NASA CR-1128

PRACTICAL RELIABILITY

Volume III - Testing

Distribution of this report is provided in the interest of information exchange. Responsibility for the contents resides in the author or organization that prepared it.

Prepared under Contract No. NASw-1448 by
RESEARCH TRIANGLE INSTITUTE
Research Triangle Park, N.C.

for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

PRECEDING PAGE BLANK NOT FILMED.

FOREWORD

The typical few-of-a kind nature of NASA systems has made reliability a premium even on the initial items delivered in a program. Reliability defined and treated on the basis of percentage of items operating successfully has much less meaning than when larger sample sizes are available as in military and commercial products. Reliability thus becomes based more on engineering confidence that the item will work as intended. The key to reliability is thus good engineering--designing reliability into the system and engineering to prevent degradation of the designed-in reliability from fabrication, testing and operation.

The PRACTICAL RELIABILITY series of reports is addressed to the typical engineer to aid his comprehension of practical problems in engineering for reliability. In these reports the intent is to present fundamental concepts on a particular subject in an interesting, mainly narrative form and make the reader aware of practical problems in applying them. There is little emphasis on describing procedures and how to implement them. Thus there is liberal use of references for both background theory and cookbook procedures. The present coverage is limited to five subject areas:

Vol I. - Parameter Variation Analysis describes the techniques for treating the effect of system parameters on performance, reliability, and other figures-of-merit.

Vol. II. - Computation considers the digital computer and where and how it can be used to aid various reliability tasks.

Vol. III. - Testing describes the basic approaches to testing and emphasizes the practical considerations and the applications to reliability.

Vol. IV. - Prediction presents mathematical methods and analysis approaches for reliability prediction and includes some methods not generally covered in tests and handbooks.

Vol. V. - Parts reviews the processes and procedures required to obtain and apply parts which will perform their functions adequately.

These reports were prepared by the Research Triangle Institute, Research Triangle Park, North Carolina 27709 under NASA Contract NASw-1448. The contract was administered under the technical direction of the Office of Reliability and Quality Assurance, NASA Headquarters, Washington, D. C. 20546 with Dr. John E. Condon, Director, as technical contract monitor. The contract effort was performed jointly by personnel from both the Statistics Research and the Engineering and Environmental Sciences Divisions. Dr. R. M. Burger was technical director with W. S. Thompson serving as project leader.

This report is Vol. III - Testing. It is not devoted just to formal reliability testing but is concerned more generally with all testing throughout a hardware program since it all affects reliability to some extent.

One of the obvious difficulties in writing on a subject so broad is that of including both breadth and depth. Considerable breadth is included even to the extent of introducing some fringe topics which aids in illustrating the relationship of testing to other activities. Equal depth could not, of course, be achieved in all topics. Hard core topics are thus treated more thoroughly than others. There was an attempt to provide enough description on each topic so that the newcomer can immediately grasp the major concepts and ideas involved in the subject. Further depth is aimed at practical considerations. To prevent a "rehash" of much material adequately treated in the literature, liberal use is made of references.

The organization of the report departs considerably from most presentations on testing. Attention is called to the major parts designated in the table of contents. The organization of the material resulted after considerable deliberation on what is really involved in testing.

W. S. Thompson is the principal author of this report. A. C. Nelson, Jr. provided much assistance in the development of material involving statistics and authored most of the Appendix. R. R. Stockard participated in the initial drafting of material. Dr. R. A. Evans assisted in outlining the text and the section on accelerated testing draws heavily from some text material on the subject previously prepared by him.

ABSTRACT

Testing is discussed from an engineering viewpoint. The subject is structured in terms of basic test types, basic problem types amenable to treatment by the basic test types, and applications of testing in hardware programs. The emphasis is on basic principles and practical problems in implementing them. Generally, the discussion emphasizes testing for reliability rather than reliability testing in the formal sense. Part I is devoted to concepts, definitions and general procedures. In particular, a section on test classifications considers the many ways of viewing tests and some of the prevalent confusion in terminology is resolved. Parts II and III collectively treat the basic test and problem types. The manner of separating these discussions into the two parts serves to highlight an important consideration in testing which is often overlooked, viz., whether or not aging is important. Part IV contains discussions on several subjects including nondestructive testing, environmental testing, and accelerated testing which are common to several or all of the basic test types. Part V surveys the major applications of testing in hardware programs and draws special attention to applications associated with reliability. An Appendix contains discussions of the mathematical topics pertinent to testing.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

Page No.

PART IV: MISCELLANEOUS TOPICS

PART V: HARDWARE PROGRAM APPLICATIONS

TABLE OF CONTENTS (Continued)

1. Introduction

A large portion of system procurement costs is designated explicitly for testing. Coupled with the many other test expenditures disguised under other program costs, testing is indeed expensive. Systems must perform successfully, however, and this requires knowledge of how they operate and verification of their ability to perform. Testing, when properly coordinated with other sound design, production, and operation practices, is a key means for accomplishing this.

Testing is thus expensive, but it can also be expensive not to test. Testing is a necessity, but it is not always easily justified. Resources allocated for testing are typically the first ones pared in budget cuts. This reflects on the difficulty of justifying the expenditure but more than ever serves to emphasize the need for making the best use of available testing resources.

Improved testing procedures such as screening, automation, and integrated testing approaches are evolving, but changes are slow and result primarily when states of urgency arise. Such improvements will continue to evolve, but so many of the real problems still occur at the implementation level. Typically, accelerated aging procedures are improperly applied, inefficient test designs are used, and damage is induced when not required. It is toward these latter problems that most of this report is aimed.

This report is devoted to promoting a better understanding of the basic principles of testing and the practical problems in implementing them. The principles are few and really relatively easy to grasp. The practical problems are many and not often obvious.

The orientation is toward the interests of the engineer who has a problem that requires testing for its resolution. It is not concerned so much with who in the organization has what responsibility nor with detailed procedures of how to carry out a test. Rather, the emphasis is on those concepts with which the engineer should be familiar, what testing approaches are available and what they can do for him, and practical considerations which he should bear in mind in planning a test.

The organization of the report recognizes a structuring of the subject summarized by the following brief comments. First, there are basic test types distinguished largely by the severity of the conditions for the test and the effects produced. For example, in one test the conditions may be irrelevant, in another they may be intended to cause instantaneous failure, and in still another they may cause aging. Inherent in these distinctions also are the nature of the observed responses and the types of measurements permissable. Second, there are basic problem types to be treated by the basic test types. For example, the investigation of failure mechanisms and the measurement of life may both require the same basic test types in some cases but different ones in other cases. Third, there are different applications of testing in hardware programs requiring certain testing approaches; the testing approaches generally are characterized by the basic problem type treated and the test type employed. For example, qualification testing employs testing approaches which generally give a better assessment of capability than does say acceptance testing. Reliability testing, a more discipline-oriented application, traditionally treats many problem areas and thus involves many testing approaches.

This structuring of testing is notably different than that associated with most presentations on the subject. The motivation for it is the desire to create a common link to all interests in testing. One has only to sample the literature to appreciate the lack of effective communication on the subject. For example, a particular designation of a test or a concept by one person may mean something quite different to another. Section 3 on test classifications treats this problem in more detail and provides further clarification of the structuring introduced above.

As indicated in the Table of Contents, the main body of the report is partitioned into five major parts. Part I is devoted mainly to concepts, definitions, and general procedures as groundwork for later discussions. Parts II and III collectively treat the basic test and problem types recognized in the structuring scheme. The manner of separating these discussions into the two parts serves to focus attention on another important distinction in testing which is often overlooked, viz., whether or not aging is important. Part IV contains discussions on several subjects common to several or all test types. Part V surveys the various applications of testing in hardware programs and draws special attention to the applications associated with reliability. The Appendix is devoted to a presentation of various mathematical topics pertinent to testing.

2. Elements of Testing

A test is a procedure used to characterize an object or substance. In a test certain conditions either exist or are purposely applied, the item responds by exhibiting some type of behavior, and there is some observation or measurement associated with the behavior. The basic ingredients of a test are thus

(1) the test conditions,

(2) the response during the test,

(3) measurements.

We can and do talk about tests representing degenerate forms of this definition; for example, when the specific test conditions are not really of interest, when the response merely represents a static property of an item, and when the measurement is simply a qualitative observation.

Note that in the above definition we talk about a test in terms of a single item. The term testing, will be used to refer to either a single test or a group of tests generally used to resolve a particular problem. Testing may thus involve a single test, a number of different tests with the same item, identical tests of different items, or different tests with different items. Reliability demonstration, for example, generally requires testing instead of just a test.

Experimentation or an experiment is used synonymously with testing to generally denote something broader than a test.

2.1 The Response During the Test

The response of an item during a test refers specifically to the cause and effect relationship between conditions and behavior. The relationship is also dependent upon properties of the tested item. The response of an item thus provides us with a physical model showing behavior as a function of certain parameters or variables.

To elaborate let y be a physically measurable parameter (or variable)* chosen to describe an item's behavior in a test. y is called the response variable and may represent a characteristic such as resistance, thrust, color, electrical continuity, explosive force, detection probability, time that a signal is above a given level, or structural deflection. Also, let the set of parameters needed to describe the test conditions and item properties be $\underline{x} = x_1, x_2, \ldots, x_n$. Those x's associated with the test conditions may represent temperature, rate of change of temperature, rms vibration level, radiation intensity, or salt concentration for example,

---

* The terms, variable and parameter, are used interchangeably through this report.

while those associated with the item properties may represent such things as impurity concentration, specific heat, stray capacitance, volume, rigidity, and absorptivity. The test response provides y, the dependent variable, as a function of $\underline{x}$, the independent variables. Usually a test involves behavior over time, thus some or all of the variables may be functions of time. Further complexity may result from concern for more than one response variable and possible correlation among these for simultaneous responses.

A familiar analogy to this physical model is the mathematical model of the response or the equation linking y to $\underline{x}$ which can be written in general form as $y = g(\underline{x})$. Often there is some correspondence between a mathematical model and a physical model; for example, structural equations derived on an analytical basis may closely conform to certain static test results. Certain tests may be conducted for the express purpose of verifying a mathematical model and also some may be devoted to estimating certain parameters of a mathematical model.

Responses can generally result in any type of behavior from no change to total destruction and can be discrete, continuous, or intermittant functions of time, the test condition parameters, or the characteristics of the tested item. Other fundamental concepts pertinent to discussion of test methods are presented in the remainder of this section.

2.1.1 Reversible Responses vs. Nonreversible Responses

A reversible response is one in which there is no permanent change in the item's properties of interest as a result of the conditions causing the response; a nonreversible response is one in which there is such a change. The distinction between reversible and nonreversible responses is not always clear-cut. For practical purposes we consider a response to be reversible if an item could be treated as the same item in a repeated application of the same conditions. Damage and aging are thus characteristic of nonreversible responses but not reversible responses. Some permanent change is, of course, always present even if at the atomic level. However, for practical purposes, many items display reversible responses.

Certain responses not normally considered reversible such as those involving hysteresis and yield of materials may, for practical purposes, be considered reversible if the item can be conveniently forced to return to its original state. Generally we say that if an item's functional characteristics are not altered in the response then the response is reversible. Damage, normally a nonreversible phenomena, can often be repaired; however, the assumption of reversibility for repairs depends upon how much repair is required. Some repair results in essentially a new item. Even certain time dependent processes such as annealing, normally associated

The major purpose of this discussion is introduction of the basic concepts applicable to all types of conditions. Thus we are not emphasizing at this point the physical environmental factors such as temperature, vibration and radiation. These are treated with greater emphasis in Sec. 10, Environmental Testing.

2.2.1 Basic Characterization

The characterization of test conditions is fundamental to their description, measurement, and control. Both qualitative and quantitative descriptions are used. Both are frequently required to completely specify conditions for a test. Although crude descriptions such as "salty atmosphere" or "low voltage" are sometimes adequate, quantitative descriptions are needed for most tests.

The basic ingredients of quantitative descriptions are (1) the characterizing parameters, (2) their values, and (3) the functional form of the model which relates them. As a simple example, a vibration condition specified for a test may read "...sinusoidal vibration, 5 mils, at 3 Hz for 25 seconds". The underlying model is $N(t) = A \sin (2\pi ft + \phi)$ where A, the amplitude, is 5 mils; f, the frequency, is 3 Hz; t, time, runs from zero to 25 sec.; and ϕ, the phase, is arbitrary. Values of parameters can also be specified by statistical characteristics such as range of parameter or values of distribution parameters such as mean and variance.

Relation to Real-World Conditions

Descriptions for test conditions have no value unless the descriptions relate to quantities that can be measured and controlled. The above description of vibration, for example, is quite typical of what a vibration table can deliver. In this case, the characteristics of the test condition can be directly related to the characteristics of what the facility can provide. That is, the vibration amplitude, frequency, duration, and sinusoidal motion are easily achieved and measured.

In some cases the relation between test conditions and facility capability is known only indirectly. For example, the characteristic of real interest in a temperature test of a circuit often is the junction temperature of each of the solid state devices in the circuit. The available parameters to be measured and controlled are ambient temperatures or, at best, device case temperatures. Some extrapolation is thus required to obtain the temperatures of interest.

Effect on Test Results

The above relation between the test description and what is practically available raises questions on validity and accuracy of test results. The test description is a conceptual model; what is achieved for the test and what exists during normal item use are each physical models. The differences among these models are related to test validity and accuracy.

The measurement of vibration table motion, for example, may reveal harmonics which for complete description of induced vibration would require a more complex model of table motion than a simple sinusoid. Whether use of the more complex model is necessary depends upon the details of the problem at hand. Typically, sinusoidal vibration is a gross approximation to the very complex vibration of a vibration table. However, a significant effort to identify and describe the shake table harmonics is not justified unless there is reason to expect that the item(s) to be tested could be more sensitive to harmonics than to the fundamental frequency. Usually one would not want, say, 25% of the total power to be contained in the harmonics because then a particular failure mode observed might be in response to a harmonic instead of to the fundamental. For some vibration tests, however, the purpose of testing an item is simply to vibrate the item to see whether something comes loose, with the fundamental frequency chosen only as a convenient number. In such cases, the harmonic content of the table motion clearly is not of interest.

Descriptions of test conditions are not always as simple as the example discussed above. Often operational profiles have to be simulated, aging must be accelerated, and many test conditions must be treated. A discussion of test conditions and how they are specified as functions of time is given below.

2.2.2 Stress

The term, stress, has its origin in the field of mechanics where it has specific meaning as force per unit area for various types of stresses such as principal, tensile, and shear. Through frequent usage in testing and reliability literature, it has become generalized (and perhaps too readily so) to mean simply any physical entity such as temperature, voltage, pressure, vibration, electrical load, and radiation which potentially causes failure whether through a nonaging or an aging response.

Many misconceptions and difficulties have resulted from this generalization, not the least of which is an understanding of what really causes failure. A good attempt at clarifying this is presented in Ref. 2-1 in which the causes of failure are dichotomized into (1) the severity of the conditions (irrelevant of time) exceeding strength, and (2) cumulative damage (resulting from conditions acting over time) exceeding endurance. In that discussion the meaning of stress is extended from the mechanics viewpoint to include other physical factors in addition to force; however, it is implied there that the term, stress, should be reserved to designate only those conditions which result in the first cause of failure while the conditions in the latter should be referred to as damagers. Stress thus would apply only to conditions causing responses in which cumulative damage can be disregarded. Even with aging, can be considered to restore an item to a form where it can be subjected to a repeated test under the same conditions as the same item.

2.1.2 Aging vs. Nonaging

Aging is a nonreversible process that occurs with the passage of time and results in the accumulation of damage. Most often, aging can be equated intuitively to expending useful life of an item. However, as in the case of annealing (negative damage), not all aging is detrimental and may thus actually improve the item and lengthen its life.

It is not always easy to discern whether aging is present because some nonaging processes, for example a reversible response which includes delay, may also require the passage of time and appear to some observers as an aging phenomena. Also, certain items under test may be actually aging without any indication of this on the basis of measured parameters.

2.1.3 Degradation, Drift, and Failure

Degradation and drift are terms typically used to denote changes in a parameter or an item's property over time. We assign no explicit meaning to them; however, degradation generally denotes a detrimental effect while drift typically signifies either a detrimental or a beneficial effect.

Failure is simply a condition of unacceptable behavior of an item based on some prespecified criteria. The criteria are subject to the needs of the designer or the investigator and may range from a simple out-of-tolerance condition to total destruction. Failures may be associated with either reversible or nonreversible responses. Items may be "born failures" as in the case of production defects or caused to fail by aging or application of test conditions.

Failures involving damage can be further grouped into simple categories. The first is similar to the voltage breakdown of a transistor: Voltages up to the breakdown point produce negligible damage (if the voltage is removed before breakdown the transistor is as good as it ever was) but as the voltage exceeds the critical value the transistor is essentially destroyed. Another case occurs, for example, in dielectric breakdown at reasonable voltages: The application of an electric field at a given temperature degrades the dielectric; if the electric field is removed the degradation still remains; if the electric field is applied again the degradation picks up where it left off; this process can continue until finally the dielectric breaks down. Obviously not all failures can be uniquely classified into simple categories; the intent is to help organize ideas.

There is often reference to specific types of failure such as catastrophic, drift, degradation, and chance or random. Misconceptions often prevail about the designation of chance failures and this is discussed further in Sec. A.9 of the Appendix. The major concern here is the distinction between catastrophic and drift (or degradation) failures. The designation, catastrophic failure, usually refers to a very rapid or abrupt change in an item's characteristics such that it ceases to function as intended. Degradation or drift failure, on the other hand, generally implies a slow change of characteristics such that at some point or level of drift or degradation the item can be said to fail. The distinction between catastrophic or drift(or degradation)failures is not always obvious. What is a slow change to one observer may be an abrupt change to another and for any one observer there are usually "in-betweens" that do not seem to fall in either category. Nevertheless failure is failure whether slow or rapid and in subsequent discussion there will usually be little emphasis on making a distinction.

2.1.4 Modes and Mechanisms

Reference is often made to modes and mechanisms of failure, degradation, drift, aging, damage, and destruction. The terms mechanisms and modes used in this respect are often confused in use and meaning. For example, it is not uncommon to find reference to an electrical short of a capacitor as both a failure mode and a failure mechanism in the same discussion. A mode is the observed way in which an item fails or degrades, etc.; the mechanism is the thing in the item responsible for the mode. For the capacitor, for example, the failure mode may be the electrical short while the mechanism may be structural deformity.

The two terms are strictly relative; the failure mechanisms for one observer may be the failure mode of another and vice versa. To an equipment supplier the loss of output of the equipment can be the failure mode with the capacitor short representing the failure mechanism. To the package designer the structural deformity of the capacitor may be the failure mode, and material fatigue may be the failure mechanism.

2.2 The Test Conditions

Some tests are concerned only with behavior under existing conditions. Some require complex generation and control capability for the conditions. Generally, the more sophisticated a test, the more attention is required of the conditions.

Test conditions in general refer to the total environment of the item during the test including signal and power inputs, loads, and the physical environment plus any special requirements for the item such as size, composition or orientation. We will, however, often refer to a single factor such as vibration, load, input voltage, or orientation as a test condition which comprises, of course, only a small part of the total environment.

though the terminology suggested in Ref. 2-1 is not followed in this report, it is recommended reading for those wishing to explore the concepts in more depth.

The generalization of the meaning of stress is allowed in this report. The term, stress, thus will refer to any condition or set of conditions which potentially cause failure or which cause any detrimental effect such as aging or degradation of performance.

When important, the cause of the responses considered in a discussion will be made clear. All responses involving aging are emphasized as a cumulative damage phenomena requiring a stress-time (stress acting over time) condition. However, stress-time conditions can also be required for certain nonaging responses as in environmental profile simulation.

Note that it is not the stress itself but the severity* of the stress with which we are really concerned. The higher the severity level the more damage is likely to be done to an item in a given time or the more likely it is to fail. The only way of knowing whether a particular set of stresses produces a higher severity level is to know whether in fact the system is more likely to fail or is being damaged at a greater rate. There are circumstances, electrical contacts for example, where increasing the voltage or the current being carried may actually improve the performance. Yet voltage and current are ordinarily considered to be stresses. There also are situations wherein increasing the temperature will improve the life of the equipment, especially if by so doing it is generally kept drier.

2.2.3 Nonstressing Conditions

Test signal inputs to an amplifier necessary to measure its gain are not normally considered a stress; neither are the nominal supply voltages and normal operating temperatures. But who is to say they are not stressing the item? As the input signal level to the amplifier is continually increased, something eventually has to be stressed. Measuring the drift rate of conventional gyros certainly causes wear of the bearings if only by a negligible amount.

Practically speaking, some conditions especially when kept within design limits, do not stress the items. But there is some point of transition when any condition can become a stress. Generally, this report will treat each factor, when considered singly or in combination with other factors, as either a stress condition or a nonstressing condition.

---

* The phrase, stress severity, is meant to be very general.

2.2.4 Single Factors vs. Multiple Factors

A fixed temperature test or a fixed altitude test of an item is concerned only with the effect of a single factor, either the temperature level or the pressure level in this case. A test to treat the combined effects of several types of conditions, say a temperature-altitude test, introduces multiple factors. Multiple factors also are introduced when more than one parameter is needed to describe a particular type of condition. Simulation of a temperature profile for example, may require being concerned with temperature levels, rates of change, and durations at given levels.

The treatment of single factors is usually quite simple. The single required parameter assumes one or, at most, several fixed values or is treated as a random variable having certain statistical characteristics such as range of the variable or mean and variance of the distribution. When the single factor becomes a time-varying function, the methods for treating it are described in Sec. 2.2.5.

Multiple factors introduce considerably more complexity. Additional consideration is required to account for the relationship among the factors and the effect of factors acting in combination, whether sequentially or simultaneously.

Consider the case of the two parameters of temperature, T, and altitude, h, having ranges (T_{min}, T_{max}) and (h_{min}, h_{max}) respectively. Figure 2-1 illustrates three of several possible ways representing relationships between them. When there is no known relationship, one can only define a region assumed or known to contain all combinations as illustrated in the upper diagram. The other extreme as shown in the lower diagram is represented by a deterministic relationship between the two parameters. Between these extremes, parameter values may only be related by known or assumed statistical properties. The central diagram illustrates one way of representing this. Each T, h coordinate pair thus has some probability of occurrence.

The extension of these concepts to more than two parameters is straightforward in principle; practically, however, sophisticated treatment of even two becomes difficult requiring special statistical and mathematical skills. More detailed discussion of treating relationships among multiple factors is presented in Sec. A.11 of the Appendix.

It should be noted that regardless of how well the relationships among parameters are known, they say nothing about their combined effect on the item to which they are applied. Temperature and/or altitude acting separately on an item may constitute a stress, each with its own severity characteristics. But the simultaneous application also represents a stress the severity characteristics of which may

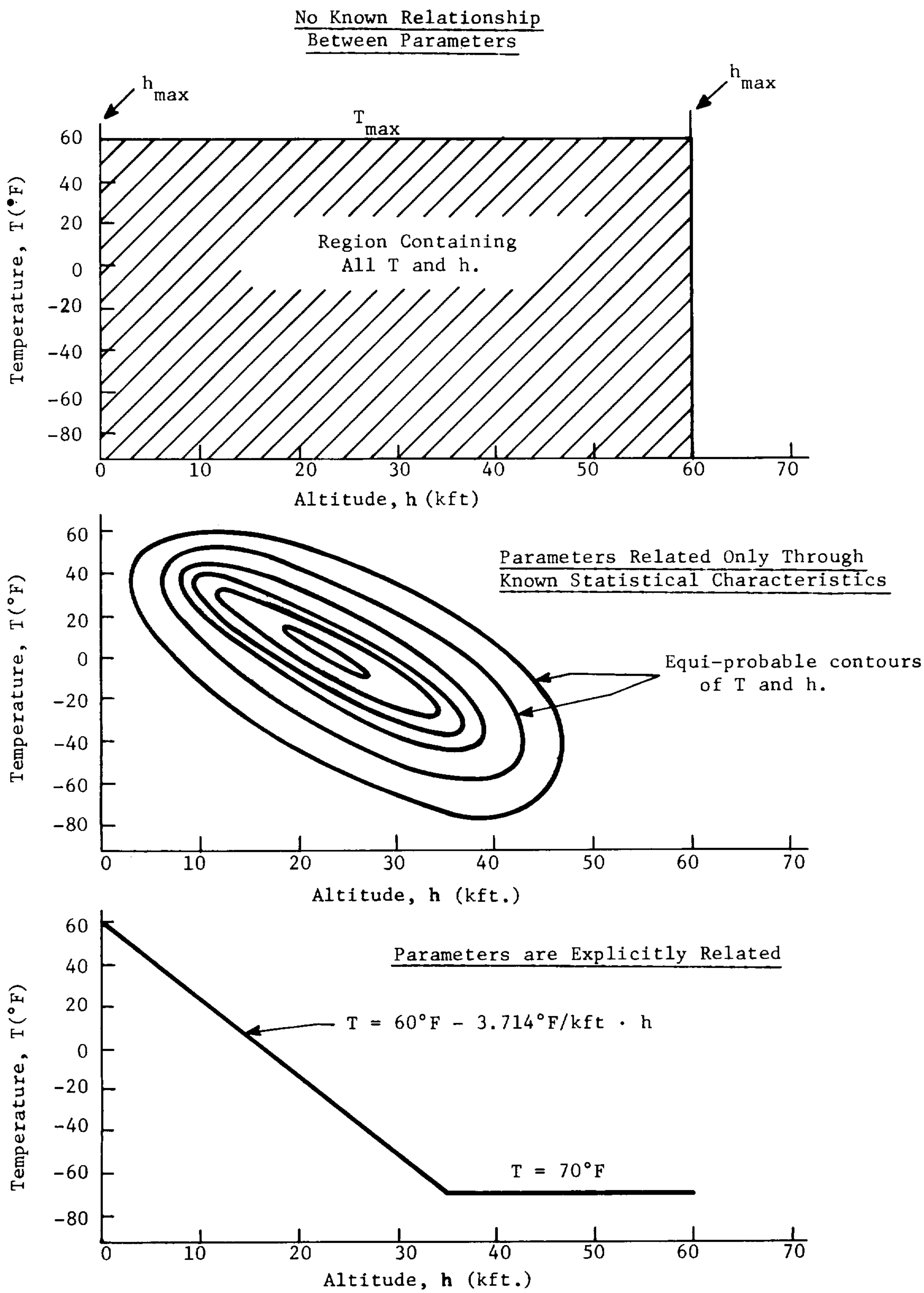


Figure 2-1. Illustration of Three Ways of Representing Multiple Factors

not be closely related to those of each acting independently. Further discussion on this effect also appears in Sec. A.11 of the Appendix.

2.2.5 Time-Dependent Conditions

The time dependency of conditions during a test becomes important when:

(1) effects of aging must be considered,

(2) there are transients or delays in the response to be measured,

(3) the response is dependent upon rate of change of applied conditions, or

(4) behavior over a profile of conditions is of interest.

Test conditions are often controlled to make some of these factors unimportant. For example, during static testing initial transients can be eliminated or minimized by a gradual transition from one level of a parameter to the next. Also, in stress-to-failure testing the stress can be applied sufficiently slowly so that strain rate is not a significant problem.

Descriptions of time-dependent conditions are considered for three time periods:

(1) the period during which initial conditions are established,

(2) the period during the test, and

(3) the period at test termination.

Figure 2-2 summarizes the various ways a single parameter may vary or be controlled during these periods. The parameter illustrated may represent one descriptive parameter of any type of condition, for example, temperature level, rate of change of temperature, power spectral density of random vibration, or saline concentration of salt spray.

A special precaution is noted when varying parameters of frequency-dependent functions. For example, random vibration is typically described by the power spectrum of a stationary random process. A simple spectrum is

$$\Phi(\omega) = \frac{\omega_c^2 \, \Phi(o)}{\omega_c^2 + \omega^2}$$

where $\Phi(\omega)$ is the proper spectral density as a function of frequency ω and ω_c is a measure of bandwidth. To change the characteristics such as power spectral density and bandwidth, $\Phi(o)$ or ω_c may be programmed to change with time. This mixing of time and frequency renders the above simple description of the spectrum invalid during changes. The true spectrum during transition is much more complex than that represented above, and the amount of departure from the given representation becomes more severe with increasing rates of change of the parameters. In certain applications, of course, any rate of change that can be achieved is acceptable. What is

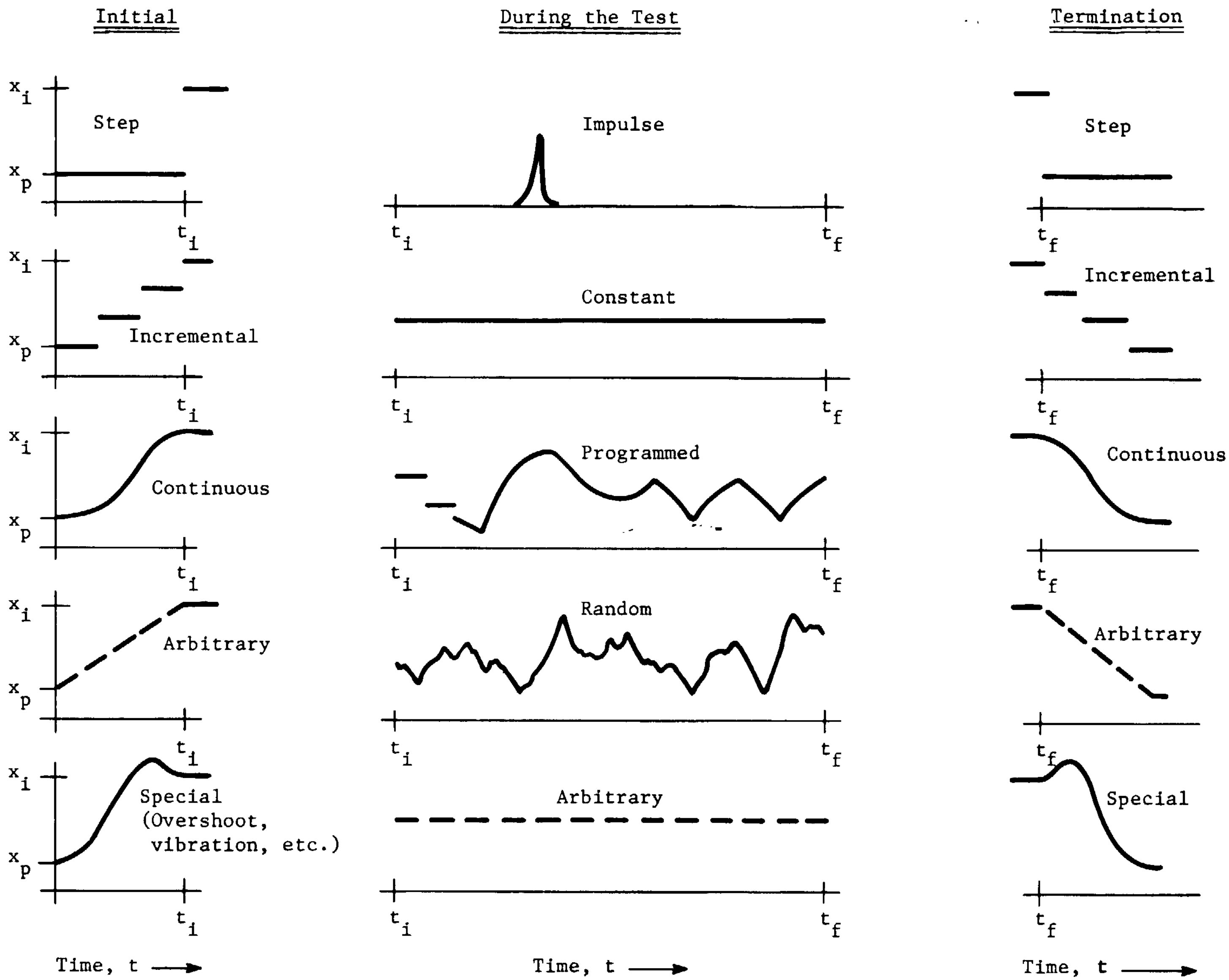


Figure 2-2. Time Descriptions of Test Condition Parameters

important from a practical viewpoint is the recognition that the specified spectrum may only be a gross approximation of what actually exists.

Initial Conditions

Some tests require no change. As illustrated in Fig. 2-2, the simplest change is a step function; however, the step is not always easily obtained nor is it always desirable. It is frequently employed deliberately to compare experimental transient responses with analytical ones for analysis of dynamic performance. A servo loop response to a step input is a typical example.

Attempts to suddenly load an item, however, can sometimes lead to undesirable effects. Consider a test to measure the structural deflection of a beam under static loading. With loads applied gradually the deflection builds up gradually to its final value. Sudden applications of large loads, however, introduce mechanical shock which may cause severe transients or a strain response greater than the yield point. Furthermore, the strain itself during the load increase may depend upon the rate of increase of load.

Undesirable effects could be similarly described for other types of equipment and test conditions. Incremental or continuous (linear or non-linear) changes, as illustrated in Fig. 2-2 are often used to avoid such effects. For example, in Ref. 2-2, the procedure specified to achieve constant acceleration requires a gradual increase to the specified value in not less than twenty seconds. Often due to physical limitations of test equipment, only gradual transitions are possible. A vibration table under heavy load, for example, may require several cycles to achieve its full amplitude of vibration. Instructions for controlling initial conditions to prevent unwanted responses appear frequently in test specification and standards.

In many cases the manner of achieving initial condition is arbitrary. When the engineer is at liberty to select his own procedures (as in informal tests) it may be beneficial to deliberately depart from standard practices. Just trying different test methods can often be informative and may aid in identifying that a specific method such as overshoot is really needed.

Conditions During the Test

Various forms of behavior for a test condition parameter are illustrated in the central portion of Fig. 2-2. Some tests involve combinations of these forms. For example, a random process variation of a parameter may be superimposed on a deterministic behavior, and certain tests may require precise control of some parameters while allowing others to assume arbitrary values.

Impulse Conditions. Impulses are typified by mechanical shock, explosive environment, transient radiation pulses, and electrical power transients. The impulse function is typically defined analytically [Ref. 2-3] to be a function having infinitesimal width, infinite height, and unit area. Clearly, the unit impulse function cannot be achieved physically, but it can often be approximated adequately to compare experimental responses with those obtained analytically. The approximated unit impulse is often substituted for the input step function when a step causes difficulty in observing the response.

In most tests employing impulse type conditions, it is not desired to approximate a unit impulse but rather to simulate the impulses occurring during normal operation. Consider a mechanical shock pulse. It is described in test specifications and standards by peak acceleration and duration, such as 50g for 6 to 10 msec. To readers unfamiliar with shock specifications and test methods, this might imply a rectangular waveform. Shock pulses, however, have various shapes which depend upon the characteristics of the shock machine. As described in Ref. 2-4, for example, shape can be controlled by adjusting damping characteristics.

A precise representation of mechanical shock pulses is often not required. For example, whether the waveform is rectangular, triangular, or a half sinusoid may not really be important; simply achieving an impulse which is approximately equivalent to the one specified may be all that is needed. The starting height of a simple drop test to obtain a shock pulse may be quickly computed with equations of motion by assuming any of several simple pulse waveforms. Calculations of this type are included in Ref. 2-4 and are extended to include damping effects.

Most shock pulse specifications are more explicit than the example presented above; Ref. 2-2 specifies half sinusoidal pulses of 30g peak amplitude with a total time between 10 and 15 msec. MIL-STD-810A [Ref. 2-2] prescribes a sawtooth shock pulse configuration. Although the tolerance requirements indicated for that waveform are more stringent than needed for many applications, it can be readily achieved by most shock machines.

Fixed Conditions. Fixed conditions are the most common conditions employed in testing. Fixed conditions have the disadvantage of rarely representing operating conditions. Their major advantage is convenience and ease of test control. However, when operating conditions are not known, tests under fixed conditions is a way to collect valuable data for extrapolation to operating conditions when they become known.

Programmed Conditions. During many tests, conditions are varied by programming values of parameters. Operational profile simulation is the most obvious example. However, the concept applies as well to other programmed conditions, for example, temperature cycling and thermal shock, varying vibration frequencies and amplitudes to scan for maximum severity of vibration stress, and step-stress or progressive-stress for accelerated testing. The functions which define parameter variations during the test may produce step changes, linear changes, non-linear changes, or periodic behavior. Theoretically, these functions can be combined to make up very complicated profiles; however, practical limitations imposed by test equipment usually limit them to fairly simple functions.

Operational profile simulation finds its largest application in reliability demonstration. When aging is involved, it is important to program the changes in the same order that they occur during operation. When aging is not involved, the order is not generally as important, and it may even be desirable to purposely change the order or eliminate some of the profile. Suppose for example that the effect of a launch vehicle acceleration profile on stable platform drift is of interest. A rocket sled can provide high uniaxial accelerations but is very difficult to program to simulate a complex acceleration profile. However, a series of rocket sled tests with the platform operated in different orientations can be used to obtain drift rates as functions of acceleration. These rates can be used analytically to predict the behavior during a launch profile.

Instead of directly simulating the temperature profile, the effect of slowly-varying temperatures on electronic equipment drift may be investigated by static measurements of drift as a function of temperature. More rapid changes may require measuring, in addition, the effects of rate of change. From static and rate measurements the predictions for temperature profile operation often can be adequately made.

Although the designer and the customer are generally much more satisfied with an item when its performance can be observed under a simulated operational profile, cost savings accrue in many cases if the profile does not have to be simulated. Intelligent planning and evaluation of alternatives are required to determine whether the profile should be simulated.

Random Conditions. Many tests assume certain random characteristics. A physical quantity (such as force, acceleration, temperature, or voltage) representing a test condition is described by a random process if, as a function of time, it possesses random characteristics of significant magnitude. The different types of random processes are detailed in Sec. A.13 of the Appendix; in summary they are:

(1) Random series - impulses having random times of occurrence, random waveform characteristics (such as rise time or amplitude), or both;
(2) Constants which have random levels - simple random variables;
(3) Deterministic random processes - functions having recognizable functional form but also certain random characteristics;

(4) Stationary random processes - functions having no recognizable functional form and for which the statistical properties do not change with time; and

(5) Nonstationary random processes - functions having no recognizable form and for which the statistical properties change with time.

From the point-of-view of controlling test conditions, the treatment of certain random conditions during a test may really be no different than for the first three types of conditions described above. Consider random vibration, for example. As noted earlier, it is typically specified by power spectral density and bandpass cut-off frequencies for a stationary random process. Controlling the specified parameters gives rise to a random process. Assuming that the vibration is typically generated by power amplifying the output of a noise tube and driving a vibration table, controlling the rms level and the bandwidth gives the desired spectrum. From the point-of-view of control this is really no different from programming simple factors like temperature and pressure, since it is the parameters of the process and not the process itself that are being controlled. That is, the parameters are controlled explicitly while the underlying random process is implicit.

Conversely, if a test is being conducted under uncontrolled conditions (as environment is uncontrolled in field testing) explicit description of the random conditions may become very important. Treatment then must be more statistically oriented; sample functions have to be obtained for estimating the statistical characteristics. With random vibration, for example, the determination of whether the random process is stationary or nonstationary has to be made.

Arbitrary Conditions. Sometimes we are not really concerned with specific values of test condition parameters. For example, in certain field tests, prelaunch checkout and production testing, we may simply assume that any condition is representative without caring whether a parameter is constant or varying according to some functional form. Under such arbitrary conditions the test becomes merely a measurement of behavior.

Terminating Test Conditions

The possible types of conditions during test terminations are similar to those for initial conditions; however, for most tests the manner of removing test conditions from the item or returning to normal ambient levels is arbitrary. The possibility of damage to the item or to test facilities is generally the only important question to be considered.

2.3 Measurements

Measurements are basic to testing in that they provide the record of the response. The only needed clarification at this point is the explanation of the basic types and the fundamental viewpoints.

2.3.1 Attributes Measurements vs. Variables Measurements

A measurement of a parameter is either an attributes measurement or a variables measurement. An attributes measurement is concerned only with determining which one of two possible results is present. For example, an item is either good or bad, it either fails a test or does not fail a test, a performance check indicates either "go" or "no-go". The result of an attributes measurement is always an either-or situation. In the terminology of the statistician, we are dealing with Bernoulli trials. Certain responses can only provide an attributes measurement; a test to determine merely whether a flash bulb operates results in a response of this type. Parameters having continuous values can be subject to attributes measurements such as being in- or out-of-tolerance or being above or below a discrimination level.

A variables measurement is concerned with determining the specific value of a parameter which can, in general, assume multiple values. There may be a finite or infinite number of possible values either distributed continuously or at discrete levels over a range of values. Measurement of resistance of a resistor or the dimension of a structural member are typically variables measurements even though, as discussed above, these can be, and frequently are, reduced to attributes measurements as when screening out items which are not in tolerance.

When one has a choice between the two types, which should he use? Variables measurements are generally more troublesome to make. For example, it takes longer to read and record the value indicated on the scale of a voltmeter than to merely determine whether it is in the green or red region. If the results needed for the immediate purpose at hand can be obtained with attributes data and there is no beneficial use of variables data, then the simplest method is obviously the one to use. But foresight on possible uses of variables data such as "post mortem" evaluation has often been a saving grace.

2.3.2 Measurement of Responses

There are only two basic points-of-view in test measurements depending upon what one is trying to learn from a test. These are:

(1) "Given the conditions, what is the response?", and

(2) "Given the response, what are the conditions?".

The first is typified by structural proof tests, measuring amplifier gain at high temperatures, measuring the life of a steel specimen subjected to fatigue stress, and power burn-in of electronic devices. With this point of view the emphasis is on the dependent variables as a function of time, the conditions, or the item's characteristics. It is often referred to as the direct response problem.

The second situation is typical of measurement of strength (i.e., the maximum severity of stress which an item can withstand without failing), measurement of the levels of temperature and input voltage acting simultaneously to give an output voltage variation of an inverter, and determining the maximum allowable vibration level for an accelerometer to survive a given duration. This viewpoint is often referred to as the inverse response problem. Another designation used is response surface determination when there is more than one independent variable of interest.

Some testing applications may involve both points-of-view. For example, there is a testing approach called gradient-seeking which is useful for determining the conditions which optimize performance. In this, the response is observed at different sets of conditions which are changed in a direction which improves performance. Both points-of-view are also commonly involved in reliability testing.

References

2-1. Evans, R. A.: Stress vs. Damage. Proc. 1967 Symp. on Rel. (Washington, D.C.), Jan 1967, pp. 633-635.

2-2. Anon.: Military Standard, Environmental Test Methods for Aerospace and Ground Equipment. MIL-STD-810A, June 1964.

2-3. Davenport, Wilbur B., Jr.; and Root, William L.: An Introduction to the Theory of Random Signals and Noise. McGraw-Hill Book Co., Inc., 1958.

2-4. Peeler, F. C.: Environmental Testing. Chapter 8 of Reliability Engineering for Electronic Systems, Richard H. Myers, Kam L. Wong, and Harold M. Gordy, ed., John Wiley & Sons, Inc., 1964.

3. Test Classifications

A significant barrier to understanding testing is the plethora of test names in common use. A given test name often means different things to different people, and two tests with the same objectives sometimes have different names.

Any form of classification is simply a way of viewing testing and there are many ways of doing just this. This section considers several ways; however, the emphasis is on the three described first since their relationship has influenced the organization of this report to such a large extent.

3.1 Three Forms of Classification and Report Organization

Most presentations of testing identify one or more major testing areas such as reliability demonstration and acceptance testing and proceed to discuss different methods and approaches for implementing these (cf., Ref. 3-1 and 3-2). Of considerable importance to a thorough understanding of testing is first a familiarity with basic test approaches and what basic problems they can solve. The understanding of their broader applications for program-oriented activities then follows more easily. Three important classifications are thus:

(1) basic test types,

(2) their basic problem uses, and

(3) their program applications.

A tabulation of typical types in each category is presented in Fig. 3-1.

The general relationship indicated in Fig. 3-1 is the basis for organization of this report. Because there is not a one-to-one correspondence among classifications, different classifications are treated in separate sections.

Note especially, however, that there is a distinction made between aging and nonaging test types. The importance of aging has a large influence on the way items are tested. When aging is present different things are generally of interest, different responses are encountered, and different considerations of the effect of test on the item are introduced. Generally, the distinction depends upon the intentional effect of the test time on the item tested. Any test can be characterized as either againg an item intentionally or aging it only unintentionally if at all. Thus an engine run-in test is an aging type, for it is the intent to alter (and in this case actually improve) the performance of the engine by operating it for an appropriate length of time. Conversely, a test to measure the gain stability of an amplifier over a temperature range is not intended to alter the gain of the amplifier due to time hence is a nonaging test although some gain change could result from aging produced by time-temperature effects.

It must be noted that it is not always immediately clear whether a response is due to aging. For example, creep of metals is a phenomena that may be viewed as

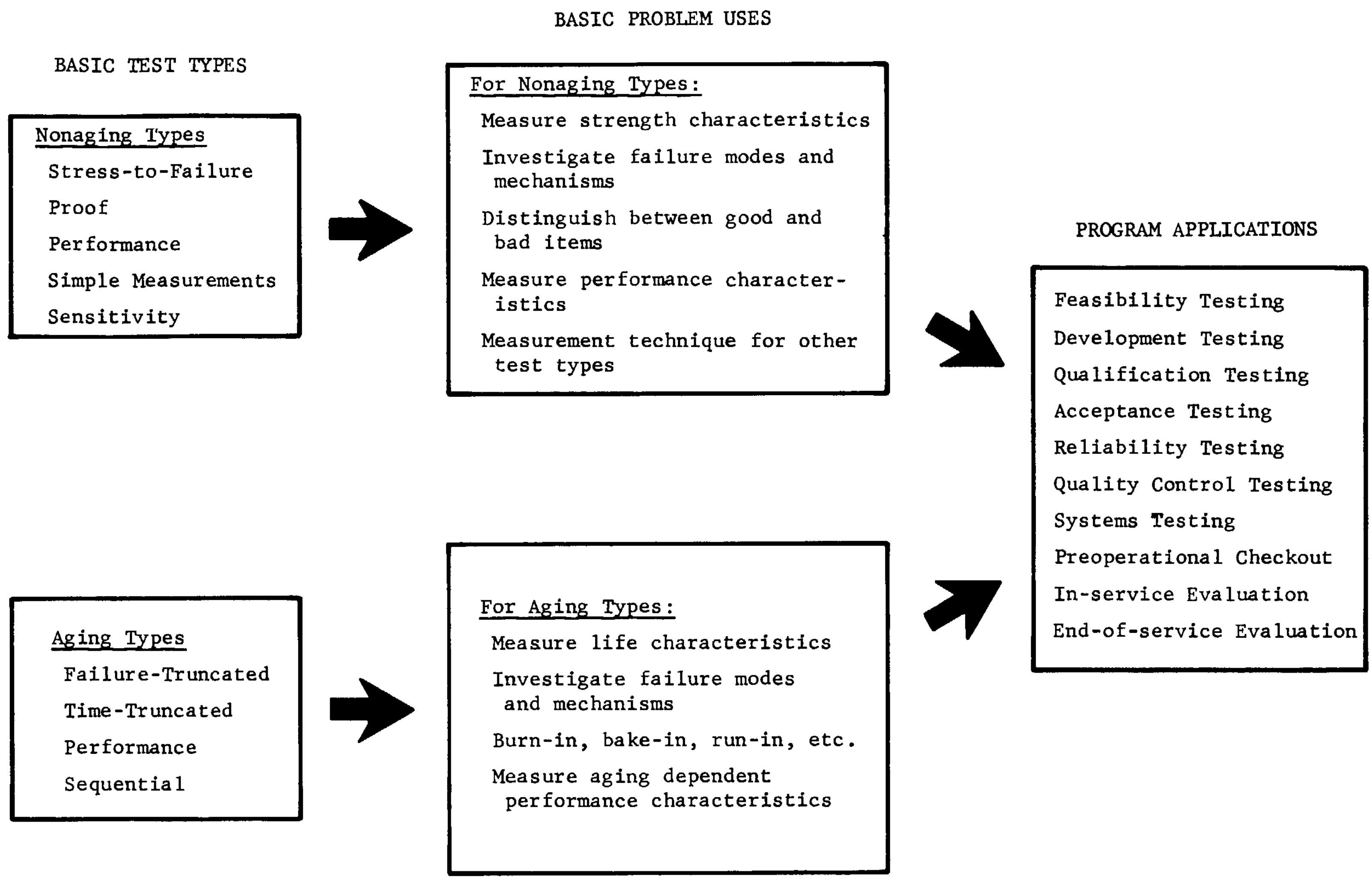


Figure 3-1. Three Test Classifications and Their General Relationship

either a nonaging or an aging response. However, tests involving such responses usually belong to one class or the other. A test to measure strength of the metal involving creep would logically be a nonaging type whereas one to measure the duration of the creep process prior to failure would be an aging type.

For report organization the earlier sections, Secs. 1 - 4, serve as introductory material common to all later discussion. Sections 5 - 8 are then devoted to descriptions of the basic test types and their uses. Sections 5 and 6 specifically treat the nonaging types and emphasis is on clarification of concepts applicable to each and practical problems in implementing them. The aging types are similarly treated in Secs. 7 and 8.

Following these discussions several sections, Secs. 9 - 12, are devoted to several miscellaneous but popular topics which do not belong to the orderly classification employed. Section 9, for example, describes nondestructive testing (NDT), the special technology developed to perform measurements more conveniently and/or less destructively than conventional methods allow. Environmental testing, described in Sec. 10, is a subject area applicable to all test types, basic uses, and program applications. Accelerated testing is treated in Sec. 11; this is applicable to all aging test types. A brief discussion of field testing in Sec. 12 completes the treatment of miscellaneous topics. Section 13 is then devoted to the various major applications of testing in hardware programs. The testing specifically for reliability is discussed separately in greater depth in Sec. 14.

Because the three classifications introduced are considered in more depth in later sections no further treatment is given in this section.

3.2 Other Forms of Test Classification

Tests can be classified in a considerable variety of ways in addition to those classifications given in Sec. 2.1. Some of these classifications and the test names likely to be found in them are outlined below. Some of the names appear in more than one classification which indicates the different viewpoints that the names may reflect. References 3-3 through 3-5 are additional sources for test designations and classifications.

3.2.1 Classification by Specific Purpose

Some test names are indicative of specific purposes and citing the test name reveals the specific problem being resolved. Test names in this category generally reflect the basic problem areas introduced in the previous section. Test names illustrative of specific purposes are:

proof test	go, no-go test	burn-in test
life test	run-in test	functional check
strength test	failure mode test	performance check.

3.2.2 Classification by General Purpose

Some test names denote a more general purpose and do not indicate what specific problem is being treated. For example, an environmental test indicates that environments are being treated but does not indicate whether performance, life, strength, or just a functional check is important. A number of test names of this type are:

environmental test	flight test	acceptance test
reliability test	research test	quality control test
development test	assurance test	systems integration test
engineering test	specification test	feasibility test
screening	storage test	in-service evaluation
tolerance test	evaluation test	end-of-service evaluation.

3.2.3 Classification by Severity of Test Conditions

Some test names indicate the level of severity of test conditions but do not help in identifying the purpose of the test. Typical names are:

overstress test	accelerated test	normal stress test
peripheral test	understress test	high temperature test
low voltage test.		

3.2.4 Classification by Test Design

The statistical design of a test often influences the name designated. Some of these are self-explanatory while with others the name may merely indicate the name of the person who developed the technique or may just generally designate the approach. Some typical test designations pertaining to test design are:

nonsequential	random balance design	matrix testing
sequential	20 zil design	factorial design
up-and-down method	Langlie design	Alexander design
Robbins-Monroe design	split-the-difference design	probit design.

3.2.5 Classification by Method of Analysis

Names in this category are not really test names but rather indicate methods of treating data from tests of various types. This classification category is included to dispel confusion in this respect. Typical names pertaining to the type of analysis and not to the type of test are:

linear discriminant	probit method
sequential analysis	hypothesis testing
Dixon-Mood analysis	maximum likelihood analysis.

3.2.6 Summary of Additional Forms

Several additional test classifications are possible. Those given below plus those already given include all that are likely to be encountered.

Classification by Method of Applying Test Conditions

constant stress test
step-stress test
progressive stress test
profile test.
programmed stress test
random stress test
stress-to-failure (test-to-failure)

Classification by Proportion of Population Tested

sampling
100% screening
screening (title alone implies 100%)

Classification by Level of Assembly

material test
part test
component test
subassembly test
assembly test
subsystem test
system test.

Classification by Degree of Formality

formal test
informal test
exploratory test (indicates informality).
probe test (informal)
search test (informal)

Classification by Degree of Destruction

destructive test
nondestructive test (general sense)
nondestructive test (special methods to circumvent destruction).

Classification by Type of Hardware Function

electrical test
structural test
mechanical test.

Classification by Method of Terminating Test

time-truncated test
failure-truncated test
sequential test (implies that termination depends upon cumulative results)
stress-to-failure (test-to-failure) test.

Classification by Type of Measurement

variables test
attributes test
parameter test
go, no-go test.

Classification by Item Maturity

breadboard test

engineering model test

production model test.

boiler-plate model test

prototype test

References

3-1. Lloyd, David K.; and Lipow, Myron: Reliability: Management, Methods and Mathematics. Prentice-Hall, Inc., 1962.

3-2. Myers, Richard H.; Wong, Kasu L.; and Gordy, Harold M., ed.: Reliability Engineering for Electronic Systems. John Wiley & Sons, Inc., 1964.

3-3. Simley, Robert W.: Testing Programs. Section 8 of Reliability Handbook, W. Grant Ireson, ed. McGraw-Hill Book Co., Inc., 1966.

3-4. MacCarley, John A.: Testing for Reliability. Part II of Reliability Control in Aerospace Equipment Development, SAE Technical Progress Series, Vol. 4. The Macmillan Co., 1963.

3-5. Leary, J. P. : General Outline of Principles Involved in Testing. AF Report No. TOR-469(5303)-5 (DDC No. 484 375). Aerospace Corp., El Segundo, Calif., Feb. 1965.

4. Perspective of Testing

Testing is the experimental approach to problem solving. Its relation to the analytical approach is illustrated in Fig. 4-1. The experimental approach in engineering has traditionally been associated with "cut-and-try" or "build-and-test" procedures. In recent years, however, analytical techniques have become more helpful because of improvements in computing capabilities. This development in no way decreases the importance of testing because the physical prototype itself remains its own most accurate model. Actually, the two approaches are complementary and best serve the engineer in a coordinated role as illustrated by the center path in Fig. 4-1.

4.1 Relation of Testing to the Hardware Development Process

Testing is a key ingredient throughout hardware development. As a design progresses from planning and definition to final operational form, problems continually arise which must be resolved by analysis or testing. The role of testing at any stage of the development cycle is illustrated in Fig. 4-2. At any of these stages the approach to solution of defined problems is formulated and this includes making decisions concerning the part that testing will play.

If adequate physical models are available, testing is generally the more reliable procedure. Even when an analysis approach is pursued, experimental data is a major input.

Good results serve to verify that the design, procedures, and applications of the design are adequate or provide the necessary information to determine the

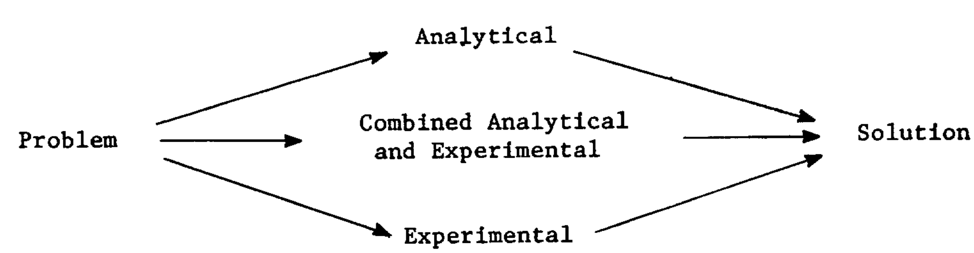


Figure 4-1. Approaches to Problem Solution

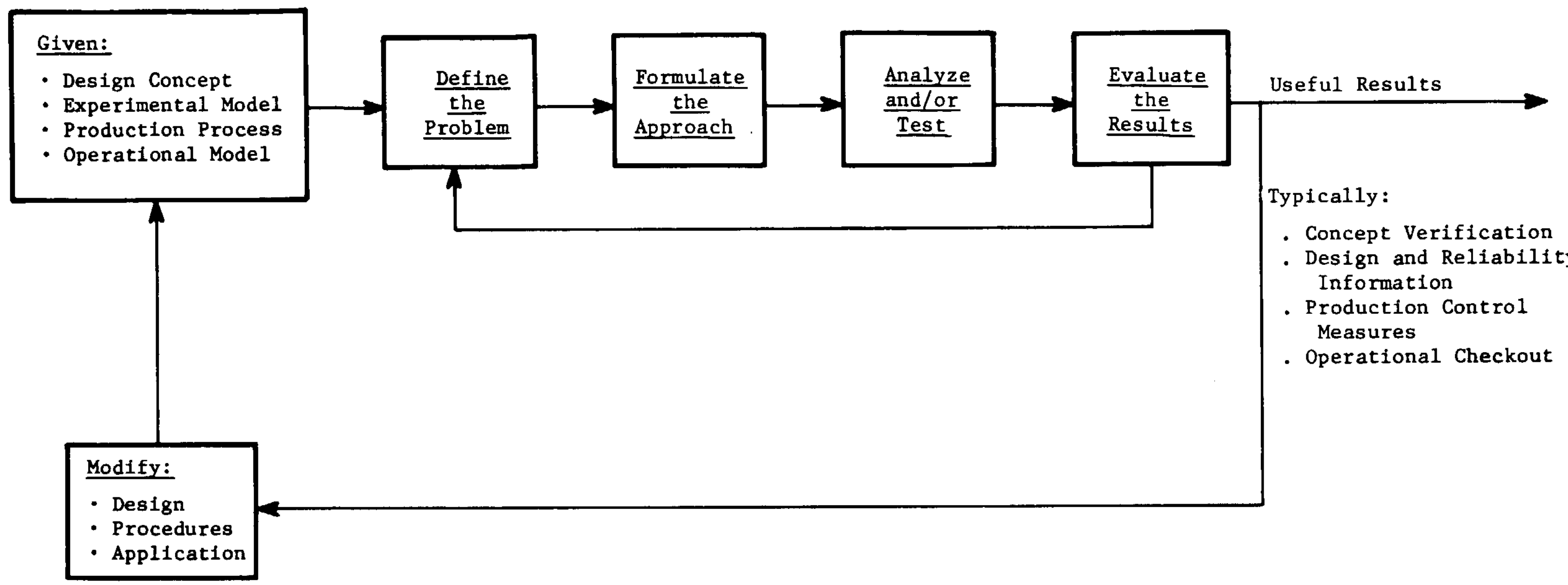


Figure 4-2. Role of Testing in Hardware Development

needed changes to progress the design toward maturity. The results must first be evaluated to establish their validity. Where the evaluation reveals inadequacies, modifications of the problem solving process itself may be in order. As illustrated by the inner loop in Fig. 4-2, this may result in modifying the approach by redefining the problem, reformulating the approach to solution, or revising or continuing the test to improve the information.

4.2 General Test Procedures

Even though there are many purposes for testing, all tests have a common basis in the general procedure followed. This procedure is illustrated in Fig. 4-3, which basically is an expansion of the inner loops in Fig. 4-2 and emphasizes testing rather than analysis.

One of the most practical aspects of testing, often downgraded in most discussions of testing, is that of forming a notion (often preconceived) of how an item behaves; this is just the conceptual model designated by the second block in Fig. 4-3. The soundness of all subsequent test procedures and the test evaluation depend heavily upon this. The model can range from only a qualitative concept of behavior to a precise mathematical description.

With reference to Fig. 4-3, the formulation of specific test objectives is a necessary prerequisite to selecting the appropriate testing approaches and designing the test. For example, the objective of measuring strength of a pressure vessel design will require stress-to-failure testing of several specimens. The test design and analysis of data are often closely linked with certain assumptions. The design of the pressure vessel test, for example, may specify the sample size required to achieve an estimate of the median of the strength distribution at a given confidence level. Furthermore, both the design and analysis may be based on the assumption that the strength distribution is cumulative normal (Gaussian).

Evaluation of results is a very important stage of the procedure. In this, queries pertain mainly to whether the results fit the original concept of behavior and to whether the original problem definition is adequately resolved with the achieved results. Rejection of either requires further decision about how to proceed.

Note that both questions must be answered since it is possible that the results may fit the concept but still not solve the problem at hand--the case of getting the right answer to the wrong problem.

The order of these queries is also important because it is sometimes easy to think that the problem is solved even though the original concept is erroneous. For example, an exposure of a resistor to high temperature for measuring its temperature coefficient (the ratio of reversible resistance change to the change in ambient

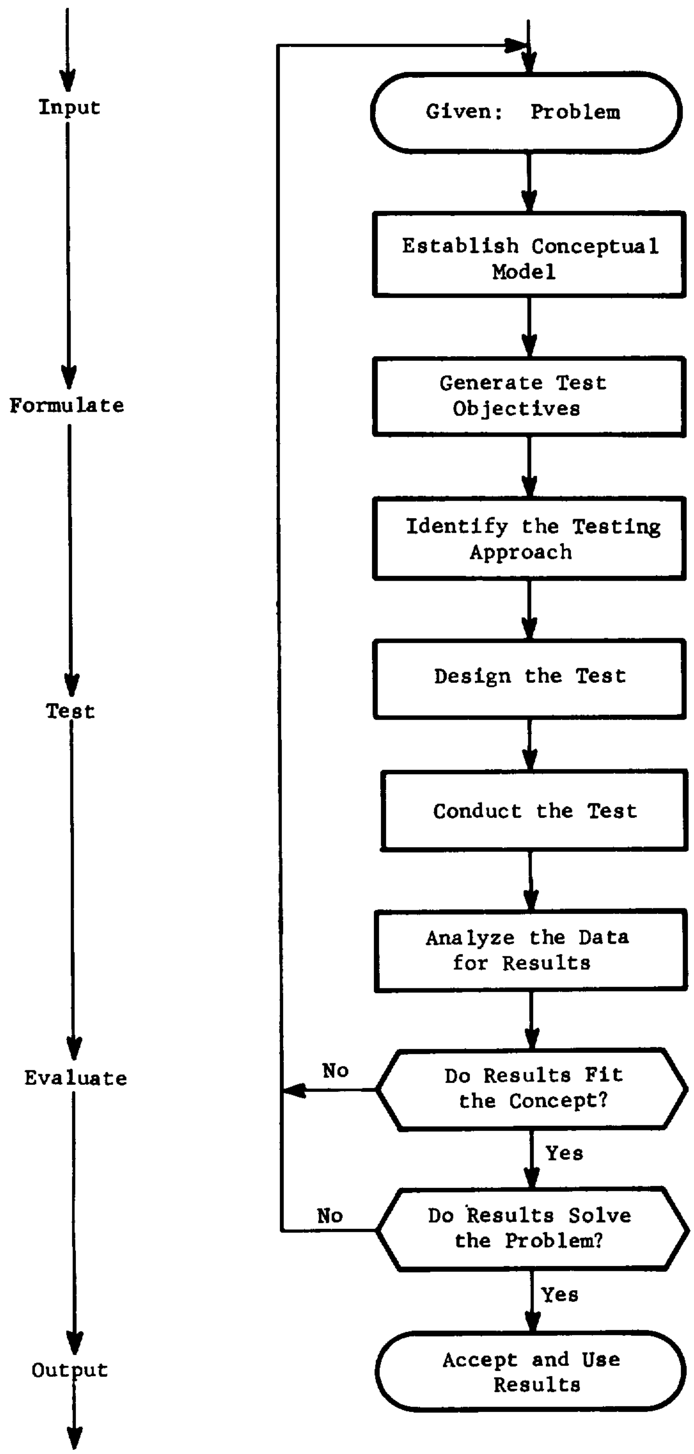


Figure 4-3. The General Test Procedure

temperature) may reveal a true change, but was all of the change reversible? If the conceptual model is found invalid, the problem definition itself may be no longer meaningful. Then obviously, a modification of the problem is in order. Such a modification may take one of several different forms: redefine the problem, retest with a different model, resort to an analytical approach; or abandon the whole situation. If only the model is in error, then a modification of the model and perhaps a retest is in order.

Note that the above decisions do not automatically reject negative results but only those which do not fit the original concept of how the item behaves. Negative results may still fit and in some cases even solve the problem (by contradiction).

It is also possible that conforming results do not solve the original problem, e.g., when too few data points are obtained or when the wrong test conditions are chosen. Before specifying more tests, however, the validity of the original problem definition should be questioned. Problem modification may be required at times even though the original concept is valid. This is more a fault of the experimenter's statement of the problem than improper procedure.

Each block in Fig. 4-3 could be expanded considerably. Conducting the test, for example, involves the tasks of obtaining test specimens, lining up test facilities, setting up and instrumenting the test, and actually taking the measurements. Test conduction activities form a large portion of the effort devoted to testing. Since such activities are fairly routine, the emphasis in this discussion has been on the importance of the conceptual model and the practical evaluation of test results.

4.3 Test Planning

Sound planning is critical to efficient testing. It deserves more depth and breadth than can be inferred from the brief discussions in the previous section. Facilities, personnel skills, instrumentation, and methods of reducing and analyzing data are typical of things considered, but has anyone considered, for example, what happens if the power fails during the test?

Test planning involves the prior consideration of as many of the practical test factors as possible. There are many of these and Table 4-1 illustrates the magnitude and complexity of them.

Amount of Planning Effort

A basic question concerns the proportion of total test effort to devote to planning. The answer itself depends on schedules, experience, and cost and complexity of test specimens. It is often stated simply in testing literature that more

Table 4-1

Practical Factors Related to Testing

<u>Responses</u>

- Reversible vs. nonreversible
- Aging important vs. aging not important
- Single vs. multiple
- Correlation of responses
- Duration
- Continuity (Discrete, intermittent, or continuous)
- Degree of destruction

<u>Nature of Items</u>

- Interruptable response vs. non-interruptable response
- One-shot vs. repeated shot
- Repairable vs. nonrepairable
- Hardware type (electronic, electro-mechanical, structural, etc.)
- Level of assembly (material, piece-part, system, etc.)
- Level of design maturity (bread-board, experimental model, pro-duction item, etc.)
- Stage of life cycle (during fab-rication, pre-application, during application, etc.)

<u>Use of Results</u>

- Estimation vs. hypothesis testing
- Basic use (measure strength, performance check, burn-in, etc.)
- Program related application (development, reliability, acceptance testing, etc.)
- Proportion of items tested
- Degree of formality

<u>Test Conditions</u>

- Stress vs. nonstressing condition
- Effect of stress
- Severity of Stress
- Relation to conditions during normal use
- Single vs. multiple factors
- Importance of control or measurement over time
- Initial condition control or measurement
- Control or measurement at test termination
- Control or measurement precision

<u>Analysis of Data</u>

- Measurement accuracy
- Measurement precision
- Importance of variability
- Statistical model (normal distribution, constant hazard, "white" noise, etc.)
- Statistical confidence
- Engineering confidence
- Necessity of replications

<u>Measurements</u>

- Attributes vs. variables
- Duration
- Continuity (discrete vs. continuous)
- Single vs. multiple
- Precision
- Accuracy
- Frequency
- Variability and uncertainty
- Instrumentation capability
- Calibration of instruments

<u>Resources</u>

- Time
- Cost
- Personnel skills
- Test facilities
- Item availability

<u>Other</u>

- Safety
- Equipment dependability

test planning is needed; clear justification for such statements is usually missing. Nevertheless, situations have arisen when it was obvious that more though to test planning could have provided significant benefits.

Consider, for example, random drift testing of gyros. Drift measurements in all three orthogonal orientations are not necessary if the orientation during application is controlled by serialization. Yet, an equipment manufacturer has been observed to conduct several hour drift tests in three orientations for each gyro during receiving tests, and the fact that some gyros failed from wearout prior to system acceptance indicates that better planning of gyro testing was badly needed.

Generally, the more complex and expensive the test specimens, the greater the proportion of total test effort that should be devoted to planning. Five to twenty percent of the total cost of testing is the typical range. Less than five percent is common when certain test procedures such as blanket screening and qualification testing have become routine. Cases like the one described above for gyro testing justifies wariness of minimum effort in test planning, however. Occasionally, greater than twenty percent of the total test effort must be applied to planning for tests of expensive items. Static firings of boosters, for example, requires extensive planning.

There is no fixed criterion. Every situation has to be considered on its own merit. The decision is guided to a large extent by the importance of the considerations discussed below in the test being implemented.

The discussions that follow highlight some important considerations for illustration. These considerations are highly interrelated, and it is recognized that for numerous tests many of the considerations are obvious. However, the "obvious" has many times caused difficulty, and it is in the interest of preventing the recurrence of such difficulty that some of the discussion is given.

4.3.1 Problem Definition

When a test is to be conducted, formal problem statements do not always exist. A circuit designer may simply be wondering what will happen if he grounds the output of his breadboarded amplifier. The same question may appear more formally as a checklist item in a formal failure mode and effects analysis. Generally, the higher the level of assembly of the item and the more mature its design, the more formal will be the problem statement.

In test planning the most basic consideration regarding the problem definition is whether the problem should be resolved by testing--just how useful will the test results be? Testing may be a faster and cheaper way of getting an answer but not necessarily of getting the appropriate answer. It does not always solve the real problem, e.g., when the operational environment cannot be closely simulated.

Alternate approaches to testing are analytical treatment or "build-and-operate". The penalties for not testing should be considered in selecting trade-offs between testing and its alternatives. For example, if the test item is flight hardware, an inadvertent error in test procedures could turn out to be more costly than not testing. Every situation has to be treated on its own merit.

4.3.2 The Conceptual Model

This is one of the most important considerations in test planning, for it strongly influences many other test decisions. The anticipated responses of individual items is first considered because they aid in determining which basic test approach is required. Any important uncertainty in this behavior or significant variability among items leads then to statistical treatment for a sample of items.

With respect to the tabulation in Table 4-1 some important factors related to item behavior are:

(1) type of response,
(2) test condition severity,
(3) relation of test conditions to operational conditions,
(4) manner of controlling test parameter values,
(5) degree of item destruction,
(6) response duration,
(7) type of hardware tested,
(8) relation of item operation during test to operation during application,
(9) failure mode complexity,
(10) relation of multiple failure modes, and
(11) criteria for terminating test.

It is often said that an item should be operated in the same mode during a test as that during normal application, but such operation during a test is not always possible nor is it always desirable. A solid propellant rocket obviously cannot be flight certified by measuring its thrust and is an example of the special case of triggered response, one-shot items have a noninterruptable response, none of which can be tested in their normal operational mode without destruction. Other reasons arise for testing in modes different from the normal operational mode--due to safety, aging, cost, or limitation of facilities. While behavior during intended application may be the eventual chief interest, the different modes of operation and special conditions required for the test often lead to different concepts of how items behave. For example, attempts to accelerate aging can radically change the behavior of items during testing from that expected during intended operation. However, even when behavior during a test is different, it still can be informative about behavior during intended application if a meaningful correlation can be made between test conditions and application conditions.

When items cannot be operated in the normal mode for testing purposes, other adequate tests or measurements can often be substituted. The rocket, for example, may be inspected by special nondestructive testing techniques such as ultrasonic or X-ray scanning to obtain measurements of parameters such as the weight, the uniformity, and the dimensions of the solid propellant charge which then can be correlated with thrust or other operating characteristics of interest.

Sometimes advantage can be taken of different but related modes of operation. Operating an integrating gyro in its normal open loop mode for a long term drift test is undesirable because the drift continues to grow without bounds. A common technique to circumvent this problem is to provide compensating torque about the output axis through an electrical feedback from the output signal generator of the gyro, which in effect converts the integrating gyro into a rate gyro with constraints on the drift. The drift characteristics of the rate gyro can be correlated with those of the integrating gyro via the transfer functions of the two configurations. Similarly, shear strength of metals may be estimated from tensile strength tests and known correlating functions. It is clear that one should make doubly sure when using such correlations that the functions used for establishing correlation between the different modes adequately reflect the real physics of the situation.

The testing of expensive items such as large rockets and satellites having a noninterruptable response, one-shot type of operation have notably posed difficult problems. Operational tests of such items are expensive, and typically for aerospace systems there are at most several-of-a-kind of such items. Several testing approaches are used to test such items. One is to rely heavily on testing components most of which can be treated conventionally. Another is to specifically design the test item to permit interruptable responses for test purposes. For example, a liquid propulsion engine is designed to operate in this manner. Another approach is to design the test in such a manner that the property of interest can be measured, without the necessity of operating in the destructive mode. The problem of correlating responses in such different modes was discussed earlier.

4.3.3 The Test Objectives

Specifying the objectives of a test typically entails clarifying the problem definition based on knowledge of item behavior. For example, the original problem of needing to know the expected life of an item may get translated to estimating the failure rate of a population from a 1000 hour life test of several items at 125°C. Not all problems can be reduced to single simple objectives. The requirement for adequately demonstrating an item's reliability may, for example, demand life tests, strength tests, performance tests, and simple measurements of parameters. Such situations usually result in a multiplicity of simple objectives, each of which defines a unique problem whose solution contributes toward solution of the original problem.

An explicit statement of objectives is usually the first item in a test specification or test procedures document. Typical objectives are:

(1) to determine the minimum discernible signal sensitivity of the AN/SPC 401 receiver,

(2) to estimate the shear stress of the Type IIA rivet, and

(3) to burn in 1N619 diodes.

Even in informal testing, such statements typically become entries in technicians' notebooks.

4.3.4 Testing Approaches

A significant portion of the remainder of this report is devoted to describing basic testing approaches. With familiarization of these, the specific approaches to be used follow from the objectives in that the types of response, type of measurement, and way of applying test conditions are (at least implicitly) already identified. For example, the first objective stated above is typical of a simple performance test; the second, a strength measurement by a stress-to-failure approach; and the third, a time-truncated aging test.

Testing approaches employed for flight accepting a satellite are limited to those which do not damage or significantly age the system. Obviously, a stress-to-failure test or a long duration test under severe aging-conditioning is not suitable in this case. Certain proof tests and some brief aging to reveal defects are typically very appropriate for flight acceptance.

Testing approaches are often predetermined by organizational policy of using standard procedures. References 4-1 through 4-3 are cited to simply illustrate the type of NASA standardization that has evolved for testing procedures; additional current NASA standards of this type can be identified in Ref. 4-4 (NASA SP-9000). References 4-5 through 4-7 similarly illustrate standardization in military development programs. The approaches in such standards are often adequate, but caution is always in order. For example, there are situations when a stress-to-failure approach is much more useful than a standard proof test approach. Special applications, new design techniques, new suppliers are all suspects for departing from standard approaches.

4.3.5 Test Design

A satellite once toppled several feet from its mount onto a concrete floor during a high temperature test of a flight acceptance test series because someone had failed to consider the effect of high temperature on the mount. This illustrates just one of many practical aspects of proper design for a test.

Test design is much more than the statistical design. It is that stage requiring utmost attention to detail, especially when the penalty for a "goof" in the test will be large. It is the stage when use of the checklist is most valuable. Some typical considerations are:

(1) availability and adequacy of test facilities,
(2) costs (including possible penalties),
(3) time (set-up, test duration, data reduction, etc.),
(4) skills,
(5) order and combination,
(6) analysis techniques,
(7) benefit of prior information,
(8) required accuracy and precision,
(9) desired form of final results,
(10) failure definition, and
(11) criteria for terminating tests.

Many of these considerations are resolved by the designer or production engineer himself. Others require special skills such as test specialists, the systems analysts, or statisticians.

The majority of tests are very informal and often consist of exploratory measurements just to learn more about an item such as a breadboard. These do not usually follow any carefully planned experimental design. On the other hand, there are many instances where statisticians can be invaluable in helping decide such issues as how many runs or trials to make, how many items should be in a sample, expected distribution of failures, method of analysis, etc. These tests are usually formal tests such as acceptance testing, sequential life testing, etc.

A discussion of statistical design of experiments is presented in Sec. A.11 of the Appendix and from this the engineer can obtain an appreciation of what role the statistical skills can play. It is again worth noting, however, that only a small portion of the testing effort is directly influenced by sophisticated statistics. Whenever statistical specialists are used, there needs to be good communication between the engineer and the statistician. It is up to the engineer to control the test planning and supply real world practicality.

In planning for tests, appropriate parameters have to be selected which truly characterize the effect of interest. Although the interest may be in determining some estimate of a quantity not directly measurable, the test itself concerns only parameters that are directly measurable. For example, direct measurements of basic physical quantities such as voltage, temperature, force, and time may eventually get translated to quantities such as strength, life, safety margin, and failure probability.

Measurements are concerned with both the test conditions themselves and with responses to these conditions. Too often, test procedures fail to include the appropriate measurements of test condition parameters for later correlation with measurements. Two basic types of measurements are attributes (go/no-go) and variables (parameter values). Even when the major analysis is concerned with attributes data, there is potential benefit in recording the tested parameters first as variables data; the need for additional analyses may occur later. For example, there is a notable increase in measuring part parameter values in all formal tests instead of just checking for conformance to tolerance. This trend has enabled statistical distribution characteristics to be obtained, providing valuable information for use in analyzing for effects of parameter variations.

One of the most often confused aspects of testing concerns units of measurement; measurement units during a test are frequently different from those eventually needed. For example, the inverse of temperature may be the correct units rather than the temperature level itself.

Automated test facilities aid in measurement and control of test conditions, but they cannot handle all situations. Even with automated facilities, manual overrides during the test are sometimes necessary. On-line computer operation aids in handling larger volumes of data than is possible manually, but care must be taken in designing and executing the test to insure that the larger data volume enables more results or more useful results to be obtained.

For most measurements that have become routine, standard procedures have evolved. Reference 4-1 through 4-7 contain typical standard procedures. Such standards serve a valuable role in reducing the planning effort and providing common procedures. Even when standard procedures appear applicable, caution is still in order--do the standard procedures really provide the information that is needed? Special tests are often required which depart from routine procedures only in the sense of their being conducted on unfamiliar equipment, used to measure a new uncertain effect, or applied to conditions which are different from usual.

4.3.6 Test Conduction

Test conduction is the phase where previously laid plans are put into effect; test conditions are applied according to the design and measurements are made. But "the best laid plans of mice and men...". Has anyone stopped to consider, for example, what happens if the main power supply fails? Can the test be resumed from where it stopped without affecting results? Will the test item be damaged? Will the test results already achieved be invalidated?

Many unexpected incidents are possible, but prior considerations can circumvent the detrimental effects of some of these. Good control documentation and proper training of personnel can alleviate many problems.

4.3.7 Analysis of Data

The considerations to be extended to data analysis during planning are highly interrelated to those for the statistical design of a test. For example, a designer wanting to know the effect of ambient temperature variations on his equipment may design his test using a proof testing approach to merely check the response at worst-case conditions. The analysis in this case is simple, either the item is good or bad at these conditions. However, there may be some concern about accuracy and precision of the test conditions and whether the tested item is representative of similar items in a population. Alternatives to the above objective and approach are measuring variation in behavior as a function of temperature or determining the temperature level which causes a tolerance to be exceeded. Either introduces more complexity in the test design and data analysis.

The degree of formality of the test often has a large influence on the extent and complexity of the analysis to be performed. An acceptance test specified by the customer may demand that extensive testing and analysis be performed to demonstrate a certain statistical confidence. Whereas for in-house verification purposes the designer may settle for considerably less because he has benefit of engineering confidence that the customer does not have. The designer, for example, may be quite satisfied with only simple functional checks or a simple graphical analysis.

The required extent and complexity of the analysis depend upon many other factors but certainly one should refrain from "analysis simply for analysis sake". Compatibility with the intended use of results is the foremost guideline. Will the procedure really resolve the problem? Are statistical procedures cluttering up your knowledge of the physics underlying the behavior? Is variability really as important as it seems?

4.3.8 Use of Results

Prior consideration of the intended use of results should exert a major influence on test planning from problem definition through data analysis. Certainly the test objectives should be derived directly from this.

Several viewpoints affect how the intended use of results influence planning. A general but basic viewpoint is that of learning versus verification. Any test involves some of both but some tend more toward one than the other. Development testing is generally associated more with the learning function than verification even though the designer may often be more intent (and possibly erroneously so) on verifying his paper design than in learning how the thing really works. Acceptance testing on the other hand, tends more toward verification or getting yes-no answers even though many are the cases where important things have been discovered at that late date. Generally, test planning should take advantage of the opportunity to learn but within reasonable constraints. Could you benefit in later analyses, for example, by making variables measurements in incoming testing rather than attributes measurements?

The viewpoint of the statistician on use of results is typically that of estimation versus hypothesis testing and this influences the phases of test planning in which he may be involved. The engineer on the other hand, is primarily interested in getting correct solutions to real-world problems. His interpretation of the problem and translation to the statistician has often failed to bring about the results really needed. What use does the engineer have for a statement about statistical confidence and how often does he really understand what it means when he has it provided? The engineer is really interested in engineering confidence*. Such problems resulting from these different viewpoints are most often attributed to communication.

Much discussion could be presented on the influence of the intended use of results on test planning. Much of it, however, would require simply reiterating much of what is said in the above sections. The essential task in planning for a test is matching the approach to problem solution to the real problem at hand. There is no simple step-by-step procedure for doing this.

---

* Discussion of these and other concepts are presented in Sec. A.5 through A.7 of the Appendix.

References

4-1. Anon.: General Specification, Semiconductor Devices, Established Reliability. NASA MSFC-SPEC-438, 1964.

4-2. Anon.: Electronic Equipment, General Specification for NASA ARC-SPEC-302, 1964.

4-3. Anon.: Standard, Environmental Test Methods for Ground Support Equipment Installations at Cape Kennedy. NASA KSC-STD-164(D), 1964.

4-4. Anon.: NASA Specifications and Standards. NASA SP-9000,1967.

4-5. Anon.: Military Standard, Test Methods for Semiconductor Devices. MIL-STD-750A, 1964.

4-6. Anon.: Military Standard, Electromagnetic Interference Test Requirements and Test Methods. MIL-STD-826A, 1966.

4-7. Anon.: Military Standard, Environmental Test Methods. MIL-STD-810B, 1967.

5. Nonaging Test Types

When one does not have to account for aging in a test, the approaches to testing are generally simpler and less time consuming. The relative importance of aging and how it affects the testing approach and classification of tests was introduced in Secs. 2 and 3. The test types considered in this section are those which treat the item basically as if aging is not a factor of importance in the response.

Certain tests are clearly devoid of aging to the tested item. A structural proof test and an amplifier gain measurement are typically nonaging types.

In some cases items simply have to be aged in order to measure certain characteristics. A drift test of a gyro, for example, requires operating time and hence wear (equivalent aging) of the bearings. In such tests, however, the results of the test do not generally depend upon the aging sustained during the test.

It is also possible that certain responses known to depend almost altogether on aging may be treated by certain nonaging test approaches. A mechanical load applied to a metal specimen at high temperature will result in creep of the material, a cumulative damage phenomena. Thermal runaway of transistors is an analogous problem in electronics. Such effects are often treated by a nonaging test approach. This basically involves ignoring the cumulative damage process per se and being concerned with the end result and its dependency upon the stress. For example, a stress may be applied at some given level and the observation continued for several seconds to see if thermal runaway is initiated.

Even long duration cumulative damage effects may be investigated by the nonaging test types. Typically, one may measure a fatigue strength by stress cycling a material for a fixed number of cycles at several levels of stress.

If, on the other hand, the duration of the response or the amount of cumulative damage for a given duration is of more importance, the aging test type described in Sec. 7 would be employed.

Test methods discussed in this section are identified in Table 5-1. Descriptions are presented there to distinguish among them. These distinctions are based primarily on the relative importance of stress. Remember that stress, by definition, is something that causes (at least potentially so) a detrimental response.

The order of the test types in Table 5-1 is generally such that the effects of stress become progressively more important. The latter type, sensitivity testing, is a special approach for testing one-shot items which have only two possible levels of a response; a familiar example is the acceptance sample testing of flashbulbs.

It is not easy to assign a known, specific test to one of these categories nor is it important that one be able to do so. That is, a specific test such as a check of transistor current gain at low temperature may be a proof test to some but a performance

Table 5-1

Comparison of Nonaging Test Types

Test Types	Examples	Distinguishing Features
Simple Measurements	Visual inspection Measurement of part dimensions Electrical continuity checks Measurement of amplifier gain in a bench test	Items are <u>not</u> purposely stressed; emphasis is only on measuring inherent properties of items.
Performance Test	Measurement of amplifier drift due to temperature Rocket sled test of a stable platform	Items are purposely stressed; emphasis is on performance subject to the conditions.
Proof Test	Structural proof test Diode reverse voltage check	Stress is fixed at a level to intentionally cause failure of bad items and prove the capability of good items.
Stress-to-Failure (Test-to-Failure)	Structural strength test Failure mode test of a circuit at high supply voltage	Stress severity is always increased to a level causing failure for every item tested.
Sensitivity Testing	Measurement of flash bulb strength to thermal stress Acceptance sampling of explosive bolts	Special approach for testing items which can have only one of two possible levels of response; used primarily for one-shot items.

test to others. The test types identified are basic test approaches and we consider it more important to first understand the basic concepts and let the reader apply them where applicable.

The emphasis in this section is then only on describing the basic test types and pointing out unique, practical problems associated with each. Basic uses of these methods are presented separately in Sec. 6 because there is not always a one-to-one correspondence between methods and uses.

5.1 Simple Measurements

This applies to the many measurements such as dimensions, weight, and inherent functional characteristics measurements where the conditions under which the measurements are made either represent usual ambient or are simply irrelevant. Even a non-stressing stimulus may be required as when measuring the nominal transient response of an amplifier, an inherent property. In a simple measurement, items are not purposely stimulated with stress to create a response and thus one is not concerned with measuring the behavior in response to stress. A standard measurement of resistance with an ohmmeter is a simple measurement; measurement of resistance as a function of current through the device or ambient temperature generally would not be a simple measurement.

A simple measurement is thus essentially a test without a response and even though it may not be considered a test in the strict sense by some readers, it is included as a type of testing because a large portion of test expenditures are devoted to it.

The following examples illustrate some typical simple measurements.

<u>Example 5-1</u>

> Instructions for incoming tests specify a visual inspection for existence of corrosion and burrs on the electrical contacts of all relays. Relays observed to have corrosion or burrs on the contacts are rejected from application.

This is a simple attributes measurement based on the quality judgment; items are either acceptable or unacceptable. A simple attributes measurement based on a quantitative measure is illustrated by the next example.

<u>Example 5-2</u>

> The resistance of resistors is checked during incoming inspection for conformance to $\pm$ 1% variation from the nominal value. Even though the resolution of the meter is 0.01%, each measurement results in either an in-tolerance or out-of-tolerance determination of each resistor.

Many simple measurements are variables measurements. The next example is typical of these.

Example 5-3

> In the fabrication of a stable platform a particular part is weighed prior to installation to determine how much compensating mass is required for balance of the gimbal. Any part mass is acceptable, but it must be recorded to the nearest 0.01 gram of the instrument scale.

The concept also applies to time-varying behavior of items so long as it truly represents an inherent property of the item as is illustrated by the next example.

Example 5-4

> A gyro is operated in a fixed orientation under laboratory conditions to measure its drift. The results of the test are illustrated in Fig. 5-1 as a continuous recording of its output signal.

The property of real interest for the gyro may only be whether the cumulative drift at time T is within tolerance or it may be something more complex such as the frequency spectrum of the recorded process. Either is still a simple measurement.

Simple measurements are applicable to all types of items both simple and complex. Measurement of characteristics of one-shot items, e.g., the measurement of light intensity of flashbulbs, can be a simple measurement if made under usual ambient conditions.

Simple measurements are prevalent throughout any hardware program. Typical reliability-related uses are:

(1) distinguishing between good and bad items,

(2) measuring performance-related characteristics, and

(3) as the measurement technique for other test types.

These are discussed in Sec. 6.

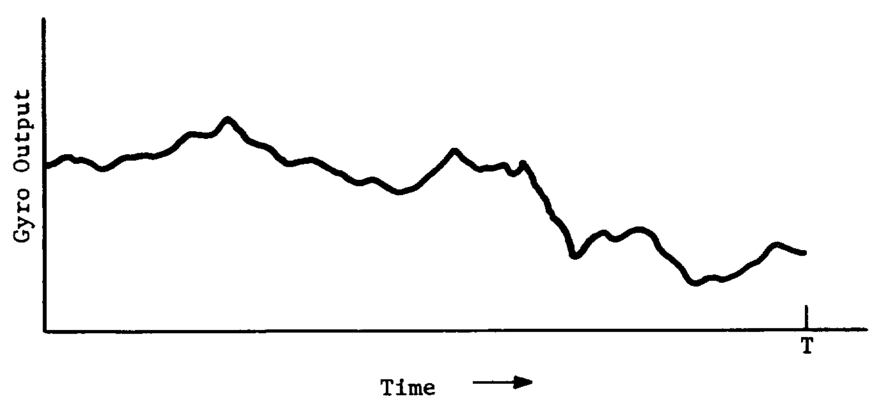


Figure 5-1. Simple Measurement of Gyro Drift

Special Considerations

A survey of practical factors related to testing was presented in Table 4-1. Even though a simple measurement is the simplest form of a test, many of the factors still apply. Some typical considerations are discussed below.

Responses. By definition simple measurements are not concerned with responses. In most cases, the item is simply devoid of a response or any response present is irrelevant. It is possible in some measurements, however, that the parameters are unknowingly responding to something. In a gyro drift test for example, the observed drift may actually be strongly dependent upon test stand vibrations transmitted via the floor and ground from a nearby highway. Even significant aging may be unavoidably or unknowingly present in the response. The required gyro operation for the drift test is certainly providing some wear of the bearings; is it significant? The control or elimination of such unwanted responses can sometimes be achieved by control of the conditions which cause it.

Conditions. It would be of little benefit to specify 30 times magnification in a visual inspection for material defects if the microscope is located on a shaky bench. It is occasionally necessary to expend considerable effort to isolate items from certain conditions. Clean rooms, vibration isolation, and thermal and humidity control are typical of such effort to eliminate unintentional effects.

Even though the conditions for simple measurements are generally less complex than those for other test types, there can be many practical considerations. Practical queries should be raised and resolved conscientiously. Typical of these are:

(1) How close the conditions should be maintained to those of normal use?

(2) Are the inputs such as supply voltage and signal inputs at the proper level?

(3) How precisely do the conditions need to be controlled?

(4) Do I really need the laboratory environment?

Measurements. Even though the emphasis is on measurements, there is really no uniqueness applicable here in contrast to other test types. A typical major consideration is whether to record information as attributes data or variables data. Most often, this is dictated by the immediate use to be made of the results as exemplified by Examples 5-2 and 5-3. In contrast Example 5-1 exemplifies the situation where the nature of the available information restricts the choice to attributes data only.

Occasionally, it proves to be advantageous to record measurements as variables data (when available) even though the immediate use of it demands only attributes data.

Example 5-4

During customer acceptance tests of the equipment employing the resistors treated in Example 5-2, it is discovered that an analytical assessment of the effect of parameter variation of circuit components is necessary. Subsequent sensitivity analysis shows that the resistance variation of these particular resistors is crucial and realistic estimates of the absolute variations are needed. But, alas! The values were not recorded during the tests.

It is, of course, impractical in all cases to simply record variables data and convert to attributes data, but there is room for intelligent foresight and discretion in this matter.

Other practical considerations regarding the measurements can be raised by queries such as:

(1) Should certain measurements be made simultaneously?

(2) Can special nondestructive testing techniques be beneficial?

(3) How important are precision and accuracy in the measurements?

Simple measurements are the simplest of all test types; the practical considerations still possess similarity in complexity to other types, however.

5.2 Performance Testing

Given some limit which defines the range of conditions within which an item is considered capable of performing (possibly even in a degraded fashion), performance testing is generally concerned with how well (or how poorly) it performs within this range or perhaps what characteristic the item should have in order to be able to successfully perform within this region. A measurement of amplifier gain as a function of temperature, load or supply voltage and a measurement of structural deflection dependency on load or vibration are typical examples of performance tests. Remember, however, that only nonaging types are being treated in this section.

It is not always easy, nor necessarily important, to clearly distinguish between a performance test and a simple measurement. We have chosen to call a test such as a standard amplifier gain measurement under laboratory or usual ambient conditions a simple measurement. If, however, under the same conditions, one is concerned with the output waveform characteristics resulting from a complicated input signal, then we assign it to the performance test category. Why? Because mainly the viewpoint is on behavior (of the output) in response to some test condition. But there is no need to quibble over such decisions because any practical principles needed for such borderline cases could be drawn from either type. Similar borderline situations occur also between performance testing and other types of nonaging test types.

In a performance test some stimulus, either stress or a nonstressing condition is applied and the subsequent behavior observed. These conditions may assume any of the characteristics described in Sec. 2.2 including fixed or time-varying characteristics and have stressing or nonstressing effects. Behavior may be discrete, intermittent, or continuous over time.

Most responses giving rise to the behavior encountered in performance testing are reversible; however, numerous examples of nonreversible (even exclusive of unintentional aging) responses exist. Measurements may be either attribute or variable types and may emphasize either viewpoint of determining performance as a function of conditions or determining conditions which cause given performance. Several examples are presented below to illustrate many of these aspects.

<u>Example 5-5</u>

> One step of a qualification test procedure for a servo amplifier specifies that the output null voltage at 60°C shall be less than 0.1v. The amplifier is operated in an oven at 60°C until is stabilizes and then checked to determine if the output null voltage is within tolerance.

Due to the attributes measurement, this is merely a performance check. Performance checks may be considered just one form of a proof test and is also recognized in Sec. 5.3. This, then, is just another borderline case for classification and may be viewed as either.

Variables measurements of the following type are very prevalent in performance testing.

<u>Example 5-6</u>

> A nominally 100KΩ resistor is known to have that value of resistance at 25°C. It is tested at 100°C to determine the change in resistance. The result is an increase of 2.3KΩ.

This test measures a performance parameter for a particular stress condition. As illustrated below it is easily extended to measure performance as a function of the stress.

<u>Example 5-7</u>

> The resistor of Example 5-6 is tested at different levels of ambient temperature between -50°C and 150°C. The result is illustrated in Fig. 5-2.

The concept is equally applicable to all types of items to measure any parameter behavior as a function of test conditions. The next example illustrates the generality.

<u>Example 5-8</u>

> In the design of a sounding rocket the aerodynamic drag as a function of Mach number and angle of attack are needed as design information. Wind tunnel tests are conducted and the reduced results for a particular configuration may appear as illustrated in Fig. 5-3. Drag is the performance parameter and Mach number and angle of attack are the conditions.

The concept also applies to consideration of performance as a function of time.

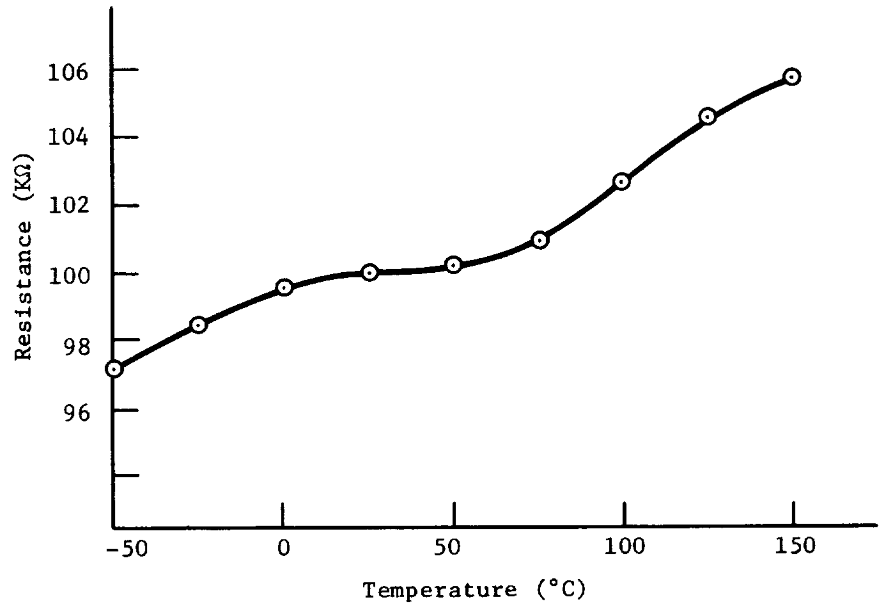


Figure 5-2. Performance Test of Resistance vs Temperature

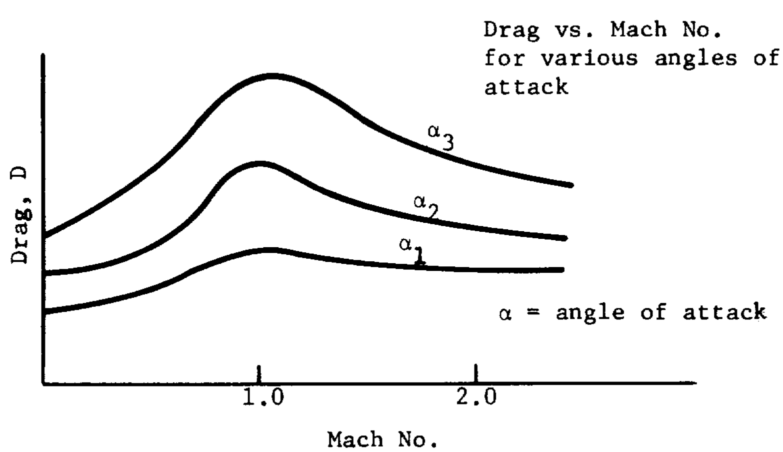


Figure 5-3. Illustration of Multiple Performance Measurements

Example 5-9

The drift of a stable platform is measured while the platform is subjected to vibration which is programmed to simulate a launch profile. The test conditions and results are illustrated in Fig. 5-4.

Even though time is an important parameter, the emphasis is not on an aging response.

There is an inference in each example above that the response is reversible. Most performance tests are, in fact, concerned primarily with reversible responses, but they can also be concerned with nonreversible responses as illustrated by the next example.

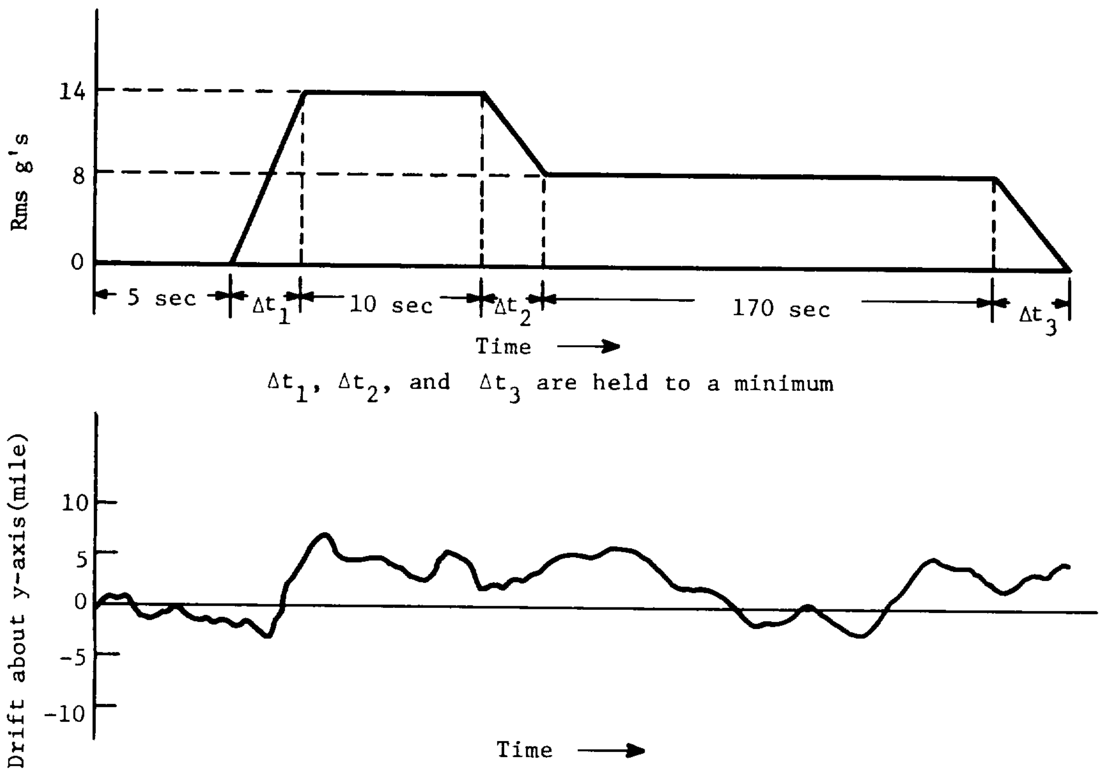


Figure 5-4. Measurement of Stable Platform Drift Performance over Time

Example 5-10

The attenuating effect of a pliable material being considered for use as an impact absorber is measured in a drop test of a load both with and without the absorber material attached. The results are illustrated in Fig. 5-5 as measured deceleration of the load. The attenuation and damping of the shock by the material is accomplished by allowing permanent deformation, a nonreversible process.

Because the deformation is permanent, a repeated test with the same specimen would yield different results. An absorber employing the material would typically be designed as a one-shot item. It is possible, however, that certain items having such nonreversible responses could be designed for several repeated uses. This seemingly introduces some type of aging and repeated tests for performance might be more properly considered an aging type.

Another viewpoint of repeated tests may be that of measuring degraded performance. This concept, however, is broader than implied by the repeated shock absorber tests; for example, an item may become degraded in one type of performance measurement such as behavior due to vibration and still another performance measurement made, say, to shock.

Behavior at overstress (but nonaging) conditions may even be of emphasis in a performance test. The absolute deflection of a structural member beyond its yield point, for example, may be a very important consideration.

All of the above discussion has emphasized the direct response problem viewpoint of "Given the conditions, what is the performance?" The inverse response problem viewpoint, viz., "Given the performance, what are the conditions?" often arises in performance testing. A very simple case is illustrated.

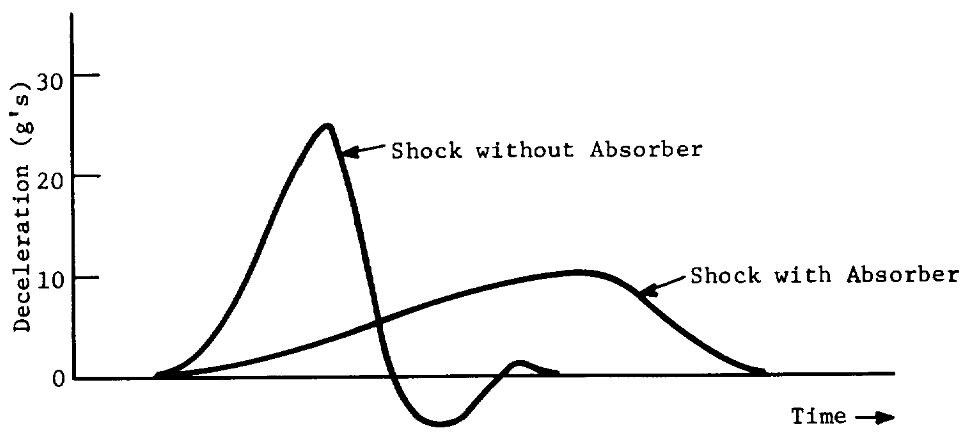


Figure 5-5. Measurement of Shock Absorber Performance

Example 5-11

A relay is tested to determine its threshold voltage (i.e., the voltage required to cause it to throw). The voltage to the coil is gradually increased until it throws and that value recorded as the threshold voltage.

A related problem is performance optimization, i.e., determining the combination of conditions which optimize the performance. As discussed in Sec. 2.3.2 this involves both viewpoints.

Performance testing is also found throughout hardware programs from feasibility testing to end-of-service evaluation. The basic reliability-related uses considered in Sec. 6 are:

(1) distinguishing between good and bad items,

(2) measuring performance-related characteristics, and

(3) as a measurement technique for other tests.

5.3 Proof Testing

The concept of proof testing is simple. A stress condition is applied and the item either fails or does not fail. The basic purpose is to determine whether the tested item can withstand predetermined stress conditions without failure. Items considered good pass the test; those that are bad fail the test. Items that pass a proof test are thus known to have a strength greater than the severity of the applied stress; items that fail, a strength less than the severity level of the applied stress. It is important to remember, however, that proof testing is basically a non-aging type.

The designation, proof testing, has most often been associated with a test in which the item is either damaged or not damaged as a result of applied mechanical stress. The concept, however, applies equally to any test designed to "prove" the capability of any type of item (including, for example, electronic parts) to any type of stress and does not necessarily have to be concerned with damage as a criteria for failing the test. Note, however, that one does not have a proof test unless a stress is present.

Example 5-12

In a qualification test of a particular transistor type the dc current gain is measured at -25°C with other conditions nominal to determine if the transistor gain meets the lower tolerance limit of 30. The measured gain is 22 and the transistor thus fails the test.

This proof test is really nothing more than a performance check because typically in such transistor tests the item is not damaged by the conditions. The test shows, however, that the capability of the item to function properly at the low temperature conditions cannot be proven.

Since damage was not induced in the tested item in Example 5-12, the response is a reversible one. Frequently, the criteria for failure is based strictly upon whether damage is induced.

Example 5-13

> A quality assurance provision for an inverter specifies a vacuum test (with the inverter inoperative) at a pressure of 0.2 mm of Hg. It is further required that the performance shall be checked following the test and that the performance shall not be impaired by the test.

Performance impairment in the above example, and hence failure to pass the test, would be indicative of damage. A similar example for proof testing to check for structural integrity is presented below.

Example 5-14

> A pressure vessel having a rated working stress of 1,000 lbs. gauge pressure is proof tested by gradually increasing the gauge pressure to 1,500 lbs. and monitoring for leaks and deformation. If no leaks or deformation are detected, then the pressure vessel or its design is considered satisfactory for normal use at 1,000 lbs. gauge pressure.

Such tests are typical of those for testing commercial gas bottles before refills.

For structural items the level of the stress used in the test is most often selected not to exceed the elastic limit of the structural material. For a good item the response for any item passing the test is definitely a reversible one and is nonreversible for any item failing. In contrast, a proof test of pliable materials could conceivably involve nonreversible responses. The concept of proof testing may also be extended to one-shot items. A flashbulb could be tested to prove the capability at a particular temperature.

The above examples are adequate to illustrate the concepts of proof testing. An excellent example of a proof test design is presented in Ref. 5-1. Note that the criteria for failure is quite important. The specific criterion depends generally upon whether emphasis is upon performance or upon integrity but may range from a simple out-of-tolerance condition to total destruction.

The basic reliability-related use of proof testing is for distinguishing between good and bad items. With this ability it is especially suitable as a screening procedure for those types of items which can be tested on a go, no-go basis. Further discussion of these applications is presented in Sec. 6.

Special Considerations

A basic problem in proof testing pertains to the selection of the maximum severity of the stress to be applied in the test. Should it be the maximum working level, 150% of it, 200% of it? Example 5-14 specified 150%. Certainly if there is a known danger of damaging all items at 150% maximum stress, then testing at this level would defeat the purpose of it.

Typically in qualification testing of purchased components the stress levels are set at the manufacturer's rated value. If the results are acceptable at this level, then the component is considered qualified for those applications in which the stress severity is lower than this (but often with derating employed to increase the safety margin). This implies that the manufacturer knows the appropriate strength of his components and that his advertised rating is less than this.

The problem with an in-house design and fabricated item is basically the same. The strength must first be known with some acceptable (engineering) confidence, either by measurement or analytical prediction, and a rated stress specified which allows the desired safety margin.* The maximum allowable severity of stress is then required to be less than or equal to this value depending upon the amount of derating desired. At any rate, the rated stress becomes the level for proof testing. Items which fail the test are thus known to have a strength less than that allowed.

The above presents the rudiments of the problem resolution; however, it may be approached in other ways. In any approach trade-offs among the factors, i.e., the design strength, the safety margin, and the maximum severity of working stress will be required. The problem is also further complicated when multiple stress factors are involved.

5.4 Stress-to-Failure** (Test-to-Failure) Testing

In a stress-to-failure test approach the severity of the stress is increased, continuously or in steps, up to the level where failure is induced. This approach is applicable only to those items which do not have to be destroyed at each level of stress in order to observe its response. This characteristic clearly distinguishes it from sensitivity testing described in Sec. 5.5.

Note that because we are talking about nonaging responses the approach is not to be confused with step-stress or progressive-stress testing approaches applicable only to accelerated testing as described in Sec. 11. Also note that even though time may be important for reasons such as delayed responses and rate of increasing stress severity being critical, there is no implication of a stress-until-failure nature of the test; thus, it is not to be confused with failure-truncated testing introduced in Sec. 7.

---

* This is only one of several ways of viewing safety margins; more detailed discussion is presented in Sec. 5.4.

** The term "stress-to-failure" is preferred to the often used term "test-to-failure" because it is more descriptive of the true nature of the testing approach.

The following example illustrates a familiar type of stress-to-failure test.

Example 5-15

A pressure vessel is tested to destruction by gradually increasing the internal pressure with all other conditions fixed while a critical dimension y is monitored. The response during the test is illustrated in Fig. 5-6 as Δy, a change in the dimension y, versus the stress of gauge pressure.

If the definition of failure for this test is rupture of the vessel then the test is literally a stress-to-destruction test. Even though the concept is most popularly associated with stressing mechanical items to induce damage it is also applicable to other types of hardware and other criteria of failure. Application to an electronic equipment is illustrated by the next example.

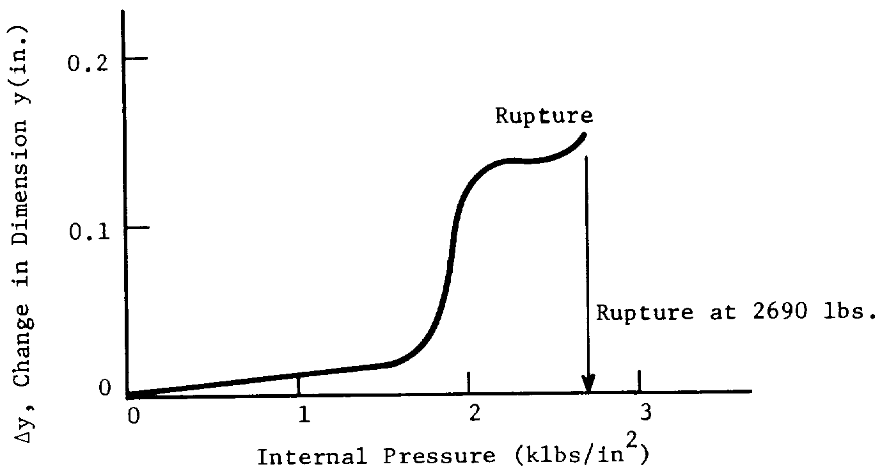


Figure 5-6. General Illustration of a Possible Stress-to-Failure Response of a Pressure Vessl to Gradually Increasing Internal Pressure

Example 5-16

A dc to dc converter is tested by increasing the load current in increments to the level where failure occurs. Failure is defined by the output voltage variation exceeding one volt. The response is illustrated in Fig. 5-7. With the incremental changes in stress levels for the above test the stress level at which failure occurs is not as precisely defined.

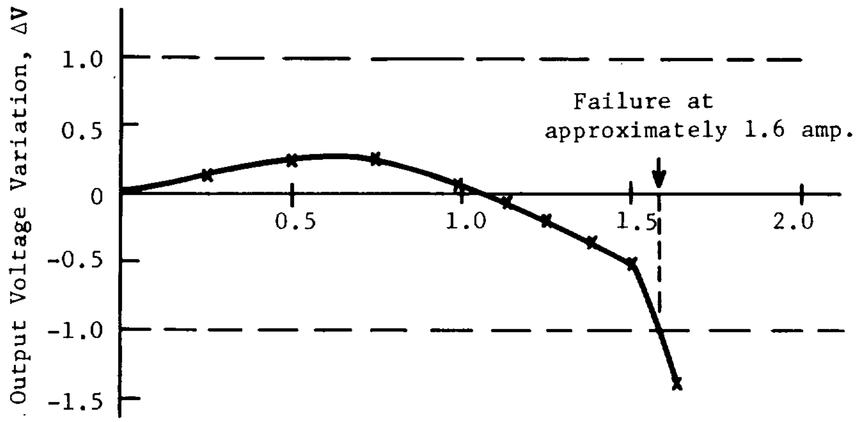


Figure 5-7. Converter Response During Strength Test

Note the difference in the failure definitions of the two examples above. As for proof testing the specific criterion depends generally upon whether emphasis is on performance capability or integrity and failure definitions can range from simple nondestructive criteria such as an out-of-tolerance to some predetermined level of damage such as total destruction or impairment of performance.

The stress-to-failure approach finds application only when there is a need to intentionally fail an item. The prime uses described in Sec. 6 are thus for measuring strength and for investigating modes and mechanisms of failure. Reference 5-2 gives a good description of an application of the approach to measuring strength distribution characteristics of electronic equipment.

5.5 Sensitivity Testing

The specific severity level of a stress, say pressure, which would cause a particular explosive bolt to fail in operation cannot be determined experimentally by a stress-to-failure approach because a test at any level requires destructive operation of the item. Sensitivity testing is an approach especially suitable for investigating sensitivity of items requiring such destructive, one-shot operation to test conditions. It has been employed for many years in connection with dosage mortality and response to bio-assay work.

In a sensitivity test an item representing a sample from a population of interest is subjected to a stimulus and either of two possible levels of a response occurs, e.g., an explosive bolt fires or does not fire. The response level is assumed to depend upon whether the stimulus exceeds some critical physical threshold for the particular sample. Furthermore, it is assumed that a distribution of these thresholds exists for the population and the aim of sensitivity testing a number of items is to describe one or several characteristics of this distribution.

The unique distinguishing features of sensitivity testing thus are:

(1) only one response is available for each level of stimulae and for each item tested, and

(2) a response can have only one of two possible levels (i.e., the measurement is an attributes measurement).

Note first that a sensitivity test of a single item is nothing more than a proof test. It is the use of these in combinations at different stimulus levels that demand the special attention here.

A general illustration of the sensitivity testing approach is illustrated by the following example.

Example 5-17

Ten explosive bolts are tested for satisfactory operation at each of five levels of pressure. The results for the total sample of fifty are presented in Fig. 5-8 as the relative number of successes at each pressure level.

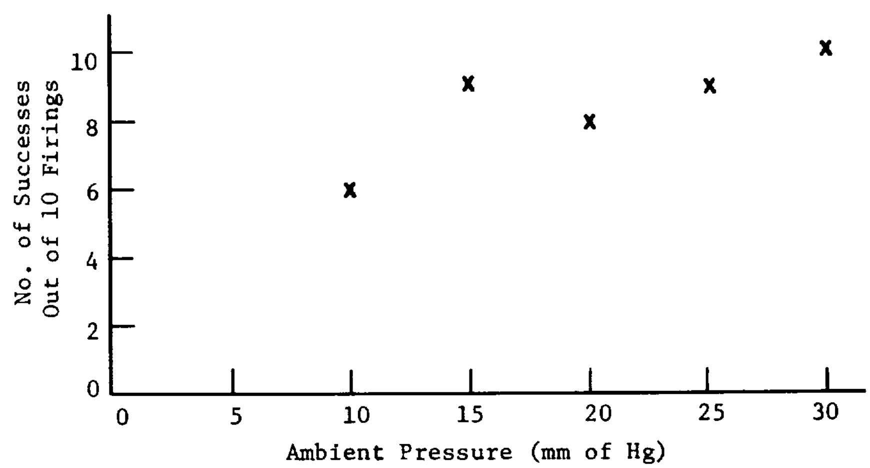


Figure 5-8. Illustration of a General Form of Sensitivity Testing

With such an approach it is easy to see how the response dependency on stress severity can be determined.

There has been considerable emphasis in the past on design of sensitivity tests for greater efficiency in terms of obtaining more information of the type needed per test sample. Many designs have evolved.

The above example illustrates a nonsequential approach in which tests of individual specimens can be performed without regard to order of stimulus levels. In a strict nonsequential approach k stimulus levels are specified prior to the test with $n_i (i = 1, 2, \ldots, k)$ items allocated for each i-th level. The probit design is a non-sequential design and was one of the earliest of all sensitivity testing designs.

Even though nonsequential approaches are still frequently used, sequential approaches have proven to be more efficient for most applications. The earliest of these was the Bruceton or Up and Down Method. In a sequential approach an item is tested at a particular level, and based on the response the next item is tested at some higher or lower level. The testing proceeds on this basis with the stimulus level for each dependent upon the results of prior tests, some for only the previous one and others for several.

Table 5-2 identifies a number of available designs for sensitivity testing. The uniqueness of each depends mainly upon the inherent assumptions about the underlying statistical distribution and the purpose of the test.

Some require no assumption about the distribution, and hence the results are distribution-free while others assume a specific type of cumulative distribution for the response function, for example, normal (or Gaussian), log-normal, or uniform distribution.

As noted earlier sensitivity testing is generally devoted to describing one or more characteristics of the distribution of threshold levels which causes the response. Actually, both the direct and indirect response problem viewpoints are applicable in sensitivity testing. Example 5-17 above and Example 5-18 later illustrate its use for the direct response problem of seeking the response (in this case the statistical frequency) dependency on stimulus level. A simple case is represented by specifying a stimulus level and estimating the fraction responding. For the increase response problem one may be interested in finding the median stimulus level at which 50% of the items would be expected to respond. For a symmetrical distribution this would be equivalent to estimating the mean and could be used in determining the safety margin for the particular item being evaluated.

Note that we do not refer to designs themselves as test types. Sensitivity testing is the basic test type; the designations refer to different designs of sensitivity testing in much the same way that there are different designs for a series of performance tests.

Table 5-2

Summary of Sensitivity Testing Design

Design	Purpose	Assumptions	Remarks
Nonsequential: Probit Design	To estimate the entire stress response function.	Gaussian or logistic distribution most often used.	Inefficient compared to sequential design.
Sequential: Up and Down	To estimate $X_{0.5}$[a] (median stress level).	Most often used with Gaussian response distribution.	Design is good only for $\alpha = 0.5$.
Up and Down for small samples	To estimate $X_{0.5}$	Gaussian response distribution.	Analysis procedures are preferred to those for large samples.
20 ZIL	To estimate "Zero threshold" stress level.	Distribution free.	Conventionally 3 sequences are run.
Run Down	To estimate X_α for $10\% \leq \alpha \leq 25\%$ or for $75\% \leq \alpha \leq 90\%$.	Gaussian response function.	200 items required for test.
Robbins-Munro	To estimate X_α for any α.	With or without distribution assumption.	Slight variations of this design are preferred, e.g., the one following.
Delayed Process (variation of the Robbins-Munro)	To estimate $X_{0.50}$	With or without distribution assumption.	Preferred to the basic Robbins-Munro when the guess for initial stimulus level is not close to the median.

[a] The designation X_α represents the stimulus level X below which the proportion α of the responses occur, e.g. $X_{0.5}$ represents the median of the response distribution.

Table 5-2 (Continued)

Design	Purpose	Assumptions	Remarks
Alexander	To estimate X_α for any α. To give an interval I such that $P(X_\alpha \varepsilon I) > P$ where P is some prescribed probability.	Distribution free.	More efficient than some other distribution free designs.
Naval Powder Factory (NPF & NPF Inverted Design)	To estimate "threshold" stress level (10 to 30% level).	Distribution free.	Similar to 20 ZIL Design.
Langlie	To estimate μ and σ of the strength distribution.	Essentially distribution free but often used by assuming a Gaussian response distribution.	Less efficient than other designs.
Rothman	To estimate X_α for any α.	Distribution free.	About as good (asymptotically) as any distribution free design.
Derman	To estimate X_α for any α.	Distribution free.	Less efficient than other distribution free procedures. Same as Bruceton design for $\alpha = 0.5$.

Since it is not possible to discuss all of the designs in depth, two are chosen for discussion below to illustrate more of the concepts. Reference 5-3 is an excellent source for more in-depth treatment of other designs. It also identifies a number of good background references. Reference 5-2 gives a good description of using the Langlie method in a systems test program.

Probit* Design - A Non-sequential Type

Probit design is one of the earliest formal approaches specified for sensitivity testing. In the design there are k levels of the stimulus with $n_i (i = 1, \ldots, k)$ items tested at each level. The choice of k and n depends upon such factors as application, the number of items available for testing and inherent assumptions. Example 5-17 is an illustration of a simple form in which n = 10 for each level which are equally spaced. The presentation of results in Fig. 5-8, however, does not conform strictly to a probit analysis.

Probit analysis was developed for use in connection with the probit design and is based on the assumption that the critical threshold levels for a population are cumulative normal. The approach is to transform the fractions of items having the same response at each stimulus level to probits and fit a line of probit value versus stimulus level. Probit is defined as the standard normal deviate corresponding to the fraction plus five. For r_i acceptable responses of n_i tests at level i the standard normal deviate is easily obtained from tables of normal error as the deviate corresponding to cumulative probability r_i/n_i. The addition of 5 is a formal procedure merely to eliminate negative numbers in the analysis. Tables giving probit values directly for fractions are readily available, e.g., Ref. 5-5.

Example 5-18

Ten flashbulbs are tested with nominal voltage and current but at different temperature levels. The number of acceptable responses at each level is tabulated below and the computation of probits is shown in accompanying columns. A plot of the computed probits versus logarithms of the stress level is presented in Fig. 5-9 with a straight line fitted to the points.

---

* The term, probit, is a contraction of "probability unit" and was introduced by Dr. C. I. Bliss [Ref. 5-4].

Example 5-18 (Continued)

Level No.	Temperature (°F)	No. of Unacceptable Responses	Fraction Responding	Standard Normal Deviate	Add	Probit
1	110	1	0.10	-1.28	5	3.72
2	120	2	0.20	-0.84	5	4.16
3	130	3	0.30	-0.52	5	4.48
4	140	4	0.40	-0.25	5	4.75
5	150	8	0.80	-0.84	5	5.84

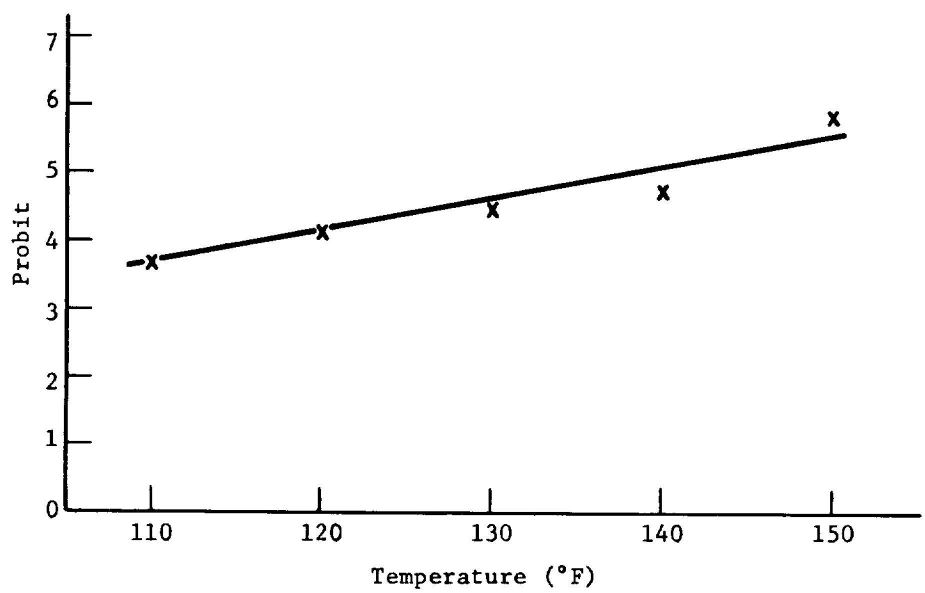


Figure 5-9. Sample Probit Graph

Straight line fits are obtained by methods described in Sec. A.8 of the Appendix. The transformation has thus provided the frequency of response as a linear function of the stimulus variable. Occasionally, experience or underlying knowledge may suggest a preliminary transformation such as logarithm of the stimulus parameter.

What advantages does the plot of Fig. 5-9 represent over the type in Fig. 5-8 which did not employ probit analysis? The only advantage is that the frequency of response is transformed to a convenient linear function of the stimulus. But why not simply fit a straight line through the points in Fig. 5-8? This is perfectly permissible and is, in fact, often relied upon without going through the probit procedure.

Remember, however, that probit analysis is based upon an underlying assumption that the response is cumulative normal. When this assumption is valid probit analysis is helpful in the mechanics of determining the dependency of fraction response on the stimulus level. This assumption sometimes introduces difficulties which are exemplified by the next example.

Example 5-19

The results of the tests described in Example 5-17 are to be analyzed by a probit analysis. An attempt to apply the same procedure as employed in Example 5-18 (especially in using tables of normal error) results in the following.

Level No.	Pressure (mm of Hg)	No. of Unacceptable Responses	Fraction Responding	Standard Normal Deviate	Add	Probit
1	30	0	0	$-\infty$	5	$-\infty$
2	25	1	0.10	-1.28	5	3.72
3	20	2	0.20	-0.84	5	4.16
4	15	1	0.10	-1.28	5	3.72
5	10	4	0.40	-0.25	5	4.75

A plot of the probits is presented in Figure 5-10. The first problem recognized is that of the probit for the zero (or unity) fraction response. As indicated by the probit at the first level this gives rise to an infinite probit by the procedures followed. Accepted practice is to rely on minimum (or maximum) working probits for purposes of fitting the line. This problem is discussed in more detail in Refs. 5-1 and 5-2.

A second problem is noted from the fact that the probit for level 3 is greater than that for level 4. There is no guarantee that the observed response function

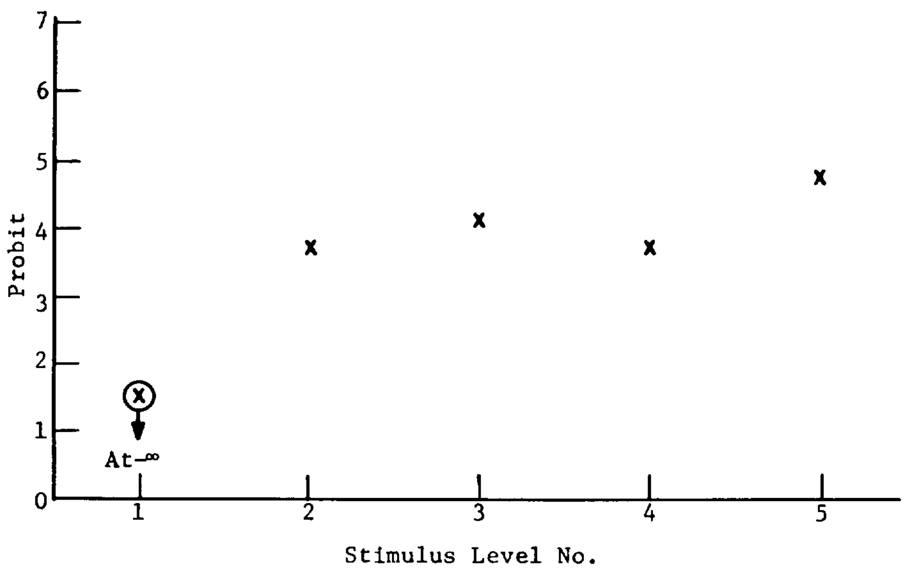


Figure 5-10. Probit Plot Illustrating Ambiguities

(i.e., the fraction responding at given levels versus the level) for certain choices of stimulus variables is monotonic. For example, in Example 5-19 one might normally think that the fraction failing would increase with decreasing pressure (increasing stimulus level); however, it may, in fact, be true that level 4 represents a less severe condition than level 3. Another way of stating this is that some of the items failing at level 3 may not have failed had they been tested at level 4. As discussed in Sec. 2.2.2, one must be careful how he defines the severity of the stimulus or stress and transformation of the stimulus variable may be in order.

Even with the proper choice of stimulus variable, it is to be remembered that one is still not observing a true cumulative response. That is, the observed fraction responding at any one level has a binomial distribution and for small sample sizes the random variation may yield nonmonotonic fluctuations in the response function.

Transformations of the frequencies of response to other than probits are sometimes helpful to permit fitting curves easily when probits do not work. Reference 5-2 cites and defines logit, log log, and angular transformations as other possible types. Tables for values of these are available in Ref. 5.5.

The efficiency of probit design may be improved to some extent by specifying different sample sizes for each level depending upon the relative weighting desired at each level. Attempts to improve the probit design, however, usually fall short of the efficiency achieved by sequential designs.

Up and Down Design - A Sequential Design

There are several possible variations of this design. The original, the Up and Down (Bruceton) design, was first implemented at the Explosives Research Laboratory at Bruceton, Pennsylvania. The design is aimed primarily at the purpose of estimating the median, $x_{0.5}$, of the distribution of the stimulus levels which cause the response of interest. The analysis is based on certain distributional assumptions. The design of the original approach consists of the following rules.

(1) A fixed step-size, d, for the stimulus levels is chosen and is selected to be as near the standard deviation of the distribution as prior knowledge permits.

(2) The first test is performed at a stimulus level as near as possible to the median.

(3) The n-th test is performed at a stimulus level

$$L_n = \begin{cases} L_{n-1} + d; \; y_{n-1} = 0 \\ L_{n-1} - d; \; y_{n-1} = 1 \end{cases}$$

where $y_{n-1} = 0$ (or 1) denotes an acceptable (or unacceptable) response.

(4) Testing continues until some specified criteria is met. This may be based on the available sample size or some criterion such as continuing until a certain number of changes of response have occurred.

An illustration of this procedure and the results is presented in Fig. 5-11. Various analysis procedures are suitable for estimating the median and variance of the stimulus threshold distribution from the test results. A method was developed by Dixon and Mood specifically for the Up and Down design results when the frequency of response can be assumed to be cumulative normal. This is described in Ref. 5-5.

Of the modifications in the Up and Down design the most useful is the Up and Down Small Sample Method, recently reported by Dixon in Ref. 5-6. The rules for conducting the tests are the same as those for the original approach except for the last step. The modification consists of continuing the testing until the total number of tests is N' where $N' = N + N_L - 1$ with N representing a prespecified nominal sample size and N_L, the number of like responses at the beginning of the series. For example,

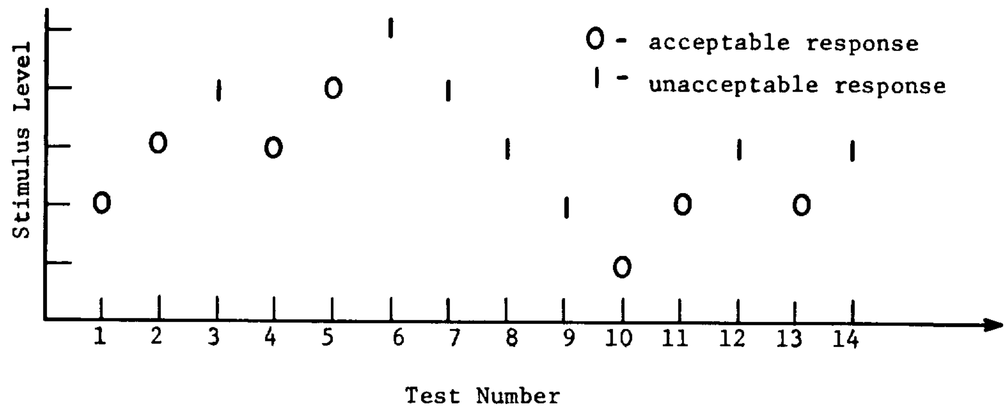


Figure 5-11. Illustration Up and Down Test Design and Results

in the series shown in Fig. 5-11 and for, say, N = 4 only 5 (i.e., N') tests are required resulting in the sequence of results of 00100.

Note that N is specified prior to the test whereas N' is known after the first change of response from the initial type. N is dependent upon the confidence level required for the estimate of the median.

For the analysis the estimate of the median $\hat{x}_{0.5}$ is computed simply by

$$\hat{x}_{0.5} = x_f + kd$$

where x_f is the stimulus level used in the last test of series, d, is the spacing of the stimulus levels, and k is a quantity dependent upon the sequence of the two types of responses. Tabulations of k for all response sequence combinations for $N \leq 6$ are given in Refs. 5-5 and 5-6.

Example 5-20

Suppose the flashbulbs in Example 5-18 were tested by the Up and Down Small Sample Method. Let N = 6 be the desired nominal sample size and let the first test be conducted at 120°F. Further, let 0 represent an acceptable response and 1 an unacceptable response. A possible sequence of responses and the corresponding stress levels dictated by the rules of the design are:

Responses	Temperature Levels (°F)
0011010	120, 130, 140, 130, 120, 130, 120, 130.

For this response sequence, a value of k = 0.952 is obtained from the tabulation in Ref. 5-6 and the estimate of the median threshold temperature is

$$\hat{x}_{0.5} = 130 + 0.952(10) = 139.5°F.$$

It is further stated in Ref. 5-6 that the estimates have error approximately independent of the chosen starting level of the tests and spacing between stimulus levels. The spacing should still be about equivalent to the standard deviation of the response thresholds but any choice between 2/3 and 3/2 of it will provide good estimates.

References

5-1. Anon: Structural Proof Test Programs. Report No. 126 (DDC No. AD 640 340), General Electric Company, Cincinnati, Ohio.

5-2. Parsons, D. W.: Reliability Demonstration Shillelagh Missile Subsystem. Paper in Annals of Reliability and Maintainability, Vol. 5-Achieving System Effectiveness, Fifth Annual Reliability and Maintainability Conference, New York, July 1965.

5-3. Rothman, D.; Alexander, M. J.; and Zimmerman, J. M.: The Design and Analysis of Sensitivity Experiments, Vols. I and II. NASA CR-62026--Vol. I and NASA CR-62026--Vol. II.

5-4. Fisher, Ronald A.: The Design of Experiments. Fifth Edition, Oliver and Boyd, London, 1949.

5-5. Fisher, Ronald A.; and Yates, Frank: Statistical Tables for Biological, Agricultural and Medical Research. Fifth Edition, Hafner Publishing Company, Inc., New York, 1957.

5-6. Dixon, W. J.: The Up and Down Method for Small Samples. J. American Statistical Association, Vol. 60, No. 312, Dec. 1965, pp. 967-978.

6. Basic Uses of Nonaging Test Types

Descriptions of nonaging test types were presented in Sec. 5. A particular test type may find several uses. An illustration of the complexity of this relationship between test types and basic uses is presented in Fig. 6-1. The purpose of this section is to describe the basic uses and illustrate some practical problems associated with using specific test types for fulfilling them. As described in Sec. 3, the basic problem uses are also encompassed in broader program applications such as reliability, development and acceptance testing. This is further discussed in Secs. 13 and 14.

6.1 Strength Measurements

The strength of an item generally refers to its ability to withstand stress without failure. A measurement of strength is a measurement of an ultimate capability of the item. Here, we are talking about strength as basically a nonaging phenomena in the sense that we are primarily interested in the level of stress (or its severity) which causes failure in contrast to a time at which failure occurs. It is permissable to talk about strength when the response to the stress is time or aging dependent as when one refers to fatigue strength.* One type of fatigue strength, for example, may be defined as the maximum peak cyclic stress for which failure does not occur for 1,000 stress cycles. Also there are strength tests (see, for example, Ref. 6-1) in which failure may not occur immediately after application but after a minute or so has elapsed. Regardless of such situations we generally consider strength as independent of aging.

Whereas in most strength testing of mechanical items the criteria for failure is some predetermined level of damage such as yield or rupture of the material, failure can be defined on any basis from a simple out-of-tolerance condition to total destruction. The strength of a resistor subject to temperature stress may thus be defined as that temperature at which the resistance exceeds its tolerance and the test to measure it may not damage it at all or age it significantly.

A formal definition of strength is provided by an appropriate stress-strength model. For the immediate purpose we will rely on the following: "There is a level of stress severity,** called strength, such that there is failure if and only if the stress severity exceeds the strength." More complex stress-strength modeling concepts are described in Refs. 6-2 and 6-3.

---

* The concept of fatigue being analogous to aging was introduced in Sec. 2.

** The importance of specifying severity instead of just stress itself was discussed in Sec. 2.2.2.

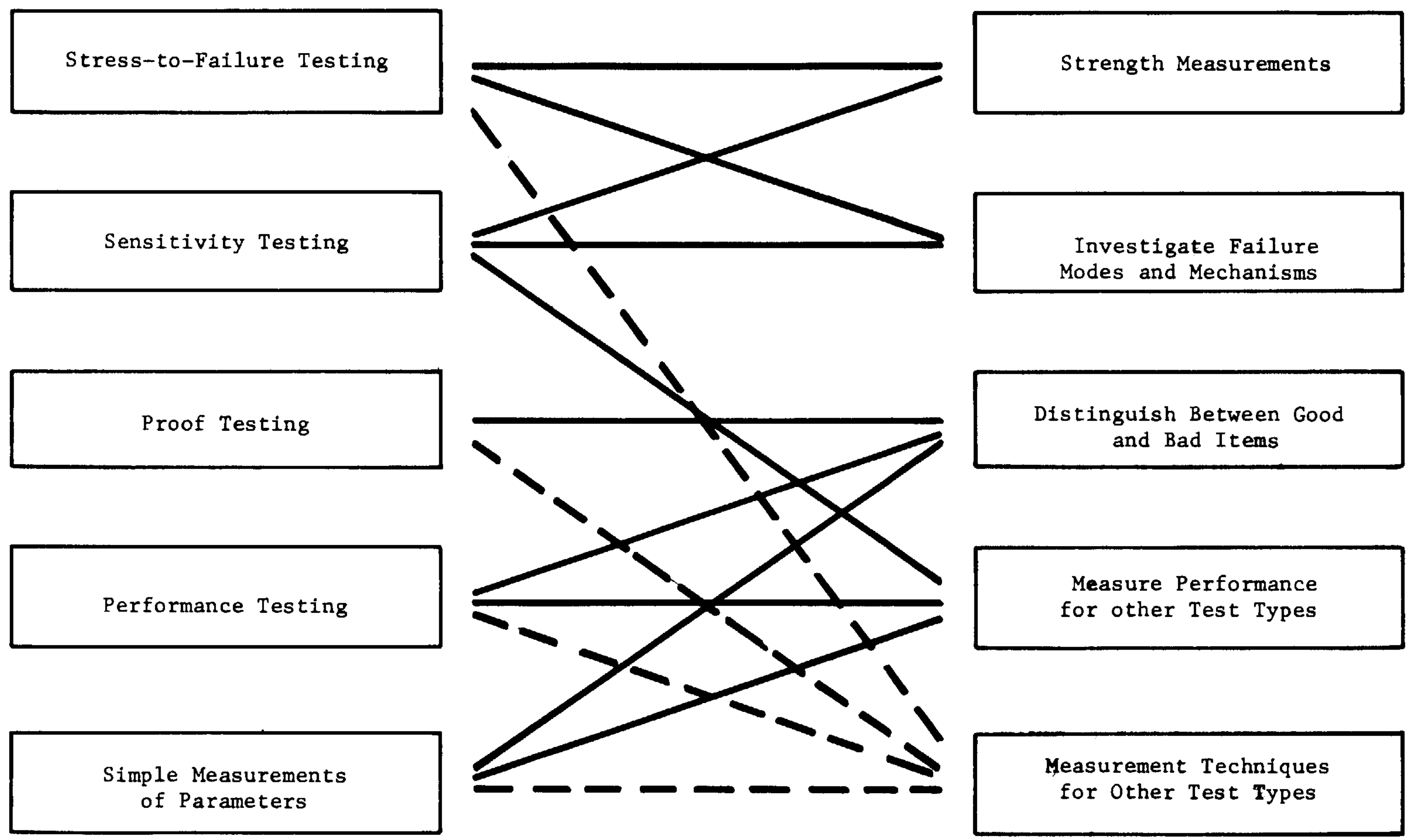


Figure 6-1. Relationship Between Nonaging Test Types and Their Basic Uses

A common example of strength in this context is the yield strength of a metal; however, even for metal one can easily define other strengths such as ultimate strength and rupture strength. Since the concept of stress in testing is extended from mechanical stresses to include other physical entities* such as temperature and radiation, there is also an analogous extension of strength to include these. Remember, however, that the units for strength are those of the severity of the stress which is not necessarily the same as those of the stress itself.

As noted in Fig. 6-1, the measurement of strength can be accomplished by either the stress-to-failure or the sensitivity testing approach. The approach depends upon the nature of the item. For those items which can be tested by the stress-to-failure approach, the strength of individual items can be measured.

<u>Example 6-1</u>

For purposes of application of a particular pressure vessel design, failure of the vessel is defined by a critical dimension y changing by 0.1 in. The strength of the vessel to a stress of gauge pressure is measured by a stress-to-failure approach. The response is illustrated in Fig. 6-2 as the monitored Δy, the change in dimension y, versus gauge pressure. The strength of the vessel is measured as 1960 lbs.

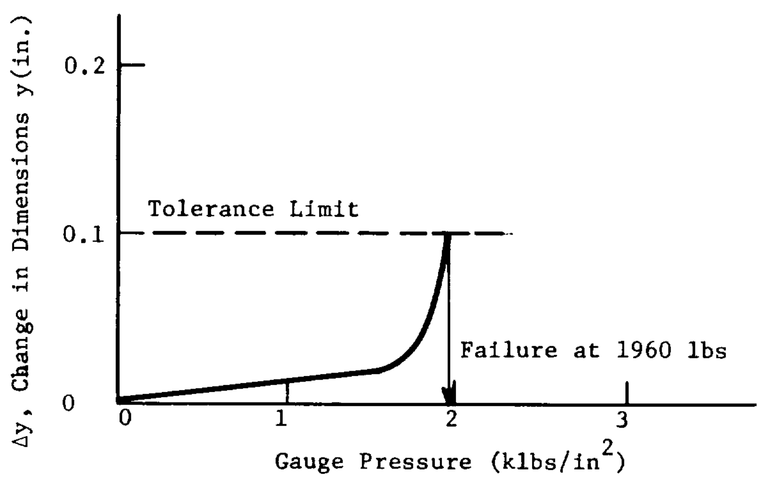


Figure 6-2. Response of Pressure in Stress-to-Failure Test where Failure is Defined by the Tolerance on the Critical Dimension

---

* The extension of stress to include other physical entities and the implication was discussed in Sec. 2.2.2.

Note first that nothing was said in this example about damage to the vessel. It is thus possible that the response can be entirely reversible, i.e., returning the gauge pressure to zero after the test would result in the monitored response retracing the response curve shown in Fig. 6-2.

Whereas, the strength of individual items can be measured by the stress-to-failure approach, it cannot be with sensitivity testing. Only characteristics of a multiple sample can be measured by sensitivity testing. Example 5-18 illustrates the use of the Probit design for measuring the cumulative frequency of response as a function of stress severity and Example 5-20 illustrates how the Up-and-Down design can be used to obtain estimates of characteristics of distributions of stress severity levels which cause failure.

The most useful concept of strength is that of a strength distribution. Fig. 6-3 shows a popular illustration of a strength distribution and its interaction with applied stress. If the strength distribution of an item is known explicitly, then the probability of the item failing due to applied stress having severity level S_1 is

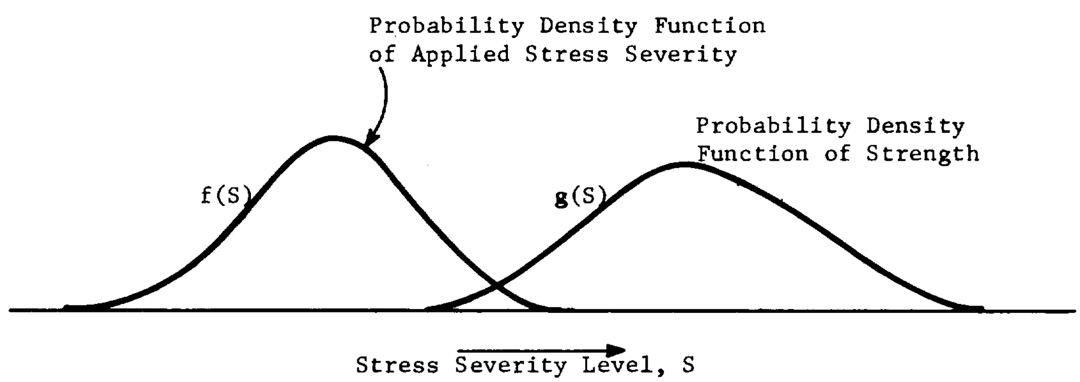


Figure 6-3. Strength vs Applied Stress Severity Distributions

$$\text{Prob. of Failure for Stress Severity } S_1 = \int_{-\infty}^{S_1} g(S)\,dS$$

where g(S) is the probability density function for the item's strength. Furthermore, if the severity of the applied stress was known to be distributed according to a probability density function f(S) then for all possible stresses

$$\text{Prob. of Failure for all Stress} = \int_{-\infty}^{\infty} f(S)\left[\int_{-\infty}^{S} g(\eta)\,d\eta'\right] dS$$

where η is a dummy variable of integration.

Typical uses of these concepts are:

(1) computing safety margins,

(2) reliability demonstration and estimation, and

(3) determining what should be the severity level of the operating stresses.

Safety margins are simply a measure of the separation between operating stress severity and strength but are typically defined in several ways. For example, with uncertainty or variability in both stress severity and strength a popular definition is

$$\text{Safety Margin} = \frac{\mu_{strength} - \mu_{stress}}{\sqrt{\sigma^2_{strength} + \sigma^2_{stress}}}$$

where μ and σ represent the mean and standard deviations. Alternately, one might assume a maximum stress severity S_{max} as a reference and define

$$\text{Safety Margin} = \frac{\mu_{strength} - S_{max}}{\sigma_{strength}} \quad .$$

The use for reliability demonstration and estimation follow from the earlier expressions for probability of failure where reliability is one minus the probability of failure. Examples and further discussion of this appear in Vol. IV - Prediction of this series and Ref. 6-4.

The allowable characteristics of the operating stresses can be determined on the basis of either a desired safety margin or a specified reliability. This first requires knowing the appropriate characteristics of the strength distribution which is an important reason for testing to measure strength.

Rarely is enough testing performed, either by the stress-to-failure or the sensitivity testing approaches to obtain a precise representation of the distribution. Most often, the results of several tests are used to obtain estimates of the characteristics such as percentiles, the median, the mean, and the standard deviation.

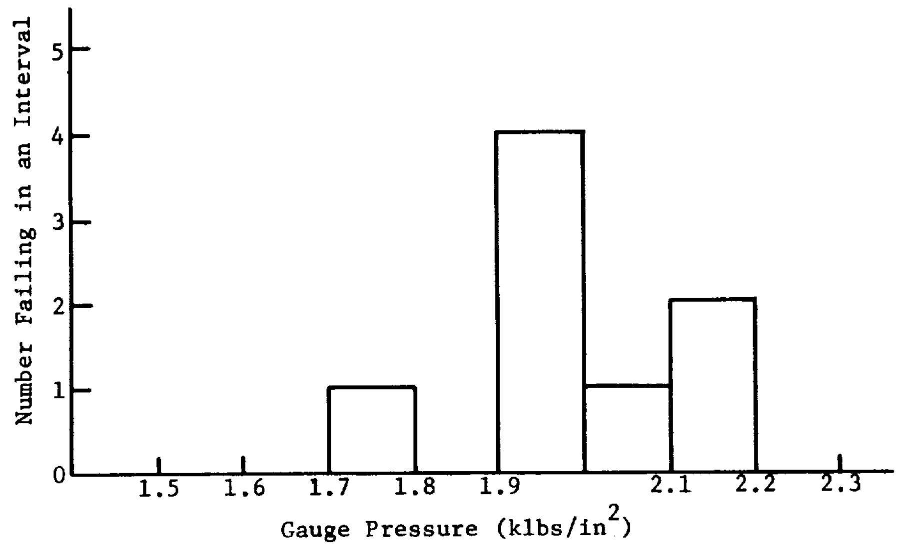


Figure 6-4. Histogram of Stress-to-Failure Results for a Pressure Vessel

Example 6-2

Eight pressure vessels are tested by the method of Ex. 6-1 in order to evaluate the strength of the design to gauge pressure. The results are tabulated below.

Test No.	1	2	3	4	5	6	7	8
Pressure Level at which Failure Occurs, x (lbs./in^2)	1960	2090	1750	1930	2110	1980	2160	1940

Sample Mean $= \bar{x} = \frac{1}{8} \sum_{i=1}^{8} x_i$

$= 1990$ psi.

Sample Std. Dev. $= s = \sqrt{\frac{1}{8} \sum_{i=1}^{8} (x_i - \bar{x})^2}$

$= 121.4$ psi.

A histogram of these results is shown in Fig. 6-3.

On the basis of a small sample of eight measurements no inferences can be made concerning the form of the distribution of the measurements. It is true that a statistical test for a specific form of the distribution, e.g., normal, can be made but the capability of detecting a departure from the specific distribution form is very small. However, it is often the case that a great deal of similar data has been obtained in the past and that certain specific distributions have been found to be adequate representations of the measurements. Two approaches are used below to demonstrate what further inferences can be made on the basis of the small sample subject to stated assumptions.

Approach 1. Assume the measurements are normally distributed with a mean μ and standard deviation σ both of which are unknown. Then on the basis of the estimates $\bar{x}$ and s of the mean and standard deviation and the use of the t-distribution (see Sec. A.1 of the Appendix) one can obtain a confidence interval estimate of the mean stress at which failure occurs. Thus, if a confidence level of 90 percent is selected a one-sided interval is given by

$$\mu \geq \bar{x} - t_{.90(7)} \frac{s}{\sqrt{n}} .$$

For $\bar{x}$ = 1,990, s = 121.4, n = 8 and $t_{0.90(7)}$ = 1.415 (Ref. 6-5), then μ = 1,929 psi with 90 percent confidence.

Another useful type of inference for a small sample of this type is a statistical tolerance interval for an individual measurement, Ref. 6-5 has more detailed discussion and Sec. A.6 of the Appendix has a limited description of the procedure. In this case the inference can be made that 95 percent of individual measurements fall between $\bar{x}$ - ks and $\bar{x}$ + ks, with a selected level of confidence, where k is obtained from tables in Ref. 6-5. In this example, for 90 percent confidence k = 3.136, thus 95 percent of the measurements fall between the values 1,609 and 2,271 with 90 percent confidence. This inference is also based on the assumption that the observations are normally distributed. These tolerance intervals can also be made one-sided as the above confidence interval. Reference 6-6 contains the appropriate tables and a description of the procedure.

Approach 2. Do not assume a specific distribution form; this is thus a distribution-free approach. From tables of Ref. 6-7 the following type of confidence interval can be obtained. For example, one is 96 percent confident that 50 percent of the individual stress-to-failure measurements (obtained under the same conditions as those of the sample) fall between the largest and the smallest value observed in a sample of eight measurements.

Similarly an inference concerning the median (as opposed to the mean for a normal distribution) can be made. For eight measurements one can infer with confidence 0.992

that the median falls between the smallest and largest measurements, i.e., 1,750 and 2,160 psi, respectively. This inference is based on the assumption of symmetry of the distribution. The confidence level is determined by the sample size and the order statistics used (i.e., the largest, next largest, etc.). See Ref. 6-5 for abbreviated tables of information needed for the inferences under this approach.

6.2 Investigation of Failure Modes and Mechanisms

Mechanisms and modes of failure are often of interest from the reliability viewpoint. To investigate these, there is no alternative but to have the item fail. Failure investigations often follow the observation or occurrence of failure during the normal coarse of some other effort such as operational use, prelaunch checkout, field testing, life testing, or proof testing. But it is sometimes necessary, even sometimes formally specifed, that items be purposely failed in order to learn more about the mechanism and modes involved.

As illustrated in Fig. 6-1, the stress-to-failure and the sensitivity testing approaches are the basic test types applicable. The stress-to-failure approach is obviously limited to those items which can be treated in that manner. All items tested by the stress-to-failure approach are available for "post mortem" investigations. Sensitivity testing is less efficient, but there is no alternative for one-shot items which have to be destroyed in order to obtain a response.

The major practical consideration here is making sure that your test gives you the correct failure mode. In the stress-to-failure approach, for example, the severity of the stress can be increased slowly or rapidly. The selection of this rate could be important. It is also possible that multiple stress factors acting simultaneously could give a different failure mode than when each is acting singly.

6.3 Distinguishing Between Good and Bad Items

The basic problem here is one of placing items on test under some specified test conditions and using the results to determine whether each is good or bad. The criteria distinguishing between good and bad can range from a simple in- versus out-of-tolerance condition to not-damaged versus damaged.

As illustrated in Fig. 6-1 the basic test types most suitable for this purpose are proof testing, performance testing, and simple measurements. With the proof testing approach one is most often concerned with whether the item is damaged or not. For example, a structural proof test may be employed in flight qualification testing of an airframe; a load is applied at some predetermined maximum level and the airframe either fails due to damage or passes because there is no damage. This is often referred to as stress-strength comparison.

When the failure criteria is less severe, say, deflection of the aircraft structure being less than some specified amount, the test approach is hardly more than a performance check. More complex criteria of performance may be employed in distinguishing between good and bad items, for example, the shape of a performance response surface or the time behavior over a complex profile of conditions.

Simple measurements are very common for this use with quality control being a very familiar area of application.

Basically, in this category one is only concerned with making a discrete value judgment about each item tested. If the item does not have to be damaged or destroyed as a result of the test, then the procedure can be used as a go, no-go test. That is, items that pass the test are available for other tests or for intended application and items that fail are no longer useful (at least for the original purpose). Specifically for items intended for operation, the procedure becomes a useful screening procedure in acceptance testing; this is discussed further in Sec. 14. If the item is not intended for operational use, the procedure will most often be a qualification test.

When the item operation is a one-shot type such as the normal mode of operation for a squib or a flashbulb and is tested by this procedure, the use emphasized here is not an end result. Rather the value judgment of whether the item is good or bad becomes an input to a strength measurement or a sampling procedure to consider populations of items.

The most basic problem with this application is in selecting the criteria for distinguishing between good and bad. For the proof testing approach this problem was discussed in Sec. 5.3 but as mentioned these can depend heavily on the maximum severity level of the working stress and/or the estimated strength of the item. For the performance testing and simple measurements approach the answer is approached through a tolerance analysis (see Vol. I - Parameter Variation Analysis of this series).

An extension of using results only from a single test to make the value judgment of good versus bad is to use the results of several different tests. The simplest procedure is to apply the different tests in sequence. This is often done in screening programs and further discussion appears in Sec. 14.

Another procedure is to combine the results from different tests according to some mathematical model. A very simple example is the computation of a gain-bandwidth product of an amplifier from separate measurements of gain and bandpass. More complex computations with analytical models are often used. For example, a value judgment for a specific structure may be made by making several different measurements and combining the results through conventional structural equations.

Empirically-based models have also been used. Reference 6-8 gives a discussion of this approach for a linear model of this type. The approach is called linear discriminant. In brief, p parameters are measured for a device and the values $x_1, \ldots, x_p$

substituted into an equation of the form $z = \lambda_1 x_1 + \ldots + \lambda_p x_p$ where the λ's are optimal numerical "weights" associated with the measured values. The equation is the linear discriminant and if z, the weighted average of the parameter values, is less than a critical value z' then the component is unacceptable. Obviously z' has to be determined beforehand and this can require considerable experimentation and analysis. The approach is described in detail in Ref. 6-8. The criteria distinguishing good from bad is arbitrary. In Ref. 6-8 it was based on life; however, it could also be based on strength or performance.

Because of the extensive experimental and analysis effort required, such an empirical approach is feasible only for situations where large numbers of components of the same type are handled.

6.4 Measurement of Performance Characteristics

As introduced in Sec. 5.2 performance is generally concerned with how well (or how poorly) an item performs within the range of conditions in which it is considered capable of performing (possibly even in a degraded fashion). The performance testing approach was addressed specifically to observing this for individual items. Results of performance measurements for individual items are often directly applicable as an end result as the performance measurements of interest. The general problem of measuring performance, however, is broader than this.

Besides performance testing per se, Fig. 6-1 illustrates that sensitivity testing and simple measurements are also applicable for measuring performance characteristics. One may apply the methods of sensitivity testing described in Sec. 5.5 to measuring the performance threshold of one-shot devices. For example, the distribution of supply current which yields a given peak light intensity of flashbulbs of a particular type can be estimated by this approach. The basic procedures and designs are no different from those for measuring strength with this approach.

Simple measurements also yield useful information on performance. The measurement of the distribution of resistance of a certain type of resistances under fixed, benign conditions can be information of considerable importance in performance estimation even under different environments. Considerations of the use of such information is presented in Vol. I - Parameter Variation Analysis of this series.

Response surface determination (i.e., the inverse response problem described in Sec. 2.3.2) for performance and optimization of performance are both frequently encountered as basic problems applications. Some good discussion of these from the experimental design point-of-view is presented in Ref. 6-4.

The general problem of measuring performance characteristics is basically one of estimation. The approach to estimation is discussed in Sec. A.5 and A.6 of the Appendix.

The most outstanding practical problem is selecting the parameters to be estimated. This is related heavily to the intended use of results.

6.5 Measurement Techniques for Other Test Types

In a proof test the measurement of response is an attributes measurement. The measurement itself in a proof test is really the same in concept as the simple measurements described as a basic test type in Sec. 5.1. Measurements are involved in all basic test approaches, even in aging types. For example, a life test of a component requires making simple measurements over time to determine when a failure response has occurred. Also, a performance test to measure the dependence of a performance parameter on a stress requires certain simple measurements.

The above is a simple illustration of how one basic testing approach can be a measurement technique for another. Other nonaging test types are applicable in this role.

Consider, for example, a measurement of life of a power supply and assume that it is desired to age it while operating at 50% load but that the desired criteria for success or failure is whether it is capable of supplying 120% load when demanded. The test can be designed to provide these aging conditions with periodic checks at 120% load. These checks at 120% are thus nothing more than a series of proof tests dispersed throughout the duration of its life.

Similarly, a stress-to-failure or a sensitivity testing approach is sometimes nployed to measure periodically over time a residual strength of items which are eing aged. This approach would assume that strength is changing with age or else here would be no benefit to testing more than once. An example of this procedure o determine a hazard rate function is described in Ref. 6-9.

It is of utmost importance in knowing what is being measured with such approaches as failure to do so can readily cause misinterpretation. For example, to age an item at one stress level and then in a very short period cause it to fail by a stress-to-failure approach may introduce a completely different failure mode than if failure had resulted later from continued aging. Failure from aging is a cumulative damage phenomena and is to be interpreted quite differently from failure resulting directly from stress. This specific concept was discussed briefly in Sec. 2.2.2; more detailed discussion is presented in Ref. 6-2.

References

6-1. Palerno, Peter M.: Strength Tests of Aircraft Securing Fittings for CVAGI. Report 970 (DDC, AD 850 306), The David Taylor Model Basin, Washington, D. C., July 1955.

6-2. Evans, R. A.: Stress Versus Damage. Proceedings 1967 Symposium on Rel., Washington, D. C., Jan. 1967, pp. 633-635.

References (Continued)

6-3. Marks, L. S.; and Baumeister, T.: Mechanical Engineer's Handbook. McGraw-Hill Book Co., Inc., 1958, pp. 5-69 to 5-70.

6-4. Lloyd, D. K.; and Lipow, Myron: Reliability: Management, Methods, and Mathematics. Prentice-Hall, Inc., Englewood, N. J., 1962.

6-5. Dixon, W. J.; and Massey, F. J.: Introduction to Statistical Analysis McGraw-Hill Book Co., Inc., 1957.

6-6. Lieberman, G. J.: Tables for One-Sided Statistical Tolerance Limits. Industrial Quality Control, vol. 14, no. 10, pp. 7-9.

6-7. Murphy, R. B.: Non-parametric Tolerance Limits. Ann. Math. Statistics. vol. 19, pp. 581-589.

6-8. Vacarro, J.; and Gorton, H. C.,eds.: RADC Reliability Physics Notebook. U. S. Air Force RADC - TR-65-330 (DDC, AD 624 769), Battelle Memorial Institute, Columbus, Ohio, 1965.

6-9. MacCarley, John A.: Testing for Reliability. Part II of Reliability Control in Aerospace Equipment Development, SAE Technical Program Series, Vol. IV, The Macmillan Co., New York.

7. Aging Test Types

When aging dependent properties of an item are to be investigated by an experimental approach, it becomes necessary to test items for sufficient durations to observe and measure these properties. For example, if one really wants to know how long a gyro will last before wearout, he has no alternative but to test it until failure occurs. The test duration may possibly be shortened by accelerating the wearout process in some way; however, it must still be operated until its useful life is consumed.

Certain aging effects can be treated by testing for some duration less than the total life the item is normally expected to have. For example, the observed performance in the early portion of the life of a component may provide some clue about how well it will perform during later life. Component failure rate estimates are often made by logging a certain number of device operating hours during the early life of a sample of items; however, this approach can have certain weaknesses depending upon the distribution of failures over time and the importance of the estimate precision.

A summary of basic test types which emphasize aging is presented in Table 7-1. The distinctions here are based generally on the relative severity of aging.* Failure-truncated testing is the most severe in that all of the useful life of all or a significant portion of the sample tested must be consumed. In the sequential approach, testing may involve failures but the test must be continued until sufficient time passes to make some value judgment about the item (or sample); the amount of aging required of tested items may thus range from small to large amounts. In time-truncated testing the relative amount of aging of each item is generally much less than for failure-truncated testing because of the shorter test duration. Performance testing is a special situation where the amount of aging depends upon how much is needed to observe the performance properties of interest. When the amount of aging becomes insignificant or irrelevant, the basic nonaging types described in Sec. 5 apply.

It is important to re-emphasize what we mean by aging. Aging is a nonreversible process that occurs with the passage of time and results in the accumulation of damage. Most often, aging can be equated intuitively to expending useful life of an item; however, as in the case of annealing (negative damage), not all aging is detrimental and may thus actually improve the item and lengthen its life.

---

* This is in contrast to the distinctions among the nonaging types which were based upon the relative severity of the stress.

Table 7-1

Comparison of Aging Test Types

Test Type[a]	Examples	Distinguishing Features[b]
Time-Truncated Test	Mission duration simulation of a spacecraft component to determine if it survives the mission Power burn-in of transistors	Items are purposely aged only for a fixed duration.
Sequential Testing	Sequential-life testing for performing accept-reject decisions of a design	Testing continues until sufficient time passes to determine whether items are good or bad; this is a special approach employing sequential analysis to determine whether to continue testing.
Failure-Truncated Test	Battery life test Failure mode investigation of a gyro for long term operation	Items are purposely aged until each fails or a specified proportion of a population fails.
Performance Test	Measurement of transistor current gain for long durations Periodic functional checking of a radar for long term degradation	Items are purposely aged by an amount dependent upon the needs of the test; the purpose is to effect aging-dependent changes in performance characteristics.

[a] Note especially that accelerated testing is not identified as a basic test type but is applicable to all of the above mentioned approaches.

[b] When the amount of aging is insignificant or irrelevant the nonaging test types apply.

A run-in test of an engine, for example, may be concerned with improving it both from a performance and life viewpoint. From a reliability point-of-view, however, we are most frequently concerned with the detrimental effects as when we are looking for failures or degraded performance.

In all of these test types the observed responses are thus assumed to result from an underlying cumulative damage process which itself may not be readily (if at all) observable. A transistor, for example, may exhibit the behavior we want for months but then, all of a sudden, fail with no indication that this was imminent. On the other hand we may be able to periodically check the wear of a bearing and predict when failure will occur. It makes little difference from the basic testing viewpoint used here whether the damage is internal (physical or chemical structure or composition changes) or external (wear, erosion, etc.) so long as it is cumulative.

Note that the classification acknowledges little about the test conditions themselves, i.e., about their severity or how they are programmed over time. Certainly the rate of the aging process is dependent upon this. This is really common to all of these aging types. Accelerated testing, for example, purposely speeds up the aging process and is applicable to all of these.

With the above classification, the descriptions of the basic test types are notably simpler than for the nonaging types.

7.1 Failure-Truncated Testing

In a failure-truncated test of a single item the item is subjected to the conditions until a failure criteria is met. The item is thus always failed. The conditions may be constant or varying and the failure criteria may range from a simple out-of-tolerance condition to total destruction.

<u>Example 7-1</u>

> A nominal six-volt battery is left connected to a constant load until failure occurs. Failure is defined by the output voltage decreasing to less than five volts.

The purpose of this time-truncated test may be anything from measuring the life under this condition to continuously monitoring the electrolytic composition.

Failure-truncated testing is often applied to samples of several items. In certain life tests for example items are placed on test and the test continued until a certain proportion have failed.

<u>Example 7-2</u>

> One-hundred resistors are tested at a fixed ambient temperature and fixed current level until a total of seventy have failed.

This simply illustrates the approach. The actual purpose of such a test may be to determine the distribution of failures over time in which case the time at which each failure occurs would be recorded. The termination criteria, i.e., the proportion of items that have to fail before test termination, depends upon such factors as what is being measured, the desired precision and confidence of the result, and the sample size.

The major disadvantage of failure-truncated testing, especially for high reliability items, is the extensive test time required. Items with very low failure rates require an unacceptably long time to fail a significant number of them. For example, failure rates on the order of 10^{-6} failures/device-hour are not at all uncommon and a single device from such a population might have to be tested for more than 100 years in order to observe failure. It is thus obvious why accelerated testing approaches for shortening the test time have been promoted.

Failure-truncated testing of long life items is often supplanted by the time-truncated approach. For example, 100 identical devices tested for 2000 hours gives 200,000 device-hours (approximately 20 device-years) of testing. For these to be "equivalent" device-hours, however, requires the distribution of failure times to be exponential. This problem is discussed further in Sec. 8.1.

The most frequent applications of the failure-truncated approach are for measuring life (and related) characteristics and investigating modes and mechanisms of failure. These uses are explored in more detail in Sec. 8.

7.2 Sequential Testing

In a sequential testing approach the duration of the test (hence the amount of aging) depends upon the results accumulated during the test. The most publicized form is sequential life testing. In this the test results are used continually from the start of the test to compute some decision function, typically the total operating time versus the total number of failures from the start of the test. Based on certain decision criteria, the test continues until a decision can be reached whether the product (or design) is acceptable or unacceptable.

The sequential life testing procedure is illustrated in Fig. 7-1. The accept and reject lines represent the decision criteria. The function containing the steps at each failure number is the decision function and in this case it illustrates that testing continues until after the eighth failure occurs with the product then being accepted. It is computed as the total operating time from the start of the test for all items placed on test.

A significant advantage to this approach is that when items are very good or very bad considerably less test time is required to reach an accept-reject decision

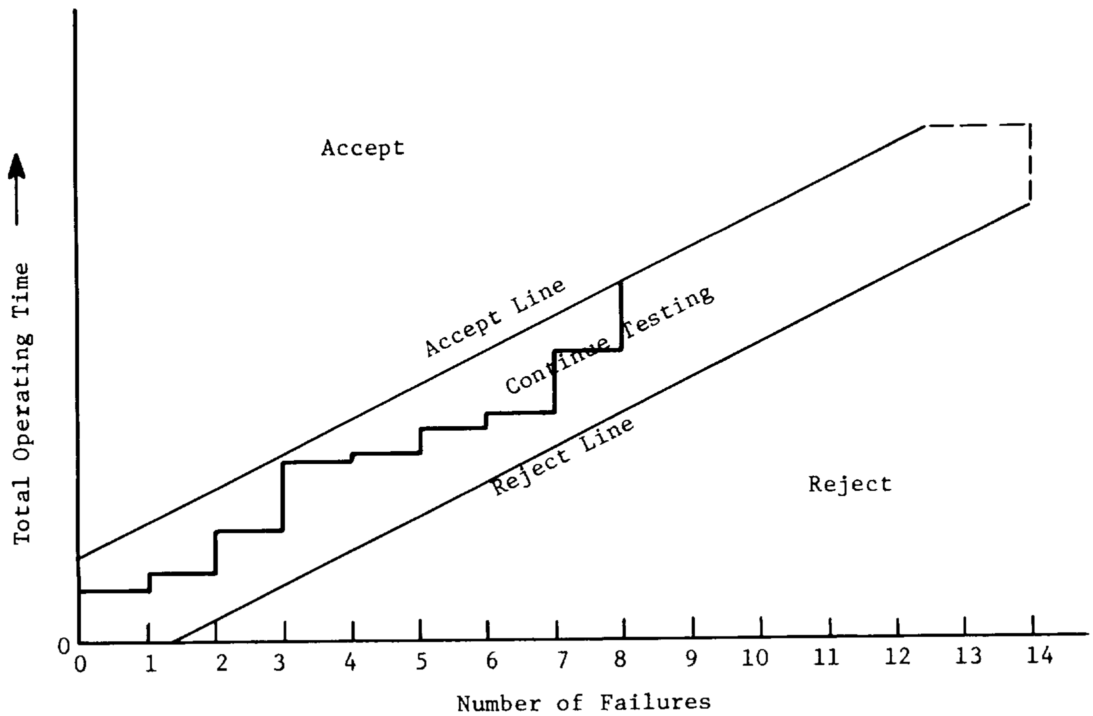


Figure 7-1. Illustration of Sequential Life Testing

than with either of the non-sequential approaches, failure-truncated or time truncated. When, however, a product is marginal, i.e., not exceptionally good or bad, the test can run on indefinitely producing indecisive results. For this reason, added conditions are often assigned to truncate the test so that the test is terminated if an accept-reject decision has not been made before a preassigned time and/or number of failures. Figure 7-1 illustrates combined truncation of both operating time and failure. Truncated tests are often referred to as truncated sequential tests. The truncation can be a very important feature in terms of prior allocation of test facilities, manpower, and specimens for test.

These procedures can be employed with single items or several items and with or without replacement of failed items. Essentially all designs for sequential life tests known to be implemented to date have been based on the assumption of constant hazard rates.* Sequential testing designs are discussed in more detail in Sec. 8.1

---

* The definition of hazard rate and description of different forms are described in Sec. A.9 of the Appendix.

but basically they consist of specifying the slope and intercepts of the accept-reject lines and the truncation limits if appropriate. This selection of accept-reject lines depends upon the acceptable mean life, the unacceptable mean life, and the risks* of accepting bad products or rejecting good products. In addition, truncation depends upon the minimum sample size and slope of the accept-reject lines.

The above discussion has focused on sequential life testing to illustrate the concept. The concept can be interpreted in a broader sense to include purposes other than for measuring life characteristics even though the testing and reliability literature also concentrates on these. An engineer monitoring the amp-sec. output of a battery in an environmental test to determine whether to terminate or continue testing is certainly using a sequential procedure.

7.3 Time (or cycle)-Truncated Testing

In a time-truncated test of an item the test ends after some prespecified duration (or number of stress cycles) of testing or when the item fails, whichever occurs first. The item is thus not always aged until failure occurs.

<u>Example 7-3</u>

> In a flight acceptance test of a satellite, exposure to an ambient temperature of -10°C is specified for a duration of 24 hours. The system is operated briefly at one-hour intervals throughout this period to demonstrate the capability to operate under this condition.

The basic purpose of this test might typically be to uncover potential failures resulting from workmanship errors during fabrication. In this case it must be assumed that only a very small portion of its useful life (if hardly any at all) is consumed unless perchance the intended mission itself is very short.

Time-truncated testing is also applicable to sample sizes larger than one.

<u>Example 7-4</u>

> Fifty transistors of a particular type are stored in an oven at 125°C for 2,000 hours. Following this exposure, the electrical properties of each are checked for conformance to tolerance. Two were determined to be failures at the end of the test.

Such a test might typically be used for qualifying the parts, estimating certain characteristics about them, or bake-in to weed out defectives from the population.

The time-truncated testing approach is a necessity when aging has to be limited as in burn-in. Also as mentioned in Sec. 7.1, one advantage of time-truncated testing over the failure-truncated approach is the shorter test time required.

---

* These are often referred to as the consumer risk and the producer risk respectively; see Sec. A.10 of the Appendix for more detailed description.

7.4 Performance Testing

Performance testing as a basic aging test type is a logical extension of the nonaging type having the same designation. The extension is not merely in time since certain nonaging performance tests also require testing over time but rather in the underlying mechanisms causing observed changes in performance. When aging is involved, the mechanism is one of cumulative damage.

The following example illustrates a test to measure aging-dependent performance changes in a single item.

<u>Example 7-5</u>

> A rate gyro is operated continuously while subjected to a known periodic positive and negative rotation with fixed amplitude about its input axis and all other conditions fixed. The rms error monitored in the output signal during the test is illustrated in Fig. 7-2.

This test may be typically one conducted by the gyro manufacturer to measure the effect of gimbal bearing wear on performance. The large increase in error after several hundred hours of operation indicates that wear indeed may be degrading performance. Also, as illustrated by the initial period, performance may improve with some aging rather than degrade.

The plot alone, however, does not indicate that the performance changes are due solely to aging. One quick check for this would be to stop and restart the test. If the performance did not change in this process, then it is reasonable to assume that aging is the cause. If performance changed back to that at the start of the test, there would be strong suspect that no significant aging was involved since the performance change appears to be reversible. Any other magnitude of change may or

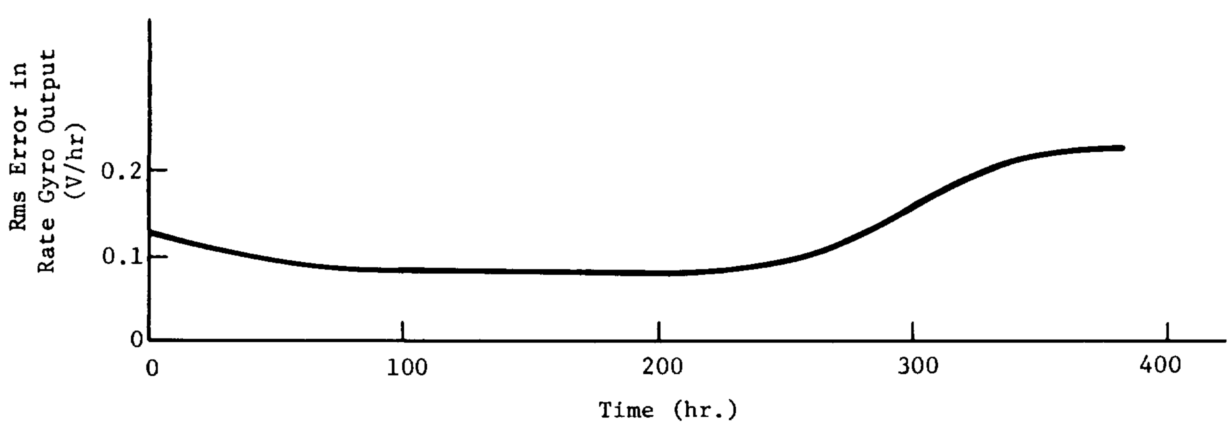


Figure 7-2. Gyro Erros Due to Aging

may not indicate significant aging. A series of stops and starts for example may actually show that there is considerable variability present. The presence of aging then may only be detected by testing to measure changes in the statistical properties over time, say the average of several such tests.

Example 7-5 dealt with performance testing a single item for a fixed amount of aging. The observed performance is characteristic of that single item and only for the period of observation. Attempts to extrapolate from performance measurements on one item to similar items and from behavior of one period to another are often made. They cannot be done legitimately on the basis of single test results alone but require additional knowledge of the aging process available from either theory or other tests. Even then extrapolation must be done with discretion.

The value of a performance test of a single item which involves aging has its greatest significance only when the item is representative of a population or is one item of a larger sample. Use of performance tests in this context are discussed in more detail in Sec. 8.

8. Basic Uses of Aging Test Types

The four basic testing approaches which emphasize aging were described in Sec. 7. The relationship between these test types and basic uses is illustrated in Fig. 8-1. Discussion of basic uses is presented below.

8.1 Life Measurements

The life of an item is an important trait from the reliability point-of-view and represents another measure of ultimate capability.* The life of an item is simply the duration that it maintains the properties that we intend for it to have. When it no longer possesses these properties we say that it has failed. The definition of failure depends upon the properties but range in complexity from a simple out-of-tolerance condition to total destruction.

When talking about failure as a result of useful life being expended, it is understood that failure results from a cumulative damage process.** The observed mode of failure of a resistor subjected to high temperature for a long duration may actually be a sudden discontinuity or "opening" between its terminal; however, we assume that this resulted from some underlying time dependent process leading to the failure.

Obviously the life of an item intended for application cannot be measured directly since this requires destruction. Repairable items can be repaired and retested repeatedly but generally they must be treated as new items. Repeated tests of the same repaired item does result in increased (engineering) confidence, however, this approach is not necessarily equivalent to testing a sample of similar items.

From the reliability viewpoint typical reasons for wanting to measure life (or related characteristics) of a sample of items are concerned briefly with using the results to

(1) predict the probability of success of similar items scheduled for application, and

(2) determine the acceptability of a product or design.

Both of these may involve aspects of both estimation and test of some hypothesized value of a life characteristic or index. Typical life-related characteristics used for this purpose are life distributions, mean life, mean-time-between-failure (MTBF), mean-time-to-(first) failure (MTTF), failure rates, and hazard rate functions.

---

* Strength is the other measure of ultimate capability and was introduced in Sec. 6.1.

** This concept was discussed in more detail in Sec. 2.2.2.

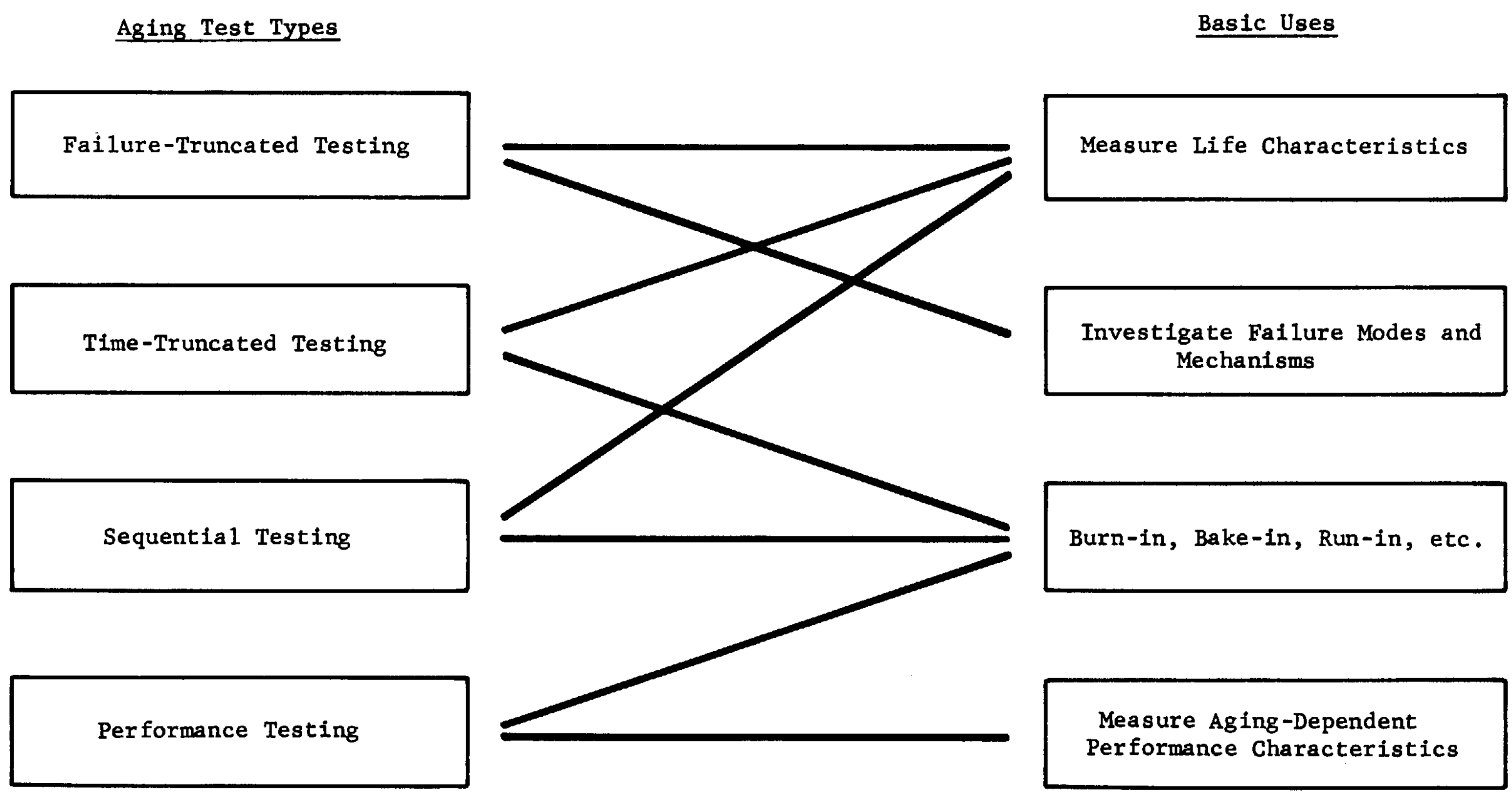


Figure 8-1. Relationship Between Aging Test Types and Their Basic Uses

One would normally have the greatest confidence in a life test if the test conditions simulate those of normal use for the item. But due to required application time or complexity of the normal use environment it is seldom possible to do this.

Accelerated testing is one popular method of shortening the test time; however, this can introduce many problems and decrease the (engineering) confidence in the results if used without familiarity of how the increased severity is really affecting the aging. Concepts and problems with accelerated testing are discussed separately in Sec. 11. At any rate, the use of accelerated testing does not basically influence the statistical design of the test.

Most often, life measurements are made under fixed conditions usually chosen with discretion to represent some typical conditions for application. A particular type of a transistor may find several different applications in a design and as a result be exposed to a different environment for each. Does one then select the most severe of these and test a sample in this environment to measure life? There is no clear-cut answer to this for it is even difficult in many cases to know a priori which environment is the most severe. Even using the most severe environments for testing will lead to pessimistic results for estimating the life of the system. This is typically illustrated by the known diversity of testing and operating environments from which generic failure rate data of parts is derived and which in turn has led to the use of application factors to adjust failure rates for each application. But the fact is that an environment has to be specified for a test and good engineering judgment is obviously needed. To the extent feasible one would match the test conditions to those of normal application.

As illustrated in Fig. 8-1, there are three basic testing approaches for measuring life. The test approaches themselves were described in Sec. 7; their use specifically for measuring life are discussed below. Two of these, failure-truncated and time-truncated are both nonsequential approaches and consequently, are treated together. The sequential approach deserves special attention and is treated later in a separate discussion.

Nonsequential Testing Approaches for Measuring Life: Failure- and Time-Truncated Testing

In either of these approaches n items are placed on test and the test continued until the appropriate termination criteria are met. The information generally required to plan a life test consists of

(1) the sample size n,

(2) the testing approach (whether failure- or time-truncated),

(3) the decision of whether to replace failed items,

(4) the termination criteria,

(5) the degree of precision (i.e., the desired statistical confidence in the case of estimation and the accept-reject risks in the case of testing an hypothesis), and

(6) the model assumed for the distribution of failures.

There is a considerable degree of dependency among these factors.

Consider, for example, that it is desired to test the hypothesis that the mean life of a particular type of item is 6,000 hours. If a particular lot of these items has an actual mean life of 3,000 hours, it would be desirable to plan the test so that there is high probability that the hypothesis, and hence the sample, would be rejected. By making certain assumptions about how the failures are distributed, the number to be tested and the testing time (or number of failures) can be determined to accomplish the desired aim of discriminating between good lots having a mean life greater than 6,000 hours and poor lots with mean life less than 3,000 hours.

It stands to reason that the probability of a lot being accepted increases as the mean life increases. A typical relationship of this type is shown in Fig. 8-2

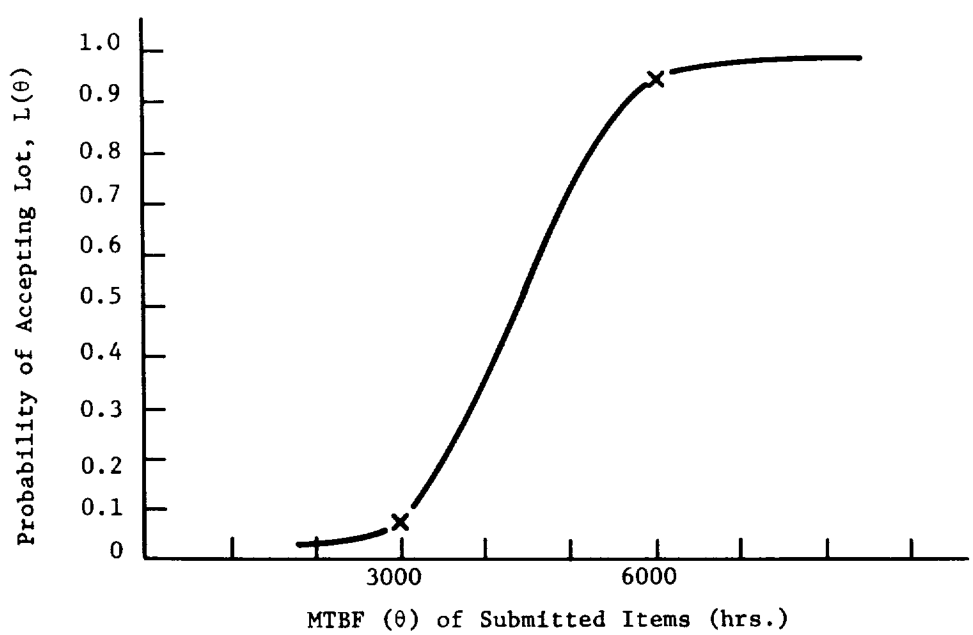


Figure 8-2. An Operating Characteristics (OC) Curve, Probability of Acceptance, $L(\theta)$, vs. Mean Life, θ.

and is called an operating characteristic (OC) curve. OC curves play a large role in implementing life tests when used to test hypotheses. The procedures are discussed later and illustrated by example.

There is a trade-off between the size of n and the waiting time to test termination. For the failure-truncated approach the trade-off is subject to a constraint on the specific number of failures required to obtain a given precision. With the time-truncated approach precision is dependent mainly upon the total test time of all items tested and the trade-off must allow for this.

In both approaches there is an option of whether or not to replace failed items during the test to maintain n constant. The decision should be heavily influenced by the relation of the test to normal application. For example, if the test is conducted to determine the proportion of spacecraft components of a given type which survive a given mission time, then it is desirable to start with n items and not replace failed items.

The termination criteria is dependent upon the desired degree of precision. For the failure-truncated approach the number of failures is the sole requirement and the greater the precision the more failures are required. For short-life items it is not uncommon to fail every item of a sample, but as noted in earlier discussion high reliability (long-life) items can demand unreasonable test durations.

For the time-truncated approach the total test time of all items tested is the basic criteria to be made compatible with the desired precision. Most often, a mean life is assumed or estimated a priori and an appropriate test duration determined from this.

The model assumed for the distribution of failures has a large influence on the test design and subsequent analysis. When there is not enough information to make an intelligent assumption about the distribution of failures, it is preferred to use a distribution-free design. This is compatible only with the time-truncated approach and entails using only the number of failures (or successes) versus the number tested as the basis for the estimate or test of hypothesis.

Usually there is enough prior knowledge to make an intelligent assumption. For example, such information as test results on an earlier design or tests on items of a similar design are available. Increased emphasis on determining such distributions on a theoretical basis has evolved in reliability physics, but this has proven to be generally an inefficient approach for this sole purpose.

A number of life test designs are summarized in Table 8-1. The specific assumption concerning the distribution of failures is the major factor used to distinguish these since it serves a key role in the test design. Since a complete tabulation of all related facts would be voluminous, Table 8-1 emphasizes scope primarily and

Table 8-1

Life Test Procedures and Related Information

Legend:

$f(t)$ = density function for time to failure

f = total number of failures

n = total number of items placed on test

R = Reliability

s = total number of successes

t_i = time of failure from start of test for the i-th item

T = total life of all items on test

T_t = total test time

θ = mean life, $\lambda = 1/\theta$ = failure rate

Note: Other symbols are defined in the table.

Failure Distribution Assumption	Special Design Features	Type of Measurement	Estimate or Hypothesis Tested	Reference
Distribution free (nonparametric)	Time-truncated Specify n, T_t Without replacement	$s = n - f$	Estimate: $\hat{R} = s/n$	8-1 p. 452
Distribution free	Time-truncated Specify n, T_t Without replacement	$s = n - f$	Estimate R: $\hat{R} = s/n$	8-1 p. 452
Distribution free under increasing (decreasing) failure rate assumption	Time-truncated Specify n, T_t Replace failed items Acceptance test single sample	s (or f) c = acceptance number p^* = confidence level desired	Test hypotheses concerning θ or some percentile of the distribution or establish with desired confidence a one-sided interval for θ. Tables of plans are given in Reference 8-2.	8-2

Table 8-1 (Continued)

Failure Distribution Assumption	Special Design Feature	Type of Measurement	Estimate or Hypothesis Tested	Reference
Exponential failure density function $f(t) = \lambda e^{-\lambda t}$, $t \geq 0$, $\lambda > 0$	Time-truncated Specify n, T_t Without replacement	s (or f) t_i's	Estimate θ: $\hat{\theta} = T/f$ where $T = \sum_{i=1}^{f} t_i + (n-f)T_t$	8-1 p. 451
	Time-truncated Specify n, T	s (or f) t_i's	Estimate θ: $\hat{\theta} = T/f$	8-1 p. 451
	Failure-truncated Specify n, f Without replacement	t_i's	Estimate θ: $\hat{\theta} = T/f$	8-1 p. 447
	Failure-truncated Specify n, f With replacement	t_f	Estimate θ: $\hat{\theta} = T/f$ where $T = n\, t_f$	8-3
Two-parameter Weibull density function $f(t) = (\gamma/\theta)(t/\theta)^{\gamma-1}$ $\exp\{-(t/\theta)^{\gamma}\}$ $t>0$, $\gamma>0$, $\theta>0$	Failure-truncated Specify f = n Specify f = n	t_i's	Estimate γ and θ: Case 1: Estimate γ with θ unknown, Case 2: Estimate γ with θ known, Case 3: Estimate θ with γ known. See ref. for estimation procedures.	8-4

Table 8-1 (Continued)

Failure Distribution Assumption	Special Design Feature	Type of Measurement	Estimate or Hypothesis Tested	Reference
Two-parameter Weibull density function (See above)	Time-truncated Specify n, T_t f must be ≤ 2	t_i's	Estimate shape parameter γ, see ref. for the procedure.	8-5
Weibull distribution $F(t) =$ $1-\exp[-(\frac{t-\gamma}{\eta-\gamma})^{\beta}]$ t = age at Failure γ = minimum life parameter η = characteristic life, the 63.2 percentile β = shape parameter	Specify n, f = n Failure truncated	t_i's	Graphical estimation of parameters of Weibull distribution; γ is estimated by trial and error; $\hat{\beta}$ and $\hat{\eta}$ are obtained from graph and a nomograph furnished on the special graph paper.	8-6
Weibull density function $f(t) = (\kappa/\theta)(t/\theta)^{\kappa-1}$ $\exp\{-(t/\theta)^{\kappa}\}$, Location parameter known or 0. Known shape parameter.	Specify n, f = m Failure-or time-truncated	t_i's	Estimate θ: $\hat{\theta}_{mn} = \{t_1^{\kappa}+t_2^{\kappa}+\ldots+t_m^{\kappa}+(n-m)t_m^{\kappa}/m\}^{1/\kappa}$; Unbiased estimates given by $\tilde{\theta} = [m^{1/\kappa}\Gamma(m)/\Gamma(\overline{m+1}/1\kappa)]\hat{\theta}$, Var $\tilde{\theta}/\theta^2$ and $\tilde{\theta}/\hat{\theta}$ are given in Table 1 of Ref. 8-10.	8-10

Table 8-1 (Continued)

Failure Distribution Assumption	Special Design Feature	Type of Measurement	Estimate or Hypothesis Tested	Reference
Two-parameter Weibull density function $f(t) = (\gamma/\theta)\ t^{\gamma-1}$ $\exp\{-t^{\gamma}/\theta\}$ $t,\ \gamma,\ \theta > 0.$	Specify n, T_t Time truncated	t_i's	Estimate θ and γ: $\hat{\theta} = \Sigma t_i^{\hat{\gamma}}/n$ and where $\hat{\gamma}$ is the solution of $\frac{\Sigma t_i^{\gamma}\ \ell n\ t_i}{\Sigma t_i^{\gamma}} - \frac{1}{\gamma} = \frac{1}{n}\sum_1^n \ell n\ t_i$	8-3
Compound Weibull distribution $F(t) = \alpha F_1(t) +$ $(1-\alpha)\ F_2(t)$ where $F_i(t) =$ $1-\exp\{-t_i^{\gamma}/\theta_i\},$ $i = 1,\ 2$	Specify n, f = n Failure truncated	t_i's	Estimation of five parameters by the method of moments; computer program is given for performing the required computations.	8-37
Generalized Gamma density function $f(t;\ a,\nu,p) =$ $\|p\| t^{p\nu-1}\exp\{-(t/a)^p\}/$ $a^p\Gamma(\nu)$ $t \geq 0,\ p \neq 0, a, \nu > 0$	Failure truncated Specify n, f = n	t_i's	Estimation of parameters a,ν,p. See ref. for methods.	8-8

Table 8-1 (Continued)

Failure Distribution Assumption	Special Design Feature	Type of Measurement	Estimate or Hypothesis Tested	Reference
Gamma density function $g_r(t;\theta) = \dfrac{e^{-t/\theta}t^{r-1}}{\theta^r\Gamma(r)}$ $t>0, \theta>0$ r assumed known.	Failure truncated Specify n, f	t_i's	Estimate θ by means of a linear function of the ordered times to failure.	8-11
Gamma density function $g(t;\alpha,\lambda,\eta) = \dfrac{\lambda^\eta}{\Gamma(\eta)}(t-\alpha)$ $\exp\{-\lambda(t-\alpha)\}$, $(\alpha \leq t<\infty, \lambda>0, \eta>0)$ λ = scale parameter η = shape parameter $g = 0$ for $y < \alpha$ α = location parameter	Failure truncated Specify n, f = n	t_i's	Assuming η known estimate θ by means of a linear function of the ordered times to failure.	8-12
Gamma density function As above with η known.	Failure truncated	t_i's	Estimate scale and location parameters using linear estimate based on k sample quantiles.	8-13
Generalized Gamma density function $f(t;c,a,b,p)=p(t-c)^{bp-1}$ $\exp\{-[(t-c)/a]^p\}/a^{bp}\Gamma(b)$	Failure truncated Specify n, f = n	t_i's	Iterative procedure given for solving maximum-likelihood estimate equations for estimates of the four parameters. Procedure is programmed in FORTRAN.	8-18

Table 8-1 (Continued)

Failure Distribution Assumption	Special Design Feature	Type of Measurement	Estimate or Hypothesis Tested	Reference
Exponential density function $f(t) = (1/\theta)e^{-t/\theta}$	Failure truncated Specify n, f = n	t_i's	Graphical procedure for estimating distribution parameter; life is plotted vs. expected value of order statistics.	8-9
Gamma density function $f(t)=[1/\theta\Gamma(a)](t/\theta)^{a-1} e^{-t/\theta}$ $(t\geq 0;\theta,a>0)$			Graphical procedure for estimating parameters of the gamma distribution; gamma variables are plotted vs. the expected values given in Tables of the ref.	
Weibull density function $f(t)=\frac{K}{\theta}(\frac{t}{\theta})^{K-1} \exp\{-(\frac{t}{\theta})^K\}$			Graphical procedure for estimating distribution parameters; life is plotted vs. expected values of order statistics for various values of shape parameters K; after K is obtained by trial and error, the intercept and the slope of the best fitting straight line are estimates of the location and the scale parameters respectively; expected values are given in the Tables of the ref.	
Gamma density function $f(t;c,\theta,\alpha)=[1/\Gamma(\alpha)\theta], [(t-c)/\theta]^{\alpha-1} \exp[-(t-c)/\theta]$ $\theta,\alpha>0,t\geq c\geq 0.$	Failure truncated Specify n, f	t_i's	An iterative procedure is given for finding the maximum-likelihood estimate of the parameters from the given data; a computer program in FORTRAN was written to perform the computations.	8-15

Table 8-1 (Continued)

Failure Distribution Assumption	Special Design Feature	Type of Measurement	Estimate or Hypothesis Tested	Reference
Normal distribution $f(t;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\{-(t-\mu)^2/2\sigma^2\}$ $\sigma>0$, $-\infty<\mu<\infty$ σ known.	Time truncated Specify n, T_t or $f = c+1$ (attribute-type plan).	f	Test hypotheses concerning the mean life, $\mu = \mu_0$ vs. $\mu(<, >, \neq)\mu_0$ Tables given in ref. 8-14	8-14
Log-normal distribution $\frac{1}{\sqrt{2\pi}\,\sigma t} \exp\{-\frac{(\ln t-\mu)^2}{2\sigma^2}\}$ $t>0$, $\sigma>0$, $-\infty<\mu<\infty$. σ known.	Time truncated Specify n, T_t or $f = c+1$ (attribute-type plan).	f	Test hypotheses concerning the median life, e^{μ} $e^{\mu} = e^{\mu_0}$ vs. $e^{\mu}(<, >, \neq)e^{\mu_0}$. Tables given in ref. 8-14.	8-14

serves to guide the reader to appropriate references for more detail. Confidence interval estimation procedures, for example, can be obtained from the reference for most types presented.

Example 8-1

> Suppose that 20 items are placed on test under simulated operating conditions for a maximum test time of 200 hours, and that 5 items failed at times 5, 30, 40, 55 and 200 hours, respectively. Assume that no prior information is available concerning the time to failure. It is desired to estimate the reliability of an item obtained under like conditions or manufacture and design.

An estimate of reliability is obtained from the distribution-free procedure given in the first row of Table 8-1.

$$\hat{R} = \frac{15}{20} = 0.75 \ .$$

Also using procedures given in Ref. 8-1, a 95 percent one-sided lower confidence interval estimate* of $\hat{R}_L = 0.545$ is obtained. Thus with 95 percent confidence the probability of survival for 200 hours exceeds 0.54.

Example 8-2

> Suppose that 10 items are placed on test under intended operating conditions until the 5th failure occurs and the failure times in hours are recorded as follows:
>
> $t_1 = 21,\ t_2 = 29,\ t_3 = 59,\ t_4 = 70,\ t_5 = 386$ hours.
>
> Assume that an exponential failure time distribution is applicable and suppose that it is desired to obtain an estimate of the mean-time-to-failure θ and give a one-sided lower 90 percent confidence limit for θ.

Using the procedures in the third row of Table 8-1, the total test time is

$$T = 565 + 5(386) = 2495 \text{ hours}$$

and the estimate $\hat{\theta}$ of the mean-time-to-failure is

$$\hat{\theta} = 2495/5 = 499 \text{ hours.}$$

Further, using procedures in Ref. 8-1, a one-sided 90 percent confidence interval $\theta_L = 312$ hours is obtained. Thus with 90 percent confidence, it is inferred that the mean-time-to-failure exceeds 312 hours.

In many practical problems one desires a statement such as "with 90 percent confidence 95 percent of the life times of individual items selected from a given

---

* Confidence interval estimation is also discussed in Sec. A.6 of the Appendix to this report.

population will exceed 200 hours, assuming this to be the mission time." Assuming an exponential distribution and applying other procedures given in Ref. 8-1, one obtains a one-sided lower confidence limit for the probability of survival $R_L = 0.527$. Hence, since only 52.7 percent of the items would survive 200 hours (under the assumptions stated), one would question the use of the item for such a mission.

<u>Example 8-3</u>

Suppose that ten items are tested under specified operation conditions until all fail. Assume that a Weibull distribution applies with the location parameter equal to zero. Suppose that it is desired to estimate the scale and shape parameters. The observed failure times in hours are as follows:

260, 600, 630, 820, 930, 950, 1270, 1590, and 1760 hours.

Using a graphical method based on the order statistics as given in Ref. 8-9, the failure times are plotted versus the expected values of the Weibull order statistics for m = n = 10 as obtained from Table 2 in the reference. Figure 8-3 clearly indicates that a shape parameter K = 2 yields a much better fit to the data than a shape parameter K = 1 (exponential case), hence K = 2 is assumed. (In this example the data were generated from a Weibull distribution having K = 2.) This slope of a fitted line yields an estimate of the Weibull scale parameter θ. Thus $\hat{\theta}$ is near 1,000 depending upon the fitting procedure used. (In this example $\theta = 1000$ was used in the Monte Carlo simulation.)

The above examples highlight the use of nonsequential life testing procedures for direct estimation of characteristics of life. Another use of considerable utility is the determination of acceptability of a product or design. Performing accept-reject decisions in qualification testing, reliability testing, and acceptance sampling are popular applications. The principals and procedures for these are well described in Ref. 8-19 from a very practical viewpoint. References 8-20 (DoD Handbook H-108) and 8-21 (MIL-STD-690A) also document a number of life testing plans and contain the OC curves for many combinations of mean life, confidence and risks and provide the information for readily determining the required test time (or number of failures) and number of items required. Both replacement and nonreplacement plans are included. Reference 8-20 is especially good for it contains a clear explanation of all aspects of the tests.

The basic concepts associated with sampling plans are summarized in Sec. A.8 of the Appendix. Two concepts, producer's risk and consumer's risk, are of particular interest and are defined as follows:

(1) Producer's risk, α, is the probability of rejecting a product with acceptable mean life. A product having acceptable mean life will thus be accepted $1-\alpha$ fraction of the time.

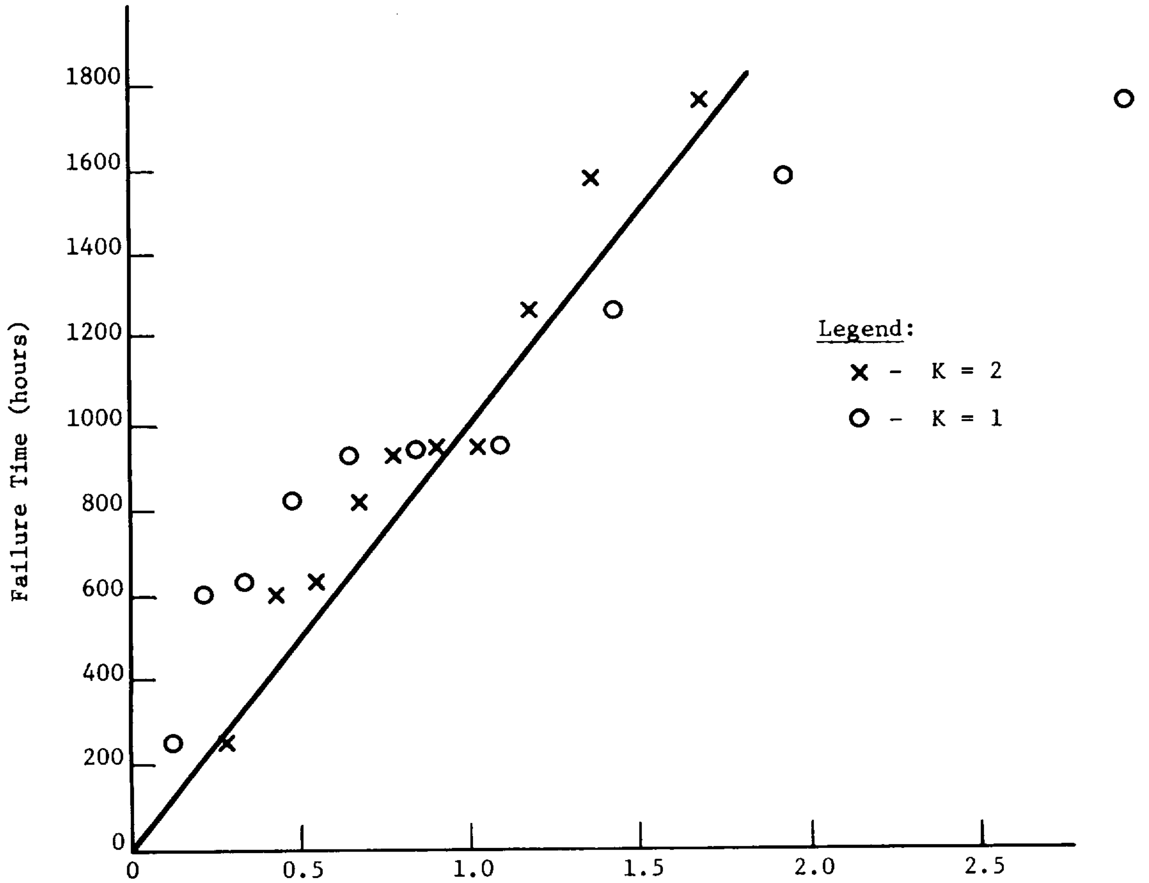


Figure 8-3. Graphical Estimation of Weibull Shape Parameter K and Scale Parameter θ

(2) Consumer's risk, β, is the probability of accepting a product with unacceptable mean life. A product having unacceptable mean life will thus be rejected 1-β fraction of the time.

An example illustrates the use of standard sampling plans.

Example 8-4

A design goal for a particular equipment is "a mean life of 2,000 hrs." The customer specifies acceptance sampling to be designed to accept 95% of the time production lots having an actual mean life of 2,000 hrs. but reject 90% of the time the lots having a mean life of 200 hrs. Further, testing without replacement is specified. The manufacturer is allowed the liberty to select the method and design the test according to standard sampling plans based on the exponential distribution for failure times.

In this case assume that scheduling does not allow for more than 200 hrs. of test time. A time-truncated procedure conforming to this constraint on test time is considered to determine the sample size and the number of failures for rejecting the lot. Using Ref. 8-20 (DoD Handbook H-108) a possible design is illustrated.

The pertinent data for designing the test is

T = test duration = 200 hrs.,

θ_0 = acceptable mean life = 2,000 hrs.,

θ_1 = unacceptable mean life = 200 hrs.,

α = producer's risk = 0.05, and

β = consumer's risk = 0.10.

The values of α and β follow from the earlier definitions and the problem statement.

With the above data the appropriate plan is located in Table 2C-3 of Ref. 8-20 and specifies a sample size, $n = 9$ and a number of failures for rejecting, $r = 3$. Thus nine items are tested and the lot rejected if three failures occur before 200 hrs. duration and accepted if less than three failures occur after 200 hrs. of testing.

Note that the design above did not employ OC curves even though it is possible to do so. OC curves have their greatest values in evaluating trade-offs between test approach, sample size, test time, etc. For example, a shorter duration test with a larger sample will accomplish the same objective.

Most life test designs described in the literature are based on the assumption that failures are distributed according to the exponential failure time distribution (i.e., a constant hazard rate*) as described by the third and fourth types listed in Table 8-2. A major reason for this is that experience has shown that many items conform to this assumption. For example, empirical justification is given in Ref. 8-19 for two systems. This does not say that it holds for all equipment however. Too often, however, the reason turns out to be simply because the procedures are much simpler than those required for other distribution assumptions with the result that gross errors are often made.

Sequential Life Testing

Sequential life testing is a procedure developed primarily to determine acceptability of a product or design more efficiently than the nonsequential procedures. When products are very good or very bad the procedure requires considerably less total test time, (i.e., the total for all items tested) that either of the nonsequential procedures. However, when a lot of items is marginal (not exceptionally good nor bad) the test could run on indefinitely without producing decisive results. For this reason provision is usually made to truncate the test on the basis of time,

* There is often confusion in the literature that a constant hazard and random failures are synonymous. Clarification of this presented in Sec. A.9 of the Appendix.

number of failures, or both. Such truncated tests are called truncated sequential life tests.

In sequential life testing n items are placed on test and the test continued until a decision to accept or reject a given hypothesis can be made on the basis of accumulated results. A sequential analysis is thus required along with the test to make a decision. This consists of generating a decision function, usually total operating time of all items tested versus the number of failures and observing whether the accept or reject bounds are crossed. A simple illustration of this was presented in Fig. 7-1; more detailed illustration is presented below by example.

<u>Example 8-5</u>

A truncated sequential life test procedure is considered for the same problem posed in Example 8-4.

Again we use the standard procedures in Ref. 8-20 (DoD Handbook H-108) with the same basic design data following Example 8-5. Using either Table 2A-1 or Table 2A-2 (this latter gives OC curves) of Ref. 8-20 for $\theta_1/\theta_0 = 1/10$, a specific plan coded as B-2 is specified. Using this designation one resorts to another tabulation in Ref. 8-20, Table 2D-1 to obtain the test design parameters. These are:

r_0 = the minimum sample size = 6

h_0/θ_0 = ratio of accept line intercept to the acceptable mean life = 0.2254

h_1/θ_0 = ratio of reject line intercept to the unacceptable mean life = 0.2894

s/θ_0 = ratio of decision line slopes to the acceptable mean life = 0.2400

Reference 8-20 also instructs the accept-reject criteria to be computed as follows:

(1) For acceptance, the acceptance line is

$$\begin{aligned} V_a &= h_0 + sk \\ &= (h_0/\theta_0)\theta_0 + (s/\theta_0)\theta_0 \cdot k \end{aligned}$$

and the truncation on total operating time is $sr_0 = (s/\theta_0)\theta_0 \cdot r_0$.

(2) For rejection, the rejection line is

$$\begin{aligned} V_r &= h_1 + sk \\ &= (h_1/\theta_0)\theta_0 + (s/\theta_0)\theta_0 \cdot k \end{aligned}$$

and the truncation on total number of failures is r_0.

Further, in testing without replacement the decision function or the total operating time of all items on test is

$$T_t = \sum_{i=1}^{k} t_i + (n-k)t$$

where n is the sample size, k the number of failures, t_i the time of the i-th failure and t is the time from the start of the test. Substitution of the test design parameters yields the design illustrated in Fig. 8-4.

Note that n does not have to be specified to design the test. This again allows for trade-off between test duration and sample size. A larger sample size for example causes the total test time to be accumulated faster (but also the failures). The dashed line in Fig. 8-4 represents a test of eight items where the six failures leading to the reject decision occur at test times of 30, 62, 165, 241, 385, and 794 hrs. from the start of the test.

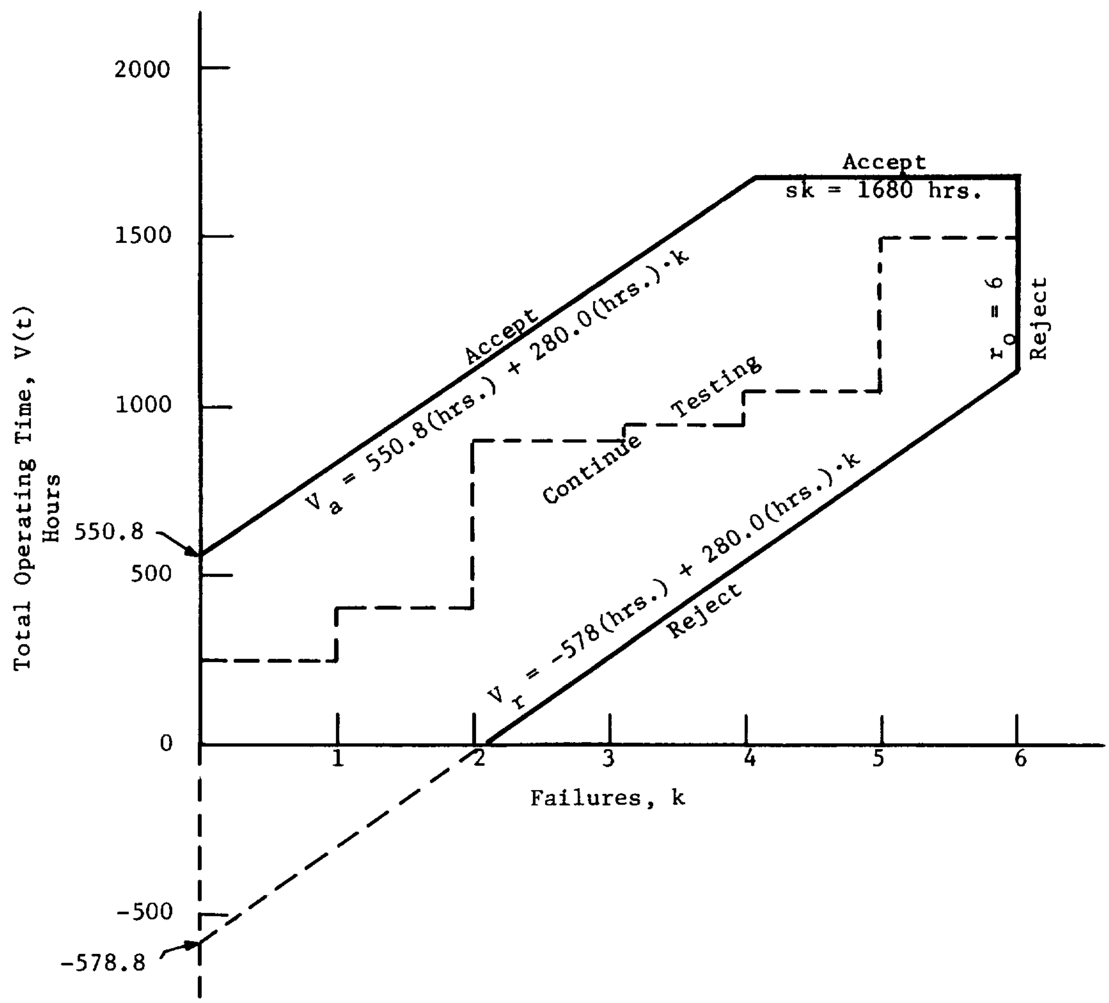


Figure 8-4. Example of Truncated Sequential Life Test

The above illustration is a straight-forward application of the standard procedures of Ref. 8-20. More detailed practical discussion on the development of sequential life test design with more examples is available in Ref. 8-19 where the development parallels that of the two nonsequential approaches discussed above.

It is especially noted that the procedures described in these references are all based on the assumption that failure times are exponentially distributed. Theoretically procedures can be based on other distributions such as the Weibull or log-normal [Ref. 8-22]. These, in general, lead to more complicated designs and have yet to be implemented on a significant scale as an improvement.

Further Perspective on Life Measurements

Of the three basic approaches to life testing, viz., failure-truncated, time-truncated, and sequential, which does one choose for his application? There are advantages and disadvantages for each and the selection should be tailored to the particular problem at hand.

For the nonsequential approaches most test requirements can be fixed in advance which can be an important factor in planning. But of these two the failure-truncated generally requires a longer test duration. For estimating life the nonsequential approaches are more suitable, but for merely judging acceptability of products the truncated sequential approach is most likely to be favored.

In making such comparisons care should be taken that they are evaluated on an equal footing. Precision of results, assumptions, mean life, replacement policy, etc. should all be equivalent for the approaches considered. Some good discussion including both qualitative and quantitative viewpoints on comparing approaches is presented in Ref. 8-19.

An even more basic question to be first answered is whether life testing itself need be used. It is generally more expensive to test in this manner; however, the penalties for not testing this way must be considered also. Could, for example, historical data from similar items provide enough (engineering) confidence? Would burn-in screening be a more feasible approach? Often the assumptions that have to be made to make the accompanying design and analysis tractable cause the results to be quite far removed from the real problem to be solved. Remember that statistical confidence when based on assumptions does not necessarily mean that you have the same engineering confidence.

Two trends in technology have tended to reduce the relative significance of life measurements. These are (1) the design and fabrication of complex, expensive, several-of-a-kind items, and (2) the development of very high reliability components such as integrated circuits.

The first of these introduces the problem of extremely small samples. It is too expensive to build a large number of satellites, for example, as a sample for life measurements. The logical alternatives to this introduce the concept of integrated testing* in which the tests at all levels from parts to the system are carefull designed to achieve maximum confidence in the adequacy of life.

The problem with measuring the life of extremely long-life components has led to greater interest in accelerated testing, reliability screening, reliability physics, reliability growth computations, Bayesian approaches, and testing larger samples for relatively short periods to obtain equivalent operating hours. Some of these are considered elsewhere in this report; the latter approach introduces special considerations treated below.

As already noted several times, many estimates and sampling plans for life measurements are based on the assumption that failure times are exponentially distributed. One should always be reasonably confident of the validity of this assumpti since actual distributions other than this can lead to unsatisfactory conclusions. For example, there is often reference to "equivalent operating hours" when short duration tests are conducted on a large number of items to compute a failure rate for items that are to be applied for longer operating periods. But this has meaning only when the hazard rate is constant throughout the long duration. If, for example, the true hazard rate is decreasing then the results of the short duration are pessimistic for longer periods of application. On the other hand, an actual increasing hazard rate can give an underestimate of failure rate over the longer period.

There has been no direct reference to application of life measurement techniques to fatigue testing. One basic difference is that failure is due mainly to the number of cycles of stress rather than duration. Ref. 8-23 is a good source for identifying practical problems associated with measuring fatigue life. A good survey discussing concepts and approaches is provided by Ref. 8-24 which also identifies other good sources.

8.2 Investigation of Failure Modes and Mechanisms

This category extends the concepts discussed in Sec. 6.2** to account for cumulative damage exceeding endurance as the predominant cause of failure. As indicated in Fig. 8-1, a failure-truncated test is the obvious approach to inducing failure when items simply have to be failed for this purpose. Failure investigations often accompany the observation or occurrence of failure in the normal course of some other

*Integrated testing is also discussed in Sec. 13.

**Sec. 6.2 has the same title but pertains only to time-independent failures caused by stress exceeding strength.

effort such as operational use, field testing,life testing, performance testing, etc. It is important in such cases to be sure which cause of failure you are investigating.

There are frequent attempts to induce failure of long-life items in short time by accelerating the aging process. One must be careful in this approach that the acceleration does not induce a different mode of failure that would be obtained under conditions of normal operation. Problems with accelerated testing are discussed in Sec. 11.

8.3 Burn-in, Bake-in, Run-in, etc.

Certain items such as electronic parts and equipment are often purposely aged for a particular duration prior to their application to induce failure in those items which have the shortest life. Such tests are called burn-in or bake-in tests. On the other hand an item such as an aircraft engine is sometimes purposely aged prior to application, not to induce failure but to improve the item. Tests of this type are called run-in or "break-in" tests. Other tests may "operate-in", "vibrate-in", or "pressurize-in" but all have the common goals of using aging purposely to results eventually in a product (the tested items themselves or other items assembled from them) that is more reliable than if the testing were not done. Whether a reliability improvement is in fact achieved or not depends on certain characteristics of the tested items.

Burn-in (or Bake-in)*

The popular "bathtub" curve often used to illustrate the hazard function** of an item is shown in Fig. 8-5. The infant mortality region is typically assumed to represent the behavior of items failing early due to manufacturing and material defects. Refs. 8-25 and 8-26 are often cited as containing adequate empirical evidence on the existence of this behavior in semiconductor devices. A later report on this behavior for integrated circuits is presented in Ref. 8-27. Empirically determined hazard functions exhibiting this property for other types of equipment are presented in Ref. 8-28.

Burn-in consists of using this property of a decreasing hazard function to improve reliability. Items can be placed on test subject to conditions which cause aging and during the infant mortality region the failure rate of the population of items not failed is decreasing.

---

*Burn-in infers aging while operating; bake-in, aging while nonoperating.

**The hazard function for an item is the conditional probability of failure in the interval (t, t+dt) given that failure has not occurred prior to time t. See also Sec. A.9 of the Appendix.

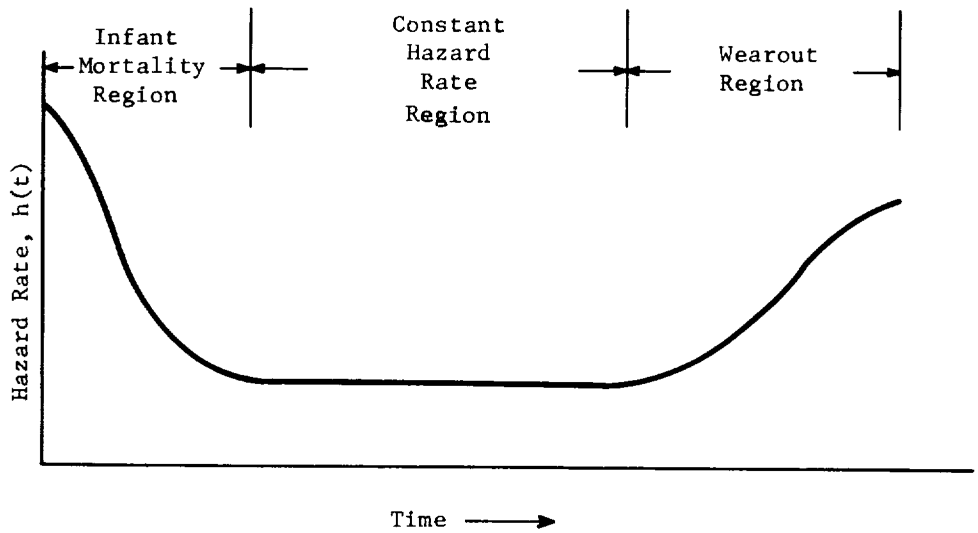


Figure 8-5. Popular Illustration of a Hazard Function

If a period of lower constant failure rate as illustrated in Fig. 8-5 does in fact exist, then a population operating during this period is obviously more reliable than if the infant mortality region were included in the period of application. In this case one would obviously conduct burn-in for the duration of the infant mortality and use the remaining nonfailed items for application.

The real situation, however, is not always this simple. First, any single hazard function represents only a particular environment; another environment may give a quite different shape. Also, a constant hazard rate is actually more rare than a nonconstant one. The argument, particularly for many reliability calculations, is usually that it is near enough constant to assume it so. Watson and Wells[Ref.8-29] considered the case where the failures are distributed according to the Weibull distribution and showed that reliability improves by eliminating those items with short lives. This has been generalized by others (see, e.g., Ref. 8-30) to show that the only requirement is a decreasing hazard rate.

The hazard function is typically constructed from the results of tests on a number of items. There are often misunderstandings when extrapolating to a single item. For a single item damage is done as time progresses; therefore, an aged item is not as good as it was when it was new. If the hazard rate is decreasing continuously we have the apparent anomaly that, even though the item itself is being degraded, as long as it has not failed it is more likely to last longer than an item which has not been operated. The explanation of course is that we do not know what the starting endurance of each item is. If we did, the endurance would no longer be a random variable and we would know the life of each part to begin with. Then the cumulative hazard function would be either 0 or 1 and the hazard rate would be a spike at the changeover. This illustrates the difference between considering an individual item and making probability statements which, even though made about an individual item, are effectively relative frequency observation about the population to which the item belongs.

A major problem is concerned with how long a burn-in test should last. For a constant hazard rate following infant mortality, one would obviously run the test long enough to encompass the infant mortality if the application duration is less than the period of constant hazard rate. For other application durations and hazard function shapes the general problem becomes one of trade-off between burn-in time and reliability of the application. For a mission starting at time t_i and extending over the interval (t_i, t_f), one would want to locate this interval on the time axis of the hazard function to minimize the area under the hazard function if he desired to maximize reliability. This can be easily seen from the general formula,

$$R = e^{\int_{t_1}^{t_2} h(t)dt} ,$$

for reliability since the maximum R is obtained by the minimum value of

$$\int_{t_1}^{t_2} h(t)\, dt.$$

The problem has been formulated by others as "how long to burn-in the components to achieve a specified reliability or mean life?" A recent paper by Lawrence [Ref.8-31] treats this with the only assumption that there is a period of decreasing hazard rate and he derives upper and lower bounds on the burn-in time to achieve a specified reliability.

Another discussion of this problem but extended to include the practical problem of selecting the test conditions is presented in Ref. 8-32*. Basically, that reference proposes that some of each lot be used for accelerated life testing to guide the choice of burn-in duration and conditions and then applying the chosen burn-in procedure to the remainder of the lot. From the discussion of the proposed procedure it appears that this would be forbiddingly expensive; however, the paper has value in the comprehensiveness of the many aspects of the burn-in problem discussed.

---

*The term, run-in, used in the title of that paper is synymous with burn-in; it should not be confused with the designation of run-in used in this section.

Suffice it to say that there is nothing to be gained by purposely conducting a burn-in test when it is known that an item possesses an increasing hazard rate during the burn-in period. There are cases, however, where burn-in of some items with increasing failure rate are unavoidable. An equipment fabricated from a number of different types of components each having a characteristics hazard function often has to be tested or "tuned up". Some of the components may thus indeed be operating in a period of increasing hazard rate*.

Generally, in such cases one makes the best selection of components and burn-in philosophy possible while considering that equipment tests are part of the mission for the components. The conglomeration of all components and their interactions actually gives rise to a hazard function for the equipment. It is sometimes possible that with past experience on similar equipments or tests on several equipments that some characteristics of the hazard rate may be determined.

Quite often, however, only one or, at most, several may exist and relatively little known about the hazard function. This is quite typical, for example, of a spacecraft or satellite. Yet testing of flight items has virtually proved to be a necessity. Evidence and discussion of this is presented in Ref. 8-33 for preflight environmental simulation as a flight acceptance test procedure. One does not normally refer to such tests as burn-in tests even though they are basically no different. An environmental test of an operational satellite, for example, has as a major purpose inducing failures to uncover workmanship errors and material defects. Test conditions are typically chosen to be considerably less severe than those for a flight qualification with nonflight hardware. Still, some cumulative damage may be occurring in the flight acceptance procedures even if no failures occur and one would want to wisely choose the duration and environments such that the success of the mission is not jeopardized by the test if it passes.

As noted in Fig. 8-1, both the time-truncated and the sequential test approaches are applicable to burn-in testing. However, all implementations of burn-in noted to date have been by the time-truncated approach; a test duration is specified prior to placing items on test for burn-in and the test terminated where this time has elapsed. It is surprising that there has been little consideration of sequential decision procedures for burn-in. Such an approach might, for example, involve a continual estimation of mean-time-between-failure (MTBF) and the test terminated at any time that this exceeds a given value. Such a procedure is described for repairable equipment in Ref. 8-35 and the problem is formulated from a dynamic programming point of view.

---

*This problem is closely related to the reliability growth problem. Ref. 8-34 gives a good introduction to reliability growth concepts.

Ref. 8-35 also gives a good discussion of burn-in testing concepts from the mathematical point of view and itself cites several good references for further study.

Most of the above discussion has focused on the problem of the burn-in duration. Two additional problems are the parameters to be measured and the severity of the aging conditions.

In most burn-in tests there is some type of performance check following the test. A failure may thus be defined simply by an out-of-tolerance or total destruction. In some cases there may be attempts to select parameters that can provide a pre-indication of later failure. Instability, noise factor, and rate of change are parameters types often employed. Such preindicator measurements are only as good as your ability to correlate the parameter behavior with underlying aging mechanisms and later performance. Such problems with extrapolating performance is discussed in Sec. 8.4.

For the test conditions the normal operating conditions cannot always be simulated nor is it necessarily desirable that they be. The basic rule is to select those which will induce the cumulative damage process of real interest. This of course is not always known and other tests may be needed to achieve this. For example, tests of two separate groups of similar items, one in a high temperature environment and the other in a low pressure environment, may yield quite distinct hazard functions; say the one for the low pressure exhibits no significant infant mortality characteristic. Discounting possible interaction effects, it may thus be desired to use only a high temperature environment for burn-in.

With the increased interest in accelerated testing, it is only natural that it be applied to burn-in. The inherent problems are basically the same as those when accelerating aging for other aging tests. It is discussed in more depth in Sec. 11.

Run-In

This use of testing is introduced only with brief discussion to illustrate its role and perspective. Everyone knows that special care to prevent overheating is required for breaking in a new automobile engine. This is typical of a run-in test. The concept applies mainly to items having moving parts with the intent to perform the initial aging under controlled conditions such as speed, lubrication, and operating temperature to improve the individual item. An analogy with electronics is operating equipment at higher than normal temperature to drive off moisture.

In a run-in test, cumulative damage is present but because the test is purposely conducted to improve the item, the damage must be beneficial (i.e., negative damage). For example, piston rings and bearings are wearing but in a manner to seat them properly.

A run-in test may employ a time-truncated approach or a sequential approach. In the case of an automobile a run-in period and speed conditions may be recommended;

the actual duration and conditions are at the customer's discretion. For more formal run-in procedures the duration may be specified prior to the test and based on life and performance test results of similar items. A good illustration of this draws from Ex. 7-5 in Sec. 7 where a long-term test of a gyro indicated performance improving throughout the first 100 hrs. of operation. If this gyro were representative of a population of similar units, then a specified run-in period for all gyros of this type would be in order.

A sequential decision procedure for terminating run-in tests may typically involve testing until some performance level or criteria such as minimum drift rate, horsepower, or operating temperature is reached.

8.4 Measuring Aging-Dependent Performance Characteristics

When aging is present and affects performance, a measurement of performance generally has explicit measuring only for the period of observation. Yet there are often attempts to extrapolate performance to later periods. A circuit designer, for example, may specify a 1% purchase tolerance for a resistor in production to obtain reasonable (engineering) confidence that it will be within 3% tolerance after 10,000 hours of operation. Even a 100 hour performance test of the resistor may provide little verification of this.

Often such extrapolations can be confidently made on the basis of experience. Even tests may have been run to establish the correlation of performance between early and later life. As described in Sec. 8.3, the ability to extrapolate can be very useful in burn-in testing when looking for pre-indicators of failure.

The concept of a performance test on a single item to measure aging-dependent properties was introduced in Sec. 7.4; an example illustrated gyro performance changing over time. There are often attempts to extrapolate the behavior of a single item in such a test to describe performance of another similar item or to a population of several such items. This is typically a necessity with large expensive items such as satellites and launch vehicles.

In some cases one may have reasonable (engineering) confidence on the basis of theory or past experience that such an extrapolation is meaningful. As is more often the case the major "trouble spots" or sources of degradation can be traced back to a few components for which test results or more than one item are available. The gyro test described in Ex. 7-5, for example, might represent one realization of observed behavior for a larger population.

With a population of items tested in this manner it may then be possible to view the performance degradation in terms of a time-changing distribution of performance parameters. For example, Fig. 8-6 illustrates the observed time behavior of a

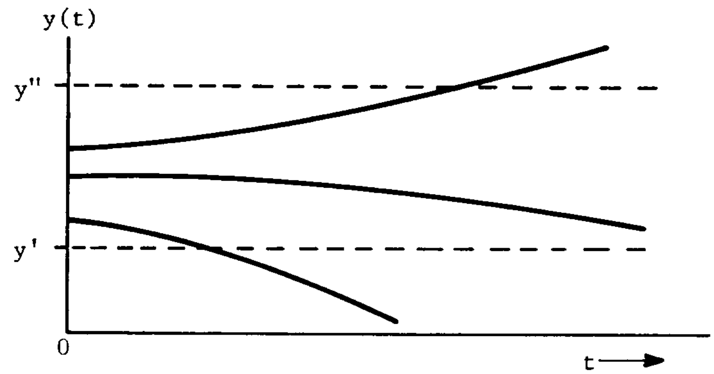


Figure 8-6. Examples of Monotonic Drift

performance parameter y for these items. In this case, we consider simple monotonic drift behavior where the parameter values may start within normal tolerance defined by y' and y" in Fig. 8-6 and some remaining within tolerance while others drift out of tolerance.

Extending this to a larger population, it is possible to construct histograms at various times to depict the time-changing nature. Numerous examples of this appear in electronic part manufacturer's literature. An illustration of this for the case of continuous distributions is shown in Fig. 8-7 which shows that as time progresses the frequency function f(y;t)* broadens with the proportion of items having parameters outside of tolerance limits increasing. As described in more detail in Vol. IV - Prediction of this series, this property is occasionally used to describe a drift reliability. This requires, however, that the tolerances or bounds chosen be able to represent failure.

When the observed behavior of a parameter is nonmonotonic in time it may not be easy from the observations to directly discern whether performance is degrading. One may then specify another quantity dependent upon the parameter behavior to more appropriately describe performance changes. Typically for a noise process the time average, variance, or certain power spectral density characteristics may be derived from the observations. When the appropriate ones for describing degradation are chosen they may be treated as the monotonic behavior described above. More discussion on this is presented also in Vol. IV - Prediction of this series.

An excellent example of performance degradation measurements and use of the results to predict reliability is given in Ref. 8-36. Performance measurements of

---

*Whereas the frequency function for the distribution of the variable y is usually written as f(y); it is represented here by f(y;t) to indicate t as a parameter.

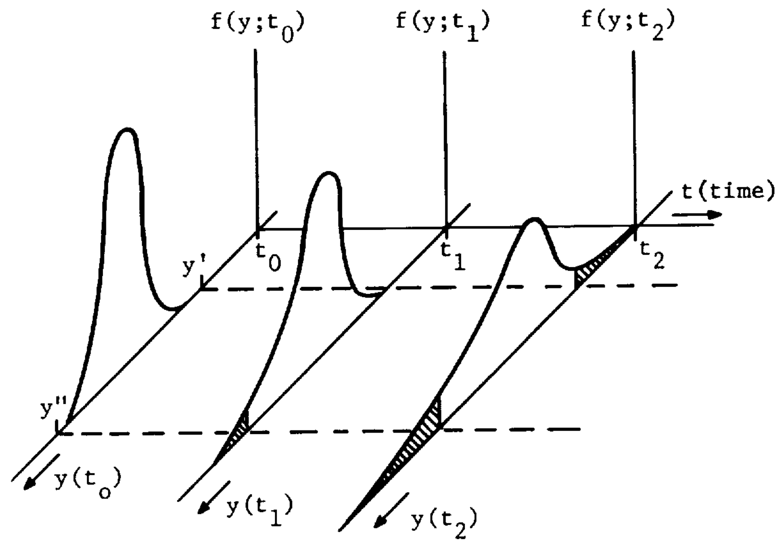


Figure 8-7. Illustration of Degrading Performance for a Population

the type described are useful for obtaining design and component application information.

References

8-1. Epstein, Benjamin: Estimation from Life Test Data. Technometrics, vol. 2, no. 4, November 1960, pp. 447-454.

8-2. Barlow, Richard E.; and Gupta, Shanti S.: Distribution-Free Life Test Sampling Plans. Technometrics, vol. 8, no. 4, November 1966, pp. 591-613.

8-3. Cohen, A. Clifford (Jr.): Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and on Censored Samples. Technometrics, vol. 7, no. 4, November 1965, pp. 579-588.

8-4. Menon, M. V.: Estimation of the Shape and Scale Parameters of the Weibull Distribution. Technometrics, vol. 5, no. 2, May 1965, pp. 175-182.

8-5. Jaech, J. L.: Estimation of Weibull Distribution Shape Parameter When no More Than Two Failures Occur Per Lot. Technometrics, vol. 6, no. 4, November 1964, pp. 415-422.

8-6. Nelson, Lloyd S.: Weibull Probability Paper. Industrial Quality Control, March 1967, pp. 452-453.

8-7. Epstein, Benjamin; and Sobel, Milton: Life Testing. J. Amer. Statis. Assoc., vol. 48, no. 263, September 1953, pp. 486-502.

8-8. Stacy, E. W.; and Mihram, G. A.: Parameter Estimation for a Generalized Gamma Distribution. Technometrics, vol. 7, no. 3, pp. 349-357.

References (Continued)

8-9. Harter, H. Leon: Expected Values of Exponential, Weibull and Gamma Order Statistics. Aerospace Research Laboratories, AD 436 763, 106p.

8-10. Harter, H. Leon; and Moore, Albert H.: Point and Interval Estimators, Based on m-order Statistics, for the Scale Parameter of a Weibull Population with Known Shape Parameter. Technometrics, vol. 7, no. 3, August 1965, pp. 405-422.

8-11. Gupta, S. S.: Order Statistics from the Gamma Distribution. Technometrics, vol. 2, no. 2, May 1960, pp. 243-262.

8-12. Wilk, M. B.; Gnanadesikan, R.; and Huyett, M. J.: Probability Plots for the Gamma Distribution. Technometrics, vol. 4, no. 1, February 1962, pp. 1-15.

8-13. Sarndal, Carl Erik: Estimation of the Parameters of the Gamma Distribution by Sample Quantiles. Technometrics, vol. 6, no. 4, November 1965, pp. 405-414.

8-14. Harter, H. Leon; and Moore, Albert H.: A Note on Estimation from a Type 1 Extreme-Value Distribution. Technometrics, vol. 9, no. 2, May 1967, pp. 325-331.

8-15. Harter, H. Leon; and Moore, Albert H.: Maximum-Likelihood Estimation of the Parameters of Gamma and Weibull Populations from Complete and From Censored Samples. Technometrics, vol. 7, no. 4, November 1965, pp. 639-648.

8-16. Gupta, Shanti A.: Life Test Sampling Plans for Normal and Lognormal Distributions. Technometrics, vol. 4, no. 2, May 1962, pp. 151-160.

8-17. Box, G. E. P.; and Hill, W. J.: Discrimination Among Mechanistic Models. Technometrics, vol. 9, no. 1, February 1967, pp. 57-72.

8-18. Harter, H. Leon: Maximum-Likelihood Estimation of the Parameters of a Four-Parameter Generalized Gamma Population from Complete and Censored Samples. Technometrics, vol. 9, no. 1, February 1967, pp. 159-165.

8-19. Myers, R. H.: Reliability Testing. Chapter IX of Reliability Engineering for Electronic Systems, edited by R. H. Myers, K. L. Wong, and H. M. Gordy, John Wiley and Sons, Inc., New York, 1964.

8-20. Anon.: Sampling Procedures and Tables for Life and Reliability Testing (Based on exponential distribution). Quality Control and Reliability Handbook H-108, Department of Defense, April 1960.

8-21. Anon.: The Military Standard Failure Rate Sampling Plans and Procedures. MIL-STD-690A, Department of Defense, July 1965.

8-22. Arorian, L. A.; and Robison, D. E.: Sequential Life Tests for the Exponential Distribution with Changing Parameter. Technometrics, vol. 8, no. 2, 1966, p. 217.

References (Continued)

8-23. American Society for Testing and Materials: Special Technical Publications
91 Manual on Fatigue Testing, 1949
92A Tentative Guide on Fatigue Tests and the Statistical Analysis of Fatigue Data, 1963.

8-24. Wood, W. A.: Experimental Approach to Basic Study of Fatigue. Technical Report No. 24, (Contract No. NONR 266(91), DDC, AD 478 481, Department of Civil Engr. and Engr. Mech., Columbia University, New York, August 1965.

8-25. Von Alven, W. H.; and Blakemore, G. J.: Reliability of Semiconductor Devices. Final Report ARINC Research Corp., Publication No. 144-6-270, 1961.

8-26. Von Alven, W. H.: Semiconductor Reliability, Vol. 2. Engineering Publishers, 1962.

8-27. Holden, Phil: Proving Integrated Circuits Reliability. Proceedings, 1966 Symposium on Reliability, San Francisco, Calif., January 1966.

8-28. Mast, L. T.: Impact of Equipment Life Characteristics on Missile Test Planning, Memorandum RM-4102-PR (DDC, AD 602 646), The Rand Corp., Santa Monica, Calif., 1964.

8-29. Watson, G. S.; and Wells, W. T.: On the Possibility of Improving the Mean Useful Life of Items by Eliminating Those with Short Lives. Technometrics, vol. 3, no. 2, May 1961.

8-30. Barlow, R. E.; Proschan, Frank; and Scheuer, E. M.: Statistical Estimation Procedures for the "Burn-In" Process. NASA CR T8131 (STAR, N66-37212) The Rand Corp., Santa Monica, Calif., Sept. 1966.

8-31. Lawrence, Michael J.: An Investigation of the Burn-In Problem. Technometrics, vol. 8, no. 1, February 1966.

8-32. Norris, R. Husmer: "Run-In" or "Burn-In" of Electronic Parts--A Comprehensive, Quantitative Basis for Choice of Temperature, Stresses, and Duration. Proceedings, Ninth Symposium on Reliability and Quality Control, San Francisco, Calif., Jan. 1963.

8-33. New, John C.; and Timmins, A. R.: Effectiveness of Environment-Simulation Testing for Spacecraft. NASA TN D-4009, June 1967.

8-34. Bresenham, J. E.: Reliability Growth Models. Tech. Report No. 74 (DDC, AD 605 993), Dept. of Statistics, Stanford University, Stanford, Calif., August 1964.

8-35. Cozzolino, J. M., Jr.: The Optimal Burn-In Testing of Repairable Equipment. Tech. Report No. 23 (DDC, AD 642 105), Mass. Inst. of Technology, Oct. 1966.

8-36. Britt, C. L.: Research in the Functional Analysis of Systems. Annual Conference Transactions, ASQC, 1965.

8-37. Falls, L. W.: NASA Technical Memorandum 53422, March 30, 1966.

9. Nondestructive Testing

In the broadest sense, nondestructive testing (NDT) pertains to any testing approach which does not impair the usefulness of an item. The discussion of NDT in this section, however, focuses on the special technology such as infrared scanning and X-ray radiography which is employed to

(1) circumvent destruction of items in the measurement of properties which would normally require destruction if measured by conventional techniques and/or

(2) permit certain measurements to be made more rapidly and conveniently than conventional techniques allow.

It is the purpose of this section to give an introductory "broad brush" review of this very extensive subject as perspective for its role in testing. In addition to cited references several sources for further reading are included at the end of this section.

The major areas of application and typical functions it can perform are illustrated in Table 9-1. In research and development NDT frequently serves as a valuable

Table 9-1

Applications, Functions, and Examples of NDT

Areas of Application	Function Performed	Examples
Research & Development	Evaluating materials, components and parts; comparing and evaluating fabrication and assembly techniques; data acquisition.	Measuring fatigue in metals, detecting cracks in welds, and non-bonds in bonded materials.
Process Control	Measuring process variables and providing control information	Radioisotope thickness gauging.
Quality Control	Detecting and locating anomalies in materials, defective parts, etc; detecting and locating fabrication and assembly defects; evaluating the production process.	Poor adhesive bonding, cracks in welds, contaminated transistors, non-uniform porosity in metals.
In-Service Evaluation	Detecting flaws, defects, wear and deterioration of items in field use without major disassembly.	Locating corrosion inside gas tanks, detecting moisture in bonded wing structures on aircraft, etc.

measurement and evaluation tool for measuring special properties of materials. In process control a useful control function can often be generated from measurements with NDT devices monitoring certain stages of the production process. The greatest benefit of NDT to date is in quality control where items and materials are evaluated without destructive sampling. NDT is also employed to measure wear and deterioration in items which are in service. It is particularly convenient here because it can often accomplish its purpose without major disassembly of the item.

Most forms of energy have been harnessed to some NDT problem or investigated as a possible NDT method. Items are observed, smelled, felt, measured, exposed to X-ray, magnetized, vibrated, acoustically excited, or heated all in the name of NDT. No one form of energy nor any one NDT method is the answer to all or even a large portion of the nondestructive testing needs. Each technique has its limitations and the methods usually compliment rather than compete with one another. It is sometimes necessary to develop a special NDT method along with development of the item to which it is to be applied.

Most NDT methods do not measure a parameter or characteristic directly but measure some more easily observed phenomenon which can be correlated with the desired characteristic. For example, the uniformity of a material can be inferred by observing magnetic flux perturbations through it or ultrasonic energy reflections from it. On the other hand, there are methods, such as X-ray radiography which permit a more direct observation. Table 9-2 summarizes typical characteristics of the more popular and most widely applied NDT methods.

9.1 Optical Methods

Optical techniques utilize optical aids such as microscopes, magnifying glasses and interferometers, to detect the presence of surface flaws, anamolies, and malfunctions in materials and items. A permanent record of surface conditions or outward appearance can be obtained by photographic means. This method can provide excellent permanent records, but can only detect and record surface phenomena.

Microscopes can provide a maximum magnification on the order of 2000 with field of view and depth range decreasing with increasing magnification. Interferometer type microscopes offer depth measurements in the low micron region. Microscopy is greatly extended by the use of electron beam microscopes, although this technique is much more expensive and requires more operator skills and specialized interpretation of results. Optical microscopy and photography are commonly joined to produce photomicrographs--photographs taken through microscopes. Fiber optics technology can be utilized to observe and record information in otherwise inaccessible areas such as the inside of fuel tanks, inside completed wing structures, etc. Wide angle and long

Table 9-2

Characteristics of Popular NDT Methods

Method	Advantages	Disadvantages	Application	Will Detect
Optical	Applicable to almost any item; both dynamic and static measurements; versatile.	Can detect surface phenomena only; sometimes requires a delay (development of photographs);	Surfaces of materials and items; interior of vessels and compartments (fiber optics);	A wide range of surface flaws; visible damage, etc:
Radiography	Can detect hidden and internal defects; both static and dynamic measurements.	Requires expensive and complex equipment; sometimes presents a radiation safety hazard.	All materials not adversely affected by the incident radiation; moving mechinery; enclosed objects; internal characteristics	Flaws in totally enclosed components, machinery etc; defects in fast moving equipment
Thermal	Provides temperature profile of a surface or area during operation; rapid; very sensitive (IR)	Sometimes requires application and removal of special materials; permanent records are difficult to obtain (except with IR)	Surfaces of material and items not damaged by application of the coating; very small surfaces such as microcircuits (IR)	Small temperature gradients on surfaces.
Magnetics	Can locate flaws or defects in assembled equipment; rapid and simple to apply; sensitive.	Applicable to only magnetic materials.	Surface of metals; wires, tubes, etc;	Flaws and anomalies on surface or within magnetic materials.
Liquid Penetrants	Inexpensive and simple.	Requires application and removal of special materials; can detect surface flaws only; a development process precedes output information.	Surfaces of materials and items not damaged by the process; nonporous materials.	Surface cracks, flaws, and defects.
Ultrasonics	Penetrates deep; can detect flaw with access to only one side of material; sensitive	Requires a specially skilled operator; permanent records and readout difficulty.	Bonded structures and materials; all solids.	Disbonds, defects, cracks, and flaws.

range photography allow recording of more information than could be obtained by the naked eye. High speed photography can produce very useful records of motion and the dynamic characteristics of a material or item. Optical equipment is available from many vendors as off-the-shelf items.

9.2 Radiography

Radiography is one method for seeing beneath surfaces. The radiographic system consists of three major components: the radiation source, the radiation detector, and the material or item to be inspected. The basic arrangements of these components generally used in radiography are illustrated in Fig. 9-1. The arrangement in Fig. 9-1(a) is the more common of the two. In this penetrating radiation is allowed to pass through the object of interest onto a film or other detector that is sensitive to the radiation. The presence of flaws, anomalies, and foreign objects are revealed by the image or detector output. A less frequently used arrangement has the source and the detector on the same side of the material as shown in Fig. 9-1(b). In this method, radiation from the source passes through the detector and strikes the material causing scatter or secondary emissions which are then detected.

Both nuclear and atomic radiation are used in radiographic NDT. Some pertinent characteristics of these are summarized in Table 9-3.

X and gamma radiation with conventionally developed X-ray sensitive films as the detector are the most widely used of present techniques. These methods can detect defects which are on the order of 1% of the material thickness. Procedures

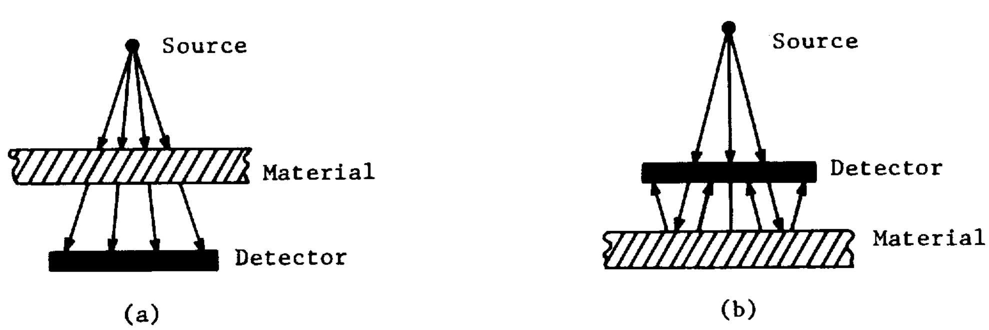


Figure 9-1. Basic Arrangements of Radiographic Measurement Components

Table 9-3

Characteristics of Radiographic NDT Methods

Radiation	Source	Material (Applications)	Detector
X-Rays	Conventional X-ray equipment	Locating foreign objects, flaws, and anomalies in parts and materials; observing machinery in operation.	X-ray sensitive films; color radiographic films; Polaroid process; Xeroradiography; fluoroscopy.
Electrons (Beta Particles)	Electron guns; secondary electrons emitted when X-rays enter sample; radioisotopes	Density measurements, thickness measurements, surface phenomena detection not distinguishable under visible light.	Photographic films; electrical detectors.
Neutrons	Fission reactors and special neutron sources.	Used in lieu of X-rays for heavy materials; for use with materials which absorb X-rays but not neutrons.	Photographic film sensitive to neutrons; neutron detectors.
Protons	Accelerators	Thickness and density measurements.	Proton detector
Gamma Rays	Radioisotopes	Used in same manner as X-rays.	Photographic films; gamma detectors.
Alpha Particles	Radioisotopes	Thickness measurements of very thin materials.	Alpha detectors with electrical readout.

have also been developed which use the faster and sometimes more economical Polaroid and Xeroradiographic processes. Color radiography has also been recently developed. This technique adds the dimensions of hue and saturation to that of brightness so that areas of opacity are easier to distinguish. Fluoroscopy can be used as the detecting phenomenon.

Electron radiation is generated by electron accelerating tubes (electron guns) and by X or gamma rays entering a material. The electron beam type radiation is used in conjunction with electron microscopes to map sub-surface phenomena. Resolutions in the submicron region are reported [Ref. 9-1]. Electron radiography also makes use of the source-detector-material arrangement shown in Fig. 9-1(b). An advantage of this method is that access to only one side of the material or item is necessary. Electron radiation can also be transmitted through the material to make such measurements as material density and thickness.

Neutrons have an advantage over X-rays and gamma rays in that they are absorbed differently by different materials. These differences can sometimes be exploited for better flaw detection discrimination. Most heavy materials do not absorb neutrons well thus thick sections can be investigated with shorter exposure times. Neutrons are very useful for examining materials (such as many plastics) which contain much hydrogen. A disadvantage is that neutrons are hard to record on films and a special process is necessary to detect them.

Both alpha and beta particles are used as thickness measuring gauges. By using a wide range of energies, thickness measurements using beta sources can be made from 1.5×10^{-5} inches of aluminum to 50 mils of steel. Reference 9-2 cites an application of alpha particles to measure 1% thickness changes in thin foils and paper.

Radiography is probably the most widely used of all NDT methods and has many more facets than are described here. Equipment for conducting such tests is readily available from many manufacturers. The topic is very thoroughly covered in Refs. 9-2 through 9-4. Specific applications are discussed in Refs. 9-1, 9-6, and 9-7.

9.3 Thermal Methods

The flow of heat through a material is altered by any discontinuities in the material. These discontinuities are reflected as variations in temperature at the surface of the item. The location and size of an anomaly can be determined by the temperature profile at the surface. Thermal methods are especially suited for evaluating bonds between two materials, i.e., for the detection of non-bonded areas.

The heat is either applied artifically or is generated in conjunction with operation of the item. For example, engine cylinders can be uniformly heated by filling with hot oil. A microcircuit in operation produces heat internally. The method of detecting and recording the surface temperature gradient varies from thermocouples to infrared scanning with various degrees of resolution and sensitivity. Table 9-4 outlines the characteristics of the techniques employed.

The frost test is a method widely used for testing the bond quality of cladded nuclear fuel elements. A chemical which has a frosty appearance and a given melting temperature is applied to the element and heat applied. A poor bond causes a change in appearance of the bond. The method can also be used on other materials.

Table 9-4

Characteristics of Thermal Methods

Method	Application	Detector	Detector Applied by:	Capabilities	Disadvantages
Frost Test	Cladding bond quality of nuclear fuel elements and other bonds.	Acenapthene or Diphenyl	Brush or spray	Can locate flaws with dimensions 0.1 $in^2 \times 0.2$ in.	A significant flaw can have low thermal resistance and go undetected.
Temperature Sensitive Phosphor	Evaluating metal-to-metal bonds, fusion bonds, etc.	Zinc-cadium sulfide phosphor in a plastic suspension viewed under ultraviolet light.	Brush or spray	Changes emissivity by 20% when temperature changes 2% over range of 40°F to 130°F.	Must be viewed under ultraviolet; emissivity change is reversible.
Tempilstik	Almost any surface.	Temperature sensitive crayons having calibrated melting points.	Marking or touching	Indicates a temperature within tolerance of $\pm$ 1% over a range of 113°F to 2000°F.	Indicates only one temperature per application; will not work at higher temperatures in a reducing environment.
Temperature Sensitive Paints	Almost any surface, i.e., metals, ceramics, stone, porcelain, plastics, wood and glass.	Paint changes color with temperature-change in permanent.	Brush or spray	104°F to 2912°F within $\pm$ 9°F; can locate flaws on order of 0.001 inches	Must dry 30 min. before use.
Infrared	Any surface emitting IR radiation, i.e., heated surface.	Photographic film and IR detector.	No contact; IR applied to detector or films by optical lens system.	Can detect temperature differences as low as 0.5°C; has a resolution as small as 0.0014 in.	The more sensitive methods require expensive equipment.
Temperature Probes	Temperature measurements of surfaces and bulk of most materials.	Thermometers, thermistors, thermocouples, and resistance thermometers.	Contact to material or medium to be measured.	Can measure temperatures between approximately -200°C to +2000°C within $\pm$ 1%.	Has low resolution for temperature profiling; lead wires conduct heat away from surfaces, etc., reducing true temperature.

The temperature profile of a surface can also be sensed by a coating of some phosphor suspended in a liquid. The phosphor changes emissivity to ultraviolet light as a function of temperature.

Tempilstiks are crayons made of a material which melts at a calibrated temperature. A specimen is marked with the appropriate Tempilstik before heat is applied. The mark melts at its calibrated temperature. There are various "tricks-of-the-trade" which extend the use of the method. Other similar Tempil products such as Tempilaq and temperature sensitive pellets are also available commercially.

Temperature sensitive paints are available which change color as a function of temperature. Some such paints change color as many as four times at four different temperature levels. The changes are permanent and provide a good permanent record. These paints are applicable over a range of 104°F to 2912°F and are accurate within $\pm$ 9°F. The paints can be used on almost any surface. Drawbacks are that the paint must dry for 30 minutes before use and must be removed after use. The surface to which it is applied must also be thoroughly clean before application. These paints are known as Thermocolor. Reference 9-2 describes these in more detail.

Methods are also available which make use of the infrared emission from heated objects. Infrared photography and photomicrography have been used for some time to record temperature profiles of surfaces. A newer and more sophisticated method is infrared scanning. Here the surface of the specimen is scanned by an optical-mechanical system which focusses small points on the surface onto an IR detector. The output signal can be IR sensitive photographic film or a voltage. Walker [Ref. 9-8] used this technique to determine the temperature profile of microcircuits. He reportedly could resolve temperature differences as low as 0.5°C and could distinguish between components separated by as little as 0.0014 inch. This method has also been used to determine bond quality in objects as large as solid-fuel rocket motor cases [Ref. 9-9]. A practical application of the IR technique is discussed in Ref. 9-10.

IR photographic equipment can be obtained from most producers of regular photographic equipment. One has also developed a sensitive solid-state IR detector.

Temperature probes are also used as temperature profile gauges. These probes utilize conventional thermometers, thermistors, resistance thermometers, and thermocouples as the temperature sensing elements. These devices measure temperature accurately, conveniently, and economically, but great numbers of them are necessary to profile a surface without loss of resolution. Also, the devices themselves, along with associated lead wires, etc., tend to lower the temperature to be measured. Thus measurements with these devices tend to indicate temperatures lower than the no-contact measurements such as IR scanning.

9.4 Liquid Penetrants

This is a method of detecting surface flaws in most materials, i.e., flaws which are open to the surface but not readily detectable by visual means. Few flaws are revealed by penetrant inspection which could not be seen visually, but penetrants make the defects much easier to locate. Penetrants are applicable to all metals as well as to glazed ceramics, plastics, and non-porous materials. A special penetrant is used on porous materials.

Flaws are rapidly and easily found by covering the surface of the material or item with a liquid having a low surface tension and a low viscosity. The liquid is drawn into the surface defects by capillary action. After the excess penetrant has been removed from the surface, a developer is applied which makes the penetrant, and hence, the flaw, visible.

There are two basic types of penetrants available; dye penetrants and fluorescent penetrants. Dye penetrants consist of a dye dissolved in the liquid penetrant. The color of the dye is chosen to give greatest contrast with the developer. One dye penetrant in general use provides a red-on-white record of defects which can be removed from the material as a permanent record or for slide projection.

Fluorescent penetrants consist of a fluorescent phosphor dissolved in the liquid penetrant. This type of penetrant works in the same manner as other penetrants. However, flaws must be viewed under near ultraviolet light with a wavelength of 3650 Å.

Some precautions associated with using these materials are that

(1) the surface of the specimen must be thoroughly cleaned before the penetrant is applied,
(2) sufficient time must be allowed for the penetrant to penetrate the flaw,
(3) the excess penetrant must be removed with care,
(4) the developer must be applied within a temperature range specified, and
(5) the results must be interpreted with care and understanding of the method used and material to which it is applied.

There are several special penetrant techniques, two of which are radioactive penetrants and the filtered particle technique. The radioactive method uses a radioactive penetrant and detects the amount of this penetrant trapped in defects by either a photographic method or with a suitable radiation detector gauge. This technique is used primarily to determine the porosity in metal alloys. The filtered particle method is used to detect flaws in porous surface such as concrete, carbon, etc. The penetrant in this case contains suspended particles. The liquid is absorbed by the defect but the particles are larger than the defect and are filtered out and left behind on the surface. These particles then give an indication of a flaw. Fluorescent particles can be used to provide more contrast.

Liquid penetrant inspection is covered in detail in Refs. 9-3 and 9-11.

9.5 Magnetics

This method is based on the fact that flaws in a magnetic material have magnetic properties different from those of the material itself. Once a magnetic field is induced in a magnetic, any flaw will perturb or distort the field because the flaw has a different magnetic permeability, and thus a different reluctance, than the material. The flaw is located by measuring these perturbations. Figure 9-2 shows how flaws affect magnetic flux lines.

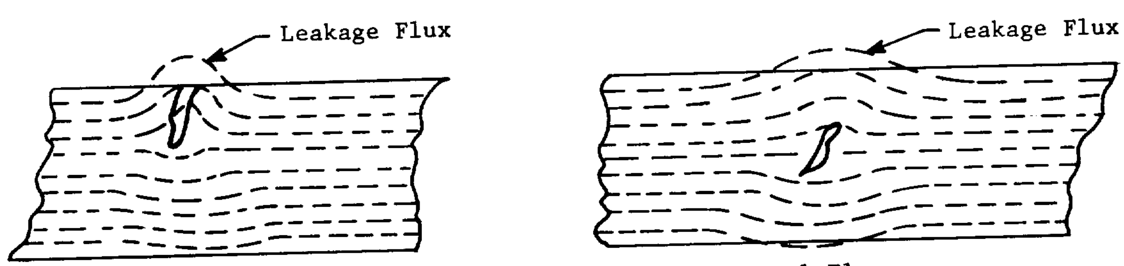
There are three basic ways of setting up a magnetic field in a magnetic material:

(1) by passing a current through all or a portion of the specimen,

(2) by passing a current through a coil surrounding or in contact with the specimen, and

(3) with magnets.

The method used depends on the type of flux lines desired. Passing a current through a specimen generates circular flux lines around the current path in accordance with the familiar "right-hand-rule". A coil around a specimen and magnets produce longitudinal magnetization. Both types of flux lines may be needed because the extent to which a flaw perturbs flux lines depends on its orientation with respect to the direction of the lines. For example, a crack perpendicular to flux lines perturbs them whereas a crack parallel to the lines may not. Thus, both flux orientations may be necessary to detect all flaws in a material.



(a) Surface Flay

(b) Internal Flow

Figure 9-2. Effect of Flow of Magnetic Flux Lines the eddy currents. As in the previously discussed magnetic method, the problem reduces to one of detecting and measuring these perturbations and distortions.

Two general types of probes are used in eddy current testing. One is an encircling coil which surrounds the specimen and investigates everything within the coil geometry. The other is a point probe which inspects only the area beneath it.

The coil type detector is affected by all the metal enclosed by the coil so statements about sensitivity are difficult to make, i.e., a long shallow crack may give the same output as a short deep one. The maximum resolution is defects with length comparable to the coil length (roughly 1/8 inches minimum) and depths of 5% or more of material wall thickness. The probes are also sensitive to the displaced volume of metal and are sensitive to defects on the order of the probe diameter. One such probe, the Probolog--developed by Shell Development Company--can detect cracks or seams 0.005 inches deep by 1/2 inches long. Generally, a defect where 1% material displaced in a 1/2 length is detectable. The probe can also detect 1% thickness changes in a 1/2 inch length. References 9-2, 9-2, and 9-12 have good discussion sections on the theory and use of eddy current testing.

9.6 Ultrasonics

Ultrasonic waves are acoustic waves above the audible range. They are employed in NDT to detect and locate flaws in composite materials and non-bonded areas in bonded materials. The impedance to ultrasonic propagation is different for a flaw or anomaly than for the basic material. Thus, a portion of the induced ultrasonic energy is reflected by a flaw just as it is by a boundary of the material. Measurement of the reflected portion or the unreflected portion is the basis for employing ultrasonics in NDT.

There are three methods of ultrasonic testing in general use: pulse, echo, transmission, and resonance.

Pulse Echo Method. In this method an applied pulse travels through the material and reflection is obtained from both a flaw and a material boundary. As the surface of the material is scanned, the appearance of a defect pulse locates the surface position of a flaw. The energy of this pulse is related to flaw size but is usually difficult to correlate with precision. By monitoring the time relationship of the initial pulse, the defect pulse, and the echo pulse, the defect can be located in depth. Many ingenious schemes have been used in the pulse echo method. For example, by introducing the initial pulse at an angle to the material surface the boundary reflection can be effectively removed in the return. Also, flaws not accessible by a simple geometry can be detected by letting the pulse zig-zag from one boundary to another until a flaw is reached. Thus, rather complex geometries can be probed by this method.

There are several methods of detecting the perturbations caused by flaws and defects. The most simple is to pass a compass over the magnetized surface. The compass needle will align with the over-all field except in the vicinity of a flaw. Although this method is crude and insensitive, the same principle gives good results when extended to distributing iron filings--either dry or in a liquid suspension-over the surface of interest. These filings are sometimes coated with a fluorescent material for a more visible pattern. These filings, of course, line up with the induced magnetic field except in the area of flaws or discontinuities in the material. Another detection method is to pass a current-carrying search coil over the surface. When the coil moves through a perturbation a voltage is generated between the coil and the inspected material. The magnitude of the voltage gives an indication of the size of the flaw. This is an especially useful method when the entire object to be inspected such as pipes, wire, etc., can be passed through a coil. A third method takes advantage of the Hall effect, which is the generation of a voltage across a current carrying material when it is placed in a magnetic field. Hall effect probes are usually made of a semiconductor material and are used by passing them over the surface of the magnetized specimen. Variations in the magnetic field due to defects and discontinuities result in a variation in the Hall voltage of the probe.

The sensitivity of this method depends on the strength of the magnetic field. All defects of consequence can usually be detected down to 0.060 inches below the surface. Defects down to 0.100 inches deep will show under ideal conditions [Ref. 9-12]. A number of other factors such as sharpness, direction and orientation of the defects also affect the sensitivity of the method.

Much of the equipment for conducting such tests can be fabricated in the laboratory based on the fundamental principles of the method. Reference 9-12 discusses several tests and test equipments put together in such a fashion. The magnetic particle technique is the most widely used. References 9-2, 9-3, and 9-12 cover magnetics in great detail. Reference 9-13 describes a special automated application of magnetic perturbation scanning.

Another method of NDT which is usually given a heading of its own but is discussed here under magnetics due to its close association is eddy current testing. It is based on the simple principle that when a coil carrying a high frequency alternating current is brought into the vicinity of an electrical conductor. These induced currents, in turn, induce a magnetic field about the conductor.

The induced currents, and thus the magnetic field, are affected by the permeability of the material. Although eddy currents can be used to test and measure such things as hardness, alloy content, uniformity of heat treatment, etc., it is largely used for flaw detection. Flaws perturb and distort the magnetic field produced by

Transmission Method. The transmission method is very similar to the pulse echo technique. A pulse is introduced at one boundary of a material but sensed at another. The energy transmitted past the flaw is attenuated due to reflection. The energy received is thus less when a flaw is present than when there is no flaw. This decrease indicates the surface position of a flaw but gives no measure of depth. This technique can also take advantage of reflections for complex geometries.

Resonance Method. The resonance method consists of exciting a material at its thickness resonant frequency. The material is driven by a transducer which is, in turn, controlled by a variable frequency oscillator. When the resonant frequency is reached, a standing wave will become established between the material faces. As the surface of the material is scanned, any change in resonant frequency not associated with material thickness is indicative of a flaw. This method is used for thickness measurement as well as flaw detection.

Both longitudinal and transverse waves are involved in propagation of ultrasonic energy in the techniques described above. Special propagation called Rayleigh waves and Lamb waves are also used but less frequently.

Rayleigh waves are surface waves analogous to ripples on water and result from control of the angle of incidence of the input ultrasonic energy. The Rayleigh-wave technique is useful for scanning across the surface of an item for flaws near the surface. A distinct advantage is the ability to investigate curved surfaces.

Lamb wave propagation applies only to thin materials and is an elastic vibration analogous to setting up ripples in the whole material. Such wave propagations have proven useful, for example, in detecting non-bonded areas in laminated structures where vibration in localized areas induced by the Lamb waves can be sensed. The Lamb-wave technique is capable of detecting cracks that extend as little as one mil below the surface of a material.

One of the major advantages of ultrasonic testing is its ability to penetrate deep into a material to locate flaws. This depends on available power and sensitivity of the detection equipment; however, the technique has been used to locate flaws as deep as 30 feet down a metallic bar. It also permits rapid measurements and is economical, relatively sensitive and reasonably accurate for measuring flaw extent and position. Accessibility to a single surface only is adequate for detecting many flaws and anomalies.

The resolution of ultrasonic test methods depends on the frequency of the ultrasonic propagation, i.e., the higher the frequency the smaller the defect that can be resolved. A limiting factor is that absorption of ultrasonic energy increases with increasing frequency. Thus a tradeoff between frequency and available energy must be made. Equipment generally available permits detection of flaws with dimensions in the 1 to 5 mil range.

The inconveniences of getting ultrasonic energy into and detection of the energy from a specimen is one of the major disadvantages. Air does not provide the needed impedance match between the transducer and specimen and liquid couplants such as oil, water, or glycerine are required. Another major disadvantage is in readout of the information. Display of pulse positions on a CRT is a much used method. Two-dimensional scanning and imaging is being employed to a limited extent but requires further development to become a practical tool. Major drawbacks with imaging are distortion and resolution. Even holographic methods are being investigated for three dimensional imaging.

Other disadvantages are the required operator skills and experience. There are also limitations posed by certain specimen geometries with regard to size, contour, and complexity. Misleading responses can also be obtained from normal internal structural characteristics such as large grains and material porosity.

Ultrasonic testing is no cure-all for NDT or even just flaw detection; however, when used with discretion in applicable situations, it is a valuable NDT tool. Good discussion on how to employ ultrasonic testing methods can be found in Refs. 9-2, 9-3, 9-12, 9-14, and 9-15. Reference 9-16 discusses the application of Rayleigh waves and Lamb waves. Specific instrumentation problems are covered well in Refs. 9-2 and 9-3. The theory of ultrasonic propagation is covered in depth in Ref. 9-15.

Ultrasonic testing equipment can be made fairly reliable and also light and portable enough to permit some on-the-site inspection. Much of the console ultrasonic equipment is developed for a specific purpose such as the ultrasonic scanning system described in Ref. 9-16.

9.7 Further Perspective

As mentioned earlier most forms of energy have been harnessed to some NDT problem or investigated as a possible NDT method. The preceding discussion by no means cover all NDT techniques but is an attempt to provide some appreciation for its capability. In searching the literature for new methods, look for topics such as color radiography, pulsed X-rays, microwaves, ultrasonic imaging, and cholesteric liquid crystals. Many of these are currently showing promise as NDT methods.

In addition to the references already cited, there are many other excellent sources of information. A good introductory survey is provided by McGonnagle [Ref. 9-17] which cites many other sources of information. Neither Ref. 9-2 nor 9-3 should be overlooked by the beginner in the field. Various military handbooks and standards contain a wealth of standard techniques; see for example Ref. 9-18. For keeping abreast of latest developments in NDT the journal, Materials Evaluation, is published monthly and describes many specific applications. Various government agency sponsored conferences specifically on NDT are frequently concerned, see for example Refs. 9-19 through 9-21.

References

9-1. Lockheed Missiles and Space Company: Parts Testing Methods Study Report. Air Force Contract No. AF 04(647)-787, July 27, 1964.

9-2. McGonnagle, Warren: Nondestructive Testing. McGraw-Hill, 1961.

9-3. McMaster, R. C.: Nondestructive Testing Handbook. 2 Volumes, Ronald Press, 1959.

9-4. Eastman Kodak Company: Radiography in Modern Industry. X-Ray Division, 1957.

9-5. Halmshaw, R., ed.: Physics of Industrial Radiography. Royal Armanent Research and Development Establishment, New York, American Elsevier, 1966.

9-6. The International Institute of Welding: Radiographs of Welds.

9-7. Berger, Harold.: Neutron Radiography. Argonne National Laboratory, 1965.

9-8. Walker, M.; Roschen, J.; and Schlegel, E.: An Infrared Scanning Technique for the Determination of Temperature Profiles in Microcircuits. IEEE, Trans. on Electron Devices, ED-10(4), 263, 1963.

9-9. Betz, Carl E.: Nondestructive Testing-Recent Achievements. Industrial Quality Control, vol. 23, no. 1, July 1966.

9-10. Selikson, Bernard; and Dimauro, Joseph: Reliability Screening Using Infrared. Rome Air Development Center, RADC-TR-66-360, October 1966, AD 642 112.

9-11. Betz, Carl E.: Principles of Penetrants. Magnaflux Corporation, Chicago, Illinois, 1962.

9-12. Hogarth, C. A.; and Blitz, J., eds.: Techniques of Nondestructive Testing. London Butterworths, 1960.

9-13. Rathmann, D. W.: Magnetic Perturbation Scanning--A New Nondestructive Testing Method. Materials Research and Standards, vol. 6, no. 8, August 1965.

9-14. Goldman, Richard G.: Ultrasonics Technology. General Electric Company, 1962.

9-15. Blitz, J.: Fundamentals of Ultrasonics. Brunel College of Technology, England, 1963.

9-16. Marshall Space Flight Center: Research Achievements Review, Series No. 10. NASA TMX-53466, Research and Development Operation, Huntsville, Alabama, 1965, N66-35979.

References (Continued)

9-17. McGonnagle, Warren; and Park, Ford: Nondestructive Testing. International Science and Technology, July 1964.

9-18. Anon.: Electromagnetic Testing. DOD Handbook H-54. One of a Series on Nondestructive Testing, October 1965.

9-19. Anon.: Nondestructive Testing: Trends and Techniques. NASA SP-5082. Proceedings of the Second Technology Status and Trends Symposium, Marshall Space Flight Center, 1967.

9-20. Anon.: Minutes of the Defense Conference on Nondestructive Testing. (DDC, AD 649 692), U. S. Army Materials Research Agency, Watertown, Mass., October 1966.

9-21. Beal, James B.: New Testing and Inspection Techniques. NASA Technology Utilization Conference, Rome, Ga., June 1967.

10. Environmental Testing

To some extent any test is an environmental test since some type of environment is always present. As discussed herein, however, environmental testing pertains to selecting and simulating the various environmental conditions of temperature, vibration, radiation, humidity, etc.,* for the express purpose of determining or verifying the capability of an item to operate satisfactorily when subjected to them. As such, it is not a basic test method itself but is a way of implementing the basic testing approaches described in earlier sections. Strength, life, and performance tests as well as other basic test types may all involve environmental testing. Some hardware development programs designate a particular phase of their testing program formally as environmental testing. In this section we are not concerned with such formality, rather with the basic problems and considerations for selecting and simulating environmental test conditions themselves.

The effect of environmental conditions, either natural or induced (man-made), on equipment is an important aspect of reliability. Environmental testing provides a method for investigating these effects. It is emphasized that environmental testing is done not because of the uncertainty of the environment but because of the uncertainty in the effects of the environment. The uncertainty of the environment can only be accounted for by conservative design practices to render it unimportant or perhaps by field testing to verify the success of the conservative design.

In environmental testing, conditions such as ambient temperature, vibration, and RF radiation are generated and controlled. In some cases there is a deliberate attempt to simulate as closely as possible the environmental profile during intended equipment operation. This is occasionally done, for example, in reliability demonstration with samples of prototype hardware. More frequently the emphasis is on simulating certain critical features of the total operational environment at specific severity levels. This is typically the approach in design qualification and flight acceptance testing of spacecraft components and systems and serves the useful purpose of uncovering design and material weaknesses and workmanship errors. In still other cases (such as development tests) the operational environment may not be known and test conditions consequently cover a wide range to explore the capabilities of an item.

All uses of environmental testing have as common objectives either determining the effect of the environmental conditions on an item or verifying that the item is

---

*The environmental conditions discussed herein generally exclude required power suppy inputs such as electrical current, hydraulic pressure, etc. even though these too may be included in a specific environmental test design.

capable of withstanding them. It is now employed in essentially all phases of hardware programs and from the parts and materials level of item complexity to fairly large systems. In programs which rely primarily on a "build and test" approach it provides the major source of confidence in operational hardware. It also remains a necessity in a complementary role in programs where greater emphasis is placed on analytical design procedures.

A logical alternative to environmental testing is testing under field or flight conditions. This alternative can provide the desired confidence but usually costs more (especially in the case of complex expensive items) and often delays the desired information. In field or flight testing test conditions are generally not as well controlled as environmental testing permits; hence, cause and effect relationships may be more obscure.

Environmental testing ranges in sophistication from very crude methods such as using an improvised temperature chamber to testing in very elaborate facilities which enable simulating many combinations of conditions. Tests may be purposely destructive (as in strength and life testing), or nondestructive (such as proof tests and burn-in).

Selecting the appropriate test conditions is the major problem associated with environmental testing. Basic factors that affect this selection are:

(1) the possible environmental conditions during intended use of the equipment,

(2) the subset of these that need to be treated by a testing approach, and

(3) the capability for generating and controlling them.

The crux of the problem lies in determining which environmental features can affect the item's behavior during intended use and in employing environmental simulation to investigate these features to the extent feasible within the constraints of cost, schedule, and testing capability. Not all environmental conditions that affect behavior can be readily simulated, and very rarely can all be generated simultaneously to account for interaction effects. Tradeoffs are thus necessary in selecting the test conditions to make the best use of available capability in obtaining environmental performance information.

10.1 Environmental Factors and Their Effects

Environmental conditions may be natural, induced, or combinations of these.* Natural environments are those which exist in nature such as the weather, solar radiation and low pressure in deep space. Induced environments are man-made and include such things as mechanical shock during transportation and handling, air

---

*In further discussion we will not generally be concerned with distinguishing whether a specific environmental factor such as temperature is natural or induced.

conditioned rooms for computers, and radio frequency interference (RFI). An example of a combined natural and induced environment is the set of conditions surrounding a space vehicle operating under the acceleration of the launch vehicle and the low pressure of space.

Table 10-1 illustrates the coverage of a number of environmental factors for a space system. The list is by no means all-inclusive for every application. Even factors of other origin (such as sub-oceanic) could possibly affect a space vehicle before recovery. Remember that the environment of on-board components may be very different due to sealing, shielding and mechanical isolation. Environmental testing is, of course, important to all types of items and not just to aerospace-oriented ones.

The set of environmental conditions in proper sequences and combinations that an item encounters during its lifetime is its environmental profile. The total profile begins during an item's fabrication and continues throughout its life. Therefore, environmental testing must consider the environments encountered in manufacturing, storage, transportation and handling as well as those experienced during operational use.

Descriptions of the environmental conditions are not always available in explicit form. No one knows precisely, for example, the environmental profile that a retro-rocket will experience throughout its life including all types of environmental factors and their severity levels. Through various sources of data on environments* it is often possible to select representative characteristics, such as averages or maximum levels of major factors, for adequately describing conditions for a test. The ways of describing test conditions and the associated concepts which were discussed in Sec. 2 are very relevant here.

Environmental conditions of greatest interest from the reliability viewpoint are those that have detrimental effects (i. e., those that cause drift, degradation, failure, wear, etc.) on equipment operation. Some conditions have no significant effect; some even may be beneficial. Table 10-2 lists some typical detrimental effects of several environmental factors. In many cases, effects not detectable when the factors are encountered singly show up when two or more are present simultaneously. For example, some electronic components function properly in either a low temperature or a vibrational environment, but when the environments are combined, component leads may break. The combined effects of several environmental factors as might apply to

---

*References 10-1 through 10-12 are sources for defining criteria on space, geographical, climatic, atmospheric, manufacturing, storage, and transportation environments. This type of information also frequently appears in Refs. 10-13 through 10-16. NASA SP-9000 [Ref. 10-17] also identifies environmental criteria documents for specific space programs and equipment.

Table 10-1

Environmental Factors for Space System Application

Environment	Mission Phases							
	Prelaunch			Flight				
	Fabri- cation	Storage, Trans., Handling	Pre- launch	Launch	Space	Re- entry	Landing	Recovery
Acceleration				x	x	x	x	
Acoustics		x	x	x		x	x	x
Aerodynam. heating				x		x	x	x
Albedo				x	x	x		
Asteroids					x			
Clouds			x	x				x
Cosmic radiation				x	x			
Dew		x	x	x			x	x
Electric atm.			x	x		x	x	x
Explosive atm.			x	x				x
Fog			x	x			x	x
Frost		x	x	x			x	x
Fungi	x	x	x					x
Gases,dissociated			x	x		x	x	x
Gases, ionized			x	x	x	x	x	x
Geomagnetism			x	x	x	x	x	x
Gravity	x	x	x	x	x	x	x	x
Hail		x	x	x			x	x
Humidity	x	x	x	x			x	x
Ice		x	x	x			x	x
Insects		x	x	x			x	x
Magnetic fields			x	x			x	x
Meteoroids					x			
Moisture	x	x	x	x			x	x
Nuclear radiation			x	x	x	x	x	x
Pollution, air	x	x	x	x			x	x
Pressure, air	x	x	x	x		x	x	x
Rain		x	x	x			x	x
RF Interference			x	x	x	x	x	x
Salt Spray		x	x	x			x	x
Sand and dust		x	x	x			x	x
Shock, Mechanical	x	x	x	x	x	x	x	x
Sleet		x	x	x			x	x
Snow		x	x	x			x	x
Solar radiation		x	x	x	x	x	x	x
Temperature	x	x	x	x	x	x	x	x
Thermal shock	x		x	x	x	x	x	
Turbulence			x	x		x	x	x
Vacuum					x			
Vapor trails				x		x	x	
Vibration	x	x	x	x	x	x	x	x
Winds and Gusts		x	x	x			x	x
Wind shear			x	x			x	x
Zero gravity					x			

Table 10-2

Environments and Typical Effects

Environment	Effects
Winds, Gust and Turbulence	Applies overloads to structures causing weakening or collapse; interferes with function such as aircraft control; convectively cools surfaces and components at low velocities and generates heat through friction at high velocities; delivers and deposits foreign materials which interfere with functions.
Precipitation: Sleet, snow, rain, hail, dew, frost	Applies overloads to structures causing weakening or collapse; removes heat from structures and items; aids corrosion; causes electrical failures causes surface deterioration and damages protective coating.
Sand and dust	Finely finished surfaces are scratched and abraded; friction between surfaces may be increased; lubricants can be contaminated; clogging of orfices, etc.; materials may be worn, cracked, or chipped.
Salt atmosphere and spray	Salt combined with water is a good conductor which can lower insulation resistance; causes galvanic corrosion of metals; chemical corrosion of metals is accelerated.
Humidity	Penetrates porous substances and causes leakage paths between electrical conductors; causes oxidation which leads to corrosion; moisture causes swelling in materials such as gaskets; excessive loss of humidity causes embrittlement and granulation
Sunshine	Causes colors to fade; affects elasticity of certain rubber compounds and plastics; increases temperatures within enclosures; can cause thermal aging; can cause ozone formation.
High temperature	Parameters of resistance, inductance, capacitance, power factor, dielectric constant, etc., will vary; insulation may soften; moving parts may jam due to expansion; finishes may blister; devices suffer thermal aging; oxidation and other chemical reactions are enhanced; viscosity reduction and evaporation of lubricants are problems; structural overloads may occur due to physical expansions.

Table 10-2 (Continued)

Environment	Effects
Low Temperature	Plastics and rubber lose flexibility and become brittle; electrical constants vary; ice formation occurs when moisture is present; lubricants gel and increase viscosity; high heat losses; finishes may crack; structures may be overloaded due to physical contraction.
Thermal shock	Materials may be instantaneously overstressed causing cracks and mechanical failure; electrical properties may be permanently altered.
High pressure	Structures such as containers, tanks, etc. may be overstressed and fractured; seals may leak; mechanical functions may be impaired.
Low pressure (High altitude)	Structures such as containers, tanks, etc. are overstressed and can be exploded or fractured; seals may leak; air bubbles in materials may explode causing damage; internal heating may increase due to lack of cooling medium; insulations may suffer arcing and breakdown; ozone may be formed; outgassing is more likely.
Gases	Corrosion of metals may be enhanced;dielectric strength may be reduced; an explosive environment can be created; heat transfer properties may be altered; oxidation may be accelerated.
Acceleration	Mechanical overloading of structures; items may be deformed or displaced; mechanical functions may impaired.
Vibration	Mechanical strength may deteriorate due to fatigue or overstress; electrical signals may be mechanically and erroneously modulated; materials and structures may be cracked, displaced, or shaken loose from mounts; mechanical functions may be impaired; finishes may be scoured by other surfaces; wear may be increased.
Shock	Mechanical structures may be overloaded causing weakening or collapse; items may be ripped from their mounts; mechanical functions may be impaired.
Nuclear/cosmic radiation	Causes heating and thermal aging; can alter chemical, physical, and electrical properties of materials; can produce gasses and secondary radiation; can cause oxidation and discoloration of surfaces; damages electrical and electronic components, especially semiconductors.

Table 10-2 (Continued)

Environment	Effects
Thermal Radiation	Causes heating and possible thermal aging; surface deterioration; structural weakening; oxidation; acceleration of chemical reactions; and alteration of physical and electrical properties.
RFI	Causes spurious and erroneous signals from electrical and electronic equipment and components; may cause complete disruption of normal electrical and electronic equipment such as communication and measuring systems.
Solar radiation	Effects similar to those for sunshine, nuclear/ cosmic radiation, and thermal radiation.
Albedo radiation	Albedo radiation is reflected electromagnetic (EM) radiation; amount depends on the reflective capabilities of illuminated object such as a planet or the moon; effects are the same as for other EM radiation.
Zero gravity	Disrupt gravity-dependent functions; aggravates high-temperature effects.
Magnetic fields	False signals are induced in electrical and electronic equipment; interfered with certain functions; can induce heating; can alter electrical properties.
Insects	Can cause surface damage and chemical reactions; can cause clogging and interference with function; can cause contamination of lubricants and other substances.
Clouds, Fog, smog, smoke, haze, etc.	Can interfere with optical and visual measurements; deposition of moisture, precipitation, etc.; enhances contamination; can act as an insulator or attenuator of radiated energy.
Acoustic noise	Vibration applied with sound waves rather than with a mechanical couple; can cause the same damage and results as vibrational environment, i. e., the sound energy excites structures to vibrate.

a particular item are illustrated in Table 10-3. As illustrated, synergistic effects do not always have adverse effects. For example, low temperature inhibits the growth of fungi and rain dilutes the corrosion effects of salt spray. A good tabulation of this type for many environmental factors and generally appropriate for many types of equipment is presented in Ref. 10-18.

A less frequent effect is when one environmental condition creates another. An example of this effect is when arcing between switch or relay contacts causes the formation of ozone, thus changing the environment and its effects.

Some conditions cause cumulative nonreversible changes in equipment; therefore, when considering equipment behavior at any point during its useful life, the history of environmental exposures should not be ignored. For example, heating from welding and soldering can cause permanent shifts in device characteristics, mechanical shock can result in permanent dislocation of a lead or a part, and nuclear radiation can cause permanent defects in semiconductor devices. The possible need for conditioning items prior to environmental testing to simulate the historical effects should not be ignored. This conditioning is sometimes necessary to assure that the response during the test is representative of that in operational use. Knowing the environmental history is not important when the effects were reversible, but whether all pertinent responses are reversible can be determined only through careful consideration. Ignoring

Table 10-3

Illustration of Interacting Environemental Effects

	Salt Spray	Vibration	Low Temperature	High Temperature
High Temperature	Accelerate Corrosion	Increase Rate of Wear	Mutually Exlcusive	
Low Temperature	Decelerate Corrosion	Intensity, Fatigue, Rupture, etc.		
Vibration	No Interaction			
Salt Spray				

the nonreversible effects which have occurred in previous testing and operations can lead to very misleading environmental test results. Admittedly they are not always easy to assess or simulate, but just knowing of their possibility can often be informative in testing.

In selecting the environmental factors and the severity levels and combinations of them to be treated, experience is usually the most reliable guide. For instance, a designer of a launch vehicle component may know that vibration and high temperature is far more likely to harm his component than low pressure and high ozone content. He may also know that the most severe conditions to consider for testing are determined by the launch environment rather than by the transportation and handling environment. Such prior knowledge and experience can help reduce the number of environmental tests needed to insure the successful operation of the item.

The problems associated with common environmental factors such as temperature, vibration, and thermal shock nearly always receive attention. Less familiar factors can sometimes be equally or even more important. Effects of albedo, for example, are more likely to be strange to most engineers than the effects of high temperature. For space application the operation of a lunar orbiting vehicle may be considerably influenced by reflected energy from the lunar surface. The characteristics of this energy for environmental testing purposes, however, are just those of electromagnetic radiation. Less common factors such as hail and insects demand special attention to determine what characteristics and severity levels to represent if indeed these factors need to be treated at all. With hail, for example, if mechanical impact damage is the major effect of interest, then the size, shape, velocity, and number per unit area of the simulated hailstones are the characteristics to worry about. On the other hand, the vibration induced by the incident hail may be the most significant factor. Insects can cause both mechanical and chemical damage and both characteristics demand consideration when insects can reasonably be expected.

When there is little available knowledge about the operational environment or its effect on an item, it is often simpler and more economical to test and see what happens instead of spending a great deal of time and money on an independent study. This is essentially the "build-and-test" approach and certainly has its limitations for large and expensive items. But when used with discretion it can be especially applicable to certain new designs or new applications of old designs.

10.2 Simulating the Conditions

The emphasis on environmental testing has led to the development of very elaborate facilities. For example, Ref. 10-19 gives a description of the huge NASA dynamic-test facility which can accommodate a six-million pound replica of the complete

Apollo-Saturn V vehicle in tests involving six degrees of motion. Other facilities have even more versatility in terms of the number of different conditions that generated simultaneously. Refs. 10-20 and 10-21 describe examples of such facilities. Some good general surveys of current environmental test capabilities are presented in Refs. 10-22 and 10-23. Ref. 10-18 gives detailed descriptions of methods for simulating various environmental conditions. Methods employed in NASA programs are described in various specifications and standards which are easily identified in Ref. 10-17 (NASA SP-9000). The most frequently cited standard for methods in military procurement is Ref. 10-24 (MIL-STD-810A). Refs. 10-15 and 10-16 are good sources for learning of new developments.

Given that certain environmental conditions need to be treated by a testing approach, it is not always possible to generate similar conditons even with the most elaborate facilities. No single facility, for example, can generate at once all of the types, energies, and intensities of Van Allen radiation for the space environment. Air turbulence, gases, and insects can typically present similar problems for environmental conditions not related to space. Many facilities are even limited in their capability to generate complex temperature profiles.

The realization of such problems has been the motivation for creating more sophisticated simulation capability. But there are often other ways of resolving the question at hand. Remember first that it is the effect of the environmental conditions that is of interest, not just the conditions themselves. Thus, is there a suitable substitute? For example, pebbles might substitute for hailstones if mechanical damage from impact is the effect of interest. Or if vibration induced by hailstones is of interest, then a vibration test already scheduled may be adequate.

Some effects are often more easily investigated from a more fundamental level. The effect of ionizing radiation is most often studied at a materials or parts level than at the level of assembled equipment. Also, the environmental conditions themselves may sometimes be separated into more fundamental components. Typically, a temperature profile is simulated by high and low levels and thermal shock; cosmic radiation may be separated into components composed separately of protons and beta particles. In such cases one must be alert that there is proper accounting for nonreversibility, interactions, and aging.

Elaborate environmental test facilities are not always needed to resolve certain problems. Simply heating individual circuit components with a soldering iron may in some cases be more informative than testing the entire circuit or assembly in an oven. And in the absence of certain capabilities, an answer from an improvised test may be better than no answer at all. For example, when concerned about mechanical shock testing capability, simply dropping the tubes from a prescribed height as a substitute may be better than ignoring the effects of shock altogether.

With the increased emphasis on generating conditions for different factors simultaneously, a very important question concerns whether to use single or combined environments. When facilities do not exist for generating combined environments, there is of course no choice but to generate single environments. Multiple environmental factors must then be treated as single environments in sequence. If the severity levels of the environments are not purposely damaging, as in a sequence of screening tests, the order of application is determined by whatever is most convenient. Tests which are purposely damaging, as in certain qualification tests and acceptance sampling, demand careful consideration of order of environments especially where only one or just a few test specimens are available. The basic criterion to employ in this case is to apply first those conditions which are least likely to damage the specimen. For a mechanical part, humidity and salt-spray tests would thus logically be applied before vibration or a mechanical load test. An electronic part would more likely be tested by applying vibration before high temperature. Such test sequencing allows the maximum amount of information to be obtained before damage occurs.

Ordering of environments for items composed of both mechanical and electrical parts is not as clear-cut. The same basic criterion still applies; however, ability to repair the item can greatly influence the ordering.

When capability exists for generating both single and combined environmental conditions, it does not necessarily follow that combined environmental testing is preferable. The decision depends mainly on what is to be accomplished with the test and is influenced strongly by factors such as time, cost, skills, and instrumentation.

Combined environmental testing has two significant advantages over single environment testing. First and most significant is the ability to investigate the synergistic effects of multiple conditions, i. e., combined testing in most instances more closely approximates the real environment. Second, several conditions can usually be applied simultaneously in a shorter time than in sequence due to savings in set-up time. Therefore combined testing often saves money. The major disadvantage is that the initial cost of the equipment for combined testing is higher.

In qualification and acceptance tests, combined environmental testing is preferable to testing with single environments. The increased confidence derived from the knowledge that synergistic effects are accounted for usually allows use of smaller safety factors in application.

In testing to relate cause and effect, combined environmental testing is used as an extension of single environment testing. Testing during development usually emphasizes learning the effects of single environments. Combined environments are employed after single environment effects have been determined and synergistic effects become of interest. Employing combined environments first can be impractical. Single environment testing can also be preferable in long duration tests due to the impracticality of committing combined environmental test facilities for long periods of time. A good discussion of the benefits and problems of testing with combined environments is presented in Ref. 10-25.

For Further Appreciation

The general survey article by Bleich[Ref. 10-23] provides easy reading and a good appreciation for environmental testing in general. Another good general discussion is presented in Ref. 10-26. The discussion by Junker [Ref. 10-27] is especially interesting in that it describes some of the chromological development of methods for specific environmental factors. Experience in the environmental testing of some spacecraft providing further justification of its effectiveness are summarized in Refs. 10-28 through 10-30.

References

10-1. Smith, Robert E.; and Otha H. Vaugh, Jr.: Space Environment Criteria Guidelines for Use in Space Vehicle Development. NASA, George C. Marshall Space Flight Center, NASA TMX-53521, Huntsville, Alabama, April 1967.

10-2. Shea L. Valley: Handbook of Geophysics and Space Environments. McGraw-Hill, New York, 1965.

10-3. Johnson, Francis S.: Satellite Environment Handbook. Stanford University Press, Stanford, California, 1965.

10-4. Cafaro, Constantino; et. al.: Engineering Study to Determine Feasible Methods of Simulating Planetary Albedo and Radiation Effects Upon the Thermal Balance of Spacecraft. Final Report for Jet Propulsion Laboratory, Contract JPL 950829, January 1965, N66-10684.

10-5. Kirby, W. G.; and Kindall, S. M.: Problems in the Laboratory Simulation of Space Particulate Radiation, RT-66-131(DDC, AD 804 214), Arnold Engineering Development Centèr, Arnold Air Force Station, Tenn., December 1966.

10-6. Leach, E. R.; Fairland, B. P.; and Bettenhausen, L. H.: The Space-Radiation Environment and Its Interactions with Matter. Radiation Effects Information Center, Battelle Memorial Institute, Columbus, Ohio, January 15, 1965, AD 456 854.

10-7. Magnolia, L. R.: The Meteoroid Environment--A Selected Bibliography. Report 9990-7156-KU-000 (STAR, N66-14539), TRW Systems, Redondo Beach, Calif., November 1965.

References (Continued)

10-8. Huscuri, L. A.: The Thermal Environment in Space. ESRO TM-49, (STAR, N20298), European Space Technology Center, Noordwijk, The Netherlands, December 1966.

10-9. Sterkin, Carol K.: Interactions of Spacecraft and other Moving Bodies with Natural Plasmas. NASA CR 70362 (STAR, 17066), JPL, Pasadena, Calif., December 1965.

10-10. Kindall, S. M.; and Kirby, W. G.: Micrometeoroid Environment and Its Simulation. RT-66-185 (DDC, AD 804 550), Arnold Engineering Development Center, Arnold Air Force Station, Tenn., December 1966.

10-11. Daniels, Glenn E.: Terrestrial Environment (Climatic) Criteria Guidelines for Use in Space Vehicle Development. 1964 Revision. NASA TMX 53023, Huntsville, Alabama, March 1964, N64-26359.

10-12. Ostrem, Fred E.; and Rummerman, M. L.: Shock and Vibration Transportation Environment Criteria. Final Report NASA-MSFC Contract NAS-8-11451, September 21, 1965, N66-34681.

10-13. Carrigan, A. L.; and Oliver, N. J.: Geophysics and Space Data Bulletin. (Published quarterly; available through DDC). Office of Aerospace Research, Hanscom Field, Mass.

10-14. Anon.: The Shock and Vibration Bulletin. (Published annually in several parts; available through DDC). A Publication of the Shock and Vibration Information Center, Naval Research Laboratory, Washington, D. C.

10-15. Sander, H. F., ed.: The Journal of Environmental Sciences. (Published bi-monthly). Institute of Environmental Sciences, Mount Prospect, Illinois.

10-16. Post, M. Z., ed.: Environmental Quarterly. (Published quarterly). Environmental Publications, Inc., Little Neck, N. Y.

10-17. Anon.: NASA Specifications and Standards. NASA SP-9000, May 1967.

10-18. Peeler, F. C.: Environmental Testing. Chapter 8 of Reliability Engineering for Electronic Systems, edited by R. H. Meyers, K. L. Wong, and H. M. Gordy, John Wiley and Sons, Inc., New York, 1964.

10-19. Kramer, Warner: Computerized Shake Facility Tests Saturn Moon Rocket, Control Engineering, vol. 14, no. 5, May 1967, p. 87.

10-20. Anon.: Procedures Manual for Structural Dynamics Testing. Goddard Space Flight Center, May 1966.

10-21. Anon.: Test Facilities Handbook. (DDC, AD 803 822), Arnold Engineering Development Center, Arnold Air Force Station, Tenn., November 1966.

10-22. Stephenson, Thomas B.: A Survey of Modern Environmental Testing. EDN, vol. 11, no. 14, November 1966.

References (Continued)

10-23. Bleich, Stephen F.: Environmental Testing, International Science and Technology, No. 33, September 1964, p. 42.

10-24. Anon.: Military Standard Environmental Test Methods. MIL-STD-810B, June 1967.

10-25. Martin-Marietta Corporation: Investigation of the Requirements for Combined-Environmental Testing of Air Force Material, Volume 1. AFFDL-TR-65-126, Contract AF 33(615)-1313, Denver, Colorado, December 1965.

10-26. Askin, David: Philosophy of Environmental Testing. Report M66-21-1 (DDC, AD 633 577), Frankford Arsenal, Philadelphia, Pa., April 1966.

10-27. Junker, V. J.: The Evolution of USAF Environmental Testing. Technical Report AFFDL-TR-65-197 (DDC, AD 625 543), Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Ohio, October 1965.

10-28. New John C.: Achieving Satellite Reliability through Environmental Tests. NASA Technical Note D-1853, July 1963.

10-29. Boeckel, John H.: The Purposes of Environmental Testing for Scientific Satellites. NASA Technical Note D-1900, July 1963.

10-30. New, John C.; and Timmins, A. R.: Effectiveness of Environmental-Simulation Testing for Spacecraft. NASA Technical Note D-4009, June 1967.

11. Accelerated Testing

In accelerated testing items are subjected to conditions more severe than those for normal use in an attempt to speed up aging and hence obtain degradation and failures in less time. It is thus a technique to shorten test time but applicable only to those tests in which aging is important. It is a very loosely defined concept; attempts to make it rigorous generally run into problems.

Loosely speaking accelerated testing probably started when someone said, "Let's shoot the juice to it and see what happens". This means roughly, "Let's treat it worse than we expect it to be treated in ordinary practice and then see what happens". One difficulty is that "treating it worse" does not always mean "shooting the juice to it". For example, electrical contacts behave better as voltage and current are increased (up to a point) and some heating may improve matters for electronic equipment by driving off moisture.

Because there is a reasonably firm qualitative foundation for much of accelerated testing, it is often used beneficially and without too much difficulty in qualitative roles such as failure mode investigation. It is in the quantitative interpretation and application such as predicting performance and life of items under normal operating conditions that it begins to run into the greatest difficulty.

A major consideration in accelerated testing is concerned with what really happens when there is an attempt to speed up the aging process. Is aging truly being accelerated or are other mechanisms being excited? To provide a common basis for discussing this and for appreciating the practical problems of accelerated testing, a simple but useful definition of true acceleration is given below.

It is first recalled that a change of conditions may, in some cases, cause both reversible and nonreversible responses. For example, some of the change in resistance for a resistor when its ambient temperature is changed can be accounted for by the temperature coefficient with the remainder due to aging. Such reversible effects can usually be distinguished; for example, they often appear as initial transients. It is assumed in the following discussion that all such reversible effects have been eliminated or subtracted out of the data. In particular a reference to the state of a device will be understood to mean that only nonreversible effects are included.

> True acceleration. The aging of an item is truly accelerated if and only if the item, under the accelerating conditions passes through all of the same states and in the same order that it would have under usual conditions.

Thus at any time t under accelerated conditions the item is in the same state as it would have been at some time kt under usual conditions. k is the acceleration factor and, itself, may be a function of time.

By the above definition one never really knows whether true acceleration exists since aging, a nonreversible process, precludes observing all states of the same item under both (accelerated versus usual) conditions. At best, one can only observe similar items under both conditions, and even then not necessarily all of the characteristics that define the state down to the last orbital electron. Thus, for practical purposes the definition must be relaxed to correspond to observable states of similar items. From the practical viewpoint a more meaningful definition is given.

> "Practical True Acceleration". The aging of an item is truly accelerated if and only if the item, under the accelerating conditions, passes reasonably through all of the same states and in the same order that a similar item does at usual conditions.

Thus, at any time t the item under accelerated conditions will have, within accepted limits the same state parameter values as those for a similar item at time kt under usual conditions. k is the acceleration factor, and itself, may be a function of time.

The detailed specification of the item state will vary with our needs and desires and with the required tractability of the resulting equations. The state of an item will ordinarily have several dimensions (components); so it can be classed as a vector. For example, consider a resistor. If we are concerned only about its resistance and nothing else, then the state of the system will be given by the resistance of the device (or something equivalent thereto such as a ratio of the resistance to an initial resistance). On the other hand, we may be concerned about several parameters, such as the resistance, the temperature coefficient of resistance, the voltage coefficient of resistance, and the chemical composition of the resistive material. Then there will be several dimensions for the system state, and two states will not be the same unless all corresponding dimensions are pair-wise the same.

One would hardly expect the two sequences of states achieved by the two tests to coincide (i.e., on a state parameter pair-wise basis) with any high precision. This relates back to the term "reasonably" in the definition of "practical true acceleration" and to the lack of precision which one is willing to accept. In order to have true acceleration in the practical sense it is only necessary that the things in which we are immediately interested be close enough under the two sets of conditions. For example, it is possible that different dimensional (components) of the state are accelerated at different rates. Not all failure modes and mechanisms need be identical. Similarity of failure modes and mechanisms may help in determining whether or not there is in fact true acceleration but it is not necessary.

In addition to verifying that true acceleration exists, much of the effort is devoted to determining the acceleration factor. It is, of course, convenient if the acceleration factor is constant and depends in some tractable way on the severity level. Estimates of an acceleration factor depend on the statistical procedures used to arrive at it.

Some of the gross failure modes in aerospace systems which can be accelerated are fatigue, corrosion, creep-rupture, stress corrosion, and various combinations of them--for mechanical parts. In electronics one does not ordinarily specify the gross failure modes for acceleration but rather specifies the "stresses" which are being increased. Some of these are temperature, supply voltage, power dissipation, vibration, humidity, corrosive elements in the ambient. There is a large body of material in the mechanical and metallurgical fields dealing with those gross failure modes, both on an implied level and on a more theoretical basis. See for example the ASTM references on fatigue for a reasonably complete bibliography on that subject; a good starting point is Ref. 11-1. Since the behavior of electronic components is organized differently, there is no organized body of literature dealing with the gross failure modes which cuts across all components. A number of information sources on accelerated testing of electronic components are listed [Refs. 11-2 through 11-26], but they should be read critically because many of them contain conceptual errors of varying degrees of importance. They will, however, give a newcomer to the field an idea of what other people are doing or are suggesting should be done. A good forthcoming state-of-the-art survey for accelerated testing of electronics is Ref. 11-27.

11.1 Methods of Programming Test Conditions

The most familiar ways of programming the conditions for accelerated testing are constant-stress and step-stress. Another approach frequently recognized is the progressive-stress method; however, as later described this is no different from the step-stress approach when the steps are small.

Remember that it is the actual severity of the stress that is of interest rather than the level of a stress factor in defining the appropriate severity when multiple stresses are involved. There are also potential problems to be resolved beforehand when multiple stresses are involved.*

The different methods are discussed separately below. Statistical designs are not discussed with these because the procedures are independent of the acceleration of the tests.

11.1.1 Constant-Stress Method

This is the traditional type of test wherein the severity level remains constant throughout the life of the items on test. It is customary to run tests at

* These concepts were discussed in more detail in Sec. 2.2.2.

several severity levels and to plot a curve (showing some measure of goodness versus the measure of severity level) which is faired through the resulting points. The measure of goodness may be failure rate, time to failure, etc. A sample of several items is usually tested at a level and the test stopped when some fraction of the original sample has failed, i.e., the test at each level is a failure-truncated test. For reliability prediction purposes the early fraction that fails is most important because only the short-lived items are going to affect seriously the reliability. For engineering improvement purposes the fraction that is very long-lived may be important as an example of a design which did in fact prove quite reliable.

The measure of severity level where there is only one dimension is usually easy; some function of the parameter used to describe that dimension is plotted. As is traditional with engineering, one hopes to choose the coordinate axes so that the stress severity versus goodness line is predicted to be straight. If more than one stress is being changed, then it is up to the engineer to either

(1) find some scalar which will measure the overall severity level or

(2) plot each one of the dimensions of the severity level; this immediately creates a problem for graphical presentation but the analytic continuations can easily be written down in their generality.

11.1.2 Step-Stress Method

In this method the severity for a sample of items is simply increased in steps or increments until some criterion for test termination is met. All steps do not have to be the same size even though this is most often done.

The term step-stress as used in the literature is ambiguous. It is convenient to classify step-stressing into three categories:

(1) Large steps in which the steps are presumed high enough and long enough so that for a given step the damage accumulated at all previous steps is negligible.

(2) Small steps in which the steps are small enough so that in the analysis one can presume with negligible error that the severity level is steadily increasing. This is then just the progressive-stress case (Sec. 11.1.3).

(3) Medium steps for which the assumptions for neither small nor large steps are valid. The cumulative damage at previous steps must be taken into account but the steps are not small enough that the severity level can be considered continuously increasing.

In order to be able to refer reasonably to these three cases in further discussions the following terminology is used: large/step-stress, medium/step-stress, and small/step-stress. The size designations are not absolute but are relative to the kind of analysis that must be performed.

Large/step-stress tests are analyzed as if they were constant-stress tests being run at the severity level of the last step. Parts which are very expensive or otherwise difficult to acquire or test are often treated in this way. Often a sample of only one is used. It is wise to consider the results as "ballpark" figures since the necessary assumption of negligibility of previous steps is likely to be in error. Preliminary tests are very often run in this way to be followed by a more comprehensive set of tests later.

Small/step-stressing is analyzed the same as progressive-stressing, and in fact by definition there is really no distinction between them. Whether in actual practice the value of a severity level jumps in small but nonzero increments or rises smoothly may be only a matter of resolution of measuring instruments or of pencil lines on graph paper. In many cases there will be a large economic advantage to choosing either very small step increments or a nominally continuously increasing procedure. As an example, if extremely accurate voltage steps are desired, a stepping switch might be used with a voltage divider; otherwise a slow motor might be used to turn a multi-turn potentiometer.

The only difference in analysis between medium/step-stressing and progressive-stressing is the summation signs being required in the former and an integral sign in the latter. Discretion may call for increasing the measure of severity level in such a way that this summation or integration is very tractable (possibly replacing the need for a complicated digital computer analysis with one which can be done by hand via the evaluation of a simple equation). As an example consider a situation wherein the Arrhenius equation* is presumed and temperature is being increased. If temperature itself is increased linearly** the summation or integral will be intractable; if $1/T$ is increased linearly the equations are tractable; further, if $e^{-1/T}$ is increased linearly the analysis may be even more tractable.

Less testing time is usually the major advantage promoted for using step-stress tests instead of constant-stress tests. A direct comparison of the methods requires an assumption for some theory of cumulative damage. In the area of metal fatigue there are many theories of cumulative damage. In electronics a simple linear model is most often assumed because of both simplicity and the absence of knowledge about existing processes.

A linear model of cumulative damage is generally, at best, a gross approximation. In some circumstances it consistently underestimates and in other circumstances,

---

* The Arrhenius and other equations for acceleration are discussed in Sec. 11.2.

** In this case a linear increase in the stepped parameter means that the level for each successive step is increased linearly with time.

consistently overestimates the correct value. Regardless of these deficiencies it has the big engineering advantages of being tractable, easily remembered, and widely used.

An important parameter in step-stress testing is the ratio of severity step size to the time at each level. This controls the rate of increase of the stress severity and is the parameter which is varied when running several tests on a particular population of items.

It is possible for some kinds of items that the maximum useful severity level will be exceeded before the device fails in the proper mode. For example, on transistors which are thermally stressed there are sometimes eutectic points where melting occurs and the transistor essentially ceases to be a transistor. When this happens the rate of increasing the stress severity needs to be lessened. It is also possible to change the slope of the steps during the course of the tests--there is no law that says it has to be constant. The severity level limits (i.e., the level where the device ceases to be its usual self) are an important limitation to step-stressing. There are other cases where the failure mode changes so drastically at some level that it is senseless to continue testing above that level.

Another advantage of step-stress testing occasionally cited is the elimination of "switch-on" problems such as initial transients and failures due to high stress rates. This is because the severity level is zero at the beginning and the severity increase can be held gradual.

11.1.3 Progressive-Stress Method

The problems and considerations associated with progressive-stress tests are essentially those of medium/step-stress and small/step-stress tests. As mentioned in Sec. 11.1.2, there is no need to belabor the difference between small/step-stressing and progressive-stressing; the only difference between medium/step-stress and progressive-stress testing is in the tractability and form of the resulting analysis.

11.1.4 Other Approaches

The simplest modification of the step-stress or progressive-stress method is to start the severity level above zero. This is an endeavor to save time, and upon occasion, to reduce the amount of cumulative damage done at severity levels other than the failure level. Some kinds of programs which are concerned with investigating cumulative damage theories may change the severity level only once during a test. For example, the initial part of one test may be at a high severity level and the remainder at a low severity level; a subsequent test reverses the procedure. Not much work of this sort is done in electronics, but metallic fatigue is a field wherein these methods of programming stresses have received considerable attention. There is no reason why the programming of the severity levels in an accelerated test cannot be anything which will add to the useful knowledge about the system. The term "probe testing" has been used in the literature, but this testing is a special case of step- or progressive-stressing where the severity level is a vector with several dimensions (components).

11.2 Acceleration Equations

Temperature is the most popular and probably the most important environmental factor for accelerated testing. It has been treated extensively in the past, and its continued use appears both easy and fruitful. The equations used in the literature to describe the accelerated behavior are a matter of some controversy. There are many experimental situations wherein temperature is changed, the results recorded at each level, then the logarithm of the results are plotted versus 1/kT (or against 1/T). This is often done because the conceptual model being used to describe the process suggests that the resulting line will be nearly straight (neglecting random variations). Many of these situations have nothing to do with the Arrhenius or Eyring equations. For example, the product of the electron and hole concentrations in a semiconductor is given by

$$np = P(T) \times \exp(-E_g/kT)$$

where P(T) is a polynomial in T (or similar expression containing fractional exponents) and E_g is the bandgap energy. The form of this equation has its roots in the Maxwell-Boltzmann distribution. There are thermodynamic equations which have been put in the form

$$y = \exp(-E/kT)$$

where E is some thermodynamic energy. One of the reasons this form is preferred is that it turns up in the rather tractable analysis for perfect gases. Therefore, in other situations some new generalized parameters may be defined by an equation of that form. The energies E and E_g above are not usually constant, but the variation of either with temperature is usually quite mild compared to the T in denominator. It happens to be convenient sometimes to split the E in the above equation into two parts, E_1 and E_2, such that $\exp(-E_1/kT)$ is a polynomial in T (fractional exponents allowed*). The resulting equation is $y = P_1(T) \exp(-E_2/kT)$. Unless the data are quite accurate, much more so than usually found in engineering experiments, the variation due to $P_1(T)$ is completely swamped by the random variations measured in y. In some experiments in basic physics (e.g., determining the bandgap energy of silicon) or in chemistry (e.g., the hydrogen iodide decomposition into hydrogen and iodine) the

---

* This can happen exactly if E_1/kT is the log of such a polynomial.

results are accurate enough to behoove one to get a reasonably accurate model for the polynomial in T. Then the y's can be corrected and the nonstraightness of the resulting line made due solely to the temperature dependence of the energy in the exponent. It is worth noting that the energy of the exponent is associated with a physical quantity only to within a few kT because most models neglect variations in energy of this amount. The variations are quite small, e.g., at room temperature $kT \simeq 0.025$eV.

Other forms relating time and temperature have been used in chemistry and metallurgy (rarely in physics). In particular, a time-temperature parameter is introduced and the observed behavior is postulated to be expressible in terms of this parameter with no other time or temperature dependence.

The reasonable success of the exp(-E/kT) equation has led many people to speculate on extensions of it to include damaging factors other than temperature. These extensions have been completely arbitrary and should not be imbued with any mystical sense of theoretical soundness.

11.2.1 The Arrenius Equation

This is often cited as the classic example for temperature dependence of reaction rates and can be written*

$$rr = A \exp(-E/kT) \ .$$

While we do not have access to the personal thoughts of Arrhenius, he was undoubtedly influenced for the form of the equation by the thermodynamic forms mentioned above. The Arrhenius equation has enjoyed an appreciable amount of success for both interpolation and extrapolation.

The Arrhenius equation is often written in an approximate form when the temperature excursions are small, as

$$rr1/rr2 = \exp(-E/kT_0 \times \Delta T/T_0) \ ,$$

where T_0 is the nominal temperature of the reaction. A very common form of this approximation is the statement that specific reaction rates will double for every

---

* Sometimes an R is used in place of the k. R is the universal gas constant; k is Boltzmann's constant. Chemists tend to use the former and physicists the latter, the difference being per mole or per molecule, respectively. When R is used, E is usually given in kilocalories per mole, whereas when k is used, E is usually given in electron volts per molecule. Very often the per molecule or per mole is dropped. A is often called the frequency factor because the earliest reactions considered were of the first order. This name does not apply to reactions of other orders. The specific reaction rate is also called the reaction rate constant.

energy surface can be introduced and if there are not too many dimensions (say only one or two), the system is extremely simple, and you are lucky, this surface can be obtained from quantum mechanical considerations. Actually, "For all but the simplest systems this is not feasible. There are also semiempirical methods for the calculation of these potential energy surfaces, but these do not generally give sufficiently accurate surfaces for practical use in predicting kinetic data. In fact, for all but the very simplest reactions one examines the nature of the activated complex from experimental kinetic data."

Electronic components are complex engineering systems from the point of view of theoretical chemistry/physics and for practical purposes use of absolute reaction rate theory will offer little if anything over the Arrhenius equation. One of the biggest obstacles to its use is the tremendous scatter in the data. Another is that the specific reaction rate is not observed, but some complicated function of it is. By the time one is discussing failure rates, he is a long, long way from a specific reaction rate.

11.2.3 A Relationship for Stress Severity Dependency on a Time-Temperature Parameter

In the field of metallurgy, in particular for prediction of time to creep-rupture failure, a time-Temperature parameter (tTp) has been found to be useful. A similar tTp is presumably also useful for plastics. In this conceptual model the severity level is considered expressible as a function of a single parameter, the tTp, where

$$tTp \equiv T(A + \log t_F) \equiv T \log(t_F/t_0) ,$$

with t_F representing the time to failure and T the absolute temperature. In creep-rupture situations the constant t_0 is taken to be on the order of 10^{-15} hr, so that for 1 hr < t_F < 1000 hr, 15 < $\log(t/t_0)$ < 12. This range of times includes most test times. It can be seen that variations in failure time up to a factor of three will produce a change in the tTp of only a few percent. Thus there is an extreme compression of the time scale. Discrepancies in the data are covered up whether by intention or not.

It is easy to show by simple algebra that this equation is inconsistent with the Arrhenius equation when used as a theory for cumulative damage notwithstanding the procrustean approaches in the literature to derive one from the other. In fact, the tTp cannot be used at all as a measure of cumulative damages.

There is certainly nothing wrong with trying to fit the results of an accelerated life test of an electronic component with a tTp. The constant t_0 can be considered adjustable to give the best fit to the data. Just remember that time and Temperature can enter in no other way than through the tTp.

10°C rise in temperature. If T_0 is taken anywhere in the range 0 to 100°C, E varies between 0.5 and 0.7 eV. The uncertainties in the experimental data in many practical situations swamp out the approximation.

If the results of an engineering investigation show that the Arrhenius behavior is not followed, all that such results mean are that the system is obviously behaving in some other fashion. The results do not mean anything is right or wrong, but simply that the simple-minded conceptual model being used is inadequate.

Since its birth over 75 years ago, the Arrhenius equation has undergone modifications in endeavors to make it more widely applicable. These modifications are basically of little concern in accelerated testing; they are discussed in texts and articles on physical chemistry.

11.2.2 The Eyring Equation (The Absolute Reaction Rate Equation)

The Eyring equation,* or as it is more often known in physical chemistry, the equation for absolute reaction rates, seems to have assumed an undue, god-like image in some of the reliability-physics/accelerated-testing literature. The specific reaction rate may be written as

$$rr = \frac{\kappa kT}{h} \exp\left(\frac{-\Delta G\dagger}{kT}\right)$$

where $\Delta G\dagger$ is the Gibbs free energy of the activated complex, κ is a transmission coefficient and is usually virtually unity, and h is Planck's constant. "It should be clearly noted that the equation has been developed for an elementary reaction and that it should be applied only to such a reaction." "The absolute reaction rate theory...has been applied with success to a wide range of solid, liquid, and vapor phase reactions. It is equally useful in considering the rates of very rapid reactions which may occur in a flame, and the rates of those reactions which under ordinary conditions require geologic ages."

Now since $\Delta G\dagger = \Delta H\dagger - T\Delta S\dagger$, where $H\dagger$ and $S\dagger$ are enthalpy and entropy of the activated complex respectively, the equation can be put in the following form

$$rr = \frac{\kappa kT}{h} \exp\left(\frac{\Delta S\dagger}{k}\right) \exp\left(\frac{-\Delta H\dagger}{kT}\right) .$$

The $\Delta H\dagger$ is closely associated with the activation energy and is equal to it within an uncertainty of a few kT (depending on the exact conceptual model chosen for the reaction kinetics). It is the term $\exp(\Delta S\dagger/k)$ that gives the trouble. A potential

* An excellent short reference is Ref. 11-29. All the quotations in this section (11.2.2) are from this paper.

11.2.4 Extensions of the Absolute Reaction Rate Equation

As mentioned in the introductory portion of this section there have been extensions of the equation for absolute reaction rate (Eyring) which have been termed "generalized Eyring equations". They involve an exponential term in which two arbitrary constants and an arbitrary function of the measure of stress are introduced. The name, Eyring, applied to this equation should not mislead anyone into thinking that he and the other eminent physical chemists who developed the theory of absolute reaction rates are responsible for it. The equation itself is obviously gross empiricism, which while not bad in itself, takes some of the aura of primacy away from it.

11.2.5 Applicable Parameters

Very often in the literature, the life of an element is considered to have for example, an Arrhenius acceleration, but the exact meaning to be associated with this statement is not clear. If the life is a random variable it is difficult to know what is meant by the life following a certain law. It is much more meaningful to assert that a particular parameter in the life distribution follows a certain law. For example, if the hazard rate is constant, the life distribution has a single parameter (λ), and it can be asserted that λ has the Arrhenius form. Other one-parameter distributions can be treated similarly.

If the life distribution has more than one parameter, e.g., the Weibull distribution, then it makes no sense at all to assert that the life follows the Arrhenius acceleration formula. Rather one must assert that one (or both) of the parameters in the distribution follows the Arrhenius formula. It is most commonly (and implicitly) assumed that the reciprocal of the location parameter has the Arrhenius form and that the shape parameter remains constant. If in fact the shape parameter remains constant, a very simple time transformation will convert the Weibull distribution to the exponential, and there is no need to treat it as a special case. If the shape parameter does not stay constant, there are serious difficulties in interpreting the data unless separate forms are assumed apriori for both the location and shape parameters. In the Normal distribution which also has two parameters (the mean and variance) a similar problem arises, viz., separate equations must be assumed for each of the parameters in order to interpret the data. Rarely does anyone in the field of reliability concern himself with a life distribution which has more than two parameters because the data are usually too inadequate to make sense out of the results.

You must always ask yourself the question of the title: This is a thermal acceleration equation for what parameter?

11.2.6 Estimation of Parameters

Estimating the parameters of the Arrhenius equation from a set of life data is difficult because the equation is nonlinear. Several computer programs are mentioned in Sec. 8, Vol. II - Computation of this series which can perform this by iteration using nonlinear least squares methods. The most desirable way, if it could be done, is actually to substitute the Arrhenius formula in the probability density function (or the cumulative distribution function) and use the data to estimate the parameters directly. In using a least squares analysis attention must always be paid to the weighting of the data points in terms of the estimated accuracy. The weight given to any point is proportional to the reciprocal of the square root of the variance at that position. For example, if a Poisson distribution has been assumed, the uncertainty in the estimate of the rate is directly related to the number of failures (or, more exactly, is directly related to the true mean at that position).

Some authors have cautioned against transformation of the variables because of the effect on the weight of the points. One can transform the variables if he changes the weighting of the points with every transformation of the dependent variable. The weighting does not change at all with transformations of the independent variable. The help of a competent statistician will be worthwhile if the estimations are at all critical, e.g., in meeting a specification.

11.3 Extrapolation

Just as everyone uses accelerated testing and will continue to use it regardless of the judgements passed on it and its limitations, everyone will continue to extrapolate from the data regardless of the admonitions against and the dangers befalling extrapolation. It is not the purpose of this section to proscribe extrapolation but to show what uncertainties exist when it is done. It is presumed that some reasonable equation derived from a model of the process exists. Curves which are fit to data points by brute force with a series, such as a power series or orthogonal polynomials, are never to be extrapolated unless the true model is of that form. Those formulas are for interpolation only; they usually behave very wildly outside the data interval. It makes no difference for example whether a least squares fit or an exact fit to the data is used, the extrapolated curve will not be a smooth extrapolation of the data points nor is it intended to be. This is not an ivory tower proscription but a very realistic one.

11.3.1 Known Model

If the model is known outside the range of the data, then the problem is statistical in nature. A statistician may be able to give help on the design of the original experiment to maximize the precision of the extrapolated value. Virtually always the data are transformed so that the resulting curve is a straight line and only a straight line is considered in the following. The principles are applicable to more complicated curves however. Consider that the origin is at the "center of gravity" of the data points.* The straight line will be of the form y = mx + b where, except for uncertainty, b is zero. In virtually all situations the error in extrapolating due to the uncertainty in b is negligible. The uncertainty in the slope m is what causes the trouble as is shown in Fig. 11-1. The heavy line is the best line calculated by whatever means are desired. The dashed lines are the loci of confidence points about that line for extrapolation purposes. In the lower right of the figure the uncertainty in x for a given value of y is shown as a solid heavy line. In the upper left the uncertainty in y for a given value of x is similarly shown. The details of calculating these intervals are included in many computer programs and are available in some statistics texts. If extrapolations are made very far, and they usually are, the uncertainty can be an appreciable fraction of the value. For example where log time scales are used uncertainties in time of factors of 10 → 100 are not unknown. These are very real uncertainties; within the confidence limits stated you don't know where a point lies, and giving point estimates can be extremely misleading to the reader. Remember that this discussion presumes that the model accurately describes the behavior in the region of extrapolation. Models are often known to have a very restricted region of applicability; certainly this region should be included in any equations written down so that the limitations are kept firmly in view.

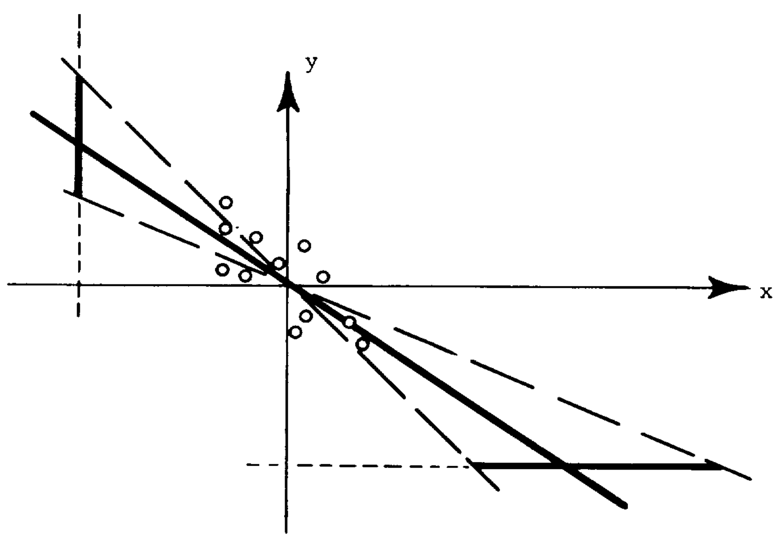


Figure 11-1. Illustration of Extrapolation

* If the points are weighted in the analysis, the origin will be the weighted center of gravity.

11.3.2 Unknown Model

If the behavior in the region of extrapolation is not known to follow a particular model, or if, as usual, one is not sure whether or not it follows that model in the region, then it is often possible to hypothesize several models and to extrapolate according to each. If the decision you make on the basis of the extrapolation is extremely sensitive to the model which has been assumed, then you are in trouble. If the decision you make on the basis of the extrapolation is not very sensitive to the model, it is generally assumed to be safe to go ahead. Fortunately, very often the latter is the situation. No one really cares what the exact prediction of life of a component or a part is, all he really cares about it whether it is long enough. This is why many acceleration and extrapolation techniques are successful. The parts are very good, so good in fact that they transcend the limitations of the analysis.

In most engineering situations one has to work in regions where decisions are not clearcut and there just aren't enough data. Generally speaking if you are in a satisfactory region, someone wants to redesign the system and put you in a questionable one; this is a consequence of getting the most for the least. Under these trying circumstances, this section can be used only as a guide, but the idea of sensitivity to the exact model is very useful and can give the engineer more engineering confidence in his decision. In this kind of situation you must beware of the statistician's use of the word confidence, since in that use it is a very technical term and certainly does not mean engineering confidence. You can easily have one without the other. It is engineering confidence in a decision that an engineer wants--not statistical confidence per se.

References

11-1. Anon: ASTM List of Publications. Published Annually by the American Society of Testing and Materials.

11-2. Vaccaro, J.: Reliability Physics Notebook. RADC Reliability Physics Notebook, October 1965, RADC-TR-65-300 USAF, AD 624 769.

11-3. Dodson, G. A.; and Howard, B. T.: High Stress Aging to Failure of Semiconductor Devices. Proceedings Seventh National Symposium on Reliability and Quality Control in Electronics, Philadelphia, Pennsylvania, January 1961, pp. 262-272.

11-4. Pratt, R. E.: Profiles in Failure Rate vs. Applications. Semiconductor Reliability, edited by J. E. Shwop and J. J. Sullivan, Engineering Publishers, Elizabeth, N. J., 1961, pp. 27-37.

11-5. Carroll, J. M.: Reliability--1962. Electronics, vol. 35, November 30, 1962, pp. 53-76.

References (Continued)

11-6. Allen, W. R.: Accelerated Life Tests of Items with many Modes of Failure. Technical Report No. 44, Statistical Techniques Research Group, Section of Mathematical Statistics, Dept. of Mathematics, Princeton University, Contract #DA-034-ORD-2297, February 1962, AD 276 743.

11-7. Bessler, S.; et. al.: An Optimal Sequential Accelerated Life Test. Technometrics, vol. 4, August 1962, pp. 367-379.

11-8. Chernoff, H.: Optimal Accelerated Life Designs for Estimation. Technometrics, vol. 4, August 1962, pp. 381-408.

11-9. Pettinato, A. D.; and McLaughlin, R. L.: Accelerated Reliability Testing. Proceedings Seventh National Symposium on Reliability and Quality Control in Electronics, Philadelphia, January 1961, pp. 241-252.

11-10. Zierdt, C. H., Jr.: Comparison of Operating Life Tests and Storage Tests. Semiconductor Reliability, edited by J. E. Shwop and H. J. Sullivan, Engineering Publishers, Elizabeth, N. J., 1961, pp. 153-167.

11-11. Starr, W. T.: Relationships of Corona to Life and Reliability of Polyethylene Insulated Cable (New Test Methods for Cable). Presented at the IEEE Winter General Meeting, New York, January 27 to February 1, 1963 IEEE Transactions Paper No. 63-180.

11-12. Davidson, K. W.; and Johnson, E. E.: Reliability Evaluation by Application Derating. Proceedings Eighth National Symposium on Reliability and Quality Control, Washington, D. C., January 1962, pp. 204-216.

11-13. Starr, W. T.,; and Endicott, H. S.: Progressive Stress--A New Accelerated Approach to Voltage Endurance. Transactions AIEE, vol. 55, part III (Power Apparatus and Systems), August 1961, pp. 515-522.

11-14. Gibson, W. C.: Life Prediction of Diffused Germanium Transistors by Means of Power Stress. Physics of Failure, vol. 4, June 1966, edited by M. E. Goldberg and J. Vaccaro, pp. 49-73, AD 637 529.

11-15. Walker, M.; et. al.: Accelerated Aging and Failure Mechanism Analysis of Thin Tantalum Film R-C Networks. Physics of Failure, vol. 4, edited by M. E. Goldberg and J. Vaccaro, June 1966, AD 637 529, pp. 179-209.

11-16. Shurtleff, W.; and Workman, W.: A Limitation to the Step Stress Testing Concept for Integrated Circuits. Physics of Failure, vol. 4, edited by M. E. Goldberg and J. Vaccaro, June 1966, AD 637 529, pp. 258-278.

11-17. Hines, L. D.: Acceleration Factor Determination on Glass Capacitors. Proceedings Eighth National Symposium on Reliability and Quality Control, Washington, D. C., January 1962, pp. 217-221.

11-18. McLean, W. E.; et. al.: Acceleration Factor Determination and Reliability Analysis of Precision Film Resistors. Proceedings 1963 Electronic Components Conference, Washington, D. C., May 1963, pp. 4-11.

11-19. Schmidt, D. F.: Performance and Reliability of Wrapped and Plotted Mylar Capacitors. Proceedings 1963 Electronic Components Conference, Washington, D. C., May 1963, pp. 142-149.

References (Continued)

11-20. Goldsmith, A.; and Yanis, E. M.: Determining the Most Probable Failure Rate by Relating Life Test Results and Application Conditions. Semiconductor Reliability, edited by J. E. Shwop and H. J. Sullivan, Engineering Publishers, Elizabeth, N. J., 1961, pp. 247-253.

11-21. Duncan, G. I.: A Report of the Proposed IEEE Test Procedure for the Evaluation of Insulation Systems for Electronic Power Transformers. Proceedings 1964 Electronic Components Conference, May 1964, pp. 279-289.

11-22. Campbell, F. J.; and Brancato, E. L.: Determination and Application of Thermal-Life Characteristics of Aerospace Wires--Parts 1 and 2. Insulation, vol. 9, a) October 1963, pp. 17-22; b) November 1963, pp. 23-30.

11-23. Brauer, J. B.: Microelectronics and Reliability Assessment. Electronic Industries, vol. 25, April 1966, pp. 67-71.

11-24. Adams, J. D.; and Holden, P. L.: What is Meaningful Integrated-Circuit Reliability? Electronic Design, vol. 13, November 22, 1965, pp. 54, 56, and 57.

11-25. Odenweller, Richard L., Jr.: An Aging and Testing Facility for High-Reliability Transistors and Diodes. The Engineer, July 1967, pp. 20-29.

11-26. Endicott, H. S.; and Zoellner, J. A.: A Preliminary Investigation of the Steady and Progressive Stress Testing of Mica Capacitors. Proceedings Seventh National Symposium on Reliability and Quality Control in Electronics, Philadelphia, Pennsylvania, pp. 229-240, January 1961.

11-27. Evans, R. A.: Literature Review Study on Accelerated Testing of Electronic Parts. Final Report, for Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, Research Triangle Institute, Research Triangle Park, N. C., under Contract #951727, (In Preparation).

11-28. Clark, George L., ed.: The Encyclopedia of Chemistry. Rhinehold, New York, 1957.

12. Testing in the Operational Environment

It is often said that the only way to really determine whether an item will function as intended in its operational environment is to put it there and see. This is basically true but it is not always possible to do this prior to the real mission nor is it necessarily desirable. One significant shortcoming of this approach, for example, is that certain cause and effect relationships may be obscured because of lack of knowledge about operating conditions.

Field tests, flight tests, commercial marketing tests, and in-service evaluation tests are all examples of testing in the operational environment. Typical reasons for wanting to test in this manner are:

(1) increase confidence in the ability of the item to perform in the actual operational environment,

(2) inability to simulate particular environmental conditions in the laboratory,

(3) the item is too large and complex for environmental simulation, and

(4) obtain response data as a basis for future laboratory tests.

Sometimes it may simply be less expensive or easier than simulating conditions. However, if the capability to simulate conditions already exists, it is usually less expensive to use environmental simulation. As illustrated by the second and third reasons above, there is often no choice but to resort to the operational environment.

There are often certain shortcomings in the approach. A simple flight test of a launch vehicle, for example, gives response to only a particular set of conditions and the behavior is not known to be representative of the population of operational items.

The extent of other shortcomings depends considerably upon how much control is maintained over the tests. If the items are merely put in the hands of customers to operate and checked from time to time, the results may only indicate how well they survived this environment. Items may get used for purposes other than those intended or operated under conditions not included in the design criteria. Unless these are known, the results can have only limited utility. An even worse situation occurs when the reporting of behavior is left to the customer. Usually he considers it the least important of his jobs and it often gets done poorly if at all.

Even when more control is maintained by the manufacturer, there can be disadvantages. Some cause and effect relationships may be obscured because of lack of detailed knowledge of the conditions. Measurements are often not as thorough or as accurate. There are typically delays in reporting results.

When the individual items are simple and inexpensive such as hand tools, turning them over to the customer for testing in his environment can be a good way of evaluating the product. When they are expensive as aircraft, usually the controlled approach is better. Basic procedures for planning such tests are no different than others. The rewards must be evaluated with respect to costs in time, money, and effort.

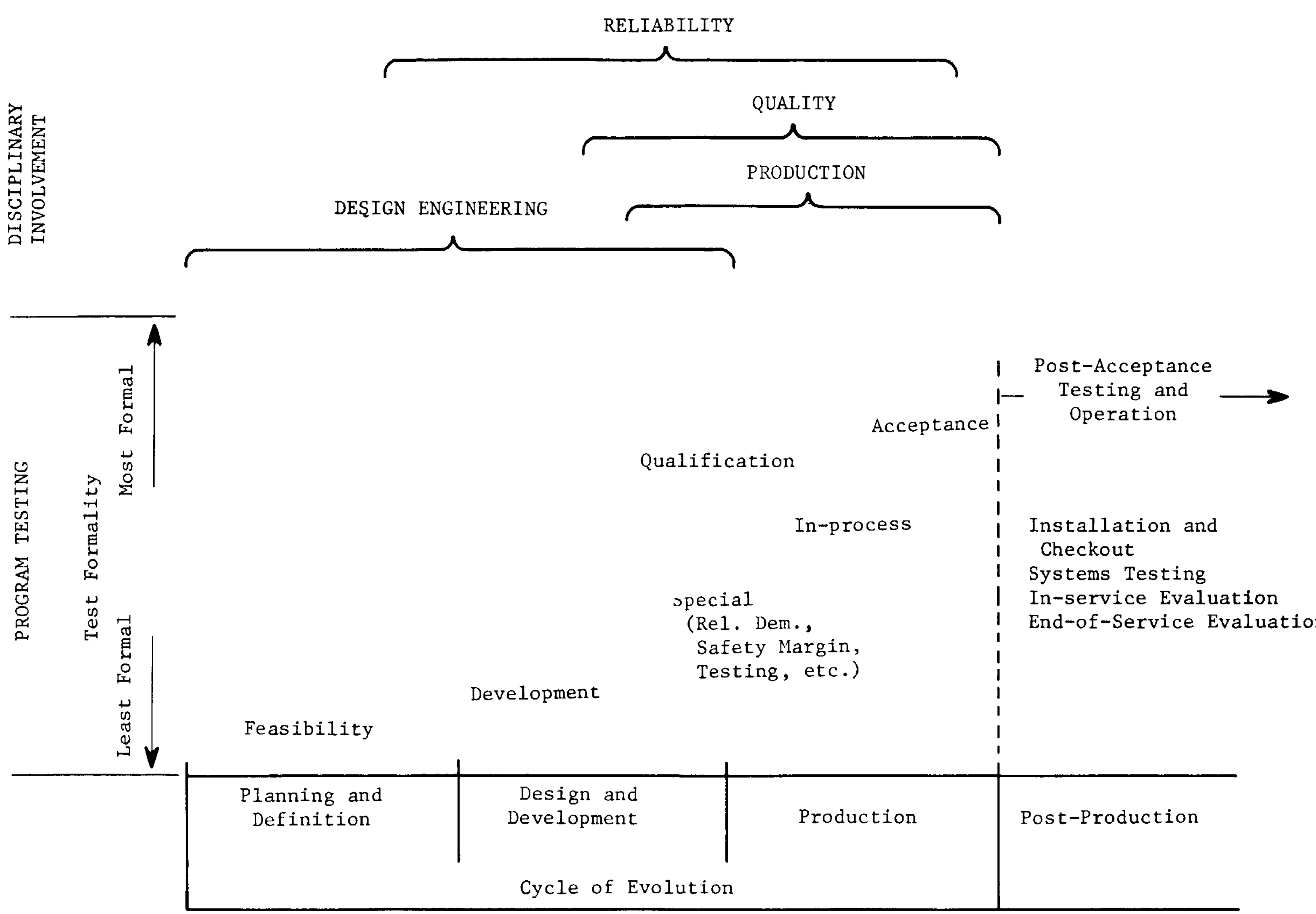


Figure 13-1 Testing in Hardware Development

13. Major Applications of Testing in Hardware Programs

Three important classifications of tests were introduced in Sec. 3.1. Two of these, the basic test types and the basic uses of tests, are discussed in Secs. 5 through 8. This section treats the third, viz., major applications of testing in a hardware development program.

As a design evolves from its initial concept through design and production to final operational form, numerous tests are required. The tests at a particular stage can usually be associated with some general problem area such as development, design qualification, and verification of the final product. Whether the program is concerned with developing a piece-part, an equipment, or a large system, the types of problem areas to be treated are basically the same.

An illustration of how these applications relate to product evolution is presented in Fig. 13-1. The designations such as feasibility, development, qualification, etc. are common terminology found in most programs. One has only to sample the literature to discover that their meaning differs quite radically from program to program. The meanings we have assigned to these is made clear in later discussion. Instead of rigid definitions we have generally compromised on concepts defined or implied by a number of sources including especially Refs. 13-1 through 13-7.

Note that in this representation the cycle of evolution ends following production; the testing and operation following production is associated with either the evolution cycle of a higher level of assembly or the operational use of the end-item. For example, installation and checkout are simply in-process activities during fabrication of higher levels of assembly. Even though the manufacturer's major attention to a product may terminate with customer acceptance, the post-production tests and operations can often provide good feedback information for improving other items in production or aiding new or modified designs. Some programs even provide for field personnel in support of this.

As illustrated, testing generally becomes more formal as the design matures. Formality indicates the degree to which the plans for carrying out the test and for documenting results are explicitly controlled by program management or the user. Early in a program the tests are usually less formal because of their required exploratory nature and their tendency to provide little information to the user or customer on the verification of the end-item. The most formal is acceptance testing which provides final verification (within its ability) to the customer that the end-item conforms to his application requirements.

The omission of test designations such as reliability, quality, and production testing may surprise some readers. These designations are mainly discipline-oriented; production testing is essentially synonymous with in-process testing but reliability and quality testing may apply to any of several applications. This depends upon which discipline is assigned the major cognizance over an application. This varies from program to program and as noted at the top of Fig. 13-1, the possible areas of involvement by the various disciplines or organizations is far-reaching.

Typically in most programs the engineering organizations responsible for design and development also maintain control over the earlier tests in the program. As the design progress toward final form, more and more influence is exerted from reliability, quality, and production personnel. Design qualification is usually controlled by quality or reliability personnel and represents the major point of transition of responsibility of the design from engineering to production.

The major influence of quality control organizations traditionally begins in the design and development stage in connection with selection and qualification of components. The subsequent design qualification, component acceptance, in-process, and end-item acceptance testing are all usually controlled by quality control personnel even though the actual conduct of these tests may be by engineering and production personnel.

The influence of the program's reliability organization is dependent upon the emphasis on reliability in the program and the approach to its treatment. Testing is a source of data for reliability assessment. Until the last few years, most of the reliability data for assessing a specific design was achieved by special tests for reliability conducted quite distinctly from others and was designed and controlled mainly by reliability personnel. More recently there has been emphasis on considering all areas of testing as potential sources of reliability. This has necessarily involved modifying some of these and incorporating some of the test design features previously provided only by the special reliability tests. The integration of these tests to provide the necessary reliability information requires some overall program test and evaluation plan. All high reliability programs now require this in one form or another. This concept of integrating tests is further discussed in Sec. 13.8. Section 14 pursues the specific problems of testing for reliability in more depth.

The nature of the testing for each area of application results from a trade-off between the information desired and the information feasibly obtained within the constraints of available models, time, cost, etc. For example, a test to obtain a reasonably accurate estimate of reliability of the end item requires test specimens similar to those intended for operational use; it may also require testing for a long duration and in a complex, simulated environment. Such test specimens will not usually be available until late in the design and development stage or early in the production stage.

Remember also that a particular application may involve several of the basic testing approaches introduced earlier in this report. For example, qualification testing may involve proof, performance, and time-truncated testing to fulfill all objectives.

13.1 Feasibility Testing

Prior to proceeding at "full speed" in developing a particular design concept, it is first necessary to verify its feasibility or to select the best of several approaches. Informal testing is often performed in support of this and such tests are usually called feasibility tests. It often involves testing existing (but perhaps refurbished) spare hardware from another program under conditions different from that originally intended. If hardware has to be fabricated for the testing it usually consists of crude breadboards and other special engineering models.

Feasibility testing may require any of the basic test types and emphasize simple functional performance in some cases, and life or strength in others. Because of its informality much of the data gets no further than laboratory notebooks. It generally is not a major source of data for quantitative reliability assessment; its major contribution to reliability is in the (engineering) confidence it provides about the accepted design approach. Statistical design of feasibility tests is rarely done.

13.2 Development Testing

Following feasibility investigations and selection of design concepts to be developed more fully, much additional testing is necessary during the design and development for generating design data or resolving specific design problems. Typically, there may be several design approaches to be further evaluated and compared on some basis or a designer may need to learn more about the operation of his design in a particular environment. Test items for such development tests are typically breadboarded circuits, boiler-plate structural models, and other special engineering models. In some programs development testing may extend to include tests with pilot production models of operational hardware.

The responsibility for design and control of much of the development testing programs usually rests with design personnel. These tests are usually very informal and tend to be exploratory or probing in nature; most data is generally retained in laboratory notebooks. There is now more emphasis in programs involving design of high-reliability and particularly few-of-a-kind items to utilize development testing as a reliability data source to the extent practical. Such integration of test applications is discussed in Sec. 13.8.

The problems treated in design and development are many and varied in nature; hence, one can readily find applications of all the basic testing approaches introduced earlier. With the exception of specially integrated tests for reliability, statistical design is generally not as prevelant as in later applications. Most of the tests are devoted to building up engineering knowledge (hence confidence) about the design.

One major word of caution to the designer is in order: Beware of nonreversible effects. After a series of tests under extreme severities of conditions plus all of the necessary handling and jiggling of wires and components, your model may be quite different than what you started with due to hidden damage and aging.

13.3 Qualification Testing

Before proceeding with full-scale production of operational items, the manufacturer and often the user want to know that the design and its component parts are inherently capable of meeting all of the application criteria. Qualification testing aids experimentally in doing this. It has long been a very formal procedure.*

In some cases designers conduct the qualification testing. It is preferable, however, if other groups conduct them to eliminate possible bias.** Separate test and evaluation groups sometimes serve this role. Some organizations consider it an integral part of reliability or quality testing. In an integrated testing approach it can be a valuable source of data for reliability assessment.

Test articles designated for design qualification testing are usually as near like the final operational item as is practical. In the qualification of a component of a space system the test article may in all noticeable respects be identical to the intended flight articles. In less critical programs special engineering models, perhaps even a refined breadboard version, may be adequate for qualification. In some cases only one specimen may be available and hence is usually treated with the utmost care. It is worthy to remind the reader that even though a small sample of one or two items may represent a large proportion of a population of kew-of-a-kind items, it is not necessarily a representative sample. With smaller, less expensive items such as piece-parts, larger samples are easier to obtain but remember here that the sample you purchase today is not necessarily representative of the ones you will purchase several weeks hence.

Most qualification tests employ conditions which at a minimum are as severe as those for usual operation. Extensive laboratory simulation of operational environments

---

* In the development of high-reliability items it is complemented by other efforts such as design reviews, part application reviews, reliability prediction, and vendor qualification for verification of design adequacy.

** As observed in several programs, special in-house tests are conducted by designers prior to qualification using identical procedures in order to assure themselves that the item will pass the qualification test.

is often involved. Field and flight testing is employed to achieve the conditions when they cannot be adequately simulated. Because of the severity of the conditions, tested items are usually considered unfit for further application as components for higher-level assemblies or as operational hardware.

Lower level components of the design are usually qualified prior to assembly. When there are special designs developed in separate programs, the basic qualification procedure for these is the same as described above. Off-the-shelf components such as piece-parts may have already been qualified per some specification and appear in a qualified (or preferred) parts list. The designs of such components generally experiences a qualification in the manufacturer's development program before release for production. The qualification for application of these in a specific design is often performed again by the user but will generally employ the same approach. The conditions for this latter qualification are typically those rated by the manufacturer for his component and due to derating by the manufacturer may not be as severe as those he originally used for his qualification. Remember, however, that not all manufacturers are equally conservative in their derating practices.

Qualification testing most commonly employs proof testing and time-truncated testing approaches. In the former, test conditions are applied at a fixed severity and the item either passes or fails the test. In the time-truncated approach the item is tested for a fixed time under certain conditions. Performance measurements are also commonly made in connection with these. A complete qualification test of an item may involve a number of individual tests and measurements under various conditions. If reliability demonstration is integrated into the testing procedure or if one-shot items are being treated, then some sampling plan (see Sec. A.10 of the Appendix) may be involved.

General guidelines for qualification testing of space system components can be found in Refs. 13-4 and 13-8; the latter also contains typical conditions for qualification. A general discussion of the philosophy of qualification testing for a specific spacecraft and the choice of conditions is presented in Ref. 13-9. Numerous NASA specifications and standards on qualification testing of specific components can be identified in Ref. 13-10; Refs. 13-11 through 13-14 are typical examples of these.

In typical qualification test specifications the specific objective of time-truncated tests is not always clear, i.e., whether the test is to look for possible aging or cumulative damage effects or is simply another form of proof testing. If the conditions closely simulate those of a mission, for example a random vibration profile representing the launch environment, then it can be either but neither does it matter. However, there are numerous examples of tests, for example one in which several high-low temperature cycles are applied, in which the conditions are very different from those for intended operation. It is usually true that if failure does occur under such conditions it does indicate a possible inherent weakness in the design. Tests of this latter type are often misleading as to the cause of failure. They are also not suitable as a source of failure-time data for estimating reliability.

It is often necessary to requalify items when the design has been modified or the fabrication process has knowingly changed. This is often the case with piece-parts and the manufacturer does not always make the customer aware, nor is he always aware, of changes in the process. Increased incidents of failure during the customers receiving tests or during later operation of the item may be the only indicator. Some high reliability programs specify a periodic requalification of certain parts; some specify tests of each lot received which if are not requalification in name are in essence. Requalification of larger items may involve waiver of a portion of the original qualification test stages.

13.4 Acceptance Testing

Whereas qualification testing verifies that a design or a component type is inherently capable of performing under intended conditions, it cannot verify that the end-items from production have the same capability. Acceptance testing plays a very formal role in providing this verification.* It is a major step in the transfer of a product from the manufacturer to the customer or user. Typically, a manufacturer of high-reliability items is required to have a very extensive acceptance test and inspection procedure for incoming parts and materials.** Also, formal acceptance tests are required for fabricated items prior to acceptance by the customers.

The major responsibility for acceptance testing is practically always assigned to quality control personnel. In some cases acceptance testing may provide data which will aid the reliability organization in refining their evaluation of the item. A well-planned integrated testing approach would consider the use of the results of acceptance tests of lower level items used in assembly for reliability assessment.

Most acceptance testing is performed using a go, no-go test approach wherein the conditions are chosen so that bad items fail and are rejected and good items pass and

---

* Important complementary efforts are quality control measures during fabrication of the design and in procurement of lower-level items; in some cases with procured items the customer may insist on his performing continual inspection of the manufacturer's process.

** There is a trend toward having the manufacturer perform more of these tests prior to shipment.

are available for application. For space equipment Ref. 13-2 specifies that the tests will be performed in such a manner that will simulate operational conditions to the highest degree practicable without damage to the end-item and which will provide a measure of over-all quality. A major intent with much acceptance testing is to reveal any possible material defect and workmanship discrepancies in the end-item.

Some acceptance tests are known to consist only of simple functional checks in a benign laboratory environment. More often, however, there is some attempt to stress the item to cause defects and deficiencies to be revealed. The procedures are usually somewhat similar to conventional qualification testing in that the basic testing approaches are proof testing and time-truncated testing. A major difference, however, is that the stresses are usually chosen to be less severe than in qualification testing to prevent damaging good items. Some good discussion on acceptance testing of a specific spacecraft is presented in Ref. 13-9 and since qualification testing is discussed there also it provides a good comparison of the procedures and conditions for the two applications. One can identify numerous specifications and standards on acceptance testing also in Ref. 13-10. References 13-11 through 13-14 are also good examples of these.

There is usually some type of acceptance test performed on every item larger and more complex than piece-parts. In high reliability programs there is even now considerable emphasis on acceptance testing 100% of all piece-parts; the term screening is used to designate this and the subject is treated in more depth in Sec. 14. When it is necessary to destroy or significantly age items in order to test them, it must be done by sampling (See Sec. A.10 of the Appendix for discussion on sampling).

13.5 In-process Testing

This pertains to all of the testing performing during fabrication. Most of it is done as part of the routine quality control inspection and testing to maintain quality standards of the process procedures and the workmanship. Production personnel typically perform most of the testing with quality control sampling periodically to maintain the checks and balances. Production engineers may need others simply as part of their assembly procedure as in the case of an alignment test of a gyro after it is installed.

For quality control purposes it is generally desirable to conduct some test after each step of assembly to check the items of hardware involved. Such tests are usually kept as simple as possible but they must be adequate to do the job. It typically involves simply attributes measurements such as visual inspection and functional checks but may in some cases require a proof test approach. Variables data is often recorded to determine the amount of variability. Sometimes the variation at one step is combined or traded off with variation of other steps to assess the overall variability and maintain it within requirements.

It is not always possible to measure the characteristics of interest by conventional methods without destroying or significantly aging something. This has promoted the use of nondestructive testing (NDT) (see Sec. 9 for a discussion of NDT principles) methods in in-process testing. When there is no alternative to destroying or aging the item, destructive testing must be done on a sampling basis. If many identical items are produced, sampling can be performed in accordance with MIL-STD-105D [Ref. 13-15]. It is often desirable, especially with low-volume production, to make the sample sizes dependent upon the results of previous tests; that is, the sample size will decrease if few or no failures are observed. Section A.10 of the Appendix of this report gives a summary treatment of sampling procedures. For description and procedures of various approaches to sampling, consult Refs. 13-16 and 13-17.

In-process testing treats many characteristics that cannot be checked by usual acceptance testing of the end-item. It is thus worth doing well. It is not traditionally a source of reliability data for input to reliability assessment models. There is no argument, however, about its potential contribution to reliability through the control measures it provides; lack of sound in-process testing and other quality control measures can only degrade the reliability that was originally designed into the item.

13.6 Special Tests

In some program there may be special applications of testing not covered in the earlier categories. Special reliability demonstration testing is an example of this for those programs not integrating this in with development or qualification testing. Reliability demonstration is just one form of reliability testing. This and other forms are discussed in Sec. 14.

It is not uncommon to find special designations of certain test programs in some hardware programs; some noted are environmental test, evaluation test program, engineering specification test, verification tests, proof test program, and safety margin test. Most often these belong to one of the other categories as defined earlier.

13.7 Post-Acceptance Tests

As noted earlier, post-acceptance tests refer to all tests conducted after customer or user acceptance. As indicated in Fig. 13-1 installation and checkout testing, systems testing, and in-service and end-of-service evaluation are typical of these. In these the product in question is simply a component of a higher-level assembly and/or is being applied in its intended application. The installation and checkout testing and the systems testing may thus represent in-process testing and qualification or acceptance testing of a higher-level assembly.

The field experience from testing and operation in the post-production period can, in many cases, be very valuable as an indicator of quality and reliability. If more items are to be produced the information may aid in determining necessary modifications in the design or in the fabrication process. Fabricated end-items are often modified as a result of the experience. The experience can also be useful for similar designs in later programs.

Most of the data from the post-production period is attributes data. It also possesses many of the same undesirable characteristics as those mentioned for data from usual field testing (see Sec. 12); typically there is often uncertainty of what really causes a failure or a particular response.

Many of the tests during the post-production period involve the item in question only indirect. For example, a systems test concerned mainly with overall performance of the system provides only a functional check of lower level components. For the most part, the specific behavior of lower-level components is only of interest if the system fails to function properly and the source of discrepancy is being traced. In some cases behavior of some individual components, especially critical ones, may be monitored separately. For example, separate drift tests of gyros after installation in stable platforms are very common and such tests provide valuable field data to the gyro manufacturer. The location of test points in a system determines also how well the behavior of an individual components can be observed.

For some programs there may be a planned in-service and end-of-service evaluation procedure. This is influenced largely by the maintenance and replacement policy. On-line testing provides an in-service evaluation of how well items are performing. "Post mortem" tests and analysis of failed or worn-out items are sometimes performed as an end-of-service evaluation to determine why and how items fail. There may also be some retest of successfully operated items after completion of a mission; however, this is usually less frequently done.

Much effort is expended on testing during the post acceptance. Many people are involved in it but there is relatively little in the literature about it except under the subject areas of the categories introduced earlier when they apply. This is largely because of the inappropriateness of statistical designs and explicit procedures. Some useful information can be found under subject headings such as failure reporting, field data and experience data. Some good sources for a start are Refs. 13-18 and 13-19.

13.8 Integrated Testing

Integrated testing pertains to coordinated planning and implementation of all testing in a program to satisfy all data requirements in the most efficient manner. It involves all major applications of testing such as development, qualification, reliability demonstration and acceptance testing and the needs of the various data users such as designers, reliability personnel, quality personnel, production engineers, system analysts, and customers. Ideally, it coordinates the tests of all levels of hardware complexity from the parts and materials level to the total system.

The major motivation for integrated testing is efficiency. Prior to its emphasis much of the testing was done as a result of piece-meal planning; individual groups planned and implemented tests mainly on the basis of their own needs. As systems became more complex and programs larger, there was simply the need to more objectively improve the return on investments in testing. Elimination of many duplications typical of conventional test programs contributes greatly toward this.

A distinct advantage in integrated testing has proven to be in communication. Through coordinated planning among different groups and sharing of information, the problems and needs of the different groups become more commonly known promoting an atmosphere of cooperation among them.

Reliability evaluation has probably benefited from integrated testing more than any other activity and has also served as a major motivating force for it. NASA Reliability Publication NPC 250-1 (Ref. 13-1) has been a major instrument in promoting integrated testing for reliability. It specifies that a reliability evaluation plan be established in a program and an integrated test program be conducted in parallel to serve as the major source of reliability data. It specifies that the plan shall include all tests of major components, except those for feasibility, and that the requirement for various types and degrees of testing shall be based on the reliability prediction and assessment models. Similar provisions for military equipment programs are given in Ref. 13-5.

Integrated test planning is thus influenced heavily by reliability data needs. This is brought out more clearly in the next section on reliability testing and several articles on further discussion and application are cited there.

References

13-1. Anon.: Reliability Program Provision for Space System Contractors. NASA Reliability Publication NPC 250-1, July 1963.

13-2. Anon.: Quality Program Provisions for Space System Contractors. NASA Quality Publication NPC 200-2, April 1962.

References (Continued)

13-3. Anon.: Inspection System Provisions for Suppliers of Space Materials Parts, Components, and Services. NASA Quality Publication NPC 200-3, April 1962.

13-4. Anon.: Apollo Test Requirements. NASA Publication NPC 500-10.

13-5. Anon.: Military Standard, Requirements for Reliability Program (For Systems and Equipments). MIL-STD-785, Department of Defense, June 1965.

13-6. Smiley, Robert W.: Testing Programs. Section 8 of Reliability Handbook, W. Grant Ireson, ed., Mc-Graw Hill Book Co., New York, 1966.

13-7. MacCarley, John A.: Testing for Reliability. Part II of Reliability Control in Aerospace Equipment Development, SAE Technical Progress Series, vol. 4, The Macmillan Co., New York, 1963.

13-8. Anon.: Qualification Test Program Guidelines. NASA Publication ASPO-RQA-11, NASA Manned Spacecraft Center, Houston, Texas, May 1965.

13-9. Shipley, W. S. and Maalay, J. E.: Mariner 4 Environmental Testing, Astronautics and Aeronautics, August 1965.

13-10. Anon.: NASA Specifications and Standards. NASA SP-9000, May 1967.

13-11. Wagner, J.: Advanced Technological Satellite Project Technical Requirements-Environmental Qualification and Acceptance Test Specification-Spacecraft Testing. NASA Spec. No. 5-320-ATS-0101, NASA Goddard Spacecraft Center, Beltsville, Md., July 1964.

13-12. Wagner, J.: Advanced Technological Satellite Project Technical Requirements-Environmental Qualification and Acceptance Test Specification-Component Testing. NASA Spec. No. S-320-ATS-0102, NASA Goddard Spacecraft Center, Beltsville, Md., July 1964.

13-13. Anon.: General Specification, Semiconductor Devices, Established Reliability. NASA Spec. No. MSFC-SPEC-438, Marshall Space Flight Center, Huntsville, Ala., Sept. 1964.

13-14. Anon.: Specification Sheet, Semiconductor Device E2N657A, Established Reliability. NASA Spec. No. MSFC-SPEC-438/100, Marshall Space Flight Center, Huntsville, Ala., May 1965.

13-15. Anon.: Sampling Procedures and Tables for Inspection by Attributes. MIL-STD-105D Department of Defense, April 1963.

13-16. Enrick, N. L.: Quality Control and Reliability. The Industrial Press, New York, Fifth Ed., 1966.

13-17. Duncan, A. J.: Quality Control and Industrial Statistics. Richard D. Irwin, Inc., Chicago, Ill., 1952.

13-18. Ireson, W. G., ed.: Reliability Handbook. McGraw-Hill Book Co., New York, 1966.

13-19. Walton, Thomas F.: Technical Data Requirements for Systems Engineering and Support. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1965.

14. Reliability Testing

In a broad sense reliability testing pertains to any test conducted to generate information on reliability characteristics. As a whole, it treats several problem areas. Typical of these are:

(1) reliability demonstration,

(2) reliability estimation,

(3) investigation of failure modes and mechanisms,

(4) reliability screening,

(5) reliability growth, and

(6) performance variation.

It it thus more than just reliability demonstration or just life testing as is often inferred in the literature.

Because of this diverse nature reliability testing overlaps in many ways with other discipline-oriented tests such as engineering, production, and quality testing. For example, a designer conducting a test to determine a cause of failure or measure the sensitivity of performance to changes in an environmental factor is generating reliability information. Also, quality control personnel designing a sequence of screening tests for acceptance testing of incoming parts are very much concerned about reliability.

A major impetus for formal reliability testing resulted from the report by the Advisory Group on Reliability of Electronic Equipment (AGREE) in 1957 [Ref. 14-1]. Of the nine task groups, Task Groups 2 and 3[*] were concerned primarily with reliability testing and more specifically, the problem of reliability demonstration. Whereas the specific recommendations of AGREE have not been closely followed in subsequent specifications and standards, their attention to the reliability testing problem served as a significant stimulus in bringing about the development of many of the procedures in use today.

The different problem areas of reliability testing introduced above are not necessarily mutually exclusive. For example, certain reliability demonstrations can involve elements of reliability estimation, performance variation, and even screening. Also, a point on a reliability growth curve results from a point or interval estimate of reliability. The treatment of each problem area requires the use of one or more of the basic testing approaches introduced in Secs. 5 through 8. These problem areas are discussed in more detail in the following sections.

---

[*] Task Group 2 was concerned with testing in the development phase and Task Group 3, the production phase.

14.1 Reliability Demonstration

When reliability is an important trait of a product, the customer or user usually wants sound evidence that the reliability meets his requirements. Reliability demonstration pertains to the testing and subsequent analysis performed to supply such evidence. It is the most popular and written-about area of reliability testing. As noted above, it was also the major area of emphasis of the AGREE groups concerned with testing.

The general problem statement for reliability demonstration is: A reliability R_L is to be demonstrated with confidence γ. R_L represents a lower one-sided confidence limit on the true reliability R such that there is a probability of γ that $R > R_L$. The quantity, $1-\gamma$, represents the risk of the statement being in error. Alternate forms of this statement can be made on the basis of mean life, strength, safety margins, failure-rate, or other reliability parameters rather than the reliability per se. Two-sided confidence limits on the parameters are also permissible; however, the one-sided limit is the most used form for this purpose.

Reliability demonstration thus consists of collecting the information necessary to accept or reject a statement about the level of reliability. It is basically an hypothesis testing procedure. As will be discussed later, it is not always possible to formulate and test an hypothesis in the formal statistical sense; such a demonstration then relies strongly on engineering judgment and engineering confidence.

14.1.1 Perspective

The general role of reliability demonstration in a hardware program is illustrated in Fig. 14-1 along with other factors that influence how it is performed. Through the closed loop flow of information, it can serve as a beneficial evaluation tool.

The Tests

The reliability demonstration tests involve any one of several basic testing approaches depending upon the nature of the reliability problem. A summary of the applicability of basic test types described in Secs. 5 and 7 to reliability demonstration is presented in Table 14-1. Generally, there are specific parameters of characteristics such as probability of success at time t, mean life, safety margin, mean value of a parameter, and per cent in-tolerance on which the demonstration or accept-reject decision is based. These are usually specified by the reliability requirements and are related to reliability of the item through the applicable

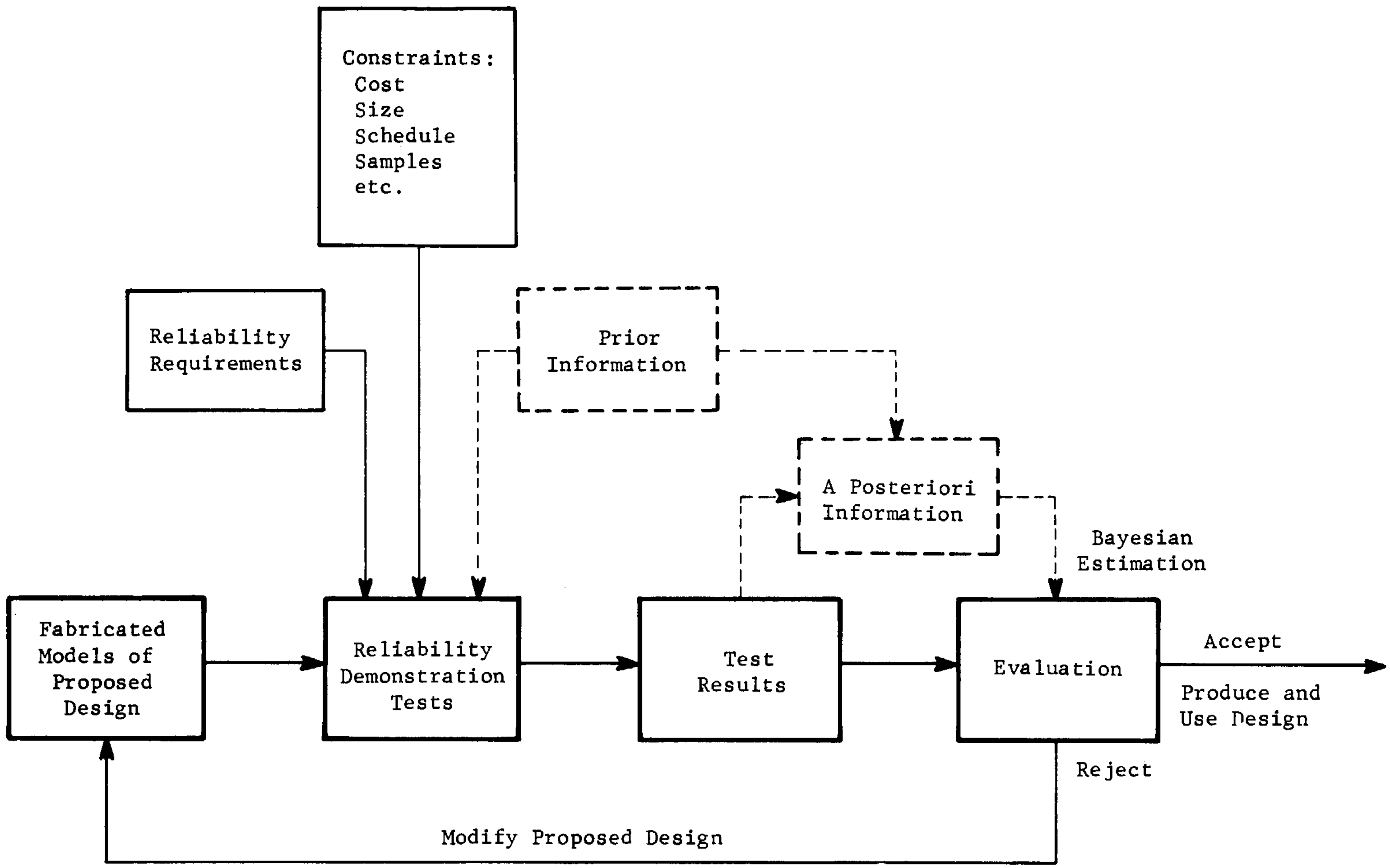


Figure 14-1. Role of Reliability Demonstration

Table 14-1

Basic Approaches to Reliability Demonstration

Reliability [a/] Model Basis	Typical Demonstration Parameters	Basic [b/] Test Types
Life	Probability of success at any time t Per cent success (or failure) at fixed time T Mean life Failure rate Mean-time-between-failure	Failure-truncated testing Time-truncated testing Sequential testing
Strength	Probability that strength exceeds stress Per cent surviving (or failing) at fixed stress S Mean strength Standard deviation of strength Safety Margin Percentile x_α of strength	Stress-to-failure testing Sensitivity testing Proof testing
Simple attributes/ variables	Probability that parameter is in tolerance Per cent nondefective (or defective) Parameter mean Parameter standard deviation	Simple measurements

a/ Different forms of models for reliability are described in Vol. IV - Prediction of this series. The categorization introduced here is slightly different but is more suitable for discussing reliability demonstration.

b/ Basic test types are defined and described in Secs. 5 and 7.

reliability model. For example, if a reliability model of the form $R(t) = e^{-\lambda t}$ is acceptable, the demonstration of an upper one-sided confidence limit λ_U on failure rate λ at confidence γ is equivalent to the demonstration of a lower one-sided confidence limit $R_L = e^{-\lambda_U t}$ at confidence γ on reliability at any time t.*

As illustrated in Table 14-1 three major categories of reliability modeling types are assumed for discussion of reliability demonstration problems. The <u>life</u> category is appropriate when failure is caused by cumulative damage;** most often, reliability is considered a function of time R(t). The <u>strength</u> category applies when failure is assumed caused by stress severity exceeding strength;*** reliability in this case is treated primarily as a function of stress and strength rather than time. The latter category is designated as <u>simple attributes/variables</u> to denote those measurements of items' characteristics when the effect of stress or time can be disregarded; typically an item's operation is represented by observed success as failure or a characteristic is represented by the value of some continuous parameter. Models for reliability employing these types of measurement are independent of stress or time.

There are obviously demonstration approaches which do not conform explicitly to the classification in Table 14-1. For example as described in Sec. 6.5 a measurement at specific points in time of residual strength of items sampled from a larger population being aged can be used to estimate a hazard function, a parameter of which, say the average, may in turn be the basis for a reliability demonstration. One should always be careful in such complicated cases to make sure he is well aware of just what is being demonstrated; otherwise, the results can be very misleading. It is also possible that a complete demonstration may require testing with two or more approaches. This would apply for example, when both adequate mean life and safety margins are required to be demonstrated.

---

* The direct substitution of a point estimate into a function to get a point estimate of the dependent variable is always valid when the argument contains only one parameter which is allowed to be variable and when the function is monotonic; it is not generally valid when the function is nomonotomic nor when two or more parameters are treated as variables.

** The concept of cumulative damage as a cause of failure was introduced in Sec. 2.2.2. Specifically, the emphasis was actually on stress and cumulative damage as distinct causes of failure. The difference has been noted numerous times in this report when appropriate and most significantly in Secs. 5 through 8 as it influenced the categorization of basic testing approaches.

*** The concept of stress as a cause of failure was also introduced in Sec. 2.2.2.

Test Conditions

It is often stated or strongly inferred in discussions of reliability demonstration that simulation of the normal operating environments is very necessary and important. Actually it depends upon how the conditions affect the reliability. For measurements belonging to the latter category in Table 14-1, the choice of test conditions is generally not very critical. Don't forget, however, that even though specific factors such as high temperature and low supply voltage may be discounted as important, other test conditions such as clean atmosphere or preconditioning of the test specimen may be very necessary.

At the other extreme there are situations when the conditions for the test may be very important. When failures can be influenced by operating conditions and result from cumulative damage, the only way of truly demonstrating reliability is by performing the test with the conditions accurately simulating those during normal operation. When this is possible, the achieved confidence is then dependent only upon the variation in the population of items tested and on the sample size. Any departure of the test conditions from those or normal operation can only tend to modify the confidence in the demonstrated reliability. It is of course impractical in many cases to simulate complex profiles or different environmental factors simultaneously. For example, a particular piece-part type such as a 1N619 diode may find many different applications in a system resulting in many different application environments for the part type. Procurement practices of most programs would specify testing for reliability demonstration with only one set of conditions. Absolute worst-case or manufacturer's rated conditions are most often used and this would tend to provide some degree of conservativeness in the confidence of the demonstrated reliability. Generally the higher the level of assembly that the item represents, the more emphasis there is on accurately simulating the conditions of normal use. Limitation of capabilities for environmental testing of very large systems increases the need for accurately representing the operational environment when testing its components. The general discussion on environmental testing in Sec. 10 applies to reliability demonstration as well as other areas of testing.

Remember that if the cause of failure is stress severity exceeding strength, there is no need to generate profiles; one need only generate the stress severity required for a stress-to-failure, sensitivity, or proof testing approach, whichever method is being used. This can still involve having to generate different environmental factors simultaneously.

Various methods of simplifying conditions are acceptable for reliability demonstration as in other areas of testing. For example, worst-case or high-low level cycling is often employed when operational profiles are not known explicitly or are too complex to simulate. Some factors may be treated separately or eliminated altogether. Application of different conditions in sequence rather than simultaneously is often used. Whatever form of simplification is used, one should always demand that adequate justification and rationale be used either by sound engineering judgment or by prior experimentation. It is rarely easy to provide the full confidence initially desired. The problem is more than a statistical one, for statistical confidence achieved under test conditions can rarely be extrapolated directly to application conditions.

Evaluation

The evaluation (see Fig. 14-1) is simply the test of the hypothesis leading to the accept-reject decisions. The basic concepts of hypothesis testing are presented in Sec. A.7 of the Appendix; however, it is important to remember here that there are two basic approaches to testing an hypothesis and hence to reliability demonstration.

One approach is characterized by the question: "Given some observations, what can I say about accepting or rejecting the hypothesis?" In this case test results would be available in the form of values of the demonstration parameters. Then along with knowledge of such factors as sample size, test duration, and numbers of failures, the appropriate estimates are obtained and compared with the hypothesis. For example an estimated lower one-sided confidence limit of 302 hours for mean life θ at 90% confidence is adequate basis for accepting the hypothesis that $\theta = 300$ hrs. if the risk of making a wrong decision is allowed to be as high as 10%. If, however, the risk is allowed to be only 5%, then insufficient evidence exists to either accept or reject the hypothesis. One then either allows a greater risk or collects more data.

A less direct approach is characterized by the question: "Given the hypothesis to be tested, what observations do I need to make in order to accept or reject it?" This approach requires development of some sampling plan which specifies such factors as sample sizes, allowable numbers of failures, and test durations which are adequate to test the hypothesis. Suppose for example, that it is desired to demonstrate a mean life θ of 300 hours (i.e., test the hypothesis that $\theta = 300$ hrs.) such that the risk of erroneously accepting a true mean life of less than 300 hours is not greater than 10%. If, say, the available test time is limited, a time-truncated test may be used. A sampling plan can then be developed using basic statistical formulas for estimation. This consists of determining for the given test time the number of samples and the maximum number of failures during the test which will satisfy the hypothesis. The only test results required then are the observed number of failures; the demonstration is successful (i.e., the hypothesis is accepted) if the observed failures do not exceed the allowable number, and conversely, is unsuccessful (i.e., the hypothesis is rejected) if the observed failures do exceed the allowable number.

There has been much emphasis on the use of sampling plans for reliability demonstration. The basic concepts of sampling and sampling plans are discussed in Sec. A.10 of the Appendix. Lloyd and Lipow [Ref. 14-2] give a good treatment of the theory. To eliminate much of the repetitive effort in developing sampling plans for each separate sampling problem, a number of standard sampling plans have been developed. Such standard plans consist of tabulations and OC (operating characteristics) curves for relationships of such factors of probability of acceptance (or rejection), sample size, acceptance criteria, acceptance parameter values and test duration. Note that plans can be developed for practically any test or measurement situation and for any assumed form of the statistical distribution of the parameters to be measured.* Some emphasize the consumer's risk; some, the producer's; some may uniquely specify either or both; and some may allow a wide choice of risks.** Some plans allow more flexibility than others in selecting factors such as risk levels, sample size, etc.

Two of the earlier plans promoted specifically for widespread reliability use were those recommended by AGREE [Ref. 14-1] in 1957.*** Both plans were developed primarily for demonstrating life; both were based on the exponential distribution for time to failure; and both employed a sequential testing procedure. They allowed only limited flexibility since the risks, test truncation times (relative to the mean life requirement), and allowable numbers of failures were fixed values.

A number of documents have been published which tabulate sampling plans for various testing approaches and different assumptions concerning the distribution of measurement parameters. Table 14-2 identifies several of these developed primarily for military procurement. A number of sampling plans appearing in the general literature prior to 1962 can be identified in Ref. 14-3. Also, many industrial organizations have extensive tabulations of sampling plans (see Ref. 14-4 for example).

Whereas the motivation for many existing tabulated plans was for quality control work, they often apply directly or are readily adaptable to reliability problems.

---

* The only limitation is the ability to treat the equations and functions involved; simulation with computers has aided this significantly.

** These terms, the consumer's and producer's risks, are explicitly defined in Sec. A.10 of the Appendix.

*** The two plans were developed for separate application to preproduction and production items. The specific recommendations of these AGREE task groups for these plans are summarized in Ref. 14-7.

For example as noted in Table 14-2, MIL-STD-105D could be used to demonstrate per cent success (or failure) at fixed time T by using a time-truncated test of duration T. An adaptation of the use of MIL-STD-105D for testing for any duration and extrapolating results to time T is described in Ref. 14-5; this requires, of course, assuming a distribution for time to failure and the Weibull distribution (for which the exponential is a special case) was chosen in that work.

Many reliability demonstration problems can be treated with established sampling plans. When they do, in fact, fit the problem at hand, their use is highly recommended as a time-saver. However, each new problem should be treated individually. A simple design change may require a new plan if for example it changes the predominant mode of failure. Sometimes the more direct approach mentioned above using data from other tests may be suitable.

Test Specimens

Generally it is desirable that the fabricated models of the proposed design serving as test specimens be as near final operational form as possible, but it is also desirable to maintain a certain degree of flexibility so that the design can be modified if the evaluation of test results indicates this is necessary. The form of the test specimens is influenced to a large extent by the stage of evolution of the design and the level of assembly of the hardware.

For off-the-shelf items such as piece-parts and certain components which are produced in large quantities and over long periods of time, the design maturity of the item has often progressed well into the production phase before specific reliability requirements of a particular customer are known. Even though the manufacturer may have already demonstrated reliability for certain application environments, the customer will typically purchase samples and conduct his own reliability demonstration tests as a part of his parts qualification program. Further reliability demonstration of such items may even be done as a part of formal acceptance testing* by sampling from procured lots during incoming tests. Rejection of the product by either indicates to the manufacturer that he must make modifications in order to remain as a supplier. Most often, a modification of the assembly or quality control procedures in his production process are all that is needed. In other cases it may entail more basic changes in the design or the material used.

At the equipment level reliability demonstration has been traditionally set up in most programs as a special phase of testing performed after formal qualification

* The term, acceptance test, is often used in the literature to denote any accept-reject testing procedure; the discussion in Sec. A.10 employs it with this meaning. As used above, however it refers to the formal area of acceptance testing as defined in Sec. 13.4.

Table 14-2

Standard Sampling Plans [a/]

Designation	Title	Demonstration Parameter	Test or Measurement Approach	Distribution Assumption	Remarks
MIL-STD-105D	Sampling Procedures and Tables for Inspection by Attributes	AQL (acceptable quality level or per cent defective)	Applies any time when each item can be classified simply as either good or bad (e.g., a device is either nondefective or defective; also, a test truncated at time T can result in simple success or failure result for demonstrating per cent success at time T.	Occurrence of defects are distribution-free	Contains single, double, and multiple plans; not tied to specific producer or consumer risks.
DoD Handbook H-108	Sampling Procedures and Tables for Life and Reliability Testing (Based on Exponential Distribution)	Mean life	Failure-truncated, time-truncated, truncated sequential.	Exponential distribution for time to failure	Plans given for various combinations of risks (α=0.10, β=0.05, etc.)
MIL-STD-414	Sampling Procedures and Tables for Inspection by Variables for Per Cent Defective	AQL (acceptable quality level or per cent defective)	Applies to any test procedures so long as the basic measurements are recorded as continuous variable values.	Normal distribution for measured parameter	Plans given for both single and double specification limits; not tied to specific producer or consumer risks

[a/] The descriptive features are limited to those adequate to give a gross illustration of the nature of the plans. The specific documents should be read for full appreciation of the detailed features.

Table 14-2 (Continued)

Designation	Title	Demonstration Parameter	Test or Measurement Approach	Distribution Assumption	Remarks
MIL-STD-690A	Failure Rate Sampling Plans and Procedures	Failure rate	Time-truncated	Exponential distribution for time to failure	Plans given for 60% and 90% confidence demonstration at six different failure rate levels.
MIL-STD-781B	Reliability Tests - Exponential Distribution	Mean life	Time-truncated and truncated sequential.	Exponential distribution for time to failure	Plans given for various combinations of risks (α=0.10, β=0.20, etc.); both low risks (10-20%) and high risks (30-40%) plans included.

testing with special prototypes or with early production models. Conducted in this manner, reliability demonstration has served as adjunct to qualification testing to provide further verification that the design meets all the requirements thus including reliability before full-scale production proceeds. More recently there has been much emphasis on integrating it with other tests such as qualification or development as a means of increasing over-all testing efficiency.

At the large system level there is the usual limitation of available test specimens. Coupled with this is the cost and difficulty associated with generating appropriate test conditions. Together these preclude the use of the traditional, direct approach to reliability demonstration. Combination of results of tests conducted on lower level parts of the system is one practice being employed to overcome this. Use of Bayesian methods as illustrated in Figure 14-1 is another.

These problems and practices are only a sampling of those associated with many modern-day systems and items of hardware. Further discussion of special problems is presented in 14.1.3.

14.1.2. Traditional* Reliability Demonstration

Three general categories of reliability demonstration, life, strength, and simple attributes/variables were introduced in Table 14-1. The approaches associated with the life category are the most celebrated forms of reliability demonstration in the traditional sense. Whereas many literature articles treat life as <u>the</u> reliability demonstration problem, approaches associated with the strength and simple attributes/variables categories are also frequently important. In some cases a single approach may not be adequate, and the total reliability demonstration may require some combination of two or more approaches.

Within all three of these categories there are certain factors to consider in detail in preparation for demonstrating reliability. Typical of these are

(1) accept-reject criteria,
(2) test approaches,
(3) hypothesis test procedure,
(4) statistical distribution assumptions,
(5) sample size,
(6) sampling plan selection,
(7) test duration,
(8) levels of risks, and
(9) costs.

---

* By "traditional" we mean those approaches where there are generally enough test specimens, test facilities, test time, etc. available to provide ample data for making required accept-reject decisions at reasonably low risks with classical statistical techniques.

In some cases certain factors may be fixed by the situation at hand; for example, program policy may specify a sequential testing approach and use of the exponential assumption for times to failure. In some cases there may be constraints imposed by certain factors; for example, program schedule may not allow the time for investigation of better sampling plans. Most often there is some flexibility available in the selection of characteristics of the demonstration and trade-offs are possible. Trade-offs often include cost as an important factor and are usually aimed at minimizing cost subject to the constraints at hand.

In later discussion on each of the three reliability demonstration categories of life, strength, and simple attributes/variables, cost is treated as the common denominator for practical considerations because of its importance. It is assumed that a simple cost model of the form

$$C = C_T + C_E$$

applies where

C = total cost of demonstration

C_T = immediate cost of the test including such factors as manpower, test specimens, facilities, and

C_E = cost of making an erroneous decision from the test results.

This simple model admittedly permits only gross qualitative consideration; however, is adequate for the purpose at hand. More sophisticated models might provide a detailed breakdown of these terms with interrelationships of the various factors and might include other terms such as the cost of not testing.

Note that the model may be the same in form (at least for this simple version) for both the customer and producer. In particular the latter term might be expanded (in certain cases) into the form

$$C_E = P_{RG} \, C_{RG} + P_{AB} \, C_{AB}$$

where the subscripts RG and AB denote events of rejecting good items and accepting bad items respectively, P is probability and C is cost. Now if this represents, say, the customer's error model, C_{AB} will typically be larger than C_{RG}. C_{AB} might represent costs due to added maintenance effort, jeopardization of the mission, creation of safety hazards, or necessity for more spares which results from the failures of bad products accepted. There are also many cases where C_{RG} can be significant to the customer if, for example, it means loss in the customer's schedule. Note especially that the values of C_{RG} and C_{AB} are essentially independent of the test itself; however, C_E is dependent on the test through P_{RG} and P_{AB} which are functions of the true reliability and the producer and consumer risks.

Generally, the larger the sample size the lower the risks in the demonstration. An increase in sample size increases the immediate cost of the test (the C_T term in the above expression) but decreases the cost of making a wrong decision (the C_E term in the above expression) by decreasing P_{RG} and P_{AB}. A logical balance of sample size with risk is thus determined by those values which minimize the total cost. In the event that the available number of specimens for test is limited, the sample size serves as a constraint. This is characteristic of special problems with many modern-day hardware items and is explored in later discussion.

It is usually less expensive to ask questions first and test later; however, the available information on cost and underlying statistical distributions may be so limited that only gross considerations are possible. If a statistical distribution can be assumed beforehand, sampling plans can help tremendously in this process by providing the designer of the test with much of the trade-off information he needs to maximize efficiency. Sampling plans also allow the advantage of being able to assess the risks before testing. Whereas sampling plans are usually developed assuming complete ignorance of the true value of the demonstration parameter, there may be some knowledge of the value which even if qualitative may provide an advantage in sampling plan selection leading to cost savings. For example, a conventional OC family of curves might indicate that a sample size of ten is needed to satisfy a given risk; whereas the sample size might be decreased if there is a strong degree of belief that the true reliability is much greater than the reliability requirement to be demonstrated.* Additional worthwhile discussion on cost considerations in sampling plan selection is presented in Ref. 14-6.

As previously recognized, there are two major risks to be considered: the consumer's and the producer's. It is traditional to explicitly specify the values of the demonstration parameter at which the risks are to be discussed. For example, a demonstration of mean life θ may specify an unacceptable level θ_1 below which the customer wants a low probability β of acceptance and an acceptable level above which the producer wants a low probability α of rejection.** In formal procurement specifications θ_1 and θ_0 (and similar respective dual levels for other demonstration parameters) have come to be associated with the minimum acceptable

---

* If such prior knowledge can be used explicitly for developing the sampling plan, this falls into the realm of Bayesian sampling plans as discussed later in this section. Lloyd and Lipow [Ref. 14-2] include some discussion related to use of such knowledge from the traditional demonstration point-of-view.

** These two levels are illustrated in Sec. A.10 of the Appendix.

reliability requirement and the design goal reliability respectively and the customer's and producer's risks are usually associated with these levels. The ratio of the two levels (θ_0/θ_1 in this case) is typically known as the discrimination ratio and is one of the major factors involved in selecting sampling plans. Reference 14-7 describes in a general manner how as θ_0/θ_1 increase, less testing is required to make an accept-reject decision. Some interesting relationships between the discrimination ratio and other factors such as test time, probability of acceptance, and allowable number of failures are presented in Ref. 14-8 and provide much insight into the effect of this parameter. It is well to remember to not insist on making this parameter lower than needed since for given risks it can only increase testing costs.

Life

Typical parameters and basic test types appropriate for life demonstration are listed in Table 14-1. The different parameters are essentially equivalent measures of reliability (pursuant to the underlying assumptions); however, convention has established in many cases which of these are used. For example, MIL-STD-690 (see Table 14-2) is oriented mainly toward piece-parts and small components and specifies a time-truncated test for demonstrating failure rates. MIL-STD-781, however, is oriented mainly toward equipments and relies mainly on sequential testing for demonstrating mean life.

The use of the different test types in life measurements was discussed in Sec. 8.1 and much of that discussion emphasized reliability demonstration. As noted previously, the life category has been treated extensively in the reliability literature. Good basic treatment is provided by Lloyd and Lipow [Ref. 14-2]. Application procedures are well described in Ref. 14-9 and good discussion of the practical facets is presented in Ref. 14-10. Each of these include some economic considerations in test plan selection. Dellinger [Ref. 14-6] treats the cost problem more directly.

One of the most controversial subjects in traditional life demonstration concerns the underlying distribution for times to failure. The exponential distribution corresponding to a constant hazard rate is the form most often assumed mainly because of simplicity. Some discussion concerning this was presented in the summary part of Sec. 8.1. Ryerson [Ref. 14-10] notes that from the customer's viewpoint the assumption is generally safe if the product passes the test in that a sampling plan based on the assumption is more likely to reject the product if the hazard rate is increasing. Myers [Ref. 14-9] takes a totally opposite stand with corresponding statements concerning the error. Actually, the risk of using a plan based on the exponential assumption when another applies varies from plan to plan. An investigation by Zelen [Ref. 14-11] using plans based on the exponential assumption for all three basic testing approaches verifies this for the specific case when the actual distribution is Weibull with shape parameters greater than unity*. If there is reasonable justification that the underlying distribution of a population is exponential (or near exponential), then the error will generally be small in either direction. There is no clear-cut alternative of what to do when there is doubt about whether the exponential assumption is valid. Recommendations are often made to first verify the assumption; this is usually easier said than done, however. Use of a distribution-free procedure is a possibility; however, this increases the immediate cost of testing. There is definitely a need for further investigation for rationale concerning the appropriateness and adequacy of the exponential distribution. Until then, each situation must be treated on its own merit.

Strength

As indicated by the list of demonstration parameters and basic test types in Table 14-1, there are several ways of demonstrating reliability in this category. The concept of strength as an important reliability characteristic and its measurement were discussed in Sec. 6.1. From the point-of-view of reliability demonstration one wishes to demonstrate with a given confidence that the probability that strength is greater than the severity of applied stress is greater than some required level, the reliability requirement.

A sketch of overlapping probability density function for stress severity and strength (see Fig. 6-3) is commonly used to depict the concept and the region of overlap defines a variable (stress minus strength) the properties of which determine the reliability. Calculations and test design based on this model often assume for simplicity and tractability that both stress severity and strength (or perhaps certain transformations of these) are normally distributed which means that their difference is also normally distributed. Lloyd and Lipow [Ref. 14-2] describe this approach in detail and illustrate lower one-sided confidence limit computations for reliability estimated from separate measurements of stress and strength. Whereas the assumption of a normal distribution is often used and is often adequate, many other distributions such as the log normal, gamma, and the Weibull are frequently mentioned as applicable types.

Since the overlapping distributions emphasize the values of stress and strength in the tails of the distribution for which occurrences of measured values in these regions with small samples are extremely rare events, there are inherent inaccuracies involved with the use of any distribution. For this reason safety margin is commonly used as the parameter to be demonstrated. Several ways of defining safety margins

---

* Weibull shape parameters greater than unity are analogous to increasing hazard rates.

were presented in Sec. 6.1; their introduction as reliability indices is generally attributed to Lusser [Ref. 14-12]. Note that the mere use of safety margin parameters does not circumvent the inherent inaccuracies resulting from events in the tails of overlapping distributions. If demonstrated to be large, however, (interpreted as large separation between stress and strength distributions), this adds confidence that the probabilities resulting from overlap of the stress severity and strength distributions is small. Some reliability demonstrations thus emphasize the direct use of safety margins as demonstration parameters and specify the accept-reject criteria in terms of demonstrated levels of these. Good examples of this approach using both stress-to-failure and sensitivity testing are described in Ref. 14-13.

Note that if the form of the distributions are known, the specific level of reliability or probability of success demonstrated corresponding to the demonstrated safety margin can be determined (at least in principle). Bombara [Ref. 14-14] presents a good description of how this is done when both distributions are normal. In that paper, tables are included which permit with given measurements of stress and strength a determination of the demonstrated reliability at various levels of confidence. It also illustrates the use of these for test design to achieve a given level of confidence.

More often, the emphasis is on demonstrating the safety margin without attempting to relate it to a specific level of probability of success. Generally, the safety margins achieved are large enough so that acceptance of the product is confidently assumed to represent essentially certainty of success during application.

Demonstration with other strength distribution parameters such as mean, standard deviation and percentiles, are, in effect, equivalent to the above approaches. Whereas the emphasis in these appears to be primarily on strength, the accept-reject criteria would typically be influenced by the operational stress severities and, similar to those for safety margin, would be specified more on the basis of achieving certainty of success rather than a specific level of reliability to be demonstrated.

One other approach is to use the proof testing method and base the demonstration on per cent surviving at a fixed stress level. For example, if the operational stress is known to be fixed at level S, then the per cent surviving is the estimate of reliability. The lower one-sided confidence limit can be obtained per usual methods. If the operational stress was not fixed but was varying, then the selection of a specific level, say the maximum working stress or the product's rated stress, could be selected as the test condition and a conservative demonstration of reliability performed. One major advantage of this approach is its amenability to an attributes sampling plan (MIL-STD-105D plans for example). On the basis of statistical theory, one could argue that it is a less efficient procedure. That it would typically require a larger sample size is true; however, it has the additional advantage that items surviving the test are not damaged and will with certainty be suitable for application. With the increased emphasis on reliability in modern systems, this is now often done on a 100% basis in which case it is one form of screening.

Of the sampling plans listed in Table 14-2, MIL-STD-105D and MIL-STD-414 are the more appropriate ones for this category of demonstration. But remember that there are other standard plans available. Having decided the demonstration parameters and test approach to use it may well be worthwhile developing a plan to assist in planning a more efficient test.

Simple Attributes/Variables

The procedures for reliability demonstration in this category are considerably simpler than for those discussed above since the effects of stress and time on individual items does not have to be treated. Reliability for products treated in this category is simply the probability of the event success; estimates of reliability are obtained simply as the per cent success when the results are attributes measurements or as the proportion in-tolerance when the results are variables measurements.

As noted in Table 14-1, the bulk of the data for demonstration is achieved with simply measurements. Sec. 5.1 illustrates measurements of this type. Many applications and approaches can be found in the quality control literature.

Sampling plans are employed extensively in this category and MIL-STD-105D and MIL-STD-414 (see Table 14-2) are frequently cited as applicable. Numerous other sampling plans are presented in Sec. A.10 of the Appendix.

14.1.3 Special Problems

Whereas the traditional approaches to reliability demonstration are still adequate for many products being designed, certain features of some modern-day systems and hardware items introduce special problems for which there is no simple, straightforward solution. Typical features causing this are:

(1) extremely high reliability (long life) requirements,

(2) high cost and limited availability of test items,

(3) tight program schedules, and

(4) large size and complexity of systems.

The problems are further compounded when several of these features apply simultaneously in a program.

The most outstanding examples of such problems are found with many NASA systems. How does one, for example, approach the reliability demonstration of a spacecraft having the following characteristics and requirements:

and demonstration problem considerably by providing more emphasis on the reliability data needs. As a result the data for reliability demonstration will often be derived from several areas of testing such as development, qualification, and acceptance testing. It also introduces the problem discussed below of combining the results of test on separate components and different levels of assembly. References 14-15 through 14-17 give descriptions of one organization's approach to the use of integrated testing in the reliability demonstration of a system. This also incorporates several other techniques and practices introduced above. Meaningfulness and compatibility of data from the different sources are the major problems in relying on integrated testing. Good planning and allocation of effort can alleviate many of the difficulties.

Combining Results of Tests on Separate Items

The problem here can be stated as one of combining demonstrated component reliabilities at different or identical confidence levels to determine a system reliability at a chosen confidence level. A number of literature articles are available on this problem, see for example, Refs. 14-18 through 14-20. Any procedure requires a system model for combining the element reliabilities. The problem rapidly gets complex with increasing complexity of the system. Interface and interaction effects are thus often neglected. There must be considerable discretion in the interpretation and use of the combined results.

Combining Data from Different Lots

This practice is mostly applicable to piece-parts and is posed in Ref. 14-10 as one of several alternatives for dealing with long-life requirements. As noted there the 38,000 series of military specifications employs this. Inherent problems here result from lot-to-lot variation.

Lowering the Confidence

In some cases the customer may be willing to accept the higher risk of accepting bad products. Potential loss to the customer should be assessed with cost models as discussed in Sec. 14.1.2. As noted in Table 14-2, MIL-STD-781 includes some provision for high risk demonstration. Reference 14-21 recognizes the advantage of lowering the required confidence during the early phases of a program when decisions should not be too rigid. In so-called "New Look" test plans described there, new decision points are defined as testing continues and the risk at each is computed independently of previous decision points.

(1) MTBF requirement of 8,000 hours,

(2) spacecraft not reusable after a mission or flight test,

(3) three spacecraft allotted for testing, and

(4) different functions required for each mission?

Traditional approaches are obviously not possible with such a system. Translation of these MTBF requirements down to one of the many piece-parts might typically reveal that years of total test time would be needed to demonstrate the required reliability at the stated confidence.

A number of practices and techniques have recently been described in the literature in an attempt to circumvent such problems. Some of the popular key words and phrases used to designate these are:

(1) integrated testing,

(2) combining results of tests on separate items,

(3) combining data from different lots,

(4) lowering the confidence (hence increasing the risks),

(5) modifying the accept-reject criteria,

(6) Bayesian methods,

(7) accelerated testing,

(8) overstress (proof) testing,

(9) reliability screening,

(10) reliability growth, and

(11) reliability physics.

No single one or combination of these has proved to be a panacea for verifying high reliability requirements. As per the usual purpose of reliability demonstration, one ideally wants enough information to make an accept-reject decision on the basis of satisfying or not satisfying some explicitly stated criteria. Some of the procedures associated with the practices and techniques listed above do attempt to provide the evidence in a form so that decisions can be clear-cut or automatic as in the traditional approaches. For example, combining the data from different lots may allow doing this. (See later discussion on this procedure). In this case there has to be some engineering judgment about lot to lot variation, hence there may be some actual loss of confidence which is not reflected in the statistical confidence. Hence there is often emphasis on supplementing statistical confidence with engineering confidence. Some procedures must rely heavily on this; some almost totally.

Integrated Testing

Whereas the initial motivation for integrated testing was mainly for improvement in efficiency in the overall testing of a program, it has aided the reliability test

Modifying the Accept-Reject Criteria

This is not intended to imply changing the reliability requirements but establishing the goal in a form more easily verified; yet, just as meaningful and sometimes better. A simple example is change from a probability of success criteria to one stated in terms of a safety margin. Also, repeated cycling of an item for the mission duration under simulated mission conditions to demonstrate a per cent success could provide a more meaningful approach than, say, continuing to test for long periods to demonstrate an MTBF. Some have suggested (see for example Ref. 14-10) that criteria be tied to either the consumer's or the producer's risk but not both. Even replacement of quantitative criteria with qualitative criteria might sometimes be a more rational approach. There are often numerous alternatives in specifying criteria. Some may require relying more on engineering judgment than others.

Bayesian Methods

There are a number of recent attempts to use Bayesian methods as a means of circumventing the small sample data problem. Prior knowledge in one form or another may be used in conjunction with designing the test or with improving the results. This is illustrated in Fig. 14-1. Reference 14-22 gives an example of making use of prior experience on success of design groups in developing sampling plans. References 14-23 and 14-24 discuss the use of reliability prediction results in planning demonstration tests. Most often the prior information is concerned with a distribution associated with the prior information. A major problem has been in the lack of good methods for quantifying the prior information. The reader is encouraged to search the literature for further applications as Bayesian methods have shown significant promise.

Accelerated Testing

There is much emphasis in using this to shorten test time. It is generally more appropriate for parts and small components than for equipments or systems. There are many inherent problems caused whenever there is an attempt to speed up aging. Whereas accelerated testing readily provides large amounts of data for statistical analyses, it should be remembered that there may be big differences between the statistical confidence achieved and the true confidence due to the underlying assumptions regarding the validity of the acceleration. The subject of accelerated testing is treated in more depth in Sec. 11.

Overstress Testing

This consists of applying conditions at severities above the normal operating or rated levels and determining whether the product is capable of withstanding them. It is not a way of demonstrating reliability at a given confidence level in the formal statistical sense. It is concerned more with achieving engineering confidence and is based on the premise that if the product works successfully at the higher than normal severities then there is negligible risk of it failing at normal levels. It is conservative with the degree of conservatism depending upon the severity of the higher-than-usual severities--are you merely proof testing the item or are you causing accelerated aging? Presence of accelerated aging does not necessarily invalidate the procedure but does complicate the picture of knowing what meaning the results have, especially when failures occur. Sections 5 through 8 should be consulted by readers not generally familiar with such differences. A good example of using overstress testing for demonstration purposes is described in Ref. 14-25.

Reliability Screening

Screening refers to testing and inspection of 100% of the items. The term generally refers primarily to piece-parts; however, it is certainly applied almost always in some at higher levels of assembly. When failure is caused by stress severity exceeding strength, a proof testing approach is appropriate. Typically, the tests are conducted at maximum severity levels of normal conditions or at rated conditions. If failures of this type were the only ones the product could have, then the procedure could be considered an adequate substitute for reliability demonstration since there is essentially 100% confidence that all items passes the test will perform satisfactorily. When failures result from aging effects, screening is required. Whereas this type of screening enhances the reliability, it is not itself a direct substitute for reliability demonstration. It is possible, however, that data for reliability demonstration can be obtained in connection with burn-in tests. More discussion on reliability screening is presented in Sec. 14.2.

Reliability Growth

The motivation for reliability growth analyses is basically different from that for reliability demonstration per se in that its major purpose is to indicate the trend in achieved reliability. Calculations of specific points for the reliability growth curve may employ data from single tests or from various sources such as items, and combining data from different lots. If the reliability growth curves include appropriate confidence levels then the results can be used for making accept-reject decisions. A good discussion of the relationship between reliability growth and demonstration is presented in Ref. 14-26. Further discussion of reliability growth is presented in Sec. 14.3.

Reliability Physics

Reliability physics is concerned with the more fundamental behavior of materials and the causes of failure which contribute to unreliability. There has been much emphasis on relying more heavily on this for achieving high reliability. A very strong argument for more use of basic studies is presented by Gollovin [Ref. 14-27]. It is yet very far from being able to provide the adequate prediction models and data on which to base decisions about fabricated hardware. Thus its role, and a valuable one, will continue to be as a supplement to other practices. Some further discussion of reliability physics is presented in Sec. 11.

Other Practices

Many other practices which, while not a part of formal reliability demonstration per se, do contribute much toward the achievement of engineering confidence that reliability goals are adequately met. Reliability prediction, design reviews, incentive contracting, and failure mode investigations, are only some of these. The achievement of high reliability, especially in the presense of constraints limiting the resources for experimentally verifying it, demands a concerted effort of all practices capable of contributing.

14.2 Screening

Screening pertains to testing every item of a particular type intended for application for the purpose of eliminating those which are defective or which potentially have shorter life than required. It is fundamentally intended to provide a supply of reliable components for fabrication of higher levels of assembly. The designation most often applies to piece-parts and small components as part of an acceptance testing program; however, practically all higher levels of assembly are subject to some type of screening test before application.

A complete screen of an item may involve a single test or two or more tests in sequence. Each test is basically a go, no-go test; items which pass a test are accepted for application or are subjected to further tests whichever is appropriate and items which fail are rejected. Some screening tests are designed to reject only those items which are damaged or destroyed; others are designed to merely make a value judgment of good or bad. Some tests are designed to cause the actual failures; some are designed only to reveal differences in characteristics that serve as indicators of weakness.

Screening has become a very popular practice as a way of achieving high reliability. Screening per se does not, however, provide quantitative measures for reliability; however, in an integrated test approach some tests may be designed to serve a dual role of screening and providing useful data for estimation. Some basic screening approaches such as proof testing may also serve as a demonstration of reliability.

14.2.1 Basic Approaches

There are several basic approaches to screening. Descriptions given below are brief as most of the underlying concepts have been presented in earlier sections.

Simple Attributes Measurements

Some items can be simply judged good or bad on the basis of a simple attributes measurement as in a visual inspection or an electrical parameter test. Measurements of this type are described in Sec. 5.1 and the uses of such measurements for making accept-reject decisions on individual items was discussed in Sec. 6.3. In this approach items are not generally damaged or aged. In some cases the characteristic being measured may be selected to serve as an indicator of stress or aging dependent failure modes and mechanisms. Measurements may be aimed at detecting physical flaws or an out-of-tolerance condition of some parameter. Special nondestructive techniques of the type described in Sec. 9 are often employed in order to measure items by this approach.

Proof Testing

In this approach items are purposely stressed to reveal a deficiency in strength. The concepts of proof testing are described in Sec. 5.3 and its uses for making accept-reject decisions of individual items was discussed in Sec. 6.3. Briefly, stresses are applied at some fixed severity level less than the strength the item is designed to have for application. Items that pass the test are not damaged and have a strength greater than the severity of the applied stress. Examples of this approach are centrifuge and pressure tests of components. Note that these tests do not age items. The criteria for failure may be either damage to the item or an out-of-tolerance condition. Stress levels for the test most often correspond to those representing maximum severity during normal operation or to rated levels; however, some screening tests of this type are known to employ severities above and below these levels. Further discussion of the rationale for selection of the severity level is presented in Sec. 5.3.

Burn-in and Bake-in

Burn-in refers to power-on aging; bake-in to power-off aging. The approach involves aging items to purposely induce failures in those which have the shortest lives. It is based on the premise that the hazard function for the items tested possesses an infant mortality region in which the probability of failure is higher than that for the period of application. Major problems to be considered consist of test conditions and test duration. Such problems were discussed in Sec. 8.2. Vol. IV - Parts of this series also discusses this subject.

Run-in or Break-in

Screening is only a secondary purpose of this technique. The major purpose is improvement of items by purposely aging. It is based on the premise that certain items can be improved, i.e., increased in strength, made to have a longer life, or simply made to perform better, by preconditioning them under controlled conditions. A very familiar example is break-in of an engine. Some piece-parts and small components can be improved by preconditioning; for example, moisture in sealed components may be driven off by heating or operating them. The technique was further discussed in Sec. 8.2. Measurements made during or at the end of the test might be used for screening. For example, if improvement from preconditioning fails to occur the item might be rejected.

Parameter Drift

This approach is based on the premise that parameter behavior observed during the early life of an item can reveal inherent weaknesses or potential unreliability in later life. The basic measurements of this type are described in Sec. 8.4. Of most importance is recognition of the difficulty and danger in extrapolating incipient effects to later time. For this reason the criteria for rejecting an item when using this approach should generally be based more on the nature, instability, and departure from the norm of the behavior rather than on the net drift during the test. Whereas such erratic behavior does not strictly imply early failure in application, experience has shown that wariness of such differences is well justified. Additional discussion of these types of measurement for screening is given in Vol. V - Parts.

Discriminants

The term discriminant generally refers to a mathematical expression* providing a criteria for the behavior of another more complicated relation. Extending the concept to screening, it simply means that the results of separate measurements on two or more characteristics of an item are combined by some mathematical model or discriminant function to obtain some figure-of-merit (FOM) which is used to judge the item good or bad. The method of discriminants is usually associated with an empirically determined function. For example, one may have reason to believe that the failure of a particular item is related to some combination of power dissipation, rise time and output noise level, and may wish to first model a FOM in terms of these test parameters and then use it in his screening program. If a linear discriminant function in these

---

* Whereas the statistical literature usually infers that the expression is a linear combination of variables, the broader interpretation used above (and obtained from Webster's Seventh New Collegiate Dictionary) is preferred since functions other than linear combinations can be used.

parameters is assumed, the modeling effort consists of observing the test parameters and whether or not the item fails, then determining the coefficients that minimize the proportion of misclassification of good and bad items. The correlations between parameters are automatically considered in the procedure; see Sec. 6.3 for further discussion of the method. Once the discriminant function is obtained, the screening procedure involves only making the measurements and calculating the value of the discriminant function FOM.

If on the other hand a discriminant function is available, no preliminary testing is required. A simple example of this approach is using the value of the gain-bandwidth product of a transistor computed from separated measurements of gain and bandwidth as a basis for determining whether it is good or bad. Existing performance of models of various types such as transistor h-parameter models, gyro transfer functions, and circuit equations may be quite adequate as discriminant functions. Thus the theoretically based discriminant assumes that one knows that failure is correlated with a certain interval of values of the FOM, and if the test parameters for an item result in a FOM in this region it is rejected, otherwise accepted.

In principle there is no limitation on the degree of complexity of the discriminant function. In practice they are usually not very complex, especially if they have to be determined empirically. The preliminary testing required to determine the emperical discriminant function is an obvious disadvantage of this approach. Even the approach with theoretically based functions finds limited use because it generally adds little improvement over using individual parameter measurements for making separate decisions in sequence. The method is advantageous only when an overall FOM obtained by combining individual measurements is more effective and efficient for determining weaknesses than using individual measurements.

14.2.2 Planning a Screen

As noted earlier, a complete screen of a particular item may involve a single test or a sequence of several tests. Most often it is the latter. Desirable features of a screen are effectiveness and efficiency. The design of a screen to achieve these qualities requires much attention to detail. Typical factors to consider in planning a screen are:

(1) intended use of the item,
(2) failure modes and mechanisms,
(3) test methods,
(4) when to screen,
(5) order of tests, and
(6) cost.

The following discussion gives brief treatment to these to illustrate typical considerations in planning a screen.

The most important requirement for developing an effective screen of an item is a thorough knowledge of the failure modes and mechanisms.* The intended use of the item dictates which failure modes are important. These can be determined with a conventional failure modes and effects analysis. Furthermore, if the relative probabilities of occurrence of the different failure modes are known it may be acceptable to neglect those having a low probability. Standard screening procedures developed for specific or generic component types generally attempt to account for all failure modes that can be reasonably treated.

The problem of understanding the failure mechanisms causing the failure modes is another story and it is really these that screening procedures must emphasize. In some cases a failure mode may uniquely determine the failure mechanism. There are numerous cases, however, where several failure mechanisms can cause the same failure mode. For example, possible failure mechanisms of electrical contacts failing to conduct adequate current could be wear, corrosion, erosion, burrs, etc. In developing a thorough screen it is necessary to consider all of the failure mechanisms even though those with known very low probabilities of occurrence could be neglected. As for failure modes, standard screening procedures generally attempt to include most known failure mechanisms that can be reasonably treated.

Thus the aim of a screen is to reveal those components which, on the basis of their failure mechanisms, are weak. A major problem in developing screening procedures then is how to determine which components of a given type have poor failure mechanism characteristics. For this one must rely on some test (or measurement) and much of the further effort in development of a screen consists of selecting the appropriate test method.

Generally, the simpler a test for this purpose the better. In terms of the testing approaches described in Sec. 14.2.1, a proof test test approach or a simple measurement approach would be preferable if applicable for either of these generally do not require aging. Usually some form of these such as electrical parameter tests, x-ray scanning, and centrifuge tests are appropriate for certain failure mechanisms; however, they rarely suffice for screening high reliability electronic parts. If nonaging methods are not adequate, then there is no alternative but to resort to aging approaches such as burn-in, bake-in, and parameter drift.

The distinction is noted here between methods which attempt to cause failure or activate behavior indicative of failure and those which rely on observing some static

* The difference between failure modes and failure mechanisms was described in Sec. 2.1.4.

property hopefully revealing failure mechanisms if they exist. Proof testing, burn-in, and parameter drift approaches are representative of the former. A major problem with methods of this type concerns the test conditions. Generally, severities above rated levels are not employed for screening for fear of harming the component. Some screening is even done using conditions representative of normal use. However, the problem with test conditions is more than selecting severity levels. It is possible, for example, that a particular failure mechanism can be excited by two different external conditions, say high temperature and load. One may thus have a choice of either or may want to employ both if there is uncertainty of equivalence.

When the screening relies on observing static properties, one is dealing only with simple measurements. Note however that it is possible that aging dependent failure mechanisms can sometimes be treated with nonaging methods. For example, measurements of current noise and infrared emission profiles have been explored for possible use as preindicators of later failure. A survey of such techniques for electronic components is described in Ref. 14-28. Don't be mislead into thinking that "tried-and-true" purposely aging is necessarily preferable to such preindicator techniques if the latter have proven successful in removing infant mortality components. When screening to remove early failures, any aging is consuming some useful life. In the face of uncertainty about the true nature of the hazard function, you may be taking less risk by assuming it to be increasing. Prior experience with similar components may aid in making judgments about whether purposely aging is beneficial.

If test methods are not available for screening certain mechanisms, the screen may have to be complemented with appropriate destructive tests on a sampling basis. A good illustration of the reasoning leading to such a decision for one component type is presented in Ref. 14-29.

The order of the different tests in a screen can be important. Table 14-3 reproduces the order of measurements and tests specified in a recently developed screening program for digital integrated circuits [Ref. 14-30]. Typical rationale for ordering the test might be as described below.

Efficiency is generally higher when the simpler tests are conducted first in that items that are rejected do not have to be subjected to more expensive testing later. Also, the nondegrading tests such as electrical parameter tests and x-ray scanning can generally be in any order.

To maximize effectiveness high temperature tests should generally precede mechanical tests since weakening of structure and other effects may occur. Also, the leak of hermeticity tests should usually follow all environmental tests to insure that damage to hermetic seals have not occurred.

Another problem associated with screening is concerned with when to screen. There are different schools of thought and obvious advantages exist for each approach.

Table 14-3

Test Sequence For Screening Digital Integrated Circuits [Ref. 14-30]

- Internal Visual Inspection (before Sealing)
- Marking Requirements
- External Visual and Mechanical Inspection
- Marking Permanency Test
- Stabilization Bake
- Electrical Tests
  - General
  - Noise Margin
  - Fan-Out (Loading) and Fan-In
  - Detail Parameter
  - Dynamic Parameter
- Thermal Cycling (Shock)
- Mechanical
  - Centrifuge
  - Vibration (Variable Frequency Monitored)
- X-Ray
- Burn-In (With Variables Data)
- Hermeticity Tests
- Final Electrical
- Lot Provision

In-process screening provides the opportunity to prevent more defects from being built into the items. End-item screening has the obvious advantage of providing a more final check of some effects that in-process screening cannot treat. Screening after assembly into higher levels of equipment provides a verification closely related to application. This latter procedure may involve a burn-in and debugging period for the equipment, however, if many component failures occur it could be a very inefficient procedure. Most high-reliability programs now employ some form of screening in all three opportunities mentioned.

The total costs of screening including penalties and risks for not screening are not easily assessed. A good discussion on screening costs is presented in Ref. 14-31 which also includes some interesting comparisons of several hardware programs. There are a number of tradeoffs possible among the various factors discussed above and these can often be beneficial in decreasing costs. Screening costs cannot be minimized, however, to the exclusion of other factors in the program. Generally, one is seeking some optimum compromise among cost, reliability, schedule, and supply of components.

There are many papers in the literature on screening. References 14-32 through 14-37 represent a sampling of these. Vol. V - Parts of this series also contains some discussion on screening.

14.3 Other Problem Areas

Reliability demonstration and screening were discussed in Secs. 14.1 and 14.2 as two major areas of reliability treated by testing. This section briefly considers other areas for contributions of testing.

14.3.1 Reliability Estimation

Reliability estimation refers to the straightforward problem of obtaining numerical indices for reliability. It can consist of point estimation or interval estimates. Testing is the major source of data for this. Usually the program reliability evaluation plan will require that a best estimate be available for assessment purposes. The estimates may also serve as a basis for making accept-reject decisions for formal reliability demonstration purposes as described in Sec. 14.1. Estimates may also provide the points on reliability growth curves (See Sec. 14.3.3).

Various reliability indices and models appropriate for reliability estimation are found in Vol. IV - Prediction of this series. Examples of reliability estimation were presented in Secs. 6.1 and 8.1. More general discussion on estimation is given in Secs. A.5 and A.6 of the Appendix. Lloyd and Lipow [Ref. 14-2] give excellent treatment of reliability estimation from a fundamental viewpoint.

14.3.2 Reliability Growth

Reliability growth pertains to the relationship of achieved reliability versus stage in the evolution of a product. The earlier tests in a program usually reveal design and fabrication discrepancies. As these are eliminated the failures per unit of equipment operating time typically become less frequent and hopefully the reliability increases to or above the original reliability goal. The inputs to the reliability growth analysis can be derived from any test, formal or informal, that gives appropriate failure-time information. Reliability growth analyses are sometimes used in connection with reliability demonstration (see Sec. 14.1.3). They are often based on a planned reliability auditing or monitoring program. It provides a valuable indication of how the reliability of the evolving hardware is progressing and thus a good measure for control. There are varying degrees of emphasis and importance placed on it among different programs. A thorough treatment of reliability growth is outside the scope of this report. There are a number of recent papers on the subject.* Many of these cite Lloyd and Lipow [Ref. 14-2] for models of reliability growth.

14.3.3 Investigation of Modes and Mechanisms of Failure

This consists of testing to isolate various causes of failure. As a reliability tool, its use relies on the premise that reliability improvement can result from knowing the causes and eliminating them. In some cases items are purposely tested in a manner to cause failure and permit the investigation while in others the investigation may be conducted adjunct to tests for other purposes. In qualification testing, for example, all failures are usually analyzed in detail to determine what corrective action is needed. This problem area was also discussed briefly in Secs. 6.2 and 8.2. Various reliability engineering texts such as Refs. 14-38 and 14-39 contain discussions on this subject.

14.3.4 Performance Variation**

Changes in performance due to stresses and aging are important in reliability since they can affect whether items perform successfully. In certain reliability tests such as those for demonstration, estimation, and screening, performance degradation may be accounted for by the definition of failure including out-of-tolerance conditions. Specific testing for performance variation is often done to identify needed design improvements. Typical measurements might involve parameter sensitivities, relationships between parameter behavior and stresses, aging dependency of parameter behavior, and statistical distribution characteristics of parameters. Performance measurements were discussed in Secs. 6.4 and 8.4. Examples of applications of performance testing

---

* Literature on reliability growth can be readily located in Reliability Abstracts and Technical Reviews (RATR).

** Vol. I - Parameter Variations Analysis of this report series describes various analytical approaches for treating performance variations; one section is devoted exclusively to a discussion of the use of physical models.

to resolve specific reliability problems in design can be found in Refs. 14-40 and 14-41. In some cases the results of tests for performance variations can be used in reliability prediction. This procedure is discussed in more detail in Vol. IV - Prediction of this report series. Reference 14-42 gives a good description of the application of this approach.

References

14-1. Anon.: Reliability of Military Electronic Equipment. The Advisory Group on Reliability of Electronic Equipment (AGREE), Office of the Assistant Secretary of Defense (Research and Engineering), June 1957.

14-2. Lloyd, David K.; and Lipow, Myron: Reliability: Management, Methods, and Mathematics. Prentice-Hall, Inc., Englewood Cliffs, N. J., 1962.

14-3. Greenwood, J. A. and Hartley, H. O.: Guide to Tables in Mathematical Statistics. Princeton University Press, Princeton, N. J., 1962.

14-4. Harter, H. Leon: Theory and Tables for Tests of Hypotheses Concerning the Mean and the Variance of a Weibull Population. Technical Report ARL 67-0059 (DDC No. AD-653 593), Ford Motor Company, Dearborn, Mich., March 1967.

14-5. Goode, Henry P.: Factors and Procedures for Applying the MIL-STD-105D Plans in Life and Reliability Inspection. Quality and Reliability Assurance Technical Report TR-7, Dept. of Defense, May 1965.

14-6. Derringer, David C.: Selecting Reliability Test Plans. Section 3 of Reliability Handbook, Grant W. Ireson, Ed., McGraw-Hill Book Co., New York, 1966.

14-7. Vol Alven, William H., Ed.: Reliability Engineering. Prentice-Hall, Inc., Englewood Cliffs, N. J.

14-8. Pinkham, R. S. and St. James, L. N.: An Approach to Systems Reliability Verification by Test. Aerospace Reliability and Maintainability Conference, May 1963.

14-9. Myers, Richard H.: Reliability Testing. Chap. 7 of Reliability Engineering for Electronic Systems edited by R. H. Myers, K. L. Wong, and H. M. Gordy.

14-10. Ryerson, Clifford M.: Acceptance Testing. Section 15 of Reliability Handbook, Grant W. Ireson, Ed., McGraw-Hill Book Co., New York, 1966.

14-11. Zelen, M. and Dannemiller, Mary C.: The Robustness of Life Testing Procedures Derived from the Exponential Distribution. Technometrics, Vol. 3, No. 1, February 1961.

14-12. Lusser, R.: Reliability Through Safety Margins. U. S. Army Ordnance Missile Command, Redstone Arsenal, 1958.

14-13. Parsons, Delbert W.: Reliability Demonstration Shillelagh Missile Subsystem. Fifth Reliability and Maintainability Conference, New York, July 1966.

14-14. Bombara, E. L.: Probability That Stress is Less Than Strength At Prescribed Confidence Levels, For Normally Distributed Data. Proceedings of Ninth Conference on the Design of Experiments in Army Research, Development and Testing, December 1964.

References (Continued)

14-15. Carpenter, R. B.: Apollo Reliability by Demonstration or Assessment. Tenth National Symposium on Reliability and Quality Control, January 1964.

14-16. Carpenter, R. B.: Demonstrating Reliability for Long Space Missions. Eleventh National Symposium on Reliability and Quality Control, January 1965.

14-17. Roelands, Donald L.: Reliability Testing in the Saturn S-II Stage Project. Fifth Reliability and Maintainability Conference, New York, July 1966.

14-18. Schick, George J.: A Comparison of Some Old and New Methods in Establishing Confidence Intervals of Serially Connected Systems. The Journal of Industrial Engineering, Vol. 18, August 1967.

14-19. Orkand, Donald S.: A Monte Carlo Method for Determining Lower Confidence Limits For System Reliability on the Basis of Sample Component Data. Report No. ORDBB-VC-4, Picatinny Arsenal, Dover, N. J., June 1960.

14-20. Blum, Marvin: Bounds of System Risk Function Induced by Subsystems. (DDC No. AD-627 654). The Rand Corp., Santa Monica, Calif., February 1966.

14-21. Allen, G. H., et.al.: Reliability Demonstration: Theory, Application and Problems. Report ESD-TR-65-198 (DDC No. AD-464 386), Air Force System Command, L. G. Hanscom Field, Bedford, Mass.

14-22. Christie, R. J.: Identification of Inadequate Subsystem - Development of Sampling Plans. 1966 Annual Symposium on Reliability, January 1966.

14-23. Smith, Anthony M.: Risk Assessment in Complex Unattended Aerospace Systems. Annals of Reliability and Maintainability - 1967, Sixth Reliability and Maintainability Conference, July 1967.

14-24. Gottfried, Paul: A Reliability Demonstration Plan with Incentives. 1966 Annual Symposium on Reliability, January 1966.

14-25. Ellis, R. R. and Pollock, J. J.: Reliability Overstress Testing of the Gemini Inertial Measuring Unit Systems Electronics Package. Proceedings, 1966 Annual Symposium on Reliability, San Francisco, California, January 1966.

14-26. Wilson, Myron A.: Risk/Cost Tradeoffs for Optimizing Reliability Demonstration Requirements. 1965 Aircraft and Missile Div. Conference, ASQC, Cocoa Beach, Florida, November 1965.

14-27. Golovin, N. E.: Reliability Engineering and Success in Space Exploration. Industrial Quality Control - Journal of the ASQC, Vol. 22, No. 9, March 1966.

14-28. Klapheke, J. W.; Spradlin, B. C.; and Easterday, J. L.: Preindications of Failure in Electronic Components. Report RSIC-445 (DDC No. AD-472 738), Batelle Memorial Institute, Columbus, Ohio, July 1965.

14-29. Hanley, L. D.; Hall, E. C.; and Partridge, J.: Progress Report on Attainable Reliability of Integrated Circuits for Systems Application. Paper presented at the Symposium on Microelectronics and Large Systems, Washington, D. C., November 1964.

References (Continued)

14-30. Lombardi, J.; McDonough, L.; and H. Padden: High Reliability Screening of Semiconductor and Integrated Circuit Devices. NASA CR-721, Grumman Aircraft Engineering Corp., Bethpage, N. Y., April 1967.

14-31. Ryerson, C. M.: Relative Costs of Different Reliability Screening Techniques, 1967 Annual Symposium on Reliability, January 1967.

14-32. Perry, James N.: Reliability Screening Techniques. 1966 Annual Symposium on Reliability, January 1966.

14-33. Solvason, J. C.: Parts Testing Methods Study Report. Report LMSC - A667574, Lockheed Missiles and Space Company, Air Force Contract AF 04(647)-787, July 1964.

14-34. Lockyear, W. H.: The Mariner Mars Parts Screening Program. Jet Propulsion Laboratory, Pasadena, California, May 1966.

14-35. Nowak, T. J.: Reliability of Integrated Circuits by Screening. 1967 Annual Symposium on Reliability, January 1967.

14-36. Davidson, K. W. and Nagy, William: Nondestructive Reliability Screening of Semiconductor Diodes. Rome Air Development Center, Rome, N. Y., Technical Report No. RADC-TR-63-526, (DDC No. AD-484 278), May 1966.

14-37. Norris, R. H.: Run-In or Burn-In of Electronic Parts - A Comprehensive, Quantitative Basis for Choice of Temperature, Stresses, and Duration. Ninth National Symposium on Reliability and Quality Control, San Francisco, California, January 1963.

14-38. Myers, R. H.; Wong, K. L.; and Gordy, H. M.; Eds.: Reliability Engineering for Electronic Systems. John Wiley and Sons, Inc., New York, 1964.

14-39. Ireson, Grant W., Ed.: Reliability Handbook. McGraw-Hill Book Co., New York, 1966.

14-40. Parker, D. C.; et.al.: Final Report, Development of Reliability Methodology for Systems Engineering, Vol. II - Application: Design Reliability Analysis of a 250 volt-ampere Static Inverter. (N66-28043), April 1966.

14-41. Tommerdahl, J. B. and Nelson, A. C.: Switching Circuit Transient Performance Prediction Using Empirical Mathematical Modeling Techniques. IEEE Transactions on Reliability, Vol. R-14, March 1965.

14-42. Britt, C. L.: Research in the Functional Analysis of Systems. Annual Conference Transactions, ASQC, 1965.

APPENDIX

APPENDIX

Statistics of Testing

The engineer attempts to build systems with as much determinism and certainty as possible. In testing and operating these systems variability and uncertainty become commonplace. Typical sources of this are:

(1) inability to build items identical,

(2) lack of precision and accuracy in measurements,

(3) variability of an item during repeated observations,

(4) lack of control of operating conditions, and

(5) inability to model and account for all real-world conditions.

Through design and application it is usually possible to render many variations and uncertainties negligible in importance. Tolerances and specifications play a large role in controlling this. However, the engineer often has to live with some of the variability and uncertainty. It is thus important in either case to determine which are important and assess them. Testing, of course, plays a large role in this.

Mathematical and statistical techniques are provided in this section to aid in planning tests, analyzing the data, and subsequently interpreting the test results. The extent and application of these formal procedures depends upon the factors introduced in the main discussion of the report. All three activities can be performed with a wide range of degrees of sophistication. For example, analyses may range from simple tabulations and plots to complex calculations with formal statistical techniques for estimation of unknown parameters or testing hypotheses concerning these parameters. The degree of the complexity of the analysis depends on such factors as the requirements of the customer, expense of the item, expense and formality of the test procedures, and the degree and importance of the variation of the test data. Similar complexities apply to planning and interpretation of results.

In order to make the written material as brief as possible summary tables have been prepared to cover specific topics such as continuous variables, Boolean algebra, calculus of probabilities, interval estimation, sampling plans, etc. Supporting each of these tables are cited references, discussion and examples demonstrating the techniques in the corresponding table. The discussion will briefly treat some of the more frequent problem areas which are often treated poorly in the reliability literature. Typical topics of this type are the meaning of confidence interval estimates, the difference between engineering and statistical confidence, and the use of the random failure law.

A thorough treatise on statistics of testing would normally begin with an introduction to the calculus of probabilities as background theory. For brevity we have chosen to start immediately with application to random variables and proceed quickly to topics such as estimation and sampling more pertinent to testing. A summary treatment of basic probability concepts is given in the appendix of Volume IV - Prediction.

A.1 Continuous Random Variables and Distributions

A continuous random variable is typically one that can take on any value in an interval. For example, the lifetime of a transistor under a certain set of test conditions could be any time greater than zero. For a very large population of transistors one would expect the lives to be scattered or distributed over a large interval of time. Continuous distribution functions are used to describe such statistical behavior. Table A.1-1 summarizes basic concepts concerning continuous random variables and their distributions. A summary of several common distributions is presented in Table A.1-2. Given a density function $p(x)$ the characteristics can be evaluated by application of the formulae in Table A.1-1.

Central Limit Theorem

One of the most important results in statistics is the central limit theorem (CLT) which states that if $x_1, x_2, \ldots, x_n$ are independent random variables all having the same distribution function $F(x)$ with mean μ and standard deviation σ, then the sum

$$s = \sum_{i=1}^{n} x_i$$

is asymptotically Normally distributed with mean $n\mu$ and standard deviation $\sigma\sqrt{n}$, i.e.,

$$P(s < s_0) = \frac{1}{\sqrt{2\pi}\ \sigma\sqrt{n}} \int_{-\infty}^{s_0} \exp\{ \frac{-1}{2\sigma^2 n} (s - n\mu)^2\} ds$$

for n sufficiently large. This result is true under very general conditions on $F(x)$; if all variables have the same distribution then it is sufficient that the second moment of x be finite. A more general form of the CLT and additional discussion of the above case appear in Ref. 52.. An important aspect of the theorem is how large n must be before the normal approximation applies. Clearly this dependence on n is conditioned by the shape of the distribution. Sums of variables having highly skewed distributions would tend to Normality more slowly than for those having symmetrical or more nearly Normal distributions. In the latter case sums of variables with n larger than 25 or 30 are very closely approximated by the Normal distribution.

Table A.1-1

Continuous Random Variables And Distributions

1. x is a random variable (r.v.) having density function $p(x)$ and (cumulative) distribution function $F(x)$.*

2. $F(x) = \int_{-\infty}^{x} p(t)dt$ (t is a dummy variable) and

$$p(x) = \frac{dF(x)}{dx}.$$

3. Property: $F(-\infty) = 0;\ F(\infty) = 1.$

4. Specifically for the range R over which x is defined

$$\int_R p(x)dx = 1.$$

5. Probability:

$$P(x \leq a) = \int_{-\infty}^{a} p(x)dx = F(a)$$

$$P(a \leq x \leq b) = \int_{a}^{b} p(x)dx = F(b) - F(a).$$

6. Expectation: For any function $g(x)$,

$$E[g(x)] = \int_R g(x)\ p(x)dx.$$

7. Mean of x (first moment about the origin):

$$E(x) = \int_R x\,p(x)dx = \nu_1.$$

8. Mean square of x (second moment about the origin):

$$E(x^2) = \int_R x^2\,p(x)dx = \nu_2.$$

9. k-th moment of x with respect to the origin:

$$E(x^k) = \int_R x^k\,p(x)dx = \nu_k.$$

---

*In precise mathematical notation, X is used to denote a random variable, then $F(x) = P(X \leq x)$, and for a continuous variable $p(x)dx \approx P(x \leq X \leq x+dx)$.

10. Variance of x (second moment about the mean):

$$E\{[x-E(x)]^2\} = \sigma^2(x) = \int_R [x-E(x)]^2 p(x)dx = \mu_2.$$

11. k-th moment of x about the mean:

$$E\{[x-E(x)]^k\} = \int_R [x-E(x)]^k p(x)dx = \mu_k.$$

12. Relationship between the first four moments:

$$\mu_0 = \nu_0 = 1$$

$$\mu_1 = 0$$

$$\mu_2 = \nu_2 - \nu_1^2,\ \nu_1 = \text{mean value of } x$$

$$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3$$

$$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1 - 3\nu_1^4.$$

13. Truncated distribution, $F_T(x)$, of $F(x)$:

$$F_T(x) = \begin{cases} F(x)/F(T) & x \leq T \\ 1 & x > T. \end{cases}$$

<u>Example</u>

Let x be a random variable with density function

$$p(x) = \lambda e^{-\lambda x}, \qquad \lambda > 0,\ x \geq 0.$$

This is the well-known Weibull density function with $\theta = 1/\lambda$ and $k = 1$ or the negative exponential density function.

Distribution: $$F(x) = \int_0^x \lambda e^{-\lambda t}dt = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}.$$

Probability: $P(1 \leq x \leq 2) = \int_1^2 \lambda e^{-\lambda x} dx = e^{-\lambda} - e^{-2\lambda}$

or

$$F(2) - F(1) = (1-e^{-2\lambda}) - (1-e^{-\lambda}) = e^{-\lambda} - e^{-2\lambda}.$$

Mean: $E(x) = \int_0^\infty \lambda x e^{-\lambda x} dx = \frac{\Gamma(2)}{\lambda} = \frac{1}{\lambda}, \; (\Gamma(k) = (k-1)!)$.

Variance: $\sigma^2(x) = \int_0^\infty (x - 1/\lambda)^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}$.

k-th moment about the origin:

$$\nu_k = \int_0^\infty x^k \lambda e^{-\lambda x} dx = \frac{(k-1)!}{\lambda^k} .$$

Table A.1-2

Continuous Density Functions and Associated Characteristics

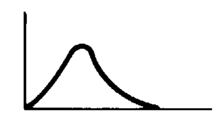
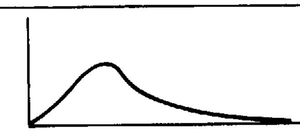
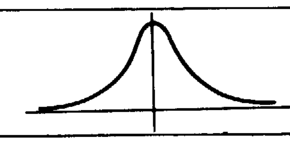
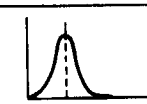
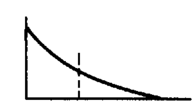
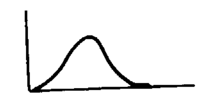
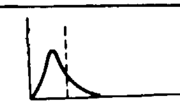
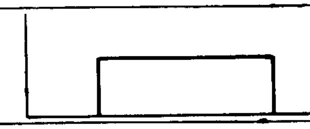
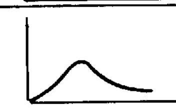
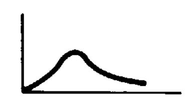
Type Distribution	Density Function f(x)	Mean E(x) and Variance V(x)	Graph of Typical p(x) or Standard Form
Beta (Standard Form)	$p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \; 0 \leq x \leq 1$ $= 0$ elsewhere $\alpha, \beta > 0$	$E(x) = \frac{\alpha}{\alpha+\beta}$ $V(x) = \frac{\alpha \cdot \beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	
F	$p(x) = \left(\frac{m}{n}\right)^{m/2} x^{(m/2)-1} / B\left(\frac{m}{2}, \frac{n}{2}\right)\left[1 + \left(\frac{m}{n}x\right)\right]^{(m+n)/2}, \; 0 \leq x \leq \infty$ $= 0$ elsewhere	$E(x) = \frac{n}{n-2}$ $V(x) = \frac{n^2[(n-2) - m(n-4)]}{m(n-2)^2(n-4)}$	
t	$p(x) = \left(\frac{1}{n}\right)^{1/2} \frac{1}{B\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \frac{1}{(1+\frac{1}{n}x^2)^{(n+1)/2}}, -\infty \leq x \leq \infty$	$E(x) = 0$ $V(x) = \frac{n}{n-2}$	
Standard Normal (or Gaussian)	$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp[-(x-\mu)^2/\sigma^2], \; -\infty \leq x \leq \infty$	$E(x) = \mu$ $V(x) = \sigma^2$	
Exponential	$p(x) = \lambda e^{-\lambda x}, \; x \geq 0, \; \lambda > 0$ or $p(x) = \frac{1}{\theta} e^{-x/\theta}, \; x \geq 0, \; \theta > 0$ where $\theta = 1/\lambda$.	$E(x) = 1/\lambda = \theta$ $V(x) = 1/\lambda^2 = \theta^2$	

Table A.1-2 (Continued)

Type Distribution	Density Function f(x)	E(x), V(x)	Graph of Typical p(x) or Standard Form
Weibull	$p(x) = \frac{kx^{k-1}}{\theta^k} \exp\left[-\left(\frac{x}{\theta}\right)^k\right], \; x \geq 0$	$E(x) = \theta\Gamma(\frac{1}{k} + 1)$ $V(x) = \theta^2\, \Gamma(\frac{2}{k} + 1) - \theta^2\, \Gamma^2(\frac{1}{k} + 1)$	
Lognormal	$p(x) = \frac{1}{x\sqrt{2\pi}\sigma} \exp\{-[(\ell nx - \mu)^2/2\sigma^2\}, \; 0 \leq x < \infty$	$E(x) = \exp[\mu + \sigma^2/2]$ $V(x) = \exp[2\mu + \sigma^2]\,(e^{\sigma^2} - 1)$	
Uniform	$p(x) = \frac{1}{h}, \; 0 \leq x \leq h$	$E(x) = h/2$ $V(x) = h^2/12$	
Gamma	$p(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \; x \geq 0$	$E(x) = \beta\alpha$ $V(x) = \beta^2\alpha$	
Generalized Gamma	$p(x;\alpha,\beta,\gamma) = \frac{\gamma}{\beta^\alpha\, \Gamma(\alpha/\gamma)} x^{\alpha-1} \exp[-(x/\beta)^\gamma],$ $0 \leq x < \infty, \alpha, \beta, \gamma > 0$	$E(x) = \beta\Gamma((\alpha+1)/\gamma)/\Gamma(\alpha/\gamma)$ $V(x) = \beta^2\Gamma((\alpha+1)/\gamma)/\Gamma(\alpha/\gamma)$ $- \beta^2\Gamma^2((\alpha+1)/\gamma)/\Gamma^2(\alpha+\gamma).$	

A.2 Discrete Random Variables and Distributions

A discrete random variable is one that takes on a finite or a countably infinite number of values. For example, a binomial variable takes on two values corresponding to a success or a failure, such as tossing a coin and the occurrence of a head being a success. On the other hand, the number of telephone calls on a given line for a specified time may be approximated by a Poisson variable for time intervals of "constant density". The number of calls might be considered to take on any one of a countably infinite number of values, 0, 1, 2, ..., etc.

Table A.2-1 summarizes the definitions and notation for the characteristics of distributions of discrete random variables. Table A.2-2 contains some of the common discrete distributions and the means and the variances. Ref. 53 contains a complete discussion of many discrete random variables and the pertinent characteristics.

<u>Example</u>

Suppose that it is desired to obtain the probability of three or fewer failures in a time interval of length t where an item upon failure is replaced by a new item. Suppose further that the exponential failure time distribution is applicable. Let the failure rate be λ = 0.01/hour and the time be 200 hours.

From the above information the mean or expected number of failures is 2 items. Furthermore the probability of x failures is given by the Poisson formula and thus for three or fewer failures the probability is expressed as

$$P(x \leq 3) = \frac{e^{-2}2^0}{0!} + \frac{e^{-2}2^1}{1!} + \frac{e^{-2}2^2}{2!} + \frac{e^{-2}2^3}{3!}$$

$$= 0.8569.$$

Table A.2-1

Basic Definitions Concerning Discrete Distributions

1.	Random variable	$x_i,\ i = 0, 1, 2, \ldots, \infty\ (x_i < x_j,\ i < j)$	(may also be finite)
2.	Probability density function	$p(x_i),\ i = 0, 1, \ldots, \infty$	
3.	Cumulative distribution function	$P(x_i) = \sum_{j=0}^{i} p(x_j)$	
4.	Expected value (mean) of x	$E[x] = \sum_{i=0}^{\infty} x_i\ p(x_i)$	
		$= \sum_{i=0}^{n} x_i\ p(x_i)$	if finite number of possibilities. In the following results the index of summation is omitted and the summation over all possible values (finite or infinite) is implied.
5.	Variance of x	$\sigma^2[x] = E[x^2] - E^2[x]$	
		$= \Sigma x_i^2\ p(x_i) - [\Sigma x_i\ p(x_i)]^2$	
	or	$\sigma^2[x] = E[x(x-1)] - E^2[x] + E[x]$	using factorial moments for an integral valued discrete variable where $E[x(x-1)] = \Sigma x_i(x_i-1)\ p(x_i)$
6.	Expected value of g(x)	$E[g(x)] = \Sigma g(x_i)\ p(x_i)$	
7.	k-th moment about the origin	$\nu_k = E(x^k) = \Sigma x_i^k\ p(x_i)$	
8.	k-th moment about the mean	$\mu_k = E[(x-\nu_1)^k]$	
		$= \Sigma(x_i-\nu_1)^k\ p(x_i)$	

Table A.2-2

Discrete Frequency Functions, Means and Variances

	Type Distribution	P(x occurrences)	Mean	Variance
1.	Binomial, $B(x;\ n,\ p)$	$\binom{n}{p} p^{x}(1-p)^{n-x},\ 0 \le x \le n.$	np	npq
2.	Poisson, $P(x;\ \mu)$	$e^{-\mu}\mu^{x}/x!\quad,\ x \ge 0.$	μ	μ
3.	Hypergeometric, $H(x;\ n,\ N_1,\ N_2)$	$\binom{N_1}{x}\binom{N_2}{n-x}\Big/\binom{N_1+N_2}{n},\ 0 \le x \le N_1.$	$N_1 n/(N_1+N_2)$	$\frac{N_1 N_2}{(N_1+N_2)^2}\ \frac{n(N_1+N_2-n)}{(N_1+N_2-1)}$
4.	Geometric, $G(x;\ p)$	$(1-p)^{x-1}p\quad,\ x \ge 1.$	$1/p$	$\frac{1-p}{p^2}$
5.	Negative Binomial, $NB(x;\ p,\ r)$	$\binom{x-1}{r-1} p^{r}(1-p)^{x-r}\quad,\ x \ge r.$	r/p	$\frac{r(1-p)}{p^2}$

A.3 Multivariate Distributions (Emphasis on Bivariate Case)

Consider the situation in which two or more measurements on a part are being obtained, e.g. the equivalent h-parameters of a transistor. These two measurements would have a joint probability density function (pdf) $p(x, y)$, say, where x and y denote the respective measurements. If the two variables are statistically independent then

$$p(x, y) = p_1(x)\, p_2(y),$$

and hence the joint density functions can be written down knowing the individual pdf's. If the variables are not independent the multivariate density function can be obtained by assuming a particular form such as the Normal density function and estimating the unknown parameters from available data.

Most of the properties of bivariate (two-variate) distributions are straightforward generalizations of the univariate distributions given earlier. The new concepts are those of conditional and marginal distributions, covariance and correlation. The generalization of these results to multivariate distributions is easily made and one should see Ref. 51 for these results.

Independent Random Variables. If two variables x and y are independent then the covariance of x and y, denoted by $\mathrm{Cov}(x, y)$ is

$$\mathrm{Cov}(x, y) = \iint (x-E(x))\, p_1(x)(y-E(y))\, p_2(y)\,dx\,dy = 0.$$

However the inverse is not true, i.e. two variables may have zero covariance (or zero correlation i.e. $\rho(x, y) = 0$) but not be independent. For example, suppose that u and v are independent variables, and let $x = u + v$, $y = u - v$. Then

$$E(xy) = E(u^2) - E(v^2) = 0, \; E(y) = 0, \text{ and}$$

$$\mathrm{Cov}(x, y) = 0 \text{ and } \rho(x, y) = 0.$$

However, x and y are dependent. See Ref. 17 for additional examples. Thus the correlation is not a general measure of dependence but rather a measure of linear dependence of two variables in physical terms; the correlation coefficient is a dimensionless covariance.

Table A.3-1

Bivariate Distributions

1. Let x, y be a pair of random variables having the joint distribution function $F(x, y)$ and density function $p(x, y)$.

2. $\iint_R p(x, y)\, dxdy = 1$, where R is the region over which x and y are defined.

3. $\int_{-\infty}^{y}\int_{-\infty}^{x} p(u, v)dudv = F(x, y)$.

4. $F(-\infty, -\infty) = 0$, $F(\infty, \infty) = 1$.

5. $p(x, y) = \dfrac{\partial F(x, y)}{\partial x \partial y}$.

6. $$P(a < x < b, c \leq y \leq d) = \int_c^d \int_a^b p(x, y)dx\, dy$$

$$= F(b, d) + F(a, c) - F(a, d) - F(b, c).$$

7. $E(g(x, y)) = \iint_R g(x, y)\, p(x, y)dxdy$.

8. $E(x) = \iint_R xp(x, y)dxdy$.

9. If x and y are independent random variables (r.v.'s) then

$$p(x, y) = p_1(x)\, p_2(y) \text{ and}$$

$$E(x) = \int x\, p_1(x)dx \text{ and } E(y) = \int y\, p_2(y)dy.$$

10. $$E(xy) = \iint_R xy\, p_1(x)\, p_2(y)dxdy$$

$$= E(x)\, E(y) \text{ if x and y are independent r.v.'s.}$$

11. $E(x - E(x))^2 = \sigma^2(x)$, $E(y - E(y))^2 = \sigma^2(y)$.

12. $E\{(x - E(x))(y - E(y))\} = \mathrm{Cov}(x, y) = $ Covariance of x and y

$$= \iint [x - E(x)][y - E(y)]\, p(x, y)dxdy.$$

13. Correlation of x and y $= \rho\{x, y\} = \mathrm{Cov}(x, y)/\sigma(x)\,\sigma(y)$
where

$$\sigma(x) = [\sigma^2(x)]^{1/2} \text{ and } \sigma(y) = [\sigma^2(y)]^{1/2}.$$

14. Marginal distribution of x is given by

$$p_1(x) = \int_{R_y} p(x, y)dy.$$

15. The conditional distribution of y for given x is given by

$$p(y|x) = \frac{p(x, y)}{p_1(x)}$$

$$= p_2(y) \text{ if x and y are independent r.v.'s.}$$

Example

Let x and y have a bivariate density function

$$p(x, y) = \frac{1}{2\pi\sqrt{1-c^2}} \exp\{-\frac{1}{2(1-c^2)}(x^2 - 2cxy + y^2)\}.$$

First of all note that

$$\int_R\int p(x, y)dxdy = 1$$

since by completion of the square of the exponent

$$p(x, y) = \frac{1}{2\pi\sqrt{1-c^2}} \iint \exp\{-\frac{1}{2(1-c^2)}(x^2 - 2cxy + c^2y^2)$$

$$+ \frac{(c^2 - 1)}{2(1-c^2)} y^2\}dxdy.$$

If the variables are transformed as follows:

$$u = (x - cy)/\sqrt{1-c^2}$$

$$v = y$$

then

$$p(u, v) = \frac{1}{2\pi} \iint \exp\{-(\frac{u^2}{2} + \frac{v^2}{2})\} du\, dv, \qquad \text{(A.3-1)}$$

using the fact that the Jacobian of the transformation is given by

$$1/ \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = 1/ \begin{vmatrix} 1/\sqrt{1-c^2} & -c/\sqrt{1-c^2} \\ 0 & 1 \end{vmatrix} = \sqrt{1-c^2} .$$

(A-3.1) can be written as the product of the integrals

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} du \bullet \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{v^2}{2}} dv .$$

Since each is the integral of the standard Normal density function the above product is unity.

Next the marginal distribution of y is given by

$$\begin{aligned} p_2(y) &= \int p(x, y)\,dx \\ &= \frac{1}{\sqrt{2\pi}} \exp\{ - \frac{y^2}{2} \} \quad . \end{aligned}$$

Hence the conditional distribution of x given y is

$$p(x|y) = \frac{p(x, y)}{p(y)} = \frac{1}{\sqrt{2\pi(1-c^2)}} \exp\{- \frac{1}{2(1-c^2)} (x - cy)^2\}.$$

<u>Mean, Variance and Covariance Formulas</u>

Let $x_1, x_2, \ldots, x_n$ be n random variables with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ respectively and correlations ρ_{12} (= $\rho(x_1, x_2)$), $\rho_{13}, \ldots, \rho_{n-1, n}$. The following results are true <u>independent of the distributions of the variables</u>. Let y be a linear combination of the variables given by

$$y = c_0 + c_1x_1 + c_2x_2 + \ldots + c_nx_n.$$

Then the mean and variance of y are denoted by μ_y and σ_y^2 and are given by

$$\mu_y = c_0 + c_1\mu_1 + c_2\mu_2 + \dots + c_n\mu_n = c_0 + \sum_{i=1}^{n} c_i\mu_i$$

$$\sigma_y^2 = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \dots + c_n^2\sigma_n^2$$

$$+ 2\rho_{12}c_1c_2\sigma_1\sigma_2 + \dots + 2\rho_{n-1,n}c_{n-1}c_n\sigma_{n-1}\sigma_n$$

or

$$\sigma_y^2 = \sum_{i=1}^{n} c_i^2\sigma_i^2 + 2\sum_{i<j}\sum c_ic_j\rho_{ij}\sigma_i\sigma_j.$$

where σ_i is the standard deviation of the i-th variable. The above formulas are true in general and one notes that the mean μ_y of y does not involve the correlations.

Now if the variables are uncorrelated (if they are independent as indicated previously) the formula for the variance reduces to

$$\sigma_y^2 = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \dots + c_n^2\sigma_n^2.$$

Now consider two functions

$$y = c_0 + c_1x_1 + \dots + c_nx_n$$

$$w = \ell_0 + \ell_1x_1 + \dots + \ell_nx_n,$$

then the covariance of y and w is given by

$$\mathrm{Cov}\{y, w\} = c_1\ell_1\sigma_1^2 + \dots + c_n\ell_n\sigma_n^2 + \sum_{j=1}^{n}\sum_{i=1}^{n} \ell_jc_i\sigma_j\sigma_i\rho_{ij}.$$

If the functions are not linear it is often possible to use a Taylor series expansion of the function $f(\underline{x})$ and then apply the mean and variance computations to this form. These formulas must be used with care, e.g. by checking the magnitude of the errors which may result in using them. Thus if

$$y = f(\underline{x})$$

then

$$y \simeq f(\underline{\mu}) + \sum \frac{\partial f}{\partial x_i}\bigg|_{\underline{\mu}} \Delta x_i + \frac{1}{2}\sum \frac{\partial^2 f}{\partial x_i^2}\bigg|_{\underline{\mu}} \Delta x_i^2$$

$$+ \frac{1}{2} \sum_{i \neq j} \sum \frac{\partial^2 f}{\partial x_i \partial x_j} \Bigg|_{\underline{\mu}} \Delta x_i \Delta x_j, \quad \Delta x_i = x_i - \mu_i,$$

and hence using only the first order terms

$$\mu_y \simeq f(\underline{\mu})$$

$$\sigma_y^2 \simeq \sum \Big(\frac{\partial f}{\partial x_i} \Bigg|_{\underline{\mu}} \Big)^2 \sigma_i^2 ,$$

where

$$\underline{\mu} = (\mu_1, \mu_2, \ldots, \mu_n),$$

and where $\frac{\partial f}{\partial x_i} \Big|_{\underline{\mu}}$ denotes the evaluation of the derivative at $\underline{\mu}$.

The above results are summarized in the following table.

Table A.3-2

Mean, Variance, and Covariance Formulas

General Case for Single Function.

If $y = c_0 + c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$,

then $\mu_y = c_0 + c_1 \mu_1 + c_2 \mu_2 + \ldots + c_n \mu_n = c_0 + \sum_1 c_i \mu_i$

and $\sigma_y^2 = \sum_i c_i^2 \sigma_i^2 + \sum_{i} \sum_{\substack{j \\ i \neq j}} c_i c_j \sigma_i \sigma_j \rho_{ij}$.

Variables Uncorrelated.

$$\mu_y = c_0 + c_1 \mu_1 + \ldots + c_n \mu_n = c_0 + \sum_1^n c_i \mu_i$$

$$\sigma_y^2 = \sum_1^n c_i^2 \sigma_i^2 .$$

General Case for two functions.

If $y = c_0 + \sum c_i x_i$

and $w = \ell_0 + \sum \ell_j x_j$

then

$$\mathrm{Cov}(y,w) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i \ell_j \sigma_i \sigma_j \rho_{ij}.$$

If x_i and x_j are uncorrelated, i.e. $\rho_{ij} = 0$ for $i \neq j$, then

$$\mathrm{Cov}(y, w) = \sum c_i \ell_i \sigma_i^2.$$

General Case for single nonlinear function.

If $y = y(\underline{x})$, $\underline{x} = (x_1, \ldots, x_n)$

then using only a first order approximation

$$\mu_y \simeq y(\underline{\mu}), \ \underline{\mu} = (\mu_1, \ldots, \mu_n), \text{ vector of means,}$$

and

$$\sigma_y^2 \simeq \sum \left(\frac{\partial y}{\partial x_i} \bigg|_{\underline{\mu}} \right)^2 \sigma_i^2.$$

A.4 Sample Statistics

The basic statistics which describe a sample selected from a distribution **or** population are briefly summarized in Table A.4-1.

Table A.4-1

1. Mean $= \bar{x} = \sum x_i/n$,

    x_i - i-th observation.

    n number of observations in sample.

2. Standard Deviation $= s = \{\sum \{x_i - \bar{x}\}^2/(n-1)\}^{1/2}$,

    $n - 1$ is the number of degrees of freedom associated with the standard deviation.

3. $x_{(j)}$ is the j-th order statistic of the sample - the j-th smallest observation.

4. Range $= x_{(n)} - x_{(1)}$ where

    $x_{(n)}$ is the largest observation of the sample,

    $x_{(1)}$ is the smallest observation of the sample.

5. Sample distribution function is the proportion of the sample observation at or below a given value plotted as a step function related to the observed values of x, i.e.,

$$F_n(x) = \frac{\text{number of observations} \leq x}{n} .$$

See Figure A.4-1 for an illustration.

The arithmetic mean $\bar{x}$ of the sample is a measure of location or central tendency of the observations and the standard deviation s is a measure of the spread or dispersion of the observations. Thus s is a measure of precision of the method of measurement or of the instrument used. The accuracy of the measurement technique is the deviation between the sample mean for an infinitely large sample and the actual or true mean. The range is also a measure of dispersion and is frequently used in quality control as an index for identifying lack of adequate control in a process. It is an efficient estimate of dispersion when the underlying distribution is Normal and the sample sizes are small, i.e., when the number of observations is about four or five. For most problems involving one variable or measurement, the mean and standard deviation are adequate to describe the sample observations. However, higher order moments would be required in fitting particular distributions or series approximations by the method of moments.

The sample distribution function is frequently plotted on probability paper which has an appropriate scale transformation on the ordinate (or abscissa) in order that the graph connecting the points $(x, F_N(x))$ will be a straight line if the particular distribution is the correct or true distribution. For small samples considerable variation of the plotted points from a straight line is expected. However, for reasonably large samples the graph paper is very useful in discriminating between possible distributions. In lieu of using probability paper it is possible to use the tables of expected values of the order statistics. The r-th order statistic is the r-th smallest value in the sample of observations.

Example

Suppose that 15 observations on the performance of a system are obtained by constructing 15 systems and measuring their performances. The results in n sec are as follows:

59.2	61.2
50.6	81.5
57.8	61.2
51.1	68.0
63.6	78.7
48.9	73.7
51.0	68.6
	81.3.

The characteristics of this sample of measurements are summarized by the following statistics.

$$\text{Mean} = \bar{x} = 63.76 \text{ n sec}$$

$$\text{Standard Deviation} = s = 11.25 \text{ n sec}$$

$$\text{Range} = w = 81.5 - 48.9 = 32.6 \text{ n sec.}$$

The sample distribution function is given in Figure A.4-1.

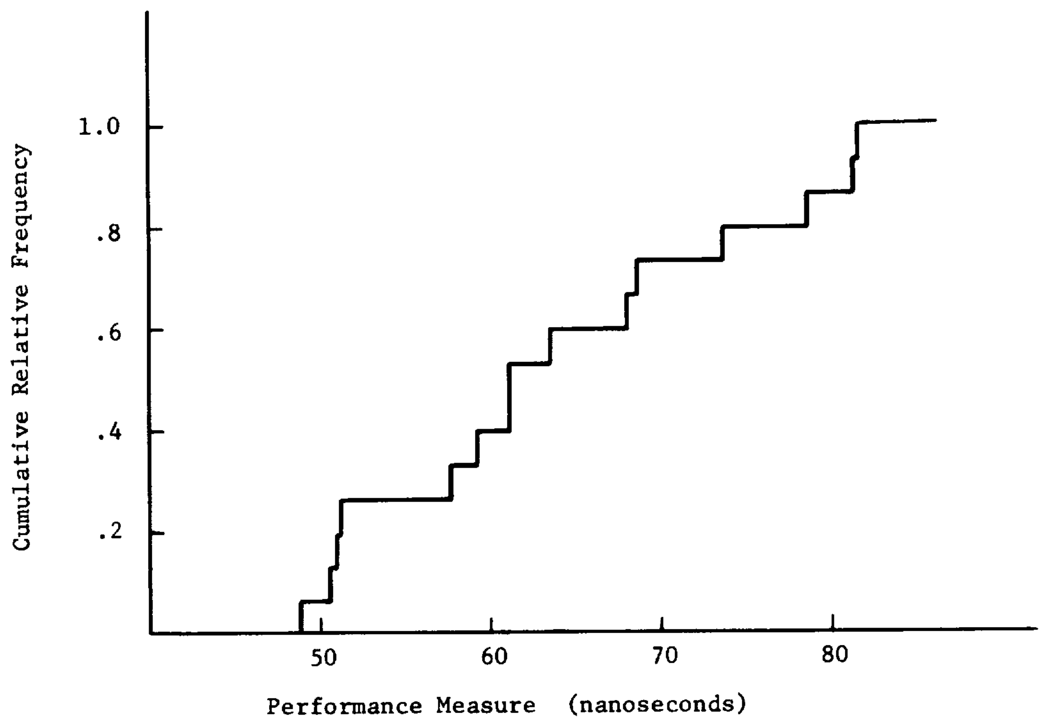


Figure A.4-1 Sample Distribution Function

A.5 Point Estimation

The primary problem of interest in point estimation is that of obtaining a function of the sample observations which estimates the parameter(s) of interest. For example, a sample mean

$$\bar{x} = \sum x_i/n = g(\underline{x})$$

is an unbiased estimate of the mean of the distribution of x. Three methods are given for the derivation of the function $g(\underline{x})$. The problem is summarized in Table A.5-1. All methods are frequently used, and in some examples some of the methods may be equivalent i.e. yield the same estimator $g(\underline{x})$. The desired properties are somewhat obvious after examination, and much of the literature on estimation deals with these properties with respect to particular estimators.

Table A.5-1

Point Estimation

1. Let $f(x; \theta)$ be a probability density function of x with a single parameter θ which represents some characteristic of the distribution such as the mean.
2. $\hat{\theta} = g(x_1, \ldots, x_n)$, a function of the sample observations, is an estimator of θ.
3. $\hat{\theta} = g(\underline{x})$ may be obtained by one of several methods, three (3) of which are given:
   (a) Maximum likelihood - Form the likelihood function of the sample, take its natural logarithm, differentiate, and equate to zero.

$$L(x_1, \ldots, x_n; \theta) = f(x_1; \theta)\, f(x_2; \theta) \ldots f(x_n; \theta)$$

$$\frac{\partial \ \ell n \ L}{\partial \theta} = 0 \text{ yields } \hat{\theta} = g(\underline{x})$$

   (b) Least squares - Form the sum of squares of deviations, differentiate with respect to the parameter and equate to zero.

Let $$S = \sum (y_i - f(x_i, \theta))^2$$

then $$\frac{\partial S}{\partial \theta} = 0 \text{ yields } \hat{\theta} = g(\underline{x}).$$

   (c) Method of moments - Equate a convenient number of sample moments to the corresponding moments of the distributions which are functions of the unknown parameters.

Desired Properties of Estimators

1. Unbiasedness: $E(\hat{\theta}) = \theta$

2. Consistency: $\hat{\theta}_n$ converges in probability to θ, that is, for any $\varepsilon \leq 0$,

$$P(|\hat{\theta}_n - \theta| > \varepsilon) \to 0 \qquad \text{as } n \to \infty.$$

3. Sufficiency: $\hat{\theta}_n$ summarizes all the relevant information in the sample with respect to the parameter θ. See A-5 for a mathematical definition of sufficiency.

4. Efficiency: $\hat{\theta}$ is an efficient estimator if it has minimum variance, i.e.

$$E(\hat{\theta} - \theta)^2 \text{ is a minimum.}$$

Example

Point Estimation by Method of Maximum Likelihood

Let x be a random variable having the Normal or Gaussian probability density function

$$f(x, \underline{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-(x-\mu)^2/2\sigma^2\}.$$

In this example $\underline{\theta} = (\sigma, \mu)$ and it is desired to estimate σ and μ with appropriate functions of the sample observations based on the method of maximum likelihood. For n samples, the likelihood function is

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\{-(x_i-\mu)^2/2\sigma^2\}.$$

$$= (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\{-\sum(x_1-\mu)^2/2\sigma^2\},$$

and

$$\ln L(x_1, \ldots, x_n; \theta) = -\frac{n}{2} \ln 2\pi - n \ln\sigma - \frac{1}{2\sigma^2} \sum (x_i-\mu)^2.$$

The partial derivatives of $\ln L$ with respect to σ and to μ are

$$\frac{\partial \ln L}{\partial\sigma} = -\frac{n}{\sigma} + \frac{\sum (x_i-\mu)^2}{\sigma^3}$$

$$\frac{\partial \ln L}{\partial\mu} = -\frac{1}{2\sigma^2} \sum 2(x_i-\mu)(-1).$$

A.6 Interval Estimation and Confidence Limits

One of the most important problems in statistical inference is that of obtaining an interval estimate of some parameter of an assumed distribution. For example, it may be desired to estimate the probability of success for a given event when the number of successes has a binomial distribution, or to estimate the mean lifetime of a particular item when lifetime has a Weibull distribution. To obtain an interval estimate, the usual procedure is to use the distribution of the estimator for given values of the parameter being estimated, (See Table A.6-1). For example, in the case of the Normal distribution with unknown mean μ and known standard deviation σ the statistic

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

where n is the sample size and $\bar{x}$ is the sample mean, has a normal distribution with mean 0 and standard deviation 1. Hence a $100(1-\alpha)$, $0 < \alpha < 1$, percent lower confidence limit can be obtained for μ by

$$\bar{x} - z_{1-\alpha}\, \sigma/\sqrt{n},$$

or the corresponding inference is expressed as

$$\mu > \bar{x} - z_{1-\alpha}\, \sigma/\sqrt{n},$$

where $z_{1-\alpha}$ is the value of the standard normal deviate below which $1-\alpha$ proportion of the values of the distribution fall. This is a one-sided interval estimate of μ in that the interval is $(\bar{x} - z_{1-\alpha}\, \sigma/\sqrt{n}, \infty)$ and it is of infinite length. A two-sided interval can be obtained by a similar procedure and thus one infers that the mean μ falls in the interval

$$\bar{x} - z_{1-\frac{\alpha}{2}}\, \sigma/\sqrt{n} < \mu < \bar{x} + z_{1-\frac{\alpha}{2}}\, \sigma/\sqrt{n}$$

with $100(1-\alpha)$ percent confidence. This interval is finite in length and its end points are symmetrical with respect to the sample mean $\bar{x}$ the center of the interval. It is also based on the underlying assumption that the n measurements are an independent sample from the same Normal distribution.

The notion of a confidence interval is simple, but it is often confused in the literature. If one constructs a 95% confidence interval estimate of a parameter based on a certain set of data, and repeats this for another set of data obtained independently, etc. for many times, then <u>over the long run</u> 95% of the statements made will be

Equating the above derivatives to zero yields

$$\hat{\sigma}^2 = \sum (x_i - \hat{\mu})^2/n$$

$$\hat{\mu} = \sum x_i/n = \bar{x}.$$

Note that the maximum likelihood estimates are not necessarily unbiased and that in particular

$$E\{\hat{\sigma}^2\} = (n-1)\sigma^2/n,$$

hence

$$s^2 = \sum (x_i - \bar{x})^2/(n-1)$$

is an unbiased estimate of σ^2.

correct as stated and 5% will be incorrect. A statement is correct if the parameter does actually fall in the interval as claimed. For example, in the case described above it is claimed that

$$\mu > \bar{x} - 1.645\sigma/\sqrt{n},$$

where $z_{0.95} = 1.645$. This statement would be correct 95% of the time that the procedure of taking a sample and making the above inference from the sample is repeated. The following table gives a brief mathematical description of the interval estimation procedure.

<u>Example</u>

Confidence Interval Estimation

Let x be a random variable (r.v.) having the Normal distribution. Suppose a sample of 10 items are selected at random from the Normal population, and a 95 percent confidence interval estimate of μ is to be obtained using the following data on survival time in hours of a certain item being tested.

1357, 1474, 1542, 1499, 1429, 1492, 1574, 1331, 1466, and 1547.

In order to obtain a confidence interval estimate of μ we use the fact that

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

has a t-distribution (see Table A.2-2) with n-1 degrees of freedom. Then using the procedures of Ref. A-6 a one-sided 95 percent confidence interval for μ is given by

$$(\bar{x} - t_{0.95(n-1)}\, s/\sqrt{n}, \infty).$$

Using the above data

$$x = 1471.1 \text{ hours}$$

$$s = 79.6 \text{ hours},$$

and

$$t_{0.95(n-1)} = 1.83$$

and thus

$$1517.2 > \mu > 1425.0$$

is a 95 percent confidence interval for μ. This particular interval either is correct, that is it includes μ, or is not correct. In the long run 95 percent of such statements would be correct if the 95 percent confidence level is used and if the statements are made on the basis of independent sets of data.

Table A.6-1

Interval Estimation

Confidence Intervals

Let $f(x; \theta)$ be the probability density function of a random variable x and let it be desired to estimate the parameter θ by a confidence interval with the property that the interval will include θ a certain proportion of the time (fixed in advance).

If the upper and lower confidence limits (end-points of the interval) are θ_L and θ_U then let

$$P(\theta_L \leq \theta \leq \theta_U) = \gamma,$$

where γ is the level of confidence. θ_L and θ_U are functions of the sample observations, $g_L(\underline{x})$ and $g_U(\underline{x})$ respectively.

One-sided intervals can be obtained in either one of the forms,

$$P(\theta_L \leq \theta) = \gamma \text{ or } P(\theta \leq \theta_U) = \gamma.$$

These intervals can be obtained by means of the distribution of $\hat{\theta}$, the estimator of θ. Let this distribution be

$$h(\hat{\theta}; \theta).$$

Then the lower limit is obtained by finding $\theta = \theta_L$ such that $P(\hat{\theta} \geq \theta_d | \theta = \theta_L) = \alpha_1$, where θ_d is the value of θ obtained from the data.

Similarly the upper limit θ is obtained from the following relationship.

$$P\{\hat{\theta} \leq \theta_d | \theta = \theta_U\} = \alpha_2,$$

where $1 - (\alpha_1 + \alpha_2) = \gamma$ is the confidence level.

It is not necessary that the problem be of the same type as the example given above. In conclusion the percent confidence attached to the interval statement is the percentage of time that such a statement is expected to be correct or more simply interpreted as one's statistical confidence in the statement.

Engineering Significance Versus Statistical Significance

Suppose that a standard technique or process for making a particular device has yielded devices with a mean performance measure of 50 units. Let a new process be evaluated, a sample of measurements made, and suppose that the inference is made that the mean performance is 54 units and the lower 95 percent confidence limit for the mean is 51 units. One would thus conclude from a statistical viewpoint that the new process is an improvement over that of the standard technique. However the engineer must look more carefully into the inference and its meaning to him. For example, does the improvement of 4 units result in an increased profit? If the profits can be improved as a result of the small increase in quality, it is of advantage for a change to be made provided the change-over costs are not too great. Thus statistical significance does not automatically imply an engineering significance and that such decisions as the above can and should be carefully considered by the engineer in all such problems. It should be pointed out that the considerations of what magnitude difference in the means which is considered of importance to the engineer should reflect in the size of the sample or number of observations to be made. Clearly the detection of large differences, say 10% of the mean, requires very few observations relative to the number required to detect small differences, say, of the order of 2 or 3% change in the mean. Some of these concepts are treated in the subject of testing hypotheses in standard statistical tests.

Tolerance Interval

Another interval estimate frequently used in reliability problems is that of a statistical tolerance interval. Such an interval also has an attached confidence level associated with it but the interval is to include a certain desired proportion of the "population" sampled rather than just a particular parameter such as the mean. In the example given above a constant k can be determined from tolerance limit table [Ref. A-6] for which the statement that "a certain proportion of the population is included in the interval

$$(\bar{x} - ks, \bar{x} + ks)"$$

is correct with a given level of confidence and a prechosen level of the proportion. The confidence has the same interpretation as above, that is, the expected proportion of statements which are correct in the sense that the interval contains at least the stated percentage of the population.

A.7 Hypothesis Testing

The treatment of hypothesis testing parallels that of estimation. However, the question as to how large a sample should be taken is usually treated in hypothesis testing rather than in estimation. This question of sample size is usually the primary one asked of the statistician; the answer is not immediate because it is necessary to know precisely what is desired. For example, suppose that the hypothesis H_0 to be tested is that the mean lifetime θ of an item is larger than or equal to 1,000 hours, i.e.

$$H_0: \quad \theta \geq 1{,}000 \text{ hours.}$$

The question of how large a sample size should be taken to test this hypothesis is not meaningful until it is stated further what one wishes to accomplish. Suppose that it is desired to <u>reject</u> H_0 if $\theta < 500$ hours with high probability, say at least 0.90. The error associated with accepting H_0 when in fact it is not true is the β error or Type II error as shown in Table A.7-1. Assume further that the exponential distribution is an acceptable description of failure times of the item under consideration. Now there are two types of error associated with testing hypotheses, namely, the probability of rejecting H_0 when it is true, i.e., when $\theta \geq 1{,}000$ hours. This error is referred to as the α error or Type I error. The following diagram illustrates the situation for this example.

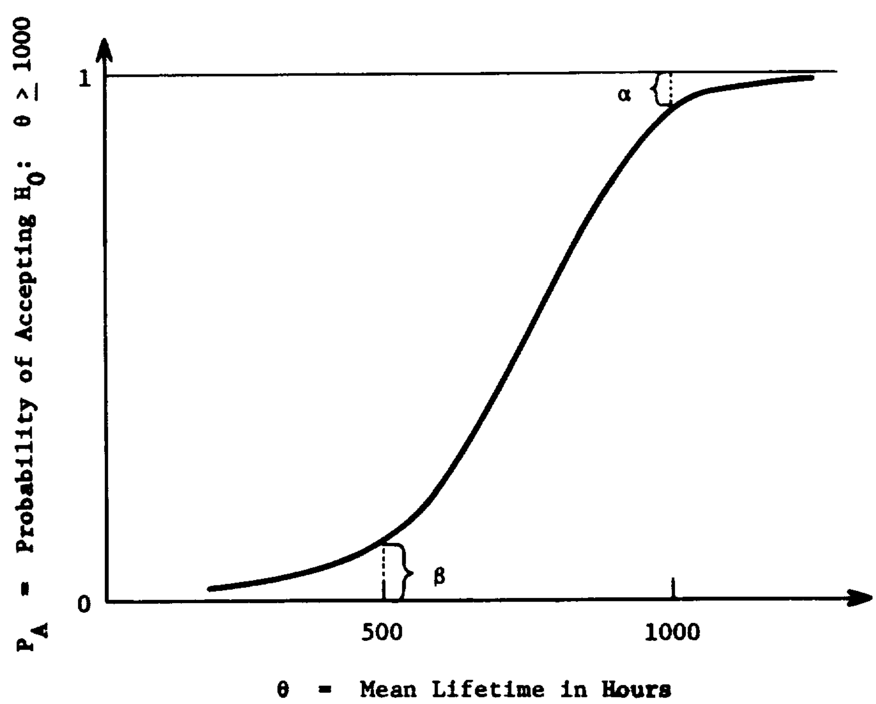


Figure A.7-1 P_A Versus θ

Some further questions must be answered prior to determination of the number of items to be tested. What type of life test is to be used? For example, n items may be tested until failure, n items may be tested until f failures occur, n items may be tested for a test time T_t, or items may be tested sequentially one-at-a-time until failure, etc. Sèction 8.1 on Life Testing in Volume 3 describes various test designs and gives references to the appropriate analyses. For this example, assume that it is desired to test n items until all items fail; then it is known from Ref. A-7 that

$$\chi^2 = \frac{2T}{\theta},$$

where T is the total test time and χ^2 has the χ^2 distribution with 2n degrees of freedom. Now if θ = 1,000 hours (or larger), then the probability of rejection should be 0.05 (or less than 0.5). Thus for θ = 1,000

$$\frac{2T}{1,000} = \chi^2_{0.05}$$

or the rejection region is given by

$$T = 500\chi^2_{0.05}(2n). \qquad \text{(A.7-1)}$$

If θ = 500 the probability of rejection should be 0.90 and hence

$$\chi^2 = \frac{2T}{500} \leq \chi^2_{.90}(2n). \qquad \text{(A.7-2)}$$

Equating the above expressions for T the following equation is obtained where 2n is the number of degrees of freedom or twice the sample size. Thus

$$500\chi^2_{0.05}(2n) = 250\chi^2_{0.90}(2n)$$

or

$$2\chi^2_{0.05}(2n) = \chi^2_{0.90}(2n)$$

and hence n = 19 items to be tested. This result can be obtained from a χ^2 table by observation. The ratio of $\chi^2_{0.90}$ to $\chi^2_{0.05}$ is found where it is near 2; then the smallest value of n is located in the table such that for 2n degrees of freedom the ratio is larger than 2. The important point to be made concerning the testing of hypotheses is that to determine the appropriate sample size other considerations such as allowed cost of testing, allowed error probabilities, etc. have to be taken into account.

Two curves are shown to indicate the effect of sample size. The α errors are assumed to be identical for each curve; however, the β or Type II errors are different as indicated $\beta(n_0) > \beta(n_1)$ for $n_0 < n_1$. Note further that β is a function of the deviation of the true θ from the hypothesized θ, $\theta = \theta_0$.

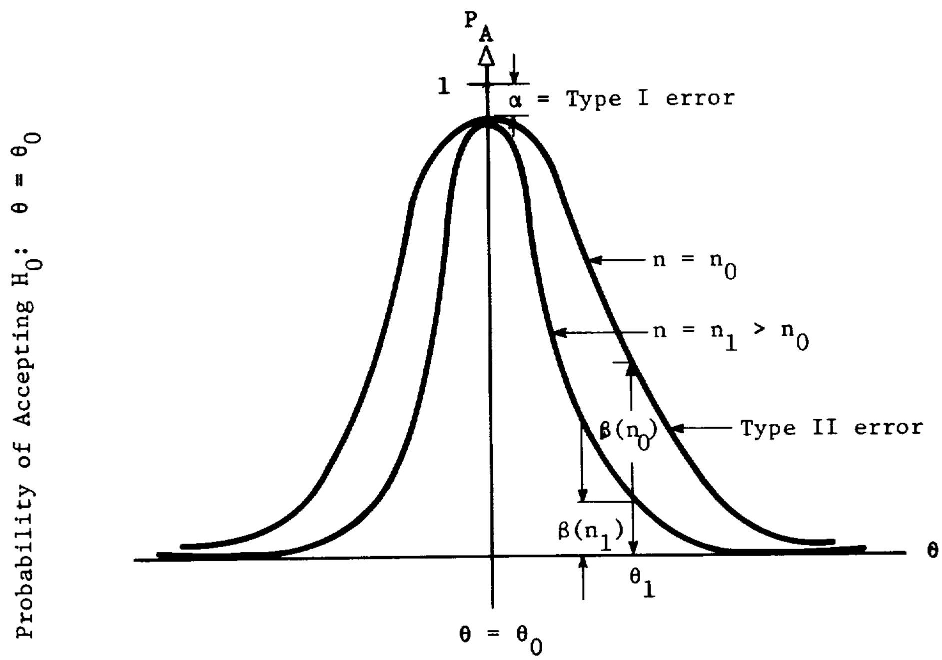


Figure A.7-2 Probability of Accepting H_0 Versus θ

Usually, one-sided tests of hypotheses are given and the curve of P_A versus θ is referred to as the operating characteristic (OC) curve. The values of α and β are indicated at $\theta = \theta_0$ and $\theta = \theta_1$, respectively. In the terminology of sampling α is referred to as the producer's risk and β as the consumer's risk.

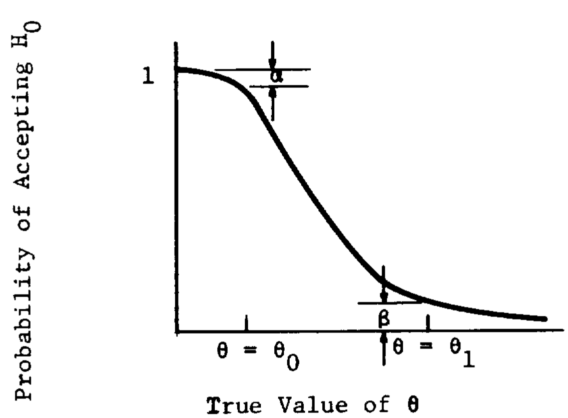


Figure A.7-3 Operating Characteristic (OC) Curve

Table A.7-1

Hypothesis Testing

Suppose that it is desired to test the hypothesis H_0 that the value of a parameter θ is equal to a specified value θ_0,

$$H_0: \theta = \theta_0$$

against an alternative hypothesis H_a that $\theta \neq \theta_0$,

$$H_a: \theta \neq \theta_0.$$

Let P_r be the probability of rejecting the hypothesis H_0, and
P_a be the probability of accepting the hypothesis H_0.

The hypothesis is rejected if the estimate $\hat{\theta}$ of θ falls in a critical region determined to satisfy desired properties.

The following tabulation gives all combinations of outcomes for the testing of the hypothesis H_0. Two types of error are possible as shown below.

	Hypothesis H_0	
Decision Concerning H_0	True	False
True	No error	Type II error = β error
False	Type I error = α error	No error

The size of the errors can be controlled by altering the sample sizes. A curve which depicts the relationship between the probability of accepting (P_A) the hypothesis H_0 versus the actual parameter value (θ) is shown below. This curve assumes a two-sided test; that is, the hypothesis H_0 is rejected if the estimate of θ differs significantly from θ_0 on either side of θ_0. The α or Type I error is shown at $\theta = \theta_0$ and it is the difference between the probability of acceptance at $\theta = \theta_0$ and unity. Thus

$$\alpha = 1 - P\ (\text{accepting } H_0 | \theta = \theta_0)$$

or

$$\alpha = P\ (\text{rejecting } H_0 | \theta = \theta_0).$$

Example

Hypothesis Testing

Suppose that the life of an item to be tested can be assumed to have the exponential distribution and that 10 items are tested until failure. The observed lifetimes are

2030, 812, 1870, 4650, 272, 1830, 45, 107, 305, and 734.

The hypothesis H_0 to be tested is that the mean life time is less than or equal to 700 hours versus the alternative hypothesis H_a that the mean life time is greater than 700 hours.

The appropriate test procedure is given in Ref. A-7 . The statistic

$$2r\hat{\theta}/\theta = 2T/\theta$$

is distributed as a χ^2 distribution with 2r degrees of freedom, where T is the sum of the life times of the items tested and r is the number of failures, which in this example is the number of items on test. Then,

$$\chi^2_{2r} = 2T/\theta = 2(12655)/700 = 36.2$$

and the probability that this value of χ^2 is exceeded is given in the table in Ref. A-8 to be ≈ .015, so H_a is the preferable hypothesis.

A.8 Least Squares

Many problems involving curve fitting with one independent and one dependent variable are solved by eye-fitting a curve to the observed set of data points. For many purposes this procedure is adequate, but it yields little in the way of quantitative information concerning how well the curve fits the data. If there are two or more independent variables one must use an analytical method for fitting.

In order to answer specific questions concerning the adequacy of a proposed relationship to estimate the dependent variable a model is first hypothesized or assumed. Then it is necessary to obtain the related statistics which describe the goodness-of-fit of the assumed model.

Of primary importance is how one selects the model to fit the data. Ideally the engineer would be able to derive a model, which, with the exception of not knowing certain constants, relates the dependent variables to the pertinent independent variables. However, in many situations it is not possible to derive the model but one has in mind a general behavior of y for given $\underline{x}$ which might be described by a specific function. This function may satisfy certain desired limit properties, but otherwise have no physical basis. In other cases it may not be possible to specify a functional form from a knowledge of the behavior of y; and hence a polynomial in the x's or their reciprocals is fitted to the data with the understanding that it represents an interpolation function and is not valid outside the range of the observations. Such a function may be considered as a Taylor series approximation to the real but unknown function. In any case let the model be denoted by

$$y = \eta(\underline{x}) + \varepsilon$$

where

$\eta(\underline{x})$ is the mean value of y for given $\underline{x} = x_1, \ldots, x_k$

and

ε is the deviation between the observed y and the mean value of y predicted by the model $\eta(\underline{x})$.

Having decided on a particular form of the model the unknown constants (parameters) may be estimated on the basis of one of several criteria. Two analytical procedures are given in Table A.8-1, viz., least squares and Chebyshev, the former being most widely used and hence described more completely in Table A.8-2.

The method of least squares assumes that the ε_i, $i = 1, \ldots, n$, are **independently** distributed with mean zero (0) and a variance σ^2 (independent of $\underline{x}$). If it is desired to test a hypothesis concerning the unknown parameters of the model it is necessary to assume, in addition to the above, information about the distribution of the ε_i. It is usually assumed that the distribution is Gaussian as the theory is well-known for this distribution. It should be emphasized that many computer programs are available to perform the required computations and that what may appear to be a complex analysis can often be done in little time on a modern digital computer. The primary problem to the user then becomes the interpretation of the results of a computer printout for several statistical descriptions are used to assess the goodness-of-fit of the model.

Assume that a given set of data has been fitted to a prescribed model by the method of least squares and that the prediction equation for the mean response (performance) for given values of $\underline{x}$ is

$$\hat{y} = b_0 + b_1 f_1(\underline{x}) + \ldots + b_p f_p(\underline{x}). \tag{A.8-1}$$

If $f_i(\underline{x}) = x_i$ then a simple linear prediction equation results, that is,

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p.$$

Similarly one may consider $f_i(\underline{x}) = x^i$ and thus a polynomial in x,

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + \ldots + b_p x^p.$$

One other example should be sufficient to show the generality of the general linear prediction equation (A.8-1). Let $f_i(\underline{x}) = 1/x_i$, then

$$\hat{y} = b_0 + b_1/x_1 + b_2/x_2 + \ldots + b_p/x_p.$$

All of the above examples are considered to be linear in the coefficients and the coefficients can be obtained by the method of least squares. Some attention will now be given to the interpretation of results of such a model and the associated statistics. First of all it is important to obtain an overall measure of dispersion of y_i, the observed values of y, from their respective predicted mean values $\hat{y}_i$. This dispersion is estimated by

$$s^2 = \frac{\Sigma(y_i - \hat{y}_i)^2}{n - (p+1)}$$

where n is the number of observations on y and p+1 is the number of estimated parameters, and hence n - (p+1) is the number of degrees of freedom associated with the estimate s^2 of σ^2. All other measures of precision are given as multiples of σ^2 and the estimates by the corresponding multiples of s^2. If the standard deviation s, the square root of the variance, is small relative to y_i and to the variation of y_i over the region in which the predictions are being made, then the predicted values are considered to be good. A usual test of goodness is made by comparing the sum of squares of deviations $\Sigma(y_i - \hat{y}_i)^2$ with a measure of the variation of y_i corrected only for the mean value, i.e.,

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \overline{y})^2} = \frac{\Sigma(\hat{y}_i - \overline{y})^2}{\Sigma(y_i - \overline{y})^2} ,$$

where R^2 is the proportion of the total variation (corrected for the mean) which is explained by the prediction equation. A large value of R (near unity) does not necessarily imply that the equation is a good fit. For example, it is possible to increase R by increasing the number of parameters in the model and consequently decreasing $\Sigma(y_i - \hat{y}_i)^2$. The same does not hold true for s^2, as it contains the number of degrees of freedom n-p-1 in the denominator. Hence as p, the number of parameters increases s^2 will eventually start to increase as unimportant terms are added to the prediction equation. Thus s^2 is a more important overall measure of precision of the prediction equation.

Other measures of precision which can be obtained from the fitted equation are those for:

(1) The individual regression coefficients, b_i, $i = 1, \ldots, p$,

(2) The estimated mean $\hat{y}$ of η,

(3) The estimate of an individual observation y_i, and

(4) Combinations of the above.

Ref. A-10 gives the procedures for estimating the above measures of precision and for obtaining related confidence limits. Let us consider here the problem of interpreting the results. The precisions of the individual coefficients $s(b_i)$ can be used to obtain a confidence interval of the form

$$b_i - t[s(b_i)] \leq \beta_i \leq b_i + t[s(b_i)]$$

where t is the value of Student t for the appropriate number of degrees of freedom, n-p-1, and the desired confidence level, see Ref. A-8 for a tabulation of t-values. Such an inference can be made for a particular coefficient. One can also obtain a confidence region for a selected set of all of the coefficients. See Ref. A-10 for the appropriate procedure for this computation.

Another important inference which can be made is that of obtaining a confidence interval for η_i using $\hat{y}_i$, the estimate of η_i, and the estimated standard deviation of $\hat{y}_i$. The width of the interval increases as one moves away from the mean point, that is, as the following distance increases,

$$|(f_1(\underline{x}_i),\ f_2(\underline{x}_i),\ \ldots,\ f_p(\underline{x}_i)) - (\bar{f}_1(\underline{x}),\ \ldots,\ \bar{f}_p(\underline{x}))|.$$

A confidence region for a future observed value is obtained in a similar manner to that for the predicted mean value with the exception that the variance of the single value includes the variance of the predicted mean and the variance of the deviation of an individual observation from its mean.

Extrapolation of results beyong the region of investigation is done with considerable risk. Although the above methods can be used outside the region of observations it is obvious that only prior technical knowledge can serve as a guide concerning the validity of the extrapolation. There are no statistical tests for this and the engineer must exercise due care. Theoretically derived models can be extrapolated and the appropriate measure of precision included in any inferences made on this basis.

Table A.8-1

Analytical Methods for Curve Fitting

1. Least Squares Fit - minimize the sum of squares of the deviations of the observations from their corresponding predicted mean values given by the hypothesized equation (model) of the curve or surface, i.e., compute estimates of the unknown parameters, β_j, $j = 1, \ldots, p$, such that

$$\Sigma(y_i - \eta_i)^2$$

is minimized, where

$\eta_i \;=\; f(x_i;\ \beta_1,\ \beta_2,\ \ldots,\ \beta_p)$,

β_j is the j-th parameter, $j = 1, \ldots, p$, and

y_i is the observed response, $i = 1, \ldots, n$.

2. Chebyshev Fit - minimize the largest absolute deviation between the observations and the corresponding predicted value, i.e., compute the estimates of the parameters such that

$$\max|y_i - \eta_i|$$

is minimized.

Table A.8-2

Method of Least Squares

Linear Models

Suppose that y_i, $i = 1, \ldots, n$, is estimated by means of a function linear in the unknown parameters to be estimated

$$\begin{aligned} y_i &= \beta_0 + \beta_1 f_1(\underline{x}_i) + \beta_2 f_2(\underline{x}_i) + \ldots + \beta_p f_p(\underline{x}_i) + \varepsilon_i \\ &= \eta_i + \varepsilon_i, \; i = 1, \ldots, n. \end{aligned}$$

where

$f_j(\underline{x}_i)$ is a function of the independent variables $x_1, \ldots, x_k$, $(j = 1, \ldots, p)$

$\beta_j (j = 0, \ldots, p)$ are the unknown parameters to be estimated from the data,

ε_i is the deviation between the observation y_i and its mean given by η_i where

$\eta_i = \beta_0 + \beta_1 f_1(\underline{x}_i) + \ldots + \beta_p f_p(\underline{x}_i)$, and

$\underline{x}_i = (x_{1i}, x_{2i}, \ldots, x_{ki})$ the i-th values of each of the independent variables used in the model.

The ε_i are assumed to be independently distributed with mean 0 and constant variance σ^2. If the variance σ^2 is some function $h(\underline{x})$, it is necessary to perform a weighted least squares computation. If $h(\underline{x})$ is known in form only and certain constants of $h(\underline{x})$ need to be estimated the problem becomes one in non-linear least squares. Such problems are treated in the following section. In the linear, non-weighted case the estimates b_i of β_i are obtained by solving the following system of equations given in matrix form:

$$(F'F)B = F'Y,$$

where

$$B = \begin{bmatrix} b_0 \\ b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_p \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

$$F = \begin{bmatrix} 1 & f_1(\underline{x}_1) & \cdots & f_p(\underline{x}_1) \\ & \cdot & & \cdot \\ & \cdot & & \cdot \\ & \cdot & & \cdot \\ 1 & f_1(\underline{x}_n) & \cdots & f_p(\underline{x}_n) \end{bmatrix}$$

and F' is the transpose of the F matrix. The j-th equation corresponding to the matrix form of the equations as given above is

$$\Sigma f_j(\underline{x}_i)\ b_0 + b_1\ \Sigma f_1(\underline{x}_i)\ f_j(\underline{x}_i) + \ldots + b_p\ \Sigma f_p(\underline{x}_i)\ f_j(\underline{x}_i)$$

$$= \Sigma f_j(\underline{x}_i)\ y_i,\ j = 0,\ \ldots,\ p.$$

Many computer programs are available to perform the solution of the above linear equations and obtain the associated measures of precision of the estimates, for example, see Ref. A-11. The least squares prediction is obtained as

$$\hat{y} = b_0 + b_1 f_1 + b_2 f_2 + \ldots + b_p f_p,$$

where f_j is some specified function of the x's, j = 1, 2, ..., p. For example, f_j might be x_j, $1/x_j$, $\ell n\ x_j$, x^j, etc.

The above remarks on inferences assume that the observations have a Gaussian distribution and are independent. These assumptions can be tested by computing the individual deviations $y_i - \hat{y}_i$ and plotting their sample distribution function on Gaussian (Normal) probability graph paper and making a subjective decision concerning the linearity of the transformed distribution function. Sometimes it is desirable to make certain tests for possible correlation of successive test data. Chapter 3 in Ref. A-10 gives several such tests.

Nonlinear Models

A model is nonlinear when the response or performance measurement is a nonlinear function of the parameters to be estimated, as in the following expression,

$$f(x_i, \underline{\theta}) = \theta_1 \exp\{-\theta_2 x_i\} + \theta_3 \exp\{-\theta_4 x_i\}.$$

The system of least squares equations obtained for a nonlinear model are nonlinear in the unknown constants to be determined, and iterative methods of solutions are required. The program NOLLES [Ref. A-11] was written for such problems. Another useful program given in the literature is a SHARE program - SDA-3094 [Ref. A-9].

Measures of precision similar to those for linear models can be made here; however, the results are approximations and the adequacy of the approximations must be considered.

A.9 The Failure-Time Distribution and Hazard-Rate Function

Failure-Time Distribution

Suppose that a large number of items are tested until failure and the life-lengths recorded. These failure-times can be ranked from the smallest value to the largest value and a sample distribution plotted as indicated in Appendix A.4. For convenience the data may be grouped according to appropriate time intervals or percentage points and only a few points plotted. It is usually preferred to plot the failure times corresponding to fixed percentage points, say 5%, 10%, 15%, ..., 100%. Similarly one can plot an estimate of the probability density function by means of a histogram. These techniques are illustrated in the first example to be given below.

In order to describe these results it is convenient to use an appropriate continuous function and fit it to the observed frequency distribution or to estimate the unknown parameters of the function which "best" fit the observations. The sense in which "best" is defined has been treated in many statistical texts and will not be discussed here except for a particular application.

Suppose that one suspects that the data can be described by a negative exponential distribution, then the continuous frequency function is defined by

$$f(t) = \lambda e^{-\lambda t}, \; 0 \leq t < \infty,$$

where t is the time in hours and λ is a constant or parameter which is to be estimated on the basis of the data. It can be shown that the estimator having the desired properties is given by

$$\hat{\lambda} = \frac{n}{\Sigma t} = \frac{1}{\bar{t}},$$

where n is the number of observations and thus $\Sigma t/n$ is the mean life time $\bar{t}$. Very often the frequency function is written as

$$f(t) = \frac{1}{\theta} e^{-t/\theta}, \; 0 \leq t < \infty$$

where θ is the mean time between failures and hence

$$\hat{\theta} = \bar{t},$$

that is, the estimate of θ is the mean life time $\bar{t}$ as one would expect.

A hypothetical set of data, based on random exponential failure times, is summarized in Table A.9-1 below. These data will be used throughout this section for the sample computations.

Table A.9-1

Observed Distribution of Failure Times

Time Intervals (hours)	Failure Frequency	Cumulative Failure Frequency	Survival Frequency
0 - 50	77	77	123
50 - 100	46	123	77
100 - 150	38	161	39
150 - 200	13	174	26
200 - 250	7	181	19
250 - 300	9	190	10
300 - 350	3	193	7
350 - 400	3	196	4
400 - 450	0	196	4
450 - 500	2	198	2
500 →	2	200	0

Hazard Function

Another function of interest in reliability problems is the hazard function which is the ratio of the probability of failure in a small interval $(t, t+\Delta t)$ given that the item has survived time t to the probability that the item survives time t. Actually the hazard function is an instantaneous conditional failure rate. Thus

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{1 - \int_0^t f(t)dt} .$$

If $f(t) = \lambda e^{-\lambda t}$ then

$$F(t) = 1 - e^{-\lambda t};$$

and hence

$$h(t) = \frac{\lambda e^{-\lambda t}}{1 - (1-e^{-\lambda t})} = \lambda,$$

or a constant when the frequency function is the exponential function. This hazard is the probability of failure on the condition that the item has survived. The function is important from the standpoint of checking for the form of the frequency function which one might use.

Example 1

Suppose that a large number of items such as transistors, electron tubes, car tires, or electric light switches, are placed on test under specified conditions until failure. Then one can obtain a sample distribution of times till failure such as that indicated below. The sample distribution function can be plotted as a function of time as follows using the hypothetical set of data given in Table A.9-1. One observes that seventy-seven (77) is the number of failures which occurred at or below 50 hours, hence 77/200 or the proportion 0.385 can be plotted as the ordinate versus the abscissa value of 50 hours. Similarly the relative cumulative frequencies can be plotted corresponding to all times at the upper end points of the respective intervals, (see Figure A.9-1).

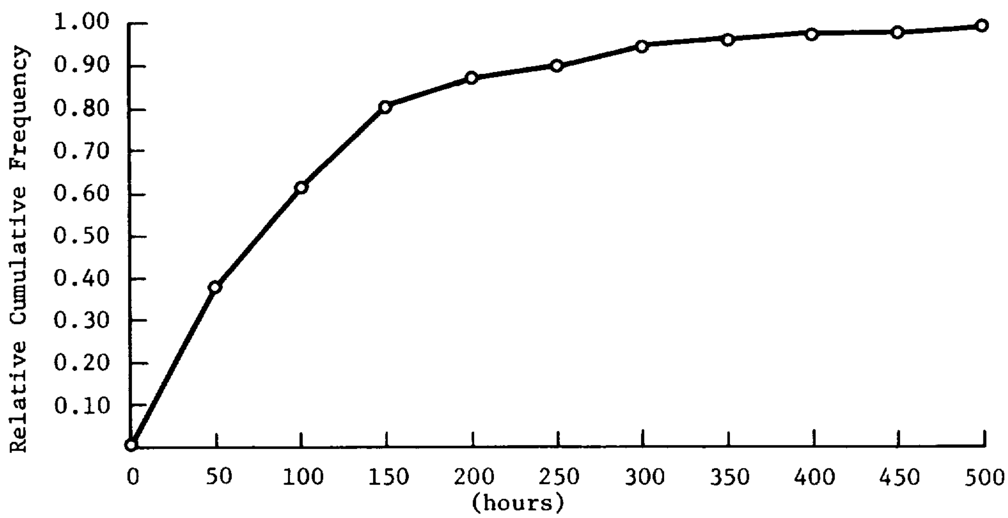


Figure A.9-1 Observed Sample Distribution

The frequencies may also be plotted as a histogram by constructing rectangles of height equal to the frequency of relative frequency, as desired on each interval as shown below.

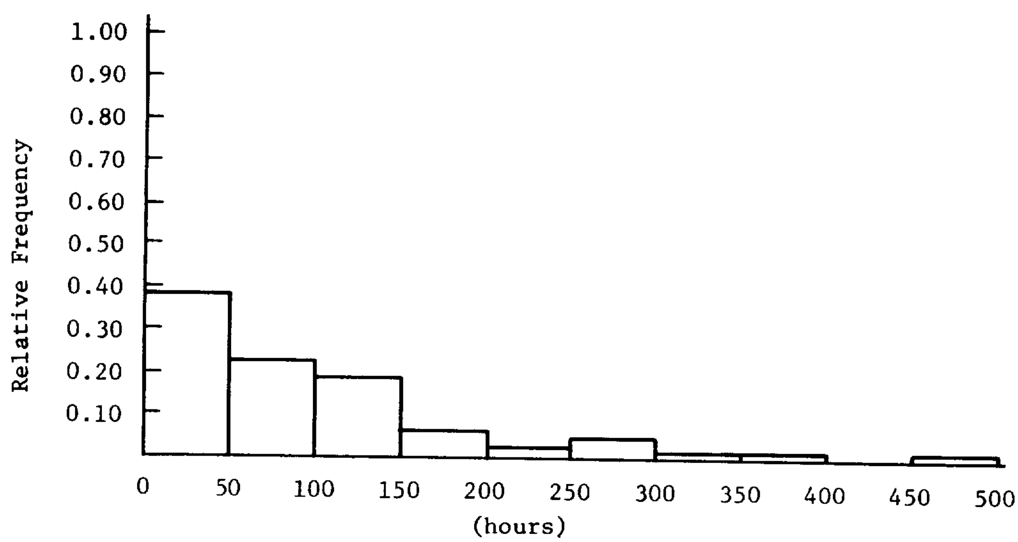


Figure A.9-2 Observed Sample Histogram

Example 2

The hazard rate is estimated for the data in Table A.9-1 for $\Delta t = 50$ hours. The following figure contains a graph of the function.

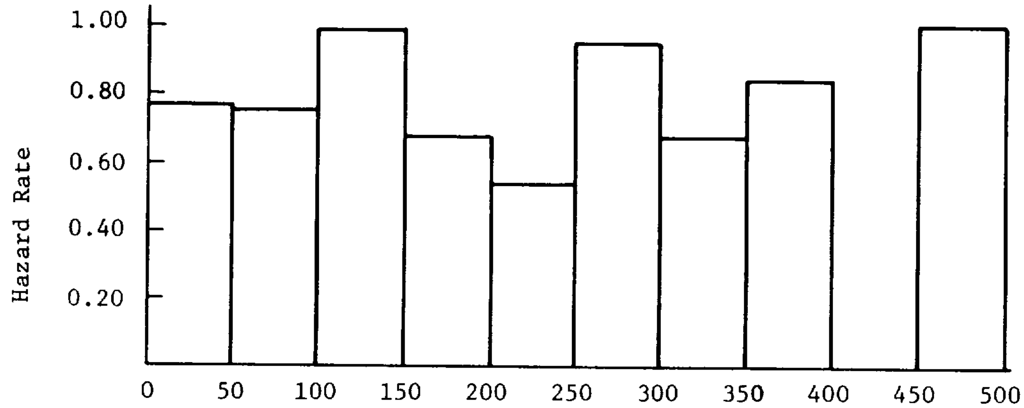


Figure A.9-3 Hazard Function (Δt = 50 hours)

From the figure it can be reasonably inferred that the hazard function is essentially a constant. There are test procedures for testing this hypothesis, see Ref. A-12.

In general, however, many authors have implied that the hazard function is a "bath tub" shaped curve as shown below.

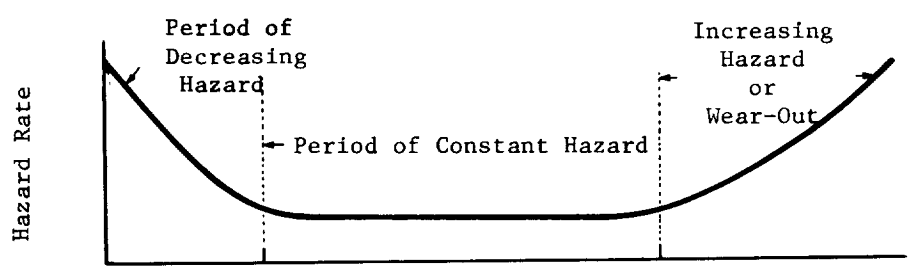


Figure A.9-4 Typical Hazard Rate Curve

The early high hazard rate is a result of manufacturing defects and removal of these from the collection of items results in a more uniform collection of items which have essentially a constant failure rate and then finally all items begin to reach a wear-out stage near the end-of-life. There is a strong parallel between the above curve and the instant mortality curve for human beings.

Much literature has evolved around the above curve or the subject area of increasing failure rate (IFR) and decreasing failure rate (DFR). For example, burn-in is often used in screening components for use in missile systems because of the early high hazard rates relative to those in the period of constant values. The question is to decide how long to test, assuming that sufficient information is available to infer the decreasing failure rate.

Random Failure Law

One of the most common misconceptions appearing in the reliability literature is the implication that the random failure law and the exponential failure law are one and the same. Assuming that the random failure law simply assumes that failures occur randomly over time then there can be several failure laws depending upon whether the log-normal, the Weibull, the gamma or some other distribution is assumed to best describe the distribution of failures. The difficulty is that random failure and the Poisson law are considered equivalent by many authors and thus the exponential law with constant hazard rate is assumed the only random failure law.

Table A.9-2

Hazard - Rate Function and Failure-Time Distribution

1. Let t be the time to failure (life length or survival time)
2. Assume t has the probability density function f(t) and the distribution function F(t), where

$$F(t) = \int_0^t f(x)dx$$

or

$$f(t) = \frac{d}{dt} F(t).$$

and zero (0) is assumed to be the threshold or location parameter.
3. The reliability R(t) is given by 1 - F(t).
4. The hazard-rate h(t) is the instantaneous failure rate and is given by

$$h(t) = \lim_{\Delta t \to 0} \frac{F(t+\Delta t) - F(t)}{\Delta t} \frac{1}{R(t)} = \frac{f(t)}{R(t)} .$$

It can also be expressed as follows:

$$h(t) = \frac{-R'(t)}{R(t)} = -\frac{d\ell n\, R(t)}{dt} ,$$

where R'(t) is dR/dt.
5. The distribution of life-length can be expressed as

$$F(t) = 1 - \exp\{ - \int_0^t h(x)dx\}.$$

Example 3

Hazard-rate Function and Failure Time Distributions

Suppose that the time to failure has the gamma density with shape parameter n and scale parameter θ, i.e.

$$f(t) = t^{n-1}\exp\{-t/\theta\} / \Gamma(n)\theta^n, \text{ for } t > 0.$$

The distribution function F(t) is the following integral

$$F(t) = \int_0^t u^{n-1} \exp\{-u/\theta\}du / \Gamma(n)\theta^n$$

and

$$R(t) = 1 - F(t) = \int_t^\infty u^{n-1} \exp\{-u/\theta\}du / \Gamma(n)\theta^n.$$

The hazard rate is

$$h(t) = \frac{f(t)}{R(t)} = \frac{t^{n-1} \exp\{-t/\theta\} / \Gamma(n)\theta^n}{\int_t^\infty u^{n-1} \exp\{-u/\theta\}du / \Gamma(n)\theta^n}$$

$$= \frac{t^{n-1} \exp\{-t/\theta\}}{\int_t^\infty u^{n-1} \exp\{-u/\theta\}du} .$$

The curves for the hazard function are given in the figure below for n = 1 (exponential case), n = 1/2 (decreasing hazard-rate function), and n = 2 (increasing hazard-rate function). For n = 1 the hazard rate is a constant and equal to $\lambda = 1/\theta$.

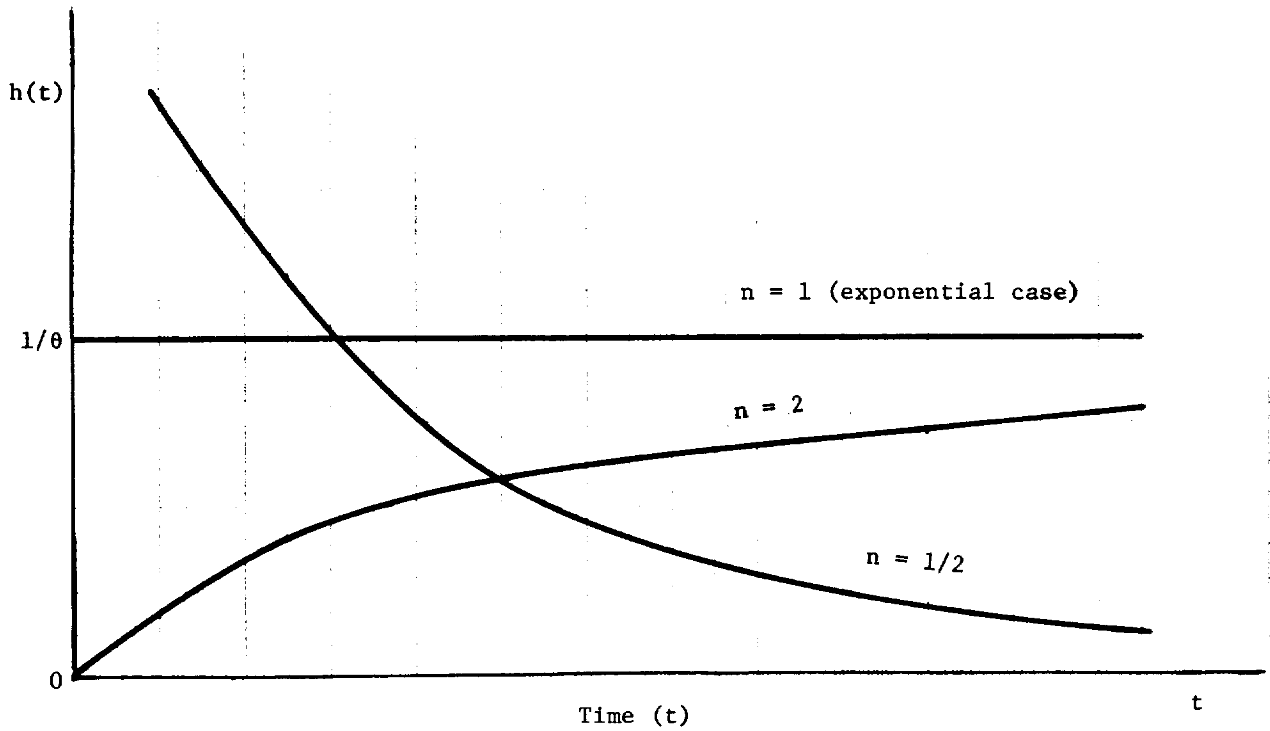


Figure A.9-5 Typical Hazard-Rate Functions for Gamma Density Function

A.10 **Sampling** and Sampling Plans

Sampling and sampling plans is used in this section primarily in connection with acceptance sampling. A long list of definitions is given in Table A.10-1 with this point of view in mind. The basic purpose behind acceptance sampling is to provide a check of incoming material or outgoing material (depending upon whether you are the consumer or supplier). A sample of the material is selected to be evaluated and depending on the results of this sample the entire lot of incoming material is either rejected, accepted, returned for rework, etc. This sampling process implies that the material is being supplied in a sufficiently large amount that a sample can be selected at random and evaluated. If the items of the sample are destroyed in the measurement process it is clear that the cost per item must not be excessive relative to the value of the information to be obtained by the sampling process. If the measurement process is non-destructive then the sample can, in the limit, include all of the items in the lot. This would be more appropriately called screening rather than acceptance sampling. Very often a combination of screening and sampling (less than 100% of the items in the lot) are used jointly in evaluating the incoming material where the measurement process is expensive for some characteristics of interest and cheap for other characteristics of interest.

There are a very large number of sampling plans depending upon the assumptions one makes concerning the distribution of the measurement. There are "distribution free" plans which assume little or nothing concerning the distribution and parametric plans in which the distribution is assumed and its parameters or characteristics are hypothesized. Table A.10-2 contains a partial summary of many plans tabulated by distribution form assumed, by whether the measurement is an attributes (discrete) or variables (continuous) measurement. Specific features of the various plans are also tabulated along with the reference.

It should be pointed out that sampling is also used in many other applications. For example, in quality control the manufacturer attempts to control or check-on a production process to detect departures from an acceptable process. He may select a sample every hour, shift, day or appropriate period of time, measure the pertinent characteristics, plot these on a chart or graph to detect any shifts in process level of defectives, mean level, dispersion level, etc.

Similarly the word sampling is used in connection with experimental designs. One example is that of nested sampling in which one is estimating the sources of variation of a process. If a process may be subdivided into stages it is often possible to estimate the variation at each stage by means of standard analysis of variance techniques. At each stage a sampling of the items is made and the desired measurements made on each of the items.

These last two types of sampling are not treated in this section of the appendix but are mentioned to give a better overall perspective of the problem.

Example

Suppose that it has been conjectured that the proportion defective of a certain production item is less than or equal to 0.02. A sample of ten items is selected from the production and one item is found to be defective. What is the probability of observing one or more defectives if the conjecture is correct? Would the observed result be in contradiction to the hypothesis that the proportion defective is 0.02?

$$\begin{aligned} P(x \geq 1) &= 1 - P(0 \text{ defectives}) \\ &= 1 - \binom{10}{0}(.02)^0 \ (0.98)^{10} \\ &= 1 - .81707 \approx 0.18. \end{aligned}$$

Thus the probability of occurrence of a result as bad or worse than that observed in the sample is not rare and one would not reject the hypothesis as stated on the basis of this evidence.

Table A.10-1

Acceptance Sampling Plans

Definitions:

1. Random Sample - a sample of n items from a population of N items is said to be random if each sample of n items has an equal chance of being selected. Thus a sample of every 100-th item from a production of 10,000 with the first item being selected at random from the first 100 items is not a random sample for such a sample cannot contain more than one item from the first 100 items or any subsequent group of 100 items.
2. α - Producer's Risk - the probability that a sampling plan will reject material of acceptable quality. For example, if a mean failure time θ of 3,000 hours is specified to be an acceptable value, then α is the probability that a particular material having this mean failure time will be rejected by the sampling plan.
3. β - Consumer's Risk - the probability that a sampling procedure will accept material of unacceptable quality. Conventionally $\beta = 0.10$.
4. AQL - Acceptable Quality Level - the quality of the material that a sampling plan is designed to accept with probability $1-\alpha$ or reject with probability α.
5. LTPD - Lot Tolerance Percent Defective - the lot quality that a sampling plan is designed to reject with probability $1-\beta$.
6. OC - Operating Characteristic - a graph of the probability of acceptance versus the quality level of the material being sampled.

7. Sampling by attributes means that the items are classified as failures (defects) or non-failures (non-defectives), and thus decisions are based only on the number of failures among the items tested.

8. Sampling by variables means that the quality of the items is measured by a continuous variable such as the time to failure, current gain of an amplifier, breaking strength of an item under test, etc.

9. The phrase with replacement implies that the items sampled are returned to the sample before the next item is selected; or if a destructive test, with replacement implies that the failed item is replaced.

10. Single Sampling Plans - a single sample of specified size is selected from the lot for testing, and the decision on the lot acceptance is based on the results of the single sample.

11. Multiple Sampling Plans - a specified number of samples (usually of different sample sizes) are selected one-at-a-time at random for testing, and the decision procedure is such that at the completion of testing of the i-th sample, one can either reject the lot, take another sample, or accept the lot. At the completion of testing of the last sample the decision is either to accept or reject the lot.

12. Sequential Sampling - items are selected one by one (or may be selected in groups of r items each) and the decision is made at the end of the test on the i-th item to either reject or accept the lot of material being tested or take another item for testing. These procedures may be either truncated or not truncated.

Table A.10-1

Tabulation of Acceptance Sampling Plans

Distribution	Sampling Procedure	Type Measurement	Index, etc.	Features	Reference*
Binomial	Single-Sampling	Attributes	Lot sizes 21 to 100,000 LTPD's 1/2% to 10% Process Av. 0 to 1/2 (LTPD)	Plans minimize totals amount of inspection at the process average.	Dodge, H.F. and Romig, H.G., 1959, <u>Sampling Inspection Tables</u>.
	Double-Sampling		Lot size 51 to 100,000 Other parameters as above.		
	Single-Sampling		Lot size 4 to 100,000 AOQL - .1% to 10% Process Av. 0 to (AOQL)		
	Double-Sampling		Lot sizes 16 to 100,000 Other parameters as in Ap. 6.		
Binomial	Single-Sampling	Attributes	Lot sizes 2 to 550,001 AQL - 0.015% to 10%, or AQL (Defects per 110 units)- 15 to 1,000 Sample sizes - 1 to 1,500	Precepts for normal, reduced, and tightened inspection are given.	<u>Military Standard</u> 105B.
	Double Sampling Multiple-Sampling		Plans chosen to match the OC curves of the single sample plans.		
	Double-Multiple		Chosen to match the OC curve of the single-sampling plan.		Freeman, H.A., Friedman, M., Mosteller, F. and Wallis, W.A., 1948, <u>Sampling Inspection</u>
Binomial	Sequential	Attributes	AQL 0.02% to 20% LTPD 0.2% to 35% $\alpha = 0.05$, $\beta = 0.10$ and 0.50.	None	Columbia Univ., 1945 <u>Sequential Analysis</u>

* References are given here rather than at the end of the section.

Table A.10-1 (Continued)

<table>
<tr><th>Distribution</th><th>Sampling Procedure</th><th>Type Measurement</th><th>Index, etc.</th><th>Features</th><th>Reference</th></tr>
<tr><td>Binomial</td><td>Single-Sample</td><td>Attributes</td><td>β = 0.05, 0.01
Process Average<table>
<tr><th>LTPD</th><th>1%</th><th>2%</th><th>3%</th><th>4%</th></tr>
<tr><td>3</td><td>⊗</td><td></td><td></td><td></td></tr>
<tr><td>4</td><td>⊗</td><td></td><td></td><td></td></tr>
<tr><td>5</td><td>⊗</td><td>⊗</td><td></td><td></td></tr>
<tr><td>6</td><td></td><td>O</td><td>x</td><td></td></tr>
<tr><td>7</td><td></td><td></td><td>x</td><td></td></tr>
<tr><td>9</td><td></td><td></td><td></td><td>x</td></tr>
</table>x- Non-destructive testing N = 10,000 and 20,000
0- Destructive testing N = 10,000 and 20,000
β = 0.05 and 0.01 with the exception of process av. of 3 and 4%.</td><td>For fixed LTPD, β, and process average minimizes the average cost of producing an item.</td><td>Horsnell, G., 1956, <u>Journal of the Royal Statistical Society</u>, Series A 120, 148-201.</td></tr>
<tr><td>Poisson Approximation to binomial.</td><td>Sequential (truncated and untruncated)</td><td>Attributes</td><td>Lot tolerance defect no.
5(5) 30(10) 60(20) 100
Consumer's risk 0.01, 0.10
Average sample number and AOQL as a function of above.</td><td>Defective articles are removed or replaced by good ones.</td><td>Anscombe, F.J., 1949 <u>Journal of the Royal Statistical Society</u>, Series A 112, 180-206.</td></tr>
<tr><td>Poisson</td><td>Single-Sample</td><td>Attributes</td><td></td><td></td><td>Greb, D.J., 1949, <u>Journal of the Americal Statistical Association</u>, 44, 62-76.</td></tr>
<tr><td>Poisson Approximation to binomial.</td><td>Single-sample</td><td>Attributes</td><td>α = 0.05
β = 0.01, 0.05, 0.1
Acceptance number C = 0(1)49</td><td></td><td>Cameron, J.M., 1952, <u>Industrial Quality Control</u>, 9, No. 1, 37-39.</td></tr>
<tr><td>Poisson</td><td>Single and Double Sampling</td><td>Attributes</td><td>$\left\{\begin{matrix}\alpha = 0.1, 0.1, 0.01 \\ \beta = 0.1, 0.01, 0.01\end{matrix}\right\}$</td><td></td><td>Kitagawa, T., 1952, <u>Tables of Poisson Distribution</u>.</td></tr>
</table>

Table A.10-1 (Continued)

Distribution	Sampling Procedure	Type Measurement	Index, etc.	Features	Reference
Binomial	Single-Sampling	Attributes	Sample sizes $n = 5(5)40$ AQL = 1% (1%) 20% LTPD = 2% (1%) 40%	Acceptance numbers are given that minimize $\alpha + \beta$.	Golub, A., 1953, Journal of the American Statistical Association, 48, 278-288.
Poisson Approximation to binomial.	Single Double Truncated Sequential	Attributes	$\alpha = \beta = 0.05$ $R = \frac{LTPD}{AQL} = 1.3$ to 1.5	Gives characteristics of the plans as a function of R	Peach, P., 1947, Industrial Statistics and Quality Control
Binomial	Double Triple	Attributes	Lot size 101 to 100,000 AQL = 1%, 1 1/2%, 2 1/2%	First stage uses narrow gauge limits	Harding, H.G., and Price, S., 1957, IRE National Convention Record, 1957, Part 10, 54-58
Binomial	Multiple	Attributes	AQL = 2 1/2% Lot size = 65 to 8,001		Jacobs, R.M., 1957, American Machinist, 101, No. 3, 113-118
Normal Known standard deviation σ	Single-Sampling	Variables	AQL = .024% to 11% Lot size = 25 to 550,000	Plans are chosen to match OC curves given in Freeman [Ref. A-25].	Grubbs, F.E., 1949, Annals of Mathematical Statistics, 20, 242-256.
Unknown standard deviation σ	Double Single Double				
Normal σ known σ unknown	Single	Variables	Plans entered by lot size and inspection level.	Precepts for inspection levels.	MIL-STD 414, Dept. of Defense, 1957.
Normal σ unknown ($s \simeq \sigma$) (sample range $\simeq \sigma$)	Single	Variables	Lot size = 40 to 110,001 Sample size = 5 to 200 AQL from 0.065 to 6.5%	Provision is made for 1 and 2-sided specification limits.	Storer, R.L., and Davidson, W.R., 1955, Industrial Quality Control, 12, No. 1, 15-18.

Distribution	Sampling Procedure	Type Measurement	Index, etc.	Features	Reference
Normal σ unknown ($s \doteq \sigma$)	Single-Sample	Variables	$\alpha = 0.05$, $\beta = 0.10$ AQL = 0.1% to 5% LTPD = 0.15% to 40%	One-sided specification	Eisenhart, C., Hastay, M.W. and Wallis, W. A., 1947, Techniques of Statistical Analysis.
Normal	Single	Variables	Sample size 3 to 200 AQL = 4% to 15%	Implements a graphical method based on Normal probability paper.	Chernoff, H., and Lieberman, G. J., 1957, Industrial Quality Control, 13, No. 7, 5-7.
Normal ($s \doteq \sigma$)	Single	Variables	Tables give 95%, 50%, and 10% points on the OC curve. Accept rule: $\bar{x} \geq L+ks$ where $k = .6(.2)\ 2(.25)2.5,3$ $n = 5, 7, 10(5)40, 50$ = sample size.	One-sided specification	Grant, E.L., 1952, Statistical Quality Control.
Normal ($s \doteq \sigma$)	Single	Variables	$\alpha = 0.01, 0.05, 0.10$ $\beta = 0.01, 0.05, 0.10$ $n = 3(1)31$	For controlling variability (variance)	McElrath, G.W., and Bearman, J.E., 1957 American Society for Quality Control National Convention Transactions, 11, 447-462.
(range $\doteq \sigma$)			$\alpha = 0.01, 0.05$ $\beta = 0.01, 0.05$ $n = 2(1)12$		
Normal	Single	Variables	$\alpha = \beta = 0.1, 0.05, 0.02, 0.01$ $n = 2(1)31, 40(10)100$	For rejecting an excessively varying product. Reject when $s^2 > k$	Burr, I.W., 1953, Engineering Statistics and Quality Control.

Table A.10-1 (Continued)

Distribution	Sampling Procedure	Type Measurement	Index, etc.	Features	References
Binomial	Continuous sampling plan (CSP). See [Ref. A-26] for definition.	Attributes	Sampling fraction (1/2% to 8%) AOQL = 0.12% to 10.7% AQL = 0.015% to 10% Number of good items needed to restore sampling after a defective = 12 to 3,200		Storer, R.L., 1956 Industrial Quality Control, 12, No. 11 48-54.
Binomial	CSP	Attributes	Classified by production rate and AQL. Production rate = 2 to 110,001 items per day. AQL = 0.015% to 10%.		Handbook 107, Dept. of Defense, 1959b.
Increasing or Decreasing Failure Rate Distribution-Free Life Testing Plan based on Binomial.	Single	Attributes	See Life Testing Summary Table 8.1		Barlow, R.E., and Gupta, S. S., 1966, "Distribution-Free Life Test Sampling Plans", Technometrics, 8, No. 4.
Binomial	Single	Attributes	Three systems are studied: a) LTPD system with $\beta = 0.10$ b) AQL system with $\alpha = 0.05$ c) IQL with the probability of acceptance = 1/2. Extensive tables are given for each system.	Minimize average costs for given point on OC curve. Inspection, acceptance, and rejection costs are assumed in p.	Hald, A., 1964, Single Sampling Inspection Plans With Specified Acceptance Probability and Minimum Average Costs, Institute of Math. Statist, Univ. of Copenhagen.
Binomial, Poisson, and Hypergeometric	Single	Attributes	n = sample size c = acceptance number	Determine (n,c) or that the probability of acceptance P_A at AQL is $P_A \geq 1 - \alpha$ and at LTPD is $P_A \leq \beta$, $(1-\alpha > \beta)$ and c is as small as possible.	Hald, A., 1966, "The Determination of Single Sampling Attribute Plans with Given Producer's and Consumer's risk", Institute of Math. Statist., Univ. of Copenhagen.

Table A.10-1 (Continued)

Distribution	Sampling Procedure	Type Measurement	Index, etc.	Features	References
Exponential		Variables (Cumulative test time and number of failures)	Probability of Qualification versus true failure rates for confidence levels 5%, ..., 99%	Establishing and maintaining FR levels at selected confidence levels. Qualification plans.	Military Standard, "Failure Rate Sampling Plans and Procedures," 1965, MIL-STD-690A.
Exponential	Sequential (PRST) Fixed Length Tests.				Military Standard, "Reliability Tests Exponential Distribution," 1965, MIL-STD-781A.

A.11 Statistical Design of Experiments

Experiments are planned to estimate or to test a particular hypothesis concerning unknown constants or parameters of a model. For example, sampling plans are selected to accept good material with high probability and reject poor quality material with high probability.

The primary statistical problems here are the selection of the number of items to test and test plan considerations such as how to terminate the test (failure-terminated or time-terminated) and whether failed items are replaced. On the other hand, an environmental test may be planned to estimate the effects on the item of the environment to be expected in application. For example, it may be desired to estimate the effect of temperature on one or more performance measures. In such cases the considerations are what temperatures to use and how many items to test at each temperature. If several environmental factors are considered simultaneously, then some specific tools in the statistical design of experiments are useful in the selection of the test levels and the number of items to be tested at each level.

In order to evaluate or compare experimental plans it is necessary to have the form of the model in mind; the model may be linear, sum of two exponentials, second degree polynomial, etc. Knowing this model form, the relative precisions of the estimates of the unknown constants or of the predicted mean performance can be determined before the actual running of the experiment.

The precisions of the estimates are given by the product of the unknown variance of the performance measures with respect to the model, which describes their relationship to the independent variables, and a constant dependent on the test plan, the relative precision. Hence in order to compare the experimental plans it is usually assumed that the variances are constant, independent of the test plan and of the values of the independent variables (environmental test levels). In reality the variance would vary from one plan to another and in some cases would depend on the environmental levels, but in practice these variations would not alter appreciably the test plan selected as "best" on the basis of the relative precisions of the estimates.

The following table gives a brief description of some important points to consider in statistical design of experiments. It would be hopeless to treat each of the various types of designs or experiments considered in the statistical literature. However, a single reference on this subject is Ref. A-13 which contains a bibliography of 110 more recent papers on the subject. Several types of designs and/or experiments are listed for familiarizing the reader with the "jargon" if he has not already been exposed to same.

Following the table a discussion on the special topic of matrix testing is given as an example of some of the considerations to be made by an engineer in a particular type of experiment which is used in reliability testing.

Table A.11

Statistical Design of Experiments

Objectives:

To estimate unknown constants or parameters of a hypothesized model form,

To test particular hypotheses concerning the parameters or the model form.

Some Criteria for Selection of Best Design:

Precisions of the estimates of the constants,

Precision of the predicted mean performance based on the model,

Quality of material being accepted on the basis of an acceptance sampling plan and OC curve, and

Ability to discriminate between (or among) hypothesized model forms.

Factors Considered in the Selection:

Number of items to test,

Methods for truncating the test,

Whether or not to replace failed items,

Number of levels of the environments to select and how many tests to be made at each level,

Hypothesized model for relating performance measurements to pertinent independent variables,

Which hypothesis is to be tested, and which alternatives are appropriate,

Magnitude of Type I and Type II errors,

OC curve,

Average sample number (ASN) and maximum number of items which may be tested in the case of a sequential experiment, and/or

The difference in performance measures which is of practical significance.

See Ref. A-13 for a list of references on the following.

Types of Experiments:

Response surface design,

Random balance experiments,

Factorial experiments,

Latin Square design,

Screening designs,

Fractional factorial experiments,

Evolutionary operation (EVOP) designs,

Sequential designs, and

Incomplete block designs.

Matrix Testing (Environmental testing with two or more factors)

Suppose that an item is being designed to operate in an environment in which the temperature may range from 25°F to 125°F, the input voltage is 28 volts dc±10 percent, and the mission time is 200 hours. If one suspects that the item is sensitive to changes in temperature and/or input voltage in that certain performance characteristics change or that the failure rates change, then it is necessary to assess the effects of these environmental changes and determine the resulting implications with respect to the item design.

Often one is concerned about more than one output characteristic or response; however, for convenience in the discussion which follows we will treat only one response variable, say y. In addition, the environmental factors will be called inputs and their respective levels will be denoted by x_1, x_2, ..., x_p, assuming that there are p such inputs, i.e., independent variables, under consideration. Most often one is interested in the behavior of y for various combinations of levels of the inputs, e.g., one set of such combinations is high temperature, low input voltage, and various times of operation under these conditions.

A test procedure may be planned in which several items are selected from the available collection of items and placed on test at several levels of the input conditions and observed for the behavior of the response y. The questions remain of how to select the number of levels to use for each independent variable and the number of observations to be made at each level. A test of the type considered here is referred to as a matrix test, that is, a test procedure which treats simultaneously several input and/or environmental factors, each at two or more levels. For example, the following figure illustrates a possible selection of testing conditions as indicated by x's.

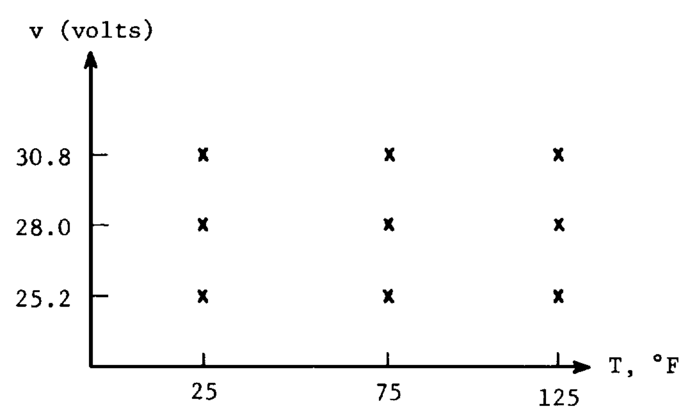


Figure A.11-1 A Selection of Test Conditions

In this example three levels of temperature (T) and three levels of voltage (V) are selected as conditions under which the tests are to be run. Depending on prior information and the objective of the test or experiment, one may wish to select either two, three, or four levels. Very seldom are more than four levels needed. The use of two levels is often adequate for a monotonic function, i.e., one which is non-decreasing (See figure below) or non-increasing.

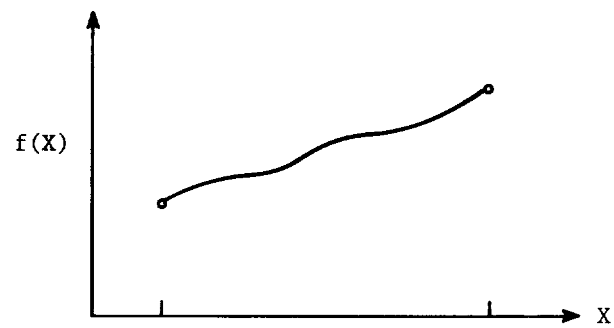


Figure A.11-2 Typical Example of Non-Decreasing Function

Mathematically a non-decreasing (non-increasing) function f(x) is expressed as one for which $x_2 \geq x_1$ implies $f(x_2) \geq f(x_1)$, $(f(x_2) \leq f(x_1))$. If in a matrix test the response increases and then decreases over the interval of interest, it is necessary to select a minimum of three levels of the particular variable. Such a situation can occur in testing an item designed for optimum operation at a nominal level, in which case the performance often degrades with a deviation of an input variable in either direction from the nominal value. In many situations one does not know the best combination of levels to use in the design, and the test conditions are to aid in identifying the "optimum combination" within a reasonable order of precision. By optimum we may mean either that the characteristic of interest will fall within given bounds or that it will exceed a given value.

A part of the planning of the test is the consideration of what type of an analysis is to be made. For example, one may wish to observe or identify the worst case condition or to estimate the optimum condition. To accomplish what is desired may or may not require a formal mathematical model to relate the response of interest to the independent variables being considered. The model may be a simple polynomial approximation to some real world model which is unknown to the experimenter, or it may be a theoretical or analytical model derived by means of physical laws and reasonable approximations resulting from plausible assumptions. The use of a model such as the latter facilitates the estimation of worst case values or the estimation of the location of the optimum combination of the variables. For example, contours of constant response may be approximated on the basis of a fitted model with the use of an appropriate computer program. The following figure shows a situation in which an optimum is a high value of response and the worst case is a low value.

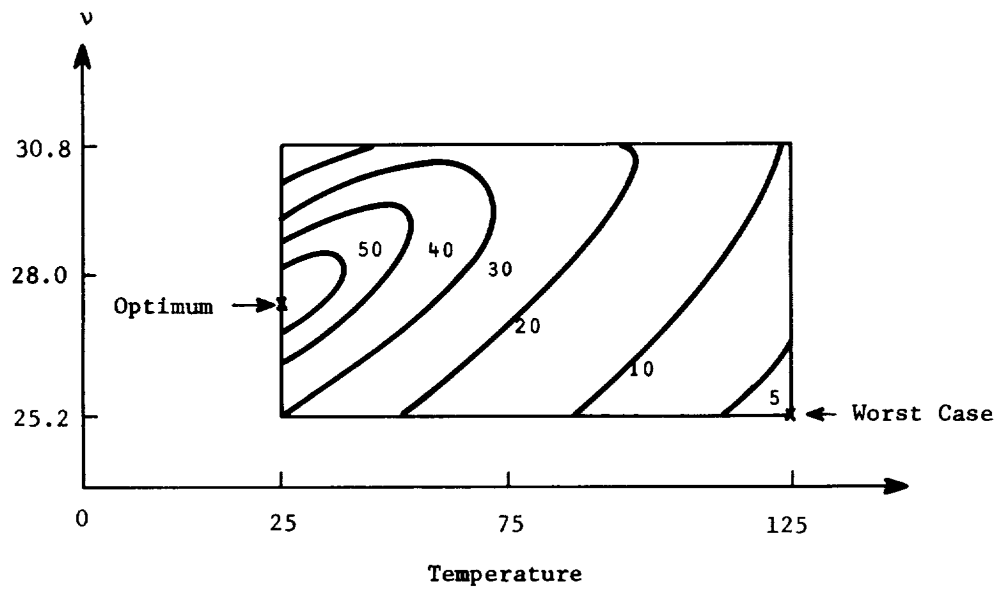


Figure A.11-3 Illustration of Use of Contours

Many approaches are possible with respect to selecting the best allocation of the experimental or test runs to obtain the desired information for an analysis. Possible approaches are given in papers with titles pertaining to: optimal designs for regression, response surface methods, factorial design of experiments, fractional factorial designs, analysis of variance, etc. For example, see Ref. A-14 through A-17.

Interaction

One of the primary reasons for considering a test procedure as described above is to answer questions concerning the effect of the simultaneous exposure to changes is two or more input or environmental variables. For example, it may be that the effect of one environment on the response is altered by the level of another variable. The diagrams below give examples of interacting and non-interacting variables. The interaction is equivalent to non-additivity of variables; in mathematical form the model in the case of interaction contains effects which might be described by products of two interacting variables. In general the interaction model is described by some function of the variables which cannot be reduced to an additive form involving the same variables. It may be possible to transform the variables to obtain an additive model in the new variables.

The presence or absence of interactions is usually determined by fitting an approximating model of polynomial form, or by an analysis of variance which breaks down the total variation of the response into that due to each of the variables and to the combined effect of the variables taken two-at-a-time, three-at-a-time, etc. See, for example, Ref. A-18 through A-20.

An example of when interactions may be important is that of life testing at each of several stress combinations. That is, suppose that an item is subjected to several stresses simultaneously and one wishes to know the mean life as a function of the stress levels. This corresponds to the problem of testing the effects of drugs in protecting an animal from death due to a certain type of bacterium or virus. A great deal of literature has evolved over the years on this problem and is of value in the estimation of mean life or the mean failure rate vs. a function of stresses such as thermal, electrical, and radiation.

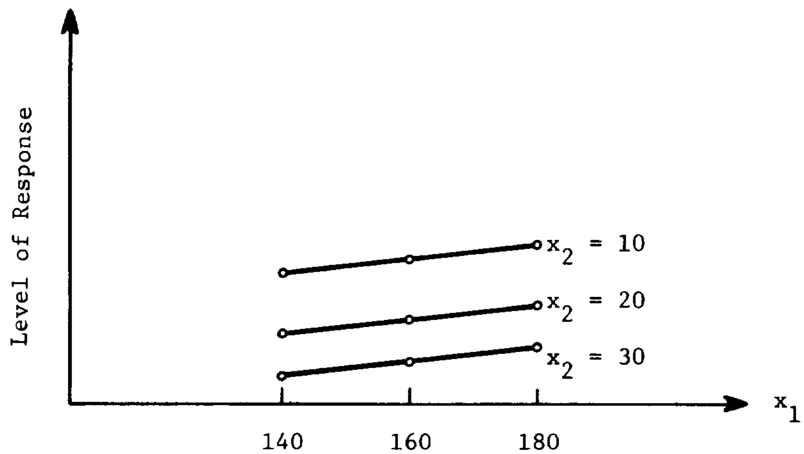


Figure A.11-4 Example of Non-Interacting Variables

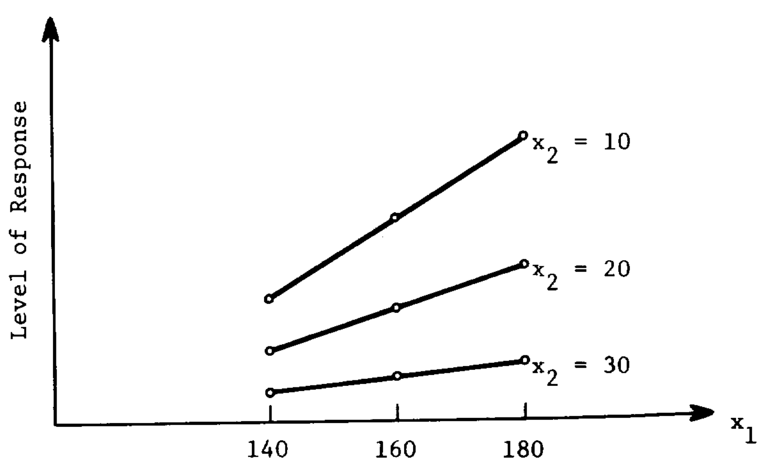


Figure A.11-5 Example of Interacting Variables

A.12 Combination of A Priori Information and Test Results

Experience with similar equipment and subjective engineering information is valuable in the analysis of proposed new equipment designs. Two ways by which this can be accomplished are described in this appendix.

The first approach is to use a model for reliability growth as the equipment design evolves from early models to advanced designs. For example, it may be assumed that the equipment reliability is at least as good now as that of all previous designs. Another approach is to assume that reliability increases according to a given functional relationship between reliability and the number of designs or number of equipments that have been produced. See Ref. A-21 for a discussion of such techniques.

Another approach is to use Bayesian decision models which use past experience to postulate prior distributions of the parameters under consideration. For example, the true failure rate may be assumed to have a probability density function, $p_0(\lambda)$, with a mean given by that observed for similar equipment. There is also an empirical Bayesian technique which uses the prior information to estimate the density function directly with observed relative frequencies and without assuming an a priori density function. See Ref. A-22 for a discussion of this procedure. The empirical Bayesian technique is not discussed in this section as its use requires large samples.

In order to compare the techniques of using prior experience with standard techniques which use no prior information, a simple example will be employed.

<u>Example 1</u>

Suppose that ten (10) equipments have been constructed and tested for T_0 hours and that no failures have occurred. Furthermore, assume that at several stages in the design cycle 20 similar equipments have been tested under the appropriate environmental conditions and that one (1) item failed. What is the reliability of the equipment?

<u>First Solution</u>: Use only the most recent test results on the equipment to be used.

The estimated relative frequency of success is 1 and a 95% lower confidence interval limit is 0.741. This lower limit $\underline{\theta}$ can be obtained by using the formula given in Ref. A-18, (page 698).

$$\underline{\theta} = \frac{x_0}{x_0 + (n-x_0+1)\,v^2_{P_2}} = 0.741$$

where

$$f_1 = 2(n-x_0+1),$$

$$f_2 = 2x_0,$$

$$x_0 = \text{number of successes observed,}$$

n_2 = number of trials made, and

v_{P_2} = the tabulated value of the variance ratio for which the probability is P_2 of not exceeding, for f_1 and f_2 degrees of freedom.

Second Solution: Use the reliability growth technique which assumes that the reliability at the last stage is no worse than it was at any previous stage.

In this case all 30 items can be treated as though they were from the same batch of items and the resulting conservative confidence interval estimate is given by the same procedure as above (1st solution) with one (1) failure and n = 30 items tested. Hence the lower limit is given by 0.850. This limit is conservative (under the assumption) in the sense that the confidence is at least as large as 95%.

Third Method of Solution: Using Bayesian method.
In this case assume that the prior density function is given by the beta function,

$$P_0\{R\} = \frac{1}{B(i,j)} R^{i-1}(1-R)^{j-1},$$

where i and j are positive integers and may be chosen to be consistent with the prior information. From previous tests it is known that the estimated reliability is

$$R = 19/20 = 0.95.$$

The above distribution has a mean

$$\mu\{R\} = \int_0^1 R \frac{1}{B(i,j)} R^{i-1}(1-R)^{j-1} dR$$

$$= i/(j+i) = 0.95 \text{ say},$$

where

$$B(i,j) = \frac{\Gamma(i)\Gamma(j)}{\Gamma(i+j)} = \frac{(i-1)!(j-1)!}{(i+j-1)!}.$$

Assume i = 19, j = 1 then the prior density function is

$$P\{R\} = \frac{1}{B(19,1)} R^{18}(1-R)^0.$$

The a posterior density function of R given r observed successes in n trials is given by

$$P\{R|r\} = P\{R\}\ p\{r|R\} \ / \int P\{R\}\ P\{r|R\} dR$$

$$= \frac{R^{18+r}(1-R)^{n-r}}{B(18+r+1,\ n-r+1)}.$$

The mean of the a posteriori distribution is

$$\hat{\mu}_B = \frac{r + 19}{n + 20},$$

which is the Bayes estimate of the reliability. Now in the example r = 10, n = 10, and hence

$$\hat{\mu}_B = \frac{29}{30} = 0.9667.$$

A lower 95% confidence interval estimate of the reliability can be obtained using the Bayesian technique given in Ref. A-23 and it is 0.902.

The results of the three solutions indicate that reliability growth and Bayesian approaches yield shorter confidence interval estimates as a result of having assumed more information. But it is necessary to assume prior information or some other relationship among the reliabilities at the various stages. However, the previous test experience should be used to the extent that it is reasonable. For better use of prior information it would be desirable to define criteria for deciding when to use test results from similar equipment. One would also be interested in how dependent the a posteriori estimates are on the a priori assumptions. See Ref. A-23 with respect to this question.

Example 2

For a second example, suppose that tests have been made on a new transistor and that 0 failures have been observed in 10^5 hours. Assume that 5 failures were observed in 10^6 hours. Furthermore, assume the hazard rate is constant. Estimate the failure rate and obtain the a posteriori distribution assuming an a priori gamma distribution

$$f(\lambda) = \frac{t_0^{r_0} e^{-\lambda t_0} \lambda^{r_0 - 1}}{\Gamma(r_0)}.$$

A 100γ percent confidence interval estimate of λ may be obtained by

$$P\{\lambda_L \le \lambda \le \lambda_U\} = \int_{\lambda_L}^{\lambda_U} f_1(\lambda) d\lambda = \gamma.$$

Consider the problem of obtaining a 100 percent one-sided confidence interval estimate. In this case let the lower limit be zero and the upper limit be determined by the solution of λ_U in the equation,

$$\int_0^{\lambda_U} f_1(\lambda) d\lambda = \gamma.$$

It can be shown that the above equation can be expressed in terms of the χ^2 distribution as

$$P\{\chi^2 \leq 2\lambda_U\ (t+t_0)\} = \gamma$$

where χ^2 has a χ^2 distribution with $2(r_0+y)$ degrees of freedom. Hence for $r_0 = 5$, $t_0 = 10^5$, $t_0 = 10^6$, $y = 0$ one obtains

$$\chi^2_\gamma = 2\lambda_U\ (t+t_0)$$

or

$$\lambda_U = \frac{\chi^2_\gamma}{2(t+t_0)} = \frac{11.1}{2(1.1 \times 10^6)}$$

$$= 5.045 \times 10^{-6}.$$

The choice of the prior distributions is primarily for mathematical convenience. However, there is considerable freedom in the choice and depending upon the quality of the prior information one can select a distribution with a large or small variance. See Ref. A-23 with respect to further discussion pertaining to this problem. One should also refer to Ref. A-24 for an application of Bayesian decision models to a problem which considers the desirability of accepting a fixed price contract to build and maintain a system of N devices for a period of T years. In addition, a problem is posed for selecting the size of an experiment (number of devices to place on test) for obtaining profit larger than zero, subject to the prior information about the failure rate λ. The mean and variance of λ having the above distribution are

$$E\{\lambda\} = r_0/t_0$$

$$\mathrm{Var}\{\lambda\} = r_0/t_0^2 = E\{\lambda\}/t_0.$$

<u>Solution</u>:
The a posteriori distribution of λ given y failures in t hours is

$$f_1(\lambda|y) = \frac{f_0(\lambda)\ e^{-\lambda t}(\lambda t)^y/y!}{\int_0^\infty [\frac{t_0^{\ r_0}}{\Gamma(r_0)}\ e^{-\lambda t_0}\ \lambda^{r_0-1}\ \frac{1}{y!}\ e^{-\lambda t}(\lambda t)^y]d\lambda}.$$

Hence

$$f_1(\lambda|y) = \frac{f_0(\lambda)\ e^{-\lambda t}(\lambda t)^y/y!}{D},$$

where

$$D = \frac{t_0^{r_0} t^y \Gamma(r_0+y)}{\Gamma(r_0) y! \ (t+t_0)^{r_0+y}} .$$

Thus

$$f_1(\lambda|y) = \frac{e^{-\lambda(t+t_0)} \lambda^{r_0+y-1} (t+t_0)^{r_0+y}}{\Gamma(r_0+y)} .$$

For the example, let $t_0 = 10^6$ hours and $r_0 = 5$, to correspond to the observed number of failures in 10^6 hours of testing, then

$$f_1(\lambda) = \frac{e^{-\lambda(t+t_0)} [\lambda(t+t_0)]^{r_0+y-1} (t+t_0)}{\Gamma(r_0+y)} \text{, with } r_0 = 5,\ t_0 = 10^6,$$

and where y is the observed number of failures in the life test on the new transistor.

The mean of the a posteriori distribution is the Bayes estimate,

$$\hat{\lambda}_1 = \frac{r_0+y}{t+t_0} = \frac{5+0}{1.1 \times 10} = 4.54 \times 10^{-6}.$$

This compares with the prior estimate of

$$\hat{\lambda}_0 = 5 \times 10^{-6}.$$

A.13 Random Processes

Suppose that an experiment consists of measuring the deflection of a structural member during a dynamic test. A continuous recording of such a measurement might appear as the function X(t) shown in Fig. A.13-1. This type of behavior is usually called a random process, also frequently called a stochastic process. The independent variable does not always have to be time; it could, for example, represent distance along the length of a bar with X(t) representing a measurement of thickness or width as a function of distance.

One of the simplest forms for the random process X(t) is a constant in time but random in value. For this case the process thus reduces to a random variable, x, methods of treatment for which are given in previous sections. A step above this in complexity would be the introduction of more complex time functions having a known form but random in the values of the parameters which define it. Such random processes are called deterministic random processes and are typified by $X(t) = x_1 + x_2t$; $X(t) = x_1 \cos(x_2t + x_3)$, etc. where the x_i parameters are random variables. Given a functional form containing random variables $x_1, \ldots, x_k$, the process can be uniquely specified by the k-variate distribution $F(x_1, \ldots, x_k)$.

As the time variations increase in complexity and lose all semblance of functional form, the random process becomes an entirely random process. The function illustrated in Fig. A.13-1 is such a process. One form of representation is to assume it adequately approximated by a finite set of values corresponding to k closely spaced real numbers dispersed over the interval of interest (0, T). Let this set of values be denoted by numbers $x_1, \ldots, x_k$ where $x_i = X(t_i)$. This in effect represents a sampling of the function X(t) and the collection of values represents a random series, often referred to in statistical literature as a time-series. Random series also appear frequently in practice and in general can have different forms. For example, they can have random values occurring regularly in time, fixed values occurring randomly in time, or random values occurring randomly in time. The Poisson process is one type of random series in which events or impulses occur randomly in time.

Now suppose the test is repeated and a new sample of X(t) taken as described above. In appearances the new X(t) would be the same, but closer observation would reveal the individual x_i's to have different values than previously. Repeating the test many times would reveal that each x_i can be treated as a random variable. In concept then the process can be described by a k-variate distribution $F(x_1, \ldots, x_k; t_1, \ldots, t_k)$ where the t_i's are included to denote the time dependency. If the normal process. Processes having other distributions could similarly be designated.

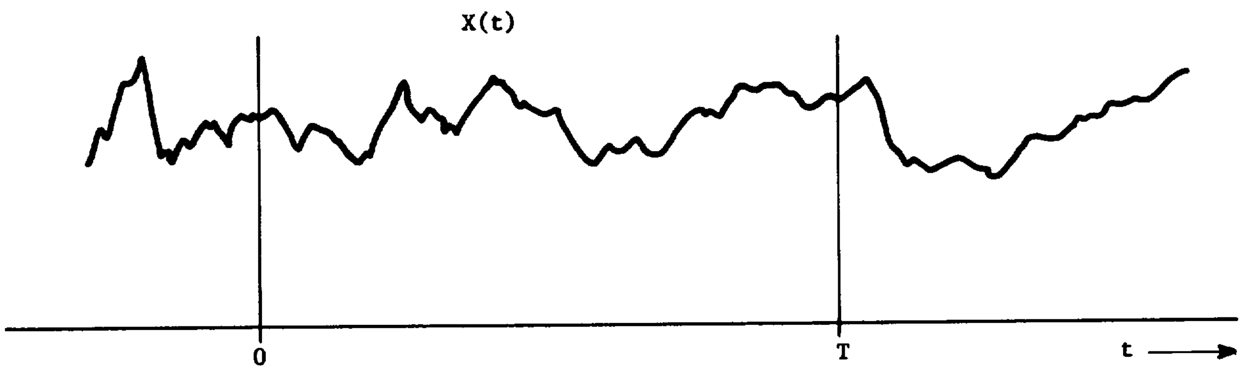


Figure A.13-1. Illustration of Random Process

Table A.13-1

Definitions of Characteristics of A Continuous Stationary Random Process*

Legend

$X(t)$ = a continuous stationary random process

t = time

x_i = $X(t_i)$

$f(x_i)$ = probability density function of x_i

$E\{x\}$ = mathematical expectation of x

$\phi(\tau)$ = autocovariance or autocorrelation function of $X(t)$

$P(\omega)$ = power spectral density function of $X(t)$

m = mean of $X(t)$

α = mean square of $X(t)$

σ^2 = variance of $X(t)$

ω = frequency in rad./sec.

Mean of X(t)

Ensemble:
$$m = E\{x_i\} = E\{x_j\}$$
$$= \int_{-\infty}^{\infty} x_i \, f(x_i) \, dx_i$$

Time:
$$m = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t)dt$$

Mean Square of X(t)

Ensemble:
$$\alpha = E\{x_i^2\} = E\{x_j^2\}$$
$$= \int_{-\infty}^{\infty} x_i^2 \, f(x_i)dx_i$$

Time:
$$\alpha = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X^2(t)dt$$

Other: a/
$$\alpha = \phi(\tau)\Big|_{\tau=0} + m^2 = \phi(0) + m^2$$
$$= 2\int_{0}^{\infty} P(\omega)d\omega + m^2$$

---

* Note indicated by a/ appears at the end of table.

Table A.13-1 (Continued)

Variance of X(t) [a]

Ensemble:

$$\sigma^2 = E\{(x_i - m)^2\} = E\{(x_j - m)^2\}$$

$$= \int_{-\infty}^{\infty} (x_i - m)^2 \; f(x_i)dx_i$$

Time:

$$\sigma^2 = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} [X(t) - m]^2 dt$$

Other:

$$\sigma^2 = \phi(\tau)\Big|_{\tau=0} = \phi(0)$$

$$= 2\int_{-\infty}^{\infty} P(\omega)\, d\omega.$$

Autocovariance or Autocorrelation Function of X(t)

Ensemble:

$$\phi(\tau) = E\{(x_i - m)(x_{i+\tau} - m)\}^{b/}; \qquad \tau = 0, 1, \ldots, k$$

$$= E\{(x_j - m)(x_{j+\tau} - m)\}; \qquad \tau = 0, 1, \ldots, k$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_i - m)(x_{i+\tau} - m) \; f(x_i, x_{i+\tau})dx_i \; dx_{i+\tau}$$

Time:

$$\phi(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} X(t) \; X(t-\tau)dt$$

Other:

$$\phi(\tau) = \int_{-\infty}^{\infty} P(\omega) \; \cos \; \omega\tau \; d\omega.$$

Power Spectral Density of X(t)

$$P(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(\tau) \; \cos \; \omega\tau \; d\omega.$$

---

[a] Note that the autocovariance and the power spectral density are defined in such a manner as to exclude the steady-state constant component. This affects the way that the latter two mean square and variance formulae are stated. Some texts may treat this slightly different. Note, however, that periodic components in X(t) are still included if present.

[b] Note that in this indexing scheme τ, the time unit, is an integer. The duration Δt of τ is defined by $\Delta t = T/k$ where k is the total number of intervals of equal length Δt into which the total interval (0, T) is divided.

In passing, one would expect there to be correlation between the x_i's depending upon the spacing between them and the rates of fluctuation of the waveform. This correlation is, of course, automatically accounted for in the joint distribution $F(x_1, \ldots, x_k; t_1, \ldots, t_k)$.

Consider now the characteristics of samples taken from a different interval $(t, t+\tau)$ of $X(t)$. Again a k-multivariate distribution is appropriate for description. If the distribution is the same for all such intervals for all t, then the process is a <u>stationary random process</u>.

There are "weaker" definitions of stationarity than that described above which are more useful in the practical sense. For example, a <u>first-order stationary</u> random process means that only $F(x_i; t_i)$ must not change over time. Definitions of any higher order stationarity are similarly permissable. In reality there is no such thing as a stationary process since everything must have a beginning and end. The concept is still very useful in practice, however, since many processes behave near enough like stationary processes long enough for us to perform an analysis.

The general treatment of nonstationary random processes is much more difficult than for stationary processes. It may be possible to transform some nonstationary processes to stationary ones, for example by subtracting out a linear trend or by multiplying by a time-varying function. Some investigations of a special class, viz., the integral of a normal stationary process, are described in Ref. A.27.

The remaining discussion is intended to provide further perspective on the treatment of random processes in practice. For brevity, the discussion treats only one general class, viz., continuous stationary random processes. Even though some of the concepts apply only to this class, the discussion provides appreciation for types of problems encountered in treating other classes.

Typical things done with random processes in practice are:

(1) measuring their characteristics,

(2) modifying their characterisitics by filtering,

(3) testing them for stationarity, and

(4) comparing them under different conditions.

Table A.13-1 lists some useful mathematical definitions of characteristics of continuous stationary random processes which are commonly employed in their treatment. Note that a distinction is made between ensemble averaging* and time averaging; however,

---

* This merely treats the process as if the population associated with a particular x_i represents the collection of values of $X(t_i)$ taken simultaneously from a large number of sample functions or realizations of $X(t)$ at time t_i.

these are equivalent if ergodicity* can be assumed and it usually is. Note especially that power spectral density and autocovariance (or autocorrelation**) are related to each other by Fourier (cosine) transforms. The power spectral density (or power spectrum***) is a very practical concept much used by engineers to represent how the power is distributed over the possible range of frequencies. The designation, power, evolved because of the analogy to power dissipated in a resistor if X(t) represented an applied voltage. Note, for example, that the variance σ^2 of X(t) is equivalent to the total power, i.e., the integral of $P(\omega)$ over all possible values of ω.

Note that the covariance expressed by the autocovariance definition is concerned only with two points of the same function displaced in time. It is very simple in concept but serves a very useful role as a step in measuring and computing the power spectrum for a waveform X(t). Prior to development of the concept, the measurement of power spectra relied upon Fourier harmonic analysis techniques. The concept is credited primarily to Norbert Wiener.

The relationship of autocovariance and power spectral density is illustrated in Fig. A.13-2 with rough sketches of measured and smoothed functions for a waveform X(t) containing both random and a periodic component. Note that the low frequency oscillation of period $2\pi/\omega_0$ in the autocovariance function gets transformed to a peak in the power spectrum. Furthermore, the peak of the autocovariance function centered at the origin indicates that a large portion of the waveform amplitude is contributed by the random component. Note that in the power spectrum there is significant distribution of power of higher frequencies. With the definitions in Table A.13-1 it can be discerned that the narrower or sharper peaks in the autocovariance function generally indicate higher frequency content in the random component, and conversely, the wider peaks synonymous with longer correlation times indicate lower frequency content. The distribution of power at the various frequencies thus become evident in the power spectrum.

---

* Because of the many subleties involved in the concept, there is no attempt to explain ergodicity. It pertains to the nature of the time behavior of X(t) and more specifically in terms of the resulting values achieved by a single realization X(t) in comparison with those of the population for x_i over the ensemble. The equivalence of ensemble averages and time averages for stationary processes follow from ergodicity. Reference A-28 gives some clear, interesting discussion of the concepts.

** Whereas most engineering texts refer to $\phi(\tau)$ as the autocorrelation function, statisticians seem to prefer to call this autocovariance function and designate $\phi(\tau)/\phi(0)$ as the autocorrelation function (see for example Ref. A-29). The distinction could be very important in calculations or in interpretation of results; however, in our discussion they are used interchangeably without fear of misinterpretation since their meanings in concept are so similar.

*** Comments similar to those given in the previous footnote can also apply to the relationship between power spectral density and power spectrum.

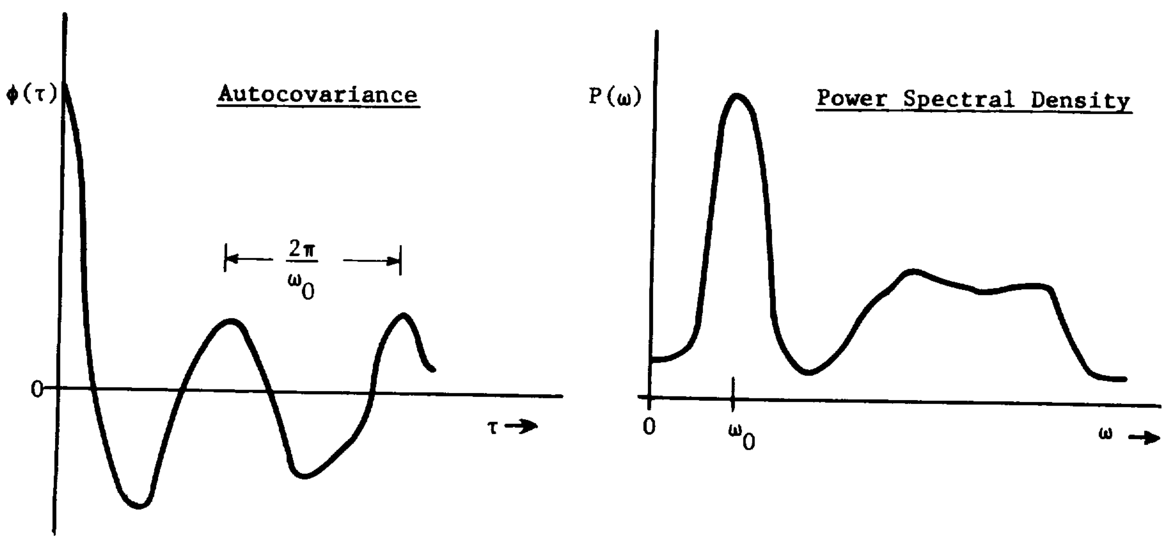


Figure A.13-2. Relationship Between Measured Autocovariance and Computed Power Spectral Density

Any practical treatment of random processes almost always involves some measurement of its characteristics. Most often these measurements are aimed toward eventual estimation of characteristics more directly associated with the power spectrum than with the autocovariance function since engineers traditionally think in terms of power and frequency. Measurements are usually made from a finite time sample of a continuous record of the process and usually assume ergodicity.

Both analog and digital computing techniques have been developed for measurement. Analog techniques treat the process on a continuous basis. Some still employ Fourier harmonic analysis procedures by passing the waveform through a bank of individual, narrow bandpass filters each tuned to a separate frequency. The power output of each is proportional to the power contained in the frequency range of the filter and a plot of the outputs versus center frequency of the filters yields a measure of the power spectrum. One of many problems here concerns the scaling to get realistic absolute measures. Some analog techniques also involve measuring autocovariance and transforming to power spectral density.

Digital computing techniques must make use of discrete data obtained by sampling the continuous waveform. This sampling generates a time-series from which the autocovariance function is typically estimated over a range of lag time τ and then converted to a power spectrum. Ref. A-29 is devoted entirely to describing the measurement of power spectra by this general approach. An excellent summary of the measurement problem is given in Ref. A-30 and is a good supplement to Ref. A-29 since it discusses some points of controversy. In addition to the problems with sampling variations and biases in the usual statistical sense, some practical problems associated with this approach are concerned with selecting:

(1) the sampling interval Δt,

(2) the total lag time in computing $\phi(\tau)$, and

(3) the spectral window in computing $P(\omega)$.

In the selection of Δt it must be remembered that no useful information about the frequency content of $X(t)$ is available between the data points once they are specified; for example, one cannot distinguish between sinusoids having half-wavelengths of Δt, $(1/2)\Delta t$, $(1/3)\Delta t$,* etc. since with proper choice of phase all of these can pass through the same two points. This effect is called aliasing and results in the power at these higher frequencies being erroneously added to the lower frequencies. Ideally, one would want to select Δt to correspond to the half-wavelengths of the sinusoid at the highest frequency of interest, i.e., if Ω_h is the highest frequency (in radians per second) of interest, then ideally $\Delta t = \pi/\Omega_h$.

* The frequency $\pi/\Delta t$ corresponding to the longest of these is called the Nyquist frequency.

Ref. A-30 suggests, however, that one should select Δt on the basis of the frequency Ω_p beyond which the total contribution of power is only 1% - 2% if it is known within any reasonable accuracy. This eliminates any problems with aliasing since the aliased power will contribute only a small amount. Even if $\Omega_p < \Omega_h$ then there is no significant loss of information since the power spectral density at frequencies greater than Ω_p will be too small to have any practical significance. If Ω_p is much larger than Ω_h then an alternative to avoid unnecessary computing is to choose Δt on the basis of the mean frequency, $\frac{1}{2}(\Omega_p+\Omega_h)$. All of the aliasing of the spectrum, essentially a folding back of the outer end, would thus occur at frequencies beyond Ω_h. The resulting spectrum would then be truncated at Ω_h for any further use.

In selecting the total lag it is suggested in Ref. A-30 that computations for autocovariance always be carried out to lags equivalent to 25% - 30% of the total length of the sample. Examination at that point may reveal that more are needed. Rationale for the decision is given in Ref. A-30. Generally, it must be based on the relative contributions of autocovariance at higher values of lag since this affects the precision of the power spectral density. It is noted that Ref. A-29 states that the truncation point for lag should be about 5% - 10% of the total sample length. The controversy is noted and clarified in Ref. A-30 however.

The spectral window is analogous to the cos $\omega\tau$ term in the Fourier transform defining the power spectral density (see Table A.13-1). It is described in Ref. A-30 how "smudging" or distortion of the power spectrum occurs due to sampling variability even when the autocovariance is known exactly. As summarized in the discussion there, much theoretical consideration has been given to trying different spectral windows in an attempt to decrease distortion. Unfortunately, less distortion is achieved at the expense of an increase in variance of the estimates. The relationship between these is complex and is discussed in detail in Ref. A-30.

The above discussion merely illustrates several of the practical problems in power spectra measurements of random processes. The reader attempting to undertake such measurements is advised to depend on the many texts and reports on this subject as it is impossible to summarize all of the important aspects of such a complex subject in a few pages. One should remember, however, that not all problems are associated with the statistical procedures. Along this vein it is appropriate to quote from Jenkins [Ref. A-30]: "In the last resort, if it is different to make sense of the spectrum from a physical point of view, then the more refined statistical considerations are irrelevant. In particular, if taking the two halves of the same series gives widely differing answers or if the next experiment produces a different spectral shape, then one has far greater problems than statistical ones."

References

A-1 Cramér, H. (1946). Mathematical Methods of Statistics, Princeton University Press, New Jersey.

A-2 Haight, F. A. (1960). Index to the Distributions of Mathematical Statistics, Journ. of Research, Vol. 65B, No. 1, Jan-Mar, pp. 23-60.

A-3 Anderson, T. W. (1958). Introduction to Multivariate Statistical Analysis, John Wiley & Sons, Inc., New York.

A-4 Feller, W. (1950). An Introduction to Probability Theory and its Applications, Vol. 1, John Wiley & Sons, Inc., New York.

A-5 Lehmann, E. L. (1959). Testing Statistical Hypotheses, John Wiley & Sons, Inc., New York.

A-6 Dixon, W. J. and Massey, F. J. (1951). Introduction to Statistical Analysis, McGraw-Hill Book Company, Inc., New York.

A-7 Epstein, B. and Sobel, M. (1953). Life Testing, Journ. of the American Statist. Assoc., Vol. 48, No. 263, pp. 486-502.

A-8 Hald, A. (1952). Statistical Tables and Formulas, John Wiley & Sons, Inc., New York.

A-9 Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters, J. Soc. Indust. Appl. Math., Vol. 11, No. 2, pp. 431-441.

A-10 Draper, N. and Smith, H. (1966). Applied Regression Analysis, John Wiley & Sons, Inc., New York.

A-11 Nelson, A. C. et al. (1967). Evaluation of Computer Programs for System Performance Effectiveness, Final Report SU-265, NASL.

A-12 Epstein, B. (1960). Tests for the Validity of the Assumption that the Underlying Distribution of Life is Exponential, Part I, Technometrics, Vol. 2, No. 1, Feb., 1960, pp. 83-102.

A-13 Ostle, B. (1967). Industry Use of Statistical Test Design, Industrial Quality Control, Jul, pp. 24-34.

A-14 Addelman, S. (1962). Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments, Technometrics, Vol. 4, No. 1, Feb, pp. 21-46.

A-15 Addelman, S. (1962). Symmetrical and Asymmetrical Fractional Factorial Plans, Technometrics, Vol. 4, No. 1, Feb, pp. 47-58.

A-16 Kiefer, J. (1961). Optimum Experimental Designs V, with Applications to Systematic and Rotatable Designs, Fourth Berkeley Symposium, 5, pp. 381-405.

A-17 Kiefer, J. (1959). Optimum Experimental Designs, J.R.S.S. (Series B), 21, pp. 272-319.

A-18 Hald, A. (1952). Statistical Theory with Engineering Applications, John Wiley & Sons, Inc., New York.

A-19 Scheffé, H. (1959). The Analysis of Variance, John Wiley & Sons, Inc., New York.

A-20 Kempthorne, O. (1952). The Design and Analysis of Experiments, John Wiley & Sons, Inc., New York.

A-21 Barlow, R. E. and Scheuer, E. M. (1965). Reliability Growth During a Development Testing Program, RM-4317-1-NASA, Contract No. NASr-21.

A-22 Press, S. J. (1965). Statistical Estimation by the Empirical Bayes Method: Some Extensions and Logistical Applications, Prepared for USAF under Project RAND-Contract No. AF 49(638)-700.

A-23 Breipohl, A. M., Prairie, R. R., and Zimmer, W. J. (1965). A Consideration of the Bayesian Approach in Reliability Evaluation, IEEE Transactions on Reliability, R-14, No. 2.

A-24 Howard, R. A. (1966). Bayesian Decision Models for System Engineering, IEEE Transactions on Systems Science and Cybernetics, SSC-1, No. 1, pp. 36-40.

A-25 Freeman, H. A., et.al.: Sampling Inspection: Principles, Procedures and Tables for Single, Double, and Sequential Sampling in Acceptance Inspection and Quality Control Based on Percent Defective, McGraw-Hill Book Co., New York and London.

A-26 Greenwood, V. A. and Hartley, H. O.: Guide to Tables in Mathematical Statistics, Princeton Univ. Press, Princeton, N. J., 1962.

A-27 Cramer, H. and Leadbetter, M. R.: Stationary and Related Stochastic Processes. John Wiley and Sons, Inc., New York, 1967.

A-28 Lee, Y. W.: Statistical Theory of Communication. John Wiley & Sons, Inc., New York, 1960.

A-29 Blackman, R. B. and Tukey, J. W.: The Measurement of Power Spectra. Dover Publications, Inc., New York, 1958.

A-30 Jenkins, G. M.: General Considerations in the Analysis of Spectra. Technometrics, Vol. 3, May 1961.

NASA-Langley, 1968 —— 15