RL-TR-93-249
In-House Report
December 1993

ACCELERATED RELIABILITY TESTING UTILIZING DESIGN OF EXPERIMENTS

Barry T. McKinney

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Materiel Command
Griffiss Air Force Base, New York

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

Although this report references limited documents * on page 131, no limited information has been extracted.

RL-TR-93-249 has been reviewed and is approved for publication.

APPROVED: Edward J. Jones

EDWARD J. JONES, Acting Chief
Systems Reliability Division
Electromagnetics & Reliability Directorate

FOR THE COMMANDER: John J. Bart

JOHN J. BART
Chief Scientist
Electromagnetics & Reliability Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1993	In-House May 93 - Sep 93

4. TITLE AND SUBTITLE
ACCELERATED RELIABILITY TESTING UTILIZING DESIGN OF EXPERIMENTS

5. FUNDING NUMBERS
PE - 62702F
PR - 2338
TA - 02
WU - TK

6. AUTHOR(S)
Barry T. McKinney

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Rome Laboratory (ERSR)
525 Brooks Road
Griffiss AFB NY 13441-4505

8. PERFORMING ORGANIZATION REPORT NUMBER
RL-TR-93-249

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Rome Laboratory (ERSR)
525 Brooks Road
Griffiss AFB NY 13441-4505

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Barry T. McKinney/ERSR (315) 330-2608

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report documents a system-level Accelerated Reliability testing methodology. The method requires no specific assumptions of a Time-to-Failure distribution nor a stress/performance model.

The methodology results in a multi-stress environmental test based on Design of Experiments, specifically a one-third fractional factorial design. The test data are modeled by the method of orthogonal polynomials.

Although the data are collected in a high-stress environment, an operational performance estimate can be established without extrapolating beyond the test data limits.

14. SUBJECT TERMS
Accelerated, High Stress, Combined Environment Reliability Test, Design of Experiments

15. NUMBER OF PAGES
150

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	U/L

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std. Z39-18
298-102

EXECUTIVE SUMMARY

In the early 1950s and 1960s accelerated testing of military hardware (characterized by extreme levels of stress) was primarily focused at the part-level. The stress/performance relationships developed at that time were predicated on demonstrated theory and utilized to extrapolate extremely high-stress test data to operational parameters. As weapon designs became more sophisticated, the need for system-level accelerated reliability testing arose.

System-level testing has advanced to fairly efficient multi-stress environmental tests; Unfortunately, traditional accelerating techniques have proven inappropriate. The part-level theories do not apply to the higher levels of assembly.

Several system-level accelerated testing techniques have been proposed, however, most methods require assumptions of a specific time-to-failure distribution and a stress/performance relationship function and, generally extrapolate well beyond test data limits. Very often these assumptions are unfounded which, when combined with a lengthy data extrapolation, lead to very questionable results.

By incorporating Designed Experimentation, this research demonstrates a combined environment, system-level reliability test technique requiring as little as 30% of the test time expected for standard MIL-HDBK-781 test plans. It is further demonstrated that the accelerating properties of high stress environmental tests can be effectively modeled without any specific assumptions and, performance predictions for operational levels of stress can be made without extrapolating beyond the test data. In addition, by utilizing the method of orthogonal polynomials, each stress included in the combined environment test can be individually modeled, providing enormously valuable information to those responsible for the system development.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

α	probability of rejecting a true null hypothesis
A_i	effect of factor A
A_L	linear effect of factor A
A_Q	quadratic effect of factor A
β	Weibull shape parameter
β_0	grand mean
B_j	effect of factor B
B_L	linear effect of factor B
B_Q	quadratic effect of factor B
C_k	effect of factor C
C_L	linear effect of factor C
C_Q	quadratic effect of factor C
δ	Weibull shape parameter
δ'	estimate of δ
Γ	Gamma function
Γ_i	resistance parameter for stress environment i
e_{ijk}	error
eV	electron volt

E_i	exponent appearing in *ith* factor of defining contrast
γ	Weibull location parameter
γ_1	coefficient of skewness
γ_2	coefficient of excess
γ_1'	coefficient of skewness for a distribution of means
γ_2'	coefficient of excess for a distribution of means
G_{rms}	G-force root mean square
ξ_j'	polynomial of degree j
λ_1	term associated with first order orthogonal polynomial
λ_2	term associated with second order orthogonal polynomial
σ	standard deviation
Ω	Ohms
n	sample size
$t_{h,m,l}$	minimum unit-time for the high, medium, and low test cells
μ	expected response, mean
μ_3	third moment of the population distribution
μ_4	forth moment of the population distribution
u	coded stress level of factor
U_j	polynomial term

Chapter I

INTRODUCTION

A. Preliminary

Significant advances in the materials, designs, and manufacturing processes of modern weapons have elevated the performance of today's systems to levels unimaginable a few years ago. With the advent of new technologies, such as photonic microprocessors, there is virtually no limit to the performance and reliability of tomorrow's weapons. However, due to cost and schedule constraints, the rapid advance of weapons' performance has left in its wake virtually no acceptable means of quantifying reliability parameters. Traditional testing methods, originally established to qualify or verify reliability, have become unrealistic due to their limited predictive properties and due to the demands of rapid development schedules.

Consider, for example, a high reliability component that has an estimated Mean-Time-Between-Failure (MTBF) of 5,000 hours. A traditional military oriented reliability test would require a minimum of 8,600 unit-hours of failure free operation in a combined stress environment. If failures occur, required test hours can easily double. In addition, current test methods do not provide or recommend convenient, or even practical, techniques to predict the effects that individual test stresses have on the unit's performance.

Industry and the Department of Defense (DoD) have recognized these deficiencies and are beginning to acknowledge that reliability testing is more than just a step in the acquisition process. Testing has become a very involved discipline, requiring serious considerations. It has become apparent that organizations involved with weapon system acquisitions are seeking a universally robust test process, characterized by a short test envelope while providing an efficient method of quantifying reliability and performance. For today's systems, this must be a highly integrated effort, possessing three very important attributes:

1. Test capabilities must be independent of the test article.
2. Test parameters must be economically acceptable.
3. Test results must be statistically valid and traceable.

1. Combined Stress Environment

Environmental stresses acting on electronic components contribute to the degradation of reliability. This phenomenon is recognized by both industry and government. Many experimenters have attempted to capture these relationships in closed form mathematical models. Bazu and Tazlauanu [1] document a "generalized" form of the widely accepted Arrhenius equation (which they have used to describe the relationships between unit reliability and a varying number of stress factors). Also discussed in their work is how this model corresponds with previously developed stress relationships, including the Hakim-Reich, Lawson, and Peck models. Models of this type, however, are generally limited to part-level devices at extreme levels of stress, and therefore have not been overly successful in predicting the operational reliability of the higher level assembly. While component manufacturers generally test for and understand the impacts of stresses at their extreme levels, few understand the relationship between stresses and reliability at operating levels.

In a few instances, the effects of individual stresses are specifically quantified; this is generally limited to very unique or specialized applications and device types. Most practitioners simply develop combined stress/reliability models via a linear regression of the natural log of failure data collected on parts exposed to elevated levels of stress. The assumption is that the stresses, most often temperature [2], increase the rate of failure by a linear constant. This assumption, therefore, allows for a means to gather "accurate" failure characteristics in a short amount of time. This gives rise to the term *accelerated testing*. To estimate the component's reliability, experimenters use a straight line fit of the high stress test data, then simply project the regression line to the nominal or operational stress environment.

For anything other than the simplest devices, current accelerated test methods do not accurately translate the high stress test data to operating reliability parameters. This is due, in part, to the reckless development and utilization of the acceleration model. For part-level testing, when studying single modes of failure, theoretical models such as the Arrhenius relationship may adequately fit the data and provide for some legitimate extrapolation. However, as the test articles become more complex, the number of unique failure modes increases. This threatens and often violates the theoretical premises of the assumed model. Experimenters often neglect the related theory and resume testing as usual [1].

Stress related performance thresholds inherently suggest a non-linear relationship between stress and unit reliability. This non-linearity could create considerable error when experimenters attempt to force the data into a linear form. Further, by utilizing an extrapolated "best fit" line, the models predict the reliability of the test article at levels of stress *below* which the data for the model was collected and do not necessarily represent accurate or legitimate estimates (Figure 1.1).

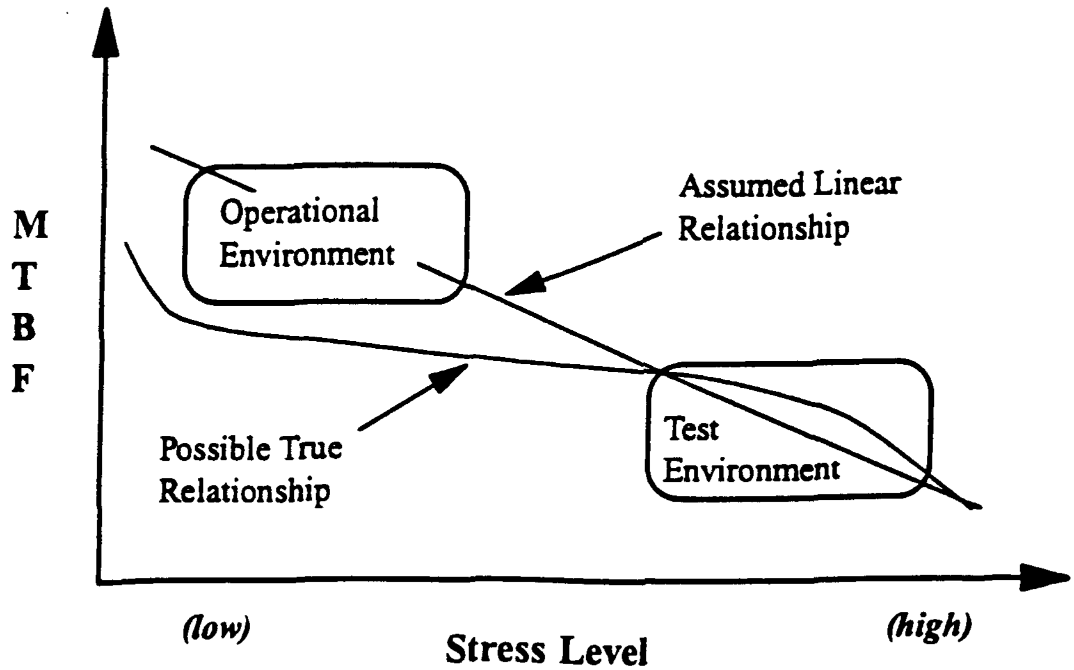


Figure 1.1 Common Accelerated Tests.

Utilization of model coefficients outside the stress range upon which they were calculated can lead to erroneous conclusions. For instance, if the true relationship were, in fact, non-linear and the range of the dependent variables was unknowingly limited to part of the curve which was somewhat linear, predictions outside this range could be disastrous (Figure 1.1). To insure the legitimacy of the stress model, it is imperative that its utilization be limited to the range of the test data.

Another significant deficiency of many accelerated reliability test methods (as well as traditional sequential testing), is their inability to identify the individual stresses that significantly affect the test article. It is not uncommon for a test procedure to include humidity and voltage stress simultaneously with temperature and vibration. Yet, it is typically assumed that the elevated temperatures and levels of vibration are the main factors degrading the unit's reliability [2]. The impact of the individual stresses on the failure rate are rarely studied. Therefore, they are generally misunderstood. Consequently, the focus of this effort is the development of the fundamental knowledge concerning each stress's contribution to the degradation of a unit's performance.

The method of testing developed and proposed herein provides an economically feasible and efficient test methodology which quantifies the test article's reliability and performance, as well as the effects that each stress has on the estimate.

2. TQM, Statistics, and DOE

This research is a product of an emerging philosophy throughout the DoD and contractor community. Total Quality Management (TQM) has evolved into a contemporary new approach, a philosophy securely built upon a statistical foundation. TQM, however, is nothing more than an educated and efficient means of management, which, in many cases, simply means doing more with less. Statistical tools decades old are just now becoming important to corporate America. Threatening foreign and domestic competition is forcing a new way of thinking throughout entire organizations.

Statistics, often thought of as a decision making tool *in the light of uncertainty* [3], has become a major thrust worldwide. One of the most resourceful and efficient tools is the Design of Experiments or DOE. Design of Experiments is by no means new. Nearly a century old, DOE was first developed by R.A. Fisher [4] in his studies of agriculture. Fisher perfected scientific shortcuts for data analysis to identify cause-and-effect relationships. Fisher's developments (designed experiments) are the key to this proposed methodology of accelerated reliability testing.

The method of testing to be discussed and developed in this research is unique in six ways:

a. The method utilizes a combined stress environment for the specific purpose of modeling all effects, not just temperature.

b. The stress range of the test overlaps the operational environment of the test unit as opposed to being well above.

c. There are no assumptions concerning the specific shape of the life distributions or stress relationship functions.

d. The method requires no extrapolation.

e. The method has been specifically developed to test and model all levels of assembly.

f. The method employs the efficiency of Designed Experimentation as a contributing "accelerator" of the test.

B. Literature Search

The discipline of reliability engineering slowly developed after World War II (sometimes referred to as "The Wizard War") and was formally recognized by the agencies of the Federal Government on 7 December 1950 when the Advisory Group of Electronic Equipment (AGREE) was established. Early emphasis of the group was focused on reliability enhancements to the vacuum tube (the Wizard). Reliability demonstration, testing, and prediction were not immediate concerns. However, published works from this new scientific discipline *were* treating the issues of test and prediction.

In 1951, W. Weibull published "A Statistical Distribution Function of Wide Applicability" in *The Journal of Applied Mechanics* [5]. Weibull's density function can be applied to nearly all facets of reliability and maintainability engineering, and is highly significant to the development of this proposed methodology. Also in that year, Epstein and Sobel delivered the paper, "Life Testing," to *The Journal of the American Statistical Association* [6]. These works were followed in 1954 by "Truncated Life Tests in the Exponential Case," also by Epstein, published in *The Annuals of Mathematical Statistics* [7]. The latter has become the basis for the standard system-level reliability test used today. Reliability prediction techniques, however, were not widely addressed until 1956. In November of that year, RCA published TR-1100, "Reliability Stress Analysis for Electronic Equipment" [8], the pioneering predecessor of the universally applied MIL-HDBK-217, "Reliability Prediction of Electronic Equipment" [9].

The AGREE report [10], published on 4 June 1957, recognized the importance of test and predictions. Soundly covering all relevant issues, the document represents the birth of reliability engineering. The report covered all aspects of reliability including: minimum performance guidelines, reliability allocation, associated cost benefits, and the effects of dormancy. Most importantly, though, it delivered to the armed services the assurance that reliability could be specified, demonstrated, and predicted.

1. Reliability Testing/Demonstration

The earliest forms of system-level reliability testing were bench type operational tests that employed the methods developed by Epstein. Unit reliability was established as a function of failures versus operating time. In the mid 1950s and 1960s, due to the poor reliability of the early electronics, methods of this type were acceptable with respect to cost and schedule. Lengthy tests were not required to establish and verify the unit's Mean-Time-Between-Failure (MTBF). Most of the advances of early reliability testing were focused on part-level devices.

A myriad of papers, specifications, and military standards seemingly flooded the industry in the late 1950s and early 1960s. Engineers were busy enhancing and testing devices such as tubes, motors, relays, semiconductors, and numerous other parts. The DoD was also involved with the part-level reliability movement, publishing documents covering all aspects of reliability. Specifications and standards were available on nearly all devices, addressing minimum performance levels, the management of the reliability effort, reliability assurance procedures, reliability test methods, and reliability sampling plans for parts. Eventually testing and prediction at the system-level, or at least at the sub-assembly level, became the focus. The sequential testing methods developed by Epstein were being proved inadequate, and were slowly losing ground to the relatively efficient testing methods for parts.

2. System-Level Concern

In 1973 Lt. Col. Ben Swett completed an Air Force staff study on reliability, investigating the poor correlation between original factory demonstrated reliability (bench testing) and that observed in the field. He concluded that, by altering the traditional sequential tests to include stresses, system-level reliability testing could better represent an operational environment. Lt. Col. Swett's recommendation was implemented in 1977. By combining the environmental tests of MIL-STD-810 with the reliability test plans of MIL-HDBK-781, the Space and Naval Warfare Systems Command published what is now the current Military Standard for reliability testing, MIL-HDBK-781C, "Reliability Test Methods, Plans, and Environments for Engineering Development, Qualification, and Production."

Within the guidelines of 781C, units are tested over a vast range of various stresses, better representing the operational environment. Although much more realistic, the minimum times required for the prescribed tests are lengthy and very demanding. For the standard test plans, the minimum test times, in multiples of the minimally acceptable MTBF, range from 2.67 to 4.40 hours, during which time there can be no failures. In the event that one or more failures occur, test times can easily double or triple. For a system with a 5000 hour MTBF, this translates into test times of 13,350 - 22,000 unit-hours minimum (18 to 30 months). Further, the *expected* test times, again in multiples of the minimally acceptable

MTBF, range from 3.42 to 11.4 unit-hours. Although 781C is still considered a contemporary means to verify reliability, assembly and system-level testing have lost considerable ground to the parts testing techniques.

During the later years of the reliability boom, as the technologies matured and unit life times grew, part-level accelerated reliability testing philosophies emerged. Theoretical stress/performance relationships were (and still are) utilized as a means to compress test schedules. The models allowed for the interpretation (i.e., justified extrapolation) of high stress test data. In contrast, efforts to develop accelerated testing methods for higher levels of assembly were short lived. In the early 1960s Rome Air Development Center experimented with assembly-level accelerated tests with components of the 412L Aircraft Warning and Control System. Early demonstrations appeared promising, yet did not prove to be successful. For the more complex assemblies, the part-level accelerated methods were inadequate.

Nelson [11] developed one of the simplest (subsequently the most useful) of the early accelerated test methods, still effectively used today for part-level screening and verification. Building on the assumption that the occurrence of normal part failures can be hurried (or accelerated) by elevated levels of stress, Nelson devised a technique that utilizes two graphical plots for the data analysis and modeling: a part-life distribution plot and a theoretical stress/relationship plot. Using techniques of this type, test analysis can be performed very quickly and easily.

The general procedure for Nelson's part-level tests is to first identify the stress (typically one) that most dramatically affects the component's reliability. In nearly all cases, temperature is selected. Following the determination of the test stress, a maximum upper bound of the stress is calculated (many times, it is simply a guess). This upper bound is the point at which, if exceeded, abnormal modes of failure would occur. These failures would not occur at use levels of stress. The upper bound identified for tests of this nature are quite often as much as five times higher than normal operating levels. Nelson recommends testing at two additional points, at levels below the upper bound but well above the use environment. A number of units are then tested at each level until there are enough failures to estimate the reliability at those stress levels. The failure times for each stress level are plotted on probability paper corresponding to the assumed life distribution of the parts. The 50% points of each data set are then transferred to an assumed relationship graph (generally plotted on Arrhenius paper). A line drawn through these points is then projected down to the operating levels of stress.

Although not generally recognized, the very high stress methods of reliability testing require at least four assumptions pertaining to just the modeling of the data:

a. The time to failure distribution of the units is known and is the same for all levels of stress.

b. The translation function of the high stress data is theoretically applicable.

c. The parameters of the translation function remain unchanged through all test levels.

d. The failures being modeled at the extreme levels of stress are the same as those occurring at operational levels.

Under Nelson's approach, test levels for parts began to soar (sometimes well exceeding practical limits), allowing experimenters to gather failure data in a very short amount of time. Although legitimate in many cases, this approach has been plagued with abuse [11].

To compress today's demanding schedules, some experimenters (neglecting the earlier attempts of Rome Air Development Center) apply the part-level theories and procedures to test assembly units in a similar fashion: well above the operational design limitations. In many cases, the actual failures being modeled physically could not occur at normal levels of stress, thereby leading to a completely invalid analysis. Further, the limits of the relationship model, be it the Arrhenius model or an inverse power function, are often neglected. It is not uncommon to see the utilization of particular theoretical models outside the premises of the related theory. The Arrhenius model is a prime example. Originally developed to model temperature dependent chemical reactions, this relationship is frequently used for failure models far removed from any possible temperature/chemical relation [1]. Some experimenters utilize these models in an empirical fashion, applying them simply because they fit the data, not necessarily the theories. At best, extrapolation based on solid theory is a good guess. But to utilize a model simply because it fits the data, and not the underlying theory, is inviting disaster. If any of these abuses have occurred, lengthy extrapolation of the data will compound the associated errors.

The specific development of system-level accelerated reliability testing was all but abandoned until the early '80s, when authors such as Derringer and Cassady [12], and Pollack and Mazzuchi [13], proposed testing methods. There are two basic techniques for assembly and system level accelerated testing: step stress testing, and fixed or constant stress testing.

Pollack and Mazzuchi proposed a step stress technique that incorporates a Bayesian analysis methodology. Under their step stress approach, assumed values of the conditional success probabilities must be estimated for each successive step of the test. These values are estimated by the utilization of classical Bayesian analysis. The *a priori* survival estimates required for this analysis are obtained through MIL-HDBK-217. Including the additional estimate of "one's strength of belief," a variance component, this method requires three estimated parameters for each test cell. The general procedure is to test at a given level of stress for a specified time, then "step" up to the next higher level. The stepping is continued until there is adequate failure data for modeling and estimation purposes. The technique has shown some promise; however, it has yet to be validated.

The other approach to assembly level accelerated testing is a constant stress test. The general procedure for this approach is to test at two or three constant levels of stress for a specific amount of time, or until there is an adequate number of failures for modeling purposes.

Both of these techniques have, unfortunately, been developed for levels of stress beyond the unit's operational environment, and therefore require some extrapolation. Pollack utilizes a Bayesian technique, while others revert back to a linear model for simplified extrapolation.

Derringer, employing the constant stress approach, does not hesitate to extrapolate, and has developed an expression for an "acceptable" range of the ensuing extrapolation as a function of the variance about the linear model.

The method of assembly and system-level testing presented in this research removes the seemingly necessary extrapolation. An important step for the development of a robust technique is to eliminate or minimize the number of assumptions and estimates. This can be accomplished by bringing the test levels of stresses down to more realistic or common levels while preserving the accelerated properties.

Chapter II

PROBLEM FORMULATION

A. Preliminary

Acquisition professionals are seeking a universally robust reliability test methodology for the assembly and system levels. The following chapters develop and demonstrate a reliability test method specifically designed for the higher order assemblies. The most pronounced contributions of this proposed method are its capability to quantitatively partition the individual stress effects of those factors commonly included in a combined environment test, and the ability to predict the unit's reliability and performance without extrapolating beyond the limits of the accelerated test data.

This methodology will be most beneficial if the testing is performed during the early stages of the design effort. The early identification of deficiencies allows the efficient development of system enhancements. These enhancements may not only affect the design of the system or assembly, but also the design of manufacturing processes and possibly the actual operation of the system. This testing technique can be applied to specifically explore the impact of competing design materials and various manufacturing processes, as well as to improve the operational and maintenance practices.

B. Problem Formulation

The proposed methodology was realized by the utilization of Designed Experiments. Fisher's achievements were predicated on the theory of linear models. Fisher's classical expression, utilized to describe the effects of the factors studied, has given way to the more general expression adapted by Searle [14]:

$$Y_{ij\ldots} = \mu + A_i + B_j + \ldots + e_{ij} \tag{2.1}$$

where

Y_{ij} = dependent variable

μ = expected response

A_i = effect on Y_{ij} from factor A

B_j = effect on Y_{ij} from factor B

e_{ij} = error

i,j,. = levels of factors A, B, ..

This linear expression can be further expanded to explore non-linear responses by individually considering the higher order effects of the factors studied. Depending on the number of test levels of each factor, the main effects can be partitioned so that the linear, quadratic, and cubic effects can be investigated. The effects (A_i, B_j, etc.) are tested in a classical sense, including a Null Hypothesis, according to a standard Analysis-of-Variance (ANOVA). The effects found significant are then modeled. The modeling technique utilized for this effort is the method of orthogonal polynomials.

Orthogonal polynomials are among the simplest curve fitting techniques. The methodology is well documented and therefore will not be developed here. The advantages of this method become far greater as the degree of the polynomial increases. Since the terms are orthogonal, higher order terms can simply be added, independent of those already considered. The addition of terms ends with the highest degree polynomial that shows significance.

Hicks [3] gives a clear discussion on the method of orthogonal polynomials. Further development and examples can also be found in Davies [15], and Draper and Smith [16]. In addition, Fisher and Yates [17] also give an indepth discussion as well as tables of the required terms.

The basic procedure is to consider the expression

$$Y'_u = U_0\xi'_0 + U_1\xi'_1 + \ldots . \tag{2.2}$$

where each ξ_j' is a polynomial of degree j, orthogonal to all other ξ's. The subscript value of Y′, u, is a coded term equal to

$$u = (X_j - \overline{X}) / I \tag{2.3}$$

where I is the interval width of the test variable X. The U_j's are given by

$$U_j = \frac{\sum_{i,j} Y_{ij}\xi'_j}{\sum_{i,j} (\xi'_j)^2} \tag{2.4}$$

which reduces to

$$U_j = \frac{Contrast}{n * \sum_{j'} (\xi'_j)^2} \tag{2.5}$$

The ξ_j' equations are straight forward. For the first three values, they are:

$$\xi'_0 = 1 \tag{2.6}$$

$$\xi'_1 = \lambda_1 u \tag{2.7}$$

$$\xi'_2 = \lambda_2 [u^2 - (\frac{k^2-1}{12})] \tag{2.8}$$

where k is the number of levels of the test factor. The λ's were developed so that the ξ'_j values are integers for the u values associated with the test.

Therefore, the development of the polynomials reduces to a simple matter of adding the linear terms. Once each of the individual stress terms has been developed, simply adding all of the terms to a single mean value renders the overall stress model.

1. Assumptions

The assumptions necessary for this testing and modeling methodology do not deviate appreciably from the assumptions of common reliability testing techniques. By not requiring specific assumptions of a time to failure distribution and a stress/performance relationship function, the assumptions required for this methodology are considerably less restrictive. They include:

a. The factors being studied are quantitative, that is, they can be described as points on a scale.

b. The interactions are negligible (discussed in detail later).

c. The factors can be equally spaced from one level to the next.

d. The errors are independent and normally distributed with mean zero and common variance σ^2; i.e., $N(0,\sigma^2)$.

e. The design limits of the test article can be determined (or approximated).

f. Multiple, identical units are available for test.

g. The test stresses can be applied simultaneously.

2. Relevant Issues

The specific concerns addressed in the following chapter, Proposed Methodology, are integral to all testing techniques. Each section of the Development provides a clear discussion of the issue at hand. The topics follow a logical progression through the development of this research. The issues discussed include:

a. The performance criterion for the unit.

b. The stresses being studied.

c. The levels of stress for the test.

d. The effects of non-normality on the ANOVA.

e. The number of test units required for the various test conditions (cells).

f. The repair and reuse of the test articles.

g. The time required for each test condition.

h. Analysis and Modeling.

i. The consumer and producer risks of the test.

j. An economic analysis of the projected test.

k. Missing data.

l. The limitations of the methodology.

m. The expected benefits of the methodology subject to the limitations.

It is demonstrated in the remaining chapters that the specific application of designed experimentation allows for the effective and much more efficient testing and modeling of system-level reliability.

Chapter III

PROPOSED METHODOLOGY

A. Preliminary

Traditional system-level reliability testing, as discussed in the previous chapters, is aimed at verifying an MTBF, generally a specified contract value. The process, therefore, is not employed to disclose the unknown; it is more an exercise to demonstrate contract compliance. However, advancements in hardware technologies have all but outdated traditional reliability testing. The advances in materials, designs, and manufacturing processes have slowly driven performance and reliability into the realm of the unknown. The time between failures of modern, high reliability electronics is so great that traditional reliability testing is simply not feasible in terms of both time and money. Reliability testing must now determine the true merits of modern, highly complex military electronics.

Reliability testing for modern systems is becoming more challenging and sophisticated. Verification is being superseded by quantification; the goal of testing today, and particularly this research, is to efficiently estimate the reliability and performance of the test unit within operational parameters by testing at elevated levels of stress.

B. The Methodology

The experimental process can be broken down into three phases [3]: the planning phase, the design phase, and the analysis phase. The planning phase is very likely the most important.

Nelson [11] acknowledges that few accelerated tests are statistically planned, resulting in, at the very minimum, less accurate information and, sometimes, no information at all. Nelson also points out that most books, even his own, overemphasize data analysis.

> *"... Few books devote enough attention to test planning, which is more important. For most well planned tests, the conclusions are apparent without fancy data analysis. And the fanciest data analysis often cannot salvage a poorly planned test." [6, pp. 210]*

The planning phase includes the determination of the performance parameter(s) of interest, the types and levels of stress used for the test, as well as the analysis technique used to study the test data. The design phase addresses the determination of the type of experimental design most suitable and efficient for the specific purposes of studying reliability, the amount of time and number of test units that will be required, and a simple tradeoff analysis between test time and the number of test units. The analysis phase of the method identifies and quantifies the demonstrated effects of stress.

The experimental process is employed for the purpose of discovering undisclosed truths. It is therefore necessary for the experimenter to be clear as to what is sought and to understand the experiment. Haphazard approaches can be costly, inefficient, and, in many cases, fruitless. For this effort, the response or dependent variable of interest will be the unit's performance, as discussed below.

1. Test Planning

a. What to Measure

The most important facet of test planning is determining what is sought and how "it" is to be measured. Performance can be measured hundreds of ways. For very simple devices, such as light bulbs, performance can be easily considered "Go/No Go". It is a simple matter of determining whether or not the unit functions. For more complex systems, measuring performance is not always this easy. Many systems are multi-functional. Considering a car stereo, the unit may be designed to receive AM and FM stations as well as play cassette tapes. If the entire unit does not work, there has obviously been a failure. But what if the situation were different? The AM reception is fine, but on hot days FM is *noisy*, or the tapedeck operates fine on smooth roads but skips on rough roads. Are these failures or not? Further, can performance of this nature be defined and classified, let alone quantified? As systems become more complex, so does defining and measuring performance.

For the purposes of this effort it is imperative that a metric of performance be established prior to testing. In addition, it is assumed that the performance of the test units can not only be defined but can also be effectively measured within the test environment. Performance, with respect to reliability, is generally defined as the number of hours prior to a failure. However, the method can also efficiently study more physical parameters such as power or accuracy. The "acceleration", in a classical sense, will apply to the MTBF measurements, while the testing of the physical parameters will be accelerated by the efficiency of DOE.

b. Identify Stresses

Chapter I introduced the concept of accelerated testing as being characterized by exposing the test unit to at least one stress at levels higher than would occur during normal use. Stresses, environmental and physical, typically include temperature, humidity, vibration, voltage, shock, and pressure. Therefore, as part of the test planning, the experimenter needs to identify:

1. The stresses the unit is exposed to under use conditions.
2. Which of those stresses most probably degrade the unit's performance.
3. Which of the identified stresses can be effectively controlled under a test environment.

Almost exclusively, temperature is the single stress chosen for conventional accelerated reliability testing.

Most large scale test facilities have adequate equipment to stress nearly any device, large or small, in a multitude of combined environments. However, small manufacturers (the targeted customer of this research) are generally limited to thermal chambers, vibration tables, and voltage supplies. Assuming that three stresses can be controlled for the testing, the next step is the determination of the various test levels.

c. Stress Levels

The proposed research was originally conceived to address pure reliability testing. However, it became apparent that the methodology can also be applied to common parametric testing as well. For this reason, there exists a need to investigate the levels of stresses for both types of testing.

The range of the various stress levels is chosen (or designed) so that the failures occurring during the test are of the same nature as those seen or expected at normal operating levels. To assure failures of this nature, the test levels will range from operational conditions to the maximum design limits (Figure 3.1). The maximum design limit is the point at which, if exceeded, the occurrence of abnormal failure modes would be expected. Because the stress range of the test overlaps the operational environment, the accelerating properties of the proposed methodology will not be driven by consistently high levels of stress. A combined multi-stress environment will be utilized to achieve the desired level of acceleration. There are three important aspects pertaining to the combined test and the stress levels:

1. The overall range of the test stresses and the number of test levels.
2. The consideration of the eventual customer.
3. The process being formulated as a traditional designed experiment.

The range of the test stresses is important for a number of reasons, not the least important of which is that failures can be propagated unique only to very high levels of stress. If the failures modeled are not those seen in the field under normal use, it is quite obvious that the testing is of little value. It is important to emphasize that for accurate and legitimate modeling, stress levels must be near or overlapping the operating range of the unit. In addition, by overlapping the operational stress range of the unit, the methodology does not require any extrapolation for an operational reliability estimate and thus permits the use of an empirical stress/performance model as opposed to a theoretical model. This eliminates common theoretical assumptions. Relative aspects of the specific stress test levels are equally important.

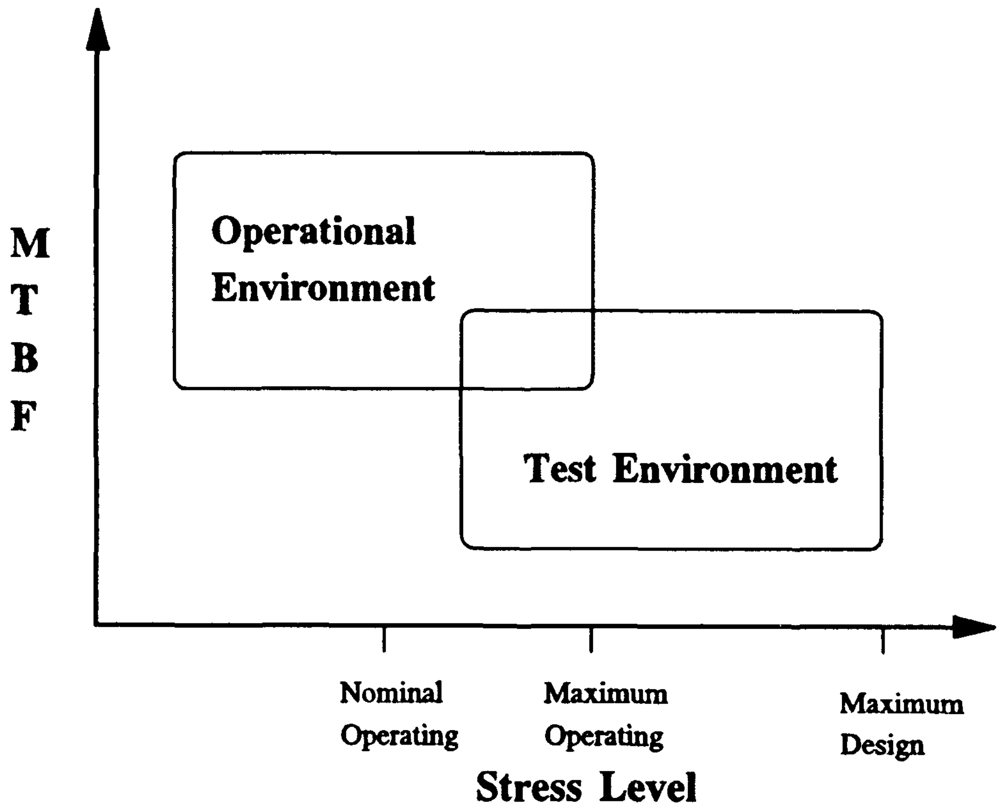


Figure 3.1. Proposed Stress Range.

The amount of data collected at any given level of the test is a function of the number of failures at that level. Recognizing the limitations of time and test units, it is desirable to maximize the expected number of failures at each level by maximizing the number of test units at each level. Stipulating fewer stress levels (some methods have proposed as many as ten test levels, [13]) ensures more test units per level, thereby increasing the expected data per cell. However, assigning too few levels, only one or two, will limit the predictive capability of the data analysis to a point estimate (similar to traditional methods) or a linear function. By considering three levels of stress, the possibility of a non-linear stress/performance relationship can be investigated and modeled. For a three level test, the ensuing analysis will be limited to a quadratic model. However, weighing the impact on the number of test units per level (data) and the additional number of estimated parameters for higher order models, the limitation to a simple curve does not appear to be costly. Considering the trade off between data and the number of test levels, as well as the anticipated benefits of a combined environment test, three levels of each test stress will be shown to be adequate.

The second important aspect of defining stress levels for accelerated reliability testing is that the results, inevitably, need to be "sold" to someone at a later date. Data collected at very high levels of stress, well beyond any conceivable operating range, is a tough sale. Customers are reluctant to believe the "black magic" a statistician can perform with, what may appear to be, irrelevant data. It is therefore imperative that the ranges of the test stresses are legitimate with respect to failure modes as well as acceptable to the customer.

The third consideration addresses a concern typically discussed when suggesting lower stress levels for accelerated testing. It is generally perceived that the lower stress range will slow down the accelerating properties of the stress test. The uniqueness overlooked is the contribution derived from the concurrent application of Design-of-Experiments: The accelerating properties of the proposed method will be preserved, and possibly enhanced, by the combined, multi-stress environment.

As stated previously, this method can also be applied to study the impacts of operational stresses on physical performance parameters. These may include items such as output power, speed, and accuracy. For experimenters interested in these types of measures, the stress levels chosen for the test may range from the minimum operational stress levels to the maximum operating stress levels.

2. Design Phase

It is important that the design phase remain free of typical accelerated testing shortcomings, such as assuming a specific time-to-failure distribution, assuming a stress/performance relationship function (e.g., the Arrhenius relationship), and utilizing lengthy extrapolations. A Design-of-Experiments approach, using a combined stress environment will avoid these pitfalls.

Traditional methods for studying the effects of various factors on a response follow a conservative approach. A typical scenario would see a test process systematically step through a one-at-a-time change in each factor through all levels. The tests are generally initiated at the lowest levels and proceed through to the maximum, quite a different process from a designed experiment.

A fundamental characteristic of DOE is the use of randomization. That is, the order of the trials is determined by a random process. The randomness of a properly executed experiment contributes significant credibility to the data analysis. The effects of various biases can be reduced through randomization. For example, consider an experiment studying two circuit card soldering processes. If five circuit cards are soldered by one process followed by five cards soldered by a second process, and there occurred a general drift in the solder temperature, it may appear that one process was superior to the other when, in fact, the difference was caused by the change in solder temperature. Random ordering of the trials allows time trends to average out. In addition, randomness supports assumptions regarding the independence of various errors, particularly measurement error.

DOE randomness assures that test cells will be examined, without bias, where one stress factor may be at an operational level while the others are at the maximum design level. This allows for data to be collected at lower levels of one stress while higher levels of the other stresses are accelerating the occurrence of failures. This methodology, therefore, also relies on high levels of stress to drive failures, yet provides data collection at operational stress levels (which eliminates an entirely unfounded extrapolation). In addition, the mere efficiency of DOE can be considered an accelerating property of this method. To avoid testing at the lowest stress conditions (where failures are unlikely to occur), only a subset of the possible combinations that excludes the lowest stress cells is used for the test.

To demonstrate the efficiency of DOE, consider a simple situation: designers are concerned with the weight of a diskdrive in a militarized system. There are two drives available for the system that seem quite comparable; the decision will be based on the weight of the units (Figure 3.2). The proposed test employs a conventional method: each diskdrive is weighed twice (A_1, A_2 and B_1, B_2), then the average weight is calculated. Thus, four individual runs (or experimental trials) are required, resulting in the following estimate of the average weight of each diskdrive.

Average weight A = $(A_1 + A_2)/2$

Average weight B = $(B_1 + B_2)/2$

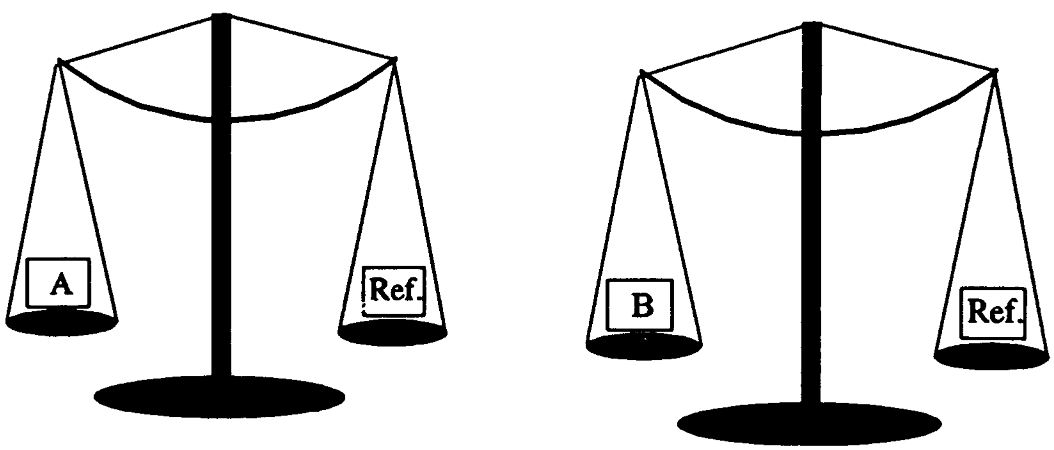
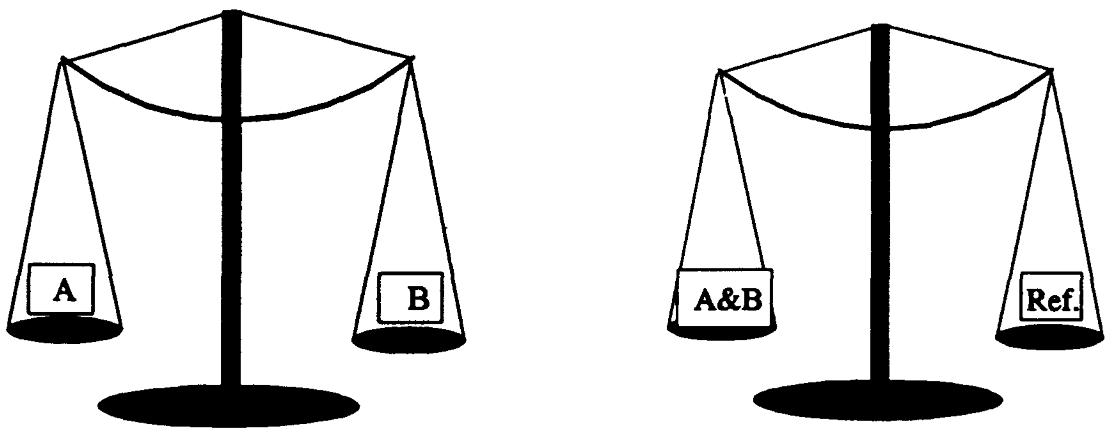


Figure 3.2a. Proposed Method.

Consider an alternative utilizing the design principles of DOE, specifically, orthogonal contrasts. One experimental run would establish the difference between A and B, and a second run would sum their weights.



(a) Run 1 (b) Run 2

Figure 3.2b. DOE Method.

As with the proposed methodology, each unit is weighed twice. The average unit weight can then be calculated as:

$$\text{Average weight A} = ((A-B)+(A+B))/2 \ ;$$
$$= (A_1+A_2)/2$$
$$\text{Average weight B} = ((A+B)-(A-B))/2 \ ;$$
$$= (B_1+B_2)/2$$

where (A-B) represents the first run, and (A+B) represents the second run. The results are identical, having the same accuracy and number of independent weighings of each diskdrive, yet the DOE approach required half as many trials. Test reductions of this nature are common, and in fact, become far more pronounced as the number of factors increase.

a. Test Design

The experimental design developed for this effort is a one-third replicate of a 3^f factorial having three test stresses each at three levels. This design will require nine test cells. Table 3.1 illustrates one of the replicates developed for this research.

The block, or replicate selection for the reliability testing was constrained to minimize the number of low stress cells. From Table 3.1 it is seen that the stress combinations involving the lowest levels were specifically avoided. Data are a function of failures, which, in turn, are a function of stress; therefore, in an attempt to maximize data, stress is maximized where possible. By avoiding the lowest stress levels, it is possible to maximize the expected amount of test data (contrary to traditional reason, it is desirable to witness failures during reliability quantification testing).

The blocks were determined by confounding all possible defining contrasts. The defining contrast is an expression stating which effects are to be confounded with the blocks [3]. Confounding of effects is necessary for fractional designs.

Table 3.1. One-Third Fractional Factorial.

	Volt.								
	1.5eV			2.0eV			2.5eV		
	Vib (grm)			*Vib (grm)*			*Vib (grm)*		
Temp	2	4	6	2	4	6	2	4	6
80°C			■		■		■		
120°C		■		■					■
160°C	■					■		■	

By considering the I and J components of the two-way interactions and the W, X, Y, and Z components of the three-way interactions, there exist 13 effects that could be utilized as the defining contrast. These components of the interactions have no physical significance, yet do prove useful in complex designs [3]. Representing the three stresses as A, B, and C, the contrasts become: A, B, C, AB, AB^2, AC, AC^2, BC, BC^2, ABC, AB^2C, ABC^2, and AB^2C^2, where AB and AB^2 represent the J and I components of the AB interaction. By using a linear relationship developed by Kempthorne [18], the determination of the blocks was straight forward.

$$L=E_1X_1+E_2X_2+...+E_iX_i \tag{3.1}$$

In Equation (3.1), E_i is the exponent appearing on the *i*th factor of the defining contrast, and X_i is the level of the *i*th factor (0, 1, or 2 representing the low, medium, and high levels, respectively) for a given treatment combination. Using this technique, all treatment combinations with the same L value, modulus 3, are placed in the same block. For a three level test, there are three possible L values: 0, 1, and 2. For example, if the defining contrast were AB^2C, the L value for treatment combination 012 (stress A at low, stress B at medium, and stress C at high), is

$$\begin{aligned} L &= 1*0+2*1+1*2 \\ &= 4 \\ &= 1(\textit{modulus } 3) \end{aligned}$$

Therefore, it was merely a task of considering all possible defining contrasts, calculating an L value for each treatment combination, then placing the same values in the same block.

Following the determination of the blocks, it was essential that the aliases be calculated and examined for reasonableness. Because only a fraction of the complete factorial is executed, the main effects and the interactions cannot be estimated independently. The situation then arises that an estimate of a required effect also estimates one or more other effects. Effects that have the same numerical value are called aliases. The aliases for the replicates are determined by multiplying the effects by the defining contrast, modulus 3. Because the design is a one-third replicate, there are two aliases for each effect. Therefore, to find the second alias, the first alias is multiplied by the defining contrast (or the effect multiplied by the defining contrast squared). For the above block, having A, B, and C represent the stresses, the aliases are:

$A*(ABC)$	$= AB^2C^2$	
$A*(ABC)^2$	$= BC$	A or (AB^2C^2 or BC)
$B*(ABC)$	$= AB^2C$	
$B*(ABC)^2$	$= AC$	B or (AB^2C or AC)
$C*(ABC)$	$= ABC^2$	
$C*(ABC)^2$	$= AB$	C or (ABC^2 or AB)
$AB^2*(ABC)$	$= AC^2$	
$AB^2*(ABC)^2$	$= BC^2$	AB^2 or (AC^2 or BC^2)

From these calculations, it is seen that the main effects are aliased with the two-way and the three-way interactions. Since main effects are not aliases with each other, this pattern is considered acceptable, assuming the effects of interaction are negligible. The other blocks for the method were established in the same fashion. There are twelve blocks that have acceptable alias patterns (Appendix A), four of which do not include the lowest stress cells (Blocks 3, 5, 9, 11). One block is randomly selected when testing is initiated. For the physical performance testing, all blocks are subject to the random selection; for reliability testing, only the four blocks that do not include the low stress cells are considered.

The assumption of negligible interaction must be considered. This assumption is common in nearly all reliability testing scenarios. However, the wide spread commonness is not in itself justification for its inclusion in this methodology.

As discussed in Chapter II, this testing technique is being developed for application during the design and development phase of the acquisition process as well as for final demonstration/validation. Recognizing the vast amount of unattainable data during the early stage, it would be a major accomplishment to identify and individually model even a few of the stresses affecting system reliability. Further, considering the hundreds of possible factors working against a fielded military system, as well as the fact that numerous case studies [19] have concluded that as many as 50% of all failures are either false alarms, RETOKs (retest OK), CNDs (can not duplicate), or maintenance induced, the quantification of the two- and three-way interactions is, for all practical purposes, not considered worthy of the required investments of time and money. In addition, designers would face an enormous task to understand the interaction, particularly the three-way, let alone make the design changes necessary to effectively eliminate their detrimental effects. Also, it is presumed that the interaction, if present, is not completely unpredictable. It is reasonable to assume that the interaction among stresses is somewhat additive or synergistic in nature. That is, if the individual effects of two stresses tend to decrease unit performance, then their interaction would further contribute to the demise of the unit. Although this may not be the case, considering the time and cost of quantification, as well as the small value added, this methodology assumes test stress interaction is negligible.

b. Test Units

This methodology was developed with the specific goal of being widely applicable, appropriate for part, package, circuit card, subassembly, and system-level testing.

Therefore, the task was not to identify a test unit, but instead, to determine the number of units required for the test. It was previously mentioned that test data are a function of failures. Generally, the expected amount of test data will be directly proportional to the number of units placed on test at various stress levels. The limiting case, of course, is the result of the tradeoff between budgeted test time and the number of test units. As will be later demonstrated for reliability measures, the more units the less test time, and conversely, the less units the more test time. Therefore, a logical starting point is to establish the minimum number of test units under the assumption that there is no limitation on time.

The test data is analyzed according to a standard Analysis-of-Variance, or ANOVA. A premise of ANOVA is that the observations are normally distributed [3, 15]. This requirement stems from the fact that the standard tables of various distributions, including the t, χ^2, and F, are developed based on normal theories. Therefore, to assure the accuracy of the percentile calculations, it is required that the test data be drawn from a normal population. To assure normality, and to avoid assumptions of a specific shape of a unit's life distribution, mean values of performance at each test combination will be collected.

According to the Central Limit Theorem (CLT), a distribution of sample means approaches normality as the sample size increases. Therefore, by collecting data on the means, point estimates can be established without knowing the exact shape of the parent distribution.

For experimenters interested in physical characteristics of their system, the mean performance will be assumed normal and calculated in the conventional arithmetic fashion. However, for reliability testing, the possibility that the units on test may not fail (i.e., provide sample data) must be considered. In this case, an estimate other than the arithmetic mean must be utilized. Considering the expression for the mean, or expected value, of a continuous random variable (time to failure), Equation 3.2,

$$E(X) = \int x \cdot f(x) dx \tag{3.2}$$

suggests that a specific probability density function (pdf) must be assumed. According to the CLT, the central tendency of a distribution of means is a function of the parent distribution as well as sample size. Thus, there is a lower bound on the sample size for each distribution for which the assumption of normality is appropriate.

Most reliability practitioners assume that systems fail according to the exponential probability density (failures are independent of time). This assumption is generally made for the convenience of the calculations rather than being suggested by the data. Some experts, Nelson, *et al*, have speculated that only 15% of systems fail exponentially, while most demonstrate some time dependency on system degradation. The necessary specifics concerning these types of assumptions can be avoided by considering the Weibull density function.

The Weibull distribution is defined by a three parameter probability density function (pdf) that can approximate the shape of many continuous pdfs (Figure 3.3). In fact, the Weibull reduces to the exponential and reduces to the Rayleigh [5] distributions with certain parametric values. In addition, Nelson has shown nearly identical results comparing analysis using the Weibull and lognormal densities. The Weibull density function is given as

$$f(t)=\frac{\beta}{\delta}\left(\frac{t-\gamma}{\delta}\right)^{\beta-1}e^{-\left(\frac{t-\gamma}{\delta}\right)^{\beta}} \tag{3.3}$$

where,

γ is the location parameter ($-\infty < \gamma < \infty$)

β is the shape parameter ($\beta > 0$)

δ is the scale parameter ($\delta > 0$).

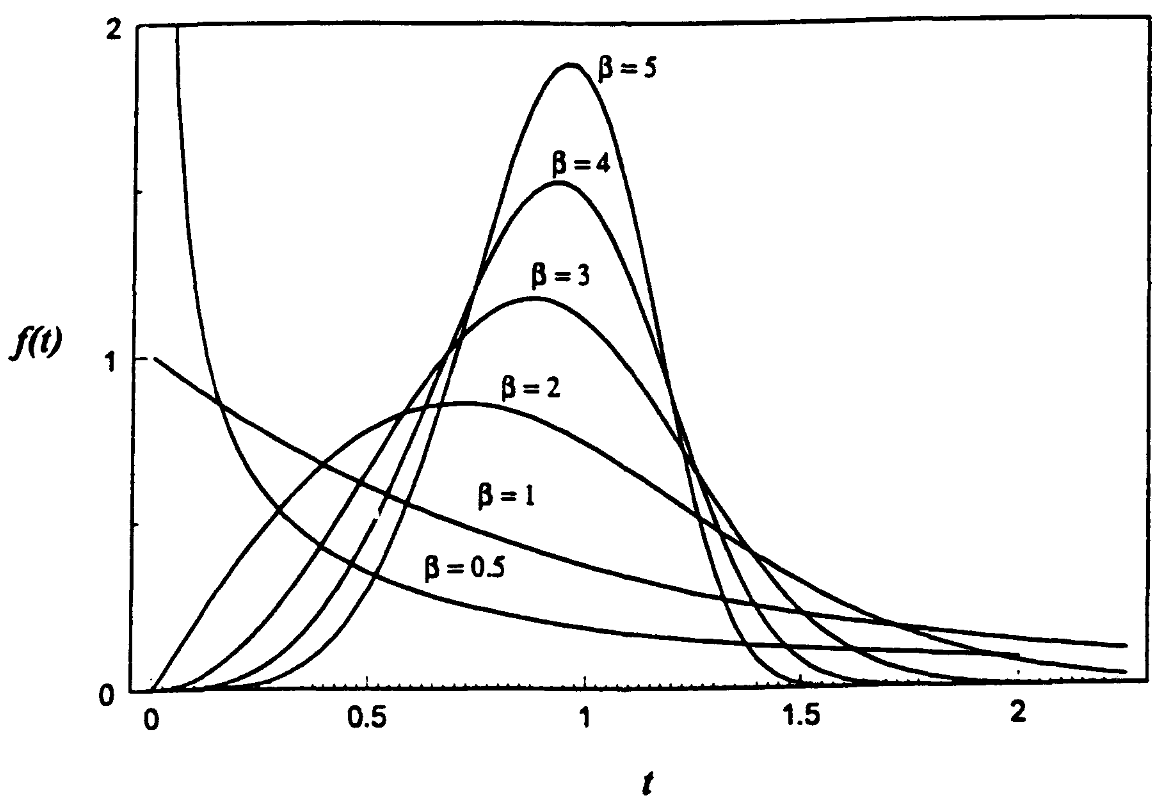


Figure 3.3. Weibull Density For γ=0, δ=1, and β=0.5, 1, 2, 3, 4, 5.

The location parameter, γ, for this effort will be set to zero as it represents the first point at which the probability of a failure is greater than zero. For the purpose of reliability testing it is recognized that failures can, in fact, occur at the instant the testing begins.

Beta (β) is the shape parameter of the function, and has been shown to be dependent on the system being studied. For electronics, β is generally in the range 0.5 - 5.0. The actual β value (i.e., shape of the distribution) will not inhibit the ability to estimate the means in any way. A range of the estimated MTBF can be established for likely β values. The Weibull can take on virtually an infinite number of shapes as a function of β, closely approximating nearly all applicable distributions. Values of β less than 1.0 coincide with a decreasing unit failure rate as a function of time. With the exception of infant mortality failures, increasing unit performance as a function of time is an unlikely situation. Figure 3.3 depicts the extreme flexibility of the Weibull density function.

Delta (δ), the scale parameter, is the characteristic life, and is always the 63.2 percentile [11]. That is, 63.2% of the failures will occur at a time $\leq \delta$. Therefore, without assuming a specific shape of the unit's failure distribution, the characteristic life of the unit can be estimated. The value of δ could be estimated by tabulating the cumulative percentile of failed units; however, for this research, δ is conservatively estimated by setting β equal to 1, giving:

$$f(t)=\frac{1}{\delta}e^{-(\frac{t}{\delta})} \tag{3.4}$$

By using the method of maximum likelihood, an estimate of δ can be established in terms of time and failures. The general expression is

$$L=\prod_{i=1}^{n}\frac{1}{\delta}e^{-(\frac{t_i}{\delta})} \tag{3.5}$$

However, the form for the likelihood function for censored data (not all units have failed) becomes

$$L=\prod_{i=1}^{r}f(t_i)\cdot\prod_{j=r}^{n}R(t) \tag{3.6}$$

where R(t) is the reliability function, n is the total number of units on test, r is the number of failed units, t_i is the failure times of the failed units, and t is the fixed time for the test cell. Equation 3.6 reduces to

$$L=(\frac{1}{\delta})^r\cdot e^{-\frac{1}{\delta}(\sum_{i=1}^{r}t_i)}\cdot e^{-\frac{(n-r)t}{\delta}} \tag{3.7}$$

Taking the natural logarithm of both sides gives

$$lnL=-r\cdot\ln(\delta)-\frac{1}{\delta}\sum_{i=1}^{r}t_i-\frac{(n-r)t}{\delta} \tag{3.8}$$

Taking the partial derivative with respect to δ, setting to zero and solving, gives

$$\hat{\delta} = \frac{\sum_{i=1}^{r} t_i + (n-r)t}{r} \tag{3.9}$$

Inspecting the numerator, it is clear that this term is simply the total amount of unit test time. Consequently, it is not necessary for all units to fail to find the maximum likelihood estimate of δ. Estimating δ for a β of one results in a conservative value. If the true β were greater than one, the resulting δ estimate would represent a value less than the 63.2%. Hence, a conservative estimate. In the event that there are no failures during a particular test condition, a χ^2 estimate of δ can be established (this will be addressed later). Once δ is estimated, a range on the expected value, or the mean time to failure, can be developed as a function of β (Table 3.2). For γ equal to zero, it can be shown that the expected value of a Weibull distributed variate is [20]

$$E(t) = \delta\Gamma(1 + 1/\beta) \tag{3.10}$$

Table 3.2. β Versus E(t) as Multiples of δ.

BETA	*E(t)*
$\beta = 1/3$	$6\ \delta$
$\beta = 1/2$	$2\ \delta$
$\beta = 1$	δ
$\beta = 2$	$0.886\ \delta$
$\beta = 3$	$0.893\ \delta$
$\beta = 4$	$0.906\ \delta$
$\beta = \infty$	δ

Having established the expected value (or mean) of the Weibull distribution in terms of δ and β, the minimum sample size can be determined. As previously stated, the central tendency of a distribution of means is dependent on the sample size and the parent distribution. For "well-behaved" parent distributions, samples of three or four have been shown to adequately approximate the normal [20]. The task was to determine the meaning of "well-behaved", and "adequately approximate" in terms of the impact on the ANOVA. Both of these questions can be addressed by considering the values γ_1 and γ_2, the coefficients of skewness and excess (or kurtosis), respectively.

An investigation of the sensitivity of the ANOVA restriction on normality resolves the minimum sample size issue. The worst case Weibull shape is used to determine a minimum sample. By minimizing β, the most "ill-behaved" or non-normal shape can be generated. Considering the lowest reasonable value of β results in the exponential distribution. The values of γ_1 and γ_2 are given as [15]:

$$\gamma_1 = \mu_3 / \sigma^3 \tag{3.11}$$

$$\gamma_2 = \mu_4 / \sigma^4 - 3 \tag{3.12}$$

where μ_3 and μ_4 are the third and fourth moments of the parent distribution. For the exponential these values can be obtained by differentiating the moment generating function.

$$M(t)=(1-t\delta)^{-1} \tag{3.13}$$

$$M'''(t)=6\delta^3(1-t\delta)^{-4} \tag{3.14}$$

$$M''''(t)=24\delta^4(1-t\delta)^{-5} \tag{3.15}$$

which, when evaluated for t=0 gives

$$\mu_3=6\delta^3 \tag{3.16}$$

$$\mu_4=24\delta^4 \tag{3.17}$$

Since σ (the standard deviation) for the exponential distribution is δ, the calculations for γ_1 and γ_2 yield

$$\gamma_1=6$$

$$\gamma_2=21$$

For random samples of size n, Davies [15] defines the values γ_1' and γ_2' as the measures of skewness and excess for the distribution of the means. These values are approximated by

$$\gamma_1'=\gamma_1/\sqrt{n} \tag{3.18}$$

$$\gamma_2'=\gamma_2/n \tag{3.19}$$

which, for the exponential gives

$$\gamma_1' = 6/\sqrt{n} \tag{3.20}$$

$$\gamma_2' = 21/n \tag{3.21}$$

It is now a matter of investigating the impact of non-normality on the F-test of the ANOVA. By considering various values of γ_1' and γ_2', the actual significance levels of the F-test can be compared to the values based on true normality. Gayen [20] gives an example demonstrating the impact of non-normality on the F-test. At the five percent normal theory significance points, Gayen illustrates the true significance levels of a variance ratio test of five groups of five observations. Having four and 20 degrees of freedom, Table 3.3 illustrates the actual significance levels as a function of γ_1' and γ_2'. The column headings, 0.0, 1.0, and 2.0, are the values of $\gamma_1'^2$. The row values, 0.5, 1.5, and 2.5, are the values of γ_2'.

Table 3.3. Effects of Non-Normality on ANOVA. Significance Levels Associated with 5% Normal Theory Values for γ_1' and γ_2' [20].

		$\gamma_1'^2$		
		0.0	1.0	2.0
γ_2'	0.5	4.88	4.98	5.07
	1.5	4.64	4.74	4.83
	2.5	4.40	4.50	4.59

Gayen indicates that even extreme non-normality has little effect on the F-test [20]. For this research, the impact on the F-test was investigated for samples as small as three. Depending on the degrees of freedom left for the error, the true significance levels associated with the five percent normal theory values were found to range from 5.0% to 11.7% for the worst case Weibull shape. Therefore, allowing a maximum deviation of five percent (this will be addressed later), if an analysis demonstrates a likely chance of a low value of β (1 or 2) the minimum number of units per cell is four. Higher β values require a minimum of three units.

To determine a reasonable range for β, the complexity of the test unit and severity of the normal use environment must be considered. The more complex the unit and the more severe the operational environment, the more likely a lower β value. As systems become larger, it is more likely that failures will appear to occur randomly. The individual components of a large system will probably not fail exponentially; but as part counts creep into the hundreds and thousands, the mass jumble of combined failure rates tends to become constant, independent of time [21]. The same situation holds true for the operational environment of the unit. The more severe the environment, the more difficult it is to predict. Therefore, when considering β, it is recommended that a value of one only be considered for systems that have historically been shown to fail randomly (if the data exists) or if the test unit is fairly complex and designed for a hostile environment. From MIL-STD-781 unit complexity has been loosely defined as follows:

100 parts- Simple

500 parts- Moderately Complex

2000 parts- Complex

4000 parts- Very Complex

The final consideration to determine the *total* number of test units must account for their repair and maintenance, not an easy task.

The repairability of the test units as well as the effectiveness of repairs will play a significant role in the determination of the total number of units required for the test. If the units are non-repairable, the problem is somewhat simplified.

For the so called "one shot" devices, the range on the minimum number of units required for the entire test would be three to twenty-seven (assuming a minimum of three units per test condition) depending on the number of failures and the effects of operating time on non-failed units. That is, if all units failed, then obviously 27 units would be the minimum required for the test. If no units failed and there was no evidence of accumulated wear or damage to the units as a result of the testing, then the same three units could be used for all test conditions. Obviously these extremes are unlikely. The problem arises when unfailed units must be evaluated for possible continued use at successive test cells, which forces consideration of test order. If the lower stress cells were run first, then it would be reasonable to assume a higher probability of some test articles reuse at the expense of losing randomness. The situation is tempting. However, for this research the tradeoff between test randomness and test-unit reuse was not considered. This returns the concern back to the effects of accumulated test time on the units. This analysis will be wholly the responsibility of the test engineers on an individual basis. If it is concluded that cumulative test time has not detracted from the performance of unfailed units, then obviously the reuse of the test article should be exercised.

The same type of consideration and failure analysis should be exercised for repairable units. The effects of accumulated test time *and* the effectiveness of the repair and maintenance activities must be considered. Again, this evaluation will be left to the test engineers.

c. Test Time

Having established the minimum number of test units per cell to be either three or four, an evaluation of the required time per cell can be made. For this evaluation an initial "paper" estimate of the system's MTBF must be calculated. This is a common practice throughout the industry, and there exist numerous methods for calculating preliminary MTBF estimates. One of the most widely utilized methods is Military Handbook 217F, "Reliability Prediction of Electronic Equipment". Users establish a preliminary estimate of the unit's performance so that a "ball park" estimate of the maximum test time can be generated. This is quite necessary for the purposes of a preliminary economic analysis of the resulting test plan.

It is important to note that a starting point must be at least estimated. If a 1,000,000-hour system (114 year MTBF) were being studied, would reliability testing add any value to the system's development? On the other hand, for a fairly poor system, efforts and funds could probably be better spent on obvious improvements rather than on testing. In addition, for any realistic economic analysis of the test plan, preliminary estimates of the required test time must be established.

A primary objective of this research is to develop a test method that can specifically quantify the performance relation for each test stress. However, it was assumed that, in the event that very few or no failures occur during a test, a validation of the unit's estimated MTBF will be satisfactory. Based on the objectives of the test engineers, a definite end point of the test can be established; however, not all the benefits of this methodology may be realized. This situation may occur in the event that the system's performance is far superior to that expected.

The approximate test time for each cell of the designed experiment is based on the preliminary estimate of the MTBF. The time estimate for each test condition is made by considering a relatively high probability of a failure. The most skewed shape of the Weibull density will give the maximum time estimate (i.e., conservative time estimate). Therefore, the Weibull density will be evaluated for a β value of 1. By utilizing an estimate of δ, a nominal environment test time will be approximated that coincides with a reasonable chance of observing a failure. It is desirable to witness failures for the purposes of establishing a "hard" estimate of unit performance; however, it is not essential. This will be discussed later. The nominal environment time estimate is made as follows:

For the Weibull density function having a β equal to 1

$$F(t)=1-e^{(\frac{1}{\delta}*T)} \tag{3.22}$$

where T is equal to total unit time under test, which is equivalent to the product of the trial test time, t, and the number of units on test, n, which gives

$$F(t)=1-e^{(\frac{1}{\delta}\cdot n\cdot t)} \tag{3.23}$$

Assuming that δ can be approximated by the estimated MTBF, δ', gives

$$F(t)=1-e^{-(\frac{1}{\delta'}\cdot n\cdot t)} \tag{3.24}$$

Solving this equation for t gives

$$t=\frac{\delta'}{-n}\ln(1-F(t)) \tag{3.25}$$

Considering a 70% chance of witnessing a failure to be reasonably high, t can be estimated directly from Equation 3.25.

$$t=\frac{\delta'}{-n}\ln(1-.7) \tag{3.26}$$

Therefore, according to Equation 3.26, for a system having an estimated MTBF of 2000 hours, tested with three units per trial, the nominal (or operational) environment test condition would require approximately,

$$=802 \; hours$$

Recognizing the probability of failure for each of the three units as being

$$p(failure) = 1 - e^{-(t/\theta)} \tag{3.27}$$

the expected number of failures for the test condition becomes

$$E(x) = 3 * (1 - e^{-(802/2000)}) \tag{3.28}$$

$$= 1$$

If four units were placed on test, the resulting test time for the nominal environment would be 601 hours. This value represents the maximum amount of time expected to elapse while providing a single failure. Therefore, it should also be considered an upper bound for all stress environment test conditions.

Having established the maximum test time for each cell by examining the nominal environment (Equation 3.26), the opposite extreme will be investigated to determine the minimum test time for each condition.

For the purposes of life testing, hours collected at a more extreme test environment would be equivalent to at least that many hours at a less severe condition. Therefore, to calculate a minimum test time per trial, it is assumed that stresses have no effect on the reliability of the test unit, and all test hours can be considered equivalent to hours at an operational environment. Since there are nine test conditions, Equation 3.26 becomes

$$t = \frac{\delta'}{-(n*9\ trials)} \ln(.3) \tag{3.29}$$

The rationale behind this calculation follows. If the units do not fail at the highest degrees of stress, failure at the lower stress conditions will not occur, thereby collecting a minimum of (t*n*9) unit-hours. This results in an expected number of failed units approximately equal to one, allowing for at least a single legitimate point estimate in the event the stresses, in fact, have no effect on the unit. With a minimum of three units per trial and a 2000-hour system, t becomes 89 hours (67 hours with four units).

Based on these maximum and minimum bounds, established for an operating (nominal) environment, time estimates for the actual test conditions (all of which are more severe) are approximated. It should be emphasized that testing at each cell can be terminated the instant the first legitimate failure occurs. The bounds established here represent the maximum trial test time in the event of no failures.

To partition the range of the trial test times given by Equations 3.26 and 3.29 into reasonable test times for the various test conditions, an investigation of the stress at each test condition is required. Of the overall block previously specified (Table 3.1), three of the test conditions involve two of the three stresses at their upper extreme. For three other trials, one stress is at the upper extreme, and for the remaining three trials, two of the three stresses are at a levels well above the nominal. It is clear that all of the test cells are, in fact, well in excess of the nominal environment, two-thirds of which are at the maximum design limit for at least one stress. The nine test conditions are equally divided into three stress categories: high, medium, and low.

Due to the vast diversity and unique performance characteristics of modern equipment, the determination of which test conditions represent the various degrees of stress (within the selected block) will be left to those applying the method. This determination may be unique to each test unit due to the possibility of a dominating stress.

Recognizing the fact that the stresses may *not* affect the test units, the time estimates from Equation 3.29 should be specified as the maximum test time for the three most extreme conditions. The test times for the mid and lower level conditions are established by considering the possibility that the higher stress cells may have been much too harsh of an environment to collect reasonable data, thereby effectively reducing the experiment by three trials. Therefore, the calculation for the mid level conditions recognizes (t*n*6) as the maximum possible number of test hours. Substituting six trials into Equation 3.29 results in 100 hours for the three mid level test conditions, while the lowest three trials yield 200 hours. If no failures occur throughout the entire test, the total amount of test time is approximately 2.2 times the estimated MTBF (δ'). At the very worst, this method compares favorably to the traditional MIL-HDBK-781 test plans.

Having established which trials of the experiment represent the high, medium, and low stress conditions, the equations for the individual trial times are:

$$high \quad t = \frac{\delta'}{-(n*9)} \ln(.3) \tag{3.30}$$

$$\textit{medium} \quad t = \frac{\delta'}{-(n*6)} \ln(.3) \tag{3.31}$$

$$\textit{low} \quad t = \frac{\delta'}{-(n*3)} \ln(.3) \tag{3.32}$$

The minimum test time for the standard MIL-HDBK-781 test plans (IID-VIID) is 2.67 times the lower (least acceptable) MTBF; the minimum *expected* test time is 3.42 times the lower MTBF. Therefore, in the event that additional test time appears necessary under the proposed method, the values returned by Equations 3.30 - 3.32 may be increased by a factor of 1.55. This value represents the ratio of the minimum expected time of MIL-HDBK-781 and the maximum (standard) time of this method, 3.42/2.20. This increase will allow the total test time of the proposed method to approach the minimum expected time of the "quickest" MIL-HDBK-781 test plan. This situation may arise in the event that the factors studied are not expected to, necessarily, accelerate the occurrence of failures.

By comparing these maximum time estimates to over 30 years of reliability "corporate knowledge" [9] compiled by Rome Laboratory, it becomes obvious that the proposed method favors the conservative versus aggressive approach (i.e., the time estimates for the individual trials were established for worst case conditions, therefore, they will be higher than the time generally required to produce failures).

Rome Laboratory, since the original RCA TR-1100, has collected and efficiently compiled historical performance data on nearly every electronic part used in military systems. This data has been specifically edited and reduced for direct utilization by engineers as design tools and reliability guidelines. Some of the more valuable information has been included in the publication, "The Rome Laboratory Reliability Engineer's Toolkit" [22].

From topic A11, *Reliability Adjustment Factors* (Appendix B), the historical knowledge clearly supports the conservative nature of this proposed method. Specifically, Tables A11-2 and A11-3 provide the estimated impact on reliability for different environments and the effects of temperature. From Table A11-3, Temperature Conversion Factors, it is estimated the MTBF of a unit is reduced, on average, 20% for the first 10°C increase in temperature, and as much as 60% for a step of 20°C.

Utilizing Table A11-3, not considering the impact of the other stresses, a system having a 2000 hour MTBF at a nominal environment of 40°C would have an 800 hour MTBF at the maximum design limit if it were 80°C. Using the proposed method, with a minimum of three units on test, solving for the most extreme stress conditions gives 89 hours. This results in 802 hours of system time at the extreme in the event of no failures. Based on MIL-HDBK-217, the expected number of failures at this level is approximately one. Similarly, for the mid-level conditions, Equation 3.31 results in 133 hours/trial. Referring to Table A11-3, the MTBF of the system would fall to 1400 hours at 60°C (mid level). These values result with an expected number of failures of approximately one for mid level test cells as well. It is therefore evident that the proposed methodology will initially give conservative test times.

The amount of test time required for conventional, physical performance testing is obviously dependent on the parameters being studied. Due to economic limitations, testing should be minimized, yet not restrict the collection of legitimate measures. This determination is left to the engineers performing the testing.

d. Trade Off Analysis

Knowing the minimum number of units required for the test and the approximated total test time, an economic analysis of the entire testing scenario can be performed. The analysis is fairly straight forward; test time is a function of test units. If the initial test duration is too long, additional units should be added. The most advantageous assignment of additional test units would be to the lowest stress conditions. Two reasons justify this assignment policy: 1) lower stresses contribute to longer test times, and 2) the more data collected at levels closer to the operational environment, the more credible the resulting analysis and predictions. The addition of test units must be restricted to units that are identical in all aspects to those already on test. The number of units on test at the various levels do not need to be equal.

The single most important aspect of this type of testing is an analysis of what can be expected upon the conclusion of the test, and if this possible return is worth the total investment. Simply, how much will it cost? And what can be expected? These concerns must account for the possibility that testing may exhaust the system budget, leaving no resources for design enhancements. In a classical sense, both the consumer's risks as well as the producer's risk should also be considered; unfortunately, the consumer's risk cannot be calculated until after the data is collected.

Once an experimental design is formulated, the next steps are implementation and analysis, followed by the interpretation of the test data.

3. Analysis

Assuming that the objectives of the testing are clear (give hard estimates of the effects of stress or, in the event of no failures, or very few, justify and support the initial estimate), the analysis is straight-forward.

When the test article performs better than expected, prompting few if any failures, the analysis will be elementary. In the event there are two or fewer failures, a simple estimate of the expected time to failure for an appropriate range of β is calculated. By considering the maximum likelihood equation for δ, a point estimate for the characteristic life is calculated as the quotient of total test time at all levels divided by the total number of failures. This value is then utilized in the expected value expression, Equation 3.10, to estimate the MTBF. Results of this nature are by no means the objective of this method; however, the eventuality of no failures must be considered.

The optimal situation would be at least one failure at each test cell, giving a single "hard" point estimate of the unit's performance at that particular combination of stress. For non-failure test cells, Welker [23] indicates that a 50 percent χ^2 confidence limit with two degrees of freedom is an appropriate estimate of the mean.

Since the data are collected as a single replicate of a one-third fractional factorial, there is no independent estimate of the error. Therefore, the error will be estimated with the second order interactions. For the block previously designated (Table 3.1), this becomes

Effect	Degrees of Freedom
A (or AB^2C^2 or BC)	2
B (or AB^2C or AC)	2
C (or ABC^2 or AB)	2
AB^2 (or AC^2 or BC^2) (error estimate)	2

providing an ANOVA model of the form:

$$Y_{ijk}=\mu+A_i+B_j+C_k+e_{ijk} \tag{3.33}$$

Each factor will be tested according to a standard ANOVA, having the Null Hypothesis (H_0) that its effect is zero. The producer's risk, α (the probability of rejecting H_0 when it is true), for the F-test is set at 0.15. This value may seem excessive; however, the factors chosen for the test were specifically assigned because they (presumably) have already demonstrated an effect on the test article. It is the objective of this methodology to model the effects of stress. Therefore, to show any significance with so few data, this level of risk may be required. However, it was recognized that the factors studied simply may not affect the test unit. Prior to considering such a large risk, the penalty incurred for rejecting H_0 when it is true (type I error) was investigated. If H_0 is rejected in error, an additional model term will be developed at a minimum cost of at least one degree of freedom for the confidence bound on the resulting point estimate. This will lead to a wider interval than actually exists. This "price" was not seen as excessive when considering the inherent, limited nature of these types of predictions. The penalty for an error of the second type (failing to reject H_0 when it is false) was also investigated. In this situation, a factor will not be modeled that does, in fact, affect the system. However, the high producer's risk associated with this test virtually eliminates the possibility of not rejecting H_0 if there is any indication of an effect.

Equation 3.33 is investigated further by considering both the linear and quadratic effects of each main stress. This not only identifies the factors significantly affecting system performance, but also prescribes the degree of the resulting model.

Because the main factors are assumed to be independent and without interaction, the individual effects can be modeled separately, then combined in a linear fashion to produce a single stress/performance relationship. This model can describe unit performance as a linear combination of three independent, quadratic equations. Therefore, this combined environment equation not only predicts non-linear performance, but also can individually quantify the contribution of each stress to the performance of the units, quickly leading designers to the most detrimental stresses.

This particular strategy of model development is significantly different from a traditional technique as well as most methods currently utilized for reliability modeling of military hardware. The most common approach to stress modeling is a single linear regression of log-transformed life data. Even if the test included several operational stresses, only one is generally modeled [2, 11].

Each test stress shown to be significant is individually modeled as an orthogonal polynomial. From the onset, the stresses have been assumed to be independent and without interaction. The only additional restriction placed on the overall methodology for the development of the polynomials is that the levels of the stresses being tested are equally spaced. This, of course, is not a burden. The upper and lower bounds of the test have already been identified. The center test condition will simply be midway between the two limits.

The actual data utilized for the ANOVA and model development are the natural logarithm of demonstrated performance. This transformation was chosen for three reasons: 1) a natural logarithm transform is an acceptable technique to remove any variance correlation to the mean (if it may be present) [15], 2) the modeling of natural logarithm transformed data has shown to be more sensitive to the subtle effects that may be present, and 3) a natural logarithm transformation has been a common practice for decades (new methodologies are far better received if they appear to be enhancements to the old methods rather than totally new).

Once each of the factors has been tested for significance, model terms will be developed for those showing an effect. These terms will be combined to estimate the unit's performance at the lowest test condition. Since this point will be near the upper bound of the system's operational performance envelope, it will certainly provide a legitimate worst case performance estimate. Finally, this estimate can be bounded by a confidence interval by utilizing the t-distribution. The general expression for this interval is

$$\hat{Y} \pm (t_{\alpha/2,\nu})*(MSE(\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0))^{.5} \tag{3.34}$$

where MSE is the mean square error from the ANOVA. The matrix notation was used due to the number of cells in the experiment. The matrix **X** represents the independent variables from the test. The vector $\mathbf{x}_0$ represents the values of the independent variables used for the estimate. The value $(\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)$ is a direct function of the geometry of the test matrix. Because the tests were established for equal stress intervals, the test matrices are *rotatable designs*. Rotatable designs have the property that the standard deviation of the fitted value is the same for any given distance from the center [24]. Figure 3.4 illustrates the values of $(\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)$ for the various positions in a one-third replicate of a 3^3 factorial. In the event that a factor does not show significance, the values associated with only two axes are considered. If two factors are non-significant, then only the single dimension array passing through the center is considered.

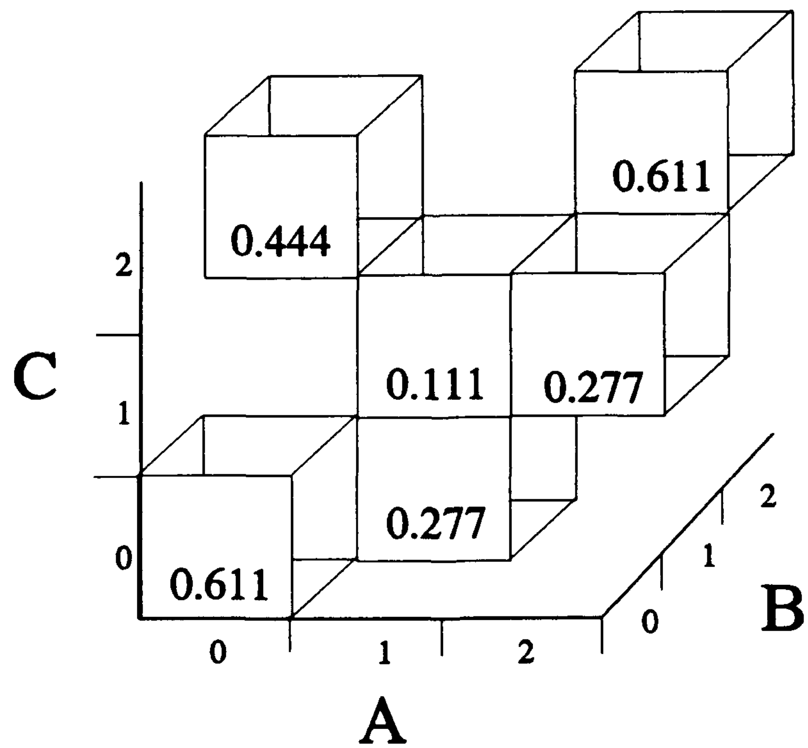


Figure 3.4. Variance Components Relative to Position.

Chapter IV

TEST PROCEDURE

A. Preliminary

Prototype software has been written to accommodate the proposed methodology. The program queries the user concerning the test article and lab/test facility. This input is of a very general nature, including:

1. Device type, complexity, and operational environment.
2. Estimated operational MTBF.
3. Types of test equipment.
4. Unit design limits.
5. Number of units available for test.

1. The Design

a. Reliability Testing

The initial data is used to establish the parameters of the test. The test levels are determined by dividing each stress range, based on the unit's design, into two equal intervals resulting in three test points. This range will span from the operational to the maximum stress environment.

The program then generates the random order of the test. As previously discussed, this involves the random selection of a one-third replicate followed by the random assignment of the nine test cells. Following the basic setup of the design, the software estimates a β value based on the unit complexity and operational environment. As discussed in Chapter III, the more complex the unit and more severe the operational environment, the more likely a low β. This value is presented to the user as a default. Once β is established, the minimum number of test units (assuming no reuse) is calculated, either 27 or 36. The program then calculates the individual trial test times. The test times are established according to Equations 3.30, 3.31, and 3.32 as discussed in Chapter III.

A simple economic (or schedule) analysis is then prepared. The options at this point are limited to the addition of test units to individual cells, reducing the total amount of test time. Once an acceptable test is established, the user implements the resulting test design, recording the total number of unit-hours and failures experienced in each cell. These data are entered back into the program for analysis and model development.

b. Parametric Tests

The procedure for more conventional physical testing requires only three units per cell and differs only in the amount of time required at each test cell. Once adequate measurements have been recorded for a cell, the test should proceed to the next cell. The analysis and model will not be affected.

2. Data Analysis

The analysis begins by calculating the experimental MTBF (or average physical parameter) demonstrated in each test cell. The natural logarithm of this data is subjected to an analysis of variance. The ANOVA identifies which factors had a significant impact on the performance of the test unit. The data supporting the significance of the various stresses is utilized for the development of the orthogonal polynomials. These curves, independently describing each stress/performance relationship, are ultimately combined by adding the model terms to a single mean value, rendering a cumulative stress acceleration model. The final step will be a calculation of the inherent MTBF at operational stress levels.

Test Procedure Example.

Consider a development team anxious to test their device. It was speculated that temperature, vibration, and voltage were the most detrimental factors affecting the unit's reliability. A MIL-HDBK-217 reliability estimate resulted in a 2000-hour MTBF for a worst case operational environment. This value was established for stresses near the maximum operational envelope. The maximum operational stress values were estimated to be 120°F, 3 G_{rms} multi-axis vibration, and 0.5 eV. In a subsequent analysis the designers determined that the maximum design stresses were approximately 220°F, 7 G_{rms} vibration, and 1.5 eV. Beyond these values abnormal failures (those not experienced under normal use) would be expected. Based on these estimates the levels of the test stresses were determined. For temperature, converting to Celsius, the range was established to be from 40° Celsius (104°F) to 100° Celsius (212°F). Therefore the three test levels were established as 40°C, 70°C, and 100°C. For vibration the levels were established as 3 G_{rms}, 5 G_{rms}, and 7 G_{rms}. And for voltage, the test levels established were 0.5 eV, 1.0 eV, and 1.5 eV.

A history and failure analysis of the parts used in the design of the unit suggested that the device would fail according to the normal distribution, that is, dependent on time (a β value greater than one). The software therefore suggested a minimum of three units per cell. However, having an abundance of test articles, the engineers decided on a minimum of four units per test cell.

The methodology developed in Chapter III, particularly the equations for the cell test times, resulted in the following experimental design. Block 3 from Appendix A was randomly chosen and the order assigned. The relative severity of the nine test conditions was then estimated. Utilizing equations 3.30, 3.31, and 3.32 provided the following test times for the trials:

lower stress conditions: 011, 101, 110

$$\text{time} = \ln(0.3)*(2000)/-(4*3)$$
$$= 200 \text{ hours}$$

medium stress conditions: 002, 020, 200

$$\text{time} = \ln(0.3)*(2000)/-(4*6)$$
$$= 100 \text{ hours}$$

highest stress conditions: 122, 212, 221

$$\text{time} = \ln(0.3)*(2000)/-(4*9)$$
$$= 67 \text{ hours}$$

Table 4.1 illustrates the actual test design; the numbers represent the maximum possible unit-hours for the designated cells.

Table 4.1. Test Procedure Example Test Plan.

	***Stress* A**								
	0			1			2		
	Stress B			***Stress B***			***Stress B***		
Str.C	0	1	2	0	1	2	0	1	2
0			400		800		400		
1		800		800					268
2	400					268		268	

Stress A represents temperature, stress B represents voltage, and stress C represents vibration. Each test cell was then run for the maximum time or until the first legitimate failure occurred, resulting in the Table 4.2.

Table 4.2. Test Procedure Example Demonstrated MTBF Values.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0			240		720		360		
1		1151*		1151*					33
2	575*					81		121	

* χ^2 50% point estimates of the MTBF.

Since the testing in each cell was terminated with the first failure, the demonstrated MTBF was simply recorded as the total amount of unit-hours collected. For the three cells that did not produce a failure (*), the following χ^2 , 50% point estimate was utilized to estimate the demonstrated MTBF value:

$$\text{MTBF} = 2(\text{total unit-time})/1.39$$

The natural logarithms of this data were then calculated. These values, depicted in Table 4.3, were utilized for the ANOVA and final model development.

Table 4.3. Natural Logarithm of Test Procedure Example Data.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0			5.480		6.579		5.886		
1		7.048		7.048					3.496
2	6.354					4.394		4.796	

In order to facilitate modeling the data with orthogonal polynomials, the ANOVA was performed by the method of linear contrasts. The procedure and all of the necessary values are well documented [3, 17]. The responses are summed over the low, medium, and high levels of each stress. These values are then multiplied by the corresponding coefficients for the linear and quadratic effects, then summed to give the contrasts as shown in Table 4.4a.

Table 4.4a. Test Procedure Example Contrasts.

Effect	Σ(low)	Σ(med)	Σ(high)			Contrast		
Temp	18.882	18.021	14.178					
Volt	19.288	18.423	13.370					
Vib	17.945	17.592	15.544	$\Sigma(\xi')^2$	λ	A	B	C
linear	-1	0	1	2	1	-4.704	-5.918	-2.401
Quad.	1	-2	1	6	3	-2.982	-4.188	-1.695

The sum of squares is calculated as

$$SS_{effect}=\frac{(contrast)^2}{3*\Sigma(\xi_j')^2} \tag{4.1}$$

The linear and quadratic sum of squares for each of the three stresses were calculated in this fashion, each having a single degree of freedom. The error sum of squares was found by subtracting the sum of squares of each effect from the total sum of squares. The total sum of squares is calculated from the data in Table 4.3 as

$$SS_{total}=\Sigma(response)^2-\frac{(\Sigma response)^2}{9} \tag{4.2}$$

These values are given in Table 4.4b.

Table 4.4b. Test Procedure Example ANOVA F-Tests.

Effect	df	SS	MS	F
$Temp_L$	1	3.688	3.688	40.9*
$Temp_Q$	1	0.494	0.494	5.5*
$Volt_L$	1	5.837	5.837	64.8*
$Volt_Q$	1	0.974	0.974	10.8*
Vib_L	1	0.961	0.961	10.6*
Vib_Q	1	0.159	0.159	1.76
Error	2	0.180	0.09	
Total	8	12.293		

*Significant effects $\alpha=0.15$

This analysis indicates significant factor effects for all three stresses. Further, the quadratic effects of temperature and voltage are also significant. Therefore, the terms modeled were the linear effects of all three stresses and the quadratic effects of temperature and voltage.

The terms for the polynomials were then calculated. Substituting β_0 for the grand mean and the corresponding designator for the stresses (A, B, or C) into Equation 2.2 for the U_j values, the model became:

$$\ln(\hat{Y}) = \beta_0 + A_L\xi'_1 + A_Q\xi'_2 + B_L\xi'_1 + B_Q\xi'_2 + C_L\xi'_1$$

where,

β_0 = grand mean

U_j = (contrast)/n* $\sum (\xi'_j)^2$

$\hat{Y}$ = an estimate of MTBF

These values were calculated as follows:

Stress A:

A_L(temp) = (-4.704)/3*2

= -0.784

A_Q(temp) = (-2.982)/3*6

= -0.165

Stress B:

B_L(volt) = (-5.918)/3*2

= -0.986

B_Q(volt) = (-4.188)/3*6

= -0.232

Stress C:

C_L(vib) = (-2.401)/3*2

= -0.400

From Equation 2.7, the linear terms for the ξ' values are equal to u since λ_1 equals 1.

$$\xi'_1 = \lambda_1 u$$

The quadratic terms for temperature and vibration are calculated by Equation 2.8:

$$\xi'_2 = \lambda_2*[u^2 - (k^2-1)/12]$$

where k is the number of test levels, in this case three. Therefore, ξ'_2 becomes

$$\xi'_2 = 3*(u^2 - 2/3)$$

$$= (3u^2 - 2)$$

For the three test levels (u values -1, 0, 1), the ξ'_2 becomes 1, -2, 1 as presented in the above table of contrasts.

The product of the U_j and ξ'_j terms were summed in a linear fashion to give an overall stress/reliability equation for the test unit:

$$\ln(MTBF) = 5.676 - 0.784u_A - 0.165*(3u_A{}^2-2) - 0.986u_B - 0.232*(3u^2{}_B-2) - 0.4u_C$$

This equation can be used to estimate the performance of the test unit under any combination and level of the test stresses between the maximum and minimum test values. The value of most interest was, of course, the estimated performance of the test article at the lowest combined stress environment, i.e., the worst case operational environment. Using u values of -1, the above cumulative stress model evaluated to:

$$\ln(MTBF) = 5.676 + 0.784 - 0.165 + 0.986 - 0.232 + 0.4$$

$$\ln(MTBF) = 7.449$$

$$MTBF = 1718 \text{ hours}$$

This MTBF estimate, equal to the Weibull δ parameter for β equal to 1, was then utilized in the expected value expression of the Weibull density (Equation 3.10) to calculate MTBF estimates for more likely β values (it was pointed out in the example that the failure mechanism demonstrated some time dependency). The range of the expected MTBF, as a function of β, became:

β	MTBF
0.5	3436 hours
1	1718
2	1522
3	1534
4	1556

Based on β, the MTBF estimate of interest can now be bounded by the desired confidence interval. In accordance with the t-distribution, the confidence bound for observed cells is:

$$\ln(\mathrm{MTBF}) \pm t_{\alpha/2,\nu} * (\mathrm{MSE}(\mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0))^{.5}$$

However, because the stress cell of interest was not one studied during the test, the point estimate is considered to be for a *future* observation. For future observations, the *prediction* interval is [24]:

$$\ln(MTBF) \pm t_{\alpha/2,\nu} * (MSE(1 + \mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0))^{.5}$$

In the above expression, υ, the degrees of freedom, is 3. This value is calculated as (n-p), the number of test points minus the number of estimated parameters. Therefore, if β were assumed equal to 3, an 85% confidence interval would be

$$\ln(1534) +/- (t_{.075,3}) * (0.09(1 + 0.611))^{.5}$$

$$7.336 +/- (1.924) * (0.381)$$

$$6.603 < \ln(MTBF) < 8.069$$

$$738 \text{ hours} < MTBF < 3193 \text{ hours}$$

Considering the conservativeness of the three estimates in the absence of hard failures, the resulting point estimate and prediction interval is considered worthy.

In addition, the nature of the combined-stress acceleration model allows for the exact quantification of each stress' effect on the reliability as the system shifts from one environment to another. This capability is highly unique. For example, if the temperature increased from 40°C to 63°C, and the vibration increased from 3 G_{rms} to 5.5G_{rms} the model becomes:

$$\begin{aligned} \ln(MTBF) &= 5.676 - 0.784(-0.233) - 0.165(3(-0.233)^2-2) - 0.986(0.25) - \\ &\quad 0.232(3(0.25)^2-2) - 0.4(-1) \\ &= 6.736 \end{aligned}$$

This shift in the environment caused approximately a 50% decrease in the MTBF from the first calculation ($\beta=1$). Traditionally, the analyst would recognize the increase in both stresses, yet would model the data as a function of only one stress (typically temperature). Using the proposed method, the exact effect of each stress can be obtained.

By inspecting the stress/acceleration model, the individual effects of each stress can be quantified. The estimate for the operational MTBF was:

$$\ln(\text{MTBF}) = 7.449$$

For the new environment, this became

$$\ln(\text{MTBF}) = 6.736$$

This results in a total decrease of 0.713. A closer examination of the model indicates the individual impacts of temperature and vibration, which can easily be calculated.

For temperature, the net change for both the linear and quadratic terms was:

$$\begin{aligned} \text{operational} &= -0.784(-1) - 0.165(1) \\ &= 0.619 \\ \text{new environment} &= -0.784(-0.233) - 0.165(-1.837) \\ &= 0.486 \end{aligned}$$

This represents a net decrease of 0.133. For vibration the total net change was:

operational $= - 0.986(-1) - 0.232(1)$

$= 0.754$

new environment $= - 0.986(.25) - 0.232(-1.8125)$

$= 0.174$

This represents a net decrease of 0.580. Therefore, the effect of temperature represents 18.65% of the total decrease in the model estimate. The effect of vibration represents 81.35% of the total decrease in the model estimate. Therefore, the direct impact on the MTBF, which fell by 50%, was calculated as:

Temperature impact $= 0.1865(0.5)$

$= 9.33\%$ decrease in the MTBF

Vibration impact $= 0.8135(0.5)$

$= 40.67\%$ decrease in the MTBF

This type of analysis can quickly lead designers to the weaknesses of their systems.

If there is a desire to enhance the results of an initial study, a second replicate can be run in a classical sense. This subsequent study would allow for a more through understanding of the interaction effects as well as provide more degrees of freedom for both the ANOVA F-test and the confidence interval on the point estimate.

Chapter V

METHODOLOGY VERIFICATION

A. Preliminary

Verification and validation are two of the most abused terms in the modeling and simulation arena. The words are quite similar in meaning. However, particularly in reliability testing and analysis, they cannot be interchanged. Verification is associated with the developer of the model and is loosely defined as an exercise that satisfies this person that a model (or method) works as he or she had hoped or anticipated. Validation involves *outside* expertise that, independent of the developer, confirms the model (or method) is accurate and appropriate. In light of these definitions, the example of Chapter IV and the following examples serve as this author's verification.

There are three general methods [24] to validate a model:

1. Collect new data to check the model's predictive ability.
2. Compare results with theoretical expectation, other independent analysis, and simulation results.
3. Use a hold-out sample to check the model and its predictive ability.

The collection of additional data is, by far, the preferred method to validate a model. However, this is often impractical due to limitations of time and money. Because of the demonstrated lack of theory for system-level testing, comparing model results to theoretical expectations is not available for validation purposes. For this research, the validation technique is limited to a data-split or hold-out sample.

The following examples demonstrate the predictive accuracy and accelerated properties of the proposed test methodology. In each case, a point estimate of the parameter studied was generated. These estimates were then bounded by confidence or prediction intervals. For each case, the estimate and resulting confidence interval *adequately* approximated the previously estimated or demonstrated value.

Adequacy is generally regarded as an estimate within 25%-35% of the true value. This range may appear excessive, however, it does represent the minimum expected accuracy of credible reliability prediction techniques. The Reliability Analysis Center (RAC), a Defense Department support organization, accredits various prediction techniques within this range [25]: MIL-HDBK-217 is generally within 33% of the true MTBF, and the RAC's Reliability Parameter Translation Model II generally estimates within 25% of the fielded performance. Reliability estimates of this nature are referred to as inherent estimates. That is, experimenters estimate the effects of the most detrimental stresses the unit will see in operation, not *all* of the stresses. It has been pointed out that literally hundreds of factors may be working against a fielded military system. The models and prediction techniques account for only a fraction of these. Therefore, the models are limited to predict the inherent performance of the unit as if the true environment were limited to the stresses modeled.

B. Method Verification.

1. Example 1.

The data for this example [26] were collected in 1991. The data were performance parameters collected on hermetically sealed, space qualified resistors. The actual parts tested were RNR55E style, MIL-R-55182 resistors. The objective was to establish estimates of the maximum resistance change over time as a function of stress (including passive conditions). To obtain these values, an estimate of a time constant was established for each stress combination. The time constant, Γ_i, represents the average time to reach the maximum resistance change and remains constant for a fixed stress environment (i). This parameter is analogous to the MTBF estimate for exponentially distributed failures and was, therefore, utilized for this validation.

The tests were conducted under two stresses, temperature and power, at several resistance levels. Both temperature and power were held at three levels. The resistance values were studied at seven levels. Although the resistance is not a *stress*, it is a factor possibly having an effect on Γ_i. For the purposes of validating the proposed methodology, the resistance level will be considered the third stress, and the parameter Γ_i will be referred to as the MTBF.

The levels of temperature in the published report [18] were 85°C, 125°C, and 140°C. Unfortunately, the levels were not equally spaced, a requirement of the proposed method. The data for the 112°C test cells (the mid-point of the temperature range) were obtained directly from the program office.

The resistance levels of the original test ranged from 20Ω to 1000Ω. The levels closest to being equally spaced were 20Ω, 100Ω, and 200Ω. The discrepancy between the two interval widths posed little concern as resistance values generally range ±5% of the rated values, creating an overlapping range of equal intervals. These levels were utilized for the validation. The power levels for the test were 0.0 watts, 0.05 watts, and 0.1 watts. The 0.0 power level represents a passive condition. The data utilized for this validation is presented in Table 5.1. Five units were tested at the 85°C cells, 13 units were tested at each of the other cells.

Table 5.1. Example 1 Data Set.

	Temp								
	85°			112°			140°		
	Power			***Power***			***Power***		
Res.	0.0	0.05	0.1	0.0	0 .05	0.1	0.0	0.05	0.1
20	3703	1562	1052		801*		291	176	149
100	4166	1923	2222	1526*			305	125	142
200	3571	1210	847			595*	236	153	131

*Obtained directly from program office.

To verify the proposed methodology in its entirety, the data from Table 5.1 was used to established the MTBF estimates as though the test were rerun according to the method developed in this research. This was done to demonstrate the appropriateness of the test-cell time calculations as well as the overall modeling and prediction development. For this verification, Block 3 was randomly selected from Appendix A. Nine data points are required for the fractional replicate. The remaining data from the 85°C and 140°C trials of the original effort represents the hold-out data for the verification.

Stress A represents temperature, stress B represents power, and stress C represents the resistance level. Historical data on resistor performance suggest that temperature is the main detriment to performance [21]. Therefore, with respect to the test time equations from Chapter III (3.30, 3.31, 3.32), the three 140°C trials of the fractional block were considered the high stress conditions, while the 112°C and 85°C trials were the mid and low stress conditions, respectively.

To calculate the test time for each trial, an initial estimate of the nominal environment MTBF (δ') had to be established. Referring again to the organization testing the resistors, it was estimated that the MTBF, although very ill-behaved and erratic, was between 2,000 hours and 4,000 hours for the near maximum operational temperature of 85°C. The estimate, δ', was set at 4,000 hours. The minimum number of units per test cell, n, was set at four (failure times are known to be exponentially distributed). The time estimates for the individual test cells were calculated by Equations 3.30, 3.31, and 3.32. The estimates became:

High stress cells:

$$t_h = \ln(0.3)*\delta'/-(n*9)$$

$$t_h = (-1.204)(4000\text{ hr})/-(4\text{ units/cell} * 9\text{ cells})$$

$$t_h = 134\text{ hours/unit}$$

This gives a maximum of 536 hours at each of the 140°C test cells.

Medium stress cells:

$$t_m = \ln(0.3)*\delta'/-(n*6)$$

$$t_m = (-1.204)(4000\text{ hr})/-(4\text{ units/cell} * 6\text{ cells})$$

$$t_m = 201\text{ hours/unit}$$

This gives a maximum of 804 hours at each of the 112°C test cells.

Low stress cells:

$$t_l = \ln(0.3)*\delta'/-(n*3)$$

$$t_l = (-1.204)(4000\text{ hr})/-(4\text{ units/cell} * 3\text{ cells})$$

$$t_l = 401\text{ hours/unit}$$

This gives a maximum of 1604 hours at each of the 85°C test cells.

Table 5.2 illustrates the test plan developed for this verification. The values in the test cells represent the maximum amount of unit-time for that test condition. The MTBF estimates from Table 5.1 (the original data) were compared to these values. If the demonstrated values from Table 5.1 were greater than the total unit-time of the corresponding trial in Table 5.2, the MTBF for that test condition was estimated by the χ^2 distribution as discussed in Chapters III and IV. The test conditions calculated in this fashion were cells 002, 011, and 101.

Table 5.2. Example 1 Test Plan.

	***Stress* A**								
	0			1			2		
	Stress B			***Stress B***			***Stress B***		
Str.C	0	1	2	0	1	2	0	1	2
0			1604		804		536		
1		1604		804					536
2	1604					804		536	

Table 5.3 shows the MTBF values that were demonstrated during the original testing that correspond with the test plan. By comparing the earlier data from Table 5.3 to the test plan presented in Table 5.2, it can easily be determined which test cells *experienced* failures. Table 5.4 presents the lesser of the cell values from Tables 5.2 and 5.3. Values extracted from Table 5.2, cells 002, 011, and 101, represent those cells which did not *experience* failures.

Table 5.3. Example 1 Demonstrated MTBF Values.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0			1052		801		291		
1		1923		1526					142
2	3571					595		153	

Table 5.4. Example 1 MTBF Values.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0			1052		801		291		
1		1604*		804*					142
2	1604*					595		153	

*Did not experience failures

Table 5.5 contains the values used for the model development. Table 5.6 is the natural logarithm of this data.

Table 5.5. Example 1 Model Development Data.

	***Stress* A**								
	0			1			2		
	Stress B			***Stress B***			***Stress B***		
Str.C	0	1	2	0	1	2	0	1	2
0			1052		801		291		
1		2308*		1157*					142
2	2308*					595		153	

*χ^2 estimates for non-failure cells.

Table 5.6. Natural Logarithm of Example 1 Development Data.

	***Stress* A**								
	0			1			2		
	Stress B			***Stress B***			***Stress B***		
Str.C	0	1	2	0	1	2	0	1	2
0			6.958		6.686		5.673		
1		7.744		7.054					4.956
2	7.744					6.389		5.030	

The data from Table 5.6 was subject to a standard ANOVA and model development as described in Chapters III and IV. The Example 1 Model Development Data Contrasts are shown in Table 5.7a, while the resulting ANOVA is shown in Table 5.7b.

Table 5.7a. Example 1 Model Development Data Contrasts.

Effect	Σ(low)	Σ(med)	Σ(high)					
Stress A	22.446	20.129	15.659					
Stress B	20.471	19.460	18.303				Contrast	
Stress C	19.317	19.754	19.163	$\Sigma(\xi')^2$	λ	A	B	C
linear	-1	0	1	2	1	-6.787	-2.168	-0.154
Quad.	1	-2	1	6	3	-2.153	-0.148	-1.028

Table 5.7b. Example 1 ANOVA F-Tests.

Effect	df	SS	MS	F
A_L	1	7.6772	7.6772	157.32*
A_Q	1	0.2575	0.2575	5.28*
B_L	1	0.7834	0.7834	16.05*
B_Q	1	0.0012	0.0012	0.02
C_L	1	0.0040	0.0040	0.08
C_Q	1	0.0587	0.0587	1.20
Error	2	0.0976	0.0488	
Total	8	8.8796		

*Significant effects at $\alpha = 0.15$.

From the ANOVA it is seen that stresses A and B had significant effects on the performance of the unit, while stress C demonstrated no effect. The model, therefore, included four terms: the grand mean, the linear and quadratic terms for stress A, and the linear term for stress B. The model became:

$$\ln(\hat{Y}) = \beta_0 + A_L\xi'_1 + A_Q\xi'_2 + B_L\xi'_1$$

where,

β_0 = grand mean

= 6.470

U_j = (contrast)/n* $\sum (\xi'_j)^2$

$\hat{Y}$ = an estimate of MTBF

The model terms were calculated as follows.

Stress A:

A_L = (-6.787)/3*2

= -1.131

A_Q = (-2.153)/3*6

= -0.120

Stress B:

B_L = (-2.168)/3*2

= -0.361

The linear term for the ξ' is u since λ_1 is 1.

$$\xi'_1 = \lambda_1 u$$

The ξ'_j quadratic term for stress A was calculated as

$$\xi'_2 = \lambda_2 * [u^2 - (k^2-1)/12]$$

where k is the number of test levels, in this case three. Therefore, ξ'_2 becomes

$$\xi'_2 = 3*(u^2 - 2/3)$$
$$= (3u^2 - 2)$$

Combining these terms into a cumulative stress acceleration model gives:

$$\ln(MTBF) = 6.470 - 1.131u_A - 0.120(3u_A{}^2-2) - 0.361u_B$$

Utilizing this model to estimate the MTBF for the lowest stress combination of the test matrix (presumably the maximum operational stress environment) gives:

$$\ln(MTBF) = 6.470 - 1.131(-1) - 0.120(3(-1)^2-2) - 0.361(-1)$$
$$\ln(MTBF) = 7.842$$
$$MTBF = 2{,}545 \text{ hours}$$

Because the data is known to be exponentially distributed, the expected value of the MTBF for various values of the Weibull β parameter, as discussed in Chapter III, was not considered.

Before placing a standard confidence interval on this estimate, the Mean-Square-of-Predicted-Error (MSPE), given by Neter *et al* [24], was considered. According to Neter, if the MSPE is *fairly* close to the MSE based on the model development data, then the MSE of the fitted regression is not seriously biased and gives a good indication of the model's predictive ability. Neter does warn, however, that if the MSPE is much larger than the MSE, then the MSPE should be utilized as an indicator of how well the model will predict in the future. The MSPE is calculated as

$$MSPE = \left(\Sigma\left(Y_{ijk} - \hat{Y}_{ijk}\right)^2\right) / n^* \tag{5.1}$$

where:

Y_{ijk} is the *ijk*th value of the observed response in the hold-out set of data,

$\hat{Y}_{ijk}$ is the corresponding predicted value based on the model-development data,

n^* is the number of cases in the hold-out, or validation data set.

For this verification, only nine data points from Table 5.1 were used for the analysis and model development. The remaining values from Table 5.1 were used to test the predictive ability of the model. That is, these values became the "hold-out" data as described by Neter. The hold-out portion of data included the six data points from both the 85°C and 140°C test condition that were not utilized in the model development. The data from the 112°C cells were not included. Table 5.8 illustrates the hold-out set of data for this verification.

Table 5.8. Example 1 Hold-Out Data Set.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0	3703	1562						176	149
1	4166		2222				305	125	
2		1210	847				236		131

Summing the squared difference between the natural logarithm of the data in Table 5.8 and the corresponding predicted values from the previously developed model gives:

$$\text{MSPE} = (1.243)/12$$
$$= 0.104$$

Comparing this value to the MSE from the ANOVA, 0.0488, it is seen the MSPE is roughly twice the MSE. For this reason [14] the MSPE should be used as the error estimate for *future* observations.

The above point estimate was then bounded by a *prediction* interval. Having 5 degrees of freedom (9 data points - 4 estimates), an 85% interval became:

$$\ln(\hat{Y}) \pm t_{.075,5} * (MSE(1 + \mathbf{x'_0}(\mathbf{X'X})^{-1}\mathbf{x_0}))^{.5}$$

$$7.842 \pm 1.699 * (0.104(1 + 0.444))^{.5}$$

$$7.842 \pm 1.699 * (0.388)$$

$$7.184 \leq \ln(\hat{Y}) \leq 8.500$$

$$1{,}318 \text{ hours} \leq \hat{Y} \leq 4{,}917 \text{ hours}$$

For this validation, the predicted MTBF of the resistors was assumed to be between 2,000 hours and 4,000 hours. Based on an initial assumption that the operational MTBF was 4,000 hours, the resulting point estimate became 2,545 hours. This value was then bounded by an 85% confidence interval that clearly supported the original performance predictions.

2. Example 2.

The data for this verification was extracted from a Rome Laboratory test program. The system is a high performance, high reliability unit. The data were collected under a combined test of three stresses, each at three equally spaced levels. Table 5.9 illustrates the actual test run for the study.

Table 5.9. Example 2 Data Collected.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0	219	219	219	219	222	225	219	219	219
1	372	366	354	369	366	354	375	378	381
2	282	285	288	279	282	288	273	270	267

The test was established to study a physical performance characteristic of the unit, and therefore was not specifically designed to accelerate the measured parameter. The levels of the test represent a mid-point and the upper and lower extremes of the stress range of interest. The *acceleration* will be demonstrated as a result of applying DOE as discussed in Chapter III.

For stresses A and B, the engineers did not know how, or even if, the stresses would effect the unit. For stress C, there was considerable evidence suggesting a possible relationship.

Concerned about possible time dependencies, the engineers collected the data in a random fashion, unbeknownst to them, as a 3^3 factorial. A single replicate was run, recording the average performance values for each cell. Each test cell was run only long enough to make the necessary measurements. An undisclosed number of units were tested.

Due to the nature of the data it was very easy to partition the entire block into a model development set and a hold-out or verification set. From Appendix A, Block 8 was randomly selected as the fractional replicate for the verification (Table 5.10). Therefore, one third of the data was used for the ANOVA and model development while the balance served as the verification hold-out.

Table 5.10. Example 2 Model Development Data.

	***Stress* A**								
	0			1			2		
	Stress B			***Stress B***			***Stress B***		
Str.C	0	1	2	0	1	2	0	1	2
0		219		219					219
1			354		366		375		
2	282					288		270	

The natural logarithm of the Example 2 Model Development Data, Table 5.11, was used for the analysis and model development as described in Chapters III and IV. The Example 2 Model Development Data Contrasts are shown in Table 5.12a, while the resulting ANOVA is shown in Table 5.12b.

Table 5.11. Natural Logarithm of Example 2 Development Data.

	***Stress* A**								
	0			1			2		
	Stress B			***Stress B***			***Stress B***		
Str.C	0	1	2	0	1	2	0	1	2
0		5.389		5.389					5.389
1			5.869		5.903		5.927		
2	5.642					5.663		5.598	

Table 5.12a. Example 2 Model Development Data Contrasts.

Effect	Σ(low)	Σ(med)	Σ(high)	$\Sigma(\xi')^2$	λ	Contrast A	Contrast B	Contrast C
Stress A	16.900	16.955	16.914					
Stress B	16.958	16.890	16.921					
Stress C	16.167	17.699	16.903					
linear	-1	0	1	2	1	0.014	-0.037	0.736
Quad.	1	-2	1	6	3	-0.096	0.099	-2.328

Table 5.12b. Example 2 ANOVA F-Tests.

Effect	df	SS	MS	F
A_L	1	0.00003	0.00003	0.02
A_Q	1	0.00051	0.00051	0.40
B_L	1	0.00023	0.00023	0.18
B_Q	1	0.00054	0.00054	0.42
C_L	1	0.09028	0.09028	70.00*
C_Q	1	0.30109	0.30109	233.40*
Error	2	0.00259	0.00129	
Total	8	0.39527		

*Significant effects $\alpha=0.15$.

It was clearly evident from the ANOVA that stresses A and B had little or no effect on the performance of the unit, while stress C demonstrated very significant linear and quadratic effects. The model, therefore, included only three terms: the grand mean, and the linear and quadratic terms for stress C.

$$\ln(\hat{Y}) = \beta_0 + C_L\xi'_1 + C_Q\xi'_2$$

where,

β_0 = grand mean

= 5.641

U_j = (contrast)/n* $\sum(\xi'_j)^2$

$\hat{Y}$ = an estimate of MTBF

The model terms for stress C became:

C_L = (0.736)/3*2

= 0.123

C_Q = (-2.328)/3*6

= -0.129

The linear term for the ξ' value is u since λ_1 is 1.

$$\xi'_1 = \lambda_1 u$$

The quadratic term for stress C was calculated as

$$\xi'_2 = \lambda_2 * [u^2 - (k^2-1)/12]$$

where k equals the number of test levels, in this case three. Therefore, ξ'_2 became

$$\xi'_2 = 3*(u^2 - 2/3)$$
$$= (3u^2 - 2)$$

Combining these terms into a cumulative stress acceleration model gives:

$$\ln(\hat{Y}) = 5.641 + 0.123u - 0.129(3u^2-2)$$

Utilizing this model to estimate the observed values of the original matrix results in Table 5.13.

Table 5.13. Predicted Values Based on Example 2 Development Data.

	Stress A								
	0			1			2		
	Stress B			***Stress B***			***Stress B***		
Str.C	0	1	2	0	1	2	0	1	2
0	5.389	5.389	5.389	5.389	5.389	5.389	5.389	5.389	5.389
1	5.899	5.899	5.899	5.899	5.899	5.899	5.899	5.899	5.899
2	5.635	5.635	5.635	5.635	5.635	5.635	5.635	5.635	5.635

Each row of the predicted values in Table 5.13 is the same, as factors A and B had no effect on the response.

As discussed in Example 1, one method of measuring the predictive accuracy of a model is to consider the MSPE [24]. By utilizing the hold-out set of data, the MSPE was calculated, then compared to the MSE from the ANOVA.

Taking the natural logarithm of the values in Table 5.9, then removing the development data, results in the hold-out data set (Table 5.14).

Table 5.14. Natural Logarithm of Example 2 Hold-Out Data.

	***Stress* A**								
	0			1			2		
	Stress B			***Stress B***			***Stress B***		
Str.C	0	1	2	0	1	2	0	1	2
0	5.389		5.389		5.403	5.416	5.389	5.389	
1	5.919	5.903		5.911		5.869		5.935	5.943
2		5.652	5.663	5.631	5.642		5.609		5.587

The MSPE, Equation 5.1, was calculated from the data in Tables 5.13 and 5.14.

$$\text{MSPE} = 0.0005$$

This value compared favorably to the MSE from the ANOVA, which was 0.00129. Further, the MSPE also compared favorably the total MSE of all non-significant effects. This value is found by adding the Sum-of-Squares of the non-significant effects to the SSE, then dividing by the associated degrees of freedom. The total MSE was calculated to be 0.00065. Therefore the resulting model was appropriate.

The model results were presented to the engineers and technicians involved with the original data collection. These individuals were asked, quite frankly, "How close do the predictions need to be to waive additional testing?" The group concluded that a maximum deviation of 10% from the demonstrated performance would be acceptable. Table 5.15 contains the actual percent of deviation from the original uncoded data, where percent deviation = 100(1-(observed/predicted)).

Table 5.15. Percent Deviation from Original, Uncoded Data.

	***Stress* A**								
	0			1			2		
	Stress B			***Stress B***			***Stress B***		
Str.C	0	1	2	0	1	2	0	1	2
0	-0.007	-0.007	-0.007	-0.007	-1.377	-2.747	-0.007	-0.007	-0.007
1	2.009	-0.364	2.927	-1.186	-0.364	2.927	-2.832	-3.655	-4.477
2	-0.693	-1.764	-2.836	0.378	-0.693	-2.836	2.521	3.592	4.663

Each of the three performance estimates in Table 5.13 could be bounded by the desired confidence interval by the method developed in Chapter III. For the cells in the model development data set, the interval becomes:

$$\ln(\hat{Y}) \pm t_{\alpha/2,\nu}*(\mathrm{MSE}(\mathbf{x'_0}(\mathbf{X'X})^{-1}\mathbf{x_0}))^{.5}$$

For the cells in the hold-out data set, the estimates are treated as predictions for *future* observations, and are bounded as follows:

$$\ln(\hat{Y}) \pm t_{\alpha/2,\nu}*(\mathrm{MSE}(1+ \mathbf{x'_0}(\mathbf{X'X})^{-1}\mathbf{x_0}))^{.5}$$

The point of most interest in this case was the mid-level stress combination; the confidence interval became:

$$5.899 \pm t_{\alpha/2,\nu}*(0.00129(0.111))^{.5}$$

Because there were three parameters estimated, there were six degrees of freedom for the interval (9 data points - 3 estimates). An 85% confidence interval about the mid-level stress is:

$$5.899 \pm (1.650)(0.012)$$

$$5.879 \leq \ln(\hat{Y}) \leq 5.919$$

which became

$$358 \leq \hat{Y} \leq 372$$

Based on the *accuracy* of this verification, the test engineers involved with the data collection concluded that the original effort could have been completed in one-ninth to one-third of the time if the proposed methodology were employed [27]. This *acceleration* translates to a system savings of approximately $370,000 to $1,000,000 .

3. Example 3.

The data for Example 3 involved capacitor reliability. Specifically, the expected shelf life of Aluminum Electrolytic capacitors [28]. This data, unfortunately, had also been previously collected and, in this case, already reduced to MTBF estimates; the original raw data could not be obtained. For this reason, the verification was established as though a second test were run (similar to Example 1). The maximum "demonstrated" MTBF for each test cell of the verification was limited to the estimated MTBF from the original experiments.

The goal of the original effort was to quantify the expected shelf life of the capacitors as a function of storage temperature and the length and diameter of the case. The only *stress* involved with the original data collection was temperature, which was studied at 65°C, 75°C, and 85°C, not an excessively stressful temperature range. Both case dimensions were studied at three equally spaced values. The original (coded) data is shown in Table 5.16. Although it is not markedly apparent from this data, the cell test times were truncated at 350 hours. For this example, stress A represents temperature, stress B represents the case diameter, and stress C represents the case length.

Table 5.16. Predicted MTBF Estimates from Original Study.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0	350	350	270	250	200	150	150	120	91
1	350	275	250	200	150	140	120	93	83
2	250	225	105	140	125	60	85	76	36

The cell times were calculated by Equations 3.30, 3.31, and 3.32. To obtain these values, the "nominal environment" MTBF was estimated from the original data. Because the cell test times were truncated at 350 hours, the MTBF value was established by calculating a 50% χ^2 estimate from the truncated test value associated with the "lowest stress" cell. This becomes:

MTBF demonstrated in cell (000) = 350 hours

50% χ^2 estimate for time truncated data = 2(test time)/1.39

Estimated "nominal" environment MTBF = 504 hours.

Utilizing this value as δ' in the cell test time equations with a minimum number on test, n, equal to four gives:

High stress cells:

time = (504)ln(0.3)/-(4*9)

= 17 hours

This results in a maximum of 68 unit-hours for the three high stress cells.

Medium stress cells:

time = (504)ln(0.3)/-(4*6)

= 25 hours

This results in a maximum of 100 unit-hours for the three medium stress cells.

Low stress cells:

time = (504)ln(0.3)/-(4*3)

= 51 hours

This results in a maximum of 204 unit-hours for the three low stress cells.

Due to the nature of the test, the occurrence of failures was not anticipated to be notably accelerated. Therefore, the cell times were multiplied by 1.55, as discussed in Chapter III. This resulted in the following test times in unit-hours:

High stress cells:

time = 68 hours * 1.55

= 105 unit-hours

Medium stress cells:

time = 100 hours * 1.55

= 155 unit-hours

Low stress cells:

time = 204 hours * 1.55

= 316 unit-hours

Block 11 was randomly selected for the verification. Because temperature was the only stress, the 85°C test cells represented the high stress cells, while the 75°C and 65°C cells represented the medium stress and low stress cells, respectively. Table 5.17 illustrates the test designed for this study. The values in the table are the maximum unit-hours for the cells.

For the verification, the model development data were the lesser of the cell values from Tables 5.16 and 5.17. This limited the MTBF estimates to the maximum time allowed by the test or the estimated value from the original effort, whichever was less. The data extracted from Table 5.17 were assumed to represent non-failure test cells as the maximum test time was less than the demonstrated MTBF from the original effort. To estimate the demonstrated MTBF for the model development data set, these data, cells 100, 201, and 210, were subject to a χ^2 transformation (Table 5.18).

Table 5.17. Test Design for Example 3.

	Stress A								
	0			1			2		
	Stress B			Stress B			Stress B		
Str.C	0	1	2	0	1	2	0	1	2
0			316	155				105	
1		316				155	105		
2	316				155				105

Table 5.18. Example 3 Model Development Data.

	Stress A								
	0			1			2		
	Stress B			Stress B			Stress B		
Str.C	0	1	2	0	1	2	0	1	2
0			270	223*				151*	
1		275				140	151*		
2	250				125				36

*χ^2 estimate for non-failure cells.

Table 5.19. Natural Logarithm of Example 3 Development Data.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0			5.598	5.407				5.017	
1		5.617				4.942	5.017		
2	5.521				4.828				3.584

The data from Table 5.19, the natural logarithm of the development data, was used for the analysis and model development. The Example 3 Model Development Data Contrasts are shown in Table 5.20a, while the resulting ANOVA is shown in Table 5.20b.

Table 5.20a. Example 3 Model Development Data Contrasts.

Effect	Σ(low)	Σ(med)	Σ(high)					
Stress A	16.736	16.177	13.618					
Stress B	15.945	15.462	14.124				Contrast	
Stress C	16.022	15.576	13.933	$\Sigma(\xi')^2$	λ	A	B	C
linear	-1	0	1	2	1	-3.118	-1.821	-2.089
Quad.	1	-2	1	6	3	0.000	-0.855	-1.197

Table 5.20b. Example 3 ANOVA F-Tests.

Effect	df	SS	MS	F
A_L	1	1.6203	1.6203	20.00*
A_Q	1	0.0000	0.0000	0.00
B_L	1	0.5527	0.5527	6.82*
B_Q	1	0.0406	0.0406	0.50
C_L	1	0.7273	0.7273	8.98*
C_Q	1	0.0796	0.0796	0.98
Error	2	0.1621	0.0810	
Total	8	3.1826		

*Significant effect for $\alpha=0.15$

Evident from the ANOVA, stresses A, B, and C had significant linear effects. The model, therefore, included four terms: the grand mean and the linear components for each stress. The model becomes:

$$\ln(\hat{Y}) = \beta_0 + A_L\xi'_1 + B_L\xi'_1 + C_L\xi'_1$$

where,

β_0 = grand mean

= 5.059

U_j = (contrast)/n* $\sum(\xi'_j)^2$

$\hat{Y}$ = an estimate of MTBF

The linear terms were calculated as follows:

Stress A:

$$A_L = (-3.118)/3*2$$

$$= -0.520$$

Stress B:

$$B_L = (-1.815)/3*2$$

$$= -0.303$$

Stress C:

$$C_L = (-2.089)/3*2$$

$$= -0.348$$

The linear term for the ξ' value is u since λ_1 is 1.

$$\xi'_1 = \lambda_1 u$$

Combining these terms into a cumulative stress acceleration model gives:

$$\ln(\hat{Y}) = 5.059 - 0.520u_A - 0.303u_B - 0.348u_C$$

This model was then used to estimate the shelf life of the capacitors at the "nominal" (lowest *stress* cell) environment. Setting the u values to -1 in the above model gives

$$\ln(\hat{Y}) = 5.059 + 0.520 + 0.303 + 0.348$$

$$\ln(\hat{Y}) = 6.23$$

$$\text{MTBF} = 508 \text{ hours}$$

The MSPE was investigated prior to establishing the confidence interval. This value was calculated to be 0.0298. Compared to the MSE from the ANOVA, 0.081, it was determined that the model is appropriate and without bias. Therefore, the prediction interval about the estimate utilized the MSE from the ANOVA. The 85% interval became:

$$\ln(\hat{Y}) \pm t_{\alpha/2,\upsilon} * (MSE(1 + \mathbf{x}'_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0))^{.5}$$

For cell (000), this was

$$6.230 \pm t_{\alpha/2,\upsilon} * (0.081(1.611))^{.5}$$

Because there were four parameters estimated, the degrees of freedom for the interval was five (9 data points - 4 estimates), giving

$$6.230 \pm (1.699)(0.361)$$

$$5.616 \leq \ln(\hat{Y}) \leq 6.843$$

which became

$$275 \text{ hours} \leq \hat{Y} \leq 937 \text{ hours}$$

Based on the relative accuracy of the point estimate and prediction interval, the methodology is considered a valid and worthy contribution.

As an additional note regarding Example 3, it was not known if the MTBF estimates of the original study took into account the possibility of a time-dependent rate of failure. If the values were simply calculated as the quotient of total operating time over failures (assuming β is 1), the proposed methodology could have influenced the MTBF estimates to represent a more likely time dependent estimate.

Considering the relative simplicity of capacitors, it could be assumed that their rate of failure in storage is time-dependent. Therefore, for analysis purposes, more likely values of the Weibull β parameter could have been used to reflect this possibility. From the expected value discussion of Chapter III, the MTBF estimate of the verification could have been calculated for more reasonable values of β (Table 5.21).

Table 5.21. MTBF Estimate for Likely β Values.

BETA	*MTBF Estimate*
$\beta = 1$	508
$\beta = 2$	450
$\beta = 3$	454
$\beta = 4$	460
$\beta = \infty$	508

This simple transformation of the data could have been performed on the original development data prior to modeling. For example, if β was assumed to be four, each of the MTBF estimates in the model development data set would have been multiplied by 0.906 prior to being analyzed and modeled. For the resulting prediction, it makes no difference which MTBF estimates (the final prediction or the model development estimates) are "adjusted." However, in cases such as this, more information may be obtained if the actual development data represented the true estimates.

C. Conclusion

Three diverse data sets, analyzed and modeled by the proposed methodology, were presented as verifications of the research. Each data set was recovered from previous efforts specifically investigating the effects of *stress* on unit performance.

In each case the method proved accurate and efficient. In addition, the method obtained performance estimates with one-third of the original data: a notable acceleration.

The data sets utilized for the verifications studied various factors, not all of which considered true stresses (e.g., capacitor case dimensions). The methodology proved very robust by accurately modeling these effects as well. Table 5.22 summarizes the Methodology Results.

Table 5.22. Summary of Results

	Example 1	Example 2	Example 3
Dem.	2000-4000	366	504
Pred.	2545	365	508
Intrv.	$1318 \leq \hat{Y} \leq 4917$	$358 \leq \hat{Y} \leq 372$	$275 \leq \hat{Y} \leq 937$

Chapter VI

CONCLUSIONS

Available accelerated reliability testing techniques have proven invaluable to the development of highly complex and reliable systems. Techniques began to emerge in the early 1950s and 1960s primarily focused at the *part-level*. The stress/performance relationship models were predicated on demonstrated theory, such as the Arrhenius relationship, and were subsequently used to extrapolate extremely high stress data to operational parameters. However, as designs became more complex, the need for *system-level* accelerated reliability testing arose. System-level testing has advanced to fairly efficient multi-stress environmental tests. Unfortunately, the existing accelerating techniques are inappropriate for the new demands. The part-level theories utilized do not apply to the higher levels of assembly. There is a need for more efficient system-level reliability testing.

The primary deficiencies of current testing techniques at the system-level are:

1. The assumption of a life distribution.
2. The assumption of a stress relationship function.
3. Extrapolating well beyond test data.

Each of these deficiencies was specifically investigated. By developing a test method based on the Design of Experiments, all three of these burdens were removed.

The product of this research is a test technique based on a one-third fractional factorial experimental design, which requires no specific assumptions and no extrapolation. By conforming to the general requirements of sound experimentation, the assumptions of a life distribution and a stress/performance relationship are simply no longer required.

An integral step in the experimental process is the statement of a Null Hypothesis. This typically states that the factors studied have no effect on the output variable. This hypothesis is tested by a conventional Analysis of Variance. A fundamental requirement of ANOVA is that the data collected during an experiment be drawn from a normal population. According to the Central Limit Theorem, a distribution of sample means approaches normality for sufficiently large samples, regardless of the parent distribution. Therefore, because expected values (means) were studied, there was no need to assume the specific shape of the test unit's time-to-failure distribution.

The stress/performance relationship and the extrapolation issues were also removed by the application of a sound experimental design. By testing three stress factors, each at three levels, the resulting geometry of the experimental design provided the necessary *acceleration* for the stress test. By confining each stress range to include a test level within the operational envelope of the system, the resulting one-third fractional factorial design designated test cells which called for the maximum level of one or more stresses while the remaining stresses were at operational levels. This situation provided an "accelerator" stress while allowing data to be collected at operational values. Therefore, because the overall test matrix overlapped the operational stress range of the unit, there was no need to extrapolate the resulting stress/performance model beyond the test data.

The method developed in this research was verified with three separate data sets. Verification set 1 included reliability data on space qualified resistors subjected to temperature, power, and resistance level constraints. The actual MTBF of the resistors was unknown; however, preliminary estimates ranged from 2,000 hours to 4,000 hours for the maximum stress environment. The methodology of this research quantified the existing stress relationships and estimated an MTBF of 2,545 hours. This was accomplished with approximately 32% of the original test data.

The second verification data set included a physical performance test on an Air Force system. The data for this effort were collected under three stresses, each at three levels, representing the operational environment of the unit. The results of applying the proposed technique clearly demonstrated the value of the methodology. The predicted performance values of the unit for each combination of the original test were extremely close to the demonstrated values. Based on these results, it was speculated that if the proposed methodology were originally implemented, the test would have been completed in one-ninth to one-third of the original test time at a system savings of $370,000 to over a $1,000,000.

Validation set 3 involved capacitor reliability. This particular data set demonstrated the flexibility of the proposed method. In this case, the dormant reliability of capacitors was investigated for a single stress, temperature, and two case dimensions. The estimated reliability of the resistors under optimum storage conditions was 504 hours. The proposed methodology estimated this value to be 508 hours, while also successfully modeling the demonstrated relationships between dormant reliability and the three variables studied. The application of the methodology arrived at the estimates utilizing only one-third of the collected data.

It has been demonstrated that this methodology has the potential to save time and money, and should be considered the preferred choice for system-level reliability testing .

Future Research

As systems become more complex, and as the economy of test procedures becomes increasingly important, future accelerated testing techniques must contend with additional issues.

1. Additional Cost Demands of Future Testing.

The cost of reliability testing has been shown to be very high. This not only includes the monetary demands but also the time requirements. This research specifically addressed the issues of time. By reducing the testing requirements, in some cases by nearly 70%, there will obviously also exist considerable cost savings. However, as the cost demand continues to play a larger role in testing, specific steps should to be taken to investigate the individual costs of the treatment combinations included in such a test.

2. Qualitative Factors.

The methodology of accelerated reliability testing developed in this research was specifically designed for quantitative factors. A technique allowing for quantitative factors as well as qualitative factors would result in a much broader realm of testing. For example, MIL-HDBK-217F recognizes 11 operational environments, from Ground Benign to Space

Flight, each qualitatively defined. In addition, modeling parameters such as "Quality Class" are simply defined as Commercial, Ruggedized, and Full Military. Factors such as these are obviously very important to reliability engineering; however, in order to effectively test and model factors such as these, a qualitative technique must be developed.

3. The Elimination of System Level Testing.

By applying the test methodology developed in this research, experimenters can better understand their systems and the effects various stresses have on them. However, due to the costs associated with reliability testing, there does exist the goal of eliminating testing (at least at the system-level) in its entirety. A fundamental knowledge of component reliability is an essential step in reaching this goal. The development of a modeling technique capable of accurately incorporating the results of several experiments on different system components may ultimately allow for an accurate system model without system-level testing.

APPENDIX A

One-Third Fractional Factorial Replicates for 3^3 Designs

Defining Contrast: ABC

Block 1.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0	■					■		■	
1			■		■		■		
2		■		■					■

Block 2.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0		■		■					■
1	■					■		■	
2			■		■		■		

Block 3. Appropriate for Reliability Testing.

Stress A

	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0			■		■		■		
1		■		■					■
2	■					■		■	

Defining Contrast: AB^2C.

Block 4.

Stress A

	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0	■				■				■
1		■				■	■		
2			■	■				■	

Block 5. Appropriate for Reliability Testing.

Stress A

	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0			■	■				■	
1	■				■				■
2		■				■	■		

Block 6.

Stress A

	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0		■				■	■		
1			■	■				■	
2	■				■				■

Defining Contrast: ABC^2

Block 7.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0	■					■		■	
1		■		■					■
2			■		■		■		

Block 8.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0		■		■					■
1			■		■		■		
2	■					■		■	

Block 9. Appropriate for Reliability Testing.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0			■		■		■		
1	■					■		■	
2		■		■					■

Defining Contrast: AB^2C^2

Block 10.

	Stress A								
	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0	■				■				■
1			■	■				■	
2		■				■	■		

Block 11. Appropriate for Reliability Testing.

Stress A

	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0			■	■				■	
1		■				■	■		
2	■				■				■

Block 12.

Stress A

	0			1			2		
	Stress B			*Stress B*			*Stress B*		
Str.C	0	1	2	0	1	2	0	1	2
0		■				■	■		
1	■				■				■
2			■	■				■	

APPENDIX B

The Rome Laboratory Reliability Engineer's Toolkit

Topic A11

SOURCE: Reproduced with the permission of Rome Laboratory, United States Air Force Materiel Command, Griffiss Air Force Base, New York. *The Rome Laboratory Reliability Engineer's Toolkit*, pp. 105-107, 1993.

Topic A11: Reliability Adjustment Factors

"What if" questions are often asked regarding reliability figures of merit. For a rapid translation, tables for different quality levels, various environments and temperatures are presented to make estimates of the effects of the various changes. The data base for these tables is a grouping of approximately 18000 parts from a number of equipment reliability predictions performed in-house on military contracts. The ratios were developed using this data base and MIL-HDBK-217F algorithms. The relative percentages of the part data base are shown as follows:

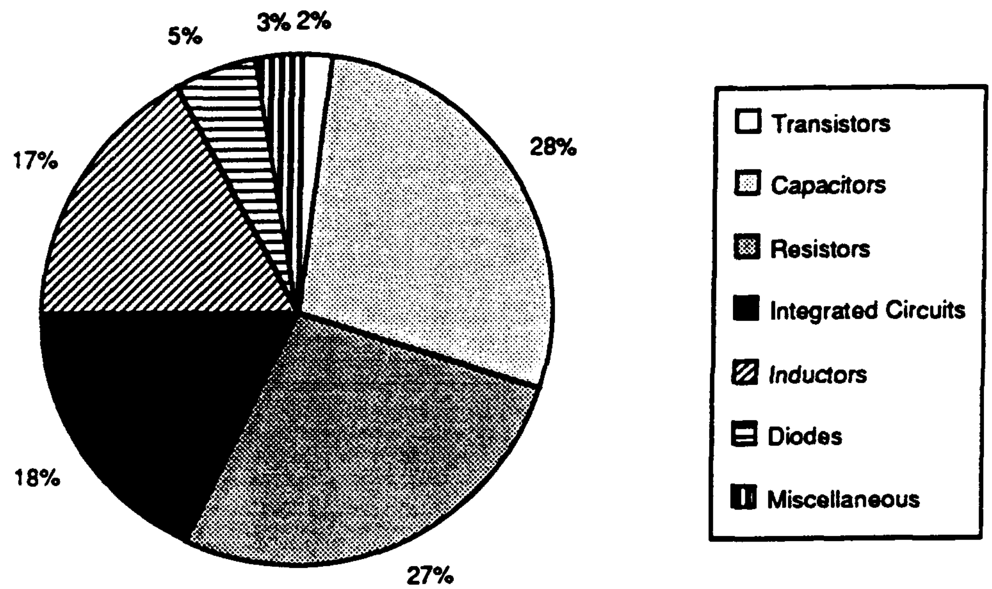


Table A11-1: Part Quality Factors (Multiply MTBF by)

		To Quality Class			
		Space	Full Military	Ruggedized	Commercial
From Quality Class	Space	X	0.8	0.5	0.2
	Full Military	1.3	X	0.6	0.2
	Ruggedized	2.1	1.6	X	0.4
	Commercial	5.3	4.1	2.5	X
	IC	Class S	Class B	Class B-1	Class D
	Semiconductor	JANTXV	JANTX	JAN	NONMIL
	Passive Part	ER(S)	ER(R)	ER(M)	NONMIL

CAUTION: Do not apply to Mean-Time-Between-Critical-Failure (MTBCF).

Table A11-2: Environmental Conversion Factors (Multiply MTBF by)

From Environment \ To Environment	G_B	G_F	G_M	N_S	N_U	A_{IC}	A_{IF}	A_{UC}	A_{UF}	A_{RW}	S_F
G_B	X	0.5	0.2	0.3	0.1	0.3	0.2	0.1	0.1	0.1	1.2
G_F	1.9	X	0.4	0.6	0.3	0.6	0.4	0.2	0.1	0.2	2.2
G_M	4.6	2.5	X	1.4	0.7	1.4	0.9	0.6	0.3	0.5	5.4
N_S	3.3	1.8	0.7	X	0.5	1.0	0.7	0.4	0.2	0.3	3.8
N_U	7.2	3.9	1.6	2.2	X	2.2	1.4	0.9	0.5	0.7	8.3
A_{IC}	3.3	1.8	0.7	1.0	0.5	X	0.7	0.4	0.2	0.3	3.9
A_{IF}	5.0	2.7	1.1	1.5	0.7	1.5	X	0.6	0.4	0.5	5.8
A_{UC}	8.2	4.4	1.8	2.5	1.2	2.5	1.6	X	0.6	0.8	9.5
A_{UF}	14.1	7.6	3.1	4.4	2.0	4.2	2.8	1.7	X	1.4	16.4
A_{RW}	10.2	5.5	2.2	3.2	1.4	3.1	2.1	1.3	0.7	X	11.9
S_F	0.9	0.5	0.2	0.3	0.1	0.3	0.2	0.1	0.1	0.1	X

Environmental Factors as Defined in MIL-HDBK-217

G_B - Ground Benign; G_F - Ground Fixed; G_M - Ground Mobile; N_S - Naval Sheltered; N_U - Naval Unsheltered; A_{IC} - Airborne Inhabited Cargo; A_{IF} - Airborne Inhabited Fighter; A_{UC} - Airborne Uninhabited Cargo; A_{UF} - Airborne Uninhabited Fighter; A_{RW} - Airborne Rotary Winged; S_F - Space Flight

CAUTION: Do not apply to MTBCF.

Table A11-3: Temperature Conversion Factors (Multiply MTBF by)

From Temperature °C \ To Temperature °C	10	20	30	40	50	60	70	80	90	100
10	X	0.9	0.8	0.8	0.7	0.5	0.4	0.3	0.2	0.1
20	1.1	X	0.9	0.9	0.7	0.6	0.5	0.4	0.2	0.2
30	1.2	1.1	X	1.0	0.8	0.7	0.5	0.4	0.3	0.2
40	1.3	1.2	1.0	X	0.8	0.7	0.6	.04	0.3	0.2
50	1.5	1.4	1.2	1.2	X	0.8	0.7	0.5	0.3	0.2
60	1.9	1.7	1.6	1.5	1.2	X	0.8	0.6	0.4	0.3
70	2.4	2.2	1.9	1.8	1.5	1.2	X	0.7	0.5	0.3
80	3.3	3.0	2.7	2.6	2.1	1.7	1.4	X	0.7	0.4
90	4.9	4.5	4.0	3.8	3.2	2.5	2.0	1.5	X	0.6
100	7.7	7.0	6.3	6.0	5.0	4.0	3.2	2.3	1.6	X

CAUTION: Do no apply to MTBCF.

REFERENCES

[1] Bazu, M. and Tazlauanu, M., Reliability Testing of Semiconductor Devices in a Humid Environment, *Proceedings 1991 Annual Reliability and Maintainability Symposium*, pp. 307, 1991.

[2] Hobbs, G. K., Stress Screening: Progress, Setbacks and the Future, *Reliability Technology, Theory & Applications*, Elsevier Science Publishers B.V. (North Holland), 1986.

[3] Hicks, C., *Fundamental Concepts in the Design of Experiments,* 3rd Edition, Holt, Rinehart, and Winston, Inc., New York, 1982.

[4] Fisher, R. A., *The Design of Experiments*, Oliver & Boyd, London, UK, 1935.

[5] Weibull, W., A Statistical Distribution Function of Wide Applicability, *Journal of Applied Mechanics*, pp. 293-297, 1951.

[6] Epstein, B. and Sobel, M., Life Testing, *Journal of the American Statistical Association*, 1951.

[7] Epstein, B., Truncated Life Tests in the Exponential Case, *Annuals of Mathematical Statistics*, 1954.

[8] Reliability Stress Analysis for Electronic Equipment, TR-1100, Radio Corporation of America, Defense Electronic Products, Camden, NJ, 1956.

[9] MIL-HDBK-217F, Reliability Prediction of Electronic Equipment, 1990.

[10] Advisory Group on the Reliability of Electronic Equipment (AGREE), U.S. Department of Defense (R&E), Final Report, June, 1957.

[11] Nelson, W., *Accelerated Test; Statistical Models, Test Plans, and Data Analysis*, John Wiley & Sons, New York, 1990.

[12] Derringer, G., Considerations in Single and Multiple Stress Accelerated Life Testing, *Journal of Quality Technology*, Vol. 14, No. 3, pp. 130, July 1982.

[13] Buoni, F. B., Mazzuchi, T. A., Pollack, L.R., and Welch, A Wide-Parametric, Bayesian Methodology for System-Level, Step-Stress Accelerated Life Testing, Unpublished Draft Proposal.

[14] Searle, S. R., *Linear Models*, Wiley & Sons, New York, 1971.

[15] Davies, O. L., *The Design and Analysis of Industrial Experiments*, Hafner Publishing Company, New York, 1967.

[16] Draper, N. R. and Smith, H., *Applied Regression Analysis*, 2nd Edition, Wiley Series in Probability and Mathematical Statistics, Wiley & Sons, New York, 1981.

[17] Fisher, R. A. and Yates, F., *Statistical Tables for Biological, Agricultural, and Medical Research*, 4th Edition, Oliver & Boyd, London, UK, 1935.

[18] Kempthorne, O., *The Design and Analysis of Experiments*, Wiley & Sons, New York, 1952.

* [19] Haller, K. A., Anderson, K., Zbytniewski, J. D., Bagnall, L., Smart BIT, Grumman Aerospace Corporation, RADC-TR-85-148. USGO agencies and their contractors; critical technology, dtd Aug 1985.

[20] Gayen, A. K., The Distribution of the Variance Ratio in Random Samples of Any Size Drawn from Non-Normal Universes, *Biometrika*, Vol. 37, pp. 236-255, 1950.

[21] Denson, W. K., Reliability Assessment of Critical Electronic Components, IIT Research Institute, RL-TR-92-197, 1992.

[22] *The Rome Laboratory Reliability Engineer's Toolkit*, Systems Reliability Division, Rome Laboratory, Air Force Materiél Command, Griffiss AFB NY, 1993.

[23] Welker, E.L., and Lipow, M., Estimating the Exponential Failure Rate from Data with no Failure Events, *Proceedings 1974 Annual Reliability and Maintainability Symposium*, pp. 420, 1974.

[24] Neter, J., Wasserman, W., and Kuter, M. H., *Applied Linear Statistical Models*, Irwin Inc., Boston, MA, 1990.

* [25] Chit, D., Recchio, A. J., Russel, D., and Wrisley, R., Reliability and Maintainability Operational Parameter Translation II, IIT Research Institute, RADC-TR-89-299, 1989. USGO agencies and their contractors; critical technology, dtd Dec 1989.

[26] Sampson, M. J. and Lee, S. M., Reliability Advantages of Hermetically Sealed Metal Film Resistors, *CARTS*, pp. 208, 1992.

[27] Jones, E., Personal Communication, Rome Laboratory, 1993.

[28] Lauber, J. A., Aluminum Electrolytic Capacitor - Reliability, Expected Life, and Shelf Capability, TP83-9, Sprague Electric Company, 1983.

[29] Angus, J. E., Saari, A. E., and VanDenBerg, S. J., Environmental Stress Screening, Hughes Aircraft Company, RADC-TR-86-149, 1986.

[30] Box, G. E. P., Hunter, W. G., and Hunter, J. S., *Statistics for Experimenters*, Wiley Series in Probability and Mathematical Statistics, Wiley & Sons, New York, 1978.

[31] Box, G. E. P., Non-Normality and Tests on Variances, *Biometrika*, Vol. 40, pp. 318-335, 1953.

[32] Burkhard, A. H., Accelerated Testing - A New Vision, *Proceedings 1992 Institute of Environmental Sciences*, pp. 388, 1992.

[33] Confer, R., Canner, J., Trostle, T., and Kurtz, S., Use of Highly Accelerated Life Test (HALT) to Determine Reliability of Multilayer Ceramic Capacitors, IEEE Transactions on Reliability, 1991.

[34] Coppola, A., Reliability Engineering of Electronic Equipment: A Historical Perspective, *IEEE Transactions on Reliability*, pp. 29, April 1984.

[35] Dhillon, B. S., *Reliability Engineering in Systems Design and Operation*, Van Nostrand Reinhold Company, New York, 1983.

[36] Epstein, B., Tests for the Validity of the Assumption that the Underlying Distribution of Life is Exponential: Part II, *Technometrics*, Vol. 2, No. 2., pp. 167, May, 1960.

[37] McDonnell Douglas Astronautics Co-East, Evaluation of Microcircuit Accelerated Test Techniques, RADC-TR-76-218, 1976.

[38] Fornell, G. E., Lt. Gen., Life Testing Advances Prove High Reliability C^3I Systems, *Signal*, pp. 62, 1991.

[39] Gayen, A. K., Significance of Difference Between the Means of Two Non-Normal Samples, *Biometrika*, Vol. 37, pp. 399-408, 1950.

[40] Hahn, G. J. and Meeker, W. Q., Jr., Pitfalls and Practical Considerations inProduct Life Analysis-Part II: Mixtures of Product Populations and More General Models, *Journal of Quality Technology*, Vol. 14, No. 4, pp. 177, October 1982.

[41] Hahn, G. J. and Meeker, W. Q., Jr., How to Plan an Accelerated Life Test - Some Practical Guidelines, *American Society for Quality Control*, Statistical Division, Milwaukee, WI, 1985.

[42] Hines, W. W. and Montgomery, D. C., *Probability and Statistics in Engineering and Management Science,* 2nd Edition, Wiley & Sons, New York, 1972.

[43] Kuehn, R., E., Four Decades of Reliability Experience, *Proceedings 1991 Annual Reliability and Maintainability Symposium*, pp. 76, 1991.

[44] Shooman, M. L., *Probabilistic Reliability: An Engineering Approach,* 2nd Edition, Krieger Publishing Company, Malabar, Florida, 1990.

[45] MIL-HDBK-781C, Reliability Test Methods, Plans, and Environments for Engineering Development, Qualification, and Production, 1987.

[46] MIL-STD-810E, Environmental Test Methods and Engineering Guidelines, 1990.

MISSION

OF

ROME LABORATORY

Mission. The mission of Rome Laboratory is to advance the science and technologies of command, control, communications and intelligence and to transition them into systems to meet customer needs. To achieve this, Rome Lab:

a. Conducts vigorous research, development and test programs in all applicable technologies;

b. Transitions technology to current and future systems to improve operational capability, readiness, and supportability;

c. Provides a full range of technical support to Air Force Materiel Command product centers and other Air Force organizations;

d. Promotes transfer of technology to the private sector;

e. Maintains leading edge technological expertise in the areas of surveillance, communications, command and control, intelligence, reliability science, electro-magnetic technology, photonics, signal processing, and computational science.

The thrust areas of technical competence include: Surveillance, Communications, Command and Control, Intelligence, Signal Processing, Computer Science and Technology, Electromagnetic Technology, Photonics and Reliability Sciences.