AD NUMBER: AD0448079
LIMITATION CHANGES
TO: Approved for public release; distribution is unlimited.
FROM: Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 30 Sep 1964. Other requests shall be referred to Air Force Avionics Laboratory, Wright-Patterson AFB, OH 45433.
AUTHORITY
AFAL ltr dtd 14 Aug 1969

UNCLASSIFIED

AD 448079

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION ALEXANDRIA, VIRGINIA

UNCLASSIFIED

AL TDR 64-235

13

ACCELERATED LIFE TESTING OF GUIDANCE COMPONENTS

TECHNICAL DOCUMENTARY REPORT NO. AL TDR 64-235

30 SEPTEMBER 1964

AIR FORCE AVIONICS LABORATORY
RESEARCH AND TECHNOLOGY DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

PROJECT NO. 3181, TASK NO. 318108

(PREPARED UNDER CONTRACT NO. AF33(615)-1157
BY THE AUTONETICS DIVISION OF NORTH AMERICAN AVIATION, INC.
ANAHEIM, CALIFORNIA
AUTHORS: B. B. WINTER, C. A. DENISON, H. J. HIETALA, F. W. GREENE)

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Qualified requesters may obtain copies of this report from the Defense Documentation Center (DDC), (formerly ASTIA), Cameron Station, Building 5, 5010 Duke Street, Alexandria, Virginia, 22314.

DDC release to Office of Technical Services (OTS) not authorized.

Copies of this report should not be returned to the Research and Technology Division, Wright-Patterson Air Force Base, Ohio, unless return is required by security considerations, contractual obligations, or notice on a specific document.

FOREWORD

This report was prepared by the Reliability Engineering Unit, Systems Engineering, Navigation Systems Division of the Autonetics Division of North American Aviation, Inc., Anaheim, California, on U.S. Air Force Contract No. AF33(615)-1157, "Accelerated Life Testing of Space Guidance Components."

The contract was sponsored by the Air Force Research and Technology Division under Project 3181, "Guidance for Space Systems," Task 318108, "Space Guidance Reliability Investigations." Project Manager is Mrs. Rita Gustin, Navigation and Guidance Division, Air Force Avionics Laboratory, Wright-Patterson AFB, Ohio. The work was performed during the period December 1963 through July 1964, under the guidance of Mr. K. N. Stevens. Mr. B. B. Winter, Principal Investigator, provided detailed technical direction for the study. The following technical personnel made major contributions in the indicated areas:

Conceptual Definition and Model Development	B. B. Winter H. J. Hietala
Equipment Analysis	C. A. Denison
Properties of Materials	F. W. Greene C. G. Jennings

Specifically, Parts I, II, and III of the study are due principally to Mr. B. B. Winter and Mr. H. J. Hietala; Part IV to Mr. C. A. Denision, Part V to Mr. F. W. Greene and Mr. C. G. Jennings; Part VI to Mr. B. B. Winter and Miss S. C. Butler, Autonetics personnel in the Gyro Development Evaluation Group, in the Accelerometer Developm..t, and in the Inertial

Instruments Research Group provided valuable assistance in the analysis of interial navigation instruments.

Dr. B. Epstein (Consultant, Palo Alto) and Dr. I. Shimi (University of California, Riverside) provided consultation on statistical aspects of the study. Mr. G. Lorden of Cornell University and Mr. T. L. Johnston (University of California, Los Angeles) provided several helpful suggestions.

Forward (Cont)

ABSTRACT

The report deals with theoretical developments necessary for accelerated testing and with specific hardware considerations. In the theoretical area, the currently prevalent approach to accelerated testing is described and analyzed, and is found inadequate. A comparative discussion of various failure models is presented, and those most amenable to meaningful accelerated testing are developed in [illegible]els for devices with multiple failure modes are described, and pertinent estimation procedures are reviewed. A new method for efficient estimation of the failure distribution of repairable equipment is presented. In the hardware area, the report includes a detailed examination of several space guidance components, with emphasis on considerations pertinent to accelerated testing, and specific applications of accelerated testing to space guidance components are recommended. The phenomenon of metallic creep is examined from the standpoint of its relation to failures of space guidance components, and methods of accelerated estimation of creep behavior are examined.

TABLE OF CONTENTS

Section		Page

I. INTRODUCTION

I.A SUMMARY AND RECOMMENDATIONS

Modern technology is frequently called upon to produce devices which will, without maintenance or other human intervention, preserve their technological potential over long periods of time. An amplifier in a sea bottom telephone cable is expected to perform its function continuously over a period of many years. An emergency transmitter in an aviator's survival kit is expected to function on demand, even after years of idle storage. In many such instances one wishes to perform appropriate tests and, based on data resulting from such tests, to make inferences regarding the device's capability of adequate performance.

When a device is considered for use as a space guidance component, it is evaluated with respect to many criteria. It must satisfy certain functional requirements, i.e., barring any malfunction whatsoever, it must be capable of performing some specified function; e.g., it must be capable of detecting changes in velocity which exceed some specified minimum. It must satisfy certain system requirements; e.g., its weight and power consumption must be compatible with the propulsive and power-generating capabilities of the system in which the device is to be used. Finally, it must satisfy certain reliability requirements; i.e., it must be capable (with very high probability) of performing its function within

---

Manuscript released by authors 30 September 1964 for publication as an RTD Technical Documentary Report.

specified performance limits for a sufficiently long period of time. This report is concerned only with a device's performance with respect to reliability requirements. Ascertaining the reliability of space guidance components presents some special problems. In space guidance applications, a device will encounter operating conditions (environments) which are only approximately predictable, since present knowledge of space environments is sometimes little more than conjecture. In addition, a whole host of problems is presented by the extensive duration of space missions. Space guidance components must function properly after a long interval, or throughout a long period. Clearly, reliability requirements for long missions present a considerable technological challenge. However, this report does not deal with the problems encountered in meeting such reliability requirements. Rather, it deals with the problems in inference which arise when those requirements are apparently met; that is, we are concerned with problems of estimation for devices whose times to failure are known to be long, though their distribution is essentially unknown. Furthermore, this report does not attempt to further the state of knowledge regarding the environments likely to be encountered in space missions. Rather, it deals with methods of inference which can be applied once a plausible operating environment has been specified.

Methods of inference are briefly surveyed in order to clarify the relative setting of the body of endeavor known as accelerated life testing. The currently prevalent approach to accelerated testing, subsequently labelled the elementary approach, is subjected to critical analysis and is found to be inadequate. Thereafter, the report presents a comparative discussion of various failure models, and those most applicable to accelerated testing are developed in detail. Once this theoretical background is established, the report presents a detailed examination of several space guidance components, with emphasis on considerations pertinent to accelerated testing. Specific applications of accelerated testing to space guidance components are recommended. The phenomenon of metallic creep is examined from the standpoint of its relation to failures of space guidance components, and methods of accelerated estimation of creep behavior are examined. The report concludes with a discussion of several special topics, including the statistical consideration of devices with several modes of failure.

Recommendations for future work are given in detail following the technical discussion of Part II. In summary, the recommendations are:

(i) The currently prevalent, elementary approach to accelerated testing should not be used; furthermore, the elementary approach is not worthy of further development, except for those procedures concerned exclusively with the selection of the better of two models of a device.

(ii) It is recommended that the time-compression approach to accelerated testing, which effectively consists of "cycling" certain pertinent stresses more frequently in test than under use conditions, be applied to three typical space guidance components. It should be applied to thermal expansion compensation-bellows, to miniature pumps used for positive pressure lubrication in velocity meters, and to failures of gyroscope gas bearings resulting from on-off cycling.

(iii) It is recommended that the damage accumulation concept, extensively studied in this report, be applied to the accelerated testing of two typical space guidance components. It should be applied to the miniature precision ball bearings used in space guidance instruments, and to long-life motor brushes (and/or slip rings) used in such instruments. It is recommended that the recently discovered effect of ultrasonic energy on metallic creep be developed into a practical tool for accelerated testing of microcreep phenomena which cause failures in inertial guidance instruments, e.g., mass unbalance or loss of alignment of critical components.

(iv) It is recommended that statistical techniques required for the analysis of devices with multiple failure modes be developed further; they are required for accelerated testing of space guidance components since such components exhibit several modes of failure and only some of those modes are now amenable to accelerated testing.

(v) Methods for efficient utilization of data for complex, repairable devices were developed in the course of the current study. It is recommended that these methods be further developed, since they are the only presently available techniques for studying guidance instruments as a whole, as opposed to accelerated testing of discrete components.

The underlying motivation of the effort reported here is the reliability analysis of space guidance components. Nonetheless, frequent reference will be made to simple devices such as light bulbs or automobile tires. Since the authors and the prospective readers of this report have a sound intuitive grasp of the behavior of such devices, they will be used to illustrate certain concepts, even though they have no relation whatsoever to space guidance.

I.B METHODS OF INFERENCE

I.B.1 Introduction

In this study one is not concerned with minor variations in a device's performance but, instead, one deals with two large classes of performance levels: the class of performance levels at which the device is considered satisfactory and the class of performance levels at which the device is considered unsatisfactory, or failed. To avoid unnecessary complication, the discussion in this report will generally be oriented to irreversible (sometimes called catastrophic) failures. That is, any device under consideration can at any time be unambiguously described as good or as failed and, once failed, it remains permanently in the latter state. The length of stay in the good state is customarily called the time to failure (TTF). One generally cannot state, a priori, the TTF of a particular device and hence, the TTF is considered a random variable. The distribution function of the TTF random variable is customarily called the failure distribution.

The most direct approach to making inferences regarding a device's capability of adequate performance over a specified period of time consists of the statistical analysis of data gathered in simulative tests. That is, sufficiently many samples of the device under consideration are subjected to a test which simulates, to the maximum extent possible, the conditions under which the device is going to be used. In particular, the length of use is also simulated; i.e., the device is tested for a period whose length is comparable to the duration of the proposed mission or the duration of the proposed standby status. We shall henceforth refer to this approach as the totally simulative testing (TST) approach. Clearly, if the proposed mission is very long, the TST approach can be very costly in terms of testing time. Furthermore, if the device is very reliable for the proposed mission, it follows from elementary statistical considerations that the demonstration or the precise estimation of that reliability will require that many devices be tested. Thus, one is faced with a

> dilemma: one would like to make inferences regarding a device's capability of adequate performance without engaging in a testing program which is prohibitively expensive, either in terms of time, or the number of devices required for test, or the costs resulting from erroneous inferences.

Attempts at the resolution of that dilemma are the subject of this report.

I.B.2 Life Testing

A very prevalent alternative to totally simulative testing (TST) is in the field of endeavor known as life testing. In life testing, one no longer requires that every device be tested for a period comparable to the length of the proposed mission, though one does require that all other use conditions be simulated to the maximum extent possible. The TST approach essentially amounts to estimating directly the probability of success in a specified mission. On the other hand, the life testing approach is essentially concerned with the distribution function of TTF; the probability of success in a specified mission is left to be derived from the failure distribution. The mission-length simulation requirement is generally removed by making assumptions regarding the form (i.e. family) of the failure distribution. For instance, if TTF is assumed to be exponentially distributed, then estimation of reliability for a mission of some specified length can be performed with results from tests which are either shorter or longer than the mission duration. However, distribution-free (nowadays often called nonparametric) methods exist for estimating the failure distribution. Of course, if a nonparametric estimate of the failure distribution is to be used to make inferences regarding a device's capabilities in a mission of some specified duration, the mission-length simulation requirement must be partly reinstated. That is, not all, but at least some devices must be tested for a period whose length is comparable to the duration of the proposed mission or standby status. It was shown in Subsection I.B.1 that the TST approach to making inferences regarding a device's capability of adequate performance leads to a dilemma. Life testing indeed is an approach to that dilemma since it allows occasional reduction in test duration; when distributional assumptions are made, a certain reduction in sample size can also be achieved. Furthermore, possible costs resulting from erroneous inference are somewhat controlled since the assumptions underlying the life test are subject to statistical testing, and since the statistical properties of the methods of inference are known. Nonetheless, life testing, per se, will not be considered in this report, since it is extensively studied in contemporary statistical literature. However, mention will be made of methods which are not currently in use in life testing but which can improve one's ability to extract information from experimental data.

I.B.3 Accelerated Testing

Approaches to the previously stated dilemma which do not fall exclusively within the province of life testing will be collectively referred to as accelerated testing. Three broad approaches are discussed in this report.

The first approach exploits the fact that the reliability (in the broad sense of the word) of a device inevitably depends on the operating conditions (environments) which it will encounter; e.g., the lifetime of an automobile tire is clearly not the same on modern freeways as on more primitive roads. The distribution function of the time-to-failure random variable (TTF) is different in the presence of different experimental conditions, i.e., different environments. The gist of this approach is the exploitation of whatever knowledge one has, regarding the variation of failure behavior with environment, in order to allow testing at environmental conditions which differ from the operating environment; the purpose being, of course, to produce the requisite information at less cost, either in terms of time or in terms of devices placed on test, than the cost of testing under conditions which simulate the operating environment. Where this approach needs to be singled out from other approaches to accelerated testing, it will be termed "over-stress testing" (OST).

A second approacn is suitable when the device under consideration fails under the influence of a stress which, in use, is not constantly present. In that case, the random event called failure cannot be meaningfully identified with the mere passage of time; rather, it corresponds to the accumulated number (and kind) of stress cycles. In such an instance, tests can be performed either by applying the stresses in a manner resembling their occurrence in practice, or one can apply them more rapidly. Clearly, the latter method will yield the requisite data sooner. This approach shall be termed the time-compression (TC) approach to accelerated testing. If the failure behavior is assumed to be governed exclusively by the accumulated number of stress cycles, then an over-stress test - in the sense of the preceding paragraph - would consist of changing the experimental conditions so as to modify the number of stress cycles required for the onset of failure. In contrast to this, the TC approach explicitly depends on the assumption that the number of stress cycles required for the onset of failure is not changed when the rate of stress-cycling is changed.

The over-stress testing (OST) and time-compression (TC) approaches will be collectively referred to as accelerated life testing (ALT). Both approaches are attempts at life testing under conditions which accelerate the onset of failures, hence they are well described by the term accelerated life testing.

A third approach consists of variations on methods of testing properties of materials, and frequently involves techniques akin to those of OST and TC. This approach is advisable when a relationship can be established between certain failures of devices and some corresponding properties of materials. For instance, the center of mass of a gyroscope shifts due to creep which results from the relaxation of residual stresses or from applied stresses. It may not be possible to directly relate the creep properties of materials to the behavior of gyroscopes made of such materials. Nonetheless, it is a fair presumption that the material which creeps less under laboratory test conditions is the better material for gyroscopes. Thus, it becomes desirable to develop methods of obtaining creep data in reasonable time.

It should be noted that the methodology of life testing has extensive applications in accelerated life testing. The OST approach frequently depends on life tests performed in various environments. The TC approach relies on life tests at various stress-cycling rates. Estimating the distribution of the amount of creep under specified conditions is theoretically akin to life-testing.

I.C DEPENDENCE OF FAILURE BEHAVIOR ON ENVIRONMENT

I.C.1 Introduction

The over-stress approach to accelerated testing (OST) is predicated on some knowledge of the "dependence of failure behavior on environment." This section contains a discussion of that concept, and introduces some related definitions which are used in subsequent discussions.

Generally, time to failure (TTF) has a different distribution function if the experimental conditions are changed. The "dependence of failure behavior" is the relation between experimental conditions and the distribution function of the TTF random variable. For instance, the distribution function of the time to failure of light bulbs depends on the voltage at which the light bulbs are operated. The distribution functions for three different operating voltages might be as shown in Figure I-1.

Note that in most instances, existing theoretical and experimental knowledge is insufficient to specify the dependence of the failure distribution on experimental conditions. However, it is possible in some instances to have information regarding the regression function, i.e the dependence of the mean on experimental conditions. In the physical sciences, and in most engineering endeavors, knowledge of the regression function is adequate, since one is usually interested in the "average" behavior of devices. But, from the standpoint of reliability considerations, knowledge of only the regression function is totally inadequate;

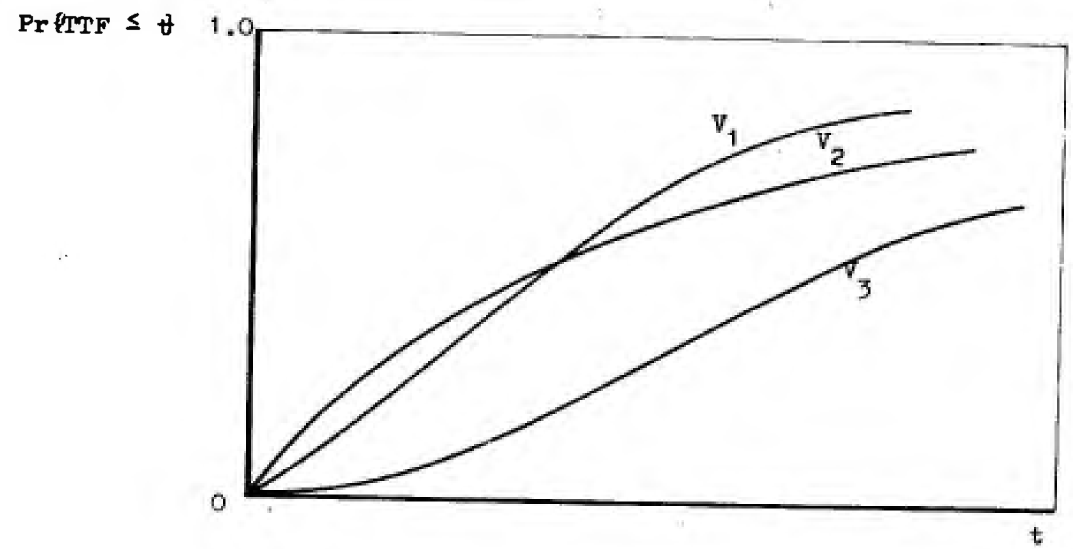


FIG. I-1. POSSIBLE VARIATION IN FAILURE DISTRIBUTION one is not interested in average behavior, but in the occurrence of certain rare events, namely failures. The expected time to failure is not nearly as significant as some low fractile of the distribution, say the first or fifth percentile (i.e., a time such that the probability of failure prior to that time is one or five percent). In the special case when the failure distribution is exponential under all experimental conditions considered (e.g. at any operating voltage for light bulbs), then the regression function contains all the necessary information since the exponential distirbution is fully described by its mean.

In the context of the above discussion, one can say that accelerated life testing is the making of inferences regarding certain properties of the failure distribution for some conditions at which no experiment was performed, based on results from experiments at some other conditions. Accelerated life testing is essentially an extrapolation to untried experimental conditions. Clearly, when no knowledge about the relation between the failure distribution and the experimental conditions is available, accelerated life testing (in the specific meaning of Subsection I.B.3) cannot be performed and experimentation must be performed at all conditions of interest. When some knowledge is available, the methods of experimentation and inference are related to the form of that knowledge, i.e., life testing is based on judicious exploitation of the dependence of a device's failure behavior (as described by the distribution function of TTF) on experimental conditions.

I.C.2 Definitions

Let the symbol E denote the experimental conditions, or environment; denote distinct environments by adding a subscript to E, e.g., E_i or E_j. For any device under consideration, there is a set of conditions under which it presumably will be used; call those conditions the operating or use environment, and denote that environment by E_u.

Now, the dependence of failure behavior on environment can be conveniently formalized as follows:

> Denote by F_i the distribution function of the time-to-failure random variable under conditions E_i. If, for some environments E_i and E_j, the distribution functions F_i and F_j are everywhere continuous, then there exists a function a_{ji} such that
>
> $$F_j(t) = F_i[a_{ji}(t)]$$
>
> The function a_{ji}, called the <u>time transformation function from environment E_j to environment E_i</u>, fully describes the dependence of failure behavior on environment.

At this point, two comments are in order. First, the name, time transformation function, indicates clearly the role of that function. If one knows the value of t for which $F_j(t)$ has some specified value, then $t^* = a_{ji}(t)$ is the time for which F_i takes on the same value, i.e. $F_j(t) = F_i(t^*)$. For instance, if t_α is the time such that the probability of failure by that time is α under environment E_j, then $t^*_\alpha = a_{ji}(t_\alpha)$ is a time such that the probability of failure by that time, under E_i, is also α. Thus a_{ji} transforms a fractile of the distribution function for environment E_j to the same fractile under environment E_i.

Secondly, the function a_{ji} introduced above was considered previously by Allen, [I-1] and [I-2], who called it the accelerating function. The desirability of standardized terminology notwithstanding, the authors of this report chose to introduce a new name; this was done in order avoid confusion with another function, introduced below, which it is appealing to call the acceleration function.

For example, suppose that the TTF is exponentially distributed in two distinct environments, say E_i and E_j; let θ_i and θ_j denote the mean lives under those two environments. Then, using the notation introduced above,

$$F_i(t) = 1 - \exp(-t/\theta_i) \quad ,$$

$$\begin{aligned} F_j(t) &= 1 - \exp(-t/\theta_j) \\ &= 1 - \exp(-t\theta_i/\theta_j\theta_i) \\ &= 1 - \exp[-t(\theta_i/\theta_j)/\theta_i] \\ &= F_i(t\frac{\theta_i}{\theta_j}) \quad . \end{aligned}$$

Thus, in this case, the time transformation function is just a constant multiplier,

$$a_{ji}(t) = Kt$$

where

$$K = \theta_i/\theta_j$$

If E_i is in fact the use environment E_u, then the constant K obtained above corresponds to what is commonly called the acceleration factor (or, sometimes, the "K-factor") in current reliability practice.

Now consider variations in only one aspect of the environment, so that the various environments can be described by a scalar; that is, the set of environments under consideration can be mapped one-to-one into the real numbers. For instance, consider life tests of capacitors at various operating voltages; then the environment E_i, which consists of testing with some applied voltage V_i, is adequately described (among all environments which differ in voltage only) by the scalar V_i.

An interesting special case is one in which

(i) variations in environment can be described by a scalar

and

(ii) the time transformation function is specified by an acceleration factor, i.e. for any two environments, E_i and E_j, there exists a constant K_{ji} such that

$$a_{ji}(t) = K_{ji}t \qquad \text{for all } t.$$

In that case, the function which relates the acceleration factor to the scalar index of environment is called the <u>acceleration function</u>. For instance, it is frequently assumed that capacitors have the exponential failure distribution, and that the mean life is proportional to some power of the applied voltage, i.e.

$$\theta(V) = CV^b$$

where $\theta(V)$ is the mean time to failure under applied voltage V. Then denoting by V_u the voltage at which the capacitor is normally used, the acceleration function, say K(V), is

$$K(V) = \theta(V_u)/\theta(V)$$

$$= CV_u^b/CV^b = (V_u/V)^b$$

since

$$F(t|V) = 1 - \exp(-t/\theta(V))$$

$$= 1 - \exp[-t\theta(V_u)/\theta(V)\theta(V_u)]$$

$$= 1 - \exp[-tK(V)/\theta(V_u)]$$

$$= F_u[tK(V)]$$

I.D STATE OF THE ART

The authors of this report have no evidence that an accelerated life test was ever performed which, in their opinion, is completely valid. A few instances of over-stress testing are known in which all steps are correct, provided certain more-or-less reasonable assumptions are taken for granted. However, the tests cannot be said to be completely correct, because those assumptions were not explicitly checked, although it was possible to test some of them. Several instances are known in which the time compression approach was used in a rather convincing manner, but precautions, needed to assure the complete validity of the test, were not taken. However, in general, the use of over-stress testing (OST) is essentially a matter of folklore. This does not mean that OST was never used, but only that it was hardly ever used meaningfully.

The folklore (i.e. presently prevalent belief) regarding over-stress testing pertains to two ideas. It presents a presumed reason for the application of OST and it deals with the method of application.

Regarding the reason for application, the folklore is that "accelerated testing must be done to save time." In fact, if saving time is the only concern, it is not desirable to rely on OST; the use of OST becomes desirable only if both the cost of devices placed on test and the cost of test time are considered. This matter is explored further in Sections II.A, II.B, II.C, and II.D.

Regarding the method of application, the folklore is essentially the following:

The failure distribution is exponential in all steady-stress environments; the "K-factor," i.e. the value of the acceleration function for a particular environment, is known (e.g., an estimate based on previously obtained data is used); by use of that K factor, all results obtained under accelerated conditions can be translated to use conditions.

In this report we shall refer to that method as the elementary method (of over-stress testing). In Section II.A, it is indicated that the method potentially applies to three distinct problems. The method is then studied in detail for each of those three problems in Sections II.B, II.C, and II.D. In particular, the effect of small errors in the assumed value of the K-factor is shown to be disastrous. Since the K-factors are never known very precisely, this renders the elementary method of OST worthless. The uncertainties in the assumed values of K-factors are an inevitable consequence of the method by which they are generated. Recall that a "failure-rate curve" merely presents the K-factor for a variety of environments, i.e. it corresponds to what is called the acceleration function in this report, and consider the following statement by Adams [I-3]:

"A component part failure rate curve is generated by gathering all failure rate data available for the component part and for related types... the data points are used to produce a graph of Failure Rate vs. Thermal Stress... This graph will have a family of Electrical Stress curves, fitted as smoothly as possible to data points available... The relationships expressed by the smoothed curves are then replotted as a family of Thermal Stress curves and anomalies in the relationship are uncovered... The plotted curves are then adjusted until both [sets of] curves are smooth..."

Clearly, acceleration functions produced in this manner are subject to considerable errors.

Some variations of the elementary approach to OST can be found in the work of Levenbach [I-4] and Allen, [I-1] and [I-2]. Though still assuming that the failure distribution is exponential at all levels of stress, they only assume the form (and not the actual values of the constants) of the acceleration function. They present statistically meaningful estimates of the important parameters of the acceleration function. In [I-2], Allen obtains the maximum likelihood estimate of the failure rate at usage conditions, under appropriate assumptions, for a test in which stress increases continuously in time. However, these publications date back to 1958, and no work of comparable significance has been published since that time.

An interesting contribution, though not quite correct, was made by Endicott and Zoellner [I-5]. They suggested that a damage accumulation concept of failure provides a potentially fruitful framework for accelerated testing (in the OST manner). That approach has been studied further in this report (Subsections III.B and III.C). Though it is not yet completely developed, it is felt that it is indeed a promising approach.

It should be noted that all of the approaches discussed above pertain to accelerated testing of fairly simple components. Accelerated testing of complex devices, in which the device is tested as a unit rather than component-by-component, does not appear feasible. This conclusion is essentially based on the current lack of knowledge regarding proper statistical handling of competing failure risks. All but the simplest components exhibit many modes of failure. Since the various modes do not respond comparably to changes in stress level or changes in the stress-cycling rate, there is need to consider separately the processes leading to the various failure modes.

When the distinction among different failure modes is significant, there are certain theoretical results required to handle the data; these results are missing. The data pertaining to various failure modes must be "decoupled" for separate analysis and separate extrapolation to use conditions. Some effort along those lines is reported by Berkson and Elveback [I-6], Epstein [I-7], Allen [I-8], and Berman [I-9]. A complete discussion of the matter of mixed failure modes for complex devices and some components was prepared during the study which is the subject of this report. The discussion, together with some new results derived in the course of this study, is presented in Section VI.A.

For complex devices which are repairable, present estimation techniques fail to utilize all available data efficiently. Considerable progress in that direction was made during the study which is the subject of this report (see Section VI.B).

II. CRITICAL ANALYSIS

II.A INTRODUCTION

II.A.1 Scope

Part II is essentially a detailed exploration of the elementary approach to accelerated life testing. In Section I.D , the elementary approach was defined as follows:

> The failure distribution is assumed to be the exponential distribution under all environments; the K-factor, i.e. the value of the acceleration function for a particular environment, is assumed to be known; by use of that K-factor, all results obtained under accelerated conditions can be translated to use conditions.

Since the method is predicated on the assumption of exponentiality, each section of Part II includes a discussion of methods of inference applicable to devices with exponential failure distribution. The assertion (made in Section I.D) that OST is not advisable if the failure distribution is hypothesized and a reduction in test-time is the only desideratum, is illustrated in each section of Part II.

As an introduction, Subsection II.B.1 contains some discussion not exclusively related to the elementary approach. An improvement on the elementary approach to a particular problem in acceptance testing is discussed in Subsection II.D.3. Since most of the publications dealing with accelerated testing are, in fact, based on the elementary approach, Part II concludes with a survey of the pertinent technical literature.

ALT is potentially applicable to three distinct problems in inference, namely selection, estimation, and decision. In the selection problem, one wishes to choose the better of several classes of devices (better in the sense of higher reliability for a specified period of time). In the estimation problem, one wishes to estimate the reliability (over a specified period of time) of a class of devices. In the decision problem, one wishes to decide whether a class of devices meet a specified reliability criterion. To what extent ALT is applicable to the above three problems depends on the strength of the assumptions one is willing (i.e. justified) to make regarding the failure distribution, and regarding its relation to the experimental conditions. Strong assumptions must be made if ALT is to be applied to the estimation problem. Very strong assumptions (hardly ever justifiable in practice) are required for the application of ALT to the decision problem. The application of ALT to the selection problem also requires strong assumptions; however, the assumptions required in the case of selection are generally somewhat weaker than those required in the case of estimation, and are much weaker than those required for the decision problem. The hierarchy of assumptions required for the application of ALT to the three problems of inference is further discussed and illustrated throughout Part II of this report.

The terminology introduced above (delineating the selection, estimation, and decision problems) is not strictly in keeping with current statistical usage. What is here termed the selection problem does, in fact, deal with a decision: either one decides that Product A is the better one, or that Product B is the better one. In fact, even some estimation procedures are based on decision-theoretic considerations; e.g. the estimation procedure discussed in Subsection II.C.2 is a minimax best estimate with respect to a certain loss function. Nonetheless, in this report, the term "decision" will be used in a restricted sense; it is clear, from the way the term is introduced above, that in the usage of this report the term essentially corresponds to what is known (in current reliability practice) as acceptance or demonstration testing.

II.A.2 Applicability of ALT

Ideally, one knows a device's failure distribution for all conceivable experimental conditions. In that instance, no life-testing whatsoever is required. On a lower level of knowledge, one knows the form (e.g. gamma, or Weibull) of the failure distribution under operational conditions - or one is willing to assume a form. In that case, statistical techniques exist, at least in principle, for treating questions of selection, estimation, or decision. [The statement "in principle" is used because ~~there are~~ some families of distributions for which the procedures are not yet developed because there was no practical demand; there also are some special problems (e.g. non-homogeneous censoring of data) which have not yet been studied for some families; in some cases procedures exist, but they are not optimal. That is, there does not exist a complete handbook of statistical methods, but past statistical developments indicate that those procedures which do not yet exist can be devised if it becomes practical to expend the requisite effort.] As a rule, when the form of the failure distribution is known (or assumed, i.e., taken for granted), then questions of inference in any of the above three categories can be resolved in arbitrarily short time; this statement is amplified and illustrated in Sections II.B, II.C, and II.D. However, this is only feasible if appropriately large samples of specimens are put on test.

As a rule, the application of ALT requires assumptions both regarding the form of the failure distribution, and regarding the relation of that distribution to the environment. Since assumptions regarding only the failure distribution are sufficient to resolve all questions of inference in arbitrarily short time, ALT clearly is not required if one assumes the form of the failure distribution and if one merely wishes to reduce test time. However, since the resolution of questions of inference in very short time may require exhorbitant sample sizes, it may be desirable to apply ALT to reduce the sample size. Thus, when assumptions regarding the failure distribution are made, the question whether it is desirable to apply ALT is fundamentally a question of cost. If one is willing to assume the form of the failure distribution, then ALT is undesirable (since ALT is based on additional assumptions) when the cost of samples (and the life-tests performed on them) is very low. But if the cost of samples (or the cost of life-tests performed on them) is significant, then it appears desirable to apply ALT even though it requires assumptions beyond those regarding the form of the failure distribution. However, an additional cost must be considered, namely the cost of errors in inference which can arise from errors in the assumptions.

This question is explored further in Subsections II.B.4, II.C.3, and II.D.2.

The above discussion started from the idealized situation in which the failure behavior of a device is completely known, and then concerned itself with the somewhat more realistic case in which only the form of the failure distribution is assumed known. In the worst case, one may be faced with a situation in which one has no good reason for making assumptions regarding the form of the failure distribution. In that case it is generally true that neither ALT nor increased sample size can be used to reduce test time. Though nothing can be done in this case to reduce test time, the use of ALT for reduction of sample size is possible under certain circumstances in connection with the selection problem. Those circumstances (i.e. the requisite assumptions) are discussed in Subsection II.B.1. However, one case is known in which test time reduction is possible even when no distributional assumptions are made; this exceptional case is discussed in the next paragraph.

Certain assumptions, discussed in Section III.B, specify the so-called Simple Wear model of failure. If these assumptions can be made then, without any explicit assumptions regarding the form of the failure distribution, ALT may be applicable to the reduction of test time. For instance, under the Simple Wear model, ALT may be usable for the estimation of a low fractile of the failure distribution by a test whose duration is relatively short in comparison to that fractile. This approach to accelerated testing is relatively unexplored in the open literature but is discussed in considerable detail in Sections III.B and III.C of this report.

Finally, note the differences between the over-stress testing (OST) and time compression (TC) approaches to ALT. The TC approach is based on the simple assumption that the distribution of the appropriate random variable (e.g. cycles-to-failure) is not changed at all by the change from operational to testing conditions (e.g. increase in cycling rate). If that assumption is valid, then the TC method is absolutely correct. If however there is an error in that assumption, i.e. some time-dependent phenomenon is present, then results based on the TC approach are in error - in some unknown fashion. The effect of possible errors in the TC approach is not explored in detail in this report. On the other hand, the assumptions underlying the OST approach are rather complex; they pertain to the manner in which the failure distribution changes when some environmental factor ("stress") is changed. However, errors in this approach are more amenable to analysis. Therefore the discussion in Sections II.B, II.C, and II.D pertains to the OST manner of accelerated life testing.

II.B SELECTION

II.B.1 Preliminary Example

A frequently encountered situation in life-testing is that in which one cannot reasonably make any a priori assumptions regarding the failure-behavior (i.e. distribution function of TTF) of the device under consideration. In this case, one must test the device under conditions which simulate, as closely as possible, the use-conditions intended for the device. In particular, the duration of the test on at least some devices in the sample must at least equal the mission duration. One's inferences, in this case, must be based on methods of distribution-free statistics, and no extrapolation can be made in good faith for periods longer than the test duration. In such an instance, accelerated life testing cannot be used to reduce the duration of the test. However, there is one problem to which accelerated testing can potentially be applied. Namely, under appropriate circumstances, the sample size required to choose the better of two classes of devices can be reduced.

For instance, assume that a light bulb is required for a function such that it is highly desirable that the light doesn't fail for at least one month. There is a choice of two classes of bulbs, say A and B. It is of interest to choose the better class, i.e. the one which has a smaller probability of failing within one month. However, both classes are so reliable that it is highly unlikely that any failure will be observed within one month, even in fairly large samples. Intuition (as well as statistical theory) tells us that such circumstances will make the selection of the better model quite problematical because, if no failures are observed, there is no information on which to base the choice. Several potential solutions of the dilemma might come to mind.

Even though the light is required for only one month, why not test the bulbs for a longer period of time, until enough failures are observed to make a choice possible? Since it was asserted, at the outset of this subsection, that no assumptions could be made regarding the failure distribution, this approach is not valid. Say that one class is selected on the basis of the number of failures observed during a year-long test. All one can say then is that the class selected has (with appropriate statistical confidence) the lesser probability of failure over a one-year period; but it does not follow that the same class has the lesser probability of failure over a one-month period! This is illustrated in Figure II-1 which shows two failure distributions; one of them is preferable in terms of one-month reliability, but the other one is preferable in terms of twelve-month reliability.

Another possible approach is to test, for one month, very large samples of bulbs. This approach is always feasible. The probability of observing some failures (and, hence, having a logical basis for decision) can be made arbitrarily large by sufficiently increasing the sample size. If the bulbs are sufficiently plentiful and sufficiently cheap, this approach is an adequate solution. However, economic considerations usually preclude this approach because the sample sizes required to discriminate between two small probabilities (of failure) are very large.

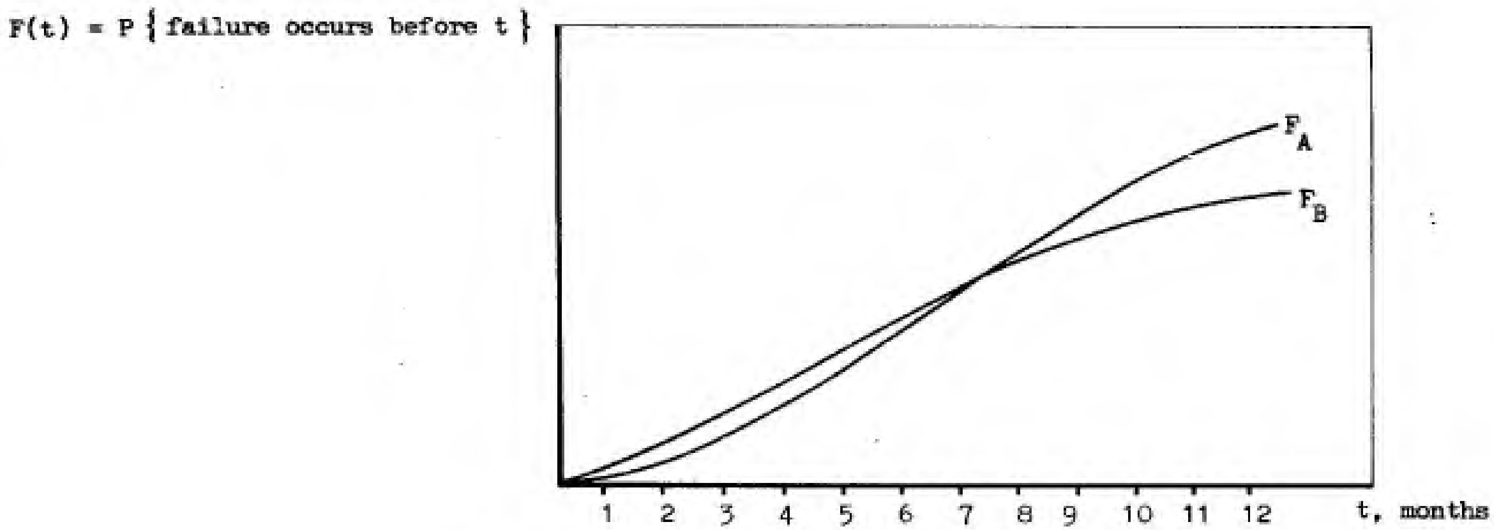


FIG. II-1. ILLUSTRATION OF EXAMPLE IN II.B.1

Faced with this quandary, one turns to accelerated testing. Indeed, the following approach is intuitively appealing. If the operating voltage of the light bulbs is increased, more failures can be expected to occur, and selection of the better model should be easier. Clearly, performing the test under such conditions fits the definition of accelerated life testing in the OST manner (see Subsection I.B.3).

Unless one knows (quantitatively) how the probability of failure in a one-month period varies with operating voltage, no precise statement can be made regarding the validity of a choice based on accelerated conditions. Nonetheless, a condition which clearly is sufficient for the validity of a choice based on accelerated conditions, is that

> whichever class of devices has the higher one-month reliability under normal operating conditions, also has the higher one-month reliability under accelerated conditions.

Note that the purpose of accelerated testing, in this instance, is not a reduction of test time; accelerated testing is done for economic reasons, in order to obtain more information, but without unduly increasing the sample size.

Two important classes of questions remain open for further investigations.

(i) Under what circumstances is the above condition likely to be satisfied? What is the nature of possible deviations from that condition? What are the consequences of such deviations in terms of effect on the statistical significance of the test?

Faced with this quandary, one turns to accelerated testing. Indeed, the following approach is intuitively appealing. If the operating voltage of the light bulbs is increased, more failures can be expected to occur, and selection of the better model should be easier. Clearly, performing the test under such conditions fits the definition of accelerated life testing in the OST manner (see Subsection I.B.3).

Unless one knows (quantitatively) how the probability of failure in a one-month period varies with operating voltage, no precise statement can be made regarding the validity of a choice based on accelerated conditions. Nonetheless, a condition which clearly is sufficient for the validity of a choice based on accelerated conditions, is that

> whichever class of devices has the higher one-month reliability under normal operating conditions, also has the higher one-month reliability under accelerated conditions.

Note that the purpose of accelerated testing, in this instance, is not a reduction of test time; accelerated testing is done for economic reasons, in order to obtain more information, but without unduly increasing the sample size.

Two important classes of questions remain open for further investigations.

(i) Under what circumstances is the above condition likely to be satisfied? What is the nature of possible deviations from that condition? What are the consequences of such deviations in terms of effect on the statistical significance of the test?

(ii) When the above condition is satisfied, what additional conditions must be imposed to guarantee that accelerated testing is indeed economical?

The second question is motivated by the fact that the sample size required to detect a difference in probabilities (of failure) depends on that difference as well as on the actual values of those probabilities. For instance, it is evident from some brief calculations by Hald([II-1], pp. 696 and 707), that accelerated testing would not be economical in the above example if both failure probabilities are increased (by increasing the voltage) in such a manner that their difference remains constant. However, if they are both increased in a manner such that their ratio remains constant, accelerated testing could indeed be used to reduce the sample size required for selecting the better class of bulbs.

II.B.2 Elementary Approach to Selection

II.B.2.1 Non-Sequential Method

If the form (e.g. gamma, or Weibull) of the failure distribution is known, statistical techniques exist (or can be devised) for the selection problem. Such techniques allow a selection to be made in arbitrarily short (expected) time provided the sample size is large enough.

For instance, assume that one is to choose one of two functionally equivalent products on the basis of reliability considerations. For purposes of illustration, assume that both products have an exponential distribution of time to failure. Then the following type of test can be used to select the product with the higher mean time to failure (which, under the exponential assumption, implies higher reliability for any interval).

> Place on test n items of type A and n items of type B. Test until r failures of type A or r failures of type B have been observed. Select the product which, at the end of the test, has not yet accumulated r failures.

The value of n in no ways affects the statistical properties of the test, and the value of r can be chosen so as to make the probability of wrong selection arbitrarily small. Figure II-2 shows approximate operating characteristic ("OC") curves for several values of r. The ordinate shows the probability of error, i.e., the probability of selecting the product with the smaller MTBF, as a function of the ratio of the larger MTBF (θ_{max}) to the smaller MTBF (θ_{min}). (The test described above is derived from a more general test discussed by Epstein [II-2], and data for Figure II-2 come from the same source).

As indicated above, the value of n does not affect the statistical properties of the test (i.e. the OC curve). However, a selection is made at a time T at which one of the products first accumulates r failures; hence, the duration of the test (i.e. T) is a random variable. The distribution of that random variable (with parameters θ_{max}, θ_{min}, r, and n) could be derived. For the current illustration, it is sufficient to consider the expected value of T. When testing is done without replacement, then the expected test duration E(T) satisfies the strict inequality

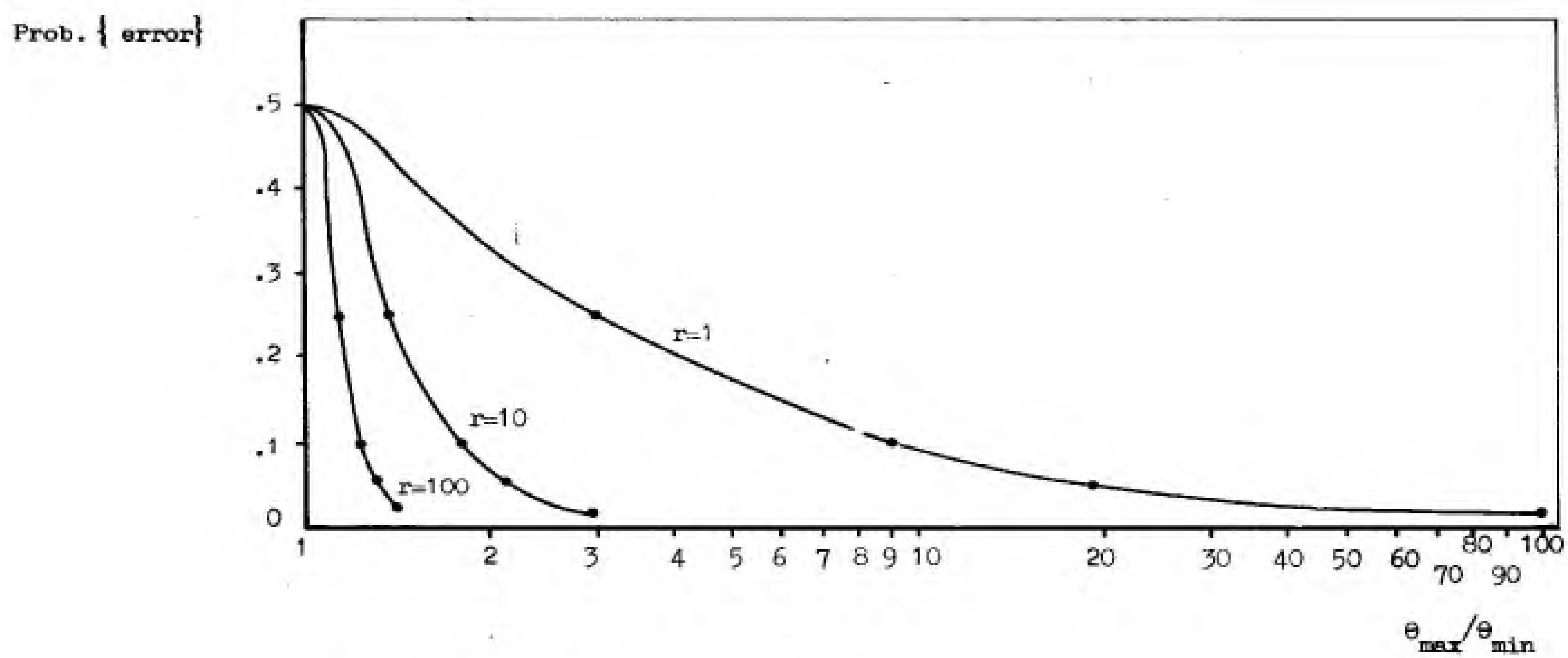


FIG. II-2. OPERATING CHARACTERISTIC CURVES FOR NON-SEQUENTIAL TEST

$$0 < E(T) < \theta_{max} \sum_{i=1}^{r} \frac{1}{n-i+1} .$$

(The upper bound follows from a relationship derived by Gumbel [II-3], and subsequently derived independently by Epstein and Sobel [II-4].) Clearly, for a fixed value of r (determined by the desired OC curve), E(T) can be made arbitrarily small by sufficiently increasing n; in fact, the upper bound for E(T) tends to zero as $n \to \infty$.

Intuitive consideration of the described test procedure may lead one to assert that E(T) satisfies a tighter inequality than the one stated above, viz.

$$\theta_{min} \sum_{i=1}^{r} \frac{1}{n-i+1} \leq E(T) \leq \theta_{max} \sum_{i=1}^{r} \frac{1}{n-i+1}$$

Unfortunately, this assertion is incorrect. For proof, it suffices to consider the test in which r = n = 1, and consider the special case in which $\theta_{min} = \theta_{max} = \theta$. Then the last-stated inequality implies $E(T) = \theta$ whereas, in fact, $E(T) = \theta/2$ in that case. The result $E(T) = \theta/2$ follows from the fact that the test terminates as soon as either product exhibits a failure, and the failures are initially generated by a Poisson process with rate $2(1/\theta)$.

II.B.2.2 Sequential Method

The same problem in selection can also be handled by a sequential procedure which requires fewer observations, on the average, to reach a decision than does the above non-sequential procedure. The sequential procedure is described and analyzed by Epstein [II-5].

Let samples of size n be taken from each class of devices and let life testing be carried out simultaneously (and continuously in time) on both classes. The sample size n is kept the same throughout the test, by replacing immediately any item which fails by a new item from the same class. Let $X_i(t)$, i = A or B, be the number of failures of items of class i by time t. Let us adopt the following rule:

experimentation is continued as long as

$$-a < X_A(t) - X_B(t) < a \qquad \text{(II.B.2.2-1)}$$

where a is a positive integer specified in Eq. (II.B.2.2-2) below. As soon as $X_A(t)$ and $X_B(t)$ are such that inequality (II.B.2.2-1) is not satisfied, stop testing. If, at the time of stopping, $X_A(t) - X_B(t) = a$, select product B as having the higher mean life. If, at the time of stopping, $X_A(t) - X_B(t) = -a$, select product A as having the higher mean life.

Denote by u the ratio $\theta_{max}/\theta_{min}$. Select a value of u, say u_o, different from unity. Impose the restriction that when $u = u_o$, the probability π_{u_o} of making an incorrect selection is $\leq \alpha$. If this restriction is imposed, it can be shown that a is the unique integer meeting the condition

$$\frac{\log\left(\frac{1-\alpha}{\alpha}\right)}{\log u_o} \leq a \leq \frac{\log\left(\frac{1-\alpha}{\alpha}\right)}{\log u_o} + 1 \quad . \qquad \text{(II.B.2.2-2)}$$

π_u, the probability of making an incorrect assertion, is given by

$$\pi_u = \frac{1}{u^a + 1} \quad , u \geq 1. \qquad \text{(II.B.2.2-3)}$$

Since the selection procedure is sequential, the number of failures observed prior to making a selection is a random variable, say Z. The expected number of items failed in the course of reaching a decision depends on the actual value of u, and is given by

$$E_u(Z) = a\left(\frac{u+1}{u-1}\right)(1 - 2\pi_u), \text{ if } u > 1$$

and by (II.B.2.2-4)

$$E_u(Z) = a^2 \quad , \text{if } u = 1 \quad .$$

In Table II-1 we give $E_1(Z)$ and $E_{u_o}(Z)$ for α = .01, .05, and u_o = 3/2, 2, 5/2, 3.

It can be shown that E(T), the expected waiting time to reach a decision is given by

TABLE II-1. VALUES OF $(E_{u_o}(Z), E_1(Z))$ FOR VARIOUS VALUES OF α AND u_o.

α \ u_o	3/2	2	5/2	3
.01	(59.1, 144)	(20.7, 49)	(13.9, 36)	(9.92, 25)
.05	(37.0, 64)	(14.1, 25)	(8.88, 16)	(5.57, 9)

$$E(T) = \frac{E_u(Z)}{n} \bigg/ \left(\frac{1}{\theta_{max}} + \frac{1}{\theta_{min}} \right)$$

where $E_u(Z)$ is given by Eq. (II.B.2.2-4).

For this sequential procedure, the dependence of E(T) on n is even clearer than in the non-sequential procedure, since E(T) is inversely proportional to n. Thus, for the sequential procedure as well as for the non-sequential one, a decision can be reached by a test whose expected duration can be made arbitrarily short by sufficiently increasing n.

The above examples illustrate (though they don't prove) our assertion that, when the form of the failure distribution is assumed, a selection test can be performed in arbitrarily short (expected) time - provided the sample size is sufficiently large. Note that "sample size" refers to the number of specimens initially placed on test and not to the number of specimens actually consumed by failures.

II.B.2.3 Accelerated Testing

Now assume that, for each product, an accelerated test can be pe_formed by changing some appropriate environmental variable, such as the operating voltage in the case of capacitors. When the environmental variable is at level S, let $\theta_A(S)$ and $\theta_B(S)$ be the MTBFs for products A and B. The elementary approach to accelerated testing is as follows:

Apply one of the previously described selection procedures to data obtained under accelerated conditions (i.e. at a stress level more severe than the usage level); whichever product has been selected in this way is also considered to be better at usage conditions.

Since the OC curve (for either the sequential or the non-sequential procedure) depends only on the ratio $\theta_{max}/\theta_{min}$, and not on the actual value of either MTBF, the OC curve is not changed if both θ_A and θ_B are increased, provided their ratio remains constant. Thus, the elementary approach to selection under accelerated conditions is perfectly correct when the ratio $\theta_A(S)/\theta_B(S)$ is constant in S. That is, if the problem at hand is one of selection, one needn't assume values of K-factors. One only need assume that, with respect to the increased environment used in the test, both classes of devices have the same K-factor.

Thus, assuming the constancy of the ratio $\theta_A(S)/\theta_B(S)$ is sufficient to allow the reduction of sample size by means of accelerated testing. This reduction of sample size comes about as follows: Both θ_{max} and θ_{min} are decreased by increasing the environmental variable. However, as was already demonstrated (under the exponentiality assumption), E(T) can be made arbitrarily small without changing the environment. Thus it would not be correct to state that a change in the environment is used to reduce the duration of the test. But, for a given (arbitrary) expected test duration, the required sample size can be reduced by accelerated testing if the appropriate assumption can be justified.

Note well the assumption required for the correct application of ALT to the selection problem. It states that $\theta_A(S)/\theta_B(S)$ is constant in S. In other words, it requires that the value of the acceleration function (see Subsection I.C.2 for definition) for each product be the same for each value of stress. In fact, the procedure is applicable if the requirement is reduced to the minimum, namely that the value of the acceleration function for each product be the same only for the specific environment at which accelerated testing is to be performed. However, no assumption is made regarding either the form of the acceleration function or regarding the values of its parameters. As will be seen in Subsections II.C.1 and II.D.2 this assumption is much weaker than the assumption required for the application of ALT to estimation or decision.

Note that the assumption required in this case is formally identical with the assumption required in accelerated lot acceptance (see Subsection II.D.3), which is formally a decision procedure. Nonetheless, the selection process appears to be more robust to minor errors in the assumption than the accelerated lot acceptance procedure. This question of robustness is discussed in Subsection II.B.2.4 below.

II.B.2.4 Error Analysis

We now consider the possible errors which can be induced by errors in the assumptions required for the application of ALT to the selection problem. The discussion remains illustrative in nature, and is based on the example established at the outset of this discussion of selection via accelerated testing. Namely, it is taken for granted that both classes of devices have the exponential failure distribution, and it is desired to select the one with the larger mean time to failure.

Let S_u be the usage level of stress, and consider the possible errors arising from performing an accelerated test at some stress S_a. Denote $\theta_A(S_u)$ by θ_A^u and $\theta_A(S_a)$ by θ_A^a, and similarly for θ_B^u and θ_B^a. Then there exist constants, say k_A and k_B, such that

$$\theta_A^a = k_A\theta_A^u$$

$$\theta_B^a = k_B\theta_A^u$$

The previously described application of ALT is predicated on the assumption that $k_A = k_B$. If one formally derives the OC curve which results when $k_A/k_B \neq 1$, i.e. when the above assumption is incorrect, one is left with the impression that very serious errors can occur. Figure II-3 shows the nominal OC (replotted from Figure II-2 but in terms of θ_A/θ_B instead of in terms of $\theta_{max}/\theta_{min}$), and the actual OCs which result if $k_A/k_B = 1.25$ or .80. The new curves are derived as follows:

Let k be the true value of the ratio k_A/k_B. Let $h(\rho)$ be the ordinate of the nominal OC (middle curve in Figure II-3) when the abscissa has value ρ. Recall that selection is based on observed failures from populations with means θ_A^a and θ_B^a, and note that

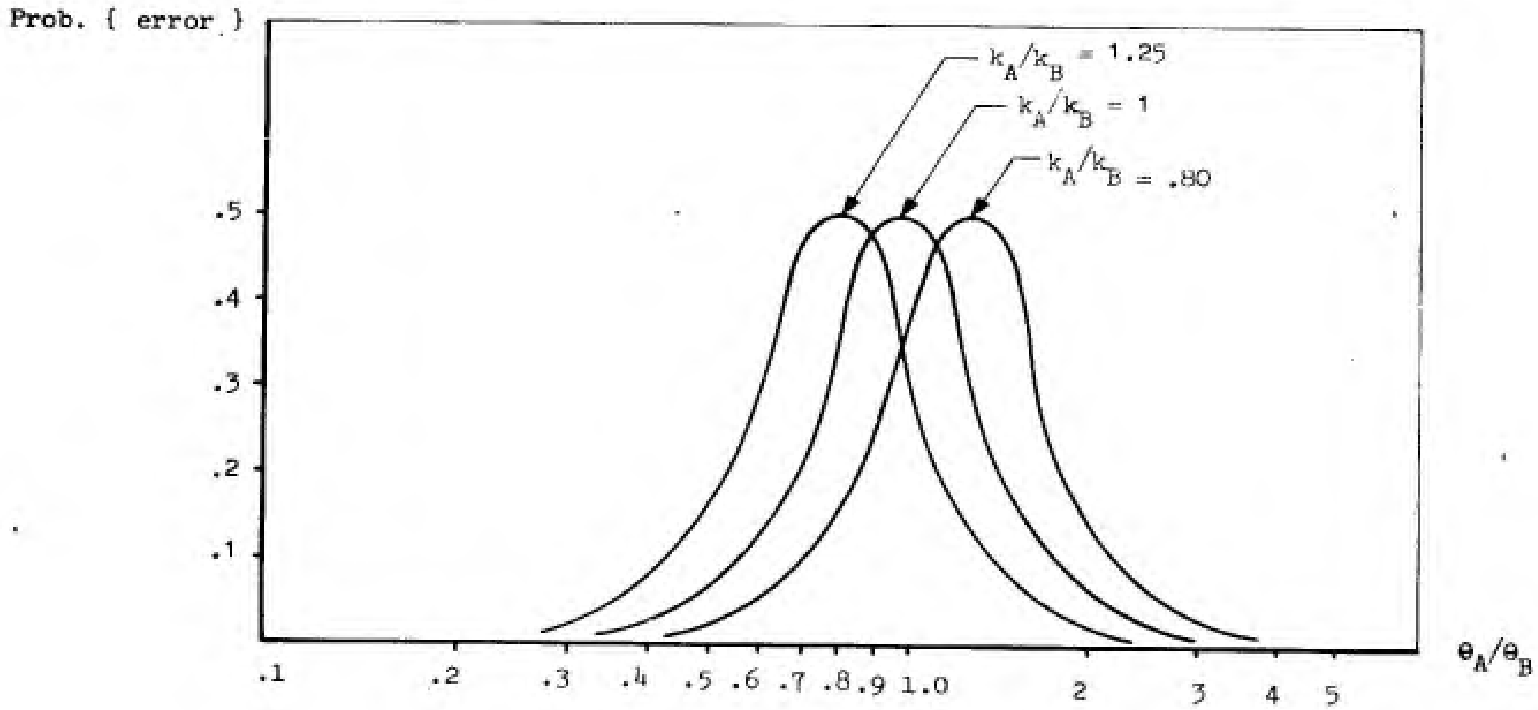


FIG. II-3. OC CURVES RESULTING FROM ERROR IN ASSUMPTIONS

$$\frac{\theta_A^u}{\theta_B^u} = \rho \quad \text{iff} \quad \rho = \frac{\theta_A^a/k_A}{\theta_B^a/k_B} = \frac{1}{k} \; \frac{\theta_A^a}{\theta_B^a} \quad ,$$

$$\theta_A^u/\theta_B^u = \rho \qquad \text{iff} \quad \theta_A^a/\theta_B^a = k\rho \quad .$$

Therefore

$$P\left\{\text{error} \;\middle|\; \theta_A^u/\theta_B^u = \rho\right\}$$

$$= P\left\{\text{error} \;\middle|\; \theta_A^a/\theta_B^a = k\rho\right\}$$

$$= h(k\rho) \quad .$$

Examination of the OC curves in Figure II-3 leads one to believe that gross errors can result from moderate discrepancies in the acceleration factors for the two classes of devices being compared. This impression is, however, mistaken. It results from the fact that, in the example, the ratio k_A/k_B is independent of the value of θ_A^u/θ_B^u. A more precise analysis of the possible errors should take account of the fact that two products with nearly equal usage MTBFs are more likely to have the same acceleration factor than are two products with very different usage MTBFs. This would then modify the results shown in Figure II-3, by preserving some discrepancies in the tails of the OC curves but by almost eliminating discrepancies near the center (i.e. near $\theta_A/\theta_B = 1$). Hence, possible errors in the underlying assumption $k_A/k_B = 1$ would have little operational effect.

There is another possible form of dependence of k_A/k_B on θ_A^u/θ_B^u. It is very likely that, in practice, the better product (i.e. the one with the higher operational MTBF) is also less sensitive to variations in environment. Formally, this means that

$$\theta_A^u < \theta_B^u \quad \text{iff} \quad k_A < k_B .$$

In that case, it can be shown than an error in the assumption on which the accelerated test is based (i.e. $k_A/k_B = 1$) leads to a test which is actually better than one believes. That is, the OC curve which actually describes the test in that case is uniformly below the OC curve corresponding to $k_A = k_B$.

The remarks in the preceding paragraph indicate that selection based on accelerated tests may be fairly robust with respect to errors in the assumptions regarding the nature of change in failure behavior with changes in environment. Thus it appears plausible that accelerated selection may be a practical endeavor. Therefore it is recommended that future effort be devoted to a more detailed discussion of this question of robustness.

The development of valid selection procedures is very desirable. They do not yield actual estimates of the failure characteristics of the device under consideration, but they provide a means for selecting the best available device. In practical terms, selection of the most reliable device should always be the prime consideration (provided various system and cost constraints are met). Actual estimation and/or demonstration of reliability should then be applied to the class of devices already selected as being the best. This philosophy eliminates the waste of performing estimation and/or demonstration tests on classes of devices which are not as good as some other class of devices. Note in passing that the customary selection procedure in current reliability practice is to perform selection by comparing estimates of the reliability of several classes of devices. In terms of statistical significance, such a procedure is wrong or, at best, inefficient.

II.C ESTIMATION

II.C.1 Confidence Interval Approach

If the form (e.g. gamma, or Weibull) of the failure distribution is known, statistical techniques exist (or can be devised) for the estimation problem. Such techniques enable one to make estimates based on data resulting from tests of arbitrarily short (expected) duration. However, in order to get nontrivial results, large sample sizes are required when the tests are of short duration.

For instance, suppose that one wishes to estimate the mean time to failure (MTBF) θ of a device which is assumed to have exponentially distributed time to failure. The confidence limits for θ can be expressed as the product of the maximum likelihood estimate $\hat{\theta}$ and an appropriate multiplier. Tables of that multiplier are given in [II-6] for a time-truncated test, i.e. a test which is stopped after a predetermined amount of cumulative operating time has been observed. Of course, for such a test, the number of failures observed is a random variable and the multiplier depends on the number of failures actually observed. When the number of failures is large (e.g. near 200) the multipliers are practically constant. In fact, for a time truncated test, and with the multiplier rounded to one significant figure, the 90 percent lower confidence limit (one-sided) for θ is $.9\hat{\theta}$ when the observed number of failures is between 70 and 750 (see [II-6]).

The observations of the last paragraph allow one to determine beforehand the approximate precision of the estimate, provided one can make a reasonable guess of the true value of MTBF. If one wishes the 90 percent lower confidence limit for θ to be approximately $.9\hat{\theta}$, one might plan to perform the life test so that the expected number of failures is 300. If θ_o is an a priori guess of the MTBF, the truncation point for the test should be chosen as follows: if T is the cumulative operating time at which the test is truncated, then the expected number of failure is T/θ; therefore T should satisfy $T/\theta_o = 300$. Even if θ_o is in error by as much as a factor of 2, T/θ remains between 150 and 600 so that one can be fairly certain that the desired precision will be achieved by the test.

Say then, that one plans to run the test so that the expected number of failures is 300. If θ_o is an a priori guess of the MTBF, the test should be truncated when the total observed operating time T is such that $T/\theta_o = 300$. Assuming that one tests with immediate replacement, and that n items are originally put on test, the actual (calendar) duration of the test is T/n. Clearly, this can be made arbitrarily short by sufficiently increasing n. Say that a test duration of t hours is considered practical. Then the size of the initial sample should be $n = T/t = 300\ \theta_o/t$. Thus it is again illustrated that, if the form of the distribution is assumed, a test (leading to estimation, in this case) can be performed in arbitrarily short time. The test is reduced to an acceptably short duration by picking a correspondingly large sample size. Under appropriate assumptions, this sample size can be reduced by accelerated testing.

In the elementary approach to ALT, it is assumed that the value of the acceleration function (i.e. K-factor) is known for the particular environment to be used for accelerated testing. Recalling that θ_o is the best a priori guess for use conditions MTBF, the a priori guess under accelerated conditions is then $\theta_o^* = \theta_o/K < \theta_o$. Then the initial sample size for the accelerated test should be $n^* = 300\ \theta_o^*/t < n$, by the same argument as above.

In the elementary approach, standard statistical techniques (see [II-7]) are applied to the data resulting from the accelerated test and the results are then translated to equivalent use-conditions results.

For example, confidence limits established for the mean time to failure at accelerated conditions, multiplied by the K-factor, then yield the confidence limits for mean time to failure at use conditions.

In the example discussed above, both the normal and the accelerated test "consume" approximately three hundred samples (since that is the expected number of failures). However, if the acceptable test duration (t) is less than the guessed MTBF (θ_o) -- and this is certainly the case when testing very reliable items -- then the initial sample size is larger than three hundred. That is, the number of samples that must be put on test is greater than the number of samples that must be "sacrificed" by failing.

Since the suggested initial sample size is proportional to the a priori guess of the MTBF value (θ_o^*), this initial sample size can be appropriately reduced by testing under an environment which yields a sufficiently low θ_o^*.

II.C.2 Bounded Relative Error Approach

Another approach to the estimation problem is possible. That approach, known as bounded relative error estimation, differs from the Neyman (i.e. confidence interval) approach in that the precision of the estimate is fixed beforehand. In confidence interval estimation, the length of the interval depends on the outcome of the experiment. On the other hand, the length of the interval is fixed beforehand in the bounded relative error approach.

Consider a life-test of a device with exponentially distributed time to failure, with mean θ. Suppose that a predetermined number, r, of failures are observed and let the associated total life be T_r. Then the maximum likelihood estimator of θ, i.e. $\hat{\theta} = T_r/r$, is also a bounded relative error estimator of θ, as shown by Epstein [II-8]. In fact,

$$Pr\left\{\left|\frac{\hat{\theta} - \theta}{\theta}\right| \leq \delta\right\} = 1 - \alpha$$

for appropriate values of r (which are derived in [II-8] and are shown in Table II-2).

The precision (i.e. δ) and the confidence (i.e. $1 - \alpha$) with which $\hat{\theta}$ achieves that precision, depend only on the predetermined number of failures r at which the test is stopped.

TABLE II-2. VALUES OF r SUCH THAT $Pr\{|\frac{\hat{\theta} - \theta}{\theta}| \leq \delta\} = 1 - \alpha$

δ \ α	.01	.05	.10
.01	66,400	38,400	27,100
.05	2,654	1,537	1,082
.10	664	384	271

Now, the expected duration of the test is $E(T_r)$, which can be made arbitrarily short by using a sufficiently large sample. Specifically, if one promptly replaces failed items, so that n of them are on test at all times, the occurrence of failures is a Poisson process with rate n/θ. Then $E(T_r \mid n) = \theta r/n$. Now one can apply the argument which was used above in connection with confidence-interval estimation. Based on an a priori guess of θ, say θ_o, one determines n so that $E(T_r)$ is acceptably small. Then one uses accelerated testing to decrease θ, thus allowing a corresponding decrease in n without affecting the already satisfactory value of $E(T_r)$.

II.C.3 Error Analysis

The above discussion of the possibil. ,y of reducing sample size by accelerated testing was oriented to the approach in which the dependence of failure behavior on environment is assumed to be known in detail. In this subsection, an example is given of the kind of error this approach can introduce.

The example will be carried out in relation to the bounded relative error approach. In the confidence interval approach, an error in the acceleration factor would cause a shift in the confidence limits; such a shift would not show very clearly the effect of the error on the precision of the estimate. On the other hand, in the bounded relative error approach, the precision of the estimate is evident at a glance, as is the modified precision resulting from an error in acceleration factor. Note in passing that no detailed error analysis need be carried out for so-called point estimates. If $\hat{\theta}_a$ is the maximum likelihood estimator (MLE) of θ_a, then the elementary approach dictates the use of $K\hat{\theta}_a$ for the MLE $\hat{\theta}_u$ of θ_u. Since $\hat{\theta}_a$ is an unbiased estimator of θ_a, any error in acceleration factor would cause $K\hat{\theta}_a$ to be a biased estimator of θ_u.

Assume that the MTBF of a device with exponential failure distribution is to be estimated. Suppose further that the MTBF (θ) is related to some environmental variable (denoted by S) by the relation

$$\theta(S) = CS^b \qquad \text{(II.C.3-1)}$$

where the constants C and b are assumed known. Let S_u be the so-called operating (or use conditions) environment and S_a the accelerated environment and, for brevity, let $\theta_u = \theta(S_u)$ and $\theta_a = \theta(S_a)$. Assume, for example, that θ_a will be estimated by the bounded relative error method. Then, in the notation used previously,

$$\Pr\left\{1 - \delta \le \frac{\hat{\theta}_a}{\theta_a} \le 1 + \delta\right\} = 1 - \alpha \qquad \text{(II.C.3-2)}$$

But the real interest is in estimating θ_u. Since

$$\theta_u = \theta_a \frac{CS_u^b}{CS_a^b} = \theta_a \left(\frac{S_u}{S_a}\right)^b$$

one would obviously use the estimator

$$\hat{\theta}_u = \hat{\theta}_a \left(\frac{S_u}{S_a} \right)^b \qquad \text{II.C.3-3)}$$

If the functional relation is indeed correct, and the value of the exponent is known exactly (note that C "drops out"), then one can state that

$$\Pr \left\{ 1 - \delta \le \frac{\hat{\theta}_u \left(\frac{S_a}{S_u} \right)^b}{\theta_u \left(\frac{S_a}{S_u} \right)^b} \le 1 + \delta \right\} = 1 - \alpha \quad .$$

That is, if the value of the exponent is known exactly, then the estimator of θ_u specified by Eq. (II.C.3-3) is as precise as the estimator of θ_a; i.e., no precision is lost by testing at accelerated conditions.

However, even if the functional relationship is indeed correct in form, the exponent may not be known correctly. That is, one may believe that the correct value of the exponent in Eq. (II.C.3-1) is b when, in fact, it is b + ε. Then estimation is based on the relationship

$$\hat{\theta}_u = \hat{\theta}_a \left(\frac{S_u}{S_a} \right)^b$$

whereas, in fact,

$$\theta_u = \theta_a \left(\frac{S_u}{S_a}\right)^b \left(\frac{S_u}{S_a}\right)^{\epsilon} .$$

and, therefore, the estimation procedure yields an estimate of θ_u such that

$$\Pr\left\{1 - \delta \le \frac{\hat{\theta}_u \left(\frac{S_a}{S_u}\right)^b}{\theta_u \left(\frac{S_u}{S_a}\right)^b \left(\frac{S_a}{S_u}\right)^{\epsilon}} = 1 + \delta\right\} = 1 - \alpha ,$$

$$\Pr\left\{(1 - \delta)\left(\frac{S_a}{S_u}\right)^{\epsilon} \le \frac{\hat{\theta}_u}{\theta_u} \le (1 + \delta)\left(\frac{S_a}{S_u}\right)^{\epsilon}\right\} = 1 - \alpha .$$

It is now clear that one possible effect of using a slightly erroneous value of b is a change in the precision of the estimate $\hat{\theta}_u$. To illustrate the possible magnitudes of error, consider the following example.

For capacitors, it is frequently assumed that the failure distribution is exponential and that it depends on the applied voltage in the manner indicated in Eq. (II.C.3-1). According to Levenbach [II-9], the value of b is approximately -5, and a deviation of 10 percent in that value would not be surprising. Assume that an accelerated test with b = -5 is performed with a voltage which would cause a tenfold reduction in MTBF, i.e. $S_a/S_u = 1.6$. If the estimate is based on the assumed value b = -5 and, in fact, b = -5.5 then

$$\Pr\left\{(1-\delta)(1.6)^{-.5} \le \frac{\hat{\theta}_u}{\theta_u} \le (1+\delta)(1.6)^{-.5}\right\} = 1-\alpha \quad .$$

In bounded relative error estimation, a value of 0.1 for δ is not unreasonable. With $\delta = .1$, the above statement becomes

$$\Pr\left\{.71 \le \frac{\hat{\theta}_u}{\theta_u} \le .87\right\} = 1-\alpha \quad .$$

Thus one uses $\hat{\theta}_u$ believing that, with confidence $1-\alpha$ (say .9), it is within 10 percent of the true value θ_u when, in fact, it almost certainly grossly underestimates θ_u.

Similarly, if b in fact is -4.5, then it follows that

$$\Pr\left\{ 1.14 \le \frac{\hat{\theta}_u}{\theta_u} \le 1.39 \right\} = 1 - \alpha ,$$

so that $\hat{\theta}_u$ is almost certain to be a gross overestimate of θ_u.

Somewhat more generally, if the relationship between failure behavior and environment is assumed to be known, say $\theta(S) = f(S)$ for an exponential failure law, then this assumption implies that

$$\theta_a = \theta_u / k$$

where

$$k = f(S_u)/f(S_a)$$

Let us call k the assumed acceleration factor.

Now, it is possible that the assumed acceleration factor is in error. Specifically let K be the correct value of the acceleration factor. Then, since $\hat{\theta}_u$ is derived from $\hat{\theta}_a$ by means of the assumed acceleration factor,

$$\frac{\hat{\theta}_u}{\theta_u} = \frac{\hat{\theta}_a k}{\theta_a K} = \frac{k}{K} \frac{\hat{\theta}_a}{\theta_a} .$$

If $\hat{\theta}_a$ is a bounded relative error estimate, then the precision of $\hat{\theta}_u$ is obvious from the relation

$$\Pr\left\{(1-\delta)\frac{k}{K} \le \frac{\hat{\theta}_u}{\theta_u} \le (1+\delta)\frac{k}{K}\right\} = 1-\alpha \tag{II.C.3-4}$$

which follows from Eq. (II.C.3-2).

It can be seen from Eq. (II.C.3-4) that if the relative error in k is of magnitude comparable to δ, then $\hat{\theta}_u$ is almost certain to either underestimate or overestimate θ_u. Specifically, let $\delta = .1$ and assume that $K = (1.1)(k)$. Then it follows from Eq. (II.C.3-4) that

$$\Pr\left\{.82 \le \frac{\hat{\theta}_u}{\theta_u} \le 1\right\} = 1-\alpha \quad .$$

Similarly, if $K = (.9)(k)$, then

$$\Pr\left\{1 \le \frac{\hat{\theta}_u}{\theta_u} \le 1.22\right\} = 1-\alpha \quad .$$

Thus, when an estimate is desired which is within 10 percent of the correct value (with confidence $1-\alpha$), then a 10 percent error in k yields an estimate which almost certainly (i.e. with confidence $1-\alpha$) is biased.

The illustrative calculation above is indicative of the gross errors that can arise when estimation from accelerated tests is done by the elementary method. Some variations on the elementary approach to estimation can be found in the work of Levenbach [II-9] and of Allen [II-10] and [II-11]. Their results are discussed in Section II.E.

Finally, note that the assumption underlying the elementary approach to accelerated estimation is much stronger than the assumption required for accelerated selection. Here an actual value of the K-factor must be known whereas, in the case of selection, this was not necessary.

II.D DECISION

II.D.1 An Acceptance Test

This subsection deals with an example of so-called acceptance (or demonstration) testing. Such a test is a decision procedure, i.e. it specifies a test, and an algorithm whose inputs are the test results and whose output is a decision, either to accept the class of devices tested or not to accept it. The statistical properties of such a procedure are described by the operating characteristic (OC), a function which gives the probability of acceptance for various possible states of nature.

To fix the ideas, assume that the test is to be applied to a class of devices with exponential failure distribution. Specifically, consider the test designated as C-11 in [II-12]. The test is intended for testing the hypothesis that the mean time to failure of the device at hand is no less than some imposed criterion. The test has the operating characteristic shown in Figure II-4. In that figure, θ is the actual MTBF and M is the imposed criterion, i.e. the desirable MTBF. The procedure embodies a consumer's risk of 0.10 at an MTBF which is approximately half the desired MTBF, and a producer's risk of 0.10 at the specified MTBF. The test procedure is as follows: Place n items on test; replace failed items promptly; terminate the test after elapsed time T (i.e. total accumulated test time nT); accept the class of devices tested if the number of failures observed by time T is less than 15,

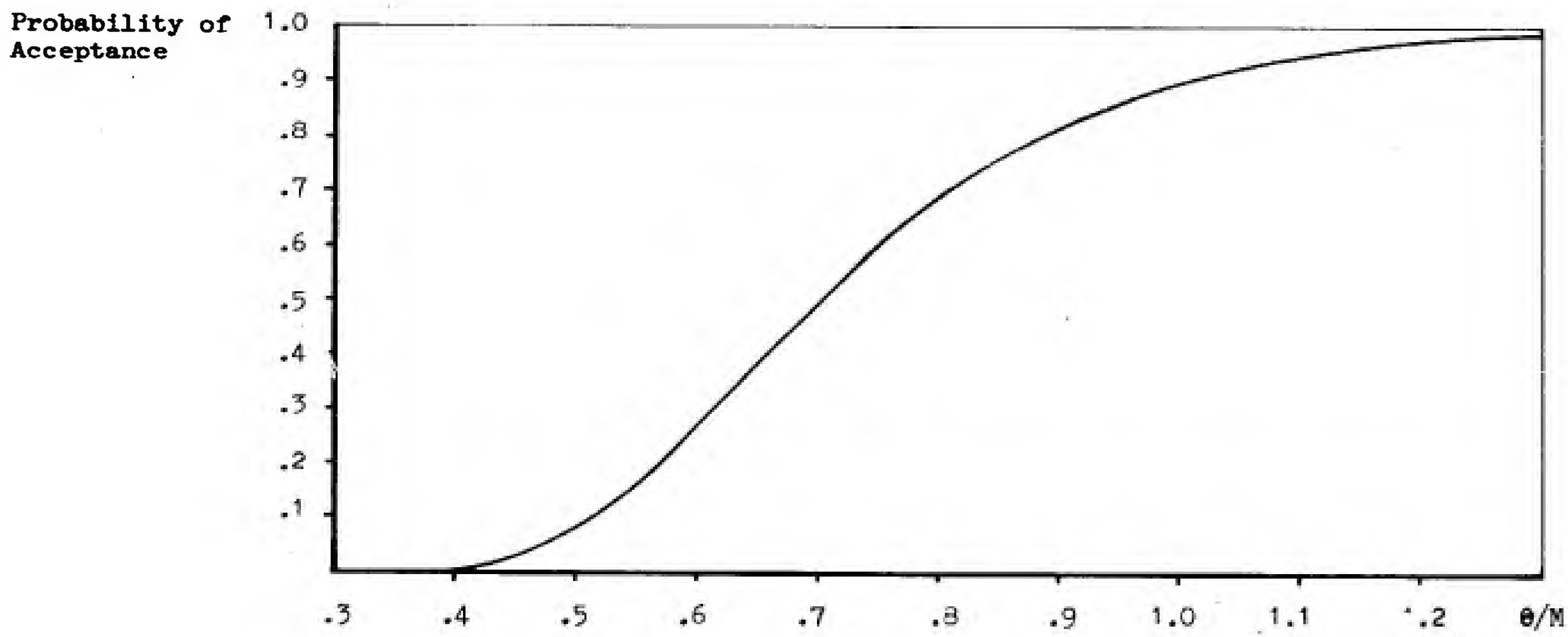


FIG. II-4. OC CURVE FOR ACCEPTANCE TEST C-11 reject otherwise. The choice of T and n is arbitrary, provided they are related as shown in Figure II-5 (see [II-12] page 2.48). That is, an appropriate sample size and termination time are selected from Figure II-5.

Three remarks are in order. First, note that "accepting the class of devices tested" corresponds to accepting the hypothesis $\theta \geq M$. Accepting (or rejecting) the devices is the operational, i.e. physical, outcome of the test procedure. Accepting (or rejecting) the hypothesis $\theta \geq M$ is the formal equivalent of that operational outcome. The two are related by the fact that M is the desired MTBF. Secondly, note from Figure II-5 that this test procedure can be applied in arbitrarily short time T by sufficiently increasing the initial sample size n. Finally, it should be noted that this procedure was obtained from [II-12] but that, in fact, all the procedures in that reference are based on the work of Epstein [II-13].

II.D.2 Accelerated Test and Error Analysis

Assume that the devices at hand can be subjected to an "accelerated" environment. Continuing with the example of the last subsection, let θ_u be the MTBF at use conditions and θ_a the MTBF at accelerated conditions. In the elementary approach to ALT, the K-factor relating the two MTBFs is assumed known, say

$$\theta_a = \theta_u/k \quad .$$

Clearly, if the assumed value of k is correct, the test described in the last subsection can be applied directly to data from the accelerated test;

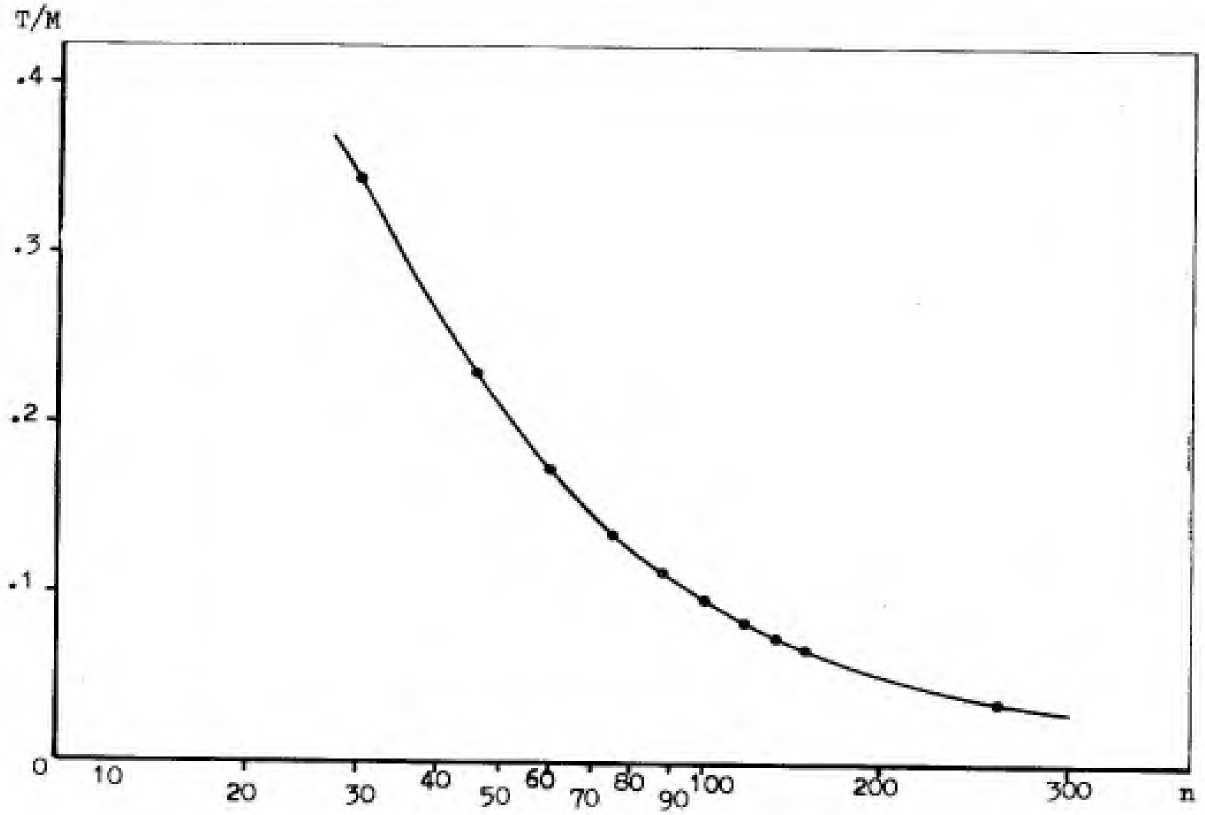


FIG. II-5. RELATION BETWEEN TEST DURATION AND SAMPLE SIZE throughout, use $\theta_a = \theta_u/k$ for θ and, if M_u is the desired use-conditions MTBF, use $M_a = M_u/k$ for M. That is, select an appropriate sample size and termination time from Figure II-5, using M_a for M. Accept or reject the hypothesis $\theta_a \geq M_a$ according as the observed number of failures is less than or greater than the acceptance number (15, in this case.)

The OC for the test of the hypothesis $\theta_a \geq M_a$ is that shown in Figure II-4. If k is indeed the correct accelerating factor, then this hypothesis is equivalent to $\theta_u \geq M_u$, and the same OC curve corresponds to the latter hypothesis. If the correct accelerating factor is K, then the OC for the hypothesis $\theta_u = M_u$ is not that of Figure II-4. The correct OC can be derived as follows:

Let $h(\rho)$ be the ordinate of the OC curve of Figure II-4 when the abscissa is ρ. Then

$$\Pr(\text{accept hypothesis } \theta_a \geq M_a \mid \theta_a/M_a = \rho) = h(\rho).$$

In the belief that k is the correct value of the accelerating factor, one accepts $\theta_u \geq M_u$ whenever one accepts $\theta_a \geq M_a$. Therefore

$$\Pr(\text{accept hypothesis } \theta_u \geq M_u \mid \theta_a/M_a = \rho) = h(\rho).$$

But

$$\frac{\theta_a}{M_a} = \frac{\theta_u/K}{M_u/k} = \frac{k}{K}\,\frac{\theta_u}{M_u} \qquad ;$$

therefore

$$\Pr(\text{accept hypothesis } \theta_u \geq M_u \mid \frac{\theta_u}{M_u} = \rho \frac{K}{k}) = h(\rho).$$

Let

$$\xi = \rho \frac{K}{k} \qquad .$$

Then

$$\Pr(\text{accept hypothesis } \theta_u \geq M_u \mid \frac{\theta_u}{M_u} = \xi) = h(\xi \frac{k}{K}) \quad .$$

Thus the correct OC is such that its ordinate, corresponding to the abscissa-value ξ, is the ordinate of the original OC corresponding to the abcissa-value $\xi k/K$. Figure II-6 shows the original OC, together with the correct OCs resulting when $K = .9k$ or $K = 1.1k$.

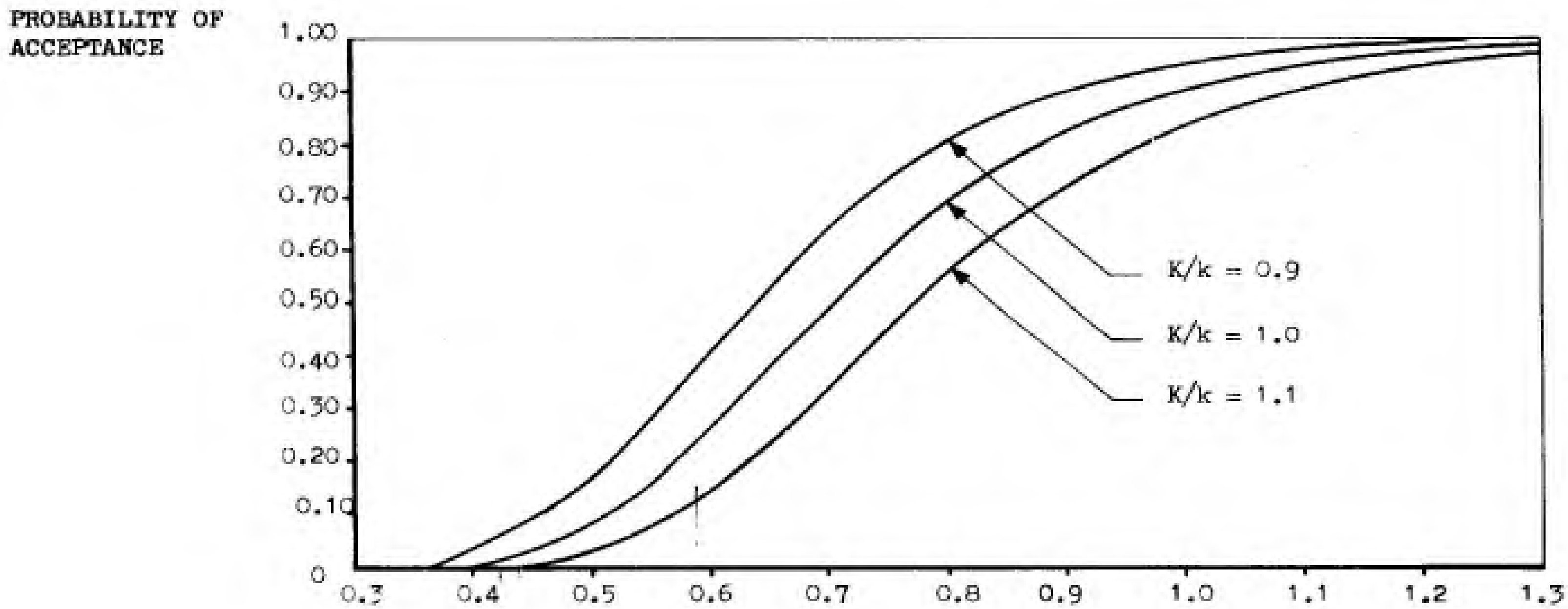


FIG. II-6. ACTUAL OC'S WITH WRONG K-FACTOR

Examination of the correct OCs for the two errors considered shows the disastrous effect that a 10 percent error in the acceleration factor can have. In one case ($K/k = 0.9$), the consumer's risk at $\theta/M = .5$ is approximately doubled. Ignoring the error in the acceleration factor leads the customer to believe that the probability of accepting a product with θ such that $\theta = M/2$ is only .10 when, in fact, it is .20. In the other case ($K/k = 1.1$), it is the producer's risk that is approximately doubled. Ignoring the error in the acceleration factor leads the producer to believe that the probability of rejection of a satisfactory lot (i.e. such that $\theta = M$) is only .10 when, in fact, it is .20. Since one cannot know beforehand whether the postulated acceleration factor is too high or too low, the elementary approach to accelerated acceptance testing is potentially very harmful to either party.

II.D.3 An Accelerated Lot Acceptance Test

The following modification of the elementary approach to accelerated acceptance testing presents a noticeable improvement. As in the elementary approach, the failure distribution is assumed to be exponential at all environments of interest. However, instead of assuming a value for the acceleration factor, one uses the data from a pilot experiment. Tests are run on a control lot both at use and at accelerated conditions; later an accelerated test is run on a subsequent lot; data from all three tests are used in making inferences about the reliability of the latter lot at use conditions. The procedure is based on the assumption that the acceleration factors for the control lot and for the subsequent production lot are identical.

Let E_a and E_u denote the accelerated and use environments repectively. Denote by λ_a and λ_u the failure rates (i.e. inverse of mean time to failure) at accelerated and use conditions respectively for the control lot. Similarly, let λ_a^* and λ_u^* pertain to the subsequent production lot.

Let the control test have sample sizes n_a and n_u, and accumulated test times T_a and T_u, under conditions E_a and E_u respectively. Let the production lot be tested only under E_a, with sample size n and accumulated test time T. If r failures occur in this last test, with times to failure $t_1, t_2, \ldots, t_r$, then

$$T = \sum_{j=1}^{r} t_j + (n - r)t_r$$

when testing is without replacement and all items are put on test simultaneously and the test is stopped at the r- th failure. Similarly for T_a and T_u, using r_a and r_u to denote the number of failures occurring in their respective tests.

It is well known (see [II-14]) that

$$2\lambda_u T_u \text{ is distributed as } \chi^2_{(2r_u)}$$

$$2\lambda_a T_a \text{ is distributed as } \chi^2_{(2r_a)}$$

$$2\lambda_a^* T \text{ is distributed as } \chi^2_{(2r)}.$$

Since $2\lambda_u T_u$, $2\lambda_a T_a$, and $2\lambda_a^* T$ are independent random variables, the random variable Z defined by

$$Z = \frac{2\lambda_u T_u/2r_u}{2\lambda_a T_a/2r_a}\,[2\lambda_a^* T/2r]$$

is distributed as the product of an F variate with a chi-square variate over its degrees of freedom. This distribution has been plotted by Monte Carlo techniques, for varying combinations of degrees of freedom, in [II-15].

Denote the distribution function of Z by $G(z; r, r_u, r_a)$, and the α -th fractile of Z by

$$Z_\alpha (r, r_u, r_a) \quad .$$

Note that G is symmetric in its first two parameters, that is to say

$$Z_\alpha (r, r_u, r_a) = Z_\alpha (r_u, r, r_a) \quad .$$

Define the random variable W by

$$W = \frac{T_a r_u r}{T_u T\, r_a} \quad .$$

Since

$$\Pr\left\{Z \leq Z_{1-\alpha}(r,r_u,r_a)\right\} = 1 - \alpha \ ,$$

it follows that

$$\Pr\left\{\frac{\lambda_u}{\lambda_a}\lambda_a^* \leq WZ_{1-\alpha}(r,r_u,r_a)\right\} = 1 - \alpha \ ,$$

How does this allow one to make statements about λ_u^*, the failure rate for the new lot at use conditions? If the acceleration factors for the control lot and the production lot are identical, then

$$\lambda_u^* = \frac{\lambda_u}{\lambda_a}\lambda_a^*$$

Thus, under this assumption,

$$\Pr\ \{\lambda_u^* = WZ_{1-\alpha}\ (r,r_u,r_a)\} = 1 - \alpha$$

from which one can set up tests of hypotheses concerning the value of λ_u^*, place confidence intervals on λ_u^* and, even if $r_u = 0$, one can place upper confidence intervals on λ_u^* (see [II-15]).

No investigation was performed regarding the robustness of this procedure to errors in the assumption $\lambda_u^* = \lambda_u \lambda_a^* / \lambda_a$. However, this procedure is clearly an improvement over the previously discussed elementary method since

(i) only the constancy of the acceleration factor is assumed, without assumptions regarding its actual value

and (ii) all the data are explicitly used, rather than merely relying on an estimate of the acceleration factor.

II.E LITERATURE SURVEY

At the very outset of the effort which culminated in the present report, an extensive survey of the technical literature was made. The primary purpose of the literature survey was to gain maximum benefit from past efforts by selecting past work which was correct and could provide an advanced starting point for the current investigation. A further purpose of the literature survey, and the subsequent detailed review of selected items, was to establish a definitive statement of the present state of the art in the area of accelerated testing. Such a statement is presented in Section I.D.

The scope of the survey and conclusions regarding the applicability of past work ar stated below, and the remainder of this subsection comments on certain publications directly related to accelerated testing. The exhaustiveness of the survey notwithstanding, some publications were probably overlooked, and are therefore not mentioned. Some other publications are not mentioned because they are not even subject to meaningful criticism. On the one hand, there exists past work which appeared to deal with accelerated testing and the related questions of the dependence of failure behavior on environment. On the other hand, there is the ever growing work in statistical theory and methodology, some of which is potentially applicable to problems likely to be encountered in the context of accelerated testing.

In the first area, the major initial sources were the Proceedings of the National Reliability and Quality Control Symposia, the IEEE Transactions on Reliability and Quality Control, the IEEE International Convention Records, and various publications of Battelle Memorial Institute. In the second area, emphasis was placed on the Journal of the American Statistical Association, the Annals of Mathematical Statistics, and Technometrics.

The survey indicated that a great deal of work had been done in the area of accelerated testing. However, the bulk of that work was either unrealistic (in a physical sense), or too restrictive in application or, in some cases, incorrect. Thus it was necessary to examine even the most fundamental questions of accelerated testing.

Only three publications were singled out as providing significant direct contributions to accelerated testing. The publications of Allen [II-10] and [II-11] provide a fruitful and statistically correct discussion of the consequences of stress variation when the devices at hand fail in accordance with a certain model. His work is mentioned frequently in this report (e.g. Subsections III.B.1, III.B.3, and III.B.4) and needs no further discussion here. The other significant contribution is that of Endicott and Zoellner [II-16]. They adapted the ideas of Prot [II-17] to accelerated testing, and attempted to incorporate the requisite statistical considerations. Their article provided the impetus for the development of the "Simple Wear" model which constitutes a major part of this report.

Endicott and Zoellner's approach to the verification of the postulated model is based on valid considerations, although they did not use valid techniques for the estimation of unknown parameters in their model. In equation (4) of their paper T' should have been used in place of T, in keeping with the notation introduced on page 231 of [II-16]; then a correct version of their equation (8) would follow. The particular model, as proposed by Endicott and Zoellner, is considered in detail in Subsection III.C.2, where valid estimation and verification techniques are derived. A subsequent publication co-authored by Endicott, i.e. Starr and Endicott [II-18], fails to even mention statistical considerations. However, this article is accompanied by an interesting discussion by Dakin [II-19] which, if reinterpreted in the terminology of this study, suggests that dependence only on damage accumulation may fail to provide an adequate description of breakdown of insulators. Dakin assumes that, at time t, the dielectric is characterized by a breakdown voltage V(t), and that V(t) is a decreasing function of time. The (instantaneous) rate of decrease depends on the applied voltage S(t), and "failure" occurs when the breakdown voltage first equals or is exceeded by the applied voltage. This assumption, though reasonable, is inconsistent with the pure damage accumulation approach of Starr and Endicott. For, under Dakin's hypothesis, a sufficiently high (but finite) applied voltage will cause immediate failure, which is impossible under the pure damage accumulation concept used by Starr and Endicott. If placed in the appropriate setting, Dakin's hypothesis essentially corresponds to Model 7 in Subsection III.B.1 of this report.

Some variations on the elementary approach (discussed in Section I) to accelerated testing can be found in papers by Bessler et al. [II-20] and by Chernoff [II-21]. These papers are based on the assumption that the failure distribution is exponential, the form of the acceleration function (see Subsection I.C.2 for definition) is known exactly, and that values of the constants in the accelerating function are known to a very good approximation. Questions of estimation and decision are considered from the standpoint of designing tests which minimize the cost of testing. The approach is indeed a variation on the elementary approach since they do not rely exclusively on an a priori "acceleration factor." Instead they allow for correction of the assumed values based on observations. Furthermore, under the elementary approach, the accelerated test is performed at only one increased environmental level, whereas the tests described in [II-20] and [II-21] may involve several levels. However, the results are obtained under excessively unrealistic and restrictive assumptions. In fact Chernoff himself states ([II-21], p. 408) that "In practical applications, the experimenter would be well advised to use non-optimal designs which ... yield estimates which are not sensitive to the underlying assumptions and also provide data which can be used to check these assumptions ..."

II.F DETAILED RECOMMENDATIONS

The elementary approach to the application of ALT to estimation or decision is not recommended. It requires beforehand knowledge of the appropriate acceleration function or, in particularly simple cases, the appropriate acceleration factor. As has been illustrated (Subsections II.B.2.4, II.C.3, and II.D.2), small uncertainties in one's knowledge of the acceleration factor can completely invalidate inferences drawn from an accelerated life test which is based on the elementary approach. However, the application of ALT to selection may be fairly insensitive to uncertainties in one's a priori knowledge of the time transformation function (see Subsection II.B.2.4). Therefore, it is recommended that this application of accelerated testing be studied further. More detailed study is required regarding the robustness of the procedure, both with respect to the errors in the assumed failure distribution, and with respect to errors in the hypothesized time transformation relationship. Also, further study is required regarding conditions which the failure distribution and the time transformation must satisfy in order to make selection under accelerated conditions economically justifiable (see Subsection II.B.1).

Some variations of the elementary approach have been proposed, e.g. by Bessler et al. [II-20], Chernoff [II-21], and Allen [II-10], [II-11]. The Bessler and Chernoff approaches (see Section II.E) are predicated on fairly accurate beforehand knowledge of the acceleration function (i.e. the function which specifies the acceleration factor throughout a range of environments). Such knowledge is generally not available; therefore it is not recommended that their approach be further explored in the immediate future, though their results should be considered whenever an approximate acceleration function is known for a device. Allen's approach (see Subsection II.B.3) requires less a priori knowledge than the approaches of Bessler and Chernoff; therefore his approach is closer to being potentially applicable. Since Allen's approach assumes the form of the failure distribution, the only possible justification for using it would be a potential reduction in sample size, i.e. the justification would be an economic one, (see discussion in Subsection II.A.2). His approach merits further investigation, in order to determine whether it indeed is economically justifiable.

The time-compression approach to accelerated testing is potentially very fruitful (see Subsection I.B.3). It presents no conceptual problems and, hence, no further general development is required. However, each individual application of this technique presents special problems. It is recommended that the technique be applied to one or more typical space guidance components, in order to explore in detail the implicit problems. The application should be such as to exemplify the techniques of physical, chemical, engineering, and statistical analysis required in devising a valid application of time-compression. In addition to illustrating the method of devising an accelerated test by means of time-compression, the application should illustrate methods of actually demonstrating the validity of a particular application. It is recommended that the time-compression technique be applied in this spirit to thermal expansion compensation bellows (see Subsection IV.C.5), to pumps (see Subsection IV.C.4), to gas bearing failures induced by start-stop cycling (see Section IV.D), and to ball bearings (see Subsection IV.C.3).

The damage accumulation approach to failure processes (see Sections III.B and III.C) has received little attention in the open literature, but was developed extensively in this study. It appears to be a very promising approach to accelerated testing in some instances. Therefore, it is recommended that the use of damage accumulation models in connection with accelerated testing receive further emphasis, both in general development and in specific application. It is recommended that this approach be applied to estimating the fatigue life of rolling element bearings (see Subsection IV.C.3.2) and to estimating the life of motor brushes (see Subsection IV.C.2.3). In the case of motor brushes, the Ehrenfeld-DeCicco [II-22] version of the damage accumulation model (see Subsection III.B.1) should be examined for applicability. Finally, in connection with damage accumulation, further investigation is recommended regarding questions of robustness, and it is recommended that the requisite statistical techniques receive further emphasis.

Two specific areas of statistical investigation are recommended for further development. The statistical treatment of devices with multiple failure modes must be developed further if accelerated testing is to be done on any but the simplest, single failure mode, devices (see Section VI.A for details). Estimation procedures for so-called restoration processes, which allow the efficient use of data for repairable equipment, should receive further emphasis. Those techniques (see Section VI.B) effectively provide an "analytic acceleration."

The recently discovered effect of ultrasonic energy on metallic creep should be studied further, since it appears possible to develop it into a practical tool for accelerated testing of microcreep phenomena. Dorn [II-23] has developed methods for extrapolating creep data on the basis of data from tests at temperatures higher than use-environment temperatures. His approach is predicated on the existence of a material constant, called activation energy; the existence of such a constant was demonstrated for several pure crystals and polycrystalline materials. The possible existence of such a constant for engineering materials should be investigated. If it exists, then both Dorn's methodology and extrapolations obtained from experiments with ultrasonics can be used effectively providing a check of each other. For a detailed discussion of this subject, see Part V of this report.

III. DISCUSSION AND USE OF FAILURE MODELS

III.A INTRODUCTION

III.A.1 Background and Scope

In Subsection I.C.2 the concept of a time transformation function was introduced. Recall that the time transformation function a_{ji} from environment E_j to environment E_i fully describes the dependence of failure behavior on environment. Moreover the function a_{ji} transforms a fractile of the time to failure (TTF) random variable for environment E_j to the same fractile of the TTF random variable for environment E_i. Hence a sample drawn from the failure distribution for environment E_j transforms into a sample drawn from the failure distribution for environment E_i in such a manner that both samples possess the same statistical properties (e.g., if the observed sample is a random sample then so is the transformed sample). Thus, given complete knowledge concerning the time transformation function (the ideal situation), accelerated life testing (ALT) becomes a trivial problem, for then one needs only to conduct testing under environment E_j, transform the sample to environment E_i, and then utilize available life testing analysis techniques on the transformed sample.

In general, the time transformation function is not known except in situations where the time compression (TC) approach is applicable. Under the TC model, the time transformation function is simply kt where k is the ratio of cycles per unit time at accelerated conditions to cycles per unit time at use conditions. Thus if the TC model holds, then ALT becomes simply a problem of life testing. For this reason, the term ALT will only refer to the over-stress testing (OST) approach throughout the remainder of Part III.

Only environments which can be adequately described by a scalar valued function will be considered in the remainder of Part III (see Subsection I.C.2 for a discussion of scalar environments). Moreover, the devices considered for application can be either of the single failure mode variety, or of the multiple failure mode variety restricted in the following sense: If the device is of the multi-mode variety then what follows either applies to the device as a whole, with no distinction being made between failure modes, or to a single mode isolated from all other modes of failure. When the failure modes of a device are partitioned into several classes of failure modes and accelerated life testing procedures such as those contained in this report are applied to one or more such classes, then the device as a whole must be considered through the application of special techniques. For a discussion of these techniques see Section VI.A.

When the time transformation function is not known, the problem of constructing ALT procedures becomes fairly difficult. Subsection III.A.2 considers a completely theoretical approach (the ideal situation). The problem with this approach is that it is hardly ever, if at all, applicable in reality. Subsection III.A.3 presents a strictly empirical approach to the solution of the problem, along with its pitfalls. It is found, as would be expected, that the strictly empirical approach is far from satisfactory and is hardly worthy of consideration.

Hence the most realistic and practical situation occurs when the conjunction of the two approaches is taken. This approach is considered in Subsection III.A.4. The approach taken in Subsection III.A.4, however, requires some knowledge concerning at least the form of the time transformation function. Although one may not have explicit knowledge concerning the time transformation function, it turns out that through the introduction of some auxiliary concepts and through heuristic considerations, one can in many cases generate a plausible time transformation function. In Subsections III.B.1 through III.B.5 some fundamental failure models are considered. These models partition a large class of failure models into 8 subclasses. Each particular class of models has its own identifying characteristics, and some of the classes have been considered in detail by other investigators. A particular variety of damage accumulation models is considered in great generality, and one particular subclass of such models has been considered extensively by the authors of this report. That particular subclass will be referred to as "the SW model," whereas "damage accumulation" refers to the entire class of which the SW model is but a part. It appears that the SW model covers a broad class of devices on which it may be desirable to conduct accelerated life testing. It is found that if the dependence, in general terms, of the hazard function on environment is known, then in many cases the time transformation function becomes readily determinable through heuristic considerations. Particular cases of the SW model which appear to have a wide range of applications are considered in Subsections III.C.1 through III.C.4. These particular cases, as is shown in Section III.C, lend themselves to the determination and partial verification (as discussed in Subsection III.D.1) of the time transformation function so that feasible accelerated life testing does in fact result. The combination of modeling and heuristic considerations will be seen to be a very fruitful approach to ALT.

In Subsections III.D.1 through III.D.3, verification techniques for testing the validity of postulated time transformation functions are given.

As indicated in Parts I and II of this report, the need for accelerated life testing arises when economic considerations preclude reliance on ordinary life testing. It has been mentioned that, to successfully solve the problem of ALT, one needs information concerning the time transformation function. Knowledge of that function forms the basis for accelerated implementation of the inference procedures discussed in Part II. Therefore, Part III is essentially devoted to problems of formulation (through the combination of modeling and heuristic considerations), estimation, and verification of time transformation functions and their unknown constants.

This section is concluded by introducing some notation and terminology that shall be used throughout the remainder of Part III.

III.A.2 Definitions and Comments

Consider a population of devices with a failure distribution F, which is dependent on the stress environment, say E. One wishes to make inferences concerning F at some sets of stress environments, to be called normal operating environments. If there is only one normal operating environment, then it shall be called the use conditions environment. Suppose further that one randomly draws samples of devices from the population, and subjects each sample to life tests under k distinct environments, all of which need not be normal operating environments. From the outcomes of these tests one wishes to make inferences concerning F at the normal operating environments. The tests conducted under the non-normal operating environments shall be called accelerated life tests and their respective environments shall be called accelerated environments. Without loss of generality, only those situations where there are k distinct accelerated test environments and only one normal operating environment shall be considered. It is not necessarily assumed that a life test is conducted at the use conditions environment. To a given environment, one affixes an index "i", denoting the environment. Generally, the index i will take values on the positive integers, though the value i=u will be used to denote the use conditions environment; that is to say, E_i denotes the i^{th} accelerated test environment and E_u denotes the use conditions environment. The following notation shall be utilized throughout the remainder of this report.

Let F_i denote the failure distribution in the i^{th} environment, let ϕ_i denote the corresponding hazard function (HF), and let Φ_i denote the corresponding integrated hazard function (IHF). Recall the connection between the three characterizations of the population of time to failure.

$$\phi_i(t) = \lim_{\Delta t \to 0} \left[\frac{Pr(t<T\leq t+\Delta t \mid T>t)}{\Delta t} \right] = \left[\frac{d}{dt} F_i(t)\right]/[1-F_i(t)]$$

$$\Phi_i(t) = \int_0^t \phi_i(s)ds$$

$$F_i(t) = 1 - \exp\{-\Phi_i(t)\} \quad ,$$

where T is the time-to-failure (TTF) random variable.

Hereafter the subscripts "i" and "j" are considered distinct unless otherwise specified. The following definitions are now formulated.

Definition 1: Suppose there exists a function, a, dependent on E_i and E_j, denoted by

$$a_{ji}(t) = a(t;\ E_i, E_j) \ ,$$

such that

$$F_j(t) = F_i[a_{ji}(t)] \quad .$$

The function a_{ji} shall be called the time transformation function of TTF under E_j to TTF under E_i.

For example, suppose that F_i and F_j are both exponential distributions with means θ_i and θ_j respectively. Then

$$F_i(t) = 1 - \exp(-t/\theta_i)$$

and

$$F_j(t) = 1 - \exp(-t/\theta_j)$$

$$= 1 - \exp(-kt/\theta_i) = F_i(kt)$$

where $k = \theta_i/\theta_j$. Thus, in this case, the time transformation function is defined by

$$a_{ji}(t) = kt \quad .$$

If $i = u$, then E_u is the use conditions environment and k is equivalent to what is commonly called the acceleration factor (or K-factor) between the accelerated environment E_j and the use conditions environment.

<u>Definition 2:</u> The function a_{ji} shall be called a dominant time transformation function iff

$$a_{ji}(t) \geq t \text{ for all } t$$

and

$$a_{ji}(t) > t \text{ for a set of positive measure in } t.$$

If F_i and F_j are both continuous functions, this is equivalent to requiring $a_{ji}(t) \geq t$ for all t, and $a_{ji}(t) > t$ for at least one interval (a,b] of finite length on the positive reals. For example, if F_i and F_j are related as shown in Figure III-1, then the time transformation function satisfies the definition of dominant time transformation function. In the case of Figure III-1,

$$a_{ji}(t) = t \quad \text{for all } t \leq T_{.05}$$

$$a_{ji}(t) > t \quad \text{for all } t > T_{.05} \quad ,$$

thus satisfying definition 2.

If no confusion arises, the compound relation " ≥ for all t and > for a set of positive measure in t" shall be denoted simply by $\sim$. In this notation, definition 2 states that a_{ji} is a dominant time transformation function iff

$$a_{ji}(t) \sim t \quad .$$

Note in passing that there always exists a time transformation function between two continuous cdf's but it does not have to be dominant. For example, if F_i and F_j are both continuous this would be the case if the functions intersected at some finite number of points in t, as shown in Figure III-2.

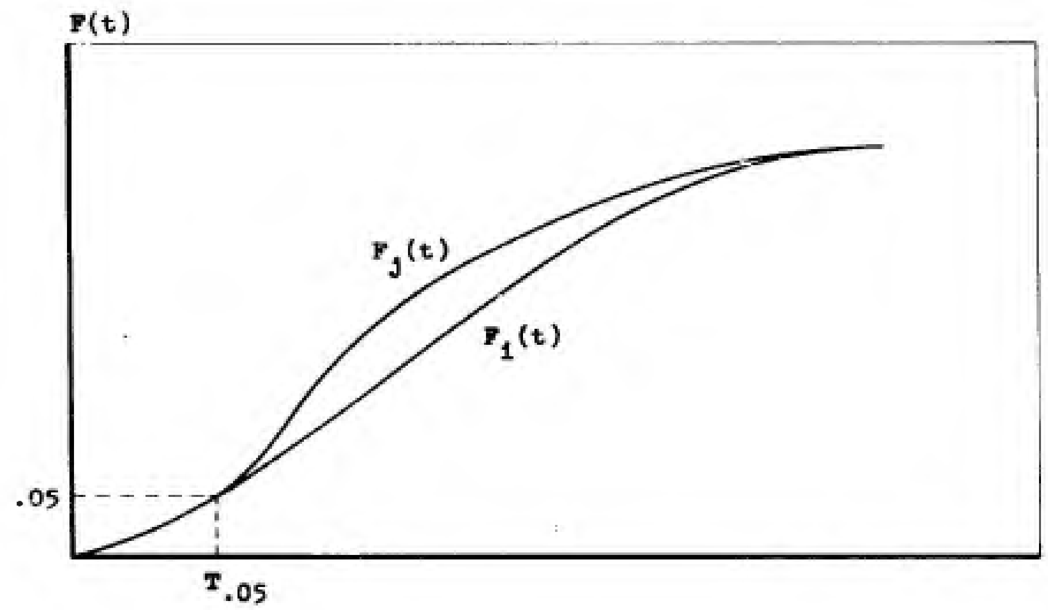


FIG. III-1. DISTRIBUTIONS RELATED BY A DOMINANT TIME TRANSFORMATION FUNCTION

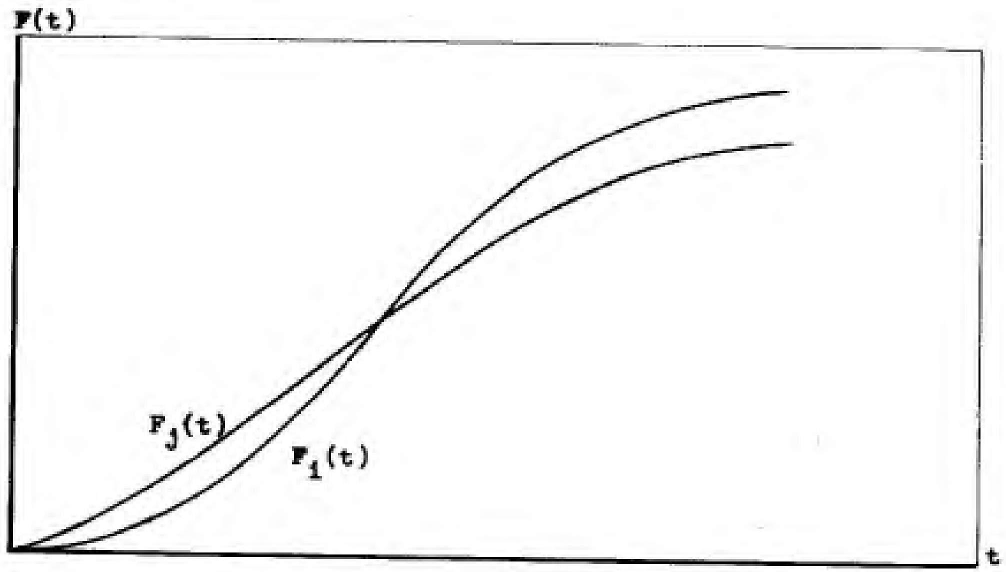


FIG. III-2. DISTRIBUTIONS RELATED BY A NON-DOMINANT TIME TRANSFORMATION FUNCTION

The time transformation function fully specifies the dependence of failure behavior on environment. Thus, if the time transformation function is known, statistics of a test under E_j yield the corresponding statistics under E_i without the necessity of testing under E_i. For instance, the empirical distribution function of data from a test under E_j estimates F_j; by means of the time transformation function, this empirical distribution function can be transformed to yield a statistically equivalent estimate of F_i. Specifically the accelerated data is transformed to use conditions as follows:

If $(t_1, \ldots, t_r)$ denote the first r times to failure in a test conducted at E_j, and if a_{ju} denotes the time transformation function from E_j to E_u, then the statistically equivalent sample $(t_1^*, \ldots, t_r^*)$ at use conditions is obtained by means of the relationship

$$t_i^* = a_{ju}(t_i) \quad , i=1, 2, \ldots, r .$$

Consider briefly the problem of conjecturing whether or not there exists an environment E_j such that a dominant time transformation function a_{ju} exists. Let A be the set of all dominant time transformation functions from environments E_j to the use conditions environment E_u. In customary mathematical notation,

$$A = \{a_{ju} | a_{ju} \text{ is dominant}\} .$$

From definition 2 it follows that, if F_u and F_j are continuous, then

$$a_{ju}(t) \sim t \quad \text{iff} \quad F_j(t) \sim F_i(t)$$

which is equivalent to

$$a_{ju}(t) \sim t \quad \text{iff} \quad \Phi_j(t) \sim \Phi_i(t) \ .$$

Thus a sufficient condition for a_{ji} to be dominant is that

$$\phi_j(t) \sim \phi_i(t) \ .$$

The concept of dominance can be related to intuitive considerations as follows. Put on test a device drawn randomly from the population of devices under consideration; as soon as it fails, replace if by another device which has not yet failed, etc. The sequence of observed failures forms what is called a restoration process (see Section VI.B). If $N_{t,i}$ denotes the number of failures observed before time t under E_i, then it can be shown that (see, e.g., [III-1], pp. 125-6)

$$E[N_{t,i}] = \Phi_i(t) \quad .$$

Hence the time transformation function a_{ji} is dominant iff

$$E[N_{t,j}] \sim E[N_{t,i}] \ .$$

That is, the time transformation function from environment E_j to environment E_i is dominant iff a restoration process in environment E_j is expected to produce at least as many failures as a restoration process under environment E_i for any length of observation, and actually more failures in some interval.

The attempt to make inferences concerning F_u is fairly difficult even when testing is conducted at environment E_u unless large amounts of data are obtainable. In general it is specifically assumed that one is dealing with very reliable items such that large scale testing at environment E_u is not economically feasible. See Subsection I.B.3 for further discussion of the need for ALT. However, suppose one could conduct life tests at environment E_u such that information about the lower tail of F_u was obtainable. Is there then any basis for attempting accelerated life testing? The answer in many cases is in the affirmative, for the data obtained from the accelerated tests could be used to further strengthen the inferences regarding the tail of F_u as well as to make additional inferences concerning the whole of the life distribution. A specific example of testing at use conditions as well as at accelerated environments arises in the attempt to construct accelerated lot acceptance test plans, considered in Subsection II.D.3.

Without loss of generality only the situation where it is desirable to make inferences concerning F_u through tests at k accelerated environments (without testing at use conditions) shall be considered. Obviously if testing is also conducted at use conditions the problems become less complex.

III.A.3 Empirical Approach

In Subsection III.A.2 it was noted that if the time transformation function a_{ju} from environment E_j to E_u is known, then accelerated life testing immediately reduces to ordinary life testing. This, of course, is an extreme situation which hardly ever, if at all, exists in reality. In this subsection, the other equally ridiculous extreme is considered, i.e., the case where absolutely no information is available concerning the time transformation function. That is to say there is no information available concerning physical theories, failure mechanisms and their relation to stress, etc., from which one could infer something about the form of time transformation functions.

One possible approach would be to first choose k steady stress accelerated test environments. For lack of information to the contrary, suppose the Weibull family (or some other family of cdf's characterizing life distributions) characterizes the failure distributions at the various environments. That is, the failure distribution under E_i is of the form

$$F_i(t) = 1 - \exp(-t^{\beta_i}/\alpha_i) \quad . \qquad \text{(III.A.3-1)}$$

Suppose it is further assumed that, the higher the index i, the more severe the stress environment; that is to say

$$F_1(t) < F_2(t) < \ldots < F_k(t) \quad \text{for all t.}$$

Suppose now that tests are run under each of these steady stress environments and that estimates of the shape parameters $\beta_1, \ldots, \beta_k$ and of the scale parameters $\alpha_1, \ldots, \alpha_k$ are obtained. If S_i denotes the scalar description of stress under E_i, then the scale parameters can be plotted versus the stress levels, as indicated in Figure III-3.

The same procedure can be used for the shape parameters. Recall that the mean of a random variable distributed in the Weibull family is given by

$$E(t|S_i)=[\Gamma(1+\frac{1}{\beta_i})][(\alpha_i)^{\beta_i}] \qquad \text{(III.A.3-2)}$$

If the shape parameters remained relatively constant, then one would suspect from Eq. (III.A.3-2) that the scale parameters would be related to the stress levels in a monotonically decreasing fashion. One could now estimate the Mood-Brown median regression line (see Subsection III.D.3) and test whether or not the stress-scale parameter relationship is linear, against an alternate choice of a convex relationship. If the hypothesis that the true relationship is linear is not rejected, then one could estimate the scale parameter at use conditions by

$$\hat{\alpha}_u = \hat{a}S_u + \hat{b}$$

where $\hat{a}$ and $\hat{b}$ are the estimates of the constants in the linear functional relationship based on the Mood-Brown median regression line.

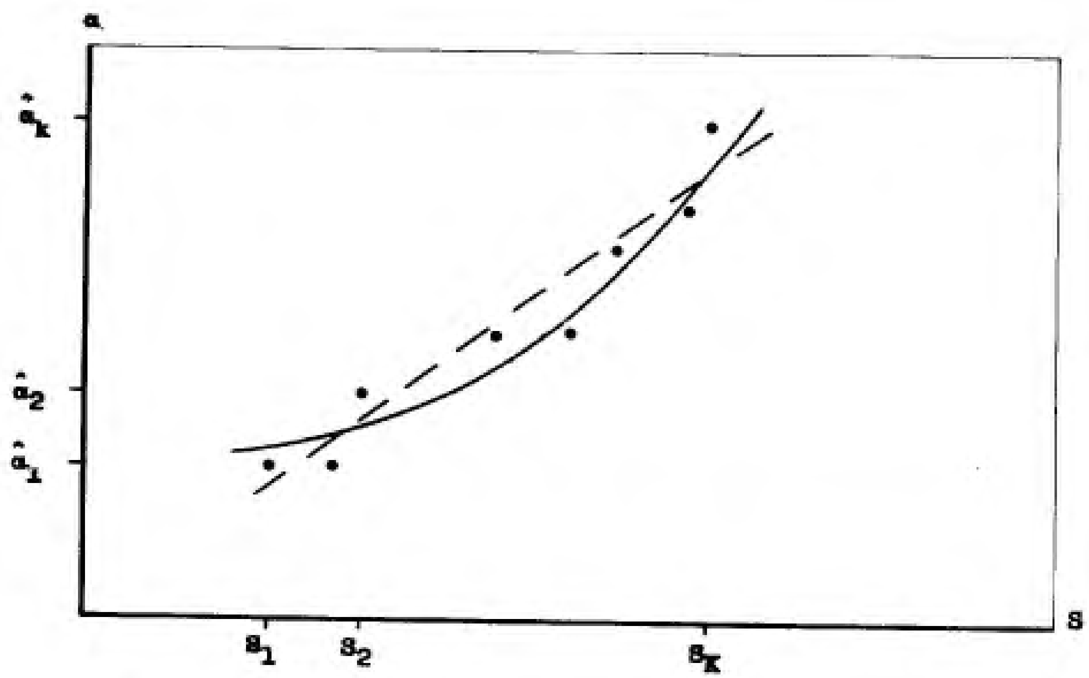


FIG. III-3. POSSIBLE REGRESSIONS OF α ON S

Another possible approach might be to assume a relationship

$$\alpha = aS^b \qquad \text{(III.A.3-3)}$$

or equivalently

$$\log \alpha = \log a + b \log S \qquad \text{(III.A.3-4)}$$

This relationship is also linear and through the use of regression techniques (say a weighted least squares approach based on the standard error estimates of $\hat{\alpha}_1, \ldots, \hat{\alpha}_k$), one could again attempt to extrapolate to the use conditions environment.

Either of these approaches could be used, as well as infinitely many others, and probably every approach would yield different results. Thus, there is no unique regression relationship which can be confidently used for extrapolation to use conditions, since there is no a priori information which would give preference to some particular form of regression over all other forms.

The following sections present certain concepts which do at times lend themselves to the construction of accelerated life testing tools and which, moreoever, lend themselves to the construction of tests of validity.

III.A.4 Semi-Theoretical Approach

Two extreme approaches to accelerated life testing were considered in the last two subsections. In reality, one is likely to encounter the intermediate situation in which one knows (or is willing to assume) the functional form of some aspect of the dependence of failure behavior on environment. For instance, past experience strongly suggests that the mean life of capacitors varies in proportion to some power of the applied voltage. If the functional form is really known (rather than merely conjectured), unknown constants can be estimated to any desired precision, thus providing a sound basis for extrapolation of accelerated test data to use conditions. If one only conjectures the functional form, a statistical test of that conjecture can be performed on the basis of data from several distinct accelerated tests. If the conjecture is not refuted by the test, the functional form, together with estimates of the appropriate constants, can then be used to transform all the accelerated test data to presumably equivalent use conditions data. If the transformation to use conditions is the correct one, the transformed samples will be independent samples from the same parent population. Thus, the correctness of the transformation can be tested by applying a so-called k-sample test (a number of such tests, nonparametric in nature, are discussed in Subsection III.D.2). If this test fails to reject the hypothesis that all the transformed samples come from the same population, one would feel justified in using the postulated functional form and estimated constants for extrapolation of accelerated test data to use conditions. Such a semi-theoretical approach will always yield more fruitful results than will a purely empirical approach.

In order to further the art of applying procedures such as the ones discussed in the above paragraph, the remainder of Part III of this report explores the relationship between heuristic considerations, mathematical modeling, and statistical verification procedures. The approach is illustrative and general in nature.

III.B FAILURE MODELS

III.B.1 General Framework

Depending on how the hazard function (HF) relates to its corresponding environment one can frequently choose an appropriate model which lends itself to the development of analytical tools for feasible accelerated life testing. The adoption of a particular model for the purpose of generating analytical tools to make accelerated life testing feasible requires an extreme amount of caution on the part of the experimenter. Therefore this subsection explores in detail the implications of several failure models. To begin with, consider the following four examples.

Example 1: Consider an automobile tire, and assume that its only possible failure is a blow-out, i.e. the tire is not removed from service due to excessive wear. Such failure results directly from a random shock environment, e.g. nails on the road. Certainly the probability of failure depends on the amount of wear and on the resilience of the rubber at the instant a shock occurs, e.g. a nail is encountered. Thus the hazard function at a given time depends on the nature of the environment at that time (e.g. local density of nails) as well as on the amount of wear up to that instant and, if the rubber deteriorates with age, it also depends on age.

Example 2: Consider a ball bearing that fails due to fatigue (pitting on the surface of the bearing deteriorates the surface until seizure occurs). In this case the hazard function at a given instant depends only on the amount of damage sustained up to that instant.

Example 3: Consider a device which fails only in a shock environment, the probability of failure depending on the magnitude of the shock. The device hazard function may or may not be dependent on the age of the device: if the probability of failure, for a given magnitude of shock, is not constant in time then the hazard function is dependent on age.

Example 4: Consider the marginal failure distribution of an organism for death caused by radiation. The organism might fail due to

(i) a dosage of radiation, acquired almost instantaneously, exceeding some critical level, or

(ii) a radiation dosage accumulated in time exceeding a different critical level.

For both causes of death the critical dosage is probably a monotonically decreasing function of age. That is to say the older the organism, the smaller the critical dosage required to induce death (e.g. a man at the ripe old age of 99 certainly would not be able to withstand as high a dosage of radiation as a young man of 24). This example characterizes a life model dependent on wear (an accumulation of radiation), instantaneous stress (instantaneous intensity of radiation), and age.

The above examples suggest that a hazard function can conveniently be categorized on the basis of the following three characteristics.

(i) the hazard function may be independent of or dependent on the age of the item.

(ii) the hazard function may be independent of or dependent on a physical or abstract property of the item intuitively identifiable with "wear" or "accumulated damage."

(iii) the hazard function may be independent of or dependent on the instantaneous stress.

Since each of three dependence properties either is or is not present, there result $2^3 = 8$ failure models. Those models, together with the names used for them in this report, are listed in Table III-1. The digits in the table refer to HF independence of, or HF dependence on, the identifying characteristic appearing in the column heading, according as the digit takes on a value "0" or "1."

The notion of dependence was treated above in a very intuitive fashion. In general, one can say that the HF is independent of some

TABLE III-1. FAILURE MODEL CATEGORIZATION

Model No.	Wear	Stress	Age	Model Designation
1	0	0	0	Exponential
2	0	0	1	General cdf
3	0	1	0	Allen
4	0	1	1	Generalized Allen
5	1	0	0	Damage Accumulation
6	1	0	1	Generalized Damage Accumulation
7	1	1	0	Wear-Stress
8	1	1	1	Total Dependence

characteristic if knowledge of the actual state of that characteristic in no way affects the hazard function. This notion will be further clarified in the ensuing discussion of the various models.

Models 1 and 2:

These two models characterize the situation where the HF is independent of the stress environment and independent of any wear processes or damage accumulation phenomena. Model 1 characterizes the situation where the HF is furthermore independent of age and thus the same exponential failure distribution holds for all environments. Model 2 characterizes the situation where the same failure distribution (not exponential) holds for all environments. Neither model appears plausible in the real world. If a situation occurred where it appeared that the HF did indeed obey one of these models, then it might be that incorrect acceleration factors were considered. For example, consider accelerating the life of an electronic amplifier by conducting the test in Los Angeles rather than in New York. Certainly, this is a far fetched example. More realistically, suppose an attempt is made to accelerate the life of light bulbs by increasing the room temperature from 20° C to 30° C. It is highly unlikely that an observable acceleration will result. Hence Model 1 or 2 would be indicated erroneously.

Model 3:

This model characterizes the situation where the hazard function is independent of the age of the item, independent of any wear or damage accumulation phenomena and is dependent only on the instantaneous applied stress. This situation was considered by Epstein [III-2] and subsequently, with respect to its application to ALT, by Allen [III-3]. Since the HF is independent of the age of the item, then certainly the failure distribution under a time-invariant stress environment (also called steady stress) must be exponential. Thus if two environments coincide at some time, say t_o, then the probability of failure of an item in the near future, given that it has survived until time t_o, is independent of which environment it was subjected to prior to time t_o. However, observe that under this model TTF under non-steady stress environments is distributed non-exponentially.

Model 4:

This model is simply a generalization of Model 3. That is, instead of having an exponential distribution of TTF under steady stress environments, the cdf is general. For example the defining cdf may be one of the gamma, Weibull, or Lognormal types. An example would be a device subjected to a random shock environment where the conditional probability of failure from a given shock is dependent on the age of the device. This type of model will be considered in some detail in Subsection III.B.3.

Model 5:

A HF belongs to Model 5 when it is independent of the age of the item as well as the instantaneous applied stress, and is dependent only on the "amount of change" (not necessarily known or measurable) which has occurred in the physical or abstract properties of the item. The "amount of change" shall be subsequently called wear or damage accumulation. This model shall hereafter be termed the damage accumulation model.

The damage accumulation model characterizes a situation where an item fails due to an accumulation of damage. The dependence of failure on wear can be either deterministic or probabilistic. That is to say, the item either inevitably fails when a certain critical amount of damage has been accumulated (and never fails prior to that event), or the item has a probability of failing for each value of wear.

Consider the first case, where the failure event is in a one-to-one correspondence with the reaching of a critical wear level. There are two possible sources for the randomness usually observed in failure times. First, the wear process itself may be stochastic in nature. And, secondly, the total wear permissible may vary from item to item. The case where randomness of failure times arises exclusively from randomness in the wear process (the critical wear level being the same for all items) was studied extensively by Ehrenfeld and DeCicco [III-4]; this model is discussed further in a subsequent paragraph. The deterministic wear case, where randomness of failure times is due exclusively to randomness in critical wear level, is discussed extensively in Subsections III.B.2 and III.B.4. A combination of the Ehrenfeld-DeCicco model and the random critical wear level assumption may be required in some practical instances.

The second case, in which failure depends on wear in a probabilistic sense, corresponds to a simplified form (i.e. without age dependence) of a model suggested by Mercer [III-5]. It is immaterial, in this case, whether the wear process is deterministic or stochastic; furthermore, the concept of critical wear level doesn't apply. However, the probability of failure, given the value of the wear variable, might also depend on age or the instantaneous applied stress or both. If there exists such a dependence on age, then the resulting model is of the type referred to as Model 6. If there exists such a dependence on the instantaneous applied stress, then the resulting model is of the type referred to as Model 7. If there exists a dependence on both age and instantaneous stress, then the resulting model is of the type referred to as Model 8. Models 6, 7, and 8 shall be discussed following the present discussion of Model 5.

Ehrenfeld and DeCicco [III-4] considered the wear process as being random, and they assumed that an item failed as soon as a critical level of damage (which does not vary from item to item) was accumulated. Moreover the amount of wear at any point in time was assumed to be measurable. From these assumptions and through heuristic considerations they were able to develop a life model (based on the gamma distribution) which allows one to make probability statements regarding the lifetime of a randomly selected item. If in fact the wear process is random, as it is in a model of this type, then one must be able to relate the wear process to various environments in order to use this model for accelerated life testing. That is to say, the Ehrenfeld-DeCicco model is or is not applicable in the field of accelerated life testing according as the relationship between the random wear process and the environment is or is not understood.

On the other hand, under the deterministic wear model, the TTF for a given item is a fixed quantity. The TTF of a given device could be determined precisely if the allowable amount of wear prior to failure and the rate of wear as a function of instantaneous stress were known. A device which obeys the deterministic wear model was discussed in Example 2 at the onset of this section. The deterministic wear concept was proposed by Endicott and Zoellner [III-6] for the accelerated life testing of capacitors; they, in turn, adapted the idea from a metallurgical application by Prot [III-7]. For a critique of their work see Section II.E. The deterministic wear model is treated extensively in Subsection III.B.2 and, for some particular cases of the model, estimation and verification techniques are derived in Section III.C. Those techniques make possible the development of accelerated life test plans.

It is interesting to note the implications of assuming an exponential failure distribution for steady stress environments. It turns out that, under an assumption of exponentially distributed TTF for steady stress environments, the probability of failure in the immediate future, given that the item has not yet failed, is independent of its previous stress history. It further happens under this assumption that the Allen Model (Model 3) is indistinguishable from the deterministic wear model.

A proof of this property is given in Subsection III.B.4. Thus, under this assumption, the HF is independent of previous stress history - which is contrary to one's intuition. This apparent paradox is discussed further in Subsection III.B.2.2. Finally, note that an exponential distribution for steady stress environments is the only exception to the following rule: under the restrictions that the HF is continuous and the probability of failure in the next instant of time is not zero or one (i.e. the wear rate does not become infinite or non-positive), it is the case that

$$\phi_j(t_j) = \phi_i(t_i)$$

if and only if

$$W(t_j \mid E_j) = W(t_i \mid E_i)$$

where ϕ_i and ϕ_j are the HF's under environments E_i and E_j respectively and $W(t_i \mid E_i)$ and $W(t_j \mid E_j)$ are the values of the wear variables at times t_i and t_j under environments E_i and E_j respectively.

Model 6:

Model 6 is a generalization of the damage accumulation model (Model 5) in that the hazard function is dependent on the age of the device as well as on the amount of wear accumulated. In the damage accumulation model, the probability of failure in the immediate future is independent of how much time it took to accumulate the damage, and depends only on the accumulated amount of damage. In Model 6, however, the "age" (i.e. length of time needed to accumulate the present amount of wear) also affects the probability of failure in the immediate future. In other words, age is of no value as a predictor of failure under Model 5, but it is of value under Model 6. A special case of this model was treated by Mercer [III-5] where the essence of Mercer's Model is contained in his summary which starts as follows:

> "The probability that a component fails is assumed to be linearly dependent on a physical wear variable, which can be measured, as well as on the age of the component. A wearing mechanism is represented by the extended Poisson process. ...."

The removal of the linearity assumption would not affect the model as such. Note well that the assumption of the wearing mechanism could also be modified to accept other forms, such as deterministic wear processes.

The earlier comments regarding the applicability of the Ehrenfeld-DeCicco model to accelerated life testing also hold for Mercer's model.

Model 7:

Model 7 characterizes a situation where the HF is dependent on wear and on instantaneous stress, but not on age. An item obeying this type of model is described in Example 1 at the onset of this discussion, provided the rubber is assumed to be non-deteriorating in time. This model includes one suggested by Dakin [III-8] for dielectric breakdown of capacitors, though Dakin's discussion is restricted in that he takes a completely deterministic approach. In Dakin's model, the dielectric strength of an insulator decreases in time under the influence of applied voltage, and failure occurs if and only if the applied voltage at some instant exceeds the dielectric strength at that instant. The influence of wear on the HF is the change in dielectric strength. The influence of instantaneous stress on the HF is clear, since the occurrence of a failure at any instant depends on the value of voltage at that instant. The observed randomness in failure times is probably mostly due to randomness in initial values of the dielectric strength.

Model 8:

This complex model allows for HF dependence on all three factors, i.e. age, wear, and instantaneous stress. It is very realistic, as is indicated by the fact that it includes both Example 1 (automobile tire) and Example 4 (living organism in a radiation environment) given at the beginning of this section. Its realsim notwithstanding, its complexity is such as to preclude any immediate useful applications of this model.

All of the models listed in Table III-1 have been considered to some extent in this discussion. The remainder of Section III.B is oriented to the discussion of the deterministic wear and generalized Allen models which, in view of the present state of the art, are considered to be the models most readily applicable to accelerated life testing.

III.B.2 The Simple Wear Model

III.B.2.1 General Theory

As mentioned in the discussion of Model 5 in the preceding section, it appears difficult to construct an accelerated life testing model when it is assumed that the wear process is random. Therefore, Subsection III.B.2 explores in detail a specific case of a deterministic wear model. In particular the case where the allowable wear prior to failure is a random variable and the wear process is strictly dependent on the environment is considered. It is not unreasonable to make such an assumption since many examples can be offered which illustrate that the amount of wear an item can withstand prior to failure does vary from item to item. Such a model is particularly worthy of consideration since it was used with apparent success for accelerated metallurgical testing (see [III-7]) and for accelerated testing of capacitors (see [III-6]). It also appears to be appropriate for the study of the fatigue life of rolling-element bearings.

It is here assumed that failure of a device is the direct result of an "accumulation of damage" over time (e.g., wear). Once an item in a population has endured a specific amount of damage (characteristic of that particular item), the item fails.

Suppose the environment, i.e. a time-ordered sequence of stress conditions, can be characterized by a scalar-valued function of time. Furthermore suppose that the amount of damage accumulated in a device is a non-negative numerical function of only the environment.

Since the amount of damage accumulated is assumed to be a non-negative function of only the environment, then it follows that the rate at which damage is accumulated cannot be a function of the amount of damage already accumulated. This restriction imposed on the deterministic wear model of Subsection III.B.1 yields the specific case mentioned above. The deterministic wear model with this restriction imposed on it shall hereafter be called the simple wear (SW) model. It should be noted that the general deterministic wear model can be handled by the SW model but it is convenient not to do so (see the last paragraph of Subsection III.B.2.1).

Let the accumulation of damage correspond in a natural way with the dissipation of some conceptual, non-negative quantity μ. This quantity μ can be perceived as corresponding to attribute(s) of the device. For example - energy to be dissipated (storage battery), mass to be disseminated (length of brushes in a motor), the number of molecular dislocations to be generated (fatigue fracture of metallic bar elements), etc. might typify this quantity. The amount of μ at time zero, $\mu(0)$, is the amount of damage that the item can withstand before failure, and the amount of μ at failure is zero.

In mathematical terms, it is assumed that the rate of dissipation of μ is given by the relation

$$-\frac{d[\mu(t)]}{dt} = -\mu'(t) = \begin{cases} w[S(t)], \text{ when } t \geq 0 \text{ and } \mu(t) > 0 \\ 0 \quad , \text{ when } t < 0 \text{ or } \mu(t) = 0 \end{cases} \qquad \text{(III.B.2-1)}$$

where $S(t)$ is a scalar description of the environment at time t and w is a non-negative integrable function with argument $S(t)$. Hereafter, w shall be referred to as the wear rate function.

Failure of an item occurs when the initial amount of μ, i.e. $\mu(0)$, is exhausted. Denoting the time to failure of a particular device by τ, there results the definition

τ is the smallest value of t
such that $\mu(t) = 0$.

In conventional mathematical notation,

$$\tau = \inf_t \{ t \mid \mu(t) = 0 \} \quad . \qquad \text{(III.B.2-2)}$$

Therefore, the time to failure of a particular device with initial amount of μ equal to $\mu(0)$, is related to the environment by

$$\tau = \inf_{t} \{t \,|\, \int_0^t w[S(x)]dx - \mu(0)\} \qquad . \qquad \text{(III.B.2-3)}$$

Since $w[S(t)]$ was assumed to be non-negative, the integral

$$\int_0^t w[S(x)]dx$$

is a strictly increasing function of time, thus assuring that (for a specified environment) there is a one-to-one correspondence between the failure time, τ, of a particular device and its initial amount of μ, i.e. $\mu(0)$. Thus it follows that the probability law of T, the time to failure of an item randomly drawn from the population, is completely specified by the probability law of $\mu(0)$, provided the environment is specified.

Once an environment is specified, the value of the above integral is a function of time. Denote that function by h_i when the specified environment is some E_i, $E_i = \{ S_i(t), t \geq 0 \}$.

That is

$$h_i(t) = \int_0^t w[S_i(x)]dx \qquad .$$

Let t_i be the time to failure of a particular device, in the environment E_i, and let μ^* be the value of $\mu(0)$ for that device. Then t_i is the unique solution of the equation

$$\mu^* = h_i(t_i) \quad . \qquad \text{(III.B.2-4)}$$

Since h is continuous (having been defined by an integral) and strictly increasing, it possesses a unique inverse h^{-1}.

Therefore

$$t_i = h_i^{-1}(\mu^*) \quad . \qquad \text{(III.B.2-5)}$$

Utilizing the notation heretofore specified let $E_1, \ldots, E_k$ represent k accelerated test environments and E_u the use conditions environment. Let S_i denote the scalar description of stress under E_i, $i = u, 1, \ldots, k$. For the randomly chosen item with $\mu(0) = \mu^*$ let t_i denote the TTF of the item had it been subjected to the i-th test environment until failure. From relation (III.B.2-4) it follows that

$$\mu^* = h_i(t_i),\ i = u, 1, \ldots, k, \qquad \text{(III.B.2-6)}$$

and from Eq. (III.B.2-5) $t_u = h_u^{-1}(\mu^*)$.

Hence from Eq. (III.B.2-6) it follows that

$$t_u = h_u^{-1}[h_i(t_i)] \quad , i = u, 1, \ldots, k \ . \qquad \text{(III.B.2-7)}$$

Now observe that if one has two random variables, say X and Y, with respective distributions F_X and F_Y, which are related by a continuous and strictly increasing function g given by

$$X = g(Y) \quad ,$$

then

$$F_Y(y) = F_X[g(y)] \quad .$$

Therefore, if the distribution of TTF under environment E_i is denoted by F_i, it follows from Eq. (III.B.2-7) that

$$F_i(t) = F_u(h_u^{-1}[h_i(t)]) \quad , i = u, 1, \ldots, k. \quad \text{(III.B.2-8)}$$

Thus the time transformation function between environments E_u and E_i is given by

$$a_{iu}(t) = h_u^{-1}[h_i(t)] \quad \text{(III.B.2-9)}$$

Similarly it follows that

$$a_{ij}(t) = h_j^{-1}[h_i(t)]$$

for all i and j.

This subsection is concluded by observing that, although it has been assumed that $\mu'(t)$ is not a function of $\mu(t)$, the situation where $\mu'(t)$ is a function of $\mu(t)$ can be satisfactorily handled by the SW model providing the differential equation (III.B.2-1) can be solved. It has been assumed here that $\mu'(t)$ is not a function of $\mu(t)$ for, if it were, it would merely complicate the exposition.

III.B.2.2 Steady Stress Environments

Consider the case when the environment E_i is a steady stress environment, that is $S_i(t) = s_i$ for all t. Then

$$\mu^* = h_i(t_i) = t_i w(s_i) \quad . \qquad \text{(III.B.2-10)}$$

Therefore, if the environment E_j is also a steady stress environment,

$$\mu^* = t_i w(s_i) = t_j w(s_j) \qquad \text{(III.B.2-11)}$$

so that

$$t_j = h_j^{-1}[h_i(t_i)] = \frac{w(s_i)}{w(s_j)} t_i \quad .$$

Thus it follows from (III.B.2-8) that, if E_u is a steady stress environment,

$$F_i(t) = F_u\left[\frac{w(s_i)}{w(s_u)} t\right] \quad . \qquad \text{(III.B.2-12)}$$

It immediately follows that, if all the environments E_i, i = u, 1, ..., k are steady stress environments, then Eq. (III.B.2-12) is valid for i = u, 1, ..., k. Hence the failure distributions for all those environments are members of the same family of distributions (e.g. the family of log-normal cdf's) and they differ only by a scale parameter.

It is interesting to note the consequences of assuming that TTF is exponentially distributed under steady stress environments. Under that assumption, it follows from Eq. (III.B.2-11) that the random variable $\mu(0)$ is also exponentially distributed. The converse is obvious: if $\mu(0)$ is exponentially distributed, then TTF in a steady stress environment is exponentially distributed.

Still assuming that TTF is exponentially distributed under steady stress environments suppose that E_o is an arbitrary environment; then by Eq. (III.B.2-4) and in view of the comment preceding Eq. (III.B.2-8),

$$F_o(t) = F_{\mu(o)}[h_o(t)]$$

where $F_{\mu(o)}$ is the distribution of $\mu(0)$.
Hence the hazard function at environment E_o is given by

$$\begin{aligned}\phi_o(t) &= \phi_{\mu(o)}[h_o(t)]\frac{d}{dt}[h_o(t)] \\ &= \lambda_{\mu(o)}w[S_o(t)]\end{aligned}$$

since $\mu(o)$ is exponentially distributed with mean equal to, say, $1/\lambda_{\mu(o)}$. Clearly it follows in this case that, for any stress environment, the probability of failure in the immediate future given that the item has not failed is independent of its previous stress history and depends only on the applied stress at that time. Certainly this is contrary to one's intuition of damage accumulation. This is a notable paradox because intuition dictates that for any damage accumulation phenomenon, the probability of failure in the immediate future is dependent on how much damage has already been accumulated.

But it is difficult to conceive of a "wear" variable whose critical level (i.e. $\mu(o)$) is exponentially distributed. Thus, it is doubtful that a simple wear (SW) model applies when TTF is exponentially distributed under steady stress since this occurs if and only if $\mu(o)$ is exponentially distributed. Finally, note that the SW model is indistinguishable from Model 3 of Subsection III.B.1 when TTF is exponentially distributed under steady stress, as will be shown in Subsection III.B.4.

Return now to the conclusion made in the first paragraph of Subsection III.B.2.2, namely that in the SW model, the failure distributions for all steady stress environments belong to the same family. Some consequences of that observation are now examined for the case when that family of cdf's is the two-parameter Weibull. Denote the scale parameters by α_i and the shape parameters by β_i. Then it follows from Eq. (III.B.2-12) that the following relations must hold:

$$\left.\begin{array}{l} \beta_u = \beta_1 = \dots = \beta_k = \beta \quad \text{(the common value)} \\ \alpha_u = \left[\dfrac{w(s_1)}{w(s_u)}\right]^{\beta} \alpha_1 \quad = \dots = \left[\dfrac{w(s_k)}{w(s_u)}\right]^{\beta} \alpha_k \end{array}\right\} \quad \text{(III.B.2-13)}$$

That is, if the Weibull distribution is the underlying cdf, then the shape parameters must be equal and the scale parameters must be related as in Eq. (III.B.2-13), for all steady stress environments.

If some failure process is assumed to follow the SW model, and this assumption is to be verified, the conclusion of the first paragraph of Subsection III.B.2.2 can be used to test whether or not the underlying family of cdf's (under steady stress) is some particular family. For instance, if the Weibull family is assumed but the hypothesis that the scale parameters obey Eq. (III.B.2-13) is rejected by a statistical test, then the Weibull assumption must also be rejected.

Denote by T_i the TTF random variable under steady stress environment E_i. Then it follows from Eq. (III.B.2-11) that the expectation

$$E[\mu(o)] = w(s_u)E[T_u] \qquad \text{(III.B.2-14)}$$

$$= w(s_1)E[T_1] = \ldots = w(s_k)\ E[T_k]$$

from which it follows that

$$E[T_i] = \frac{E[\mu(o)]}{w(s_i)} = K[w(s_i)]^{-1} \quad . \qquad \text{(III.B.2-15)}$$

From Eq. (III.B.2-15), it follows that the regression of TTF on stress is inversely proportional to w(s). That is,

$$E(T|s) = K[w(s)]^{-1}$$

for steady stress environments $S(t) = s$. Hence, depending upon how much knowledge is available regarding the mean lives for steady stress environments, it might be possible to use regression techniques to estimate the unknown constants in $w(s)$. It also follows from relation (III.B.2-11) that a similar relationship (proportionality to $[w(s)]^{-1}$) holds between the fractiles of the distributions and the expected values of the order statistics.

III.B.2.3 Estimation and Verification

Consider now the main problems at hand, namely

(i) How is the choice of w to be accomplished?

(ii) Once having postulated w, can the unknown constants in w be estimated and can the function be empirically verified?

This subsection shall deal with the second of the two problems; however, some comments dealing with the first problem are in order.

If a function w for a particular application can be found, then accelerated testing will be feasible. The form of w will generally be chosen through knowledge of the physical mechanisms of failure, physical theories, or empirical results of previous investigators. Once such a function has been chosen, it remains to verify the theoretical model (function) so that an extrapolation of reliability, resulting from data at accelerated test conditions, to reliability at use conditions will indeed be justifiable. The verification of the theoretical model would of course follow the estimation of pertinent unknown constants in the theoretical model.

If the form of $w[S(t)]$ is completely unknown, then a somewhat empirical study can be implemented by conducting k steady stress accelerated pilot tests. From Eq. (III.B.2-15), one might then take a reasonable guess concerning the form of w. Finally using the postulated functional form w, more extensive testing should be conducted in order to more precisely estimate the unknown constants in w and to verify the form postulated from the results of the pilot experiments.

Returning to the second problem, it turns out that in many cases an estimation and verification procedure for the postulated function w does exist. In fact, an estimation and verification procedure exists, with respect to steady stress environments, when the wear rate function, w, has the following property:

for an arbitrary steady stress, there exist functions b_o, b_1, and f such that

$$\log [w(s)] = b_o + b_1 f(s),$$

s is the stress level,

b_o and b_1 are independent of s and are in one-to-one correspondence with the unknown constants of w,

f is independent of those constants and depends only on s.

The procedure is based on the so-called Hill test, which is introduced and described at this point. Proof of the above statement is given after the discussion of the Hill test.

Suppose that it is desired to test whether a random variable Y is related to an experimental variable x in linear fashion, i.e.,

$$H_o: Y = \beta_o + \beta_1 x + \epsilon$$

where ϵ is a random variable, as against the alternative hypothesis that the relationship is either convex or concave, i.e.,

$$H_1: Y = \xi(x) + \epsilon$$

where ξ is a strictly convex (or strictly concave) function. The parameters β_o, β_1 and the function ξ are unspecified. A convex function is pictured in Figure III-4 along with a linear function to portray graphically what the hypotheses indicate. Suppose observations are made on

$$Y_i = f(x_i) + \epsilon_i \quad .$$

Observations under the null hypothesis H_o take the form

$$y_i = \beta_o + \beta_1 x_i + \varepsilon_i \quad .$$

If the ϵ_i's are independently normally distributed with zero means and common variance, then the construction of the test is simply a matter of utilizing known results from regression theory. Unfortunately, it shall be found that the situation encountered will be one where the ϵ_i's in general are distributed independently and non-normally in a continuous cdf with median at zero. The only exception to this rule occurs when TTF, T, is log normally distributed [see Eq. (III.B.2-24)] which occurs when and only when log T is normally distributed. Hill [III-9] treated the non-normal case in great generality. From here on, his test regarding H_o and H_1 above, with observations on

$$Y_i = \beta_o + \beta_1 x_i + \epsilon_i$$

under H_o such that ϵ_i is distributed in a continuous pdf, f_ϵ, with

$$\int_{-\infty}^{0} f_\epsilon(x)dx = F_\epsilon(0) = 1/2$$

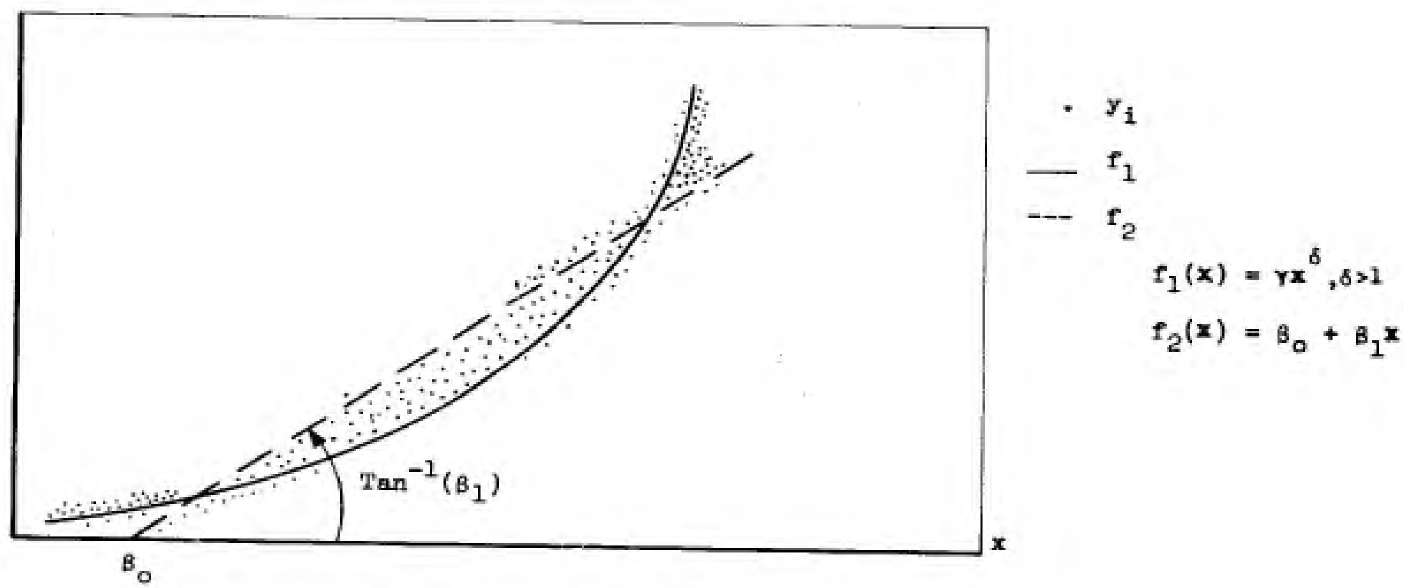


FIG. III-4. ILLUSTRATION FOR HILL TEST

shall be referred to as the Hill test of linearity versus convexity. This test is considered in more detail in Subsection III.D.3.

Now if TTF, T, is distributed in any continuous cdf, F, then it immediately follows that the random variable T^* defined by

$$T^* \equiv \log(T/T_{median})$$

is distributed in a continuous cdf F^* given by

$$F^*(t) = F(T_{median}e^t)$$

and therefore

$$F^*(0) = 1/2.$$

Thus if the postulated functional form w requires that, in some specified type of environment, the TTF random variable be related to environment through the relation

$$\log(T) = \beta_o + \beta_1 x + T^* \qquad \text{(III.B.2-21)}$$

where x is some function of environment only, then it immediately follows that the Hill test of linearity versus convexity applies to Eq. (III.B.2-21) for those environments for which T^* is identically distributed.

The applicability of the Hill test is now investigated. From the definition of the time transformation function, it follows that, for any γ-th fractile of the TTF random variables,

$$T_{u,\gamma} = a_{1u}(T_{1,\gamma}) = a_{2u}(T_{2,\gamma}) = \ldots = a_{ku}(T_{k,\gamma})$$

or equivalently from Eq. (III.B.2-9),

$$h_u(T_{u,\gamma}) = h_1(T_{1,\gamma}) = \ldots = h_k(T_{k,\gamma}) \quad . \qquad \text{(III.B.2-22)}$$

Recall that

$$h_i(t) = \int_0^t w[S_i(x)]dx, \qquad \text{(III.B.2-23)}$$

Thus if S_i and S_j describe the same types of environment (e.g. E_i and E_j are both steady stress, both linearly increasing stress, etc.), then it follows that h_i and h_j will be the same type of functions differing only in known environment scalar description constants (i.e. constants in S_i and S_j). Suppose, without loss of generality, that $h_i, \ldots, h_m$, $m \leq k$ are of the same type (in the sense of the preceding sentence). Now if the $\{h_i\}$ are of the same type, it follows that the $\{h_i^{-1}\}$ are of the same type. From Eq. (III.B.2-22) it follows that

$$T_{i,.50} = h_i^{-1}(k_{.50}) \qquad \text{(III.B.2-24)}$$

where $k_{.50}$ is the common value of Eq. (III.B.2-22) evaluated at $\gamma = .50$ (i.e. the 50-th percentile of the distribution of $\mu(o)$). Let t_{ij} denote the j-th observation in a random sample drawn from F_i (e.g. if n failures from an initial sample of size n are observed in a non-replacement type test under environment E_i and then are randomly permuted, the resulting observations are equivalent to a random sample of size n); then taking logs on both sides of Eq.(III.B.2-24) and adding $\log(t_{ij}/T_{i,.50})$ to both sides yields

$$\log(t_{ij}) = \log[h_i^{-1}(k_{.50})] + \log(t_{ij}/T_{i,.50}) \qquad \text{(III.B.2-25)}$$

Therefore it becomes clear that the wear rate function w indeed satisfies Eq.(III.B.2-21) if h, for the specified type of environment, satisfies the following requirement:

$$\left.\begin{array}{l}\log[h_i^{-1}(k_{.50})] \\ \text{is of the form} \\ \log[h_i^{-1}(k_{.50})] = b_o + b_1 f_i\end{array}\right\} \qquad \text{(III.B.2-26)}$$

where b_o and b_1 are unknown constants in one-one correspondence with the unknown constants in w, and f_i is independent of the unknown constants but dependent on the known constants. When this requirement is satisfied, then Eq. (III.B.2-25) becomes

$$y_{ij} = \beta_o + \beta_1 x_i + \varepsilon_{ij}$$

where

$$\left.\begin{array}{lll}y_{ij} &= \log(t_{ij}) & \\ \beta_o &= b_o & \text{from Eq.(III.B.2-26)} \\ \beta_1 &= b_1 & \text{from Eq.(III.B.2-26)} \\ x_i &= f_i & \text{from Eq.(III.B.2-26)} \\ \varepsilon_{ij} &= \log(t_{ij}/T_{i,.50}) & \end{array}\right\} \text{(III.B.2-27)} \quad .$$

Therefore the Hill test of linearity versus convexity applies for those environments for which $\log(T_i/T_{i,.50})$ is identically distributed. Let ϵ_i denote the random variable $\log(T_i/T_{i,.50})$. It can easily be shown that ϵ_i and ϵ_j are identically distributed if and only if

$$h_i(T_{i,.50}e^{\varepsilon}) = h_j(T_{i,.50}e^{\varepsilon}) \text{ for all } \varepsilon. \qquad \text{(III.B.2-28)}$$

Now for a particular type of environment it follows that the Hill test applies whenever Eq.(III.B.2-26) holds and Eq.(III.B.2-28) holds for all environments of that type. (Note that Eq.(III.B.2-28) holds for all steady stress environments.) Hence the postulated function w can frequently be subjected to a test of validity and the unknown constants in w can be estimated.

In particular, consider the situation where the environments $E_1, \ldots, E_m$ are all of the steady stress variety. In this situation

$$h_i(t) = w(s_i)t \qquad \text{(III.B.2-29)}$$

where s_i is the scalar description of stress at E_i, so that

$$h_i^{-1}(k_{.50}) = \frac{k_{.50}}{w(s_i)} \qquad \text{(III.B.2-30)}$$

From Eq.(III.B.2-26), if

$$\begin{aligned} \log[h_i^{-1}(k_{.50})] &= \log k_{.50} - \log w(s_i) \\ &= b_o + b_1 f_i \end{aligned} \qquad \text{(III.B.2-31)}$$

where b_o and b_1 are unknown constants in one-one correspondence with the unknown constants to be estimated in w, then the Hill test of linearity versus convexity is applicable. From Eq.(III.B.2-31) it follows that if $w(s_i)$ is such that

$$\log w(s_i) = b_o' + b_1' f(s_i) \qquad \text{(III.B.2-32)}$$

where b_o' and b_1' denote functions in a one-one correspondence with the unknown parameters in w and $f(s_i)$ is independent of the unknown parameters, then Eq.(III.B.2-27) takes the form

$$
\left.\begin{aligned}
y_{ij} &= \beta_o + \beta_1 x_i + \epsilon_{ij} \\
\text{where} \quad y_{ij} &= \log(t_{ij}) \\
\beta_o &= \log k_{.50} - b_o = b'_o \\
\beta_1 &= b'_1 \\
x_i &= f(s_i) \\
\epsilon_{ij} &= \log(t_{ij}/T_{i,.50})
\end{aligned}\right\} \qquad \text{(III.B.2-33)}
$$

Hence the Hill test of linearity (linearity implied by the postulated SW wear rate function) versus convexity applies.

Hence if Eq.(III.B.2-32) holds then, by conducting k accelerated tests at steady stress environments, one can perform a test concerning the validity of the postulated functional form w by way of the Hill test of linearity versus convexity. The application of the test requires that certain unknown constants first be estimated. The Mood-Brown estimation procedure [III-10], which provides a necessary ingredient for the Hill test, applies whenever the Hill testing procedure is applicable. Therefore, the above conditions for Hill test applicability are also sufficient conditions for valid estimation.

A general function which may hold for a great many cases is

$$w[S(t)] = A[S(t)]^b \qquad \text{(III.B.2-34)}$$

It turns out that the Hill test is applicable for Eq. (III.B.2-34) for steady stress and some linearly increasing stress environments. It also turns out that the Hill test applies for the exponential and Arrhenius type wear rate functions defined in Section III.C, when the environments are of the steady stress variety. Section III.C can be read at this point, leaving Subsections III.B.3, 4, and 5 for later reading.

III.B.3 Generalized Allen Model

In this section Model 3 of Subsection III.B.1 shall be considered in the same light as Allen [III-3] considered Model 2. Allen made the following simplifying assumptions with respect to the accelerated life testing of capacitors, viz.

(i) TTF under steady stress environments is distributed in the negative exponential distribution (a necessary assumption for Model 2)

(ii) The ratio of the hazard function at the j-th accelerated environment to the hazard function at use conditions is equal to the n-th power of the ratio of the applied voltages, for n a real number, viz.

$$\frac{\phi_j(t)}{\phi_u(t)} = \left[\frac{V_j(t)}{V_u(t)}\right]^n$$

These assumptions are generalized here to some extent. Replace V(t), the applied voltage at time t, by S(t), a general scalar description of stress at time t. Allen's model can thus be generalized by replacing (i) and (ii) by

(1) TTF under steady stress environments is distributed in some not necessarily specified family of cdf's (e.g. - the Weibull family.)

(2)
$$\frac{\phi_j(t)}{\phi_u(t)} = \frac{w^*[S_j(t)]}{w^*[S_u(t)]} \qquad \text{(III.B.3-1)}$$

where w^* is some function.

It is noted that although Eq.(III.B.3-1) is of the form

$$\frac{\phi_j(t)}{\phi_i(t)} = \frac{w^*[S_j(t)]}{w^*[S_i(t)]}$$

this does not force a loss of generality; for, if it is assumed that

$$\frac{\phi_j(t)}{\phi_i(t)} = g[S_j(t),S_i(t)]$$

for some function g, then there exists a function w^* such that

$$g[S_j(t),S_i(t)] = \frac{w^*[S_j(t)]}{w^*[S_i(t)]}$$

The second assumption of the generalized Allen model clearly states that the ratio of the hazard functions at time t for two distinct environments is dependent only on the stresses being applied at time t for the two environments. It should be pointed out that the original model was used with apparent success in the accelerated life testing of mica capacitors, so it is quite possible that the general model could be used with success for the accelerated life testing of some other electrical devices.

Suppose then that some relationship given by Eq. (III.B.3-1) is postulated. Assuming all unknown constants in the relationship can be estimated, then Eq. (III.B.3-1) can be sometimes used to find a time transformation function. It follows from the definition of time transformation function that a_{ji} satisfies the relationship

$$\Phi_j(t) = \Phi_i\left[a_{ji}(t)\right] \quad , \qquad \text{(III.B.3-2)}$$

that is to say

$$\phi_j(t) = \phi_i[a_{ji}(t)] \frac{d[a_{ji}(t)]}{dt} \quad ; \qquad \text{(III.B.3-3)}$$

using Eq.(III.B.3-1) and Eq.(III.B.3-3) it follows that

$$\frac{w^*[S_j(t)]}{w^*[S_i(t)]}\phi_i(t) = \phi_i[a_{ji}(t)] \frac{d[a_{ji}(t)]}{dt} \quad . \qquad \text{(III.B.3-4)}$$

Equation (III.B.3-4) indicates the relation between the time transformation function and the dependence of hazard on environment. No general method of finding a_{ji} via Eq. (III.B.3-4) is apparent. However, if the failure distribution at some environment E_i is exponential, then a_{ji} is given by Eq.(III.B.3-5), derived as follows. By virtue of exponentiality, $\phi_i(t)$ is a constant, say λ, for all values of the argument; hence

$$\phi_i(t) = \phi_i[a_{ji}(t)] \quad , \text{ all } t;$$

furthermore, from $F_j(0) = F_i(0) = 0$ it follows that $a_{ji}(0) = 0$. Therefore integration of Eq.(III.B.3-4) yields, in this special case

$$a_{ji}(t) = \int_0^t \frac{w^*[S_j(x)]}{w^*[S_i(x)]} dx \qquad \text{(III.B.3-5)}$$

Now recall that, in the SW model, the regression of time to failure on values of stress for steady stress environments fully specifies the model. The remainder of this section examines whether a similar statement is correct for the generalized Allen model. In fact, it appears that knowledge of only the regression function is insufficient to fully specify the model, when the model is of the generalized Allen type.

Consider a collection of steady stress environments E_i, such that $S_i(t) = s_i$, and assume that the failure distribution is of the Weibull form for each such environment. Let α_i be the scale parameter and β_i the shape parameter of the failure distribution F_i. From Eq. (III.B.3-1), it follows that

$$\frac{\phi_i(t)}{\phi_j(t)} = \frac{\beta_i t^{\beta_i - 1}/\alpha_i}{\beta_j t^{\beta_j - 1}/\alpha_j} = \frac{\alpha_j}{\alpha_i} \cdot \frac{\beta_i t^{\beta_i - 1}}{\beta_j t^{\beta_j - 1}} = \frac{w^*(s_i)}{w^*(s_j)}$$

Since $\frac{w^*(s_i)}{w^*(s_j)}$ is independent of t, it follows that $\beta_i = \beta$ for all steady stress environments E_i; hence

$$\frac{w^*(s_j)}{w^*(s_i)} = \frac{\alpha_i}{\alpha_j} \quad . \qquad \text{(III.B.3-6)}$$

It is known that the expectation of a Weibull-distributed random variable T, with scale parameter α and shape parameter β, is

$$E(T) = \alpha^{1/\beta}\Gamma(1 + \frac{1}{\beta}) \quad . \qquad \text{(III.B.3-7)}$$

Denote by $E(T \mid s)$ the expectation of TTF in a steady stress environment such that $S(t) = s$. Then, since $\beta_i = \beta$ for all steady stress environments, it follows from Eq.(III.B.3-7) that

$$E(T \mid s_i) = \alpha_i^{1/\beta}\Gamma(1 + \frac{1}{\beta}) \quad . \qquad \text{(III.B.3-8)}$$

Hence, from Eq. (III.B.3-6) and Eq. (III.B.3-8),

$$\frac{w^*(s_j)}{w^*(s_i)} = \left[\frac{E(T \mid s_i)}{E(T \mid s_j)}\right]^{\beta} \quad . \qquad \text{(III.B.3-9)}$$

From Eq. (III.B.3-2), it now follows that

$$\frac{t^{\beta}}{\alpha_i} = \phi_i(t) = \phi_j[a_{ij}(t)] = \frac{[a_{ij}(t)]^{\beta}}{\alpha_j} \quad .$$

Thus, from Eq. (III.B.3-6) and Eq. (III.B.3-9)

$$a_{ij}(t) = \left[\frac{\alpha_j}{\alpha_i}\right]^{1/\beta} t = \frac{E(T|s_j)}{E(T|s_i)}\, t \quad . \qquad \text{(III.B.3-10)}$$

Equation (III.B.3-10), which yields the time transformation function in terms of the regression function, was derived under the specific assumption that the TTF is Weibull-distributed at all steady stress environments. In view of results for the SW model, one might conjecture that the Eq. (III.B.3-10) is, in fact, valid for any form of steady-stress failure distribution and any given regression. This conjecture is false, and is disproved in the next paragraph.

First note that, for steady stress environments, Eq. (III.B.3-1) becomes

$$\frac{\phi_i(t)}{\phi_j(t)} = \frac{w^*(s_i)}{w^*(s_j)} \quad ;$$

hence

$$\frac{\Phi_i(t)}{\Phi_j(t)} = \frac{w^*(s_i)}{w^*(s_j)}$$

and, by using Eq.(III.B.3-2), one obtains

$$\frac{w^*(s_i)}{w^*(s_j)} \phi_j(t) = \phi_j[a_{ij}(t)]$$

Denote $E(T \mid s_i)/E(T \mid s_j)$ by C_1, and denote $w^*(s_i)/w^*(s_j)$ by C_2. Then the conjecture of the above paragraph is equivalent to conjecturing that

$$C_2\Phi_j(t) = \Phi_j(C_1 t) \quad \text{for all } t \qquad \text{(III.B.3-11)}$$

regardless of the form of F_j. As a counterexample to the conjecture, consider the failure distribution F with hazard function $\phi(t) = a + 2bt$, $a \neq 0$ and $b \neq 0$. Then Eq.(III.B.3-11) requires that

$$C_2[at + bt^2] = aC_1 t + bC_1^2 t^2 \quad \text{for all } t$$

which is satisfied if and only if

$$\left[\frac{C_2}{C_1} - 1\right] \cdot at + \left[\frac{C_2}{C_1} - C_1\right] \cdot bt^2 = 0$$

which, in turn, is satisfied if and only if

$$C_1 = C_2 = 1$$

since $a \neq 0$ and $b \neq 0$. But, $C_1 = 1$ means that the expected time to failure is independent of the stress level, and $C_2 = 1$ means that the failure distribution is independent of the stress level. Thus it is seen that the conjecture does not hold for arbitrary failure distributions and arbitrary, non-trivial, regression.

III.B.4 <u>SW vs Generalized Allen Model</u>

The model due to Allen [III-3], generalized in Subsection III.B.3, and the SW model developed in Subsection III.B.2 are inherently different with respect to their basic structures. A comparison of these two models therefore deserves a fairly general treatment and comprises the substance of this Subsection. The comparison is based on a study of the distinguishability of the two models. That is, could either model be used to account for a given form of dependence of failure behavior on environment? It is shown that the two models are indistinguishable if and only if the failure distribution under steady stress is exponential. Thus, in that case, it is immaterial whether one works within the framework of the SW model or of the (original) Allen model. Furthermore, if one has a non-exponential failure distribution at steady stress, one can decide (on the basis of observations) which of the two models applies - if one a priori restricts the choice of models to those two.

Let the dependence of failure behavior on environment be specified by stating the hazard function for all possible environments. The proof is based on the consideration of two arbitrary environments, say E_i and E_j; without loss of generality, one of them, say E_j, can be taken to be a steady stress environment. Let S_i be the scalar description of E_i and let the steady stress under E_j be s_j. Let the hazard functions under E_i and E_j be ϕ_i and ϕ_j respectively. The proof is developed by considering the relation which must hold between ϕ_i and ϕ_j if the model is the SW model, and the relation which must hold if the model is the generalized Allen model.

If the model is the generalized Allen model, it follows from Eq.(III.B.3-1) that

$$\phi_i(t) = \frac{w^*[S_i(t)]}{w^*(s_j)} \phi_j(t) \quad . \qquad \text{(III.B.4-1)}$$

If the model is the SW model, it follows from Eq.(III.B.2-8) and Eq.(III.B.2-10) that

$$\phi_i(t) = \phi_j\left[\frac{h_i(t)}{w(s_j)}\right] \frac{w[S_i(t)]}{w(s_j)} \qquad \text{(III.B.4-2)}$$

where

$$h_i(t) = \int_0^t w[S_i(x)]dx \quad .$$

No., the dependence of failure behavior on environment is explainable by either model if and only if there exist functions w* and w such that the two above expressions for ϕ_i coincide.

That is, the two models are indistinguishable if and only if

$$\frac{w^*[S_i(t)]}{w^*(s_j)} \phi_j(t) = \left[\frac{w[S_i(t)]}{w(s_j)}\right] \phi_j\left[\frac{h_i(t)}{w(s_j)}\right] \quad . \qquad \text{(III.B.4-3)}$$

If the model is the generalized Allen model, $\phi_j(t)$ must be independent of past stress history, i.e. of the values of $S_j(\tau)$ for $\tau<t$. For the two models to be indistinguishable, $\phi_j(t)$ must also have this property if viewed in terms of the SW model. For the SW model, it was shown in Subsection III.B.2.2 that a HF obeying the SW model is independent of previous stress history when the distribution of TTF is exponential under steady stress environments. That is, the condition of exponentially distributed TTF is

sufficient for the hazard function to be independent of previous stress history. It shall now be shown that the same condition is also necessary.

For a non-steady stress environment, $h_i(t)$ depends on previous stress history, except in the trivial (and hereafter excluded) case in which $w[S_i(t)]$ is independent of the stress $S_i(t)$. But, if $h_i(t)$ is dependent on previous stress history, then it follows from Eq.(III.B.4-2) that $\phi_i(t)$ is independent of previous stress history if and only if

$$\phi_j[\frac{h_i(t)}{w(s_j)}]$$

is independent of $h_i(t)$. However $\phi_j[\frac{h_i(t)}{w(s_j)}]$ is independent of $h_i(t)$ if and only if ϕ_j is independent of its argument, in which case ϕ_j is a constant function in time. Since E_j is a steady stress environment, the exponential distribution of TTF under steady stress is shown to be a necessary condition for $\phi_i(t)$ to be independent of previous stress history.

Therefore, to ascertain when the two models coincide, one only needs to consider the case where TTF under E_j is distributed in the exponential distribution. In this situation Eq.(III.B.4-1) becomes

$$\phi_i(t) = K_j^* w^*[S_i(t)] \quad , \qquad \text{(III.B.4-4)}$$

where

$$K_j^* \equiv \phi_j(t)/w^*[s_j]$$

is constant for all t since ϕ_j is constant.

Moreover, Eq.(III.B.4-2) becomes

$$\phi_i(t) = K_j w[S_i(t)] \tag{III.B.4-5}$$

where

$$K_j \equiv \phi_j[\frac{h_i(t)}{w(s_j)}]/w(s_j)$$

From Eq.(III.B.4-4) and Eq.(III.B.4-5) it follows that the two models coincide only when $w = w^*$, which proves the following: necessary and sufficient conditions for the SW and generalized Allen models to coincide are

(a) time to failure under fixed stress environments is distributed in the negative exponential distribution

and

(b) the stress-time functions w and w* under the two models are the same function.

Note that the following situation may occur and, in fact, does occur when the steady-stress failure distribution is Weibull. The SW and generalized Allen models may be indistinguishable if one is restricted to steady stress environments, though they have different implications for non-steady-stress environments. That is, though either model may be used to account for steady-stress behavior, the predictions of non-steady stress behavior are then different, depending on which model is chosen. Hence one must be very careful to choose the correct model, when one or the other of these two models applies.

III.B.5 Age-Wear Models

As mentioned in Subsection III.B.1 Mercer [III-5] considered a case of Model 6 of Subsection III.B.1 in great generality. Other wear-dependent statistical models have been developed and analyzed by Mercer and Smith [III-11], and Birnbaum and Saunders [III-12]. In particular, Mercer and Smith developed a model for the wear of conveyor belting which had the ability to assess the relative importance of discrete blows on continuous wear. Birnbaum and Saunders were able to theoretically explain the different life distributions which are commonly assumed in failure rate analyses through the consideration of fatigue dependent hazard function models.

This section contains some brief comments concerning age-wear models. The age-wear specific hazard function is customarily defined as

$$\phi(x,w) = \lim_{\Delta x \to 0} \left\{ \frac{\Pr[x < X \leq x + \Delta x \mid X > x, D(x) = d]}{\Delta x} \right\} \quad \text{(III.B.5-1)}$$

where D(x) is the value of the damage or wear variable at time x. Under the SW model discussed in Subsection III.B.2, D is a deterministic function of the stress history, but the total amount of wear required to induce failure is a random variable.

When D(x) is a stochastic process of specified structure and (III.B.5-1) is given, then a probabilistic model of wear and failure emerges. Unfortunately when D(x) is a stochastic process the probabilistic model is very difficult to apply in the field of endeavor known as accelerated life testing. In particular, one must understand how the stochastic process behaves when more severe stress environments are applied. Suppose however that wear is produced by a series of "blows" occurring in a Poisson process. Suppose further that the wear at the i-th blow is a positive random variable D_i, the sequence $\{D_i\}$ being independent identically distributed random variables, independent of the process. If wear is additive, then a cumulative wear process (i.e. damage accumulation) results. A treatment of this special case can be found in Cox [III-13], p. 95. Further information concerning processes of this type can be found in Cox [III-13], p. 115.

The most general case treated in the literature is due to Mercer. Mercer's work is far too complicated to be discussed here but a short summary of Mercer's model is given in the discussion of Model 6 in Subsection III.B.1.

III.C SPECIAL CASES OF THE SW MODEL

III.C.1 Introduction

In this section some general wear rate functions for the SW model shall be considered. In particular, given a postulated wear rate function, how does one estimate the unknown constants in the function, and how does one then verify the postulated form (i.e., statistically test its validity)? (It should be recognized by the reader that the results sought and the various functional forms studied have been chosen due to the strong influence of empirical results of previous investigators, and in view of the physical mechanisms of failure observed for various mechanical, electromechanical and electrical devices.) These topics shall be treated in detail for the four wear rate functions given below.

In Subsection III.C.2 a functional form for the SW wear rate function given by

$$w[S(t)] = a[S(t)]^{b}$$

shall be considered. This wear rate function has been proposed for the study of TTF for capacitors, rolling element bearings, etc.

In Subsection III.C.3 a functional form for the SW wear rate function given by

$$w[S(t)] = a \exp[bS(t)]$$

shall be considered.

In Subsection III.C.4 a functional form for the SW wear rate function given by

$$w[S(t)] = K_1 \exp[-K_2/S(t)]$$

shall be considered. This function is closely akin to the well known Arrhenius chemical reaction rate model.

Most of the detailed emphasis will be plased in the derivation of the results in Subsection III.C.2 and thereafter the results shall be stated without derivation. To derive the results in Subsections III.C.3 and III.C.4 one merely needs to follow the procedure used in Subsection III.C.2 which shall be delineated in detail in this section. To avoid possible confusion, the following general terminology and methodology (introduced in Section III.B) shall be restated.

For a general environment E defined by the set of scalar descriptions of stress at time t, $\{S(t), t > 0\}$, the function h_E is defined as

$$h_E(t) = \int_0^t w[S(x)]dx. \tag{III.C.1-1}$$

Of course, the environment alone does not fully specify the function h; it also depends on the wear rate function which governs the effect of the environment on wear. However, this dependence of h on w will not be explicitly denoted. Three types of wear rate functions are considered in Subsections III.C.2, III.C.3, and III.C.4, one type in each subsection; throughout a subsection, h_E will denote the function h specified by the environment E and by a wear rate function of the type considered in that subsection.

If one is considering some collection of environments, each identified by a particular value of the indexing variable i, then the symbol h_i shall be used to denote the h_E function corresponding to the environment E_i. As previously, i=u will denote the use environment. In Subsections III.C.2 through III.C.4 three types of environment (defined precisely in Subsection III.C.2) are of particular interest: steady stress environments, linearly increasing environments, and step-stress environments. The indexing variable i will take on values C, L, and I, respectively, to denote arbitrary environments of such type.

Now for a specified environment E, there corresponds to any given item the pair of real numbers (μ^*, t) where μ^* is the realization of the random variable $\mu(o)$ for that item, and t is the corresponding realization of the TTF random variable T under the environment E for that item. Equation (III.B.2-4) states that

$$\mu^* = h_E(t)$$

which, in terms of the random variables $\mu(o)$ and T, means that

$$\mu(o) = h_E(T)$$

Thus the necessary framework is established for the calculation of the functions h_E and h_E^{-1} when the environment is completely specified. The procedure initially alluded to, which could be used to investigate other types of wear rate functions, is the following:

(i) Enumerate the types of environments (e.g., steady stress, linearly increasing stress, step-stress) to be considered.

(ii) Calculate the functions h, defined by Eq.(III.C.1-1), for environments of each type enumerated in (i); that is, calculate h_E, where h_E is an arbitrary environment of the specified type.

(iii) Calculate the function h_u corresponding to the use conditions environment E_u as well as its inverse h_u^{-1}.

(iv) Calculate the time transformation functions $\{a_{iu}\}$ which, from Eq.(III.B.2-9), are known to be

$$a_{iu} = h_u^{-1} h_i \quad ,$$

that is to say,

$$a_{iu}(t) = h_u^{-1}[h_i(t)] \quad .$$

(v) Check to see whether the Hill test of linearity versus convexity is applicable for those types of environments (see (i)) which shall be considered. This investigation can be carried out by applying the techniques in Subsection III.B.2-3. In particular, the discussion starting with the paragraph containing Eq.(III.B.2-23) and ending with the paragraph containing Eq.(III.B.2-28) gives a general procedure from which the investigation can be implemented. If all environments considered are steady stress environments, then the procedure is greatly simplified, and if Eq.(III.B.2-23) and (III.B.2-27) are replaced by Eq.(III.B.2-29) and (III.B.2-33), respectively, in the above sentence, the sentence holds for the simplified procedure based on steady stress environments. If the Hill test can be applied, then test(s) for the validity of the postulated function w result, along with estimates of the unknown constants in w. In particular, it is shown in this section that the Hill test applies for steady stress environments, linearly increasing stress environments, but not for step stress environments when the wear rate function is given by

$$w[S(t)] = \vec{a}[S(t)]^b$$

Similar results are stated, without proof, for two other types of wear rate functions.

(vi) If the Hill test does not apply, search for alternate tests and estimation procedures. New estimation procedures must be found as well as alternate tests since the assumptions that must be satisfied to make the Hill test applicable are the same assumptions needed for the valid utilization of the estimation techniques used in conjunction with the Hill test.

(vii) If the tests of linearity versus convexity are not rejected (it is likely that not more than two such tests would be applied in practice), the data can then be extrapolated, based on the estimated time transformation function, to the use conditions environment and a K-sample test can be used to partially test for the validity of the extrapolation.

This procedure gives an outline for the estimation and verification techniques for a postulated wear rate function.

III.C.2 The Power Wear Rate Function

In this subsection the form of the SW wear rate function is assumed to be

$$w[S(t)] = a[S(t)]^{b} \quad , \quad b>0$$

with a and b not known a priori. Here only three types of stress functions shall be considered, namely those corresponding to steady stress environments, linearly increasing stress environments, and step stress environments. In situations where an upper bound on applied stress exists(e.g., a stress tolerance limit determined by design characteristics of the devices) and an increasing stress test is desirable, a stress function should be used which assymptotically in time approaches some value less than the upper bound. This topic shall be discussed in a later paragraph. It shall be found that a stress function of the step stress variety has inherent limitations with respect to its application in the SW model. Let i = C denote an arbitrary steady (constant) stress environment, i = L an arbitrary linearly increasing stress environment and i = I an arbitrary step stress (incremental) environment. Thus h_C, H_L, and h_I denote functions h_i for typical environments E_C, E_L, and E_I, respectively. Similarly let T_C, T_L, and T_I denote the TTF random variables under environments E_C, E_L, E_I, respectively. The stress functions S_C, S_L, and S_I are formally defined as follows:

$$S_C(t) = s_o \quad \text{for all } t$$

and

$$S_L(t) = \delta t, \ \delta \text{ fixed and known}$$

$$S_I(t)=\begin{cases} s_1 & \tau_o = 0 \le t < \tau_1 \\ s_2 & \tau_1 \le t < \tau_2 \\ \vdots & \vdots \\ s_K & \tau_{K-1} \le t < \tau_K \\ s_{K+1} & \tau_K \le t < \tau_{K+1} = \infty, \ \tau\text{'s fixed and known.} \end{cases}$$

$$
S_I(t) = \begin{cases} s_1 & \tau_o = 0 \le t < \tau_1 \\ s_2 & \tau_1 \le t < \tau_2 \\ \vdots & \vdots \\ s_K & \tau_{K-1} \le t < \tau_K \\ s_{K+1} & \tau_K \le t < \tau_{K+1} = \infty, \ \tau\text{'s fixed and known.} \end{cases}
$$

Graphically, the step stress function might appear as given in Figure III-5.

Throughout this section the use conditions environment E_u shall be assumed, without loss of generality, to be a steady stress environment with scalar description s_u. It is recalled that step (iii) of the procedure outlined in Subsection III.C.1 required the inverse function h_u^{-1} to be calculated. Since the use conditions environment might not actually be a steady stress environment, the inverse functions h_C^{-1}, h_L^{-1} and h_I^{-1} shall be calculated and given in Eq.(III.C.2-2). The inverse functions need to be calculated when E_u is not a steady stress environment. Using the appropriate expression for the actual h_u^{-1}, the equations (III.C.2-3) can be rederived. Using Eq.(III.C.1-1) the functions h_C, h_L and h_I shall now be calculated and their respective inverses shall be given.

Since E_u is assumed to be a steady stress environment with scalar description of stress given by s_u, it follows that

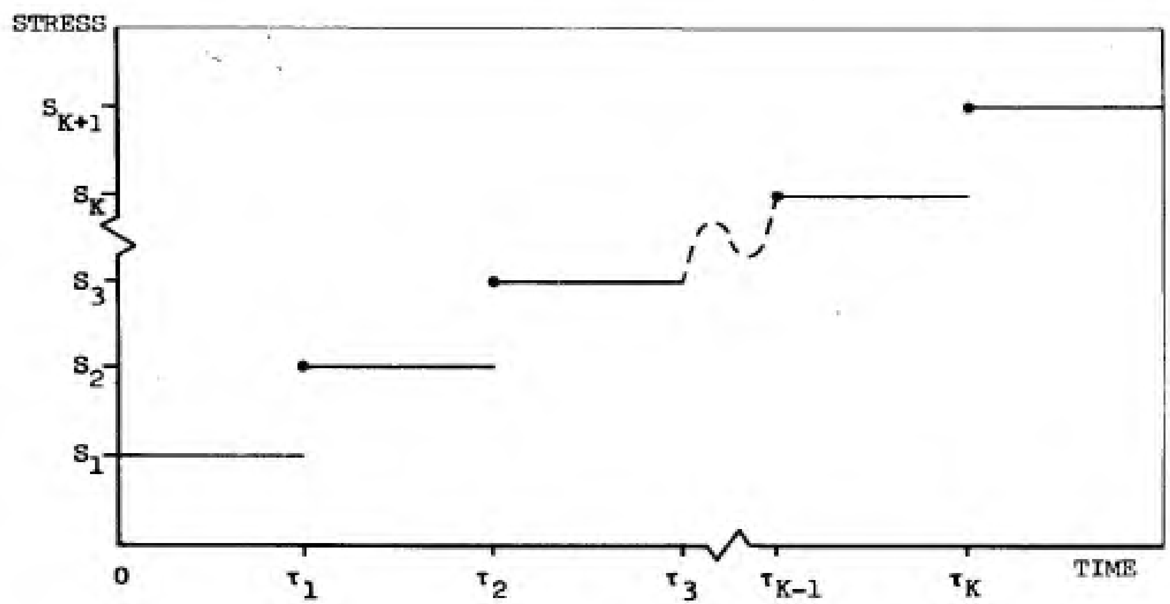


FIG. III-5. ILLUSTRATION OF STEP STRESS ENVIRONMENT

$$h_u(t) = a \cdot s_u^b t \quad ,$$

and hence

$$h_u^{-1}(\mu^*) = \frac{\mu^*}{a \cdot s_u^b} \quad .$$

Now from step (iv) one calculates the time transformation functions a_{Cu}, a_{Lu}, and a_{Iu} respectively. That is to say

$$a_{\eta u}(t) = h_u^{-1}[h_\eta(t)] = \frac{h_\eta(t)}{a \cdot s_u^b} \qquad , \eta = C, L, \text{ or } I$$

so that the time transformation functions are given by the Equations (III.C.2-3).

Equations (III.C.2-1)

$$h_C(t) = \int_0^t w[S_C(x)]\,dx = \int_0^t a s_o^b\,dx$$

$$= a s_o^b t$$

$$h_L(t) = \int_0^t w[S_L(x)]dx = \int_0^t a(\delta x)^b dx$$

$$= \frac{a\delta^b}{b+1} t^{b+1}$$

$$h_I(t) = \int_0^t w[S_I(x)]dx$$

$$= \sum_{i=o}^{j-1} \left[\int_{\tau_i}^{\tau_{i+1}} a(s_{i+1})^b dx\right] + \int_{\tau_j}^{t} a(s_{j+1})^b dx$$

$$= \sum_{i=o}^{j-1} a\, s_{i+1}^b(\tau_{i+1} - \tau_i) + a\, s_{j+1}^b(t - \tau_j) \quad \text{for } \tau_j \leqq t < \tau_{j+1}$$

$$j = 0, 1, 2, \ldots, K$$

where $\sum_{i=o}^{j-1} a\, s_{i+1}^b\,(\tau_{i+1}-\tau_i) = 0$ for $j = 0$.

Equations (III.C.2-2)

$$h_C^{-1}(\mu^*) = \mu^*/(a\ s_o^b)$$

$$h_L^{-1}(\mu^*) = \left[\frac{(b+1)\mu^*}{a\delta^b}\right]^{\frac{1}{b+1}}$$

$$h_I^{-1}(\mu^*) = \frac{\mu^* - \sum_{i=o}^{J-1} a\ s_{i+1}^b\ (\tau_{i+1} - \tau_i)}{a\ s_{J+1}^b} + \tau_J$$

where J is the unique integer such that

$$\sum_{i=o}^{J-1} a\ s_{i+1}^b(\tau_{i+1}-\tau_i) \leqq \mu^* < \sum_{i=o}^{J} a\ s_{i+1}^b\ (\tau_{i+1}-\tau_i) \text{ and}$$

where $\sum_{i=o}^{J-1} a\ s_{i+1}^b\ (\tau_{i+1}-\tau_i) = 0$ if $J = 0$.

Equations (III.C.2-3)

$$a_{Cu}(t) = c_1 t$$

$$a_{Lu}(t) = c_2 t^{b+1}$$

$$a_{Iu}(t) = \begin{cases} d_1 t & , \ o \leq t < \tau_1 \\ d_2 + d_3 t & , \ \tau_1 \leq t < \tau_2 \\ \cdots\cdots & \quad \cdots\cdots \\ d_{2K-2} + d_{2K-1} t & , \ \tau_{K-1} \leq t < \tau_K \\ d_{2K} + d_{2K+1} t & , \ \tau_K \leq t < \infty \end{cases}$$

where

$$c_1 = \left[\frac{s_o}{s_u}\right]^b$$

$$c_2 = \frac{1}{b+1} \left[\frac{\delta}{s_u}\right]^b$$

$$d_1 = \left[\frac{s_1}{s_u}\right]^b$$

$$d_2 = \tau_1 \left[\left(\frac{s_1}{s_u}\right)^b - \left(\frac{s_2}{s_u}\right)^b\right], \quad d_3 = \left(\frac{s_2}{s_u}\right)^b$$

$$d_4 = d_2 + \tau_2 \left[\left(\frac{s_2}{s_u}\right)^b - \left(\frac{s_3}{s_u}\right)^b\right], \quad d_5 = \left(\frac{s_3}{s_u}\right)^b$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$d_{2K-2} = d_{2K-4} + \tau_{K-1} \left[\left(\frac{s_{K-1}}{s_u}\right)^b - \left(\frac{s_K}{s_u}\right)^b\right], \quad d_{2K-1} = \left(\frac{s_K}{s_u}\right)^b$$

$$d_{2K} = d_{2K-2} + \tau_K \left[\left(\frac{s_K}{s_u}\right)^b - \left(\frac{s_{K+1}}{s_u}\right)^b\right], \quad d_{2K+1} = \left(\frac{s_{K+1}}{s_u}\right)^b$$

Thus it is known that

$$\left.\begin{aligned} F_C(t) &= F_u[a_{Cu}(t)] \\ F_L(t) &= F_u[a_{Lu}(t)] \\ F_I(t) &= F_u[a_{Iu}(t)] \quad . \end{aligned}\right\} \quad \text{(III.C.2-4)}$$

Now, given a sample of n failures in a test under environment E_L for example, a corresponding sample with identical properties under E_u is obtained by utilizing the time transformation function a_{Lu}, from environment E_L to environment E_u, given in Equations (III.C.2-3).

It is to be noted that all of the time transformation functions in Eq. (III.C.2-3) are independent of the constant "a" in w, hence the constant "b" in w is the only constant needed to be estimated.

Suppose now that $m_1 > 1$ steady stress environments were used for accelerated test environments and were denoted by $E_{C1}, E_{C2}, \ldots, E_{Cm_1}$, with scalar descriptions of stress given by $s_{o1}, \ldots s_{om_1}$, respectively. Now

$$\log w(s_{oi}) = \log a + b \log (s_{oi}) \quad ;$$

hence from Eq. (III.B.2-32) and Eq. (III.B.2-33) it follows that the Hill test of linearity versus convexity can be applied to the following relation

$$y_{ij} = \beta_0 + \beta_1 x_i + \epsilon_{ij} \qquad \text{(III.C.2-5)}$$

where

$y_{ij} = \log (t_{ij})$

t_{ij} = the j^{th} observation in a random sample of size n_i from the distribution of TTF under environment E_{Ci}

$\beta_0 = \log (K_{.50}/a)$

$\beta_1 = -b$

$x_i = \log (s_{oi})$

$\epsilon_{ij} = \log (t_{ij}/T_{i,.50})$,

$T_{i,.50}$ = the median of distribution of TTF under E_{Ci}

$K_{.50}$ = the median of the distribution of the random variable $\mu(0)$.

Note that, with ϵ_i denoting the random variable $\log (T_i / T_{i,.50})$

$\Pr[\epsilon_i > 0] = 1/2$

$\Pr[\epsilon_i < 0] = 1/2$

so that the ϵ_i are distributed in a continuous cdf with median at zero. Hence all the assumptions needed for the application of the Hill test of linearity versus convexity are satisfied. In passing, it is remarked that a censored sample of times to failure invalidates the test, for then t_{ij} can no longer be thought of as a random observation.

Now suppose there are $m_2>1$ linearly increasing stress environments denoted by E_{L1}, E_{L2}, ... E_{Lm_2} described by rate of increase constants δ_1, δ_2, ...,δ_{m_2} respectively. From Equations (III.C.2-1) it follows that $T_{i,\gamma}$, the γ^{th} percentile of the distribution of TTF under E_{Li} is related to $T_{j,\gamma}$, the γ^{th} percentile of distribution of TTF under E_{Lj} by

$$\frac{a\,\delta_i^b}{b+1}\,T_{i,\gamma}^{b+1} = \frac{a\,\delta_j^b}{b+1}\,T_{j,\gamma}^{b+1} .$$

For the 50^{th} percentile (the median) it follows that

$$\frac{a\,\delta_i^b}{b+1}\,T_{i,.50}^{b+1} = \frac{a\,\delta_j^b}{b+1}\,T_{j,.50}^{b+1} = K_{.50} \qquad \text{(III.C.2-6)}$$

where $K_{.50}$ is the median of distribution of $\mu(0)$. Hence it follows that

$$T_{i,.50} = \left[\frac{(b+1)K_{.50}}{a\delta_i^b}\right]^{1/b+1} .$$

for all i. Note that this equation is equivalent to the function h_L^{-1} given in Equations (III.C.2-2) with μ^* replaced by $K_{.50}$ and δ replaced by δ_i . From this result it follows that

$$\log T_{i,.50} = \frac{1}{b+1}\log\left[\frac{(b+1)K_{.50}}{a}\right] - \frac{b}{b+1}\log\delta_i .$$

Since the only constant needed to be estimated is the constant b, then from Eq. (III.B.2-26) it again follows that the Hill test of linearity versus convexity is applicable. In particular the equation given by

$$y_{ij} = \beta_o + \beta_i x_i + \varepsilon_{ij}$$

satisfies the requirements for the Hill test to be applicable where

$$y_{ij} = \log(t_{ij})$$

t_{ij} = the j^{th} observation in a random sample of size n_i from the distribution of TTF under environment E_{Li}

$$\beta_0 = \frac{1}{b+1} \log \left[\frac{(b+1)K_{.50}}{a} \right]$$

$$\beta_1 = - \frac{b}{b+1}$$

$$x_i = \log \delta_i$$

$$\varepsilon_{ij} = \log(t_{ij}/T_{i,.50})$$

The last remark in the paragraph following Eq. (III.C.2-5) also applies here.

It is obvious from the form of h_I^{-1} that such a procedure does not exist for the step stress environments. In fact the Equations (III.C.2-1) and (III.C.2-2) for the step stress environment are so cumbersome, that it is very difficult to even imagine the utilization of step stress tests in accelerated life testing.

Suppose, however, that one wishes to conduct some sort of increasing stress test but the stress must at all times remain below some stress S*. Certainly step stress testing would satisfy this requirement, but the data would not be usable; hence, a step stress should not be used.

As to linearly increasing stress functions, one cannot guarantee that they will not increase past the critical level S* before the item fails. Therefore the use of stress functions such as

$$S(t) = \left[1-e^{-t/\tau}\right]S^*$$

or

$$S(t) = \frac{t}{t+1} S^* \quad , \text{etc.}$$

should be further considered.

III.C.3 The Exponential Wear Rate Function

In this section a SW wear rate function given by

$$w[S(t)] = a \exp[bS(t)] \quad , \quad a>0,\ b>0 \qquad \text{(III.C.3-1)}$$

shall be studied. Estimation and verification techniques for this type of SW wear rate function shall be given although the details of derivation shall be omitted. The procedure used to obtain the results in this section was delineated in Subsection III.C.1. The notation for steady stress environments and linearly increasing stress environments developed in the previous section shall be utilized here. The functions h_C, h_C^{-1}, h_L, and h_L^{-1} are calculated through the use of Eq. (III.C.1-1) and are given here.

$$\left.\begin{aligned} h_C(t) &= at\cdot[\exp(bs_o)] \\ h_L(t) &= \frac{a}{b\delta}\cdot[\exp(b\delta t) - 1] \\ h_C^{-1}(\mu^*) &= \mu^*/[a \exp(bs_o)] \\ h_L^{-1}(\mu^*) &= \frac{\ln[1 + \frac{b\delta}{a}\mu^*]}{b\delta} \end{aligned}\right\} \qquad \text{(III.C.3-2)}$$

From Eq. (III.C.3-2) it is quite easy to derive the time transformation functions. In particular, if E_u is a steady stress environment with scalar description of stress s_u, it follows that

$$\left.\begin{aligned} a_{Cu}(t) &= C_1 t \\ a_{Lu}(t) &= C_2 + C_3 \exp[b\delta t] \\ \text{where } C_1 &= \exp[b(s_o - s_u)] \\ C_2 &= -[\exp(-bs_u)]/b\delta \\ C_3 &= -C_2 \quad . \end{aligned}\right\} \qquad \text{(III.C.3-3)}$$

Again utilizing Eq. (III.B.2-26) and Eq. (III.B.2-27) it follows for K_1 steady stress environments with failure distribution functions F_i and scalar descriptions of stress s_i, $i = 1, 2, \ldots, K_1$, that

$$\log(t_{ij}) = -bs_i + \log C + \log(t_{ij}/T_{i,.50})$$

where t_{ij} = the j^{th} observation in a random sample of size n drawn from F_i, $i = 1, 2, \ldots, K_1$ (III.C.3-4)

$$\log C = \frac{\text{median } [\mu(o)]}{a}$$

and $T_{i,.50} = \text{median } (T|E_i)$.

Thus Eq. (III.C.3-4) again satisfies the postulates for the Hill test of linearity versus convexity.

Unfortunately the linearly increasing stress environments do not possess the property that

$$\log(w[S(t)]) = b_0 + b_1[f\ S(t)] \quad ,$$

so that the Hill test does not apply here.

III.C.4 The Arrhenius Wear Rate Function

The Arrhenius wear rate function is defined as

$$w[S(t)] = ae^{-b/S(t)} \quad , \quad a>0,\ b>0 \quad . \qquad \text{(III.C.4-1)}$$

For steady stress environments the time transformation functions take the form

$$a_{Cu}(t) = \exp[b(1/s_u - 1/s_C)]t \quad , \qquad \text{(III.C.4-2)}$$

where s_C is the stress value at E_C.

From Eq. (III.C.4-1) it follows that Eq. (III.B.2-26) is satisfied, namely

$$\log w(s) = \log a - b/s \quad .$$

Hence the Hill test of linearity versus convexity again can be utilized to estimate the unknown constant in the time transformation functions and to test the hypothesis that Eq. (III.C.4-1) is indeed the correct function.

III.D VERIFICATION TECHNIQUES

III.D.1 Introduction

The previous sections of Part III dealt with problems concerning the postulation of time transformation functions, the estimation of unknown constants in postulated time transformation functions, and the verification of postulated time transformation functions. Here the following problem shall be considered: Given a postulated time transformation function, how does one partially justify an extrapolation of data from accelerated conditions to use conditions?

In the notation heretofore utilized, let F_i denote the failure distribution at the i^{th} accelerated environment and let a_{ju} denote the time transformation function from environment E_j to use conditions. Consider extrapolating the data from each accelerated life test through the appropriate time transformation function. That is, transform the sample obtained from F_i through the time transformation function a_{ju}. Now if the time transformation function is indeed valid, then each "equivalent" transformed sample at use conditions possesses the same statistical properties as the original sample which generated it. For example if a random sample was drawn from F_i then the resulting equivalent sample at use-conditions is a random sample from F_u. Thus if K equivalent random samples are obtained at use conditions, by transformation from K accelerated tests, it follows that a K-sample test such as discussed in Subsection III.D.2 could be utilized.

However, the time transformation functions used to extrapolate the data will use constants estimated from the data. Thus, even if the time transformation function is valid in form, random errors in the estimates of the constants will cause greater dispersion than might be expected if the constants were known exactly. At any rate, if the time transformation functions have been correctly chosen and reasonably precise estimates of the unknown constants have been made, then one would not suspect a K-sample test to be rejected. On the other hand, if the time transformation functions have been incorrectly chosen one would suspect that the K-sample test would be rejected.

The above procedure does not prove that extrapolation from accelerated tests to use conditions is valid. However, it does test whether the time transformation function provides mutually consistent extrapolations from various accelerated test environments. Certainly extrapolation is not valid if such consistency is lacking. Therefore, the K-sample tests will detect some cases where extrapolation is definitely invalid.

III.D.2 <u>Goodness of Fit and K-Sample Tests</u>

The problem of testing the hypothesis that a sample of observations has been drawn from a specified population, or that K samples ($K \geq 2$) have been drawn from the same population, has been encountered frequently in this report (in particular see Sections III.B. and III.C). Here reference to some test statistics and publications of corresponding tables, shall be given. Emphasis will not be placed on the chi-square goodness of fit statistic due to the existence of better test statistics (better in the sense of power, tabulations of the statistic for finite sample sizes, etc.).

Goodness of Fit Tests. The first of the tests considered shall be the Cramer-von Mises test. Let $\hat{F}$ denote the empirical distribution function estimating F. The test is based on a statistic developed by N. V. Smirnov given by

$$W_n^2 = n \int_{-\infty}^{+\infty} [\hat{F}(x) - F(x)]^2 \cdot \Psi[F(x)]\, dF(x)$$

where $\Psi[F(x)]$ is a weight function (≥ 1) to be selected according to some desirable property. A summary of Smirnov's work can be found in [III-14]. The weight function is considered further in the next paragraphs.

A different approach was studied by A. N. Kolmogorov in which the test statistic used was the maximum distance between F and $\hat{F}$ based on the data. The test statistic was defined by

$$D_n = \underset{-\infty<x<+\infty}{\text{maximum}} |\hat{F}(x) - F(x)| \quad . \qquad \text{(III.D.2-1)}$$

The D_n test statistic was generalized, by the introduction of a weight function, in 1952 by Anderson and Darling[III-15]. This test statistic is defined as follows

$$D_n^* = (\sqrt{n}) \underset{-\infty<x<+\infty}{\text{maximum}} \{ |\hat{F}(x) - F(x)| \Psi[F(x)] \} \ . \qquad \text{(III.D.2-2)}$$

It is noted that if

$$\Psi[F(x)] = 1 \quad \text{for all } x$$

emphasis is placed on the central portion of the distribution, for

$$\Psi[F(x)] = \frac{1}{F(x)[1 - F(x)]}$$

emphasis is placed in the tails of the distribution, and for

$$\Psi[F(x)] = \frac{1}{F(x)}$$

Emphasis is placed in the lower tail of the distribution. Life testing problems many times require investigation of the lower tail of the distribution. In this situation one is advised to use the latter form of the weight function. The distribution of Eq. (III.D.2-2) for finite sample size has been tabulated in [III-16], and [III-17]. The distribution of Eq. (III.D.2-2) with weight functions

$$\Psi(t) = 1/t$$

and

$$\Psi(t) = 1/(1-t)$$

has been investigated in [III-18], [III-19], [III-20]. For further references to goodness of fit tests and discussions of the power of these tests, the reader is referred to Sections 3.5 and 3.6 of Buckland [III-21].

Two Sample Tests. A test that two independent random samples come from the same distribution can be satisfactorily handled by a test statistic developed by Smirnov [III-22]. The test statistic is defined as follows

$$D_{mn} = \sqrt{\frac{mn}{m+n}} \underset{-\infty<x<+\infty}{\text{maximum}} \left|\hat{F}(x) - \hat{G}(x)\right| \qquad \text{(III.D.2-3)}$$

where $\hat{F}$ and $\hat{G}$ are the empirical cdf's of two random samples of sizes m and n. A table of the limiting distribution of Eq. (III.D.2-3) can be found in [III-23]. Corrections to this table can be found in [III-24], [III-25], with extensive percentage points in [III-26].

The test statistic in Eq. (III.D.2-3) was then modified by the introduction of a weight function, viz.

$$D^*_{mn} = \sqrt{\frac{mn}{m+n}} \underset{-\infty<x<+\infty}{\text{maximum}} \{ |\hat{F}(x) - \hat{G}(x)| \Psi(x) \} \qquad \text{(III.D.2-4)}$$

where $\Psi(x) = \begin{cases} 1 & a<x<b \\ 0 & \text{elsewhere} \end{cases}$

or $\Psi(x) = \begin{cases} 1 & x<a,\ x>b \\ 0 & \text{elsewhere} \end{cases}$

The latter weight function puts all of the emphasis in the tails of the distribution. The limiting distribution of Eq. (III.D.2-4) has been treated in [III-27]. Drion[III-34] showed that the asymptotic formulae for the case where m=2 were satisfactory with n as small as twenty.

For unequal sample sizes, two special cases have been studied. Case one is where the first sample is a multiple of the second sample and case two is where the two samples differ only slightly in size. For the former case Blackman [III-28] did some work; and for the latter case, work has been done recently by Reimann and Vincze [III-29].

K-Sample Tests. The large sample asymptotic distribution for three equal-size samples was calculated by David [III-30] but not tabulated. The test is a development of a Smirnov type statistic and a table with sample sizes up to n = 40 has been given by Birnbaum and Hall [III-31]. In the case of K(>3) samples, Kiefer [III-32] derived some limiting distributions for analogs of the Kolmogorov-Smirnov-Cramer-von Mises statistics. It was shown that if a homogeneity of K-samples with respect to some continuous distribution could be satisfactorily explored, then the goodness of fit statistic would follow.

Other tests such as testing whether K distributions differ only by scale parameters, or only by location parameters, can be found in Kendall and Stuart [III-33], Vol. II, Chapter 31.

The ideal test for the purpose of this study would be a K-sample test with weight function placing the emphasis in the lower tail of the distribution, with modifications to handle censored unequal sample sizes. Although much work has been done in the field of K-sample tests, it does not appear that precisely such a test has been developed. For this reason it would be of value to investigate this problem in further detail.

III.D.3 Hill Test of Linearity vs Convexity

In this section a test of linearity of a median regression curve against an alternative of convexity is discussed. This test was alluded to in Subsection III.B.2.4 and here the logic underlying the test, the estimation procedures used, and the test itself shall be considered. This test was considered in detail and in great generality by Hill [III-9]hence the details of the test shall not be given here. Instead, a broad overview concerning the assumptions needed for a valid application of the test, properties of the test, and miscellaneous comments shall be given. It is remarked in passing that, although the alternative hypothesis concerns convexity of the median regression curve, the results can be easily modified for an alternative of concavity or for a joint alternative of either convexity or concavity.

In particular the following test shall be discussed:

$$H_o: \quad Y_i = \beta_o + \beta_1 x_i + \epsilon_i \quad , \quad i = 0,1,2, \ldots, n$$

against an alternative hypothesis

$$H_1: \quad Y_i = \xi(x_i) + \epsilon_i, \qquad i = 0,1,2, \ldots, n$$

where β_o, β_1, and ξ are unspecified and ξ is a nonlinear convex function. The basic assumption needed for the test is that the ϵ_i are independent identically distributed random variables with median zero and with continuous pdf, f_ϵ, such that $f_\epsilon(0) > 0$. The x_i are observations on an experimental variable (observations on a variable controlled by the experimenter) and are fixed, whereas Y_i is a random variable corresponding to a selected level of the controlled variable x_i. It is remarked that Subsections III.B.4, III.C.1, III.C.2, and III.C.3 consider the Hill test where Y_i is the natural logarithim of TTF and x_i is a controlled variable specifying the stress environment among a group of similar environments (such as the linear coefficient of time in linearly increasing stress environments). Observations under H_o then take the form

$$y_i = \beta_o + \beta_1 x_i + \epsilon_i \quad .$$

The following description of the test is taken from page 1096 of [III-9].

"The test consists of estimating a line by the Mood-Brown procedure (using medians) from a central subset of the observations, making a weighted count of the number of remaining observations lying above the line and rejecting H_o if the number, R_n, is large."

The Mood-Brown estimation procedure referred to in the above description of the test can be found in Mood [III-10], pages 409, 410. The Mood-Brown estimation procedure defines a unique line such that half of the sample points (y_i, x_i) used to estimate the line lie above the line, and half lie below the line, hence the term median regression line. This line shall subsequently be abbreviated by $(\hat{\beta}_1, \hat{\beta}_o)$ where $\hat{\beta}_o$ and $\hat{\beta}_1$ are the estimates of β_o and β_1 under H_o found by the Mood-Brown procedure.

The Hill test of linearity versus convexity is based on the asymptotic distribution of $(\hat{\beta}_1, \hat{\beta}_o)$ and R_n, the number of points lying above the line $(\hat{\beta}_1, \hat{\beta}_o)$. As might be suspected the asymptotic distribution of $(\hat{\beta}_1, \hat{\beta}_o)$ is a Bivariate Normal distribution. In the asymptotic distribution, it is found that the results depend on a spacing function h which puts the points

$$x_o \leq x_1 \leq x_2 \leq \dots \leq x_n$$

in one-one correspondence with points in the interval [0, 1]. In general, this function should be chosen by the experimenter prior to testing. It is further found that under the alternative hypothesis the line $(\hat{\beta}_1, \hat{\beta}_o)$ converges to a line $(\beta_1(\xi), \beta_o(\xi))$ based on the convex function ξ .

With this last statement in mind the logic underlying the test, taken from page 1107 of [III-9] is as follows:

> "The logic underlying the test depends upon the fact that the line joining points of a strictly convex function lies above the function in the interval between the abscissae of the points, and lies below elsewhere. Now we have seen in the previous section that when the convex function $\xi(x)$ is the true regression curve, the Mood-Brown estimate $(\hat{\beta}_1, \hat{\beta}_o)$ converges to a line $(\beta_1(\xi), \beta_o(\xi))$ satisfying equations (2.3.1). But it is evident from these equations that if $[c_1, d_1]$ is the interval used for line estimation, then $(\beta_1(\xi), \beta_o(\xi))$ must intersect $\xi(x)$ in two points with abscissae in this interval. Hence for large n the line $(\hat{\beta}_1, \hat{\beta}_o)$ will tend to lie below the curve in the intervals $c \leq x \leq c_1$ and $d_1 \leq x \leq d$; and so if x_i lies in either of the intervals then Y_i will tend to have probability of greater than 1/2 of lying above $(\hat{\beta}_1, \hat{\beta}_o)$. Consequently R_n will tend to be large."

Equations (2.3.1) mentioned in the above quotation require the expected number of points lying left or right of the median of the x_i and above the line to be m/4 among the m points used for estimation of the Mood-Brown line.

It is remarked at this point that the foregoing has been restricted to only one Y observation at each x. For this reason the following comments by Hill made on pages 1106 and 1116 of [III-9] are now in order.

"When there are more than one Y observations, say $Y_{i1}, \ldots, Y_{ik_i}$, at each x_i, then slight modifications in the statement and proof of the theorems concerning the asymptotic distributions of R_n and $(\hat{\beta}_1, \hat{\beta}_o)$ are necessary. This situation corresponds to a spacing function which is constant within certain subintervals of [0, 1]."

" . . . That although we have assumed heretofore that the function h is strictly monotonic, thus restricting the number of Y observations at a fixed x to be one, this assumption is not essential, and the results go through for non-strictly monotonic h(t) with only minor modifications."

Apparently the minor modifications occur only in the proof of the theorems and in the statement of the theorems; in particular, the assumption of strictly monotonic function h is changed to non-decreasing function h. It is readily seen through investigation of the statements of the theorems and the proofs of the theorems that the same results hold if more than one Y observation is made at each x observation.

Consider now some properties of the Hill test. It turns out that

(i) the Mood-Brown estimator of (β_1, β_o) is a consistent estimator under H_o and, moreover, estimates a unique line $(\beta_1(\xi),\beta_o(\xi))$ based on the convex function ξ, the weighting function used in the test statistic, and the spacing function h;

(ii) the asymptotic power of the Hill test is one, against a wide class of alternatives, so that the test is consistent for these alternatives;

(iii) the optimal (in terms of maximum asymptotic relative efficiency) weighting function mentioned in (i) against a quadratic alternative is found to be of the form

$$a(t) = k_1 + \frac{1}{2}(t - 1)^2 \quad , \; k_1 < 0 \quad ,$$

thus weighting (Y_i, x_i) for small x_i heavier than (Y_i, x_i) for large x_i;

(iv) the optimal subset of observations (for the quadratic convex alternative) on which to base the Mood-Brown estimator is approximately the middle 64% of the observations;

(v) the optimal asymptotic relative efficiency was found to be comparable to that of the sign test versus the t-test (i.e., approximately $2/\pi = .636$).

Hill also carried out some Monte-Carlo studies for the robustness of the Hill test and found that, if the total number of points used to estimate the Mood-Brown line exceeded 10, then the asymptotic result was satisfactorily normal for testing H_o. Finally, it is stated by Hill on page 1123 of [III-9] that

> "Whether the conditions $0 \leq \delta \leq .20$ and $h(2\delta) > 10$ are appropriate for other distributions (other than the normal) has not been investigated. The author suspects that they will be in a wide variety of cases."

Here δ is the proportion of observations in either tail (not in the central subset used to estimate the Mood-Brown line) and n is the total number of observations.

IV. GUIDANCE INSTRUMENT CONSIDERATIONS

IV.A INTRODUCTION

IV.A.1 Preface

Part IV of this report is concerned with describing the scope of applicability and feasibility of accelerated life testing of precision electromechanical components such as those in space guidance instruments.

An instrument can be exposed to many physical conditions to cause instrument failures in less time than under normal operating conditions. Many such conditions are, a priori, plausible choices for accelerating instrument failure. Since the objective of accelerated life testing is to make inferences regarding the instrument reliability at use conditions rather than just when and how it fails at the accelerated conditions, there must be some rationale for believing that the accelerated conditions reflect the instrument reliability at use conditions. In lieu of the mere conjunction of life tests at both normal and accelerated conditions, this rationale must follow from an understanding of the physical processes involved at both sets of conditions. In particular, at least two areas of knowledge are presupposed before attempting to subject a device to an accelerated life test: i) knowledge of the modes of failure (and their specific physical causes) that occur at usage conditions, and, ii) a sufficient knowledge of the dependence of failure behavior on experimental conditions to justify an extrapolation from accelerated conditions to usage condition(s).

Clearly, by considering the physical process responsible for a particular failure, a meaningful choice of accelerating conditions for that process can be made.

At the beginning of this study, several types of instruments were investigated, i.e., accelerometers of the pendulous-integrating-gyroscopic type, the differentiating-feedback type, electromagnetic type, and gyroscopes of the single-axis, free-rotor, and case rotated types. Of these, two Autonetics manufactured instruments were selected for study in depth, the VM-4A Velocity Meter and the G6 Gyroscope. These instruments were selected for study because there was a relatively large amount of information available concerning their possible modes of failure, yet they represent a class of very reliable instruments.

After probing into various aspects of instrument failure phenomena and the implementation of accelerated life tests, it became obvious that the most appropriate level of interest is that of separate functional components which make up a total instrument. An important consideration was to lessen the possibility of unknowingly introducing spurious failure modes. Also, mere curve-fitting of data pertaining to various environmental conditions, such as the procedures described by Adams [I-3] (and quoted in Subsection I.D of this report), does not instill much confidence in the recipient of estimates based on accelerated tests. Hence it was desired to improve on this type of extrapolating by considerations more related to the physical phenomena involved.

In this light, therefore, several components of the two instruments were selected for detailed study. Some generalizations concerning the salient features of instrument failure that limit the feasibility and applicability of accelerated life tests are presented in Subsection IV.B. The opinions expressed in this section were, on the whole, formulated from studies of the two Autonetics instruments. Therefore, that section is not meant to be an expression of the state-of-the-art of knowledge concerning the failure phenomena of all inertial instruments. Descriptions of the instruments, components, and their possible failure modes are included. Some of the problems likely to be encountered in implementing any accelerated life test and the framework for their solution are illustrated in connection with the two selected instruments.

IV.A.2 Conclusions

Given a specific mode of failure (and its physical mechanism) that can occur at usage conditions, accelerating conditions affecting that particular mode of failure can generally be chosen. Depending on the type of failure process under consideration, different types of accelerating conditions are appropriate. The following three types of processes are apparent [see Section IV.B]: i) cumulative processes, ii) stabilization processes, and iii) critical limit processes. For the cumulative processes the time-compression (TC) or over-stress-testing (OST) approaches are appropriate. For stabilization processes, testing of material properties for selection of the best material appears to be appropriate, and in some cases the OST approach appears possible. For the critical limit processes, environmental tests are appropriate, and if estimation of a low fractile is desired, a stochastic approximation approach is recommended (see Section VI.C). Though consideration of specific modes of failure is required in order to implement an accelerated test, actual failures in a given mode need not be observed beforehand. For instance, if two metal surfaces are rubbing against each other, it is expected that some sort of wear process will be important, and an accelerated test can be designed accordingly. But note that there are instances of failures whose occurrence cannot reasonably be hypothesized beforehand. For instance, some failures can result from completely unexpected and unpredictable stresses. Also, failures resulting from lapses in quality control are not really amenable to accelerated testing. For the reasons discussed in this paragraph accelerated life testing does not appear applicable for evaluation of total instrument reliability with respect to a given environment in the same sense that a normal life test is employed. ALT does appear applicable for the evaluation of cumulative (or "wearout') failure modes.

The basic problem concerning the feasibility of accelerating life testing is that of justifying the extrapolations to use conditions. Failure models can be hypothesized, and their validity can be partially ascertained at the accelerated conditions. However, in order to extrapolate to the use conditions, a fairly detailed knowledge of the failure processes is needed. The limiting factor affecting the feasibility of realizing a meaningful extrapolation is the extent to which the physical mechanisms responsible for failures are amenable to being defined and controlled.

IV.B GENERAL INSTRUMENT CONSIDERATIONS

Drawing inferences from an accelerated life test generally requires a thorough knowledge of the dependence of failure behavior on experimental conditions. The type of information needed can be described best in connection with specific examples. In general, that information is required which establishes that a mechanism of failure is unchanged over the range of accelerated to usage conditions. The notion of a mechanism being unchanged implies that the constants in the model which relates the failure behavior to the experimental conditions are unchanged. Knowledge of the failure mechanisms (the influencing variables and the physical manifestations), and of the relative sensitivity of the device to the influencing variables, is desirable in order to exploit this means of justifying the extrapolations. This approach appears more feasible for electromechanical components than for electronic components because the basic failure processes associated with mechanical components are more directly observable. For example, see the discussion concerning the mechanism of brush wear in Subsection IV.C.2.2.

The change in experimental conditions can take two forms. Some environmental or operational stress (temperature, mechanical load, etc.) can be increased - the over-stress-test (OST) approach; or the rate of accumulating cycles of some pertinent stress can be increased - the time compression (TC) approach. For the OST approach, the dependence of the failure mode in question on some appropriate accelerating stresses should be known for an adequate range. Generally, this information is not available, except as a rough rule of thumb or general trend for a class of similar devices. Having tested the device at accelerated conditions and considered the nature of the physical processes involved, a suitable analytical model and the proper constants can be developed which account for both the failure dependence on stress and the variability of the data. Then comes the crucial stage: justification of the extrapolation implied by the analytical model. What is considered a valid justification is essentially a matter of judgement regarding the possible consequences of erroneous conclusions. But, if reasonable precautions are taken to verify that there are no gross changes (from accelerated to usage conditions) in the variables influencing the failure, then there is surely less risk in accepting the resulting prediction than when one merely relies on the extension (to normal use conditions) of a curve fitted to data at accelerated conditions; such precautions should also be taken when more formal failure models (e.g. the SW model discussed in Part III) are used for extrapolation. Reliance on the alternative method, i.e. increasing the frequency, requires that the distribution of cycles-to-failure be independent of age or operating time. (Though independence from cycling rate implies independence from age or operating time, the converse is not true). Again, the difficulties likely to be encountered in this approach arise from the necessity of justifying the assumption. An attempt to justify the assumption of independence of cycling rate can be based on one or both of the following. A detailed study should be made of all variables that can possibly affect the amount of damage per cycle to ascertain that these variables are unchanged at the increased frequency. The other approach for justification would be to statistically test that the distribution is unchanged at several frequencies. This would tend to indicate the possible presence of any time dependent effects. Illustrations of these considerations are found in Sections IV.C and IV.D.

In order to gain a better insight into the scope of applicability of accelerated life testing, it is instructive to consider some of the salient features of instrument failure processes that were abstracted from a study of several instrument components. A review of the various mechanisms of change that can cause instrument failure strongly suggests that such mechanisms can be grouped into the following three classes.

Cumulative Processes. Consider the following mechanisms: wear, fatigue, embrittlement, contamination, evaporation, and corrosion. Each process has at least one "cause" that is external to the material itself, and if the cause is continuously present, the process would generally continue in a nondecreasing or cumulative manner until the material or device breaks, cracks, dissipates, etc. These are cumulative processes.

The cumulative process is characterized by the functional dependence on the environmental and operational conditions and the variability of this dependence from one instrument to another.

Stabilization Processes. Now consider (1) dimensional or mass distribution changes due to residual stress relief or metallurgical instability, (2) the change in density of a material due to fluid absorption, and (3) the natural stabilization of a magnet. These processes usually progress in a cumulative manner towards an equilibrium or stabilized state. The forces approaching equilibrium are within the material for the residual relief and metallurgical processes, while the forces involved for magnet stability stem from interconnections between the flux return paths and the magnet and also from interactions within the material. The fluid absorption involves pressure differentials. Clearly, then, the environment can affect the rate of stabilization and the level of the stabilized state, but it is not a constant driving force as for the cumulative processes. Another distinguishing quality is that once the processes have stabilized it is usually possible to make an adjustment that will compensate for any past changes in instrument performance. The most important parameters for characterization of stabilization processes are the sensitivity of the stabilized state of various changes in the environment, and the rate at which the stabilization process occurs under operating conditions.

__Critical Limit Processes__. A critical limit process is one in which the strength of some functional component is exceeded by some extreme in a random environment, e.g. gas spin bearing seizure during vibration. In that example, the functional component is the hydrodynamic support system. Its strength is that bearing load which can be supported without metal-to-metal seizure. Another example of a limit process is the relative slippage between parts fabricated with sliding or force fits. The pertinent factors affecting reliability that would need evaluation are the critical strength, the changes of strength as a function of time, and the distribution of that aspect of the random environment that can cause failure.

A cumulative process can be viewed as an accumulation of some kind of damage, and there is a failure when some fixed amount of damage (unique to that particular device) has been accumulated. There are generally many variables that influence the rate at which the damage is accumulated, e.g. composition or structure type variables, or operational variables such as temperature, speed, and mechanical load. The time to failure is a random quantity, and it is frequently inferred that this randomness is due to the variability in the influencing factors; in another approach, the randomness is attributed to the difference, from device to device, in the critical amount of damage which can be sustained before failure occurs. When the time to failure is strongly influenced by one of these factors and furthermore, this particular factor is amenable to being controlled, this factor can be used to accelerate the failure in a controlled and continuous manner. This is the essence of the OST approach, whereas the TC approach assumes that the strongly influencing factors can be kept constant while the frequency of repetition of normally occurring stresses is increased. As is shown subsequently, there are several instances in which the "strong" factors are controllable. If the variables that are available for controlled change have relatively little influence on the failure process (in comparision to the effect of uncontrollable variables), then the feasibility of implementing an accelerated test in a valid manner is nil.

Now consider the stabilization processes. Of the three instances considered, only one was concluded to be amenable to an accelerated test (in the OST manner). This instance was the change in density of a material due to fluid absorption. Here, the rate of stabilization is strongly dependent on a controllable variable, i.e., the fluid pressure. For a gyroscope rotor, an important manifestation of creep is the mass unbalance, a random, reversible effect with a zero mean. Clearly, the appropriate utilization of an accelerated test in this situation would be in the selection of the best material with respect to creep behavior, instead of for the estimation of time to first mass unbalance of some given amount. The material selection approach is discussed later in Part V of this report.

It should also be mentioned that the important characterization of critical limit processes is its critical strength. The determination of such critical strengths does not, in principle, involve much time, so accelerated testing for this type of process is not meaningful.

Note that so-called "random" failures could fall in any of the above categories of failure processes. A failure is referred to as random generally because the cause of failure is unknown, or because it occurs very infrequently or unexpectedly. Clearly, an accelerated life test for "random" failures of this nature is not possible.

Another salient feature of failure phenomena concerns the notion of failure mechanisms. So far, it has been implicitly assumed that the "mechanism of failure" is an entity which can always be labeled (such as wear, fatigue, oxidation, etc.) and that, after an accelerated test, the failures can always be labeled as to mode of failure. The caution to be expressed here is that the mechanism of failure is not always definable. For example, sometimes the cause of failure is a complex combination of two or more "defined" mechanisms of failure. Such a combination might make it difficult to ascertain that the mechanism of failure at accelerated conditions is the same mechanism which occurs at usage conditions, and hence conclude that there has been no change in mechanism.

Since it is difficult to discuss the scope of accelerated life testing in terms of generalities, further discussion will be in connection with some specific components.

IV.C VELOCITY METER

IV.C.1 <u>Instrument Description</u>

The Autonetics VM-4A Velocity Meter is a singly-integrating, single-axis accelerometer. The inertial sensing element is a floated cylindrical body known as a "pendulum" since its center of mass is displaced from its axis of rotation. The pendulum is constrained axially, radially, and torsionally by a hydrostatic (liquid) bearing, which is kept pressurized by an electromagnetically operated pump. The flotation and bearing liquid is highly viscous (for damping pendulum rotation) and of sufficient density to provide neutral buoyancy for the pendulum at the proper operating temperature.

In operation, an acceleration along the input axis of the instrument creates a torque that makes the pendulum rotate slightly. The rotation is electrically sensed, amplified by an external servo controller and then converted to a d-c voltage. This voltage is fed back to the servomotor that drives a rotating six-pole permanent magnet (the "drag" magnet) with the proper speed and direction to restore the pendulum to essentially its position with no acceleration. The restoring torque is produced by eddy currents induced in a conductive portion of the pendulum by the drag magnet. The output of the velocity meter is an electrical signal from a readout encoder that measures the magnitude and direction of the magnet shaft angular rotation. Because the drag force on the pendulum is proportional to the magnet angular speed, the amount of magnet-shaft angular rotation is proportional to the velocity of the instrument. Hence, the electrical signal is a measure of the input velocity. Further technical description of the VM-4A can be found in [IV-1].

The principal component parts of the velocity meter (some of which are already mentioned above) are the pendulum, the hydrostatic fluid bearing, the pump, the capacitor pickoff, the pickoff transformer, the servo motor, the magnet, the readout encoder, a pair of ball bearings, and thermal compensation bellows. Of these components, four are susceptible to some sort of cumulative damage failure, i.e., the servomotor, the ball bearings, the pump, and the bellows. Each of the components will be described subsequently; their possible failure modes are discussed, and consideration is given to the feasibility of an accelerated life test for each component.

Brief mention is made now of some components not analyzed in detail in this study. The hydrostatic fluid bearing is nearly stationary and, in operational circumstances, there is no metal-to-metal contact between the shaft and the bearing. A small amount of contamination can cause it to stick,thus causing an erratic output. The pendulum is susceptible to density changes due to absorption of the floatation fluid. Such a process would be of the stabilization type and would possibly be amenable to an accelerated test, though it was not considered in any detail in this study. Once the magnet and its flux return path have been artifically stabilized, the magnetic field has not been observed to change

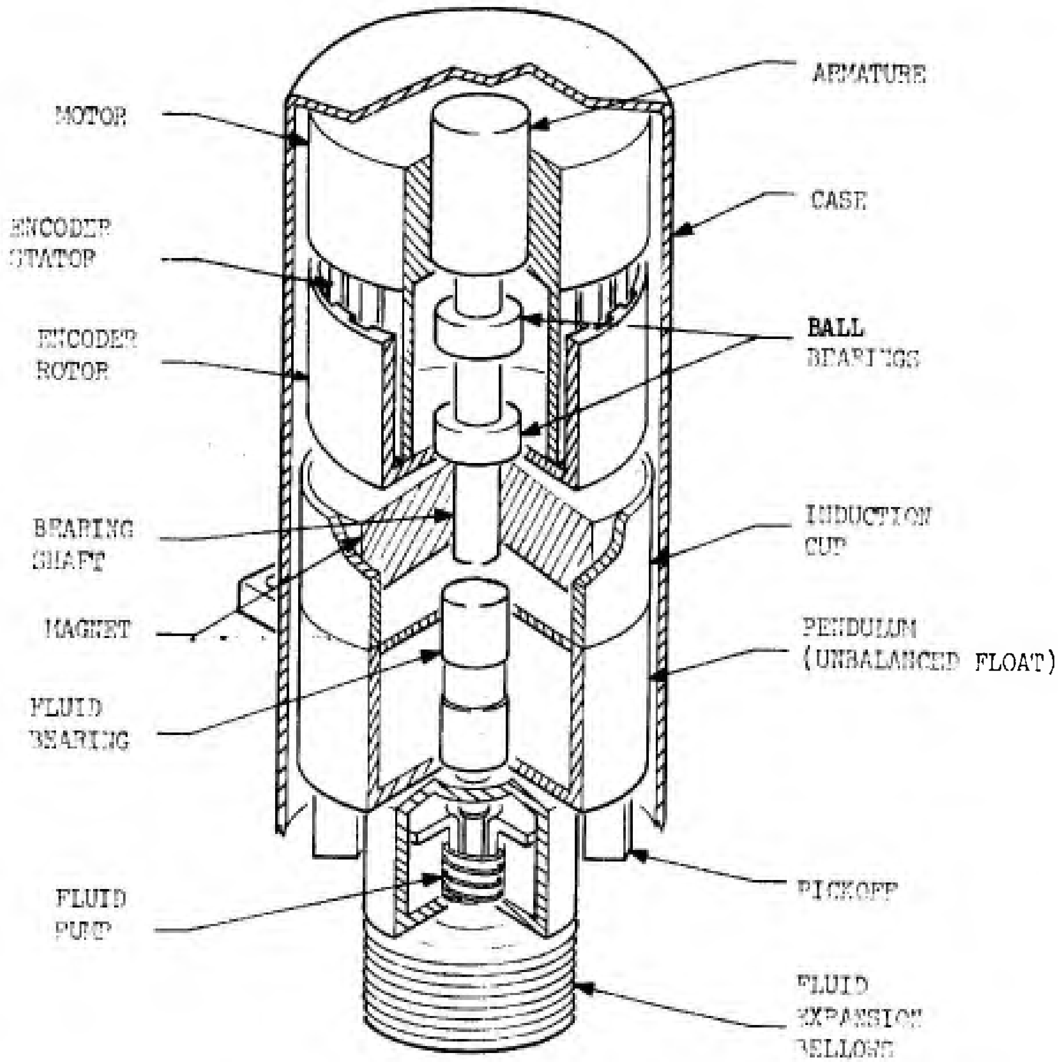


FIG. IV-1. SIMPLIFIED DRAWING OF VELOCITY METER significantly within a period of a year. No accelerated test is believed to be possible that would estimate how stable a magnet will be. The capacitor pickoff and readout encoder are electrostatic devices. All electrical connections can, in principle, eventually become open due to vibration or deterioration of mechanical bonds. Perhaps, also, radiation could have some deleterious effects on the cement joints or the flotation fluid.

IV.C.2 Servomotor

IV.C.2.1 Possible Failure Modes

The servomotor is a low speed, high torque d-c motor, employing a wound-armature rotor and a permanent magnet stator. As mentioned before, the motor is required for driving the magnet and encoder in the instrument feedback loop. The speed of the motor is proportional to the amount of acceleration along the sensitive axis of the velocity meter, approximately 1 rps per "g" of acceleration.

Besides failures of the ball bearings, which are discussed in Subsection IV.C.3.1, most failures of servomotors are associated with the brushes and commutator. The processes which occur during sliding at the interface between a carbon brush and the metal commutator are reflected in the magnitude of the coefficient of friction, the electrical contact resistance, and the rate of wear of the brush. The first two quantities affect certain operating performance variables (e.g., starting and running voltages and currents), whereas, wear of the motor brushes will eventually result in catastrophic failure of the instrument motor. Very infrequent failure modes are typically loss of spring tension against the motor brushes, contamination of the motor air gap or of the ball bearings by the brush wear particles, breakage of wire connections, etc. However, wearout of the motor brush is generally the mode that limits the life of the motor.

It is sometimes surmised that the rate of brush wear is constant for most of the brush length. Such an assumption, if true, would allow estimation of the life of a brush by simply using the initial rate of wear. However, this sort of extrapolation can result in brush life estimations that are wrong by a factor of two to ten in either direction from the observed life, judging from past experience with motor brush life tests, because of changes in wear rate during the brush life. It appears that accelerated life tests can be employed to determine how the wear rate changes with brush length (and thus time), or instead, what the time or the cycles to brush wearout is at usage conditions of motor load and speed. Of course, the reasons for the changes in wear rate are needed in order that an accelerated test can be designed which will reflect these changes. However, knowledge of these reasons is not essential for merely implementing an accelerated test. Instead, they are essential for justifying any extrapolations from accelerated conditions to usage conditions. The following discussion of the mechanism of brush wear suggests possible ways of implementing an accelerated life test for brush wearout and possible reasons for the changes in wear rate that have been observed.

IV.C.2.2 Mechanism of Brush Wear

The rate of brush wear is generally associated with the characteristics of a film on the copper commutator. The way in which the rate of wear of the brush depends upon the mechanical and electrical loads, the speed, and the apparent area of the contact was extensively investigated by Lancaster [IV-2]. The experimental investigation utilized an apparatus that allowed the conditions of sliding to be well defined, as opposed to using motors to simulate the service conditions of the brushes. In order to formulate an accelerated life test of motor brushes, it is useful to know, in addition, how the brush life is affected by motor service conditions. Therefore, some data on motor brush life from an instrument development program at Autonetics will first be discussed.

The typical manner in which brush wear in motor service conditions behaves with time is shown in Figures IV-2 through IV-5 (from [IV-3]). Each of the four plots represents one motor with four brushes. The data selected for illustration were obtained from four motors made by the same manufacturer. All of the motors were tested at the usage conditions of load, speed and temperature until brush wearout occurred or the design requirements were exceeded.

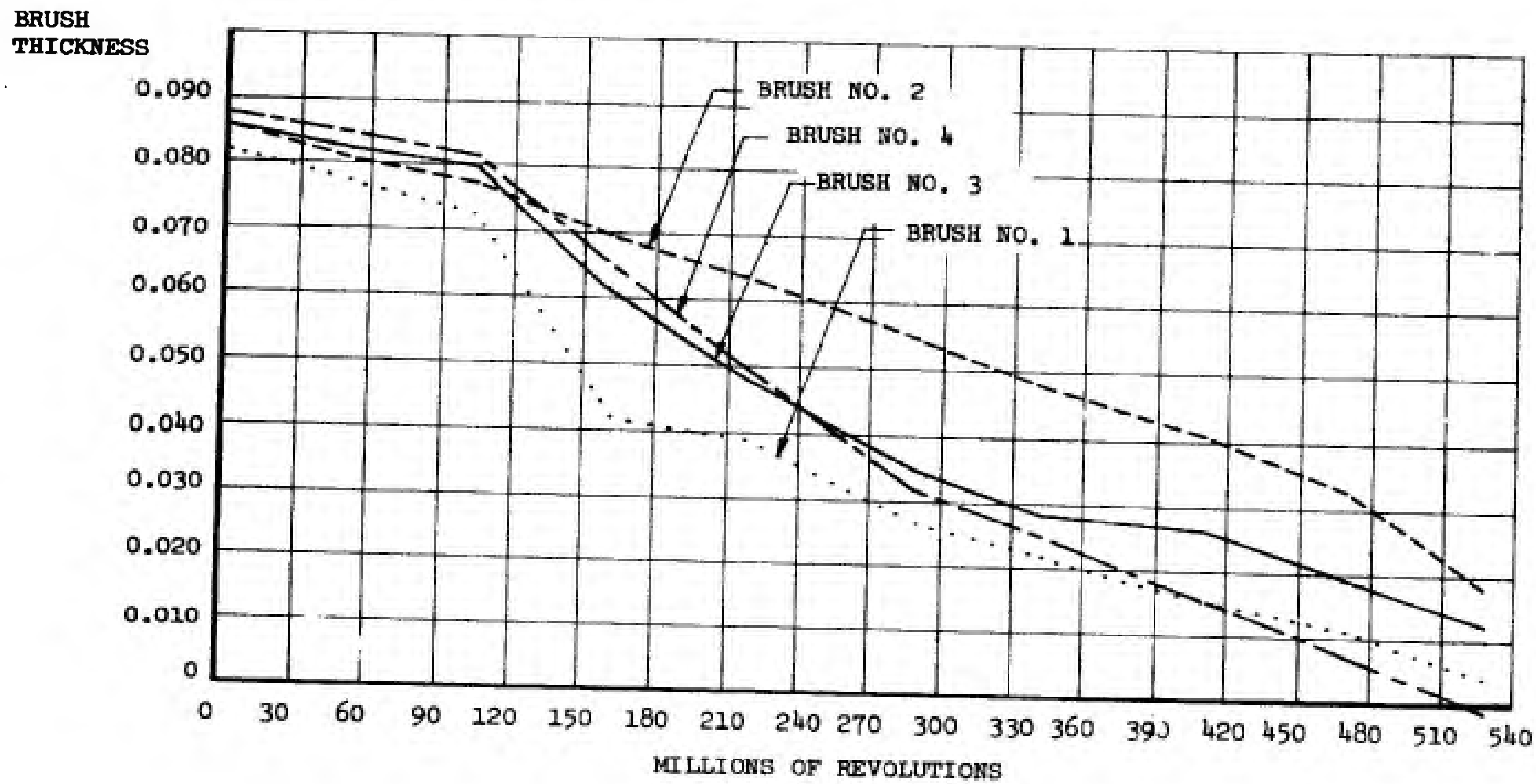


FIG. IV-2. MOTOR BRUSH LIFE TEST MOTOR NO. 1

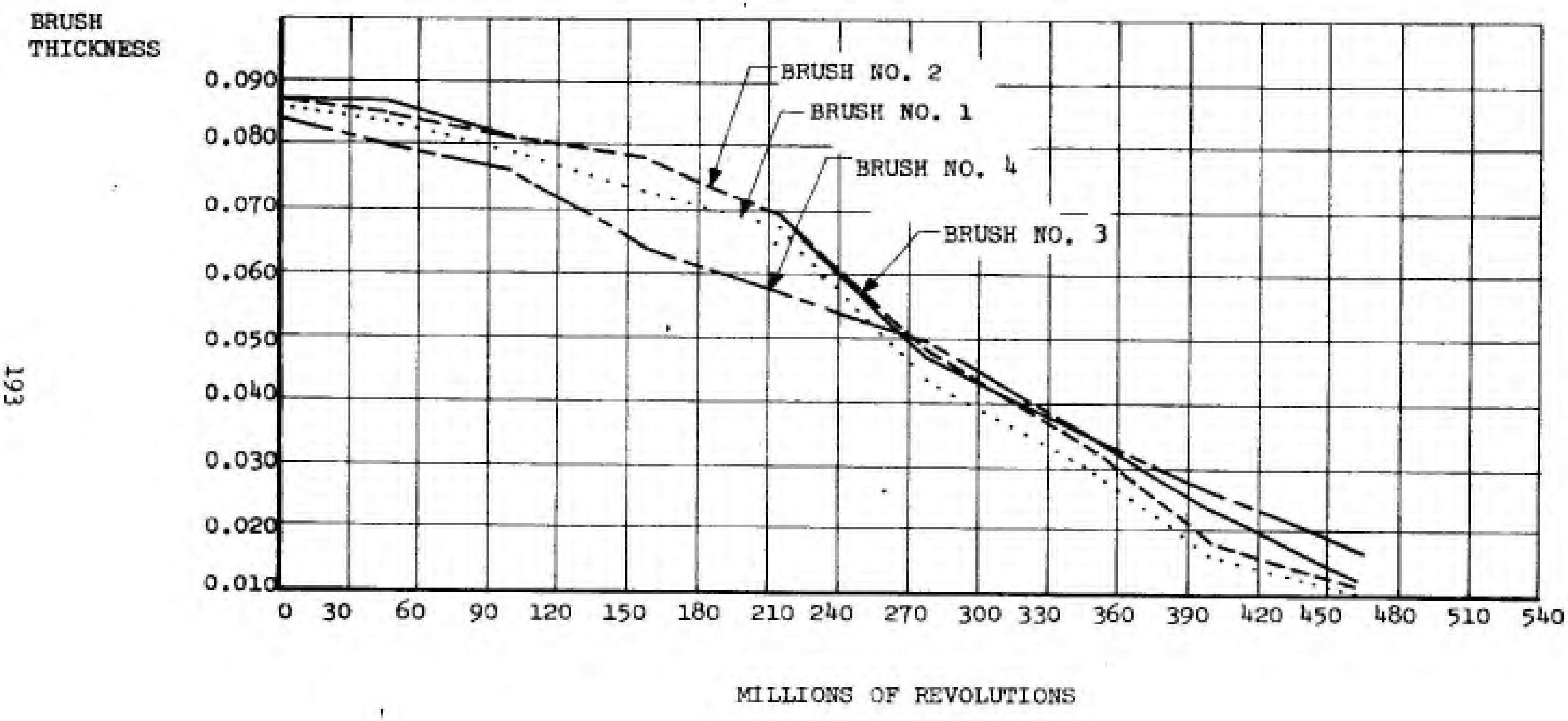


FIG. IV-3. MOTOR BRUSH LIFE TEST MOTOR NO. 2

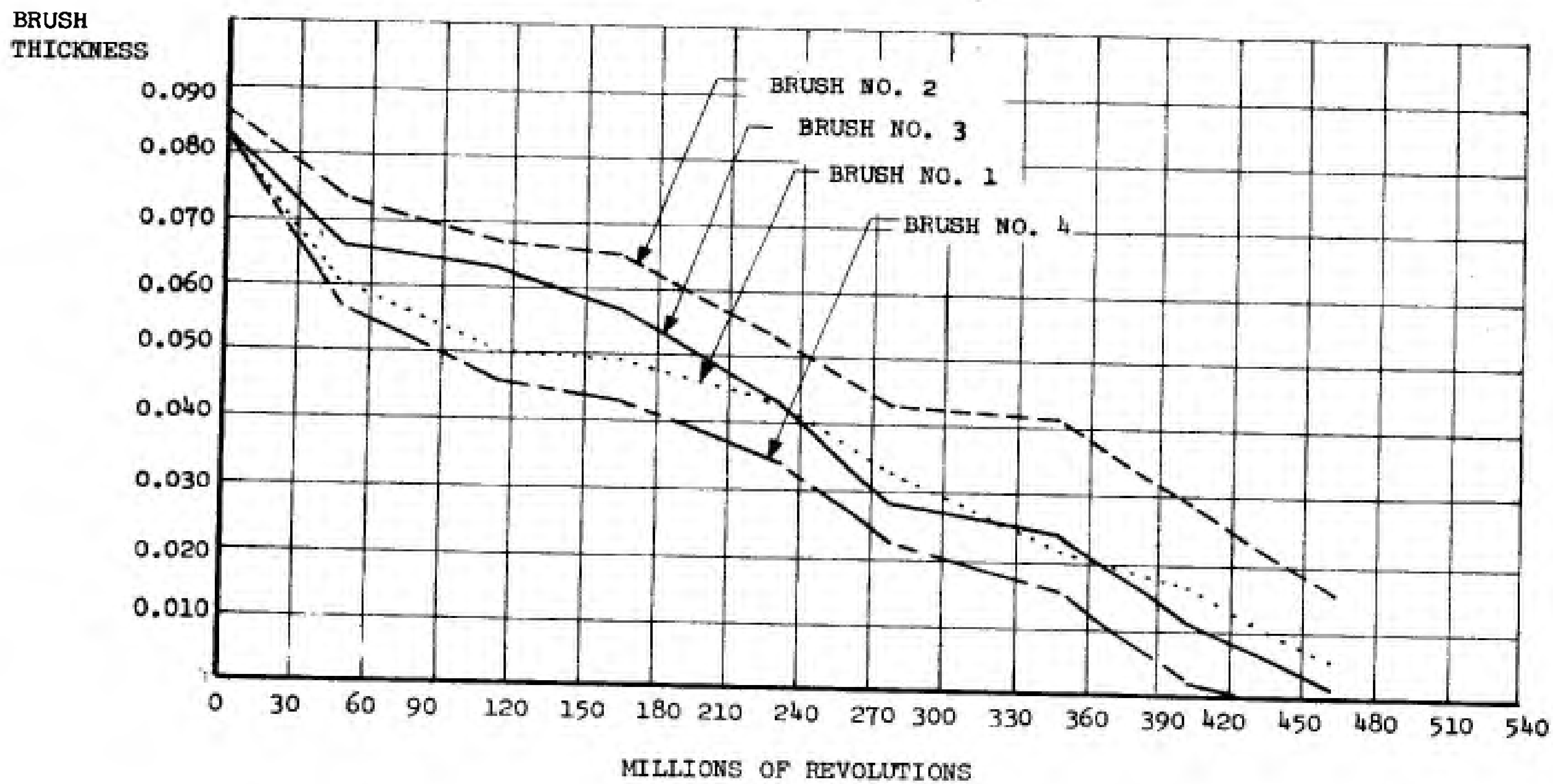


FIG. IV-4. MOTOR BRUSH LIFE TEST MOTOR NO. 3

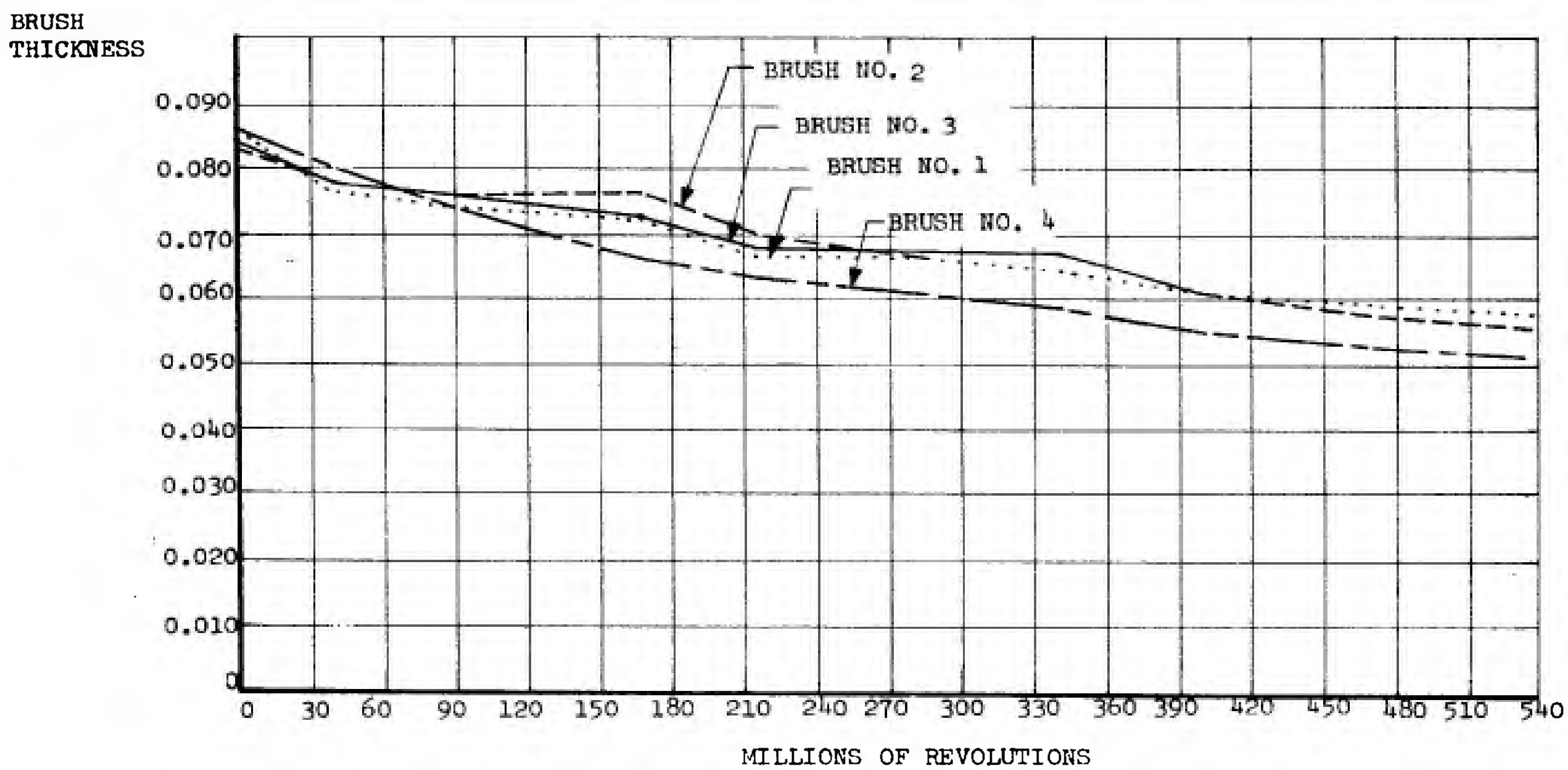


FIG. IV-5. MOTOR BRUSH LIFE TEST MOTOR NO. 4

Note in particular that the initial wear rate can be greater or less than the later wear rates. In some cases, all motors with the same brush grade had the same type of brush wearout behavior. Figures IV-2 and IV-3 show that the initial rate was always less than what existed after, say, 25 percent of the brush length was worn off. In other brush grades, e.g. Figure IV-5, the wear rate was relatively constant for the whole sample during the period observed, but only a small percentage of the brush had been used. However, the wear rates always changed for the brush grades not surviving in the observed test time. Therefore, it is quite possible (without a better understanding of the variables introduced by service conditions) for the wear rate of those barely worn away to change at some shorter brush length, or longer test time.

Some possible reasons for changes in wear rate (length used per time) are: different contact area as brush length is worn away due to varying brush cross section or changes in angle, change in spring tension, changes in commutator surface properties that affect the amount of arcing, etc. Individual study of each brush application would probably be necessary in order to attribute specific possible causes of wear rate changes to a particular brush application. In other words, the spring tension as a function of time, the causes for changes in brush contact area, and other factors affecting brush wear need to be observed under service conditions in order to characterize the pertinent aspects of the service conditions. For example, consideration of the design of the brush and brush holder might reveal that, as the brush is worn, more brush area comes into contact with the commutator, and therefore a slower thickness wear rate would result. If such a change in area could account for the observed rate change in an accelerated test, then one would be justified in extrapolating to usage conditions. Experimental studies such as Lancaster's are very useful in ascertaining how sensitive the wear process is to various aspects of the service conditions.

Lancaster's description of the mechanism of brush wear emphasizes the influence and role of the surface film generated on the copper. Hence, factors that can influence the properties of the film during long and short periods of time are important considerations in the development of valid accelerated life tests for motor brushes. According to Lancaster, the rate of wear usually decreases with time to a limiting value. Figure IV-6 reproduced from [IV-2], shows typical data that were obtained for the volume of brush wear as a function of time without current for a heavy (500 grams) load. The initial transient in the rate of wear is presumably a consequence of changes which occur in the conditions of sliding before an equilibrium is reached, such as, changes in temperature or apparent area of contact, or in the physical and chemical properties of the surfaces. In his words,

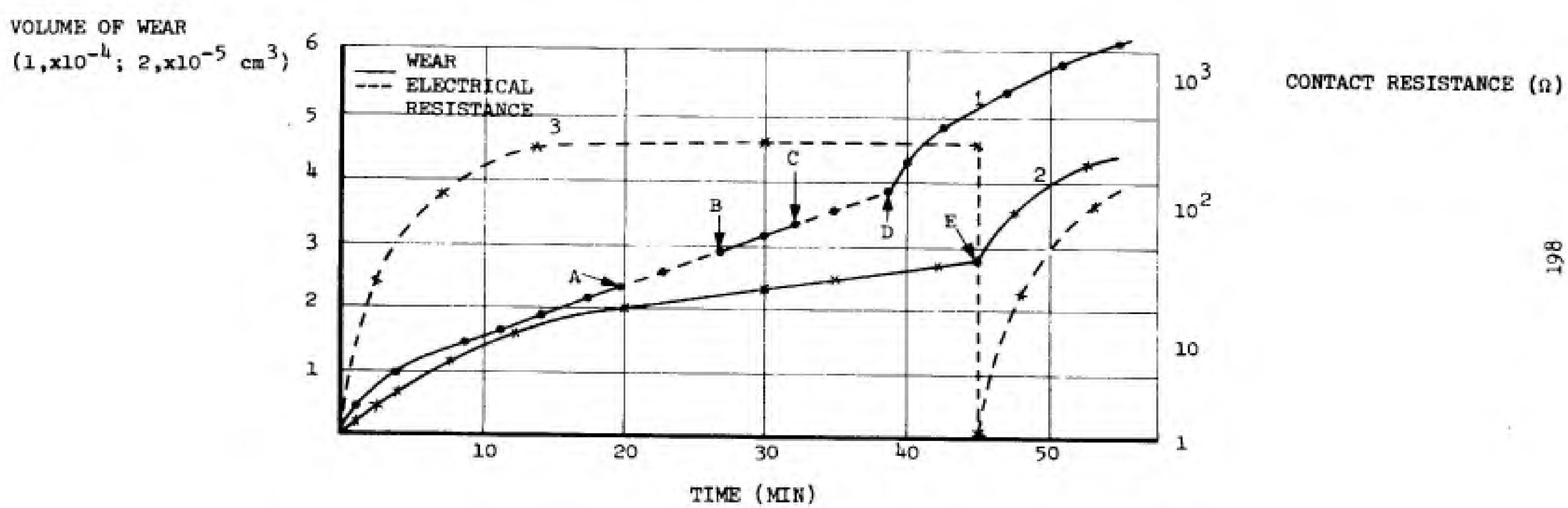


FIG. IV-6. VARIATION OF VOLUME OF WEAR WITH TIME

"To determine which of these possibilities was most important, the wear process at a load of 500g (curve 1) was interrupted at various times after the limiting rate of wear had been reached. At the point A the surfaces were allowed to cool from their equilibrium mean temperature of about 15° C above ambient; at B a freshly prepared conical brush was substituted for the original one, and at C the surface of the brush was refaced. None of these changes significantly influenced the rate of wear. At the point D, however, the surface of the copper disk was lightly abraded with precision grade silicon carbide paper and the rate of wear immediately increased. It may therefore be concluded that the gradual reduction in the rate of wear during the early stages of sliding on freshly prepared copper is caused by changes occurring on the surface of the copper."

The same conclusion was reached from similar series of experiments made at a lighter load (50 grams). Curve 2 in Figure IV-6 shows that the period during which the rate of wear decreases with time is longer for a light load than for a heavy load. In addition, curve 3 shows that the electrical contact resistance between the surfaces for the 50 gram load also increases with time; both the resistance and the rate of wear become constant at about the same time. The decrease in contact resistance accompanying the abrasion of the copper disk was more than two orders of magnitude for the light load, but no change in contact resistance was observed for the heavy load; its value remained less than 0.01 ohms.

The above experimental results indicate the nature of the short-term relationship between the volume of wear and time under well controlled conditions of mechanical load and no current through the contact. The relationships between the volume of wear and time when a current existed were reported by Lancaster to be very similar to those obtained in the absence of current.

There are two types of wear behavior typical of two ranges of loads. Figure IV-7, curve A (from [IV-2]) shows the variation of the limiting rate of wear with load. Note that there is a discontinuity in the wear rate and electrical resistance at about 100g load. The rates of wear above and below the discontinuity are directly proportional to the load. The two regimes of wear may be termed "severe" and "mild" wear. In the severe wear regime, the reproducibility of the rates of wear at any one load was about ± 10 percent; during mild wear, the experimental scatter was about ± 25 percent.

The relatively high rate of severe wear was concluded to be a direct consequence of the formation of a transferred graphite layer rather than vice versa. This conclusion was reached from the following experimental observations. Two brushes were loaded independently on the same track on the copper disk. With a load of 1 kg on one brush, the other was loaded at 150g and the volume rates of wear of each are given by the points W and X in Figure IV-7. The proportionality between these rates of wear and load shows that the surface properties of the transferred

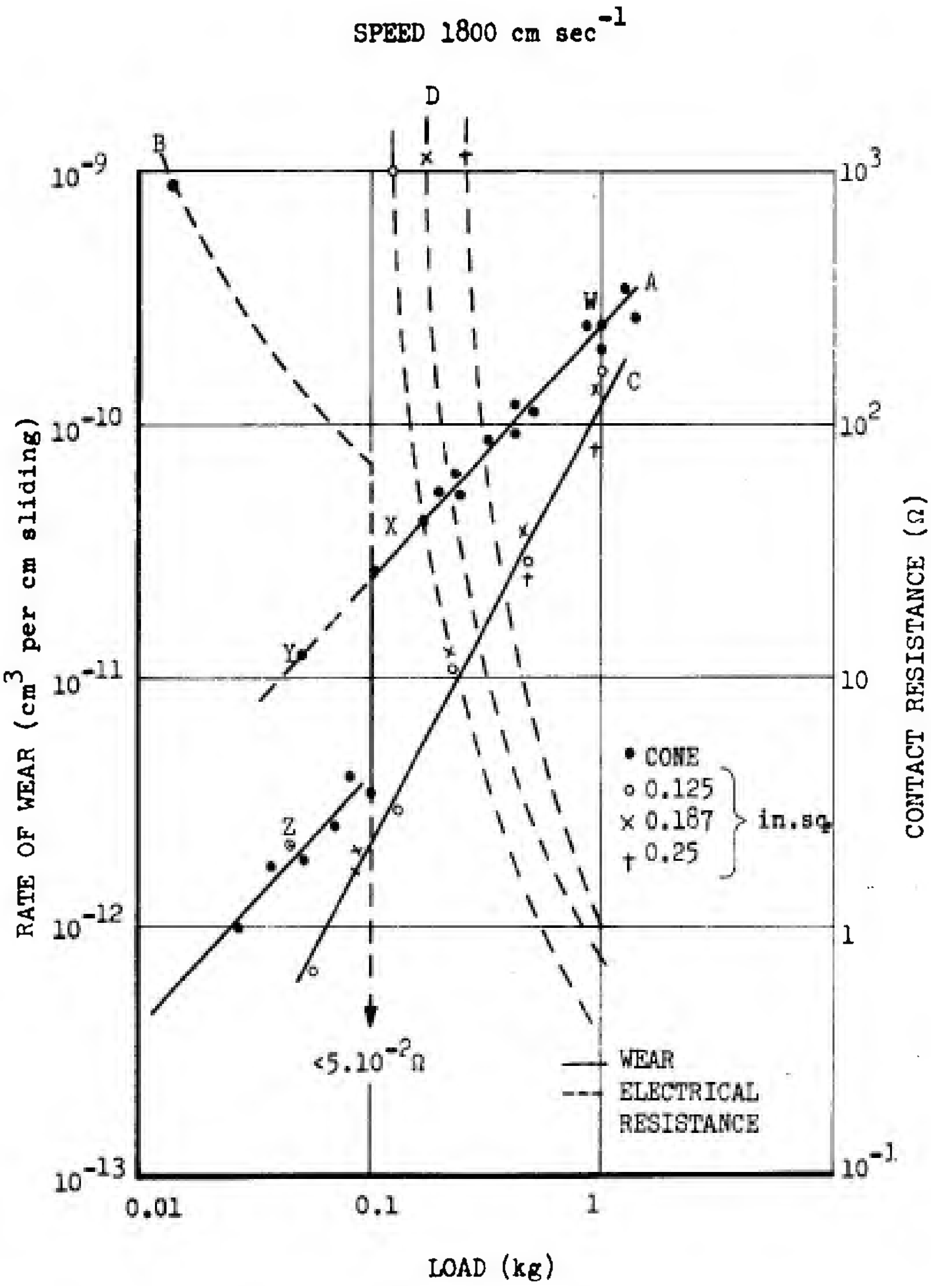


FIG. IV-7. VARIATION OF THE RATE OF WEAR WITH LOAD film, which influence the wear of the brush, remain essentially the same at all loads within the severe wear regime. The 150g load lies exactly on the extrapolation of the line WX. When the brush carrying the 1 kg load was removed, the rate of wear of the brush at 55g decreased gradually from the point Y to the normal mild wear value, point Z. Hence, the transferred layer of graphite produced by the heavy load cannot be maintained by the light load. The graphite, originally produced by the 1 kg loaded brush, gradually wears away under the low load, and the film on the copper disk becomes the typical Cu_2O, which causes only mild wear.

When a conical-ended brush was replaced by one of constant (and larger) cross section, there was no longer a discontinuity at a critical load in either the rate of wear or the electrical contact resistance. Also, rate of wear is a function of some power, greater than one, of the load (i.e., the rate of wear increases more rapidly than proportionally). See Figure IV-7, curves C and D respectively.

Lancaster [IV-2] reports that the measured wear rate characteristics described above are also correlated with physical manifestations. That is, for the conical-ended brush at all loads less than the critical value, only a small amount of graphite was present on the copper. The major constitutent of the light brown colored film was Cu_2O. At all the severe wear loads, a black layer of transferred graphite on the copper was apparent, and electron diffraction showed that very little CuO or Cu_2O was present in the layer. However, for the continuous, and greater than proportional, change in wear rate with load (Figure IV-7, curve C), microscopic examination of the wear tracks on the copper showed that, as the load increased, there was a gradual increase in the amount of transferred graphite and the color changed from light brown to black. The contact resistance measurements in curves D of Figure IV-7 also indicate that the amount of oxide present in the wear track decreases with increasing load. Comparison of the physical manifestations and measured wear rate behavior between the two brush types is suggested by Lancaster to mean that absence of proportionality between the rate of wear and the load is a consequence of the gradual changes in composition (graphite vs Cu_2O) of the surface film on the copper as the load is increased.

The conditions of sliding which are commonly used in practice in electrical machines almost invariably lead, or are intended to lead, to the regime of mild wear. The loading of such brushes is usually specified in terms of stress (force per unit area) rather than absolute load, whereas the experiments of Lancaster show that the volume rate of mild wear is dominated by the load rather than the stress. In other words, the effect of an increase in stress caused by increasing the load changes the volume rate of wear far more than an equivalent reduction of the apparent area of contact. It is also clear that there is no fundamental relationship between the volume rate of wear and load for the situation represented by curve C of Figure IV-7, the physical changes being continuous from the mild regime to the severe regime.

The effect of the oxide film generated during mild wear at light loads and low speeds is to prevent transfer of graphite to the surface, presumably by preventing contact between the brush and the copper substrate. The extent to which oxidation occurs has been shown to be determined primarily by the time elapsing between repeated contacts over the same localized region of the copper surface and by the mean temperature (as apposed to the localized "flash" temperature) of the surface during this time. Some of Lancaster's data supporting this conclusion are shown in Figure IV-8. Note als that the relationship between the rate of wear and load becomes continuous when the speed is reduced from 1800 cm/sec to 460 cm/sec or below, and furthermore that proportionality is no longer obtained.

The influence of current on the rate of wear is shown (see Figure IV-9) to be greatest when the conditions of sliding are such that appreciable oxidation of the copper can occur. The main effect of the current is then to increase the amount of graphite transferred to the copper and displace the wear behavior of the brush towards the severe wear regime. The increase in transfer suggests that the passage of current supplements mechanical breakdown of the oxide film, either by puncturing caused by a high voltage gradient or as a result of local heating and softening of copper causing an increase in the deformation beneath the oxide film. Although the above mechanism can account for the

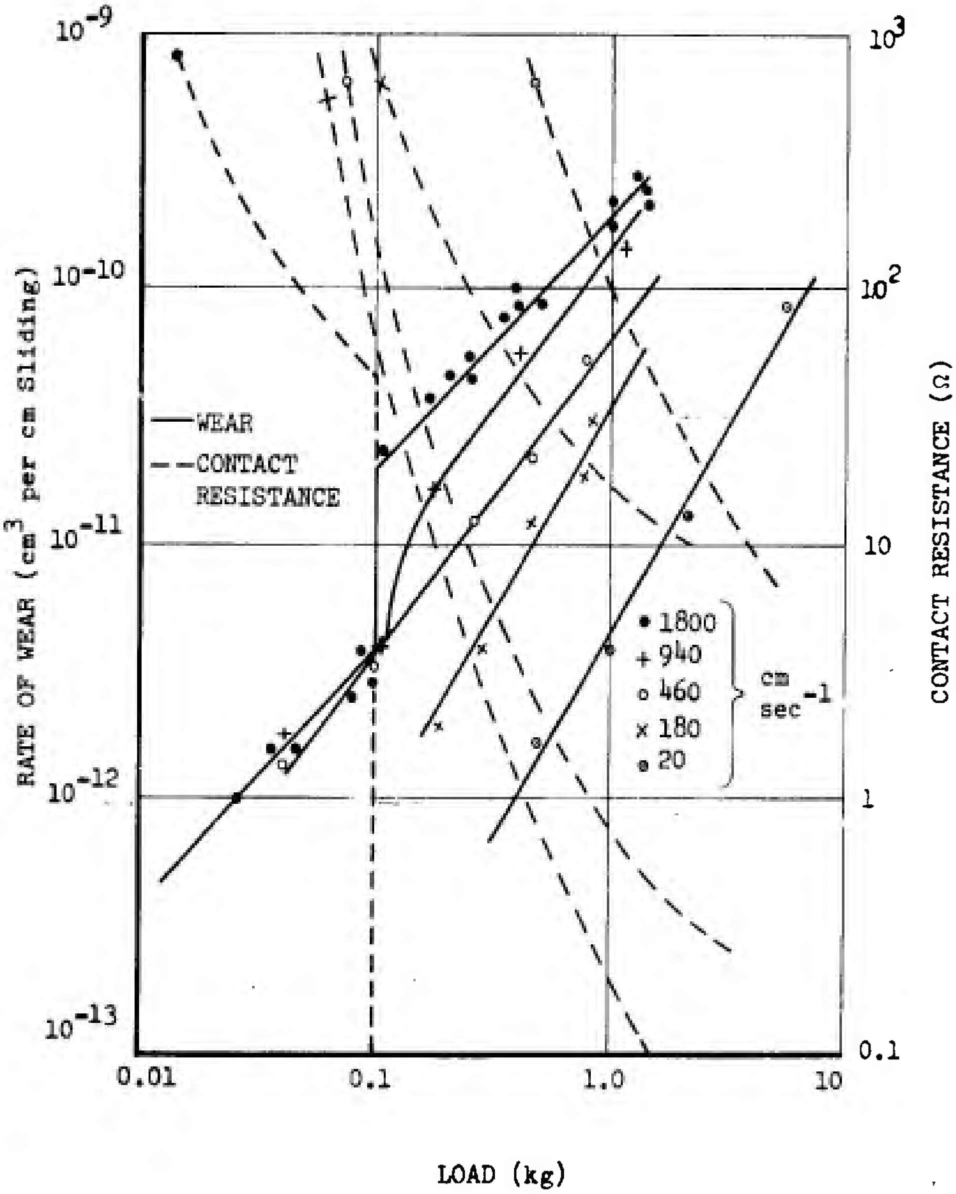
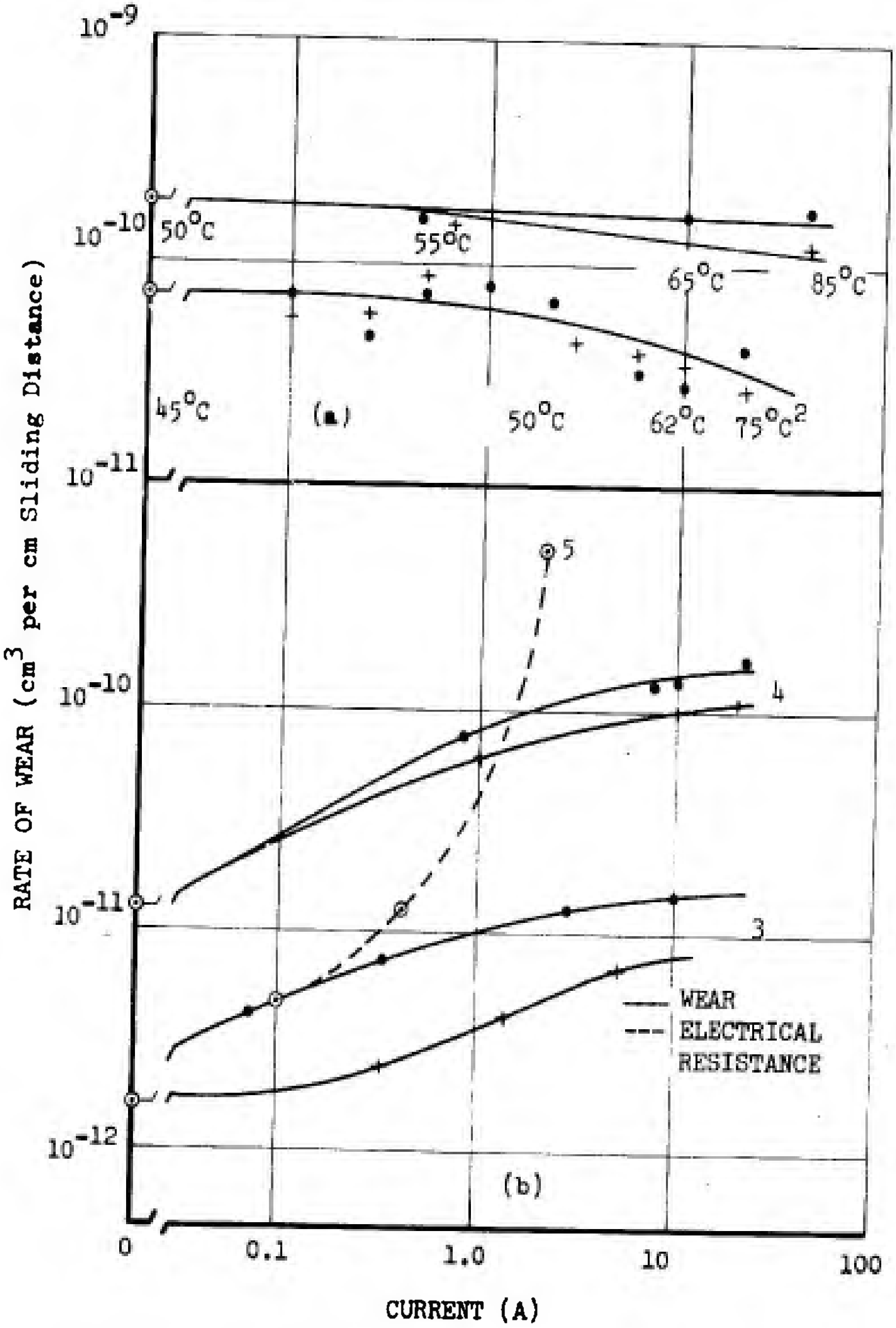


FIG. IV-8. RELATIONSHIP BETWEEN THE RATE OF WEAR AND LOAD AT VARIOUS SPEEDS (CONICAL BRUSHES)



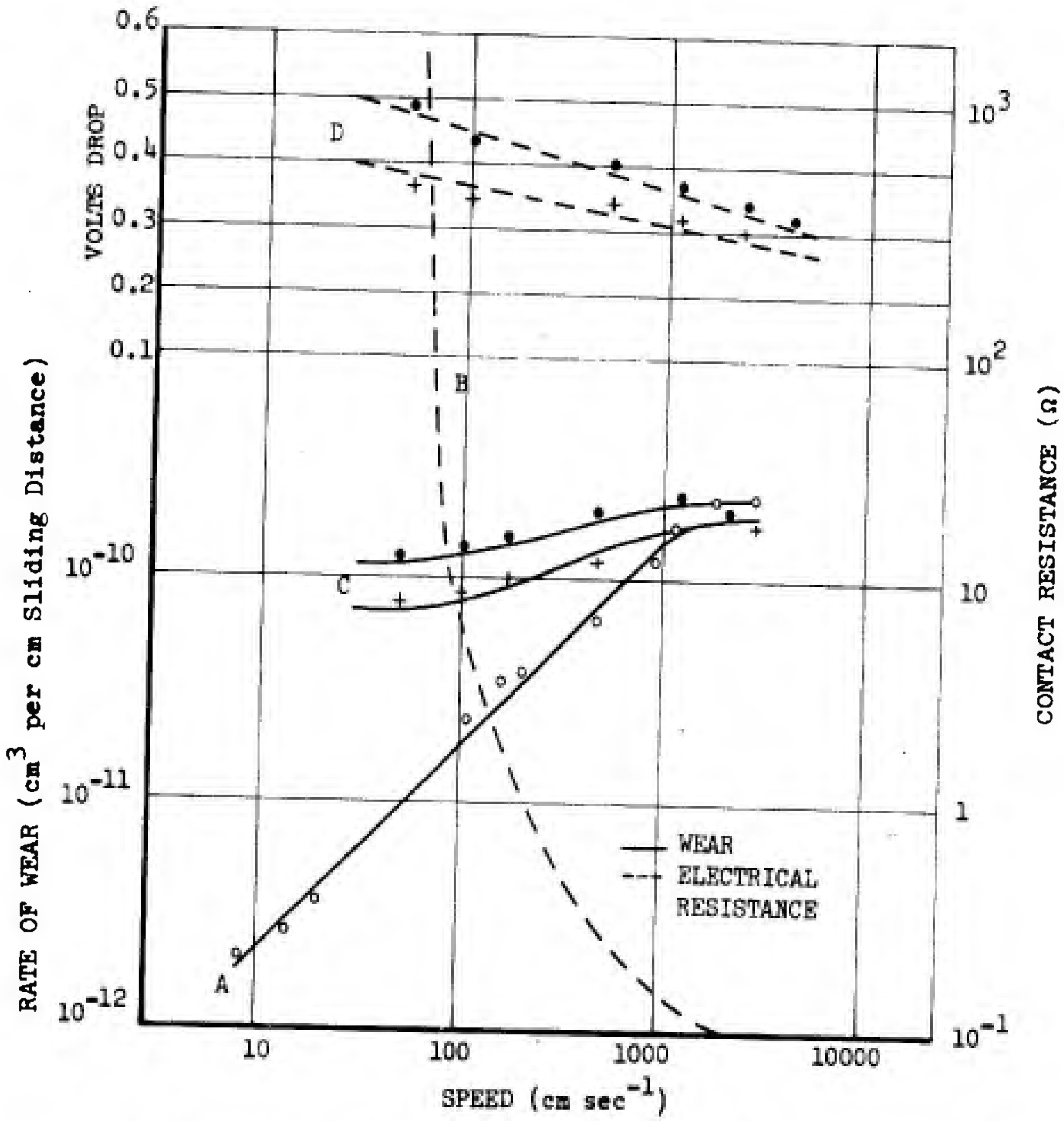
(a) Severe Wear Regine 1800 cm sec^{-1}, 1, 1 kg; 2, 280 g

(b) Mild Wear Regime. 3, 50 g, 1800 cm sec^{-1}; 4, 1 kg, 100 cm sec^{-1}; 5, 50 g, 1800 cm sec^{-1}, arcing.

FIG. IV-9. VARIATION OF THE RATE OF WEAR WITH CURRENT whole of the increase in the rate of wear at very high currents, Lancaster asserts that it does not by itself explain why, at lower currents and no arcing, the rates of wear of a positive brush and the amounts of graphite transferred to the copper are less than those obtained with a negative brush. The Cu_2O film generated on copper during sliding, however, is a semiconductor, at least in patches where the film is relatively thin (see Schröter [IV-4]). It is therefore possible that beneath the positive brush an appreciable proportion of the current may pass through the oxide without causing breakdown.

Further confirmation of a change from the mild to severe wear regime caused by the passage of current was obtained from a determination of the relationship between the rate of wear and speed, Figure IV-10. The servere wear regime is obtained at speeds in excess of about 1500 cm/sec. Curve C shows the rates of wear of the positive and negative brushes, and it is noted that the variation with speed is much smaller than without any current, curve A.

The process of material removal from brush, according to Lancaster, may be visualized as a fatigue process on a microscopic scale, i.e., the removal of a fragment from one of the surfaces is the final result of a succession of localized encounters. If the localized deformation during sliding is plastic, a few encounters will suffice to remove a fragment and the rate of wear is relatively high; for elastic deformation, however,



A, Without Current; B, Contact Resistance (without current); C, with Current (10A); D, Voltage Drop.

Load 1 kg + Brush Positive · Brush Negative

FIG. IV-10. VARIATION OF THE RATE OF WEAR WITH SPEED a much larger number of encounters will be required. In the mild wear regime, the number of local stress cycles and their magnitude is relatively low; however, in the severe wear regime, the surface of the transferred graphite layer is smooth but the coefficient of friction between the brush and the transferred layer is greater than that on oxidized copper. This suggests that the cohesion of graphite to itself exceeds the adhesion between graphite and Cu_2O, and in consequence fewer stress cycles will be necessary before a fragment is detached from the brush.

In conclusion, among all the variables studied by Lancaster, variations in load appear to have the most clear-cut quantitative effect on the variations of wear rate. By observing the physical characteristics of the film on commutator as a function of load, one can generally identify a change in wear mechanism. The fact that Lancaster was able to correctly hypothesize the result of an experiment on the basis of past experiments indicates that the mechanism of brush wear is quite well understood. However, further study is needed to better ascertain the effects of arcing and large brush consumption on the rate of wear.

IV.C.2.3 Accelerated Test

Lancaster's research study [IV-2] appears to have resulted in a coherent description of the mechanical and chemical processes that are responsible for brush wear during non-arcing conditions of sliding. However, the description by Lancaster does not include a detailed study of the influence of arcing nor any direct account of the influence of very long wear times (or large brush consumption). Knowledge of the influence of arcing is desired because motor commutators invariably cause arcing to some extent; and past experience with motor brush life tests indicates that the wear rate can change during a long period of time. Although slip rings have not been investigated in detail during this study, it would seem that they are affected by wear mechanisms similar to what Lancaster describes; particularly since a continuous ring would not cause arcing.

From considerations of the physical processes involved it appears as though a model of the damage accumulation type (Model 5 Subsection IV.B.1) can be applied to motor brush wear, since failure occurs when the brush length has been consumed; this intuitively relates the brush length to the amount of damage which can be endured prior to failure. There are two possible sources for the randomness usually observed in failure times. First, the wear process itself may be stochastic in nature. Secondly, the total permissible wear may vary from brush to brush. In Subsection III.B.2 a specific case of the damage accumulation model, called the SW model is investigated in detail. The SW model utilizes the assumption that the wear rate at time t is independent of the amount of damage accumulated by time t, although this assumption can be modified to some extent. The validity of the necessary assumptions for the motor brushes can be partly ascertained from a knowledge of the physical processes involved for a particular brush type, brush shape, brush holder, and commutator. For instance, from the changes that have been observed in brush contact area and spring tension, it appears that the wear rate at time t can, in fact, depend on the amount of wear (i.e. the length already consumed) at time t.

According to Lancaster's experiments, the brush load has a pronounced effect on the wear rate, i.e., one order of magnitude change in load causes about two orders of magnitude change in wear rate. In addition, the rate of wear appears to be affected by the load in a clear-cut quantitative manner. Therefore, implementation of an accelerated test of the brush appears to be most amenable by increasing the mechanical load on the brush. Since the amount of arcing is dependent on the amount of contact bounce, it is expected that the increases in load would decrease the amount of arcing. Therefore, as emphasized before, the conditions of arcing at the accelerated conditions should be compared with usage conditions in order to determine that there is no gross change in the factors affecting the wear process, e.g., no arcing existing at accelerated conditions but some arcing existing at usage conditions would be a gross change in mechanism.

IV.C.3 Ball Bearings

IV.C.3.1 Possible Failure Modes

The magnet readout encoder and the motor rotor are both mounted on a single shaft assembly. This shaft turns on a pair of preloaded precision ball bearings at a speed of approximately 1 rps per "g". Indicators of proper bearing performance are its compliance (shaft displacement per unit of load) and resisting torque. Excessive compliance would cause relative displacement of the magnet with respect to the induction element, and hence a change in instrument scale factor (proportionality factor relating shaft velocity to input acceleration). Erratic torque could cause an erratic velocity meter output.

The available gross evidence of past bearing failures in low-to-medium speed, low load applications tends to indicate that some form of wear has been the most frequent cause of failure. Generally, the area most susceptible to adhesive and abrasive wear is the retainer-race contact area because it is difficult to establish thick film lubrication between these elements, especially when the retainer is of the metal type. It is also possible that the two halves of a metal retainer can break at their coupling links. Impregnated retainers can shrink, peel, or crack. A bearing can be contaminated by some source such as wear particles from motor brush wear. Therefore, there are many reasons why bearings can and do fail.

However, if a bearing is well designed so that there is thick film lubrication at usage speeds and loads, it is expected that when it fails, the cause of failure will not be of the wear-induced (adhesive or abrasive) type. Instead, the expected failure mode would be rolling-contact fatigue, or lubricant breakdown due to age. Recent designs of miniature bearings for medium speed, low load applications having thick film lubrication between the balls and races and using impregnated retainers are subjected to extremely small amounts of wear. However, the rolling contact stresses are in the order of 100,000 psi or more. Therefore, it is reasonable to expect that some instrument bearing applications will be susceptible to the rolling-contact fatigue and lubrication breakdown modes of failure.

The most prominent factor that contributes to lubricant failure has usually been oxidation (see [IV-5]). The results of oxidative deterioration are often easily confused with other factors, for instance, thermal deterioration. When a lubricant starts to fail through oxidation, the result is usually the formation of both soluble and insoluble compounds that may appear as resins, sludges, or acidic compounds. These effects in turn manifest themselves as follows:

(i) Viscosity of lubricant gradually rises.

(ii) Adherent surface deposits form that many interfer in small clearance spaces.

(iii) Metal surfaces become corroded.

Certain metal surfaces and small wear particles exert a pronounced catalytic effect in promoting the oxidation rate of many lubricants. Iron may cause such an effect with one lubricant while copper alloys affect others. Freshly exposed surfaces due to wear processes are significant catalytic sources.

Oxidation is generally controlled by the addition of inhibitors that retard the rate of oxidation. A time lag, called the induction period, exists before the effect of oxidation (as measured by the pH number) becomes significant, after which a rapid rise in acidity or viscosity of the lubricant is usually noted. The effect of the oxidation inhibitor is usually one of increasing the induction-period time after which the same symptoms appear. Other approaches to controlling oxidation are to limit temperature and to eliminate potentially reactive surrounding and absorbed gases.

The inference should not be drawn, however, that all the effects of oxidation are harmful. In many lubricants, small quantities of oxidation reaction products are beneficial and even necessary to the lubrication process.

Thermal stability is a fundamental characteristic of the pure lubricant; it remains essentially invariant in a class of lubricants which may differ in physical and chemical structure due to the presence of various additives. Thermal instability is difficult to distinguish from other high-temperature effects such as oxidation.

Essentially, thermal instability refers to changes in the lubricant that are not dependent on external reactants. Therefore, to study the thermal stability properties of a lubricant, all surrounding and absorbed reactive gases or contaminants should be avoided.

Rolling-contact fatigue is manifested as a damage to the surface by local pitting or flaking. Normally, rolling-contact fatigue is a function of the number of stress cycles to which a given unit volume (or surface area) is subjected. Fatigue life clearly is not a deterministic quantity, hence the number of stress cycles to failure should be considered as a random variable. Weibull [IV-6] suggested that the distribution function of that random variable should be the one nowadays called the Weibull distribution. However, the Weibull distribution does not appear to be a good fit of experimental data in the region of relatively early failures, approximately below the tenth percentile of the failure distribution nor in the region of long lives, approximately above the sixtieth percentile (see [IV-5]).

Many investigators have been concerned with rolling-contact fatigue. Among the variables affecting rolling-contact fatigue that have been studied are load, speed, lubricant base stock, lubricant viscosity, temperature, and such processing and design variables as material composition, melting techniques, hardness, fiber orientation, and contact angle. A large portion of the investigation of these variables has been with easily fabricated speciments such as balls or cylindrical rods.

The life of bearings is frequently specified by stating either the tenth or the fiftieth percentile of their failure distribution, either of these being referred to as bearing life. Extensive experimental data exist which make plausible the contention that bearing life, with respect to rolling-contact fatigue, is inversely proportional to the third or fourth power of the load for a variety of materials and lubricants (see [IV-7] to [IV-11]). For fatigue of structures, it is generally found that the number of cycles to fatigue failure is independent of the rate at which the cycles are accumulated, except at high temperatures. Though the effect of temperature on rolling-contact fatigue has been studied often, no general rule concerning a quantitative relation of fatigue life to temperature has been established. These facts suggest that the load-life relation and the possible independence of life from speed (at least for some range of speeds) are the most feasible ways of implementing an accelerated life test for rolling-contact fatigue. However, thorough consideration needs to be given to each particular application in order to justify the implied extrapolations. Some of these considerations are discussed in Subsection IV.C.3.2, below.

IV.C.3.2 Accelerated Test

Consider first the over-stress approach to accelerated testing of rolling-contact fatigue. The increased stresses used in the accelerated tests should not cause boundary lubrication (a thin film allowing some metal-to-metal contact) or any inelastic deformations. If boundary lubrication existed between the balls and races, it would be expected that wear could take place and, hence, ultimate failure could be due to an interacting combination of wear (adhesive or abrasive) and fatigue (pitting and flaking). It might be argued that such a mix of failure modes does not preclude the possibility of accelerating one mode; one merely has to consider the occurrence of failures as a competing risks process (see Section VI.A) in which the fatigue-failure risk is increased (in a controlled fashion) by increasing the load. However, the contact stress and temperature would be considerably higher for this situation than what would exist (for the same load) during thick film lubrication, so that the fatigue life-load characteristics are likely to be different. Therefore, there would be little justification for using the fatigue life-load parameters estimated from the high stress test for extrapolation to lower stresses. Such change in failure mechanisms might be tolerable if data existed which supported the contention that fatigue life-load parameters were not significantly affected by such changes in failure mechanism.

Accelerated testing can be attempted by increasing the rotational speed of the bearing, thus (implicitly) relying on a time-compression approach. Recall that the time-compression approach is predicated on the assumption that the failure distribution (in terms of some pertinent random variable) is not changed by a change in experimental conditions. In the current instance, the assumption is that the distribution of the number of cycles to fatigue failure remains unchanged when the bearings rotate more rapidly. However, there are at least four fatigue-influencing variables which might be affected by changes in bearing speed. At high speed, centrifugal force on the balls changes the load at the ball-race contacts; bearing temperature may rise; the mechanical lubricating properties (preventing metal-to-metal contact) might change; and the chemical lubricating properties (minimizing the damage that can occur in intermittent contact) might change.

The change in the number of cycles to failure, due to some of these factors, may be insignificant, depending on the particular bearing type, materials, and application. For small bearings (such as might be used in inertial instruments), the centrifugal force will generally be less than one half pound for speeds under about 5,000 rpm. For instance, with 3/32 inch diameter steel balls in a bearing with one half inch outside diameter at 5,000 rpm, the centrifugal force is approximately one fourth pound, which is generally insignificant compared to applied usage loads, and even more insignificant if the applied load is increased, along with the speed, for accelerating the fatigue failure.

Bearing temperature influences rolling-contact fatigue life in several ways. Both viscosity and pressure-viscosity coefficients of lubricating fluids decrease with rising temperature (see [IV-12]). Carter (see [IV-13]) has shown that these physical properties have a significant effect on fatigue life. Other factors such as lubricant chemical activity and stress-induced solid-state reactions in the bearing materials can be influenced by temperature. Carter [IV-13] reports fatigue data obtained with M-1 steel balls and a sebacate lubricant at 100°, 250°, and 450° F in a fatigue spin tester. Weibull plots of these data, given in [IV-13], are reproduced in Figure IV-11.

The plots of Figure IV-11 indicate that, for reasonable changes in temperature (say 50° F or less) due to increase of speed or ambient temperature, the resultant change in fatigue life might be insignificant. However, it is noticed that, though the ten percent life is shorter at the highest temperature, the fifty percent life at 400° F is longer than at 250° F. [Observe that it is not known, at this time, whether the difference is real or merely apparent; since the curves are only a more-or-less forced fit of a Weibull distribution, and no formal comparison was performed, it is not certain whether the apparent discrepancy is statistically significant.] Carter attributes this disparity to the observed polymerization of the lubricant into a very viscous residue at the high temperature. Since viscosity affects fatigue life, the formation

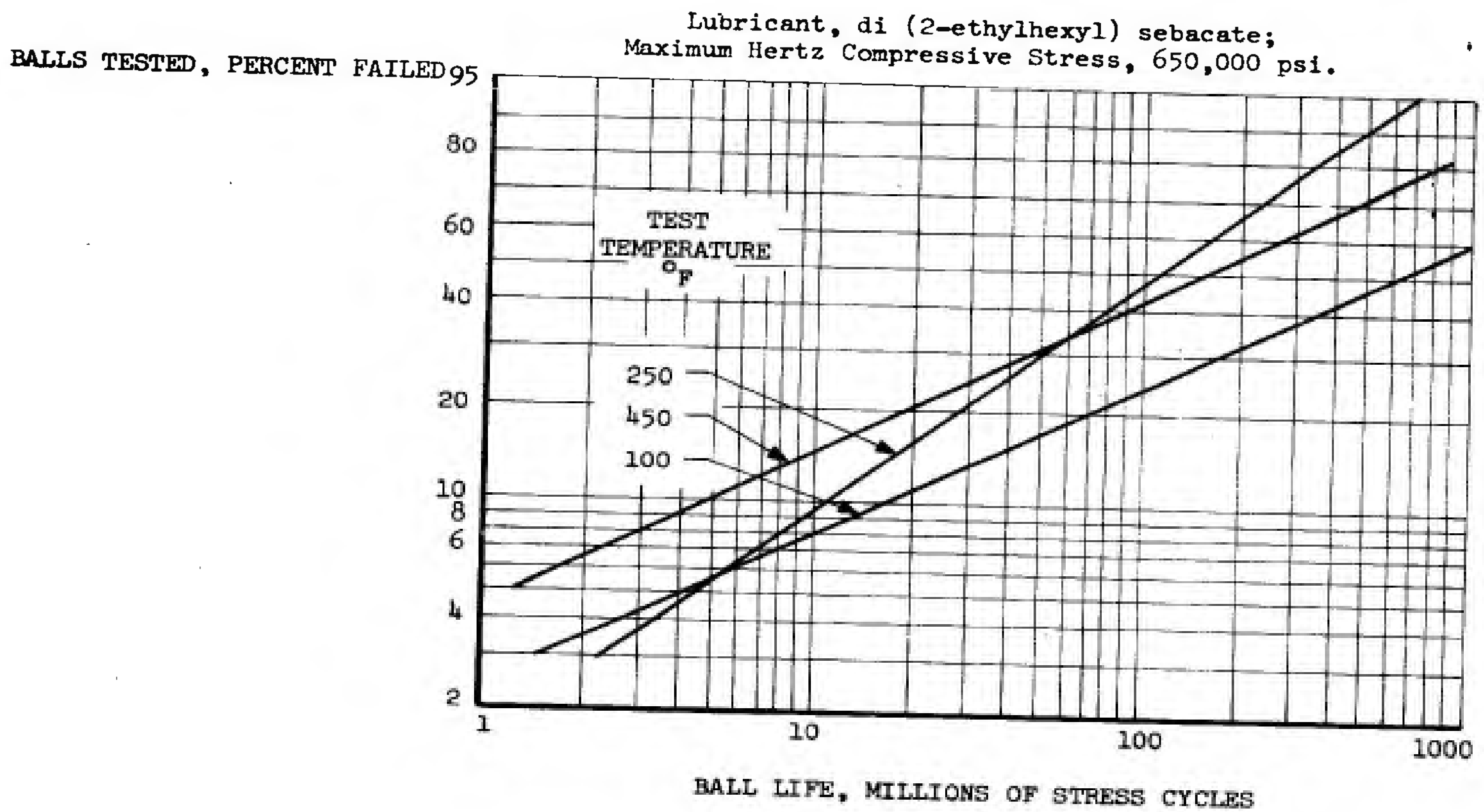


FIG. IV-11. FATIGUE LIFE FOR 1/2-INCH AISI M-1 BALLS AT VARIOUS TEST TEMPERATURES of a very viscous film with longer test time would tend to extend fatigue life; this effect would be more pronounced for longer-lived fatigue failures. Another investigator (Jackson [IV-14]), using the spin rig method of fatigue testing, found that the ten and fifty percent lives showed opposite trends in the comparison of different materials and lubricants. Whereas the results of Carter indicate a sensitivity on temperature that might be insignificant for expected small changes in bearing temperature, the anomaly makes one suspect that a test of longer time duration (hence, at a slower cycle rate) would show a wider divergence in ten percent fatigue lives for the same temperature differences because of time dependent (as opposed to cycle dependent) chemical reactions. This would suggest that in accelerated tests implemented by increasing the bearing speed, the tests should be performed at several accelerated speeds to attempt to ascertain whether any time dependent processes are of primary, or secondary, importance.

Consider next an accelerated test for lubricant breakdown. Recalling that lubricant breakdown can occur due to oxidation, thermal instability, or even evaporation, testing at higher than normal temperatures would surely accelerate the failures. However, the relative importance of the three modes at both usage conditions and accelerated conditions should be known before one attempts to make any extrapolations. Especially, the relative importance of oxidation and thermal instability should be ascertained since both processes are reported to be manifested by the same symptoms (see [IV-5]). The principal concern is that the dominant mode at the high range of temperature may not be the same mode that is dominant at the lower range.

In addition, one should attempt to take into account the extent of influence that wear particles, other contamination, or the lubricant churning action has on these modes. This would be unnecessary, to some extent, if the bearing was run at usage speed and load conditions. That is, the decreased time to failure due to the increased temperatures would tend to decrease the amount of influence possible from churning or any wear that is present under normal conditions of speed and load; however, the decreased viscosity at the higher temperatures might tend to cause more wear and churning so as to compensate for the decreased testing time.

A very practical problem clearly arises if the responsible mode of failure cannot be correctly diagnosed. One such situation, which at this point is only hypothetical, is the following. Suppose that it is desired to accelerate the lubricant breakdown mode of failure by increasing the temperature. At some stage of the lubricant breakdown, it is expected that oxidation products would raise the stress concentrations, thus increasing the possibility of fatigue pitting. Therefore, what could be a valid lubricant breakdown failure might be mistakenly diagnosed as a fatigue failure, unless it is obvious that the lubricant had deteriorated previous to the fatigue pitting.

In conclusion,the extent to which the physical mechanisms responsible for failures are amenable to being controlled or defined is probably the largest factor affecting the feasibility of realizing a meaningful accelerated life test run without a controlled life test at usage conditions. In other words, the mechanism of failure at usage conditions for a ball bearing can be a complex combination of other defined mechanisms; for example, at usage conditions, thin film lubrication between the races and balls could mean an undefined combination of wear and rolling-contact fatigue. On the other hand, if thick film lubrication existed, the failure mode would be more defined, i.e., rolling-contact fatigue. In the latter case, the failures could be accelerated by increasing the load. By means of an electrical resistance measurement, it could be ascertained that thick film lubrication existed at the test conditions. The observed failures could then be extrapolated to usage conditions with some confidence. In the former case, an extrapolation could be justified if data existed which supported the assumption that the fatigue life-load parameters were about the same at usage conditions as at the accelerated test conditions. For well designed bearings, obtaining thick film lubrication is not unrealistic. For example, for a miniature bearing design using a lithium soap grease as lubricant, Constantine [IV-15] found that the film thickness was relatively constant and seemingly provided complete support for static and medium speed (to 1200 rpm) and low load conditions. The power-law which apparently characterizes the dependence of bearing fatigue-life on load is used in Subsection III.C.2 in connection with a restricted damage-accumulation model of failure. The fact that Prot [IV-16] was apparently successful in using the damage-accumulation concept in connection with fatigue, leads one to believe that the model developed in Subsection III.C.2 can be applied to the fatigue life of ball bearings.

IV.C.4 Pump

IV.C.4.1 Possible Failure Modes

The VM4A pump supplies pressurized liquid to the hydrostatic bearing of the sensing element. The pump is of a reciprocating piston-cylinder design, but with the piston stationary and the cylinder moving. It is of a single-acting, solenoid-cocking, spring-actuated type, cocked electromagnetically by rectangular pulses.

The stationary piston of the pump has a broad flange at the top, and is hollow, with a ball check valve at its center. The pump cylinder fits over the piston closely in the usual manner. The small flange at the top of the cylinder serves as a detent or stop for a compressed helical spring, which winds closely around the cylinder and bears against the flange at all times. Outside the spring is a coaxially wound solenoid. Below the piston,in the hollow of the cylinder, is a second ball check valve. The cylinder also serves as the magnetic core for the solenoid, and is commonly referred to as the solenoid core. When the cocking solenoid is energized by a rectangular pulse, it forms an electromagnet with the core (pump cylinder); the core is pulled axially along the centerline of the solenoid, through its entire travel. During this stroke the spring is further compressed, the lower valve (in

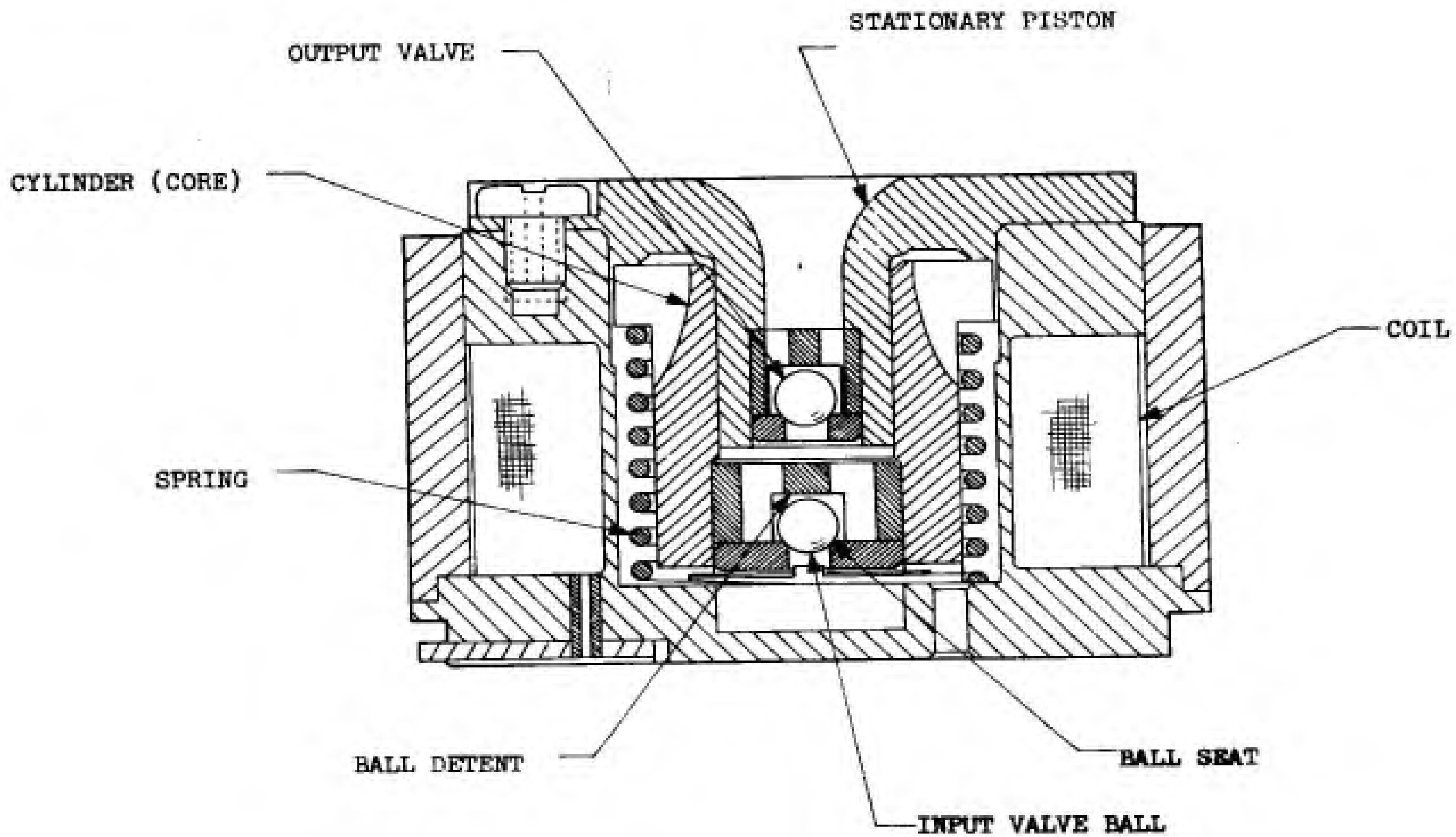


FIG. IV-12. DETAIL OF A VELOCITY METER PUMP the cylinder) opens, and liquid passes through to the piston. At the end of the stroke the lower valve closes, trapping the liquid above it. The core is held in this cocked position until released by the solenoid magnet at the end of the pulse. The spring then pushes the core upward, forcing the liquid through the piston check valve and through a filter (above the piston flange) to the bearing, where it is discharged through four openings. Return flow during the next cocking stroke is prevented by the closure of the upper check valve. The foregoing sequence is repeated several times per second.

Since the pump operates continually while the velocity meter is in service, the problem of wear becomes critical, imposing a limitation on the lifetime of the instrument. This difficulty is surmounted by using a short stroke and by making the cylinder and piston, which are the parts most subject to wear, of very hard materials and with polished surfaces. The Fluorester bearing liquid, in which the pump is totally immersed, has excellent lubricity, and thus serves the further function of lubricating the moving parts of the pump.

Excess output pressure instability and poor pressure flow characteristics are failure symptoms. The required pressure flow characteristics are determined by the specific instrument performance needs. In general, because a fluid, sliding surfaces, and impact forces are present, the failure modes could be fluid contamination, adhesive wear, abrasive wear, fatigue wear, or deformation. Specifically, the following failure modes of the VM4A instrument which are associated with the pump could be postulated:

(i) Wear of the sliding surfaces between the piston and cylinder.

(ii) Some sort of wear or local deformation at each of several surfaces subjected to impact forces, i.e.,

a. Either ball-retainer seat surface.

b. Either ball-retainer detent surface.

c. Detent surfaces between the piston and cylinder.

d. Detent surfaces between cylinder and housing.

(iii) Contamination of the fluid by pump wear particles.

(iv) Change in spring constant.

Functional test and diagnostic examination of pumps which have seen service have revealed failures traceable to deformation (of the order of several thousandths of an inch) of the ball-retainer detent surfaces of the inlet valve. Damage at the piston-cylinder detent surfaces has also been observed although no failures have been specifically attributed to it. The observed failures were either the pump failing to function at all or an instability of the output fluid pressure. It was postulated and experimentally confirmed that the failure was due to increased ball "play" resulting from deformation of the retainer at the ball-retainer detent contact surface. The increased "play" caused improper valving and decreased pump efficiency.

The deformation of the ball-retainer detent surface was discovered relatively early in the pump development history, and corrective action was taken. Based on some engineering tests with different materials, another material was selected. The test involved operating the pump at an increased frequency and using a fluid less viscous than what was normally used so that the impact forces were greater. This test was meant to be only an approximate measure of the relative lives of the inlet ball-retainer detent surface for the different materials. An accelerated life test of the pump was begun during the pump development program using the new material in order to re-evaluate the life of the same retainer part made with the new material. The life test is of the time-compression type, as described in detail below. Only one failure has occurred to date in this test. That failure was due to the same retainer part being slightly deformed.

IV.C.4.2 Accelerated Test

The design of an accelerated life test of the pump should obviously require detailed consideration of each specific mechanism of possible failure. One typical consideration would be to ascertain whether or not there is thick film lubrication between the piston and cylinder. Special model pumps or ingenious measurement techniques might have to be invoked in order to accomplish this, but it could obviate the need for an accelerated test of a wear failure mode.

Furthermore, the design of an accelerated life test for such a failure mode (i.e., wear resulting from metal-to-metal contact), would be difficult because the actual manifestations on instrument operation are not known. In other words, would the ultimate effect of wear be a seizure or sticking of the pump itself, or would the wear particles cause sticking of the hydro-static bearing that supports the pendulum float? The first possibility would allow the pump to be tested by itself, whereas the second might require that the total operating instrument be used.

Other considerations for the implementation of an accelerated life test will be exemplified by further discussion of the one failure mechanism that has been observed, i.e., the inlet ball-retainer detent surface becoming slightly deformed from the impact against the valve ball. Two alternatives for accelerating the failure will be discussed. These are increasing the stress induced by the ball impact or increasing the pump cycle frequency.

Consider first the possibility of increasing the impact stresses. Note that the impact under consideration occurs only during the cocking stroke. The motion of the detent is identical to that of the moving cylinder, and the motion of the ball before impact is affected by the fluid flow through the inlet hole. The relation between the extent of impact deformation and any description of the dynamics of the collision or the number of impacts is not known for any metal-fluid systems. Furthermore, an analytical description of the ball-cylinder dynamics involves several parameters that are unknown or only approximately known.

Therefore, it would appear advisable to simply increase, to the extent possible, the electromagnet drive current that cocks the cylinder. Due to the consequent increase in impact force, the failures would be expected to occur sooner than under normal conditions. Obviously, in order to extrapolate the failure data at accelerated conditions to usage conditions, a model relating cycles-to-failure to the electromagnet drive power would be needed. A model of the damage-accumulation type, as described in Section III.B, might prove to be applicable. The possible increase in impact force would be limited by the allowable power dissipation and the extent of magnetic saturation.

The implementation of the alternative method, i.e., increasing the frequency, requires justification of the implied assumption that the failure mechanism in question is unchanged by the increase in cycling rate. Hopefully, the only time dependent variables affecting the inlet valve retainer are those occurring within each pump stroke period. In other words, the implied assumption is that there are no chemical reactions or metal recovery effects that have a significant influence except possibly within each pump stroke. The pertinent description of the pump operation is the motion of the inlet ball and retainer as a function of time. It is necessary to ascertain that the ball and retainer motions as a function of time (the time being referenced to the start of each cycle) are unchanged at the increased cycling rate. However, it is adequate to simply describe the motion of only the retainer (or equivalently, the moving cylinder) since the ball within the valve comes to its terminal displacement in less time than the cylinder stroke displacement.

Recall that the impact in question occurs during the cocking stroke. Both the cylinder velocity and displacement at the beginning of the cocking stroke must be the same, respectively, at the accelerated frequency as at the usage frequency. Since the initial conditions for the cocking stroke are also the final conditions for the pumping stroke, the pumping stroke is required to be unchanged. Therefore, strictly speaking, the cycle time period must be no less than the time required for the cylinder to come to rest at the end of the pumping stroke plus the cocking time. The extent of time compression possible with this approach is limited by the ratio of the cycle time period during normal usage to the time period required for the cylinder to complete a cocking and pumping stroke. However, if the plunger comes to rest at the end of its pumping stroke during usage conditions, it is possible to exceed this ratio. The increased time compression is accomplished by changing the output resistance to flow (while at the accelerated frequency) until the same amount of fluid is pumped, per cycle, as at usage conditions. By this indirect compensation the plunger stroke length is made the same as at usage conditions; and as a consequence, the final plunger velocity is the same, namely zero. Hence the inlet valve damage should be unchanged. However, the damage per cycle at the outlet valve and plunger detent surface will be changed by the flow compensation.

IV.C.5 Bellows

IV.C.5.1 Possible Failure Modes

Metallic bellows are typically used for compensation of thermal expansion of fluids in inertial instruments. The purpose of the compensation is to avoid excessive fluid pressure changes accompanying temperature changes.

Failure can occur by the development of cracks through which the enclosed fluid can escape. Even though no leaks of lubricant fluid have been reported, it has been shown that such leaks are possible. Life tests have been conducted using helium gas to detect cracks. A leak rate of 10^{-7} cc/sec of helium was calculated to be equivalent to a leak rate of lubricant fluid that would cause instrument malfunction within three years. Helium leak rates of 10^{-7} cc/sec, or greater, were observed, though generally after more than fifty thousand expansion-compression cycles. The crack-causing mechanism may be one of fatigue and some time dependent process such as corrosion.

IV.C.5.2 Accelerated Life Test

If fatigue is the only significant mechanism during usage conditions, failure data can be obtained more rapidly by merely deflecting the bellows at a rate higher than that encountered in use. This accelerated test would require a pilot experiment to decide whether chemical action contributes significantly to crack formation during normal usage. This could involve performing tests at high temperature, both with and without the possibly corrosive liquid or atmosphere, since the increased temperature would accelerate the chemical action (if any).

Besides attempting to consider the significant failure phenomena occurring at usage conditions, one should give consideration to the fact that an increased cycling rate, and the manner of achieving it, may influence the mechanisms of change in a manner different than at usage conditions. For instance, if the bellows were flexed at a high frequency, the resulting increase in local temperature could possibly be sufficient to affect the metallurgical properties or any chemical reactions. Another spurious influence at high frequency could be a change in mechanical stresses due to wedging of the fluid contained by the bellows.

An alternate approach could be a time series extrapolation. This would be possible if the manner in which the helium leak rate increases with time were known with some generality. This could also offer a more sensitive criterion of failure which would be useful in empirically determining, in less time, the effects of increased cycling rate.

IV.D GYROSCOPE

IV.D.1 Instrument Description

The Autonetics G6 gyroscope is of the free-rotor type, with a gas-lubricated spin bearing. It is designed for ballistic missile application, i.e., high g level capability and high reliability. The principal components of the gyro are a rotor and spin-bearing, a stop-bearing, a motor, pickoffs, torquer, and hermetically sealed case. The rotor (a spinning wheel) spins on a bearing that permits three degrees of angular freedom about the support. The gas bearing is spherically shaped and concentric with the center of mass of the rotor. There is a thin (approximately 60 microinches) film of helium gas between the stationary ball and the cavity within the rotor. The gas pressure necessary to support the load is provided by the hydrodynamic action accompanying the relative angular motion of the rotor and ball.

The rotor has a conductive rim and is driven by eddy-current induction effect from an axial-flux polyphase stator mounted on the case. A two-axis pickoff detects angular deflection of the case relative to the rotor and controls a servo system for moving the case to nullify the deflection. To avoid excessive drift rate, due to a component of the motor torque perpendicular to the spin axis, the feedback system must hold a very close null. Since large angular accelerations could cause the rotor to rub temporarily against the shaft, or the motor induction

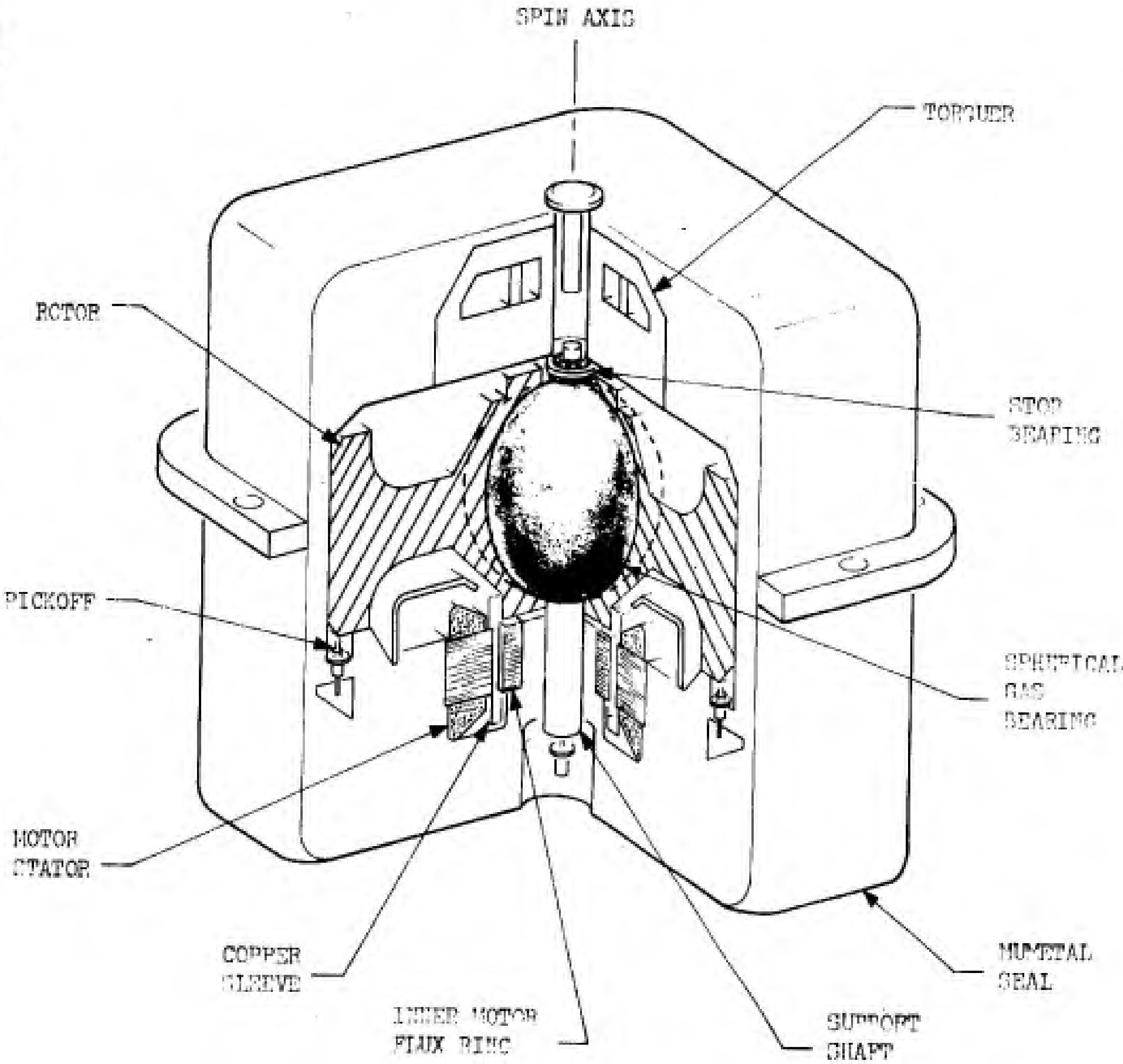


FIG. IV-13. SIMPLIFIED DRAWING OF FREE-ROTOR GYRO rim to rub the motor stator, a ball bearing (called the stop bearing) is mounted on the shaft, and this bearing is contacted before any such rubbing can occur.

IV.D.2 Possible Failure Modes

Changes in material dimensions and contamination of the bearing are two factors affecting the life and reliability of the gyroscope (with respect to the gas spin bearing). Eventual manifestations of these factors can be a failure of the rotor to start, erratic output angle, precessing (biased) output, and decreased capability of the bearing to support acceleration loads. These factors (or failure modes) are discussed next.

Changes in material dimensions (generally referred to as creep) have two basic causes. Small dimensional changes characteristically occur when any material partially or completely undergoes a solid state phase change or other change of state. This movement is determined by the phases present and their degree of instability and varies as a function of time. The other basic cause of instability results from an applied stress, a residual stress, or a combination of both.

The instability due to phase changes in metals can usually be eliminated by using pure metals which have been annealed or otherwise treated to be as close to thermodynamic equilibrium as possible.

However, pure metals so treated are invariably too weak for satisfactory design application.

Instability due to applied stresses does not need to be a significant problem in most instrument applications because it is generally possible to design the instrument so that all applied stresses are so low that a significant amount of creep is not possible. It has been reported by a number of investigators [IV-17], [IV-18], and [IV-19] that a stress level sufficient to cause one microinch of permanent set in a short-time test (which stress level is called the precision elastic limit, or PEL stress) causes no creep within a period of a year. It is not known whether this effect is real, or whether it merely reflects the investigators' inability to measure the creep on the level at which it occurs. Most of the investigators have reported an accuracy of five microinches per inch over an extended period of time (up to one year). No measurable microcreep was detected by the Autonetics Metallurgical Laboratory in beryllium tests when loaded to the PEL stress [IV-20].

Residual stresses can be developed in many ways but generally fall into two categories; i.e., those produced by fabrication and assembly of parts, and those resulting from temperature changes. Forming, machining, and joining are examples of fabrication techniques that can lead to assembly-induced residual stresses. Quenching and incompatible rates of thermal expansion of adjacent materials are examples of sources of stress induced from temperature changes. Since the residual stresses are relieved by creep, the dimensions should stabilize to some final value if no new residual stresses are induced. An extensive discussion of creep can be found in Part V of this report.

Contamination of the bearing can come from sources both external and internal to the bearing. Possible external sources are dust or other foreign material, and condensable vapors. A possible internal source is wear particles generated during the starting and stopping of the rotor. It is believed that any dust in the bearing will be detected some time during the many hours of testing to which a gyroscope is subjected before it is used in the field. However, the effects of condensable vapors may take a long time to be significant. This failure mode is due to the progress of the high and low pressure zones on the rotor cavity across the surface of the ball, which causes the pressure of the lubricating gas to pulse by a factor of about five in a manner believed to approximate the isothermal mode. Since gas is constantly entering on one side of the bearing and departing on the other, the system acts to dry condensables out of the gas stream in the bearing, and perforce, to evaporate them from elsewhere in the gyro. Materials of very low volatility should merely take a long time to get into the bearing annulus.

The vapor condensation mode of failure is dramatically observed when the gyro is run in moist air because the required motor power will rise markedly, and the rotor sometimes cannot be kept running. All known liquid substances have viscosities much too high to permit operation of a G6 gas bearing into which they have intruded sufficiently to bridge the gap. The influence of such pumping out upon formation of gummy deposits is unknown, but some appraisal may be given of the relative importance of the mode. It is surmised that five to ten milligrams of liquid contaminant such as water or oil vapors would be the normal lethal dose. For future gyros, the properly lubricated stop bearing will not, after centrifuging, retain more than three milligrams of oil. The O-ring bears about 1 milligram of removable volatiles, and the metal surfaces much less. The encapsulated motors have been a possible source of volatiles, 80 to 100 milligrams, but present understanding of how one must bake out assemblies and detail parts effectively nullifies this failure mode. Hence, accumulation of slightly volatile "gunk" in the spin bearing following long operation appears to be unlikely in future models.

The generation of wear particles of sufficient size or quantity to prevent restart is the inevitable result of start-stop events. Despite the existence of a deliberately applied boundary lubricant film, the contact of asperities on the metallic surfaces will always produce some wear. The energy of the spinning rotor at the moment of contact, or between start and lift off, must be converted into heat by some non-elastic microprocess. Processes of this type are:

(i) Viscous disturbance of the lubricant film.

(ii) Displacement or squeezing out of the absorbed final layer of boundary lubricant.

(iii) Rupture of lubricant molecular chains.

(iv) Plastic deformation of metal surfaces at points of contact.

(v) Shearing-off of over-stressed or fatigued metal at contact points.

Only the last-named process generates particles. The first three processes noted are harmless. Some of the considerations regarding the boundary lubricant are reviewed in the next paragraph.

It was learned early that gas-lubricated spin bearings would not start in practice if they were clean and dry of any lubricant. After extracting with a polar solvent and hard vacuum bakings, such a bearing did not turn with available motor torques. The same kind of behavior prevails with ceramic bearings. Thus, a molecular layer of adsorbed gas and water vapor is not adequate; the deposit left by lens paper and reagent acetone will work, and the deposit left after vapor degreasing a rubbed-in coating of chlorotrifluoroethylene (CTFE) seems about optimum. The plating of the ball is conventional hard chromium and, therefore, porous and full of cracks which pick up and retain the lubricant which is a viscous fluid. The CTFE is a powerful boundary lubricant for this system, and CTFE fluids have noteworthy stability especially in oxidative environments and in contact with metals. Gyros filled with this material as a floatation fluid have performed at 130° F for many years with no evidence of chemical degradation or fluid-induced corrosion.

From the above discussion of apparently possible modes of failure for the G6 gyroscope, it is apparent that most failure modes are essentially eliminated though the gyro's ability to sustain repeated start-stop cycles is a limiting factor. The projected operating life of the G6 gyro is extremely long when cumulative processes are considered. The helium seal appears to be essentially permanent and is amenable to leak tests that would detect leak rates of a magnitude that would decrease the gas pressure by only one psia in twenty years. The sealing material is corrosion resistant, ductile, and stress free; therefore, there is little reason at this time to believe that leaks will develop, the device having once been free of leaks. The highest operational stresses in any beryllium part are well below the precision elastic limit for "zero" creep. The possible deposition of condensable vapors in the spin bearing is controlled by rigorous exclusion of volatiles from all materials of construction in addition to the normal vacuum bake.

In long space missions, constraints on power consumption might dictate that the gyroscope be turned on and off several times. Therefore, it would be advisable to investigate the probability distribution of the number of successive starts which a bearing can perform. Little information is available regarding that distribution. In terms of data available from field usage, there has been no known case of a G6 gyroscope spin bearing failing to start while in service, except after being damaged due to excessive vibration or shock loads. Engineering tests provide some information, and indicate that some bearings can sustain in excess of 1800 start-stop cycles. However, the engineering tests were performed with small sample sizes (one to three); furthermore, no control experiments were used to determine whether the changed cycling rate (higher in test than in practice) has any effect on the distribution of cycles to failure. Note that it appears that a bearing that has failed to start can be cleaned and restored to like new condition; hence, large samples for start-stop tests can possibly be obtained without using very many gyroscopes, by merely using restored bearings in lieu of additional samples.

IV.D.3 Accelerated Test

Among the possible failure modes considered above, the failure of the spin bearing to start is the most amenable to an accelerated life test. A reduction in test time can be achieved by cycling the instrument more often. However, precautions must be taken to perform the experiment in a manner which avoids any possible spurious effects of the increased frequency of cycling. That is, the failure process may involve some time-dependent phenomena (superposed on the obvious dependence on the number of start-stop cycles). For instance, recovery of lubricating properties after a stop may require an appreciable amount of time; increasing the cycling frequency to the point where this recovery is incomplete in any cycle would modify the distribution of number of cycles to failure. Preliminary analysis has indicated that cycling at the maximum possible rate (determined by the lift-off time of the gyroscope) should be free of such spurious effects. However, a controlled experiment should be performed at a lesser frequency; data from the two experiments should be compared by testing the hypothesis that both sets come from populations with the same distribution function.

V. MICROCREEP PHENOMENA

V.A INTRODUCTION

V.A.1 Effect of Microcreep

Present and anticipated guidance needs for space travel place very stringent requirements on materials used for critical electromechanical components. In the design of the ultra precise parts used in guidance apparatus, it is essential that account be taken of minute changes in dimensions which may occur during flight. In order to design the system to prevent failure due to material movement, tests must be conducted on these materials to determine the extent of the movement (subsequently referred to as microcreep) at the operating stress. The usual tests must simulate the operating time as well as the operating stress and temperature. Operating times for space travel will extend into years; therefore, a reliable acceleration of the microcreep must be accomplished in order to obtain design information in a reasonable amount of time.

The objective of the investigation documented in Part V to survey methods of accelerating microcreep phenomena. Failure prevention can best be accomplished by the optimum selection and application of materials, based on known and anticipated requirements. The information gained from microcreep acceleration can have a threefold benefit to design, namely: the selection of best materials and processes in a reasonable amount of time, the reduction of residual stresses which are due to manufacturing processes, and the effective inspection of incoming materials.

Although the effect of microcreep is very hard to observe, its effect can usually be calculated for a given guidance system. Table V-1, taken from [V-1] depicts past, present, and future guidance requirements in terms of allowable velocity error at thrust termination, assuming the terminal thrust velocity is of the order of 10^4 ft/sec. Assuming that the present guidance capability for ICBM missiles is approximately ± 1 mile, the amount of material dimensional instability allowable in order to meet this requirement is given in Table V-2 for a few of the components. The ICBM missile was chosen because its velocity error requirements lie between those of two anticipated space missions (for which the guidance systems haven't been completely developed). The requirements shown are for a hypothetical system for which the magnitude of the error sources were taken from [V-1]. As illustrated in Table V-2, relative movements of a few microinches per inch in critical electromechanical components can result in instrument failure, i.e. failure to operate according to specification. Failure can be of a catastrophic type such as when air bearing surfaces scrape, or failure can be a more subtle type such as unacceptable mass unbalance which causes errors in the instrument performance.

Guidance instruments built for space travel must be able to operate reliably during the escape period from the earth's atmosphere (which will involve high acceleration), as well as during a cruise period in the space environment (which will be at very low stress levels but for long periods of time). Thus, the space instrument must be able to cope both with the problems encountered in ballistic systems and those encountered in cruise

TABLE V-1. RELATIVE REQUIREMENTS ON GUIDANCE SYSTEM ACCURACIES FOR MILITARY AND SPACE MISSIONS

Vehicle	Target Accuracy	Maximum Velocity Error at Thrust Termination
Scientific Satellite	±100 miles	80 ft/sec.
ICBM	±10 miles	4 ft/sec.
Moon Impact	±100 miles	1 ft/sec.
ICBM	±1 mile	0.5 ft/sec.
Mars Impact	±1000 miles	0.3 ft/sec.

TABLE V-2. GUIDANCE SYSTEM MATERIAL DIMENSIONAL STABILITY REQUIREMENTS FOR ICBM WITH ACCURACY OF ±1 MILE

Component	Type of Error	Allowable Material Movement
Gyro	Center of gravity drift	1 microinch per inch
Stable Platform	Alignment shift	5 microinches per inch
Gyro Air Bearings	Contact between parts rotating at high angular velocity	20 microinches
Computer Memories	Contact between parts rotating at high angular velocity	25 microinches
Accelerometer	Error torque	Sufficient for 0.1 dyne-cm torque

systems. More concrete examples of the two types of stability requirements are reflected in the Minuteman guidance system and in the atomic submarine inertial guidance system. For the Minuteman, fifteen microinches per inch is the maximum permissible total change between planes of the mounting flanges for gyros on the stable platform during the three month storage period between instrument calibrations (see [V-2]). If a portion of a critical gyro used in the atomic submarine guidance system creeps at a rate of 0.01 microinches per inch per day or more, resulting in an equivalent mass unbalance, then unacceptable performance will result (see [V-3]).

V.A.2 <u>Nature of Microcreep</u>

The dimensional instability previously referred to has two basic causes. On one hand, small dimensional changes characteristically occur when any material partially or completely undergoes a solid state phase change or other change of state. This movement is determined by the phases present and their degree of instability and varies as a function of time. The other basic cause of instability results from an applied stress, a residual stress, or a combination of both. It is the response of a material to a force imposed upon it and this reaction varies with the magnitude and direction of this load, the temperature, the kind of material, and the treatment to which it may have been subjected. Very often, these effects are also time dependent.

The instability due to phase changes in metals can usually be eliminated by using pure metals which have been annealed or otherwise treated to be as close to thermodynamic equilibrium as possible. However, pure metals so treated are invariably too weak for satisfactory design applications. Therefore the techniques of martensitic hardening, solid solution hardening, or precipitation hardening must be chosen in practice to provide the required strength to resist the second form of instability. Metallurgical stability has to be achieved by thermal treatments which are designed to complete, as much as possible, the phase changes or transformations, without appreciably affecting the elastic strength or inducing excessive residual stress. Such treatments have been determined for most materials used in critical guidance system components. However, an awareness of this type of instability is required to properly evaluate the second type of instability (due to stress) since it can readily be mistaken for stress induced instability.

The more universal cause of instability (and the one of interest in this study) is distortion caused by mechanical stresses. In guidance and control instruments, such applied stress is generally quite low, as it usually only arises from acceleration or centrifugal loads. Although these stress levels are low, a significant amount of material movement can occur over an extended period. Another stress causing instability is that known as residual stress. Residual stresses can be developed in many ways but generally fall into two categories. Those produced by fabrication and assembly of parts, and those resulting from temperature changes. Forming, machining, and joining are examples of fabrication techniques that can lead to assembly stresses. Quenching and incompatible rates of thermal expansion of adjacent material are examples of sources of stresses induced by temperature changes.

Engineering design practices for critical components limit the applied stresses to levels below which no gross plastic deformation occurs. This stress maximum is determined by a non-time-dependent parameter called the Precision Elastic Limit (PEL) which is defined as that stress which induces a permanent set of one microinch per inch. This amount of deformation is an arbitrary value compatible with both reliable measuring capability and guidance system requirements. It has been reported by a number of investigators that very little microcreep occurs at stresses below the PEL (see [V-4], [V-5]). It is not known whether this effect is real, or whether it merely reflects investigators' inability to measure the microcreep on the level at which it occurs. Most of the investigators have reported an accuracy of five microinches per inch over an extended period of time (up to one year). The National Bureau of Standards reports a measuring capability of 0.10 (μin/in) over a period of several years (see [V-6]). However, these measurements were made on gauge blocks without any applied loads. No measurable microcreep was detected by the Autonetics Metallurgical Laboratory in beryllium tests when loaded to the PEL stress (see [V-7]). Certainly, the amount of microcreep was not greater than five microinches per inch (which is the creep test accuracy). It should be noted that, at stresses around the PEL, short-term permanent set can be measured much more accurately than microcreep due to the inherent "noise" of strain gauges.

V.B MICROCREEP BEHAVIOR

Creep is a time-dependent deformation of a material. This time-dependent deformation may be permanent, or recoverable, or a combination of both. The recoverable portion of the deformation is called anelastic creep, and the permanent portion is called plastic creep. Both, however, are time dependent. Microcreep then is that type of creep which is induced by small strains arising from low stresses such as those encountered by space guidance components. Although anelastic creep is negligible in high stress studies, it cannot be overlooked when investigating microcreep. Different materials exhibit anelastic creep in varying degrees, and in some materials this type of creep is not evident at all. However, when present, it is usually initiated at stress levels below those necessary to induce plastic creep. The presence of anelastic creep can readily be detected by unloading the specimen and determining the time-dependent strain recovery. The operational effect of anelastic creep on guidance instruments, which contain components having constant operational stresses, is of course the same as the effect of plastic creep.

Both of the above types of creep are generally explained by the theory that any time dependent deformation (subsequently called flow) occurs principally by the motion of defects termed dislocations in the crystal structure of the material. Dislocations are generally categorized in two general types, referred to as edge dislocations and screw dislocations. The way in which these two types of dislocations manifest themselves, either independently or simultaneously, determines the mechanism by which flow takes place. The area of greatest contention concerns the nature of the dislocation mechanism which is dominant under a given condition of stress and temperature, and the equation describing the flow.

Conrad [V-8] has categorized, according to crystal structures, the types of dislocation flow he believes to be operating in certain temperature ranges. His beliefs are based on data generated by himself as well as data taken from the literature. By his own admission, the data are insufficient to fully justify the proposed flow mechanisms, but could be used as a guide. His proposed mechanisms for temperatures less than one fourth of the absolute melting temperature are given in Table V-3. These types of dislocation flow result in either the exhaustion of dislocation sources or the multiplication of dislocation sources (strain hardening). These processes are then manifested by anelastic and/or plastic creep. Where multiplication occurs, the additional dislocations start interfering with each other's motion and become locked into tangles which require additional energy to release before additional motion can occur. The two most common types of multiplication processes are referred to as the Frank-Read process and the Multiple Cross-Glide process. As more dislocations are generated and subsequently pinned by tangles as creep occurs, the resulting strain rate diminishes with time for a given stress and temperature. When creep occurs by an exhaustion of sources, there is no multiplication of dislocations. Since the potential sources of motion are exhausted during the creep, the strain rate for the exhaustion mechanism also decreases with time.

TABLE V-3. PROPOSED RATE-CONTROLLING FLOW MECHANISMS IN METALS

Crystal Structure	Metals	Most Likely Mechanism	Alternative Mechanism
Close Packed Hexagonal	Zn, Cd, Mg, Ti, Be	Intersection of dislocations	Non-conservative motion of jogs
Face Centered Cubic	Al, Cn, Ag, An, Ni	Intersection of dislocations	Conservative motion of glissile jogs
Body Centered Cubic	V, Nb, Ta, Cr, Mo, W, Fe	Overcoming the Perierls-Nabarro stress	Non-conservative motion of jogs. Overcoming interstitial precipitation cross-slip.

It appears that the mechanisms described by Conrad and the expressions derived for them cannot be adequately applied at this time to useful engineering materials. Most of the data generated to date using the dislocation model have been on single crystals or high purity polycrystalline materials, neither of which are used very extensively as basic engineering materials. The effects of alloying, as well as thermal and mechanical history, would have a complicating effect on the already involved relationships. It is concluded that it is not generally feasible to study the acceleration of microcreep by use of the existing purely theoretical exploitation of dislocation concepts.

According to the theory of dislocations, the anelastic phenomenon is dependent on the effects of heterogeneity of dislocation patterns and on the effects of the orientation of grain boundaries on dislocation flow strength. Slip occurs statistically in relatively few so-called "soft grains" of the surrounding elastic matrix. When the stress is removed, back stresses arising from the elastic strains in the matrix force the dislocations to return to their original positions. However this concept has not yet matured sufficiently to relate the time dependency of anelastic creep to other measurable parameters.

As stated before, plastic creep can be categorized into two separate types, exhaustion of dislocation sources and strain hardening. Most investigators who have studied creep at high stresses and large strains report that strain hardening predominates. Two investigators contacted ([V-9] and [V-10]) have indicated that at low temperatures and low stresses it is most likely that exhaustion mechanisms predominate.

Tinder [V-11] demonstrated in high purity polycrystalline copper, zinc, and aluminum that the strain hardening mechanism is not observed for plastic strain at low stress levels. He infers that a source exhaustion mechanism (he does not refer to it as such) is operating for microcreep in these materials at low stress levels. However, too little work has been done on microcreep in all materials of interest. If microcreep does in fact occur primarily by an exhaustion of sources, then a more straightforward extrapolation approach seems possible.

V.C METHODS OF LONG-TERM CREEP ESTIMATION

Numerous concepts have been proposed and utilized by various investigators for obtaining estimates of long-term creep behavior on the basis of short-term tests. One approach describes long-term behavior with curves that are fitted to short-term test data. Another approach for predicting creep behavior is based on the concept of trading temperature for time. One such method is predicated on the existence of a material constant called the activation energy. A third approach has recently been proposed which utilizes ultrasonic energy to accelerate the microcreep. The following subsections discuss these approaches.

V.C.1 Curve-Fitting

There exist various families of curves to fit creep data. For materials exhibiting anelastic creep behavior, Lubahn [V-12] has suggested the use of a spring-dashpot analog to fit the data from short-term tests. The long-term behavior is then inferred from the time response of the analog. In the low temperature cases of interest the total strain in tension (elastic plus anelastic) versus time can be approximately expressed by the equation

$$\epsilon + \epsilon_a = S[(1/E) + a(1-e^{-t/\tau})]$$

where

S is the normal stress

E is the modulus of elasticity

t is the time in tension

a and τ are empirical constants

and $\epsilon = S/E$ is the elastic part of the strain and $\epsilon_a = aS(1-e^{-t/\tau})$ is the anelastic part. The physical significance of a is that a is the amount of anelastic strain which would presumably occur in infinite time. The constant τ, known as the relaxation time, indicates the rapidity of approach to the asymptotic value of anelastic strain; after an amount of time τ, the anelastic strain (according to the model) reaches $1-e^{-1} \approx 63.2\%$ of its ultimate value. This description of creep behavior indicates that it is analogous to the motion of the load in the spring-dashpot assembly shown in Figure V-1.

A better fit of the anelastic creep curve can be obtained by considering several spring-dashpot assemblies in series, i.e.,

$$\epsilon_a = \sum_{i=1}^{n} a_i(1-e^{-t/\tau_i})$$

where i indexes the several spring-dashpot assemblies, n is the number of such assemblies in series, and τ_i is the time constant associated with the i-th spring-dashpot assembly. By convention the indexing is such that $\tau_1 \leq \tau_2 \leq \ldots \leq \tau_n$. In general, data can be fitted more and more accurately as one uses a larger number of terms. For a given set of data, as the number of terms is increased, the coefficient of each term decreases and the τ value of any term becomes less and less different from that in the adjacent term. The limiting case is represented by an infinite number of terms, each having an infinitesimal as its coefficient, and each having a τ value greater by an infinitesimal than the preceding term. Such an infinite series can be represented by a graph called a "relaxation

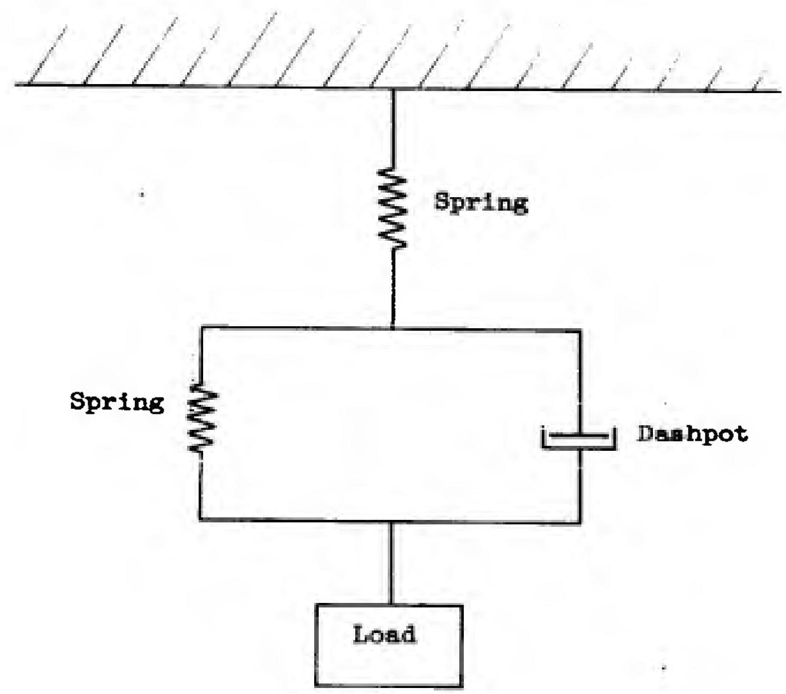


FIG. V-1. SPRING-DASHPOT ASSEMBLY

spectrum". The abscissa of the relaxation spectrum is $\log\tau$. The ordinate Ψ is such that the infinitesimal areas under the curve are equal to the coefficients a_i in the series expression for ε_a. That is, if the area under the relaxation spectrum curve is divided into infinitesimally narrow vertical rectangles, the area of each rectangle is the coefficient a of one of the terms of the infinite series, and the abscissa value at the center of the rectangle is the logarithm of the relaxation time for that term.

Zener [V-13] has shown that the relaxation spectrum can be approximated rather closely by plotting the anelastic creep versus log time, measuring the slopes of this curve and plotting these slopes against the values of log time. Either the anelastic creep at load or the recovery at no load can be plotted, as they have approximately the same time constant(s). The approximate relaxation spectrum can be further adjusted (by procedures presented by Lubahn [V-12]) until the combinations of spring-dashpot assemblies have an analogous strain-time relationship which fits the short-time test data.

Guard [V-14] recommends an empirical relationship for plastic creep which reflects the effects of both the dislocation source exhaustion mechanism and the strain hardening mechanism (source multiplication). The form of the relationship is:

$$\varepsilon = a \log t + bt^{\beta}$$

where ε is plastic strain

a, b, and β are fit constants for each stress and temperature of interest,

t is time in stress.

Obviously, log t approaches infinity as time increases without bound. Nonetheless, for the times of interest, this function has been observed to satisfactorily fit the creep test data. If the simple exhaustion mechanism is in effect (which is generally true for stresses near the PEL), b and β will be equal to zero and the plastic creep model will reduce to

$$\varepsilon = a \log t.$$

The constant a for a given temperature and stress can be easily determined from a plot of strain versus log time.

There are three major sources of extrapolation errors.

(i) The use of a functional relationship which is no longer dominant outside the time range in question.

(ii) The onset of new processes (aging, recovery, precipitation) which changes the constants.

(iii) The existence of history effects, i.e., an effect of past stress-temperature history on the current behavior.

Since it is not always possible to separate the components of the observed creep (anelastic and plastic), it is recommended that the more general model (the one used by Lubahn) be used and its fit evaluated. A check for the existence of large extrapolation errors due to the sources listed can be made by comparing results with those of other independent extrapolation methods, such as the activation energy method discussed in the next subsection.

V.C.2 Acceleration by Temperature

Reaction rate and diffusion principles suggest that the dependence of strain rate on temperature should be of the form

$$\frac{d}{dt}[\varepsilon(t)] \equiv \dot{\varepsilon} = ae^{-b/T}$$

where T is the absolute temperature, and a and b are appropriate parameters. Note that, in appearance, the above relationship suggests a constant strain rate and hence a linear increase of strain in time, which would contradict existing creep data. The paradox is however only apparent, because the parameter a in fact depends on the already realized strain. Dorn [V-15] and [V-16] has determined the parameter b for several pure polycrystalline materials. Rewriting b as $b=\phi/R$, where R is the universal gas constant, indicates an analogy between the parameter ϕ and the quantity referred to as activation energy in solid state diffusion and chemical reaction rate theories. For this obvious reason, Dorn and other workers refer to ϕ as the "apparent activation energy" (AAE). The parameter a, which Dorn refers to as the material "frequency factor", includes terms describing crystal lattice properties and movements. In the present state of knowledge of this subject, the frequency factor and activation energies must be determined experimentally.

In the case of several pure polycrystalline materials, the "apparent activation energy" (characteristic of each material) has been shown to be relatively insensitive to applied stress, strain, and strain rate. On the other hand the AAE was experimentally shown to be uniquely dependent on temperature as shown in Figure V-2. Each energy plateau is associated with a unique creep rate controlling dislocation mechanism. The AAE is determined experimentally by abruptly changing the temperature by a slight amount. A corresponding change in strain rate is observed. From the ratio of strain rates just before and after the temperature change, the AAE can be calculated.

To determine that the AAE is constant in the temperature range of interest for a constant stress, other temperature changes may be used to determine corresponding AAE's. If they are constant, as has been shown to be for aluminum around room temperature (0 to 75 degrees centigrade) [V-15] , then this means only one mechanism for flow predominates. Dorn [V-15] suggests that, under appropriate assumptions, a time-temperature tradeoff can be formalized by reducing all data to a common "equivalent time" scale specified by

$$\theta = te^{-\phi/RT}$$

where θ is the equivalent time and t is the real test time. Dorn's empirical results with aluminum [V-15] indicate excellent correlation of creep strains by this time transformation of creep strain-time data at various temperatures. The amount of acceleration for a given temperature change depends on the magnitude of the AAE. It should be noted again that this approach is applicable only if the AAE is the same for both use conditions and accelerated conditions.

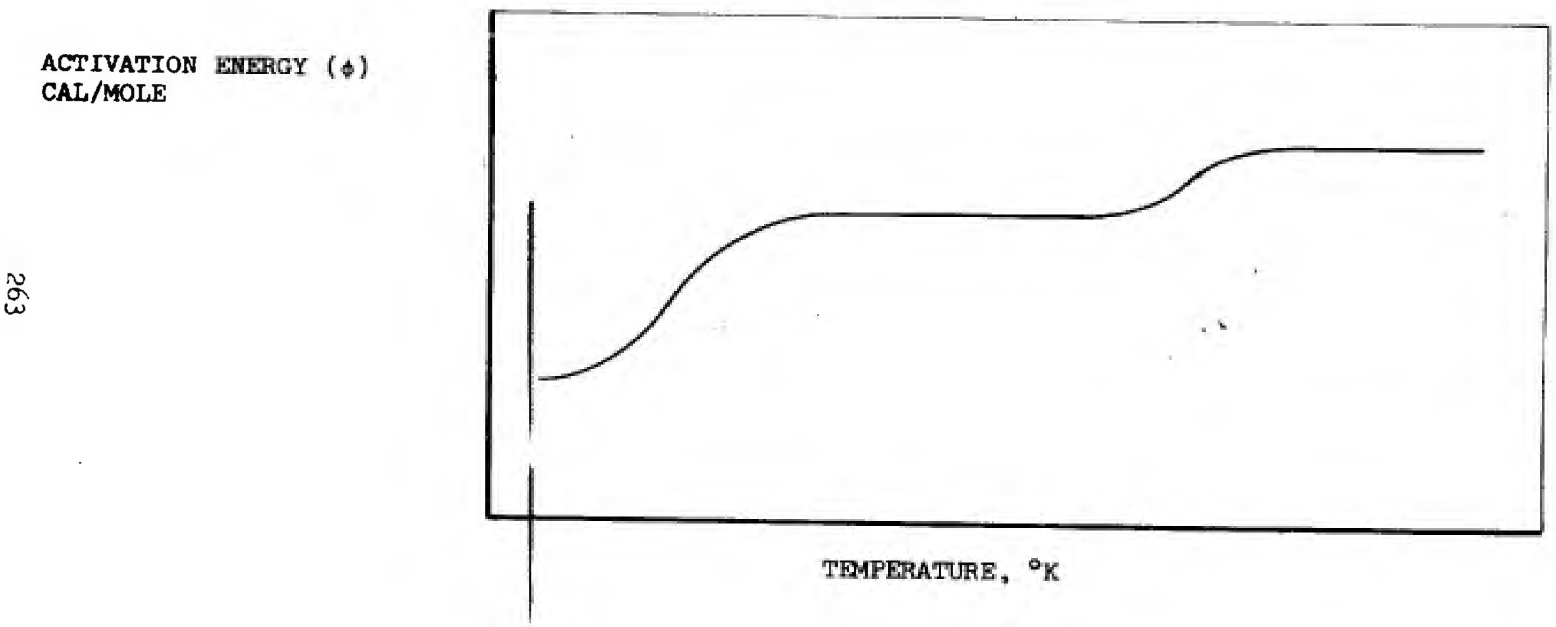


FIG. V-2. ACTIVATION ENERGY VS. TEMPERATURE

V.C.3 Acceleration by Ultrasonic Energy

V.C.3.1 Effects on Material Deformation

Intense ultrasonic energy has been shown to have very marked effects on the plastic deformation of metals. Langenecker [V-17] has demonstrated that the yield strength of a material can be reduced to zero by high intensity ultrasound (up to 100 W/cms). Strain vibrations are induced into the material by the ultrasonic transducer. A stress, referred to as an acoustical stress, develops from the ultrasound which can be readily calculated. The acoustical stress required to reduce the yield strength of some materials to zero is from one to two orders of magnitude lower than the yield stress per se. The reason for this has been attributed to the concentration of the acoustical stress waves at dislocation points. Langenecker feels that a perfect crystal (i.e., a crystal that is free of dislocations) would probably transmit ultrasonic stress waves with no attenuation. The softening effect (lowering of the yield strength) during ultrasonic vibration does not appear to have any affect on the mechanical behavior of the material after vibration up to a critical ultrasonic energy level (see [V-18]). Above this level a work hardening of the material becomes apparent: the yield strength increases after a specimen is subjected to ultrasonic energy above the critical value. Further increasing of ultrasonic energy induces failure in the polycrystalline materials studied. The fact that softening occurs at accoustical stress levels below the work hardening range supports the contention that microcreep (from low stresses in the vicinity of the PEL) occurs by an exhaustion of sources.

The benefits derived from work softening due to ultrasonic vibration have been utilized in extruding and forming of materials as well as ultrasonic machining (see [V-19]). The relief of residual stresses due to lowering the yield strength is presently under investigation by the Naval Ordinance Test Station, China Lake, California, and by Autonetics.

The benefits from higher amplitude ultrasonic vibration have been utilized both as a material strengthening mechanism [V-20] as well as for limited accelerated fatigue testing [V-21]. For fatigue testing, the specimens are vibrated at their resonance frequency resulting in large strain amplitudes which are cyclical. The use of high frequency ultrasonics looks very promising for acceleration of microcreep.

The previous discussion illustrates the effects of ultrasonics on other types of processes involving plastic deformation. Although no work to date has been reported on the effects of ultrasonics on creep, it appears evident that an acceleration of creep can be accomplished using ultrasound. The approach and anticipated problems will be discussed in the next subsection.

V.C.3.2 Ultrasonic Acceleration of Microcreep

Acceleration of microcreep at the stress levels and temperatures of interest appears very reasonable by the addition of ultrasonics to the loaded specimens. The frequency of the strain waves and their magnitude can be controlled by the selection of the transducer (including horn) and the applied voltage. Details of the theory of ultrasonic vibration transmission are given in [V-22].

The major problem anticipated in using this technique is the determination of the acceleration factor. The acceleration factor could possibly be arrived at by comparing the amount of microcreep that takes place with and without the ultrasonic waves during a short-term test. A strain amplitude and frequency would have to be determined that would not induce spurious acceleration. An indication of spurious acceleration could be obtained both by the comparison of short-time strain curves with and without ultrasonic vibration and by a comparison with results attained by either or both of the extrapolation techniques discussed in Subsections V.C.1 and V.C.2.

Uniform and controlled transmission of the ultrasonic vibration from the source (transducer) to the test specimen may create a problem; however, two methods appear feasible. A mechanical attachment (adhesively bonded or threaded) of the transducer to the specimen is the method predominantly used at the Naval Ordinance Test Station. It has been shown [V-20] that undesirable standing waves can be created if the attachment is not correct. The other approach is to rely on a source of sufficiently high energy, in order to make a liquid transmission medium usable. The energy lost by cavitation at the surface of the specimen in the liquid will reduce the high intensity ultrasonic energy to the level desired.

Because this approach is new, difficulties can be expected; however, its potential for microcreep acceleration, if proven valid, is great. Hopefully, it can be applied to the acceleration of microcreep in a manner that will allow the selection of the best materials and processes for a given application. It is also possible that ultrasonic energy can relieve part of or all the residual stresses in manufactured components.

V.C.4 Statistical Considerations

If several presumably identical specimens are subjected to a creep test, they will not all behave identically. In view of such inevitable randomness, it is clear that creep data should be handled statistically as opposed to deterministically. There are three alternative approaches to the statistical treatment of creep data. These are discussed below.

One approach is to specify a load and a time interval, and examine the randomness in the resulting amount of creep. The potential value of measuring this randomness is exemplified by an analysis of the differential creep experienced in vibrating string instruments. In one type of vibrating string accelerometer, there are two vibrating ribbons in tension. Because the ribbons are vibrating in perpendicular planes, each ribbon has its own proof mass and the masses are isolated from each other by a spring. An acceleration along the axis of the ribbons causes an opposite change in tension for the two ribbons. The change in the difference of vibrating frequencies for the two ribbons is a measure of the acceleration.

It is conjectured that, if the ribbon tension is great enough to cause some creep, then the amount of creep in the two ribbons will not be identical, and the difference frequency will change, thus indicating a false acceleration (zero-shift). Instead of attempting to estimate the probability of occurrence of excessive zero-shift by testing the whole instrument, one can derive this probability from the statistical properties of the absolute magnitude of creep strain of single ribbons. The amount of creep of each ribbon, in specified time and under a given tension is a random variable, say X; let the distribution function be F with density f. Then $Z \equiv |X_1-X_2|$ is the difference in the amount of creep of two specimens. Denoting the distribution function of Z by G, it can be readily shown (see e.g. Hald [V-23], p. 320) that

$$G(z) = 2 \int_{-\infty}^{\infty} [F(x+z)-F(x)]f(x)dx \quad .$$

A simple integration by parts shows that

$$\int_{-\infty}^{\infty} F(x)f(x)dx = 1/2$$

and, hence,

$$G(z) = 2 \left[\int_{-\infty}^{\infty} F(x+z)f(x)dx \right] -1 \quad .$$

(Since creep is a non-negative random variable, the lower limit of the integration should be taken as zero.) If the distribution function F were known, one could calculate G, and then decide whether the "zero-shift" could indeed be attributed to random differences in creep of the two specimens.

In the second approach, a certain amount of creep is specified; one then can examine the length of time required to reach that amount of creep under a specified load.

In the third approach one examines the load which will cause a specified amount of creep in a specified length of time. Since creep tests must be conducted at constant loads, this examination method must include a "trial and error" process. The method known as stochastic approximation can be used to ascertain the value of the load which will result in the specified creep with a given probability. The application of that technique to creep constitutes the second example at the end of Section VI.C.

VI. SPECIAL TOPICS

VI.A STATISTICAL CONSIDERATION OF MULTIPLE FAILURE MODES

VI.A.1 Introduction

Analysis of the reliability of complex devices (such as inertial navigation instruments) as opposed to that of simple components, requires appropriate statistical methods for handling cases with multiple failure modes. As in all statistical analysis, one must start with an appropriate model for the physical system from which the data to be analyzed arise. This model is based on past knowledge, being devised either on the basis of physical (engineering) considerations or on the basis of data from a pilot experiment.

Two models are particularly worthy of consideration as representative of the failure behavior of devices with multiple failure modes. One model corresponds to the situation in which several somewhat different classes of devices are randomly mixed to form the population of devices under consideration; this model will be referred to as the mixed populations (MP) model. The other model corresponds to the situation in which several potential causes of failure exist simultaneously, and failure actually occurs when one of those causes materializes; this model will be referred to as the competing risks (CR) model. The two models are further discussed and illustrated in Subsection VI.A.2.

To apply a statistical model to a given situation one should first understand the consequences of the model's assumptions. Therefore, both models are studied extensively in Subsections VI.A.2 and VI.A.3 with respect to some general consequences of the assumptions. Questions of inference are considered in Subsection VI.A.4. The pertinence of this investigation to accelerated life testing is discussed in Subsection VI.A.4.1.

VI.A.2 Description of Two Models

For the mixed populations (MP) model, assume that a collection of nominally identical devices can be segregated into a finite number of classes on the basis of some criterion pertinent to the failure behavior of the devices. For instance, let the devices result from several different production processes. It is plausible to assume that the distribution of time-to-failure may be different for devices resulting from different processes; the process from which a device results is then a natural criterion of classification. As another example, assume that a collection of nominally identical devices exhibits several modes of failure, but that any particular device has an inherent propensity for one mode of failure which precludes the possibility of failure in any other mode; the inherent mode of failure, though it may not be known a priori, is a natural criterion for the classification. An example of such a separation of failure modes is the following; a crack in the casing of an automobile tire, at the time of manufacture, will lead to an early blowout and will preclude the possibility of tire failure due to wear past a specified limit.

The MP model is used to analyze situations which fit the above description. The principal characteristic of the model is that each unit, e.g., device, of the population conceptually belongs to one, and only one, subpopulation (this subpopulation may not be known to the experimenter but should be determinable in principle). This property of the model is known as mutual exclusiveness of the subpopulations. The second important property is the predetermination of the subpopulation to which a device belongs, at least in principle. These two properties are inherent in the MP model. When the classification into subpopulations is done on the basis of some characteristic such as the manufacturing process (or supplier) from which a device originates, it is clear that the subpopulation to which a device belongs is determined at the outset of any experiment. When the classification into subpopulations is on the basis of mode of failure, the MP model applies only if the ultimate mode of failure for any particular device is predestined.

It may not be plausible to assume that the mode of failure of a device is predestined. Assume that each mode of failure is due to a different physical process, and that the processes operate simultaneously; actual failure occurs the first time one of these processes reaches its goal. If the different processes are assumed to operate independently, there results the model which is referred to as the competing risks (CR) model. For example, consider again an automobile tire; a flat tire can be caused by either a nail or a piece of glass, whichever is encountered first. As another example, a capacitor can fail due to a breakdown in the dielectric or due to the breaking of one of the leads, these two risks competing until one of them prevails. The competing risks model occurs frequently in biostatistics because it applies to the mortality of living organisms due to various diseases, since many diseases are potential risks until one of them finally takes its toll by causing death.

The two models can conveniently be contrasted by means of the following artificial example: Let there be a horse race among six horses; consider a failure to have occurred when the winning horse arrives at the finish line, and let the number of that horse be the failure mode. If the race is run in customary fashion, with all horses starting simultaneously, the failure process fits the competing risks model. Each horse is one risk; in any one trial, all risks are present, and even the rarest risk, i.e., slowest horse, has a chance of occurring. On the other hand, consider the following unorthodox manner of racing. First a die is cast to determine the winning horse; then the winning horse runs alone, to determine the time it requires to reach the finish line; now the failure process fits the mixed populations model. The many possible running times of the i-th horse form the i-th subpopulation, and the six subpopulations are mixed in a ratio determined by the probabilities of the various faces on the die to form the resulting population of all possible times-to-failure.

In a real situation, it is perfectly possible that both models may have to be invoked for a complete analysis. A sample of devices may consist of units from two different manufacturers, which fact suggests the MP model. However, for any particular device, there are several possible modes of failure whose occurrence may well be in accord with the CR model. In this study, the two models are treated separately (though in parallel). Their superposition, as suggested in this paragraph, is best studied in relation to specific applications.

As stated earlier in this subsection, it is imperative to be able to assume the existence of separate and mutually exclusive subpopulations among the items to which the MP model is to apply. A given item must arise from only one distinct process or be subject to only predetermined cause of failure for the model to apply. In practice this assumption seems highly implausible, expecially for complex devices. In the case of different suppliers or production processes for parts of a complex device, the physical situation might meet the assumptions of an MP model, which could then be used in conjunction with a CR model for the failure possibilities of the device.

In addition, it turns out that the two models cannot be statistically distinguished although they are quite different in their assumptions. In the MP model, for an arbitrarily chosen device, the times-to-failure (TTFs) from two or more causes are dependent upon each other: if the device fails at some time from one cause, then it is known that it could not have failed from any other cause. This dependency does not hold in the competing risks model, and it then follows from Cox [VI-1] that it is impossible to decide, from data on failure times and type of failure only, which of the two models applies. Therefore, any decision on choice of model must depend on physical considerations. A choice must be made, since inference procedures (and their results) differ significantly for the two models.

The two models are quite different but are frequently confused. For example, in a recent article by Gottfried and Roberts [VI-2] it is stated:

> "In semiconductor device testing, a variety of failure modes are commonly observed. For example, a single life test may produce emitter-open and emitter-collector short failures ... Often these are the results of different failure mechanisms. Whenever multiple failure mechanisms and/or failure modes are observed, one may suspect that a heterogeneous population exists."

The authors thus indicate that they find the mixed populations model appropriate. But, since there is no indication that a given semiconductor is predestined to fail in a particular mode, it would seem more nearly correct to interpret the authors' life test in terms of the competing risks model. Subsequently, the authors introduce a specific example in which they treat data pertaining to two modes of failure as coming from two distinct subpopulations. The data which they present does not indicate that a device is predestined to one mode of failure in a manner which precludes the possible occurrence of a failure in the other mode. Again, the competing risks model would seem more appropriate.

VI.A.3 Consequences of Model Definitions

VI.A.3.1 Population and Underlying Distributions

When studying devices whose time-to-failure can be described by one of the models discussed in Subsection VI.A.2, one may choose to consider the distribution of times-to-failure, without regard to mode of failure, the nature of the risk causing failure, or the subpopulation to which a failed item belongs. In this report, that distribution is referred to as the population failure distribution. Throughout the subsequent discussion, let the non-negative random variable X denote the time-to-failure of a randomly selected device; let F be the distribution function of the random variable X, i.e., denote by F the population failure distribution,

$$F(x) = P(X \leq x) \quad . \qquad \text{(VI.A.3.1-1)}$$

Let the random variable N (with values conveniently chosen in some subset of the positive integers) denote the mode of failure of a randomly selected device. In the case of the MP model, N denotes the subpopulation to which a particular randomly selected device belongs; in the case of the CR model, N denotes the risk which, by prevailing over all other concurrent risks, brings about the failure of a particular randomly selected device. Let the set $\{p_i\}$ be the probability distribution of the random variable N, i.e., denote by p_i the probability that a randomly selected device will fail in the i-th mode,

$$p_i \equiv P(N=i) \quad . \qquad \text{(VI.A.3.1-2)}$$

Refer to p_i as the overall probability of failure in the i-th mode. Now the population failure distribution for each model is derived in terms of some other functions which arise naturally in the course of formal definition of the models.

Consider the mixed population model first. Let F_i be the distribution function of the time-to-failure random variable for the devices in subpopulation i. The random variable N defined above now denotes the subpopulation to which a particular device belongs. In view of the definition given above for the random variable X, the distribution function F_i is clearly the conditional distribution function defined by

$$F_i(x) \equiv P(X \leq x \mid N = i) \quad . \qquad \text{(VI.A.3.1-3)}$$

Let the production (or supply) process be such that the class of a given device is stochastically independent of the class of previously produced devices, i.e., let the nature of the defect be assigned independently to each device. Then the i-th subpopulation forms a proportion p_i of the total population of devices under consideration. The population failure distribution F, i.e., the distribution of the random variable X which is the time-to-failure of a randomly selected device, is given by

$$\text{(VI.A.3.1-4)}$$

$$\begin{aligned} F(x) = P(X \leq x) &= \sum_{i=1}^{n} P(X \leq x \text{ and } N = i) \\ &= \sum_{i=1}^{n} P(N = i) P(X \leq x \mid N = i) \\ &= \sum_{i=1}^{n} P(N = i) F_i(x) \\ &= \sum_{i=1}^{n} p_i \, F_i(x) \quad , \end{aligned}$$

where n is the number of distinct subpopulations.

The competing risk (CR) model is conveniently formalized in an article by Berman [VI-3] and is restated here. Consider the following hypothetical experiment: there exist n random variables $X_i (i = 1,2,\ldots,n)$ which are continuous and can assume any non-negative value. These variables are distributed independently; let F_i denote the distribution function of X_i. In theory one observes the values of all X_i and then defines the random variable X as the minimum over i of all X_i values, i.e., $X = \min(X_1, X_2,\ldots,X_n)$. Also specified is the random variable N which takes on the value i whenever $X = X_i$. Now, assume that the information available to the observer is restricted to observations on the values of X and N; if $X = X_i$ and $N = i$, the values of X_j are unknown to the experimenter for $j \neq i$. This hypothetical experiment describes the observations of the times-to-failure of devices under the following definition: think of X_i as the TTF from the i-th cause if this cause (or kind of failure) were the only possible one. N takes on values which label the cause producing an actual failure at some time. Any given device can potentially fail from any one of the n causes, which operate independently, but all that is observed is the time X of the failure realized first and the cause i responsible for this failure. For example, let X_2 be the time to accidental death and X_1 the time to non-accidental death. If an individual has a fatal accident before he dies of natural causes $(x_2 < x_1)$, the time to death X is the value of X_2; the label for the cause of death N is 2. The value of X_1 is unknown to the actuary. One cannot state by how much X_1 is greater than X_2 because it is the time to which the individual might have lived if there had been no fatal accident.

Since the different processes producing a failure are stochastically independent in the CR model it follows that F, the population failure distribution, is given by

$$F(X) \equiv P(X \leq x) = 1 - P(X_1 > x \text{ and } X_2 > x \text{ and } \ldots X_n > x)$$

$$= 1 - \prod_{i=1}^{n} [1 - F_i(x)] \quad . \qquad \text{(VI.A.3.1-5)}$$

For either of the models under consideration, there exists for each mode (i.e., risk or subpopulation) a function which can intuitively be considered as the "distribution of time-to-failure in the i-th mode" or, briefly, the underlying distribution for failure in the i-th mode. The underlying distribution for the i-th mode is the distribution function which would result if all failure modes, other than the i-th one, were impossible. It is the failure distribution which would be observed if a cure were available for all modes of failure other than the i-th one. In the MP model, that is tantamount to considering the entire population as consisting only of the i-th subpopulation. In the CR model, the "availability of a cure" is tantamount to considering the risk of failure from cause i, identified with the random variable X_i, as the only risk. It is obvious from the discussions preceding Equations (VI.A.3.1-3) and (VI.A.3.1-5) that, in either model, F_i is the underlying distribution for failure in the i-th mode.

A word of caution is in order. Equation (VI.A.3.1-3) is the defining equation of the underlying distribution for the MP model, but not for the CR model. In the later, $F_i(x) \not\equiv P(X \leq x \mid N = i)$, as can be seen from Equations (VI.A.3.2-4) and (VI.A.3.3-3) which are derived subsequently. The notion of underlying distribution is an intuitively appealing one, but it must be handled cautiously; the relationship between the population failure distribution and the underlying distributions is strikingly different in the two models considered. The relations between F_i and F are developed thoroughly in Subsections VI.A.3.2 and VI.A.3.3 .

VI.A.3.2 Joint Distribution

Let H be the joint distribution of the time-to-failure and mode of failure. Denote by $H_i(x)$ the joint probability that failure occurs no later than x and it is in the i-th mode, namely

$$H_i(x) \equiv P(X \leq x \text{ and } N = i) \quad ;$$

it follows that

$$H(x,m) \equiv P(X \leq x \text{ and } N \leq m) = \sum_{i=1}^{m} H_i(x) \qquad \text{(VI.A.3.2-1)}$$

for m = 1, 2, ..., n.

In either of the two models considered, the various modes of failure are mutually exclusive. It follows that the population failure distribution F is related to the joint distribution H (expressed in terms of the "components" H_i) as follows:

$$F(X) \equiv P(X \leq x) = \sum_{i=1}^{n} P(X \leq x \text{ and } N = i)$$

$$= \sum_{i=1}^{n} H_i(x) \quad , \tag{VI.A.3.2-2}$$

where n is the number of competing risks or the number of subpopulations.

The relationship between the population failure distribution F and the underlying distributions F_i was derived in Subsection VI.A.3.1 . Now consider the relationship between the joint distribution H and the underlying distributions. For the MP model, the relationship is obvious in view of the definitions:

$$H_i(x) = P(X \leq x \text{ and } N = i) = P(N = i)P(X \leq x \mid N = i)$$

$$= p_i F_i(x) \tag{VI.A.3.2-3}$$

For the CR model, the relationship is less obvious. In a recent note, Berman [VI-3] showed that

$$H_i(x) = \int_0^x \prod_{j \neq i} [1 - F_j(t)] \, dF_i(t) \tag{VI.A.3.2-4}$$

and

$$F_i(x) = 1 - \exp \left\{ -\int_0^x [1 - \sum_{j=1}^{n} H_j(t)]^{-1} dH_i(t) \right\} . \tag{VI.A.3.2-5}$$

The joint distribution H is a bivariate distribution; hence, it has two marginal distributions. One of them is the population failure distribution F. The other is the overall distribution of failures by mode, namely the set $\{p_i\}$ introduced in Subsection VI.A.3.1. Clearly, $p_i = H_i(\infty)$. For either model, p_i is the ultimate proportion of devices failing in the i-th mode, regardless of time of failure. For the MP model alone p_i has the obvious interpretation given in Subsection VI.A.3.1 : the i-th subpopulation forms a proportion p_i of the total population of devices under consideration.

VI.A.3.3 Conditional Distributions

The relation between failure time and mode of failure is illustrated by the underlying distributions F_i and by the joint distribution H, i.e., by the functions H_i. Further light can be shed on that relation by examining two conditional distributions. The first is the distribution of failures by type, given that they have occurred no later than time x, namely the set $\{p_i(x)\}$ specified by

$$p_i(x) \equiv P(N = i \mid X \le x) \quad . \qquad \text{(VI.A.3.3-1)}$$

The second is the distribution of failure times, given that they are in a specified mode, which will be labelled G_i and is specified by

$$G_i(x) \equiv P(x \le x \mid N = i) \quad . \qquad \text{(VI.A.3.3-2)}$$

For both models,

$$G_i(x) = \frac{P(X \le x \text{ and } N = i)}{P(N = i)} = \frac{H_i(x)}{p_i} \quad . \qquad \text{(VI.A.3.3-3)}$$

Whenever P(N = i) is not zero, it is obvious from the definitions (VI.A.3.1-2) and (VI.A.3.3-2) that, for both models,

$$F(x) = \sum_{i=1}^{n} p_i G_i(x) \qquad . \qquad \text{(VI.A.3.3-4)}$$

For the MP model Eq. (VI.A.3.3-3) reduces, in view of Eq. (VI.A.3.3-4), to

$$G_i(x) = F_i(x) \qquad . \qquad \text{(VI.A.3.3-5)}$$

Of course, this result also follows directly from the definition of G_i and the definition (VI.A.3.1-3) of the underlying distribution F_i for the MP model. However, no similar simplification is possible for the CR model. Thus, for the MP model only, the conditional distribution of failure times for a specified mode of failure coincides with the underlying distribution for that mode.

For both models, it follows from definitions (VI.A.3.3-1) and (VI.A.3.2-1) that

$$p_i(x) = \frac{P(X \leq x \text{ and } N = i)}{P(X \leq x)} = \frac{H_i(x)}{F(x)} \qquad . \qquad \text{(VI.A.3.3-6)}$$

Note that $p_i(x)$ can be interpreted as follows. If a random sample of devices are observed from time zero to time x, let ξ_i be the fraction of observed failures which were in the i-th mode; then $p_i(x)$ is the expectation of the random variable ξ_i.

An interesting question arises in connection with the conditional probability $p_i(x)$ [for which the symbolism $p_x(i)$ would perhaps be more meaningful]. When is $p_i(x)$ constant for all x ? That is, under what conditions is the expectation of the fraction of failures in a specified mode constant for any length of observation?

For the CR model, the above question has been studied by Allen [VI-4] in the case where the underlying distributions F_i are differentiable and, hence, have corresponding failure rate (or hazard, or force of mortality) functions. Let ϕ_i be the failure rate function corresponding to the underlying distribution F_i, i.e., $\phi_i(x) = f_i(x)/[1 - F_i(x)]$ where $f_i(x) = \frac{d}{dx}F_i(x)$.

Allen's result is the following:

> A necessary and sufficient condition that $p_i(x)$ be a constant, for all x, is that the ratio
>
> $$\frac{\phi_i(x)}{\sum_{j=1}^{n} \phi_j(x)} \qquad \text{(VI.A.3.3-7)}$$
>
> be constant, for all x. In that case, the constant value of the ratio equals the constant value of $p_i(x)$ which is just the marginal (i.e., overall) probability of failure in the i-th mode, namely p_i.

The ratio (VI.A.3.3-7) will be called the hazard ratio; and when it is constant for the i-th risk, it will be said that a constant hazard ratio (CHR) exists for risk i . (Another necessary and sufficient condition for the constancy of $p_i(x)$ is discussed in Subsection VI.A.4.2, in connection with a discussion of estimation techniques.)

In the case of only two competing risks, constancy of the hazard ratio implies that the conditional distribution of failure modes $\{p_i(x)\}$ is the same for all times x. (Note that the set $\{p_i(x)\}$ specifies the distribution of probability over the various modes for those failures which occur prior to time x.) However, when more than two failure risks are involved, the hazard ratio can be constant for one risk without implying the constancy of the conditional distribution $\{p_i(x)\}$. For example, let there be three failure modes, such that ϕ_1 is an increasing function and ϕ_2 a decreasing one, but such that they have a constant sum; then, if ϕ_3 is a constant function, the ratio (VI.A.3.3-7) remains constant only for i = 3 and $p_i(x) \neq p_i$ when i = 1 or i = 2.

Allen further points out that

$$1 - F_i(x) = \prod_{j=1}^{n} [1 - F_j(x)]^{p_i} \quad \text{, for all x,} \qquad \text{(VI.A.3.3-8)}$$

is equivalent to

$$\frac{\phi_i(x)}{\sum_{j=1}^{n} \phi_j(x)} = p_i \quad \text{, for all x.} \qquad \text{(VI.A.3.3-9)}$$

Therefore, by Eq. (VI.A.3.1-5), Eq. (VI.A.3.3-8) is equivalent to

$$F(x) = 1 - [1 - F_i(x)]^{1/p_i} \quad \text{, for all x.} \qquad \text{(VI.A.3.3-10)}$$

Thus, any failure risk i for which a constant hazard ratio (CHR) exists has the special property that its underlying distribution uniquely specifies the population failure distribution, without explicit consideration of the other underlying distributions.

It follows from Eq. (VI.A.3.3-10) that, when a CHR exists for the i-th failure risk, then

$$p_i = \frac{\log[1 - F_i(x)]}{\log[1 - F(x)]} \qquad \text{(VI.A.3.3-11)}$$

In the case of the mixed populations model, the question of constancy of the conditional probability $p_i(x)$ has a very different answer. Recall that if $p_i(x)$ is to be constant for all x, it equals the overall probability of failure in the i-th mode, namely p_i. Then, using Equations (VI.A.3.3-3) through (VI.A.3.3-6), $p_i(x)$ is constant for all x if and only if,

$$p_i(x) = \frac{p_i F_i(x)}{\sum_{j-1}^{n} p_j F_j(x)} = p_i \quad . \qquad \text{(VI.A.3.3-12)}$$

This equation is satisfied if and only if $F_i(x) = F(x)$ for all x. That is, the expected fraction of failures in a specified mode is constant for any length of observation if and only if the failure distribution for that subpopulation is identical with the population failure distribution.

Consequently, the conditional distribution of failure modes $\{p_i(x)\}$ is the same for all times x if and only if each subpopulation follows the same underlying distribution.

This subsection is closed by recalling that the underlying distribution F_i results from restricting the population failure distribution by the assumption that, whenever a failure occurs, it is certain to be in the i-th mode, i.e.,

$$p_j(x) = P(N = j \mid X \le x) = \begin{cases} 0 \text{ for } j \ne i \\ 1 \text{ for } j = i \end{cases}.$$

For the MP model, substituting this value of p_i in Eq. (VI.A.3.3-4), and using Eq. (VI.A.3.3-5), yields $F(x)= F_i(x)$. For the CR model, substituting this value of $p_i(x)$ into Eq. (VI.A.3.3-6) yields $F(x) = H_i(x)$. The "availability of a cure" (as used in defining underlying distributions) is tantamount to saying that $X_j \le x$ is an impossible event in the CR model for $j \ne i$ and any finite x. Therefore $F_j(x) = 0$ for $j \ne i$, and Eq. (VI.A.3.2-4) reduces to

$$H_i(x) = \int_0^x dF_i(t) = F_i(x)$$

and hence $F(x) = F_i(x)$.

VI.A.4 <u>Estimation</u>

VI.A.4.1 <u>Need for Separation of Modes</u>

This subsection deals with the need for separate estimation of the underlying distributions F_i in the CR model. Since such separate estimation is greatly facilitated when the CHR condition holds for some risk, the implications of that condition are discussed.

The need for separate estimation of the F_i's arises when one wishes to get estimates of various characteristics of the population being tested and then predict what these characteristic will be in another situation, i.e., one wants to extrapolate from the present environment to another environment. The problems associated with extrapolation can best be illustrated by considering an insurance company which is concerned about the effect of a possible cancer cure on the life-distribution of humans. For such an analysis, they might consider only two risks; deaths from cancer and deaths from other causes. They need to predict, from current data, what the life-distribution will be in the future, after such a cure becomes available. Let F_1 and F_2 denote the underlying distributions for deaths from cancer and deaths from other causes, respectively, under present conditions; denote by F_1^* and F_2^* the distribution functions for the same risks under future conditions; i.e., the superscript "*" refers to future conditions.

The distribution of deaths in the future, using a CR model, will be F^*, such that $F^*(x) = 1 - [1-F_1^*(x)][1 - F_2^*(x)]$. In order to predict F^*, they need to predict F_1^* and F_2^* from present data. To do this they first might assume that a cure for cancer will eliminate any chance of death from this cause, i.e., $F_1^*(x) = 0$ for all finite x. For purposes of this example assume that the underlying distribution for other causes of death will not change because of a cancer cure, i.e., $F_2^* = F_2$. Given the foregoing assumptions about the relationship between the distribution functions under the two environments, the insurance company can estimate F_2 from the present data, which is broken down into types of death. Without such a breakdown of the available data into deaths from cancer and deaths from other causes, there certainly is no way of estimating F_2. Mathematical expressions are needed that either relate the distribution functions to aspects of the pertinent environment or, equivalently, transform a distribution function in one environment to the distribution function in another environment. For the insurance company example, there exist the following transformations on the distribution functions F_1 and F_2 to the future functions: $F_1^*(x) = 0$ for all finite x, and $F_2^* = F_2$, regardless of the nature of F_1 and F_2. These transformations are obtained by making reasonable assumptions. Without such assumptions there is no way at all to make predictions for any environment other than the one from which the present data come.

The above actuarial example can be translated, almost word for word, into life testing terms. The present environment is the life test; from the results of the life test one wants to predict the times-to-failure, or alternatively life expectancies, in the use environment, the one in which the type of items being tested will actually operate. Before securing estimates and predicting, it is best to have expressions that relate the underlying distribution functions to the relevant aspects of the two environments.

For non-accelerated life testing, the assumption is that the testing and use environments are the same; therefore, the distribution function for any type of failure remains the same. Estimates from the test results are the best predictors; and, unless there is a different cost attached to different types of failures, there is no necessity for separating the failures into the different types. However, for accelerated life testing, the ability to extrapolate from testing conditions to use conditions dictates a separation of the different types of failures. In the variety of ALT called over-stress testing (OST) the devices are run under stress conditions unlikely to be present in use, so the distribution functions under test differ from those in the use environment. For example, consider an OST situation in which voltages in excess of those normally used are applied to a capacitor to get the failure-time distribution for the risk of a dielectric breakdown. Since the test pertains only to failures from one risk, the distribution function for other risks, say mechanical failures, can be assumed to be the same in the accelerated test and normal use environments. Therefore, if the underlying distribution functions under this OST are F_1 (for dielectric breakdowns) and F_2 (for mechanical breakdown), in the future use environment they will be respectively $F_1^*(\neq F_1)$ and $F_2^*(=F_2)$.

For the time-compression (TC) variety of ALT, F_1^* is assumed to equal F_1 since the compression of time is not supposed to affect the underlying distribution function of the type of failure under test. One still needs to use the differences among failure risks since the other risks, which can be collectively labelled risk 2, can account for some of the failures in the TC test. Furthermore, these other failure risks may be distributed differently in this test than in other environments, i.e., $F_2^* \neq F_2$.

By distinguishing among the failure types, one can thus extrapolate from the data of the accelerated test to the separate distribution functions of the failures in the use environment. If the results of these failures differ, for example in cost of failure to the system in which the device functions, it is important to have the separate distribution functions of time-to-failure in order to make statistical or other decisions.

In practice, failures from one risk may be totally obscured by failures from another risk. This eventuality presents great dangers when the various risks are not known a priori, and one relies on experimental data to indicate their presence. Allen [VI-5] shows that, in the case of an increased-voltage test for the dielectric breakdown of a capacitor, the accelerated test can increase the failure rate for this risk relative to the failure rate for all other risks combined and thus further increase the probability of not observing any failures due to the latter risk. Without observations on the failures other than dielectric breakdown, one cannot estimate the underlying distribution for that risk and thus one cannot extrapolate to use conditions, and predictions of performance in the use environment will be invalid.

Another reason for separating the types of failures from each other is that the accelerated test may also alter the underlying distribution for the second risk in the above example so that $F_2^* \neq F_2$. In the initial example of the insurance company, the absence of a cure for cancer presumably does not change the underlying distribution of deaths from other causes (although it will increase the number of non-cancer deaths observed at advanced ages). However, in a circuit consisting of a capacitor and resistor connected in parallel, the increased voltage test used to accelerate dielectric breakdown failures of the capacitor most certainly affects the underlying distribution for resistor failures.

Unless the dependency of the underlying failure distributions upon the environment is known not to require independent estimation using the different kinds of failures, it is essential to break down the total observed failures into the different risks. To illustrate this point, consider a competing risks model with two risks, whose underlying distributions are exponential with parameters $\lambda_j = a_j b_j^{\varepsilon}$, $j = 1,2$. The environment is represented by the constant ε ; in the test $\varepsilon > 0$ and the experimenter wishes to extrapolate to $\varepsilon = 0$ for the use environment. Then the distribution of all failures in the test is also exponential with parameter $\lambda = a_1 b_1^{\varepsilon} + a_2 b_2^{\varepsilon}$, and independent estimation of a_1, a_2, b_1, and b_2 must be done in order to extrapolate. On the other hand, if $\lambda_j = a_j + b_j \cdot \varepsilon$, the distribution of all failures is exponential with parameter $\lambda = (a_1+a_2) + (b_1+b_2)\varepsilon$. To extrapolate, one can ignore the differences between failure types and estimate the sums (a_1+a_2) and (b_1+b_2). Since the relationship between test and use environments is commonly not easy to specify so precisely in practice, it is unwise to ignore the differences among types of failures.

Which assumptions about the model should be chosen, is part of the problem of finding the relation between the distribution functions and the environments of interest. The physical behavior of the devices determines the choice of assumptions about the model, e.g., the kind of underlying distribution functions that are postulated. The assumptions then are translated into functions for the various distributions and joint probabilities used in the model and required for estimation. Again from physical considerations one gets the transformations that mathematically state the relation of these functions to the environment. Sometimes the assumptions that lead to the simplest and most well-developed estimators do not fit the physical behavior of the devices and ease of estimation must be sacrificed; otherwise the extrapolation to the use environment is invalid.

Such a conflict between ease of estimation and physical considerations comes about when a decision must be made concerning the validity or invalidity of the constant hazard ratio assumption (VI.A.3.3-7) in the CR model. In a great many situations the hazard ratio is certainly not constant. A non-constant hazard ratio situation is best illustrated by a situation where one risk is due to a damage or wear process, which eventually induces failure due to an accumulation of a critical amount of damage, and all other risks are due to other kinds of causes. In the damage process, the hazard function peaks sharply or becomes very large in some interval of time. A good example of such a situation is illustrated by the time to "death" of a snowman where "death" is considered to have occurred either upon his total demise by melting, or when he is first hit by a snowball flung by a prankster. Clearly the risk due to melting is of the wear type, thus the hazard function for this risk will certainly be non-constant. The other cause is essentially a Poisson process, so that the hazard function is constant in time. Clearly the ratio of the hazard functions for the two risks cannot be constant. Although this example was constructed for illustrative purposes only, it follows that the ratio of the hazard functions may fail to be constant for any CR situation where one risk is due to a damage or wear process and the other risks, of which there is at least one, are due to other random causes.

This subsection was devoted to a discussion of the reasons for separately estimating the underlying distributions F_i. As to the possibility of actually performing such separate estimation, it must be admitted that it is essentially an open problem, unless one makes a priori assumptions regarding the form of the F_i's. Berkson and Elvebach [VI-6] have extensively studied the simplest case of competing risks, in which the underlying distributions of two competing risks are both exponential. Plausible procedures are developed in Subsection VI.A.4.2 for the more general and more realistic case where one merely assumes that a constant hazard ratio exists for at least one of the competing risks. However, as indicated above, many practical instances may fail to satisfy the constant hazard ratio condition. For that general instance, no non-parametric estimation techniques are presently known, though it appears that they could be developed from Berman's theorem [Eqs. (VI.A.3.2-4) and (VI.A.3.2-5)]. An approach for the general case, based on an a priori assumption of the form of the underlying distributions, is briefly discussed in Subsection VI.A.4.2.

VI.A.4.2 Estimation Procedures

VI.A.4.2.1 Estimation in the CHR Case

In the competing risks model, the underlying distributions of the different types of failures are not directly observable. In this subsection ,techniques for the estimation of the i-th underlying distribution corresponding to the i-th risk shall be presented for the case in which it is known that the constant hazard ratio (CHR) condition holds for the i-th risk. From Eq. (VI.A.3.3-10), it is seen that the following relation defines a plausible estimator for the underlying distribution of the i-th risk, F_i:

$$\hat{F}_i(x) = 1 - [1 - \hat{F}(x)]^{\hat{p}_i} , \qquad \text{(VI.A.4.2.1-1)}$$

where $\hat{F}$ and $\hat{p}_i$ are the estimators of F and p_i respectively.

The estimator $\hat{F}_i$ of F_i defined by Eq. (VI.A.4.2.1-1) is a plausible estimator when estimates of F and p_i are obtained from the following experiment: assume that the experimenter is testing M devices and that all M items have failed by the end of the test. Label the duration of the test as T, and let $K_i(x)$ be the number of failures of type i observed by time x. The total number of failures by time x is $K(x) \equiv \sum_i K_i(x)$; obviously, in this situation, $K(T) = M$. (The question of correct inclusion of failure times of devices which survive the test is a common problem in reliability testing; since all items fail by the end of the test described here, this problem does not arise.)

Under the assumption of a constant hazard ratio condition, it is known (see Subsection VI.A.3.3) that $p_i(x)$, the expected proportion of observed failures of type i, is a constant function in x, that is to say

$$p_i(x) = p_i \quad \text{for all } x.$$

From the laws of large numbers, it is known that the sample proportion converges with probability one to the true proportion, and the most efficient estimator of p_i is given by

$$\hat{p}_i = \frac{K_i(T)}{K(T)} = \frac{K_i(T)}{M} \quad . \qquad \text{(VI.A.4.2.1-2)}$$

The empirical distribution function $\hat{F}(x) = K(x)/K(T)$ is used to estimate $F(x)$. The values of these estimators for a given sample are substituted in Eq. (VI.A.4.2.1-1) to get estimates of $F_i(x)$.

If the CHR condition does not hold, the estimation procedure specified by Eqs. (VI.A.4.2.1-1) and (VI.A.4.2.1-2) leads to incorrect estimates of $F_i(x)$, but the degree of the error is unknown. Therefore, it is wise to test for the constancy of the hazard ratio first; tests for a constant hazard ratio condition are outlined in Subsection VI.A.4.2.2, below.

Finally note that Berkson and Elveback [VI-6] have studied extensively the estimation problem in a simple case in which a CHR condition holds. Specifically, they investigated two competing risks, each of which has an exponential underlying distribution.

VI.A.4.2.2 Testing the CHR Condition

A test for the existence of a CHR for the i-th risk depends on the following

Lemma: $G_i(x) = F(x)$ for all x (VI.A.4.2.2-1)

if and only if

a constant hazard ratio exists

for the i-th risk.

Proof: From Equations (VI.A.3.3-3) and (VI.A.3.3-1) it follows that

$$G_i(x) = \frac{P[X \leq x \text{ and } N = i]}{P[N = i]}$$

$$= \frac{P(X \leq x)\ P[N = i \mid X \leq x]}{p_i}$$

$$= \frac{F(x)\ p_i(x)}{p_i} .$$

Hence, $G_i(x) = F(x)$ for all x if and only if $p_i(x) = p_i$ for all x. This condition, in turn, is known to be equivalent to the existence of a CHR for that risk (see Subsection VI.A.3.3).

Epstein [VI-7] proved a special case of the above lemma and also suggested the general form.

By virtue of the above lemma, one can test for the existence of a CHR condition by testing two samples for equality of their distributions. Thus the existence of a CHR for a particular risk can be tested as follows. Let the risk in which one is interested be labelled "1" and let all other risks be considered collectively as risk "2". As shown in the discussion following the definition (VI.A.3.3-7), it is known that either both risks have constant hazard ratios or neither risk has a CHR when there are only two risks. Therefore it follows from the lemma that $G_1 = G_2$ if and only if the CHR condition holds. Perform two experiments in the manner described in Subsection VI.A.4.2.1. Use the data from one experiment to construct the empirical distribution function corresponding to G_1, and similarly use data from the other experiment for G_2. The equality of G_1 and G_2 can now be tested by a standard two-sample test (see discussion of such tests in Subsection III.D.2). Denote by $\hat{G}_i$ the empirical distribution function of a sample from a (hypothetical) population with distribution G_i. To obtain an expression for $\hat{G}_i$, start with the definition of H_i, which in turn, defines G_i. Recall that $H_i(x)$ is the probability that a failure has occurred by time x and that the failure is of the i-th type; i.e., $H_i(x)$ is the expected proportion of failures prior to time x which are of type i. This suggests the following estimator:

$$\hat{H}_i(x) = \frac{K_i(x)}{K(T)} \quad \text{(VI.A.4.2.2-2)}$$

for all x. Substituting the estimators for $H_i(x)$ and p_i into Eq. (VI.A.3.3-3), one obtains

$$\hat{G}_i(x) = \frac{\hat{H}_i(x)}{\hat{p}_i} = \frac{K_i(x)}{K_i(T)} \quad \text{(VI.A.4.2.2-3)}$$

Indeed, among the M observed failures, exactly $K_i(T)$ of the i-th type occurred. $G_i(x)$ is simply the expected proportion of these $K_i(T)$ failures which occur prior to time x; i.e., $G_i(x) = E[K_i(x)/K_i(T)]$.

Finally, note that it may be correct to consider type 1 failure times and type 2 failure times from the same experiment as independent samples from populations with distributions G_1 and G_2. In that case, the two-sample tests are applicable and one experiment is sufficient for obtaining all the required data. This procedure is, in fact, followed by Epstein [VI-7].

VI.A.4.2.3 <u>General Case for the CR Model</u>

When a CHR can be assumed to exist for some failure risk, considerable time and effort can be saved in the estimation of the failure distribution for that risk since the estimators described in the preceding subsection (VI.A.4.2.1) are extremely simple. However, for some failure risks, it is not plausible to postualte a CHR. For these risks the result of Berman [VI-3] in Eq. (VI.A.3.2-5) suggests the following possible estimator for F_i:

$$\hat{F}_i(x) = 1 - \exp\left\{-\int_0^x \frac{\widehat{dH}_i(t)}{\left[1-\sum_{j=1}^{n} \hat{H}_j(t)\right]}\right\} , \qquad \text{(VI.A.4.2.3-1)}$$

when the estimators, $\widehat{dH}_i(t)$ and $\hat{H}_j(t)$, exist.

Although non-parametric estimators exist for $H_j(t)$, procedures for the non-parametric estimation of integrals such as the one above are not readily available, and one must rely on parametric methods. First assume that the underlying distribution functions $F_i(i = 1, 2, \ldots, n)$ have certain forms, for instance, Weibull distributions with different shape and scale parameters for different risks. (For the Weibull distribution there exists a CHR for one risk if and only if all the shape parameters are equal and, thus, all risks have CHRs.) Now substitution of the formulas for these distributions into Eq. (VI.A.3.2-4) yields the forms for the corresponding joint probabilities $H_i(x)$, as functions of the unknown parameters of $\{F_i\}$. The failure-times observed in an experiment are samples from a population described by the functions $\{H_i(x)\}$. Therefore, one can fit the formulas derived for $H_i(x)$ to the estimates $\hat{H}_i(x)$ obtained in Eq. (VI.A.4.2.2-2), i.e, to the empirical functions $K_i(x)/K(T)$, and thus get estimates of the parameters of the distribution F_i.

Though such an estimation procedure is feasible, its properties have not been investigated. It appears that such an investigation would have to rely on Monte Carlo methods.

IV.A.4.2.4 Estimation for the MP Model

For the mixed populations model, one can use the estimators for p_i, $F(x)$, $H_i(x)$, and $G_i(x)$ which were derived in Subsections VI.A.4.2.1 and VI.A.4.2.2. The estimator for $F_i(x)$ is $\hat{G}_i(x)$ since these distributions are identical in the MP model [see Eq. (VI.A.3.3-5)]. All these estimators apply in tests similar to the ones of Subsection VI.A.4.2.1.

For a test which ends at the common truncation time T, regardless of whether all the devices have failed or not, Mendenhall and Hader [VI-8] have studied estimation procedures for situations with two subpopulations whose underlying distribution functions are exponential. Assuming that it is possible to assign a device to one of the subpopulations after it has failed, the authors derive maximum likelihood estimators for p, λ_1, and λ_2, where p is as in Eq. (VI.A.3.1-2) and λ_i is the "failure rate" of the i-th exponential subpopulation. Swami and Doss [VI-9] extended Mendenhall and Hader's results to the general case where M devices are tested and the test is terminated at some time T_j for the j-th device tested (j = 1, 2, ..., M). (This modification, i.e., unequal truncation times, was studied for the simpler case of a single exponential population by Bartholomew.) The authors assume that, if the j-th device fails before T_j, it is possible to determine the subpopulation to which it belongs;

otherwise it is not possible and the failure time is known only to exceed T_j. For two exponential subpopulations the authors derive maximum likelihood extimators for p, λ_1, and λ_2. The variances and co-variances of their estimators are given by Swami and Doss, but other properties of the estimators are not known. Gumbel [VI-10] studied a similar problem; he derived estimators for a mixture of two Poisson processes under the assumption that the mix-ratio 'p' is known.

Rider [VI-11] studies the "dissection" of a mix of two exponential distributions. He assumes that the experimenter never knows the sub-population to which a device belongs, i.e., that only failure-time data are available. Using $F(x) = pF_1(x) + (1-p)F_2(x)$ as his model and assuming that F_1 and F_2 are exponential, he derives estimators for p and for the parameters λ_1 and λ_2 of the exponential distributions by the method of moments. This approach seems most applicable to situations in which there is no a priori reason to doubt the exponentiality of the failure-time distribution but the data contradict this. In this case perhaps there is an unknown and previously unsuspected mixture of exponentially distributed populations which can be separated, for a better fit of the data, by Rider's approach. This estimation procedure can in principle be extended to mixes of other distributions and has, in fact, been studied by Rider for a mixture of two Weibull distributions with identical shape parameters.

VI.B ESTIMATION FOR REPAIRABLE DEVICES

Most statistical estimation procedures are based on random sampling: "Given a sequence of independent, identically distributed, random variables X_1, X_2, ..., X_n, ... with common distribution function F, ..., etc.". However, in life-estimation of repairable equipments, common practice usually leads to a sequence of samples which are neither independent nor identically distributed. The problem in this case is how to obtain the estimates needed for making reliability predictions about the life of the equipment. To visualize the type of failure data arising from the operation of devices which are repairable, consider a radio, which is turned on at some time t. When it "fails", i.e., ceases to work properly for some reason, the radio is repaired by replacing or fixing the part(s) responsible for the "failure", and the radio is then turned on again. This process continues for some specified time and the sequence of failure times are recorded in the form $(t_1, t_2, \ldots, t_r)$, where the last failure time t_r is less than or equal to the total observation time T. (Repair time can be disregarded when all data on failure times are recorded in terms of actual operating time only.) If one wants to make predictions about the time the radio could operate perfectly, the distribution function for the time-to-first-failure (TTFF) must be estimated. A manufacturer of radios might want to obtain such an estimate in order to evaluate the cost of the warantees he issues with the radios.

Similarly, space guidance devices are repairable when they are not in flight in space. If n such devices are in a life testing situation in which they can be repaired, each device generates a sequence of failure times in the same form $(t_1, t_2, \ldots, t_r)$ as the data obtained from the radio, and there are as many such sequences as there are devices on test. As a rule the guidance device is not readily repairable under the conditions of outer space, so an estimate of the distribution of the TTFF is needed to be able to make reliability predictions regarding the operation of the device in space.

To estimate the population distribution of the TTFFs from such observed sequences one can use the empirical distribution function of first failure times only, but this procedure is wasteful since there exists information in the subsequent times-between-failures (TBFs). Unfortunately, statistically valid procedures for using the TBFs to estimate the distribution of TTFFs are not available, except in some special cases. In general, it is not clear what effects the failure and repair have upon the failure rate, and hence, upon the distribution of times-to-failure; therefore, one cannot state that the TBFs come from the same population of failure times as the TTFFs. Thus, it is usually not clear what to do with the data on TBFs.

However, in certain cases where appropriate assumptions can be made about the repair process, it is possible to use these data. In particular, there are two classes of assumptions for which usable estimators based upon all failure time data have been derived. In the first class, it is assumed that each repair renews the device to its original condition at the beginning of observation; then the sequence of failures can be described statistically by a renewal process. When the underlying distribution of first failure times of a renewal process is known to be of a certain form, e.g., exponential, estimation procedures are available. If no form is assumed for the distribution, a non-parametric estimation procedure readily follows from results of Meier and Kaplan [VI-12]. The second class of such assumptions about the nature of the repair are described in the following paragraphs, and a non-parametric estimator for the underlying distribution is derived subsequent to this description.

Assume that the failure and repair of the device does not significantly alter its instantaneous failure behavior at time τ at which the failure occurs. For example, suppose that one of many tubes in a radio fails; then, in terms of its probability of failure in the immediate future, the repair of the radio will essentially restore it to the condition it was in just before it failed. Another way to look at such restoration is to assume that observation is started simultaneously on a large (hypothetically infinite) collection of identical devices, from which one device is randomly selected as the device on test. Whenever this device on test fails, it is replaced by another device chosen at random from those which have not yet failed. This replacement process continues until the test is terminated, and it is equivalent to assuming that repair restores the device in such a way that the occurrence of a failure does not affect future failure behavior.

Two additional assumptions require that the device be brand new at the beginning of the observation period, and that there exist no time period in which the device is known to be infallible. These assumptions are reasonable ones for radios, space guidance devices, etc. It is also true for such devices that one of them cannot fail twice or more instantaneously, so that the following assumption holds: there exists no interval Δt in which two or more failures can occur when Δt gets infinitely small. Since there can be no more than one failure at any instant τ, it follows that the probability of one or more failures in the immediate future equals the probability of exactly one failure in the immediate future. The final property required is that the number of failures occurring in any interval is independent of the number of failures in any other non-overlapping interval. This is a reasonable assumption for a repairable device which satisfies the above assumptions by virtue of the following argument: the probability of a failure in the immediate future, following a failure at t, i.e., in $(t, t + \Delta t]$, is not affected by the occurrence of the failure and so the probability of some number of failures in some interval beginning at t does not depend upon the occurrence of a failure or failures before that interval. This argument can be extended to cover any non-overlapping intervals and, therefore, the property holds.

Under the assumptions of the preceding paragraphs, the sequence of failures of the device can be described stochastically by a process which will be called the <u>restoration process</u> (also called a "minimal repair" process in Barlow and Hunter [VI-13]). To describe it mathematically, let F be the population distribution of times-to-first-failure; F will be called the <u>underlying distribution of the process</u>. Let $\phi(t) = \frac{d}{dt}[F(t)]/ [1 - F(t)]$ be the failure rate (or "hazard") function at t. Then the assumption that the repair at t restores the device to its condition just prior to the failure is equivalent to stating that $\lim_{\Delta t \to 0} \phi(t - \Delta t) = \lim_{\Delta t \to 0} \phi(t + \Delta t)$.

Furthermore, it follows from the definition of the failure rate function that

$$\lim_{\Delta t \to 0} \frac{P\ \{\text{one or more failures occur in } (t, t + \Delta t]\}}{\Delta t} = \phi(t).$$

Now, for a fixed t, let N_t be a random variable which takes on the values of the number of failures occurring in the interval (0, t]. When t varies and N_t satisfies all the assumptions described above, N_t is called a restoration process.

It can be seen that N_t is simply a non-homogeneous Poisson process described by Parzen [VI-14], p. 125.

From Parzen's Theorem 2.a, ([VI-14], p. 125), there immediately results the following

<u>Corollary</u>:

$$E(N_t) = \int_0^t \phi(u)du = \Phi(t). \qquad \text{(VI.B-1)}$$

Any observation on N_t is simply a sequence of failure-times; therefore, it is obvious that N_t is the appropriate process with which to describe the failure data from repairable devices which are restored in the manner that satisfies the assumptions stated previously. With this statistical model, it is now possible to estimate the distribution of the times-to-first-failure. First, consider a situation in which n identical repairable devices are operated independently and from which the sequence of failure-times for each device is obtained. If the sequence of failures for each device follows a restoration process, and if the underlying distribution F is the same for each process, then there exists a joint process which will be called an n-fold superposed restoration process. An example of a situation which can be described by such a process is a life test on n identical space guidance devices. Let the restoration process for the i-th device be labelled $N_t^{(i)}$.

It is common practice to perform estimation by replacing expected values by observed sample averages. Thus, the above corollary (VI.B-1) and the well-known relation

$$F(t) = 1 - \exp[-\Phi(t)]$$

suggest that the distribution function F can be estimated by the function H_n^* defined by

$$H_n^*(t) = 1 - \exp\left[-\frac{1}{n} \sum_{i=1}^{n} N_t^{(i)} \right], \quad 0 \leq t < \infty$$

$$H_n^*(\infty) = 1 \qquad \text{(VI.B-2)}$$

This function is a desirable estimator of F, as can be seen from the following theorem.

Theorem: Consider an n-fold superposed restoration process with underlying distribution F and, for $0 < t$, let $N_t^{(i)}$ be the number of failures in (0, t] for the i-th process. Then the function $H_n^*(t)$ defined by Eq. (VI.B-2) converges strongly to F, i.e.,

$$P\left\{ \lim_{n\to\infty} \left[H_n^*(t) - F(t) \right] = 0 \right\} = 1 \text{ for all } t. \qquad \text{(VI.B-3)}$$

Proof: For any fixed t, in view of Corollary (VI.B-1), the random variables $N_t^{(i)}$ $(i = 1, 2, \ldots, n)$, are independent and identically distributed with expected value $\Phi(t)$. Therefore, the Kolmogorov strong law of large numbers (see Kolmogorov [VI-15] or Fisz [VI-16] p. 226) applies to the sequence $\{Y_i\}$, $Y_i \equiv N_t^{(i)}$. From the law one obtains

$$P\left\{ \lim_{n\to\infty} \left[\frac{1}{n} \sum_{i=1}^{n} N_t^{(i)} - \Phi(t) \right] = 0 \right\} = 1 , \qquad \text{(VI.B-4)}$$

or

$$P\left\{ \lim_{n\to\infty} \left[\frac{1}{n} \sum_{i=1}^{n} N_t^{(i)} \right] = \Phi(t) \right\} = 1 .$$

Since the function e^{-x} is a continuous function of x, the negative exponential of the average in brackets converges strongly to the function exp $[-\Phi(t)]$. Rewriting Eq. (VI.B-4) then yields the following convergence:

$$P\left\{\lim_{n\to\infty}\left[1-\exp\left(-\frac{1}{n}\sum_{i=1}^{n}N_t^{(i)}\right)\right] = 1-\exp\left[-\Phi(t)\right]\right\} = 1 ,$$

which proves the theorem.

$H_n^*(t)$ does have the drawback of being biased. Fortunately, the bias can be removed; the function H_n (defined only for $n > 1$)

$$H_n(t) = 1 - \exp\left\{\left[\log\frac{(n-1)}{n}\right]\sum_{i=1}^{n}N_t^{(i)}\right\} \qquad 0 \leq t < \infty \quad \text{(VI.B-5)}$$

$$H_n(\infty) = 1$$

can be shown to be an unbiased estimator of F(t). Furthermore, the proof of the strong convergence of H_n^* to F is readily adaptable to proving that H_n converges strongly to F.

The statistic H_n ~~is a desirable estimator since~~ (with probability one) it can be made to fit F with arbitrary closeness, provided that the sample size n is sufficiently large. Since convergence in probability is implied by convergence with probability one, H_n is a consistent estimator of F. This estimator can be used for observations of a fixed length T or for observations that continue until a fixed number of failures have occurred for each device, provided that none of the devices is removed or lost from observation. One has only to count the number of failures observed at or before some time t, and use this count in Eq. (VI.B-5). The computed function $H_n(t)$ will jump at those times t at which failures occur, resembling an empirical distribution function.

A simulated test was done for the observation interval (0, 90], under the assumptions that there were ten devices on test, i.e., there was a ten-fold superposed restoration process, and that the underlying distribution of the process was the uniform distribution over the interval [0, 100]. To see how such a test is simulated recall the following: one way to look at a restoration process is to assume that each device on test that fails is replaced by a randomly selected device from the survivors of a hypothetical infinite collection of devices, started at the same time as the tested item. Then the times-to-failure of individual devices are simulated by selecting a sequence of numbers from a table of random numbers. When each device fails, at a time t indicated by a random number, it is replaced by a device whose failure time is the next random number in the sequence which is greater than t (since the replacement is assumed to have survived past t). Thus, a random number between 0 and 100 was selected and then random numbers were skipped until one was found which was greater than the last one selected, etc., until one was found which exceeded 90. The failure time sequences resulting from this restoration process are shown in Table VI-1. The last number selected in each of the ten sequences is in parentheses and is not considered part of the observed data. Figure VI-1 shows a plot of $H_{10}(t)$ based on the data

TABLE VI-1. EXAMPLE DATA

Device Number	Failure Times
1	47, 50, 67, 73, (96)
2	75, 76, 85, (92)
3	53, 74, 75, 88, (92)
4	81, (98)
5	42, (93)
6	24, 56, 70, 86, (91)
7	79, (99)
8	38, 81, (93)
9	52, 53, 72, 84, 86, (96)
10	50, (92)

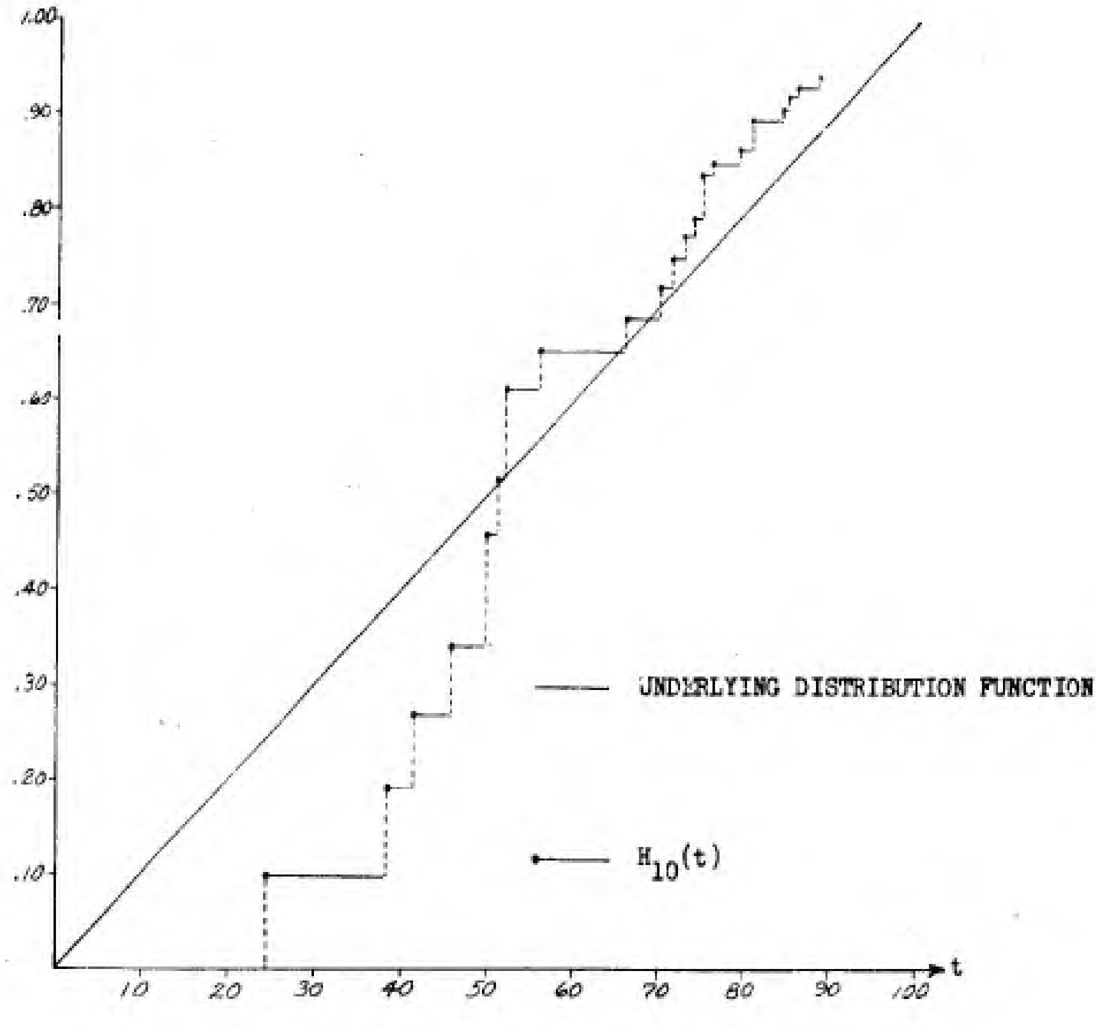


FIG. VI-1. DISTRIBUTION FUNCTION F ESTIMATED BY H_{10}

of Table VI-1, with the underlying distribution shown for comparison.

Finally, note that one of the advantages of using H_n as an estimator for the underlying distribution function is that it is non-parametric. Of course, the restoration process can be studied when the form of the underlying distribution is assumed; for example, Störmer [VI-17] has studied the restoration process under the assumption that the underlying distribution is Weibull.

VI. C STOCHASTIC APPROXIMATION

In this section, the technique known as stochastic approximation is briefly described, and reasons are given for the conclusion that it is not directly applicable to accelerated testing. The section concludes with an indication of some reliability problems to which the technique is applicable. Of course, some such problems may arise in connection with accelerated testing.

Suppose that to each value of the real-valued parameter x there corresponds a random variable $Y = Y(x)$ with continuous distribution function $H(y|x)$. Let $M(x)$ be the regression function of Y on x, i.e.,

$$M(x) = \int_{-\infty}^{\infty} y \, dH(y|x) \quad .$$

For known α, consider the equation

$$M(x) = \alpha$$

and suppose that it has a unique solution θ, i.e.,

$$M(\theta) = \alpha .$$

The technique of stochastic approximation enables one to construct a sequence $x_1, x_2, \ldots, x_n, \ldots$ which, in an appropriate sense and under certain sufficient conditions, converges to the solution θ. The technique was originated by Robbins and Monro [VI-18] and has received considerable attention subsequently; for example, in the Annals of Mathematical Statistics for the period 1952 - 1954, this subject is considered on pp. 457-461 of Volume 23, and pp. 382-386, 386-388, 463-483, 737-744 of Volume 25.

This subject was examined because of its connection with the subject discussed in Section I.C, i.e., the dependence of life-distributions on environment. If the parameter x represents the environment, and the random variable Y(x) is the time-to-failure in the environment, then M(x) is the mean time-to-failure in that environment. The sequence x_1, x_2, ..., x_n, ... is the sequence of environmental settings at which life-tests should be performed in order to find the environment at which the mean life has a specified value α. The technique, however, is not applicable to problems of accelerated life testing, for the following reasons:

(i) The usual problem is to determine (some properties of) the life distribution at a given environment. The technique of stochastic approximation is directed at the converse problem: having specified a property of the life distribution (i.e., its mathematical expectation), find the environment to which this property corresponds.

(ii) In a life testing application, the technique would be extremely time consuming: the choice of environmental level x_{n+1} depends on the outcome of the test at level x_n, i.e., the sequence of life tests have to be performed in succession, rather than simultaneously.

Though not applicable in a life-testing context, the technique of stochastic approximation can be very useful for some reliability problems such as those where "strength" rather than time-to-failure is a critical variable. The theoretical interpretation applicable in such instances is given below, followed by an example.

Let there be given a constant α, $0 < \alpha < 1$. Let Z_n be a sequence of independent random variables, each with the same distribution F, and also let z_n be the particular value of a single sample of Z_n. Assume that the actual value of z_n is not known to the experimenter but that, for a preassigned value of the parameter x_n, he knows whether $z_n \leq x_n$ or $z_n > x_n$. Then the technique of stochastic approximation specifies a sequence $x_1, x_2, \ldots, x_n, \ldots$ which, in an appropriate sense, converges to a value θ such that

$$F(\theta) = \alpha$$

For example, let Z be the current needed to explode a squib, and let a squib be tested by sudden application of a preassigned current. Assume that any squib can be tested only once, say because a former test modifies its inherent "critical current" in some unknown fashion. Here z_n is the critical current of the n-th squib tested, and x_n is the current used in testing it. The observable result is whether $z_n \leq x_n$ (squib explodes) or $z_n > x_n$ (no explosion). The stochastic approximation technique allows one to select current levels successively (based on former levels and outcomes) so that they will converge to the value at which the probability that a randomly selected squib will explode is equal to the specified value α.

Another possible application of stochastic approximation is in the determination of the load which will cause a specified amount of creep in specified time for some item. In particular one might wish to find the load such that the probability that a randomly selected item will exceed a maximum permissible creep in a specified amount of time is α (say .05), when that load is applied. That is to say, let Z be the critical load necessary to cause a maximum permissible creep in a specified amount of time. Obviously an item can only be tested once. In this case Z_n is the critical load for the n-th item, and x_n is the applied load in testing the n-th item. The observable result is whether $z_n \leq x_n$ (in which case creep exceeds maximum permissible creep) or $z_n > x_n$ (in which case creep does not exceed maximum permissible creep). Again the stochastic approximation technique allows one to select load levels (based on former levels and outcomes) so that they will converge to the value at which the probability that a randomly selected item will exceed the maximum permissible creep in the specified time is equal to the specified value α (say .05).

VII. BIBLIOGRAPHY

References for Part I

I-1 ALLEN, W. R. : "Statistical inference from accelerated processes"; pp. 85-93 of Proceedings of the Statistical Techniques in Missile Evaluation Symposium, B. Harshbarger, ed.; Virginia Polytechnic Institute, Blacksburg, Va.: 1958.

I-2 ALLEN, W. R. : "Inference from tests with continuously increasing stress"; Operations Res. 7 (1959), 303-312.

I-3 ADAMS, D.A. : "Component part failure rate curve considerations"; Ninth Nat'l. Symp. Rel. and Q.C., 301-314; 1963.

I-4 LEVENBACH, G. J. : "Accelerated life testing of capacitors"; IRE Trans. on Rel. and Q.C., RQC-10 (June, 1957), 9-20.

I-5 ENDICOTT, H. S. and J. A. ZOELLNER : "A preliminary investigation of the steady and progressive stress testing of mica capacitors"; Seventh Nat'l. Symp. Rel. and Q. C., 229-240; 1961.

I-6 BERKSON, J. and L. ELVEBACK : "Competing exponential risks, with particular reference to the study of smoking and lung cancer"; J. Amer. Statist. Assoc. 55 (1960), 415-428.

I-7 LYNCH, W. M., C. B. CLARK, and B. EPSTEIN : Advanced Vehicle Guidance Systems Reliability Investigation; Technical Documentary Report ASD-TDR-62-359; Wright-Patterson Air Force Base, Ohio; 1962.

I-8 ALLEN, W. R. : "A note on conditional probability of failure when hazards are proportional"; Operations Res. 11 (1963), 658-659.

I-9 BERMAN, S. M. : "Note on extreme values, competing risks and semi-Markov processes"; Ann. Math. Statist. 34 (1963), 1104-1106.

References for Part II

II-1 HALD, A. : Statistical Theory with Engineering Applications; John Wiley & Sons, New York; 1952.

II-2 EPSTEIN, B. : Two Sample Life Tests; Technical Report No. 8 under ONR Contract Nonr-2575(00); Wayne State University, Detroit, Michigan; 1960.

II-3 GUMBEL, E. J. : "Les intervales extremes entre les emissions radio-actives"; Journal de Physique et le Radium 8, ser. 7 (1937), 321-329.

II-4 EPSTEIN, B. and M. SOBEL : "Life testing"; J. Amer. Statist. Assoc. 48 (1953), 486-502.

II-5 EPSTEIN, B. : "A sequential two sample test"; J. Franklin Inst., vol. 260, no. 1 (July, 1955), 25-29.

II-6 SIMONDS, T. A. : "MTBF confidence limits"; Industrial Quality Control, vol. XX, no. 6 (December, 1963), 21-27.

II-7	EPSTEIN, B.	:	Statistical Techniques in Life Testing; OTS Publication PB 171580, U. S. Dept. of Commerce, Washington, D.C.; n.d. (circa 1961).
II-8	EPSTEIN, B.	:	"Estimates of bounded relative error for the mean life of an exponential distribution"; Technometrics 3 (1961), 107-109.
II-9	LEVENBACH, G. J.	:	"Accelerated life testing of capacitors"; IRE Trans. on Reliability and Quality Control, RQC-10 (June, 1957), 9-20.
II-10	ALLEN, W. R.	:	"Statistical inference from accelerated processes"; pp. 85-93 of Proceedings of the Statistical Techniques in Missile Evaluation Symposium, B. Harshbarger, ed.; Virginia Polytechnic Institute, Blacksburg, Va.; 1958.
II-11	ALLEN, W. R.	:	"Inference from tests with continuously increasing stress; Operations Res. 7 (1959), 303-312.

II-12 (No author) : <u>Quality Control and Reliability Handbook H108, Sampling Procedures and Tables for Life and Reliability Testing (Based on Exponential Distribution)</u>; Office of the Assistant Secretary of Defense, Washington 25, D.C.; 1960.

II-13 EPSTEIN, B. : <u>Testing of Hypotheses</u>; Technical Report No. 3 under ONR Contract Nonr-2163(00); Wayne State University, Detroit, Michigan.

II-14 EPSTEIN, B. and M. SOBEL : "Some theorems relevant to life testing from an exponential distribution"; <u>Ann. Math. Statist. 25</u> (1954), 373-381.

II-15 HIETALA, H. J. and F. R. MAIOCCO : <u>An Accelerated Lot Acceptance Test with Monte Carlo Plots of the Distribution of the Product of an F Variate with a χ^2 Variate Over its Degrees of Freedom</u>; Report No. TM 560-43-RSA-19; Autonetics, Anaheim, Calif.; 1964.

II-16	ENDICOTT, h. S. and J. A. ZOELLNER	: "A preliminary investigation of the steady and progressive stress testing of mica capacitors"; Seventh Nat'l Symp. Rel. and Q. C., 229-240; 1961.
II-17	PROT, E. M.	: "Fatigue testing under progressive loading, a new technique for testing materials" (in French); Revue de Metallurgie XLV (1948), 481-489.
II-18	STARR, W. T. and H. S. ENDICOTT	: "Progressive stress - a new accelerated approach to voltage endurance"; Trans. AIEE (Power Apparatus and Systems) 80 (1961), 515-522.
II-19	DAKIN, T. W.	: A discussion of paper by Starr and Endicott [II-18], Trans. AIEE (Power Apparatus and Systems) 80 (1961), 522-523.
II-20	BESSLER, S., H. CHERNOFF, and A. W. MARSHALL	: "An optimal sequential accelerated life test"; Technometrics 4 (1962), 367-379.
II-21	CHERNOFF, H.	: "Optimal accelerated life designs for estimation"; Technometrics 4 (1962), 381-408.

II-22 EHRENFELD, S. and H. DeCICCO : "Estimation, control, and verification procedures for a reliability model based on variables data"; Management Science vol. 10, no. 2 (Jan., 1964), 249-260.

II-23 DORN, J. E. : Energetics in Dislocation Mechanics; Report No. 10455; Lawrence Radiation Laboratory, University of California, Berkeley, Calif.; 1963.

II-24 JAFFE, N. and J. E. DORN : Effects of Temperature on the Creep of Polycrystalline Aluminum by the Cross-slip Mechanism; Report No. 108; Minerals Research Laboratory, University of California, Berkeley; 1960.

References for Part III

III-1 PARZEN, E. : Stochastic Processes; Holden-Day, San Francisco; 1962.

III-2 EPSTEIN, B. : "The exponential distribution and its role in life testing"; Industrial Quality Control vol. XV, no. 6 (December, 1958), 4-9.

III-3 ALLEN, W. R. : "Inference from tests with continuously increasing stress"; Operations Res. 7 (1959), 303-312.

III-4 EHRENFELD, S. and H. DeCICCO : "Estimation, control and verification procedures for a reliability model based on variables data"; Management Science vol. 10, no. 2 (January, 1964), 249-260.

III-5 MERCER, A. : "Some simple wear dependent renewal processes"; J. Royal Stat. Soc.-Series B 23 (1961); 368-376.

III-6 ENDICOTT, H. S. and J. A. ZOELLNER : "A preliminary investigation of the steady and progressive stress testing of mica capacitors"; Seventh Nat'l. Symp. Rel. and Q. C., 229-240; 1961.

III-7 PROT, E. M. : "Fatigue testing under progressive loading, a new technique for testing materials" (in French); Revue de Metallurgie XLV (1948), 481-489.

III-8 DAKIN, T. W. : Discussion of paper by Starr and Endicott [II-18] , Trans. AIEE (Power Apparatus and Systems) 80 (1961), 522-523.

III-9 HILL, B. M. : "A test of linearity versus convexity of a median regression curve"; Ann. Math. Statist. 33 (1962), 1096-1123.

III-10 MOOD, A. M. : An Introduction to the Theory of Statistics; McGraw Hill, New York; 1950.

III-11 MERCER, A. and C. S. SMITH : "A random walk in which the steps occur randomly in time"; Biometrika 46 (1959), 30-35.

III-12 BIRNBAUM, Z. W. and S. C. SAUNDERS : "A statistical model for life-length of materials"; J. Amer. Statist. Assoc. 53 (1958), 151-160.

III-13 COX, D. R. : Renewal Theory; John Wiley & Sons, New York; 1962.

III-14 VON MISES, R. : "On the asymptotic distribution of differentiable statistical functions"; Ann. Math. Statist. 18 (1947), 309-348.

III-15 ANDERSON, T. W. and D. A. DARLING : "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes"; Ann. Math. Statist. 23 (1952), 193-212.

III-16 WALD, A. and J. WOLFOWITZ : "Confidence limits for continuous distribution functions"; *Ann. Math. Statist.* *10* (1939), 105-118.

III-17 WALD, A. and J. WOLFOWITZ : "Note on confidence limits for continuous distribution functions"; *Ann. Math. Statist.* *12* (1941), 118-119.

III-18 RENYI, A. : "On the theory of ordered samples"; *Acta Math. Acad. Sci. Hungar.* *4* (1953), 191-231.

III-19 CHUNG, K. L. : "On the ratio of an empirical distribution to the theoretical distribution function"; *Acta Math. Sinica* *5* (1955), 347-368.

III-20 MALMQUIST, S. : "On certain confidence contours for distribution functions"; *Ann. Math. Statist.* *25* (1954), 523-533.

III-21 B[illegible]KLAND, W. R. : *Statistical Assessment of the Life Characteristic*; Griffin's Statistical Monographs and Courses, No. 13; Charles Griffin and Co., London; 1964.

III-22 SMIRNOV, N. V. : "Sur la distribution de ω^2 criterion"; C. R. Acad. Sci. 202 (1936), 449-502.

III-23 SMIRNOV, N. V. : "Table for estimating the goodness of fit of empirical distributions"; Ann. Math. Statist. 19 (1948), 279-281.

III-24 KUNISAWA, K., H. MAKABE, and H. MORIMURA : "Tables of confidence bands for population distribution function"; Reports of Stat. App. Res. (JUSE) 1 (1951), 23-44.

III-25 KUNISAWA, K., H. MAKABE, and H. MORIMURA : "Notes on confidence bands of population distributions"; Reports of Stat. App. Res. (JUSE) 4 (1955), 18-20.

III-26 MILLER, L. H. : "Table of percentage points of Kolmogorov statistics"; J. Amer. Statist. Assoc. 51 (1956), 111-121.

III-27 KVIT, I. D. : "On N.V. Smirnov's theorem concerning the comparison of two samples"; Doklady Akad. Nauk SSSR (NS) 71 (1950), 229-231.

III-28 BLACKMAN, J. : "An extension of the Kolmogorov limit distribution"; Ann. Math. Statist. 27 (1956), 513-520 and 29 (1958), 318-324.

III-29 REIMANN and VINCZE : Publ. Math. Inst. Hung. Acad. Sci. 5 (1960), 293-309.

III-30 DAVID, H. T. : "A three-sample Kolmogorov-Smirnoff test"; Ann. Math. Statist. 29 (1958), 842-851.

III-31 BIRNBAUM, Z. W. and R. A. HALL : "Small sample distributions for multi-sample statistics of the Smirnov type"; Ann. Math. Statist. 31 (1960), 710-720.

III-32 KIEFER, J. : "K-sample analogues of the Kolmogorov-Smirnov and Cramer-v.Mises tests"; Ann. Math. Statist. 30 (1959), 420-447.

III-33 KENDALL, M. G. and A. STUART : The Advanced Theory of Statistics vol. II; Hafner, New York; 1961; (chapter 31).

III-34 DRION, E. F. : "Some distribution-free tests for the difference between two empirical cumulative distribution functions"; Ann. Math. Statist. 23 (1952), 563-574.

References for Part IV

IV-1 (No author) : Instrument Technical Description of VM-4A Velocity Meter; Report No. EM 262-285 (Confidential); Autonetics, Anaheim, Calif.; 1962.

IV-2 LANCASTER, J. K. : "The influence of the conditions of sliding on the wear of electro-graphitic brushes"; British Journal of Applied Physics 13 (1962), 468-477.

IV-3 DOWDY, J. L. : Private communication; Autonetics, Anaheim, California; June, 1964.

IV-4 SCHROTER, F. : (Title unknown; cited by Lancaster [IV-2]); Elektrotech. Z. A76, 498.

IV-5 BISSON, E. E. and W. J. ANDERSON : Advanced Bearing Technology; Publication No. SP-38; Natl. Aeronautics and Space Admin., Washington, D.C.; 1964.

IV-6 WEIBULL, W. : "A statistical representation of fatigue failures in solids"; Trans. Royal ~~Inst. Tech.~~, No. 27; Stockholm; 1949.

IV-7 BUTLER, R. H. and T. L. CARTER : Stress-Life Relation of the Rolling-Contact Fatigue Spin Rig; Report No. TN 3930; Natl. Adv. Committee on Aeronautics; 1957.

IV-8 CARTER, T. L. and E. W. ZARETSKY : Rolling-Contact Fatigue Life of a Crystal Glass Ceramic; Report No. TND-259; National Aeronautics and Space Administration; 1960.

IV-9 BARWELL, F. T. and D. SCOTT : "Effect of lubricant on pitting failure of ball bearings"; Engineering vol. 182, no. 4713 (July 6, 1956), 9-12.

IV-10 SCOTT, D. : "Lubricants at higher temperatures: assessing the effects on ball bearing failures". Engineering vol. 185, no. 4811 (May 23, 1958) 660-662.

IV-11 CORDIANO, H. V., E. P. COCHRAN, JR, and R. J. WOLFE : "A study of combustion resistant hydraulic fluids as ball bearing lubricants". Lubrication Engineering 12 (1956), 261-266.

IV-12 (No author) : "Viscosity and density of over 40 lubricating fluids of known composition at pressures to 150,000 psi and temperatures to 425°F"; ASME 2 (1953).

IV-13 CARTER, T. L. : A Study of Some Factors Affecting Rolling-Contact Fatigue Life; Report No. R-60; Natl. Aeronautics and Space Admin.; 1960.

IV-14 JACKSON, E. G. : "Rolling-contact fatigue evaluation of bearing materials and lubricants"; Transactions ASLE 2 (1959), 121-128.

IV-15 CONSTANTINE, R. W. : Ball Bearing Reliability Program; Report No. T-66/32-4; Autonetics, Anaheim, Calif.; 1964.

IV-16 PROT, E. M. : "Fatigue testing under progressive loading, a new technique for testing materials" (in French); Revue de Metallurgie XLV (1948), 481-489.

IV-17 SNYDER, W. L. : "Beryllium stability improvement"; Internal Letter No. 63-32-244-2-FGU-96; Autonetics, Anaheim, Calif.; 1963.

IV-18 LEMENT, B. S. and B. L. AVERBACH : Measurement and Control of the Dimensional Behavior of Metals; Report No. R-95; MIT Instrumentation Laboratory, Cambridge, Mass.; 1955.

IV-19 HUGHEL, T. J. : An Investigation of the Precision Mechanical Properties of Several Types of Beryllium; Report No. MR-120; General Motors; 1960.

IV-20 MYLER, M. L. : Short Term Microcreep Tests on Hot-Pressed and Forged Beryllium; Metallurgical Laboratory Report No. TR 63-369; Autonetics, Anaheim, Calif.; 1963.

References for Part V

V-1 PITMAN, G. R., JR. : Inertial Guidance; John Wiley & Sons, New York; 1962.

V-2 (No author) N17 System Detail Specifications; Publication No. SDS-N17-01; Autonetics, Anaheim, Calif.; 1963; Pa. 3.1.3.

V-3 SNYDER, W. L. : "Beryllium stability improvement"; Internal Letter 63-32-244-2-FGU-96; Autonetics, Anaheim, Calif.; 1963.

V-4 LEMENT, B. S. and B. L. AVERBACH : Measurement and Control of the Dimensional Behavior of Metals; Report No. R-95; MIT Instrumentation Laboratory, Cambridge, Mass.; 1955.

V-5 HUGHEL, T. J. : An Investigation of the Precision Mechanical Properties of Several Types of Beryllium; Report No. MR-120; General Motors Corp.; 1960.

V-6 MYERSON, M. R., et al. : "Gauge blocks of superior stability: initial developments in materials and measurement"; NBS Journal of Research 64C (1960), 175-207.

V-7 MYLER, M. L. : *Short Term Microcreep Tests on Hot-Press[illegible] Forged Beryllium*; Report No. [illegible]-63-369; Autonetics, Anaheim, Calif.; 1963.

V-8 CONRAD, H. : *Thermally-Activated Deformation of Metals*; Report No. TDR-169 (3240-11) TN-10; Aerospace Corporation, El Segundo, Calif.; June 17, 1963.

V-9 GUARD, R. : Private Communication; North American Aviation Science Center, Thousand Oaks, Calif.; May 4, 1964.

V-10 CARNAHAN, R. D. : Private Communication; Aerospace Corporation, El Segundo, Calif.; May 22, 1964.

V-11 TINDER, R. F., and J. WASHBURN : *Factors Affecting the Initiation of Plastic Flow in Metals*; report (number unknown) under Office of Naval Research contract no. Nonr-222(52); University of California, Berkeley.

V-12 LUBAHN, J. D. and R. P. FELGAR : Plasticity and Creep of Materials; John Wiley & Sons, New York; 1961; Chapters 6 and 10.

V-13 ZENER, C. M. : "Elasticity and anelasticity"; Trans. AIME 167 (1946); 155.

V-14 GUARD, R. : "The extrapolation of microcreep behavior"; (unpublished report); North American Aviation Science Center, Thousand Oaks, California; June 1964.

V-15 JAFFE, N. and J. E. DORN : Effects of Temperature on the Creep of Polycrystalline Aluminum by the Cross-slip Mechanism; Report No. 108; Minerals Research Laboratory, University of California, Berkeley; March 1960.

V-16 DORN, J. E. : Energetics in Dislocation Mechanics; Report No. 10455; Lawrence Radiation Laboratory, University of California, Berkeley, Calif.; February 1963.

V-17 LANGENECKER, B., et al. : Effects of Ultrasound on Deformation Characteristics of Structural Metals; Report No. 8482 (Navweps), TP 3447 (NOTS); Naval Ordinance Test Station, China Lake, California; March 1964.

V-18 LANGENECKER, B. : "Work-softening of metal crystals by alternating the rate of glide strain"; Acta Metallurgica 9 (1961), 937-940.

V-19 ROSENFIELD, A. R. : The Application of Ultrasonic Energy in the Deformation of Metals; Report No. 107; Defense Metals Information Center, Battelle Memorial Institute, Columbus, Ohio; August 1963.

V-20 LANGENECKER, B. : "Work hardening of zinc crystals by high-amplitude ultrasonic waves"; Proceedings of the American Society for Testing and Materials 62 (1962), 602-609.

V-21 NEPPIRAS, E. A. : "Very high energy ultrasonics"; British Journal of Applied Physics 11 (1960), 143-150.

V-22 MASON, W. P. : Physical Acoustics, Volume 1; Academic Press, New York; 1964.

V-23 HALD, A. : Statistical Theory with Engineering Applications; John Wiley & Sons, New York; 1957.

References for Part VI

VI-1 COX, D. R. : Renewal Theory; John Wiley & Sons, New York; 1962.

VI-2 GOTTFRIED, P. and H. R. ROBERTS : "Some pitfalls of the Weibull distribution"; Ninth Nat'l. Symp. Rel. and Q. C., 372-379; 1963.

VI-3 BERMAN, S. M. : "A note on extreme values, competing risks and semi-Markov processes"; Ann. Math. Statist. 34 (1963), 1104-1106.

VI-4 ALLEN, W. R. : "A note on conditional probability of failure when hazards are proportional"; Operations Res. 11 (1963), 658-659.

VI-5 ALLEN, W. R. : "A pitfall in accelerated testing"; Naval Research Logistics Quarterly 10 (1963), 271-273.

VI-6 BERKSON, J. and L. ELVEBACK : "Competing exponential risks, with particular reference to the study of smoking and lung cancer"; J. Amer. Statist. Assoc. 55 (1960), 415-428.

VI-7 LYNCH, W. M., C. B. CLARK, and B. EPSTEIN : Advanced Vehicle Guidance Systems Reliability Investigation; Technical Documentary Report ASD-TDR-62-359; Wright-Patterson Air Force Base, Ohio; 1962.

VI-8 MENDENHALL, W. and R. J. HADER : "Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data"; Biometrika 45 (1958), 504-520.

VI-9 SWAMI, P. S. and S. A. D. C. DOSS : "On a problem of Bartholomew in life testing"; Sankhya Series A 23 (1961), 225-230.

VI-10 GUMBEL, E. J. : "La dissection d'une répartition"; Annales de l'Université de Lyon, troisieme serie--section A. fasc. 2, 1940.

VI-11 RIDER, R. P. : "The method of moments applied to a mixture of two exponential distributions"; Ann. Math. Statist. 32 (1961), 143-147.

VI-12 KAPLAN, E. L. and P. MEIER : "Non-parametric estimation from incomplete observations"; J. Amer. Statist. Assoc. 53 (1958), 457-481.

VI-13 BARLOW, R. and L. HUNTER : "Mathematical models for system reliability"; Sylvania Technologist, vol. 13, no. 2 (April, 1960).

VI-14 PARZEN, E. : Stochastic Processes; Holden-Day, San Francisco, 1962.

VI-15 KOLMOGOROV, A. N. : Grundbegriffe der Wahrscheinlichkeitsrechnung; Springer, Berlin; 1933.

VI-16 FISZ, M. : Probability Theory and Mathematical Statistics; John Wiley & Sons, New York; 1963.

VI-17 STORMER, H. : "Uber einen statistischen Test zur Bestimmung der Parameter einer Lebensdauerverteilung"; Metrika 4 (1961), 63-77.

VI-18 ROBBINS, H. and S. MONRO : "A stochastic approximation method"; Ann. Math. Statist. 22 (1951), 400-407.

Research and Technology Division
Wright-Patterson AFB, Ohio
ACCELERATED LIFE TESTING OF SPACE
GUIDANCE COMPONENTS.

Final Report, 30 September 64, 342 pp incl. illus, tables, refs., Unclass. Report

The report deals with theoretical developments necessary for accelerated testing and with specific hardware considerations. In the theoretical area, the prevalent approach to accelerated testing is described and analyzed, and is found inadequate. A comparative discussion of various failure models is presented; those most amenable to meaningful accelerated testing are developed. Models for devices with multiple failure modes are described, andpertinent estimation procedures are reviewed. A new method for efficient estimation of the failure distribution of repairable equipment is presented. For hardware, the report includes a detailed examination of several space guidance components, and specific applications of accelerated testing to such components are recommended. Metallic creep is examined from the standpoint of its relation to failures of space guidance components, and methods of ac   ated estimation of creep behavior are examined.

1. Reliability
2. Accelerated Testing
3. Space Navigation

I. AFSC Project 3181 Task 318108
II. Contract AF 33(615)-1157
III. Autonetics, a Division of North American Aviation, Inc., Anaheim, California
IV. Winter, B.B., Hietala, H.J., Denison, C.A., Greene, F.W.

Research and Technology Division
Wright-Patterson AFB, Ohio
ACCELERATED LIFE TESTING OF SPACE
GUIDANCE COMPONENTS

Final Report, 30 September 64, 342 pp incl. illus, tables, refs., Unclass. Report

The report deals with theoretical developments necessary for accelerated testing and with specific hardware considerations. In the theoretical area, the prevalent approach to accelerated testing is described and analyzed, and is found inadequate. A comparative discussion of various failure models is presented; those most amenable to meaningful accelerated testing are developed. Models for devices with multiple failure modes are described, and pertinent estimation procedures are reviewed. A new method for efficient estimation of the failure distribution of repairable equipment is presented. For hardware, the report includes a detailed examination of several space guidance components, and specific applications of accelerated testing to such components are recommended. Metallic creep is examined from the standpoint of its relation to failures of space guidance components, and methods of accelerated estimation of creep behavior are examined.

1. Reliability
2. Accelerated Testing
3. Space Navigation

I. AFSC Project 3181 Task 318108
II. Contract AF 33(615)-1157
III. Autonetics, a Division of North American Aviation, Inc., Anaheim, California
IV. Winter, B.B., Hietala, H. J., Denison, C.A., Greene, F.W.

Research and Technology Division
Wright-Patterson AFB, Ohio
ACCELERATED LIFE TESTING OF SPACE
GUIDANCE COMPONENTS.

Final Report, 30 September 64, 342 pp incl. illus, tables, refs., Unclass. Report

The report deals with theoretical developments necessary for accelerated testing and with specific hardware considerations. In the theoretical area, the prevalent approach to accelerated testing is described and analyzed, and is found inadequate. A comparative discussion of various failure models is presented; those most amenable to meaningful accelerated testing are developed. Models for devices with multiple failure modes are described, andpertinent estimation procedures are reviewed. A new method for efficient estimation of the failure distribution of repairable equipment is presented. For hardware, the report includes a detailed examination of several space guidance components, and specific applications of accelerated testing to such components are recommended. Metallic creep is examined from the standpoint of its relation to failures of space guidance components, and methods of accelerated estimation of creep behavior are examined.

1. Reliability
2. Accelerated Testing
3. Space Navigation

I. AFSC Project 3181 Task 318108
II. Contract AF 33(615)-1157
III. Autonetics, a Division of North American Aviation, Inc., Anaheim, California
IV. Winter, B.B., Hietala, H.J., Denison, C.A., Greene, F.W.

Research and Technology Division
Wright-Patterson AFB, Ohio
ACCELERATED LIFE TESTING OF SPACE
GUIDANCE COMPONENTS

Final Report, 30 September 64, 342 pp incl. illus, tables, refs., Unclass. Report

The report deals with theoretical developments necessary for accelerated testing and with specific hardware considerations. In the theoretical area, the prevalent approach to accelerated testing is described and analyzed, and is found inadequate. A comparative discussion of various failure models is presented; those most amenable to meaningful accelerated testing are developed. Models for devices with multiple failure modes are described, and pertinent estimation procedures are reviewed. A new method for efficient estimation of the failure distribution of repairable equipment is presented. For hardware, the report includes a detailed examination of several space guidance components, and specific applications of accelerated testing to such components are recommended. Metallic creep is examined from the standpoint of its relation to failures of space guidance components, and methods of accelerated estimation of creep behavior are examined.

1. Reliability
2. Accelerated Testing
3. Space Navigation

I. AFSC Project 3181 Task 318108
II. Contract AF 33(615)-1157
III. Autonetics, a Division of North American Aviation, Inc., Anaheim, California
IV. Winter, B.B., Hietala, H.J., Denison, C.A., Greene, F.W.

Research and Technology Division
Wright-Patterson AFB, Ohio
ACCELERATED LIFE TESTING OF SPACE
GUIDANCE COMPONENTS.

Final Report, 30 September 64, 342 pp incl. illus, tables, refs., Unclass. Report

The report deals with theoretical developments necessary for accelerated testing and with specific hardware considerations. In the theoretical area, the prevalent approach to accelerated testing is described and analyzed, and is found inadequate. A comparative discussion of various failure models is presented; those most amenable to meaningful accelerated testing are developed. Models for devices with multiple failure modes are described, and pertinent estimation procedures are reviewed. A new method for efficient estimation of the failure distribution of repairable equipment is presented. For hardware, the report includes a detailed examination of several space guidance components, and specific applications of accelerated testing to such components are recommended. Metallic creep is examined from the standpoint of its relation to failures of space guidance components, and methods of accelerated estimation of creep behavior are examined.

1. Reliability
2. Accelerated Testing
3. Space Navigation
I. AFSC Project 3181 Task 318108
II. Contract AF 33(615)-1157
III. Autonetics, a Division of North American Aviation, Inc., Anaheim, California
IV. Winter, B.B., Hietala, H.J., Denison, C.A., Greene, F.W.

Research and Technology Division
Wright-Patterson AFB, Ohio
ACCELERATED LIFE TESTING OF SPACE
GUIDANCE COMPONENTS

Final Report, 30 September 64, 342 pp incl. illus, tables, refs., Unclass. Report

The report deals with theoretical developments necessary for accelerated testing and with specific hardware considerations. In the theoretical area, the prevalent approach to accelerated testing is described and analyzed, and is found inadequate. A comparative discussion of various failure models is presented; those most amenable to meaningful accelerated testing are developed. Models for devices with multiple failure modes are described, and pertinent estimation procedures are reviewed. A new method for efficient estimation of the failure distribution of repairable equipment is presented. For hardware, the report includes a detailed examination of several space guidance components, and specific applications of accelerated testing to such components are recommended. Metallic creep is examined from the standpoint of its relation to failures of space guidance components, and methods of accelerated estimating of creep behavior are examined.

1. Reliability
2. Accelerated Testing
3. Space Navigation
I. AFSC Project 3181 Task 318108
II. Contract AF 33(615)-1157
III. Autonetics, a Division of North American Aviation, Inc., Anaheim, California
IV. Winter, B.B., Hietala, H. J., Denison, C.A., Greene, F.W.

Research and Technology Division
Wright-Patterson AFB, Ohio
ACCELERATED LIFE TESTING OF SPACE
GUIDANCE COMPONENTS.

Final Report, 30 September 64, 342 pp incl. illus, tables, refs., Unclass. Report

The report deals with theoretical developments necessary for accelerated testing and with specific hardware considerations. In the theoretical area, the prevalent approach to accelerated testing is described and analyzed, and is found inadequate. A comparative discussion of various failure models is presented; those most amenable to meaningful accelerated testing are developed. Models for devices with multiple failure modes are described, andpertinent estimation procedures are reviewed. A new method for efficient estimation of the failure distribution of repairable equipment is presented. For hardware, the report includes a detailed examination of several space guidance components, and specific applications of accelerated testing to such components are recommended. Metallic creep is examined from the standpoint of its relation to failures of space guidance components, and methods of accelerated estimation of creep behavior are examined.

1. Reliability
2. Accelerated Testing
3. Space Navigation

I. AFSC Project 3181 Task 318108
II. Contract AF 33(615)-1157
III. Autonetics, a Division of North American Aviation, Inc., Anaheim, California
IV. Winter, B.B., Hietala, H.J., Denison, C.A., Greene, F.W.

Research and Technology Division
Wright-Patterson AFB, Ohio
ACCELERATED LIFE TESTING OF SPACE
GUIDANCE COMPONENTS

Final Report, 30 September 64, 342 pp incl. illus, tables, refs., Unclass. Report

The report deals with theoretical developments necessary for accelerated testing and with specific hardware considerations. In the theoretical area, the prevalent approach to accelerated testing is described and analyzed, and is found inadequate. A comparative discussion of various failure models is presented; those most amenable to meaningful accelerated testing are developed. Models for devices with multiple failure modes are described, and pertinent estimation procedures are reviewed. A new method for efficient estimation of the failure distribution of repairable equipment is presented. For hardware, the report includes a detailed examination of several space guidance components, and specific applications of accelerated testing to such components are recommended. Metallic creep is examined from the standpoint of its relation to failures of space guidance components, and methods of accelerated estimation of creep behavior are examined.

1. Reliability
2. Accelerated Testing
3. Space Navigation

I. AFSC Project 3181 Task 318108
II. Contract AF 33(615)-1157
III. Autonetics, a Division of North American Aviation, Inc., Anaheim, California
IV. Winter, B.B., Hietala, H. J., Denison, C.A., Greene, F.W.

UNCLASSIFIED

UNCLASSIFIED