true is we test several items for just a few months to accumulate a greater number of test hours.

An obvious challenge with respect to quantifying stakeholder expectations for reliability and confidence is that requirements such as 90% reliability with 90% confidence, 90% reliability with 95% confidence, or 95% reliability with 99% confidence usually are nothing more than arbitrary numbers, without physical or programmatic meaning. Given the importance of having well defined requirements, however, our recommendation is not to solicit numbers from stakeholders, but ask them what type of life paradigm they expect in the context of existing, historic or even hypothetical missions.

An example of developing a life paradigm can be illustrated by hypothesizing that the new technology being considered is a radioisotope power system (RPS). Stakeholders could specify that for a mission of duration, T, they wish to have reliability, r, where r is some quantitative value (e.g., 95%). Many technical programs impose/recommend quantitative reliability requirements for projects,[9-11] so techniques for deriving/quantifying r exist. These techniques typically involve an optimization process subject to constraints, where the variables being analyzed can include safety, reliability, and project resources (such as budget and schedule). The techniques range from highly mathematical[12-19] to architecting traders where the *best* (i.e., optimal, subject to the imposed constraints) design option is selected from those analyzed and its reliability chosen as the baseline allocation for r. The conclusion, here, is that techniques are available for deriving r, but these conventional processes ignore any uncertainty in r. In order to assign a required confidence level to r, it must be viewed as uncertain due to inevitable uncertainty on any tests or analyses used to evaluate it.

Continuing the RPS example, suppose the stakeholders expected Voyager-like reliability and confidence. Both Voyagers have three radioisotope thermoelectric generators (RTGs) and over three and a half decades of successful RTG flight operation on each. Retaining the statistics with an exponential reliability model and Jeffreys' non-informative prior, Figure 2-A depicts the confidence a Voyager-like RPS will have reliability of at least r over a ten year mission. If r is required to be at least 90%, a Voyager life paradigm furnishes 99.99% probability the new RPS will satisfy its baseline reliability requirement over a ten year mission. With respect to project resource constraints, the issue is whether they are sufficient for the project to demonstrate, through testing and/or analyses, that the new RPS technology can furnish 90% reliability with 99.99% confidence during ten operating years.

The impact of a life paradigm on project resources can be illustrated by continuing the RPS example. Figure 3[1] depicts the evolution of RTG technology applied to a selection of robotic science missions. The first RTG appeared in 1954 (see Figure 4).[20] Even if the new RPS technology is a RTG, testing six units for over 37 years is simply incompatible with any technology development schedule (this is equivalent to the RTG flight history for both Voyager, since each Voyager has operated for ~37 years in space and has three RTGs). Fortunately for RTGs, the RTG technical community has developed and maintained physics-based models which have been validated against both flight and test data for a number of different RTG technologies.

Because of the high confidence in these models, instead of a life test equivalent to decades of operation, the new RTG couple material could be characterized through laboratory testing and its physical properties used to predict performance over the ten operating years imposed on the project. If the end of life performance margins are at least as large those quantified for the previous missions (e.g., Voyager) and the confidence in these margins is at least as great as the confidence in the previous margin estimates (see Figure 5), then the new technology will have the same probability and confidence that it will complete its mission successfully. Moreover, since material characterization and margin quantification are a routine part of the RTG technology development process, any resources needed to demonstrate that the new technology will satisfy its reliability requirement with the imposed confidence will be

negligible since such tests and analyses are a standard constituent of the RTG technology development budgeting and scheduling process. Consequently, for new technology which is evolutionary instead of revolutionary, the resources necessary to demonstrate satisfactory lifetime with confidence are usually negligible because the underlying physics is well understood. This has reduced our demonstration problem from evaluating the entire system to one of exploring a specific set of material properties.

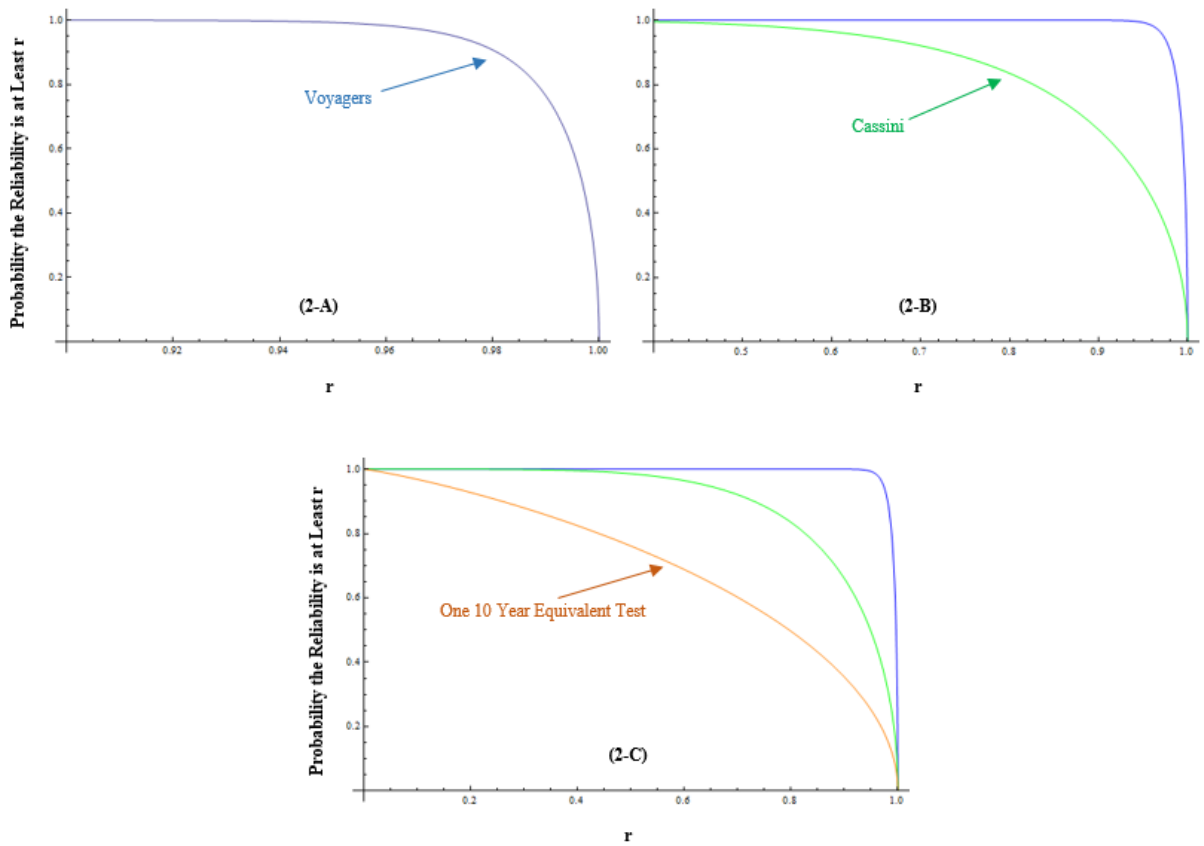

Figure 2. Confidence of Reliability for a 10 Year Mission with Various Life Paradigms. (Note that the domain of the abscissa varies, causing the shape of the curves to shift. The probability the reliability is at least r is derived in the appendix.)

Highly revolutionary concepts represent the other extreme of technology development. Imagine a new power source so novel that there is essentially no applicable test or flight data to demonstrate compliance with the life paradigm. Obviously this concept is at a very low technology readiness level so the project has the opportunity to optimize technology development with respect to performance, lifetime demonstration, and resources. If the Voyager mission is selected as the life paradigm, the project must scrupulously review the testing needs, cost, and schedule. Since a test schedule spanning decades is unacceptable, the project must consider the extent to which life testing can be accelerated and test multiple units. If an A_f on the order of 37 is achievable, life testing compatible with a Voyager life paradigm can be completed in about a year if six units are available for testing (and test interruptions are negligible). However, such an ambitious A_f may not be feasible. Rather than an accelerated test campaign, testing multiple units for a shorter time may be feasible if $M \times t = 6 \times 37$. Here, however, the cost of fabricating M units is unlikely to be affordable. For accelerated testing of multiple units the life demonstration criterion becomes $M \times A_f \times t = 6 \times 37$. An important admonition is that if $A_f \times t$ is less than the mission length for the new technology (ten years for this example), then it becomes necessary to demonstrate that aging and wear-out will not occur during that part of the mission beyond which testing had been conducted. If aging or wear-out can occur late in the mission (or their occurrence cannot be ruled out), it then becomes necessary to establish that their resultant performance degradation can be managed in a manner which will satisfy the life paradigm.

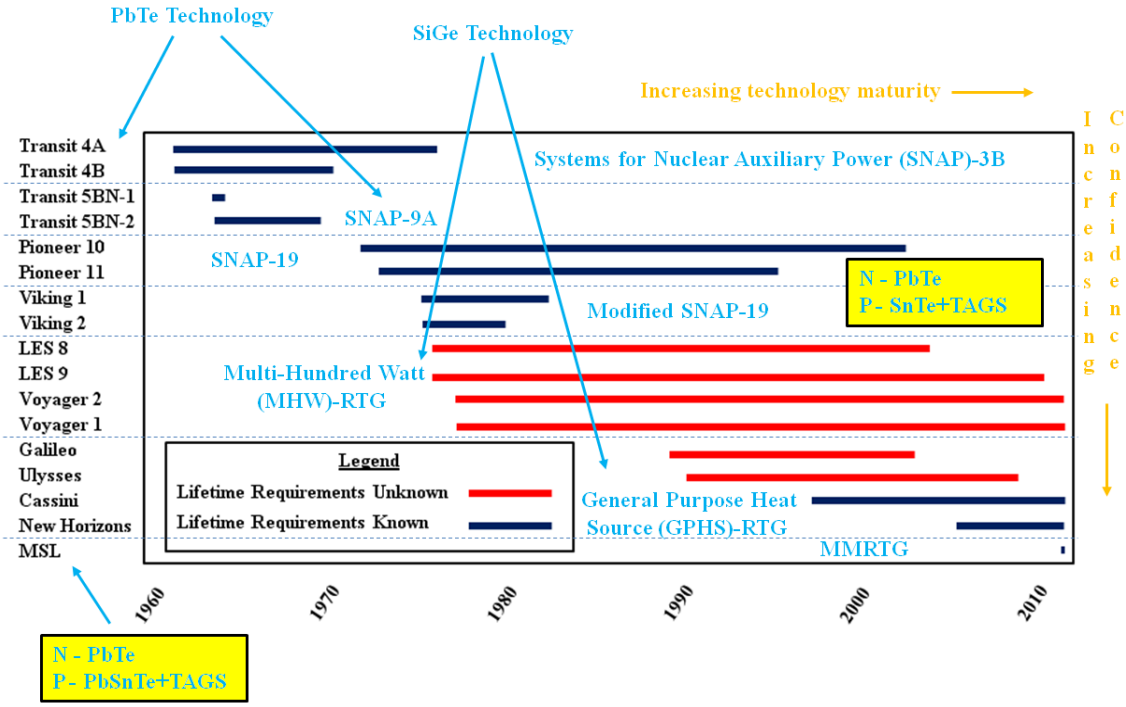

Figure 3. Synopsis of RTG Technology Flown on Selected Robotic Science Missions.

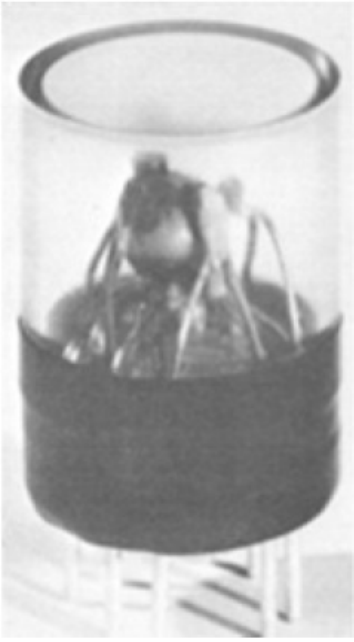

Figure 4. The First RTG (1954).

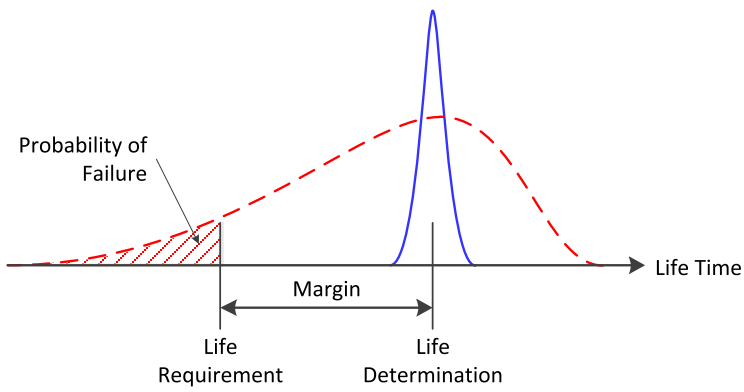

Figure 5. The Probabilistic Nature of Margins.

An alternative to testing is to demonstrate lifetime using analysis (analogous to the approach described for RTGs), but this requires that validated models for the new technology exist, or that sufficient resources are available to the project to develop and validate such models.

One could also consider descoping the technology somewhat, by relying more on established technology where such reliance does not appreciably impact performance objectives. For instance, if the new RPS technology is intended to convert thermal power from radioactive decay into electrical power more efficiently than is achievable with existing technology, it might be prudent to consider lowering the overall conversion efficiency slightly in order to achieve life objectives with confidence, and within available resources (or at least with a smaller increase in needed resources). Note that we have now entered a trade space where reliability, confidence, and performance are being traded in order to satisfy cost and schedule constraints. Increasing reliance on established technology permits the reliability of that technology to be demonstrated using applicable test and flight data, thereby reducing the need for model development and validation, or long-term life tests.

Figure 6 summarizes these types of technology development trades. Where well established technology is being applied (Subsystem A in Figure 6), existing flight and test data can be used to demonstrate lifetime. If established technology has to be modified to integrate it with the new technology, the applicability of existing flight and test data to the modification must be confirmed (Subsystem B in Figure 6). For a new design, new reliability models and more extensive testing are needed to demonstrate lifetime. From the perspective of technology development program trades, resources needed to demonstrate lifetime increase monotonically from Subsystem A to Subsystem C. By decomposing the system to components and interfaces, we can apply more rigorous testing standards to those items with little heritage and accept the wealth of experience for those with more. At the full up system level, testing within resource constraints in conjunction with the goal-based assessments is a viable option. A final option (at least for this example) is to impose a less stringent life paradigm on the new technology. In the context of Figure 2-B, it may be better (from a resource utilization perspective) to impose a Cassini life paradigm on the new technology initially, with the intention that as flight experience is gained the technology can be enhanced to a level where it satisfies a Voyager life paradigm.

Use of a life paradigm also permits consideration of hypothetical life tests, such as Figure 2-C. If the new technology is to achieve 90% reliability over ten years with a Voyager life paradigm, there will be 99.99% Bayesian confidence that this reliability requirement is satisfied. If the new technology is to achieve 90% reliability over ten years with a Cassini life paradigm, there will be 65.95% Bayesian confidence that this reliability requirement is satisfied. However, for a life paradigm equivalent to testing one unit for ten years, there is only 35.38% Bayesian confidence the new technology will achieve 90% reliability over ten years.

III. Conclusion

Using risk informed processes to demonstrate compliance with life requirements (e.g., demonstrating a certain reliability will be achieved with a specified level of confidence) is a useful technique for minimizing resources (e.g., budget and schedule) needed for developing new technologies. Moreover, using risk informed processes to establish life requirements, by directly trading reliability, confidence, and technology performance with resources, enhances the ability of projects and stakeholders to develop life requirements compatible with expectations and resource constraints. Goal based testing is consistent with the risk-informed decision making (RIDM) process,[5-7] as well as

the Constellation Program lessons learned regarding risk informed testing.[21] Risk informed processes should be applied when developing requirements for new technology, and implemented as part of the verification and validation steps needed to demonstrate that the new technology satisfies the technical performance requirements imposed.

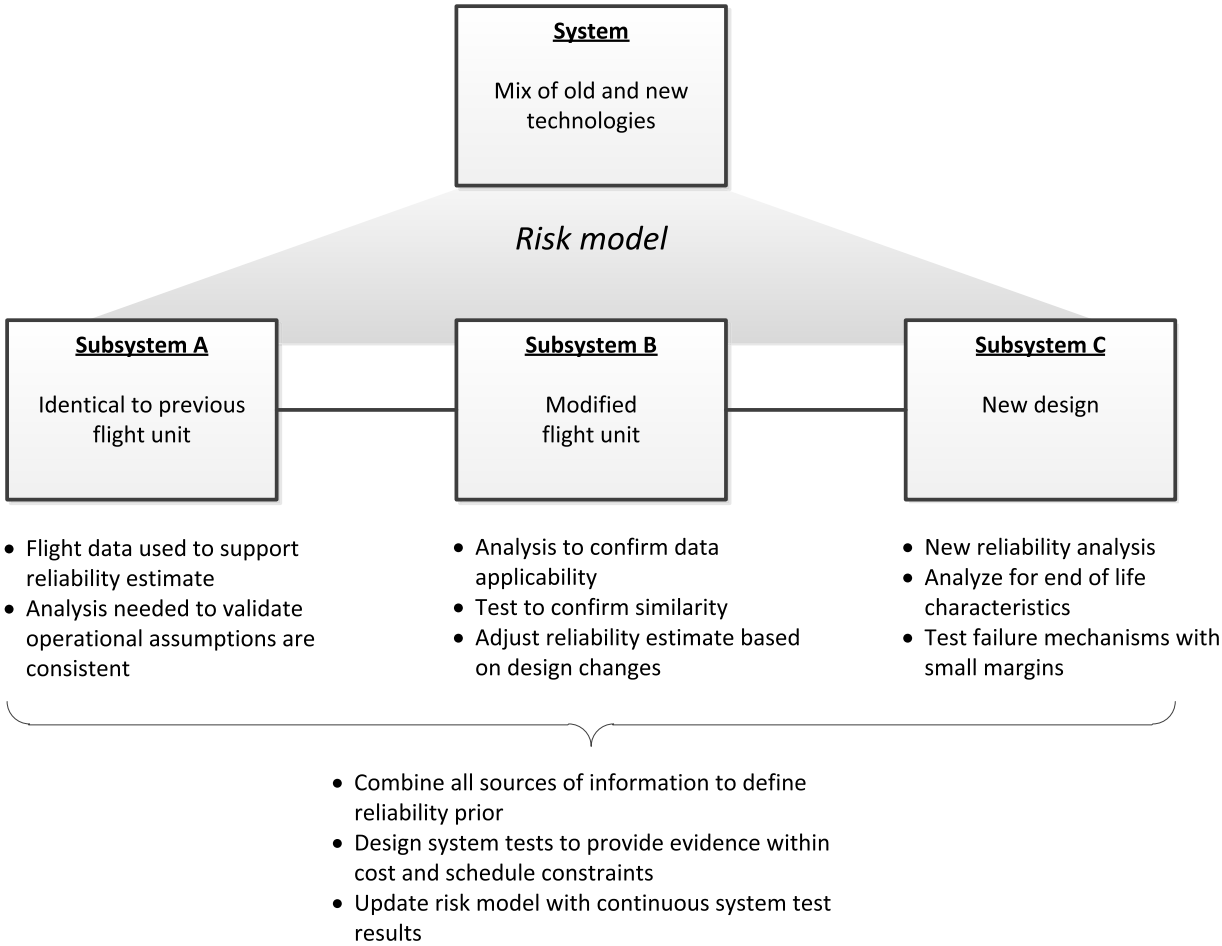

Figure 6. Illustration of the Goal-Based, Risk Informed Process.

Appendix

Postulating it can be demonstrated that λ is time-independent, the probability that no failure will occur in t test hours (or months, or years) is $e^{-\lambda t}$. This is also the Bayesian likelihood function for the new technology. If a Jeffreys' non-informative prior is selected, it has the mathematical form, $1/\sqrt{\lambda}$.

The Bayesian posterior probability density function for λ, given evidence, E, is $\dfrac{e^{-\lambda t}\pi_0(\lambda)}{\int_0^\infty e^{-xt}\pi_0(x)\,dx}$. The evidence is that no failure occurred in t test hours (or months, or years), and our knowledge of λ is conditionally dependent upon this evidence.

Reliability is simply the probability that no failure occurs over a specified interval of time, so for a mission of duration, T:

$$r = e^{-\lambda T} \tag{1}$$

Note that T, in this equation, is the mission duration and differs, conceptually, from t, which constitutes our evidence from testing.
We can derive f(r|E) from Eq. 1 and the basic rules for transformation of variables in probability density functions. Specifically, we know from Eq. 1 that:

$$\lambda = -\frac{1}{T}\ln(r) \tag{2}$$

so:

$$f(r|E) = \pi_1\left[-\frac{1}{T}\ln(r)|E\right]\left|\frac{d\lambda}{dr}\right| = \frac{1}{r\,T}\pi_1\left[-\frac{1}{T}\ln(r)|E\right] \tag{3}$$

The probability the reliability of the new technology has a value of at least r is simply, $\int_r^1 f(x|E)dx$. This is the ordinate in Figure 2. For a Jeffreys' non-informative prior:

$$\int_r^1 f(x|E)dx = \mathrm{erf}\left[\sqrt{-\frac{t}{T}\ln(r)}\right] \tag{4}$$

It is important to remember that this derivation is notional, performed in order to illustrate the overall process. A more rigorous statistical analysis, such as those described in Ref. 8, should be applied to actual technology development programs and projects.

Acknowledgments

The research described in this paper was carried out at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under a contract with NASA. If was sponsored by the Radioisotope Power Systems (RPS) Program Office at NASA John H Glenn Research Center (GRC). The authors wish to acknowledge the support and encouragement of several colleagues, including: John F. Stocky of JPL who was instrumental in initiating this research; June F. Zakrajsek, the RPS Program Planning and Assessment Manager; Ronald T. Reeve, JPL's Nuclear Space Power Office Manager; Paul Ostdiek of The Johns Hopkins University Applied Physics Laboratory; Edward Zampino from GRC; and Robert B. Gounley of JPL.

References

1. Smith, C. A. et al, "Goal Based Testing: A Risk Informed Process", Radioisotope Power Systems (RPS) Program report RPS-RPT-0104, September 2012.
2. Wasserman, G. S., *Reliability Verification, Testing and Analysis in Engineering Design*, Marcal Dekker, Inc., New York, NY, 2003, pp. 203-239.
3. MIL-HDBK-781A, "Handbook for Reliability Test Methods, Plans, and Environments for Engineering, Development Qualification, and Production," DoD 1996, pp. 37-52.
4. Abernathy, R., The New Weibull Handbook, 5th. ed., SAE Publications, Warrendale, PA, 1993, pp. 6-1 - 6-23.
5. NASA/SP-2010-580, "NASA System Safety Handbook," 2011.
6. NASA/SP-2011-3422, "NASA Risk Management Handbook," 2011.
7. NASA/SP-2007-6105, "NASA System Engineering Handbook," 2007.
8. Everline, C. J., "Bayesian Approach to Quantifying Epistemic Uncertainty in a Processor Availability Model," *Journal of Spacecraft and Rockets*, Vol. 49, No. 6, November-December 2012, pp. 1029-1031.
9. OSNP-3, Rev. 1, "Reliability Program Requirements for Space and Terrestrial Power Systems," DOE, 2010.
10. MIL-STD-785B, "Reliability Program for Systems and Equipment Development and Production," DoD 1980.
11. NASA NPR 8705.2B, "Human-Rating Requirements for Space Systems," 2008.
12. Lakey, P. B. and Neufelder, A. M., "System and Software Reliability Assurance Notebook," Rome Laboratory, RL-TR-97-XX, 1997.
13. Nakagawa, Y. and Nakashlinia, K., "A Heuristic Method for Determining Optimal Reliability Allocation," *IEEE Transactions on Reliability*, Vol. R-26, No. 3, August 1977, pp. 156-161.
14. Krevor, Z. C., A Methodology to Link Cost and Reliability for Launch Vehicle Design, Ph.D. Dissertation, School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, August 2007.

15. Tillman, F. A., Hwang, C.-L., and Kuo, W., "Optimization Techniques for System Reliability with Redundancy-A Review," *IEEE Transactions on Reliability*, Vol. R-26, No. 3, August 1977, pp. 148-155.

16. Yalaoui, A., Chu, C. and Chaˆtelet, E., "Reliability allocation problem in a series–parallel system," *Reliability Engineering and System Safety*, Vol. 90, 2005, pp 55–61.

17. Elegbede, A. O. C., Chu, C., Adjallah, K. H., and Yalaoui, F., "Reliability Allocation Through Cost Minimization," *IEEE Transactions on Reliability*, Vol. 52, No. 1, March 2003, pp. 106-111.

18. Malaiya, Y. K., "Reliability Allocation," *Encyclopedia of Statistics in Quality and Reliability*, edited by F. Ruggeri, R. S. Kenett, and F. W. Faltin, John Wiley and Sons, Inc., 2007, pp. 1-5.

19. Alexander, A. J., "The Costs and Benefits of Reliability in Military Equipment," The RAND Corporation, Santa Monica, CA, 1988.

20. Cataldo, R. and Bennett, G., "U.S. Space Radioisotope Power Systems and Applications: Past, Present and Future," in *Radioisotopes: Applications in Physical Sciences*, S. Nirmal, ed., InTech 2011, Chap. 22.

21. Rhatigan, J. L. ed., *Constellation Program Lessons Learned*, NASA/SP-2011-6127-Vol-2, 2011, pp. 63-64.