General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.

- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.

- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.

- This document is paginated as submitted by the original source.

- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Produced by the NASA Center for Aerospace Information (CASI)

FINAL REPORT

on

METHODOLOGY FOR DESIGNING ACCELERATED AGING TESTS FOR PREDICTING LIFE OF PHOTOVOLTAIC ARRAYS

to

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY

for the

ENCAPSULATION TASK OF THE
LOW-COST SILICON SOLAR ARRAY PROJECT

February 1, 1977

AUG 1977
RECEIVED
NASA STI FACILITY
INPUT BRANCH

G. B. Gaines, R. E. Thomas, G. C. Derringer
C. W. Kistler, D. M. Bigg, and D. C. Carmichael

BATTELLE
Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

FINAL REPORT

on

METHODOLOGY FOR DESIGNING ACCELERATED AGING TESTS FOR PREDICTING LIFE OF PHOTOVOLTAIC ARRAYS

to

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY

for the

ENCAPSULATION TASK OF THE
LOW-COST SILICON SOLAR ARRAY PROJECT
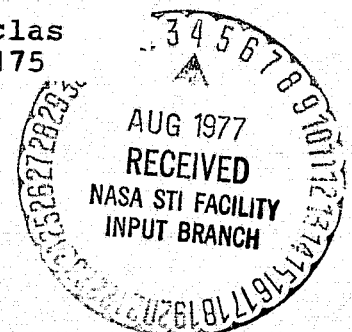
February 1, 1977

G. B. Gaines, R. E. Thomas, G. C. Derringer
C. W. Kistler, D. M. Bigg, and D. C. Carmichael

BATTELLE
Columbus Laboratories
505 King Avenue
Columbus, Ohio 43201

METHODOLOGY FOR DESIGNING ACCELERATED AGING TESTS
FOR PREDICTING LIFE OF PHOTOVOLTAIC ARRAYS

ABSTRACT

Current analyses dictate that for economic viability photovoltaic arrays should have a service life of 20 years or more. A need exists for meaningful accelerated tests to predict array life so that a relatively quick assessment can be made of the potential of present developmental designs and of the extent to which improvements in life are needed. As a part of the Encapsulation Task of the LSSA Project, this study undertook to develop a methodology for designing aging tests in which life prediction was paramount.

The methodology developed builds upon past experience with regard to aging behavior in those material classes which are expected to be utilized as encapsulant elements, viz., glasses and polymers, and upon past experience with the design of aging tests. Accordingly, these experiences were reviewed; the results constitute Part I of the technical discussion.

Part II presents the improved methodology developed in this study. The implementation of the methodology is illustrated using an example design for a solar–cell module. The developed methodology emphasizes the importance of incorporating substantial contributions at the time of initiation of the test design from statisticians, materials scientists, and test engineers in order to achieve a test design that is both statistically satisfactory and is practical in terms of the number of tests to be run.

The first six steps of the developed methodology focus on the explicit identification of necessary engineering input information, identification of possible failure modes and environmental variables (stresses) that may affect the time rates of degradation for each failure mode without changing the failure mode, estimation of expected overall severity of each combination of environmental stresses, and analysis of severity ratings as a complete factorial experiment, with the results graphically represented by a hierarchical tree. An examination of the tree makes it possible to identify those test conditions (combinations of environmental stress levels) that are expected to produce the largest changes in degradation rates. The less important tests are eliminated from further consideration by pruning the tree.

Because the remaining tests may not form a suitable experimental design, the methodology provides for the inclusion of selected additional tests to remove as many statistical deficiencies as possible within the allowable time/cost constraints. The first iteration test design that results from this procedure consists of an explicit identification of what tests should be run and how many specimens should be tested at each test condition. The methodology then requires the consideration of alternative test designs that include up to five stress levels for each kind of environmental stress. The final test design is obtained by making parametric trade–off studies among the alternative designs.

Considerations of precision, accuracy, and test sensitivity are also included in the report. It is recommended that empirical, statistical, and conceptual methods be used to analyze the data resulting from the implemented accelerated test. It is also recommended that, during the test program, predictions be made both within and between stress conditions in order to generate a "track record" for predicting degradation at lower environmental stress conditions using data obtained from higher stress conditions.

TABLE OF CONTENTS

PRECEDING PAGE BLANK NOT FILMED

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF TABLES
(Continued)

LIST OF FIGURES

Page

METHODOLOGY FOR DESIGNING ACCELERATED AGING TESTS FOR PREDICTING LIFE OF PHOTOVOLTAIC ARRAYS

SUMMARY

This study was conducted in support of the Encapsulation Task of the Low-Cost Silicon Solar Array (LSSA) Project which is managed by JPL for ERDA-Division of Solar Energy and is part of ERDA's Solar Photovoltaic Conversion Program. The 1986 goals of the LSSA Project are to develop silicon photovoltaic arrays that:

- Are priced at less than $500/kW (peak)

- Are producible in quantities greater than 500,000 kW/yr

- Have lifetimes greater than 20 years

- Have conversion efficiencies greater than 10 percent.

These goals require that long-life, low-cost photovoltaic arrays be developed over a relatively short period of time. Accordingly, a specific requirement of the project is to develop accelerated tests that can predict service life in a relatively short time.

Three other related studies on low-cost encapsulation systems to protect arrays for 20 years in terrestrial service environments are being conducted at Battelle's Columbus Laboratories. The scope of each is described briefly in the Introduction of this report.

Objectives of This Study

The objectives of this study were:

- To review and assess available literature and data regarding environmental degradation of glasses and polymers, aging-test methodology and techniques, and data-analysis methods, where prediction of life has been a major goal

- To develop a practical and credible methodology for designing abbreviated/accelerated aging tests for evaluating photovoltaic arrays and components intended for a 20-year service life in a terrestrial environment

- To illustrate the developed methodology by the design of a test program for the materials and interfaces of an example array encapsulation system.

Summary Definition of the Problem

Silicon solar cells, if adequately protected from the stresses of the terrestrial environment, can be expected with reasonable confidence to meet the LSSA 20-year conversion-efficiency and

service-life goals. The success of the LSSA Project, then, depends substantially upon the development of a low-cost array encapsulation system that will provide the requisite protection. The applicability of a candidate glass or polymeric material in the encapsulation system thus depends upon its ability to withstand 20 years of outdoor service without degrading seriously enough to result in an unacceptable level of deterioration in the electrical output/efficiency of the photovoltaic array. Since 20 years of test time is obviously impractical, credible short-term (abbreviated/accelerated) predictive aging tests must be employed in the encapsulation-system and array development efforts. At the outset of this study, a suitable methodology for conducting such tests did not exist.

Numerous accelerated and natural aging studies of glasses and polymers which had been carried out in the past, however, offered a basis for developing the methodology. Extensive analyses of the past experience revealed that several methodologies involving an assumed mathematical model based upon relevant physical properties had been used with varying degrees of success in the design of accelerated aging tests of single materials. A photovoltaic array, however, whatever its final design and configuration, has several component materials and several interfaces and must operate under widely different, interacting environmental stresses. Consequently, the test methodology cannot be based upon a single assumed aging process and the test design must be relatively comprehensive. The problem then becomes one of devising a suitable test methodology that can be carried out within practical constraints of time and cost.

The past experience also reveals that although many conventional accelerated aging tests, such as those based on full factorial designs, are started; few are completed either because of time, cost, interest, or other constraints not envisioned at the beginning of the test. This emphasizes that a suitable test methodology must allow for the exercise and verification of engineering judgments in order to reduce to a practical level the number of tests needed to qualify materials and components. In this regard, the methodology must also incorporate appropriate iterative sequences to insure the statistical quality of the data being developed in a particular test series.

It should also be noted that the methodology development problem is further complicated by the absence of information regarding the ultimate design of LSSA arrays or modules [*] and their method of manufacture. This has required the development of test-design procedures which are very broad in their applicability to various module materials, designs, and possible failure modes.

Summary of the Approach Used

A methodology was developed offering improved statistical design but which also relied heavily on *a priori* engineering input so as to permit the planning of acceptable test designs which are *practical* in terms of number of tests. Because of this reliance, a portion of the study was directed toward critically assessing and drawing from the past experience in aging behavior of glasses and polymers, mathematical models of aging behavior, and statistical treatments of experimental test designs. These reviews and analyses constituted Part I of the study. An improved methodology for desiging predictive aging tests was then developed and was applied to an example module to illustrate the procedure for test design. The methodology development constituted Part II of the

---

(*) The terms "module" and "array" are often used interchangeably. The term
  "module" has been used in this report and signifies a group of cells that are
  formed into one unit or structure by the encapsulation system.

2

study. Summary statements regarding the specific approaches used in each part of the study are given in the following paragraphs.

During Part I (detailed in Part I of the Technical Discussion of this report), available information concerning past aging test programs and materials aging research was collected. Identification of the appropriate published literature was aided by extensive computerized searching of major data banks and governmental information sources dealing with selected subject matter. Articles and documents identified as possible sources of relevant information were obtained and then reviewed by researchers specializing in the areas of the subject matter. Additional information was elicited by personal communications.

This information was critiqued to define and assess the state-of-the-art of predictive testing of glass and polymeric materials. Materials scientists and statisticians studied the collected information with the intent of defining the degradation mechanisms, rates, and effects that have been observed in glasses and polymers subjected to environmental stresses, the laboratory and outdoor test techniques that have been employed, the statistical procedures used, and the capabilities of available instrumentation. In addition to its value during formulation of test procedures, the information developed was to aid in making the engineering judgments that must be exercised to secure reasonably sized test programs. The analysis of the collected information by the statistical specialists was particularly concerned with assessing the applicability of reported aging test designs and property-versus-time models to the LSSA module test-methodology design problem.

The second part of the study (Part II of the Technical Discussion of the report) was directed to the development of an improved methodology for predictive testing of photovoltaic modules, materials, and components. Methodology development was approached with the goals of (1) securing a test acceleration factor as large as possible, (2) assuring ready applicability to the various test situations that will be encountered, and (3) assuring a maximum utilization of previous model design and data evaluation work. Much emphasis was placed upon development of procedures for quantifying and ranking engineering inputs, for the practical purpose described above of limiting the size of the test program, and upon extending the applicability of extrapolation procedures described in the literature. An example of a test design for a candidate module design and materials was generated to illustrate the methodology developed.

Summary of Results of the Review and Analysis
of Past Work on Aging and Accelerated Testing

On the basis of the optical requirements of encapsulated photovoltaic modules and the cost constraints imposed by the LSSA goals, it is anticipated that the primary constituents of the encapsulating system for these modules will be glasses and/or polymeric materials. The aging behavior of these types of materials, both singly and in various interfacial combinations, is the principal concern of this study. In Part I, the current state of knowledge on the aging behavior of glasses and polymers and the state of the art in accelerated test design methodology, prior to the original work contained in the second part of this report, are reviewed and analyzed from the point of view of the requirements of the LSSA program.

Aging Behavior of Glasses

When compared to most other materials, glasses are highly resistant to outdoor weathering. The major areas of vulnerability are attack by moisture, particularly of a strong acidic or alkaline nature, and damage by thermal or mechanical stresses. The degradation modes of most serious concern for photovoltaic applications are losses of strength and optical transmission.

Acidic attack is a leaching process involving an exchange of sodium and hydrogen ions, which results in an altered surface region. Alkaline attack involves an etching process which breaks up the silica structure, weakening the glass and degrading its optical properties. The alkaline attack rate is generally several orders of magnitude higher than the acidic attack rate; however, severe alkaline conditions are seldom encountered in normal weathering situations except in cases where strong detergents are used. Dirt accumulation can be a serious problem both as a contributing factor in chemical attack and by direct effects on optical transmission. These factors are generally highly location specific.

Extensive quantitative data on the aging of glasses under natural weathering conditions are not available in the published literature. This is due to the high weathering resistance (low rate of change) of these materials and to the acceptance of pre-use chemical durability tests as satisfactory forecasters of weatherability. A body of data on artificial weathering tests on glasses exists; however, much of this information is either not applicable to weathering of modern glasses suitable for photovoltaic module applications under normal conditions, or is not sufficiently detailed to be of value. There is a serious need for new and improved techniques for characterizing surface degradation in glasses.

In general, it is believed that failures associated with interfaces between glass and the other constituents (organics) of the encapsulation system are more likely to limit the lifetime and reliability of the module than failures associated with bulk degradation of the glass.

Aging Behavior of Polymers

The major factors affecting the service life of polymeric materials are ultraviolet (UV) radiation, oxygen, moisture, temperature, chemical pollutants (SO_2, NO_2, CO, NaCl), dirt accumulation, and abrasion. UV radiation can alter bond structures and activate chemicals which will degrade these materials. Oxidation, which can be accelerated by UV radiation, generally results in discoloration and embrittlement. Moisture tends to leach out stabilizing chemicals in polymeric materials and degrade mechanical strength, while thermal gradients and temperature fluctuations contribute to delamination at interfaces between differing materials. Chemical pollutant effects are not as well characterized as those of some of the other factors and are sometimes difficult to separate from them. In general, there is considerable interaction among degradation mechanisms in polymers, and aging characteristics are highly structurally and compositionally dependent. Nonchemical degradation (dirt accumulation, abrasion, etc.) has received limited attention to date.

Existing techniques for measuring degradation of the properties of polymers are more advanced than in the case of glasses, although still in need of improvement. The usefulness of existing data from natural weathering tests on polymers to the LSSA application is limited by the fact that many of the weather parameters contributing to degradation have either not been recorded or have been inadequately specified. Artificial weathering tests have exhibited a number of deficiences in simulating the natural environment. These deficiencies have hampered correlations of the results of natural and artificial weathering conditions. Some test results have been confounded by

manufacturers' unannounced alterations of material structures. Generally, studies of polymers in outdoor weathering experiments have been hampered by techniques limitations, and also by deficiencies in the experimental design of past aging tests.

Aging Tests and Models

Practically all of the systematic aging tests in which life prediction has been an important goal have been done with single materials, not laminates nor structures having interfaces with other materials. Mathematical model building, describing degradation as a function of time and environmental factors, has been a feature of several important studies. Unfortunately, the success has been highly mixed; many tests have been aborted for many reasons. Nevertheless, some significant information that can aid in test development and design has come out of this experience. The experience suggests that tests should not be designed or interpreted based on a single mechanistic model of failure; all feasible models must be considered.

Statistical Approach to Accelerated Testing

Detailed statistical investigations have been carried out by several researchers for selected models related to the Weibull distribution and the Arrhenius and Eyring relations. An overall methodology for developing an optimized accelerated test program for several stresses was not found in the literature. Most investigations have restricted attention to the case of a single stress. This restriction often precludes extrapolation to "normal stress" conditions, expecially for photovoltaic modules that are exposed to a variety of stresses associated with temperature cycling, relative humidity, ultraviolet radiation, and pollutants such as sulfur dioxide. Because accelerated testing involves extrapolations from overstress to normal stress conditions, the optimal extrapolation procedures of Hoel and Levine appear to be especially relevant to the design of accelerated tests. In the present study, this procedure was extended to the case of several simultaneous stresses.

Summary of the Work to Develop Improved
Accelerated-Test Methodology

In Part II of the study, an improved methodology for the design of accelerated tests was developed and the method was applied to an example photovoltaic module to demonstrate the recommended procedures. The developed approach utilizes a team of statisticians, materials scientists, and test engineers to produce a test design that has statistical validity but minimizes the number of tests to be run in order to achieve a practical test plan.

Accelerated-Test Design

In the developed methodology, the initial step in the preparation of the accelerated-test plan is an engineering examination of the module design and fabrication materials in order to identify the expected failure modes under normal stress. A complete factorial experimental design is then conceptually analyzed to estimate the relative severity of the possible test conditions, including high- and low-stress conditions for stresses associated with appropriate variables. (There are 16 possible test conditions in the complete factorial design in the example developed in the report.) The severity of each test combination is estimated in a systematic manner to account for the

separate effects of important interactions of the stresses. The severity ratings are then analyzed using Yates' method to compute the expected main effects and interactions of all test variables. Successive Yates' analyses are used to obtain a graphical representation of the expected conditional main effects. This representation, a hierarchical tree, is employed to reduce the number of tests (from 16 to 5 in the example), as represented by terminal cells of a pruned tree. As a preliminary to applying the extrapolation procedure of Hoel and Levine, the pruned tree is represented algebraically by means of Lagrangian polynomials. Numerical scales for variables are constructed to obtain coordinates for the extrapolation point associated with normal operating conditions at various geographic locations. Finally, a generalized version of Hoel and Levine's extrapolation procedure is used to identify the number of test modules that should be employed in each of the remaining tests to maximize the percision of the extrapolations to normal stress conditions.

The methodology then calls for considerations of various alternative test designs. The ideal number of stress levels for a given stress is taken to be five. The minimum number of modules to be tested at any given stress combination is also taken to be five. The spacing between stress levels is taken to correspond to the Chebychev points in the interval -1 to 1. Statistical assessments related to the underlying model, main effects, interactions, confounding, etc. are also included to determine whether supplementary test conditions are justified. These quantitative studies serve as a basis for parameteric studies which, in turn, lead to the selection of the final test design.

Design Instrumentation

Once the basic experimental design is selected, the properties to be measured for a particular module design are established. Test methods for property determinations that give the required sensitivity and precision are emphasized. The methodology provides procedures to rate test methods relative to cost, as well as precision. Since appropriate property measurements cannot be made on tested modules without destruction in some cases, individual materials and module subsystems can be placed in the environmental testing chamber to allow for these property-change measurements.

Data Analysis

Several approaches to data analysis need to be applied in analyzing property-vs.-time data. It is recommended in the methodology developed that the conceptual approach (models derived from physical laws), the statistical approach (separation of signal from "noise" arising from variability in measurement procedures, lack of control, etc.), and the empirical approach (different lots, burn-in, sample selection, etc.) should all be incorporated in the data analysis. Application of physical laws to arrive at the predictive model is most appealing but often cannot be done alone with sufficient assurance and particularly with complex phenomena. Several mathematical models are examined to arrive at the one most suitable. Procedures are provided in the methodology to aid discrimination among models.

A summary of the steps recommended in the developed methodology for arriving at the final accelerated-test design is given in Table 23 on page 119. Included in the tabulation of the recommended steps is a synopsis of the application of each step of the methodology to the example module (Figure 3). References to the portion of the report describing each of the steps in detail are also cited in Table 23.

6

Summarized Implications of the Results

The methodology for designing accelerated aging tests for predicting the life of photovoltaic modules that was developed during the study is believed, both conceptually and in actual state of development, to represent a significant advance in the state of the art. It extends, and makes very practical use of, prior research. The procedures developed and illustrated constitute an innovative method of designing manageable tests involving several types of stresses. It is recommended that the developed methodology be experimentally implemented to test innovative aspects such as the extrapolation procedure and to further detail the test method.

Owing to the complexity of the photovoltaic-module aging test requirements, the implementation of the methodology to design a test plan is not a routine task. It is recommended that the procedure should involve substantial contributions at the time of initiation of the test design from statisticians, materials scientists, and test engineers.

Although the developed procedures provide a significantly improved test design method, all aspects of accelerated test design are not yet optimized. Future work directed to the following three areas could further improve the design of predictive aging tests. First, additional effort should be conducted on the scaling factors relating the environmental stresses used in accelerated tests to the conventional climatic parameters. To do this, the principal elements of a module (e.g., interfaces between dissimilar materials, optically transmitting top covers, etc.) should be studied separately with respect to probable failure modes and appropriate scaling factors. Scaling factors and tests applicable to a wide range of module designs should result from the availability of such information. Second, it is recommended that the extrapolation procedure should be further improved to account more properly for any change in precision of measurement that may occur with a change in stress level. Third, the methodology delineates the influence of precision and costs associated with the instrumentation selected for measuring property changes, but optimization of these considerations should be explicitly included in the design. Moreover, more attention than was possible in this study needs to be given to the selection of the instrumentation, particularly in regard to what instrumentation is appropriate to measure property changes associated with failure modes. A study should be conducted, therefore, which will (a) identify direct and indirect measurements which might serve as precursors of failure modes, and (b) determine what instrumentation for these measurements is available now or can be made available with reasonable effort. Sensitivity, precision, cost, and applicability to in situ measurements are among the characteristics of the instrumentation that need to be evaluated.

INTRODUCTION

This report presents the final results of a study conducted in support of the Low-Cost Silicon Solar Array (LSSA) Project[1], sponsored by the Energy Research and Development Administration (ERDA), Division of Solar Energy, and managed by the Jet Propulsion Laboratory (JPL). The 1985 objectives of the LSSA Project are to develop the technology and manufacturing capability to produce 500,000 kW/year of photovoltaic arrays at a cost of less than $500/kW, with an efficiency of greater than 10 percent and a service life of 20 years. The Encapsulation Task of this project is concerned with the development of the encapsulation systems for terrestrial photovoltaic arrays. Within this task, four interrelated studies were assigned to Battelle's Columbus Laboratories:

Study 1: Review of World Experience and Properties of Materials for Encapsulation of Terrestrial Solar-Cell Arrays. Available data defining the state of the art of encapsulation-system materials and processes were collected and analyzed to provide a credible basis for defining materials evaluation and development efforts for the Encapsulation Task.

Study 2: Definition of Terrestrial Service Environments for Encapsulation Materials. Environmental conditions to which a terrestrial solar array will be exposed over a 20-year lifetime were characterized to aid definition of realistic test programs for encapsulation system materials.

Study 3: Evaluation of Encapsulation Designs and Materials for Low-Cost Long-Life Silicon Photovoltaic Arrays. The properties and aging behavior of candidate encapsulation materials and the performance of encapsulated photovoltaic cells before and after aging in various environments are being evaluated.

Study 4: Development of Methodology for Accelerated Aging Tests for Predicting Life of Photovoltaic Arrays. Available data from past aging and accelerated environmental testing of glass and polymeric materials, as well as test designs and models, were reviewed, and a methodology for predictive accelerated testing of photovoltaic arrays, materials, and components was developed.

This report presents the final results on Study 4. Separate reports have been issued on Studies 1 and 2 and preparation of the report on Study 3 is in progress.[2, 3, 4]

OBJECTIVES

The specific objectives of this study were:

(1) To review and assess present knowledge and experience regarding degradation of glasses and polymers, aging-test methodology and techniques, and data-analysis methods where prediction of life has been a major goal

(2) To develop an improved methodology for designing aging tests for photovoltaic modules

(3) To give an example of the application of this methodology to specific materials and/or interfaces in candidate encapsulation systems.

The study emphasis was on photovoltaic modules, not the total conversion system. Even more specifically, the emphasis was on the module encapsulation system, although the interactions between the encapsulant and the cells had to be taken into account. The methodology describes the important factors that must be considered in the test design, including suitable statistical and experimental techniques for treating data analysis and predictive methods. The test design includes specimen design and number, environmental parameters and operating conditions, properties to be measured, instrumentation, and data analysis methods.

For reference purposes, it is to be noted that a portion of the objectives had to be changed near the beginning of the program. Originally, an improved methodology was to be developed and applied to existing aging data on a selected material to provide limited validation of the methodology. However, because appropriate data were not found, the improved methodology developed has been illustrated for a model photovoltaic module in which the materials are specified.

BACKGROUND AND NEED FOR STUDY

At the current stage of assessment of the economics of photovoltaic modules, it appears that a module lifetime of 20 years or more is required for photovoltaic conversion to be viable. Experience with photovoltaic modules in terrestrial environments with documented conditions and performance is very limited to date.[2] Experience from the microelectronics area with p-n junction devices of a similar nature, however, suggests that such a life for a solar cell might be realized, provided the cell can be protected from the influence of terrestrial atmospheres by appropriate encapsulation. There is substantial experience with encapsulation of devices in the microelectronics industry to be drawn upon. On the other hand, encapsulation of solar cells for the environments to be reckoned with adds some dimensions not found in the experience to date. The first such dimension is allowed costs, both dollar costs and energy costs. Encapsulation costs for microelectronic packages must be constrained to be sure, but not to the extent of encapsulation costs for photovoltaic modules, either in materials or processing considerations.

Another added dimension is the required lifetime. Substantial experience with electronic devices over a period of 20 years does not exist. Evolution of new and improved devices, along with the relative infancy of the microelectronics industry, has generally precluded such experience.

A third added dimension is the outside terrestrial environments that solar cells are expected to face. Encapsulation microelectronic packages have been exposed generally to somewhat controlled environments. Hail, rain, large temperature excursions, and high insolation have not been a part of the environment.

Thus, despite substantial experience with encapsulation of electronic devices, these added dimensions will force the photovoltaic conversion industry in large measure to start anew. In this situation, several critical questions arise. One of these is: *"How can one determine if he has made*

appropriate choices of design, materials, and processing for the array to meet the lifetime require-ment of 20 years?". Clearly, accelerated testing is called for. Testing for 20 years prior to install-ment of photovoltaic devices in the field is certainly an unacceptable alternative. Essentially then, the subject of this report is to answer the above question.

Aside from the choices of design, material, etc. per se, there is the question of the method-ology of accelerated testing itself. The goal of such testing is to expose the module to a known overstressed environment for a short period compared to its lifetime, measure selected properties, and from appropriate analysis predict with acceptable confidence limits the lifetime under normal conditions (normal stress levels). That is, one desires a "true" accelerating factor. Unfortunately, it seems that a "true" accelerating factor is known for very few, if any, devices.

The reasons for this situation are not surprising. In most cases, obtaining a true acceleration factor is time consuming and expensive, especially for long service lives. Although "accelerated testing" is as old as device development itself, the goal most often sought is to find a failure mode through the use of overstressed environments and then to "fix" the "weak link" and, hopefully, lengthen the service life under normal stresses. Testing and analyses are not generally expanded to the extent that service life can be predicted within acceptable confidence limits.

So, given that the photovoltaic module is in a large sense a new "device", that a long service life is required, and that a complete methodology for aging tests does not exist, there is a special need to develop a design methodology for such tests. It will be recognized that these develop-ments constitute an ambitious and challenging undertaking. The "ultimate" method might never be found; this program, however, attempts to extend substantially the methodology for designing tests in an acceptable and systematic manner.

ORGANIZATION OF THE REMAINDER OF THE REPORT

The main body of this report, which follows the next section on the approach to the study and to accelerated testing in particular, is presented in two major parts. Part I is devoted to a review and analysis of the experience to date in several selected areas of materials aging and testing. Part II treats the improved accelerated test–design methodology developed in this program.

To draw as much as possible on the experience in accelerated testing, the following subjects were reviewed: aging behavior and testing techniques of polymers and glasses in various environ-ments, major aging studies where life prediction was a major goal, aging models used to describe behavior of selected materials, and statistical approaches to accelerated testing, with the princi-pal emphasis on the number of stress levels and number of modules to be tested at each level. These subjects are discussed in Part I.

Part II describes the accelerated test methodology developed in this study. It extends the state of the art in several ways. To make its development clearer, the improved methodology is applied to a sample module, which is used to illustrate evolution of the methodology. Included are the basic elements of the experimental design, instrumentation of the design, and data-analysis procedures.

The report ends with a summary tabulation (Table 23) of the recommended steps in the developed test–design methodology and with conclusions and recommendations.

PROGRAM APPROACH

ACQUISITION AND REVIEW OF INFORMATION

For Part I of the study, published information on the results of aging studies of glasses and polymers, on aging test models, and on statistical approaches to accelerated testing was collected by computerized and manual searching of the technical literature, bibliographic documents, and governmental publications. To efficiently search the very large volume of literature of potential interest to this study, the relevant published material was identified to the extent possible by inter-rogating the following computer-accessible data bases in various organizations:

CHEMCON	INSPEC
	(Science Abstracts)
CIRC	
(Air Force)	NASA
DDC	NTIS
Engineering Index	PLASTEC
ERDA RECON (File 1)	Reliability Analysis Center
	(RADC)
ERDA RECON (File 9)	SSIE
ERDA RECON (File 10)	(Research in progress)

Unpublished information was obtained from company literature and by private communications. Battelle researchers specializing in the various applicable technical areas helped define the search approach (which was continuously expanded). These same researchers reviewed and evaluated the information with regard to the objectives of this study. The results of this analysis constitute Part I of the technical discussion of the report.

THE APPROACH TO ACCELERATED TESTING AND SELECTED DEFINITIONS

As preparation for the main text of this report, the purpose of this section is to outline the important features of an accelerated test, describe the recommended general approach to accel-erated testing, and state some definitions of terms used in discussing the methodology developed in Part II of this work.

Distinction Among Accelerated Life Tests, Qualification Tests, and Acceptance Tests

The design objectives of an accelerated life test are difficult to meet for a variety of practical reasons. The more important difficulties include the following. Accelerated tests are most de-sirable for newly designed systems for which very little data exist even under laboratory conditions.

This results from the fact that the first kind of testing performed in a newly fabricated system usually consists of performance testing (does the system operate as desired under controlled laboratory conditions?) as mentioned previously. After a number of "fixes" have been made to improve performance, the next state frequently consists of attempting to make the system more robust in order to withstand more abusive operating conditions, uncontrolled environmental (field) conditions, poor maintenance and repair procedures, etc. The basic approach at this second stage consists of operating the system under highly abusive stresses for short durations in order to reveal design weaknesses. Again, fixes are made, or relative comparisons are made among competing designs, and a suitable system design is identified. These kinds of tests are frequently labeled as *qualification or acceptance tests.* In general, by passing such tests successfully, it may be claimed that the tested system operates as intended, during, or after being subjected to, a wide range of abusive conditions that have been deliberately selected to reveal system weaknesses. If these conditions are more abusive than normal operating conditions, it is reasonable to expect that such a system would operate successfully in a normal field environment at the time it is installed. That is, based on these tests, successful operation of the system is expected at time t = 0.

It is not known from such tests, however, how long the system will continue to operate successfully in the field. It is this distinction that sets accelerated life testing apart from qualification and acceptance testing. Given that a system operates successfully at t = 0, an *accelerated life test* is directly concerned with generating a valid prediction for how long the system will continue to operate successfully in a specified environment. In this sense, accelerated life testing is seen to be related to the reliability of the system.

It should also be noted that this terminology is not universally accepted. Some tests are referred to as accelerated tests simply because a system is operated under stress conditions believed to be higher than normal. Thus, qualification and acceptance tests are sometimes called accelerated tests. It seems preferable, however, to refer to such tests as over–stress tests, and reserve the label "accelerated test" for those tests that are directly concerned with predicting how long a system will operate under normal conditions using test data obtained at over–stress conditions. This is the definition used in this report.

A Prototype Accelerated Test

A general theoretical basis for the design of an accelerated life test does not exist. Several partial or total concepts have evolved which correspond to particular needs and testing philosophies. A concept is outlined below which corresponds to experiences of the authors and which will serve to identify and introduce the important features of accelerated tests to be treated in this report, that is, the general approach to accelerated testing recommended in this report. In this statement of the concept, intuitive definitions are used first; more specific definitions are then given to describe a more general concept of the methodology.

Experimental Approach for Accelerated Testing

The experimental approach for the prototype accelerated test is:

- For a given module design and specific operating conditions, replicate modules are placed on test at several higher–than–normal stress levels.

- Nondestructive measurements of module quality are made for each module at specified times.

- At selected measurement times, one or more modules are removed from the test and subjected to complete diagnostic tear-down (destructive) analyses.

- The nondestructive and destructive measurements are made in order to identify all observable changes in material properties and/or module performance characteristics from initial base-line measurements made on each module. The results are expressed as time-rates-of-change which, in turn, are used to predict the material properties and performance characteristics of each module at the next scheduled measurement time.

- The most stressful test is chosen to maximally increase the time rate of degradation of the observed failure mechanism for the module, consistent with the requirement that the failure mechanism observed at the high-stress condition is identical to the dominant failure mechanism that holds under normal conditions.

Data Analysis for Accelerated Testing

The data obtained from a prototype accelerated life test are analyzed as outlined by the major steps and schematic diagrams given below:

- Measure the degradation rate for each quality $x_1(t)$ associated with observed changes in material properties and performance characteristics for generalized stress levels $S_0, S_1, \ldots S_n$, where S_0 denotes normal stress, and S_1 through S_n denote successively higher-than-normal stress levels. The results are represented as plots with degradation rates measured as a function of time. A separate plot is obtained for each measured quality (property).
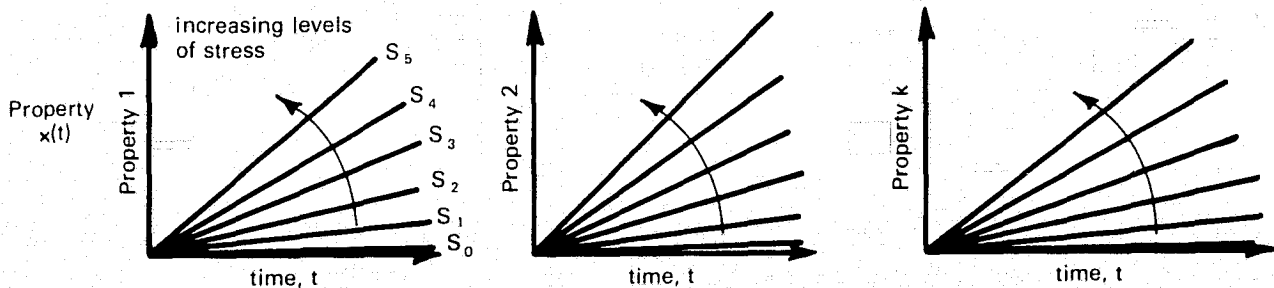
- Relate degradation rates to measures of the environmental and/or operational stresses. An Arrhenius-type plot showing a linear relation between $\log \overset{o}{x}$ and reciprocal stress is often a candidate relationship, especially when the stress is temperature.

13

DEGRADATION RATE of the Material or Performance Property $\dot{x}(t)$

Property 1
Property 2
Property k
1/Stress   Normal stress
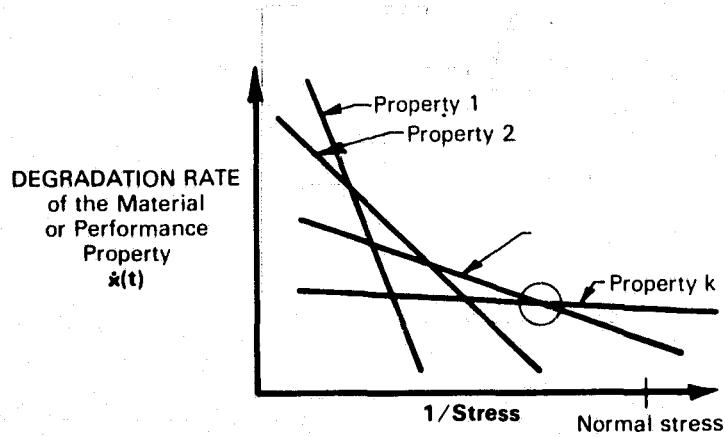
- **Predict** amount of degradation of each quality expected to be observed at the next measurement time. Predictions are made both within a given stress level, and from higher stress levels to lower stress levels.

Within Stress Level:

Property $x(t)$

S_5

Predicted Amount

Predicted time

time, t

Between Stress Levels:

Property $x(t)$

S_5

S_o

Predicted

t_1 t_2

time, t
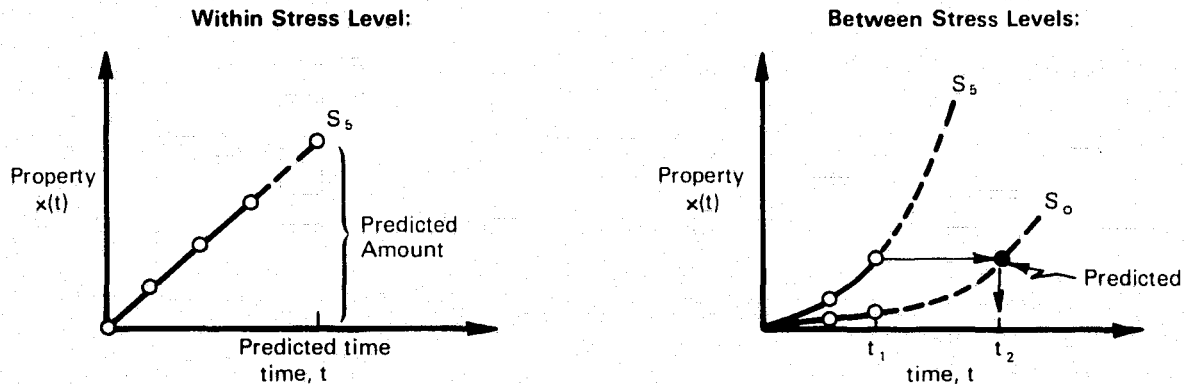
- **Verify** predicted degradation amounts using data obtained from subsequent measurement times. Where predictions are not verified, reassess the data obtained to date and revise and hopefully improve the procedure used for predictive extrapolation. The measure, relate, predict, and verify scheme outlined above provides a feedback loop for updating and improving predictive capabilities. It also provides a documentation procedure for establishing a "track record" for making valid predictions.

Generalized Stress

The term "stress" as used in this report is intended to denote a general measure of the severity of the combined environmental and operational conditions experienced by a solar-cell module. Such a generalized stress, of course, is not restricted to the concept of stress as associated with mechanics. As the environmental or operational conditions change over time, the stress is viewed as changing in a corresponding manner. Specifically, it is required in the definitions that increased stress must always be associated with increased rate of degradation of some failure mechanism. Roughly, increased stress levels must yield more rapid rates of degradation, and hence, shorter lifetimes.

14

As described in this way, the concept of stress is essentially an "undefined term". To be useful for accelerated test design, it must be possible to identify specific stresses and their associated degradation processes. This must be done for each module design, installation design, operating condition, and geographic location or environmental condition. That is, to be useful, each generalized stress must ultimately be reduced to one or more specific measurements.

Degradation Rates

In this concept of accelerated testing, primary emphasis is placed on the accurate and precise determination of degradation rates of quality (property) measures, especially for the degradation rate of the quality measure associated with the failure mechanism believed to be dominant under normal stress conditions. That is, degradation of quality over time is taken to be the key output measure for an accelerated test. In general, the degradation rates obtained from the quality measures at successive measurement times are used to generate the extrapolated predictions of quality at the next measurement time, and more importantly, for a time period of 20 years under normal stress conditions.

Under these assumptions, it is clear that in order to have acceptable high quality at the end of 20 years, ideal modules would be expected to have (1) high initial quality and (2) low degradation rates for quality under normal stresses. Under real-world conditions, it is conceivable that some interactions may occur between these two requirements. For example, a module with relatively high initial quality but correspondingly high degradation rates may have lower quality after 20 years than a competing module with relatively low initial quality but low degradation rates.

Definition of Failure

In the preceding discussion, quality degradation rates are emphasized. This procedure contrasts with many common approaches to accelerated testing that emphasize definitions of failure. There are several reasons for preferring that attention be focused on degradation rates rather than failure. The best reason stems from the fact that whether or not a module has failed depends on the application. That is, the same module with a relatively low quality measure after 20 years of operation may constitute a "failure" in some system designs, but would not constitute a "failure" in a different system design. This is true even though the initial quality and degradation rates of quality for the module are identical in both cases. Thus, degradation of quality appears to be more inherent to the module itself, and focusing on quality and the degradation of quality avoids the paradox of having the same module quality indicating failure or nonfailure, depending on the application.

A second reason for focusing on degradation rates of quality stems from the fact that these rates are clearly required in order to make predictions of future quality levels. In the simple case of linear degradation over time, the degradation rate is constant and the amount of additional degradation at a future time is easily predicted by the product of the degradation rate and the projected time interval. If the predicted quality level is below that required for a more specified application, then the module would be predicted to fail; otherwise, no failure would be predicted. In this way, failure definitions and their applications can still be made appropriately after the degradation rates are determined.

The net result of this emphasis on degradation *rates* is that the results obtained from the accelerated test program can be used repeatedly for many different types of applications, and with correspondingly different failure definitions. This lends more generality and usefulness to the results of the accelerated test program.

Acceleration Factors

In an ideal case, it would be determined that 1 hour of operation at a high stress level is equivalent to, say, 10 hours of operation at normal stress levels. This would correspond to an "acceleration factor" of ten, and would allow a 2-year test under the higher stress to be interpreted as equivalent to a 20-year test under normal stress. In conceptually simple cases, the acceleration factor is simply equal to the ratio of the rate of degradation of the dominant failure mechanism at the high stress level to the rate of degradation at normal stress. Accordingly, the larger the acceleration factor, the more efficient and desirable the accelerated life test, provided that the dominant failure mechanism is the same for both the accelerated and normal operating conditions. However, it must also be noted that considerable controversy tends to be associated with every acceleration factor, and the larger the acceleration factor, the greater is the controversy. Acceleration factors of the order of ten, as desired for module testing, are large. Consequently, the utmost attention and care are required to preclude major deficiencies in the design, implementation, and analysis of the required accelerated test program.

RESULTS AND TECHNICAL DISCUSSION —
PART I: REVIEW AND ANALYSIS OF PAST WORK
ON AGING OF GLASS AND POLYMERIC MATERIALS
AND ON ACCELERATED TESTING

RESULTS AND TECHNICAL DISCUSSION —
PART I: REVIEW AND ANALYSIS OF PAST WORK
ON AGING OF GLASS AND POLYMERIC MATERIALS
AND ON ACCELERATED TESTING

THE INFLUENCE OF THE ENVIRONMENT, TESTING TECHNIQUES,
AND OPERATING CONDITIONS ON THE AGING OF
GLASSES AND POLYMERS

In designing any aging experiment under overstressed conditions (accelerated testing), a major consideration is choosing a set of environmental parameters that will accelerate those degradation rates of materials or interfaces that control failure under normal stresses. Most often, it is impractical to set up and control all environmental parameters in a test. Ideally, one would like to know a priori the dominant failure mode in the device or material. Under such circumstances, the more that is known about the behavior of the material or device, the better one can pick those environmental parameters and their magnitudes that will bring about the normal-stress failure. In an absolute sense, an acceptable accelerated test design does not require a knowledge of cause-and-effect directly; only an acceptable correlation between service life and some measurable quantity(ies) is required. This circumstance, however, applies only when sufficient field data are available. Field data on module performance, against which correlations might be attempted, are extremely limited to date. Therefore, to make the test design tractable under reasonable time and cost constraints, existing knowledge of the behavior of materials in various environments must be factored into the design in the beginning. In this light, brief summaries of the aging behavior of glasses and polymers of interest to the encapsulation problems are presented, along with some discussion of test methods and procedures.

Degradation Effects and Tests for Glasses

Although glasses are known to exhibit long lifetimes, they are susceptible to some weathering by chemical attack, especially by moisture. Moisture not only lowers the strength of glass (as evidenced by its susceptibility to static fatigue), but also, on prolonged exposure, can chemically leach and/or etch the surface. In those glasses with poor durability, these effects result in a permanent haze or translucency. In some glasses, solarization (UV-enhanced darkening) has been known to occur, but this can be avoided by compositional modifications.

Acidic Attack

Glass is susceptible to moisture attack by a leaching mechanism (selective attack or exchange of sodium ions by hydrogen ions). The more acidic the solution, the higher the rate of attack. Because the leaching process results in a depleted surface layer, subsequent attack is controlled by diffusion through this layer, so that the rate of attack may vary with the square root of time, assuming no change in the physical properties of the skin.[5,6] Weatherability of glass is sometimes judged initially on the basis of resistance to attack by

17

water, as measured by one or more alkali–extraction tests. But visual changes can occur in some glasses in which alkali extraction is not significant.[7] Weatherability is also dependent on relative humidity, although alkali–extraction data may not correlate with visual changes occurring on surfaces in different humidities.[7]

The chemical nature and thermal history of the glass surface also have pronounced effects on the chemical durability, as evidenced by surface treatments used to de-alkalize the interior of soda–lime glass containers, e.g., treatment with hot SO_2 or fluorine compounds.[8] The SO_2 treatment can increase the chemical durability to leaching by a factor of 25, but has little effect on chemical attack by alkali, where etching is the predominant attack mechanism.[5] The tin–rich surface of float glass is also more weather resistant than the silica–rich top surface, although specific data have not been reported.[9]

Alkali Attack

Glass can also be attacked by alkaline solutions, which essentially dissolve the silica structure of the glass. If no residue buildup occurs, the reaction rate proceeds linearly with time, and will increase by a factor of two or three with each change in pH unit.[5] In accordance with the difference in corrosion mechanism, the rate of attack for most glasses is several orders of magnitude higher in alkalies than in acids. Alkali attack is not normally encountered in weathering situations unless strong detergents are used for cleaning purposes.

Property Changes Resulting from
Surface Degradation

Loss of transmittance and strength are two factors of major concern for terrestrial encapsulants. These properties are known to be affected by washing of the glass surface, weathering products allowed to accumulate under static exposure conditions[7], and surface abrasion. Although abrasion can reduce the normal transmittance, solar cells can utilize some portion of diffuse light. Thus, surface abrasion may not be as critical to total light transmittance as in other applications.

Loss in transmittance by weathering can be measured by a haze test but this test is not commonly used to follow corrosion, probably because of the lack of an instrument as sensitive as the eye.[10] Moreover, the results are not reproducible.[7] Ellipsometry, a technique for measuring the refractive index and thickness of coatings on glass[11], has been used to follow weathering of high–lead glasses in laboratory experiments.[12] A variety of analytical techniques that characterize chemical changes occurring at surfaces[12] have been used to show that "as-fabricated" glass containers are weathered to some extent.[13] Attenuated-total-reflectance (ATR) spectroscopy has also been applied to the study of glass surfaces.[14] Usually, however, chemical-leach tests are used to predict the performance of materials in corrosive environments; appearance evaluations of exposed samples are used to confirm laboratory predictions.

The time dependence of glass strength in a moist environment is a well–known mechanical characteristic called static fatigue which is attributable to chemical attack of the surface. Depending on the environment, the long–term strength of glass for any one surface condition may be reduced to approximately 20 percent of the "dry" strength[15]. Beyond that the fatigue curve becomes horizontal and failure does not occur regardless of stress duration. This "endurance limit" is attributed to the healing of flaws by chemical etching at a rate higher than

that at which a "crack" can propagate.[15] Depending on the type of glass, aging or storage of stress–free glass in various moist environments can either increase or decrease the strength.[16] Unlike chemical attack, static fatigue is caused by the formation of a hydrated layer on the surface (rather than solution or ion exchange). Strength is relatively independent of pH, except at very high values, where it is increased (presumably by chemical etching of flaws at a higher rate than growth can occur), or at very low values, where it is decreased.[15]

The physical nature of the surface also affects the strength of glass. For this reason, glass in an abrasive environment is often abraded by one of several techniques[15, 17, 18] before testing. Because surface and environmental factors affect the strength of glass, strength data are usually reported only in the technical literature. Manufacturers usually report conservative "working" values reflecting the strength of abraded samples subjected to long-term loads, unless a particular application warrants otherwise.

Because static fatigue is caused by slow crack growth, fracture mechanics techniques using proof testing have been used to study and predict the failure time (life) of glass for any stress condition.[19] Crack–velocity measurements (made in solutions of varying pH), used in calculating glass strength, have shown that at low crack velocities (long–duration loads) the pH of the solution can affect crack–propagation rates. At high–crack velocities (short–duration loads), the crack propagation rate, independent of pH, is governed by the environment (acidic in water) at the crack tip.[20] Fracture–mechanics principles may, therefore, prove useful in the selection or evaluation of materials, and in prediction of the life of ultimate system components.

Solarization and Dirt Effects

Loss of transmittance can occur by solarization and dirt accumulation, as well as by weathering. Solarization is defined as the phenomenon whereby exposure to natural sunlight changes the transmission, and possibly color, of a glass. The transmission decrease in the visible range is attributed to a change in oxidation of multivalent cations in the glass by the following type of reactions:

$$4MnO + As_2O_5 = 2Mn_2O_3 + As_2O_3 \qquad (1)$$

$$4FeO + As_2O_5 = 2Fe_2O_3 + As_2O_3 \qquad (2)$$
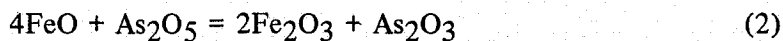
The first reaction imparts a purple discoloration; the second accounts for aging of UV-transmitting glasses[21]. Sb_2O_5 functions in the same manner as As_2O_5. Obviously, the best method of avoiding solarization is to minimize iron and other multivalent ions in the glass. Because arsenic and antimony oxides usually are not used as firing agents in soda–lime glasses, solarizatioi is not common.

Dirt accumulation is perhaps a more serious problem, depending on the inclination angle of the surface.[22] This effect is, however, related to a specific environment and is relatively independent of material type unless the surface is etched or abraded either prior to or during exposure.

Natural Weathering Tests

Because of the long natural life of glasses and the conventional use of chemical durability tests to predict the weatherability of various glass compositions before use, relatively little information has been published on the weatherability of modern-day glasses. Polymeric coatings have been used to protect these glasses but with little apparent success judging from results of accelerated humidity-cabinet exposures.[23]

With respect to strength, the stress-free aging of both annealed and toughened (method unspecified) soda-lime plate glass has been studied for times up to 5 years.[24] The effect of outdoor exposure on the strength of chemically strengthened glass panels has been reported as insignificant at the conclusion of 1 year of a 5-year test program.[25] As might be expected, lot-to-lot variability, surface-abrasion conditions, edge-initiated failures, and testing techniques were other factors affecting strength results.[25] U.S. Navy researchers have exposed chemically strengthened and thermally tempered glass to seawater in both the laboratory and ocean for periods up to 3 years. They found no fatigue effects at the 50 percent stress level; exposure actually made the glass stronger.[18] U.S. Army researchers have compiled static fatigue data on some glasses.[26]

Generally, a real or simulated system has been tested, rather than testing just a glass material, to evaluate materials-compatibility, design, and assembly factors — the areas in which most problems occur.

Artificial Weathering Tests

The glass industry has found that results of ASTM alkali-extraction tests conducted on either containers or powdered samples correlate quite well with outdoor weatherability.[27] For glasses with extremely long-term durability requirements, dynamic-corrosion-test techniques are also available.[28] The thermal history of a glass must also be considered, since heat treatment can result in phase separation and a drastic decrease in chemical durability.[29] In Morey's book, *The Properties of Glass,* the chapter on chemical durability of glass includes a dated but good discussion of test methods and compositional factors for bulk glass.[30] In Tooley's book, *The Handbook of Glass Manufacture,* the section on chemical durability emphasizes liquid-media test techniques which yield quantitative data for predicting durability at any service condition on the basis of accelerated test data plotted in Arrhenius form[31], but the usefulness of this technique for glasses exposed to natural weather environments is questionable.

For coatings on architectural glass, acidic attack[32], as well as salt-spray, relative-humidity, or weatherometer "cabinet" tests may be employed to evaluate the durability of a product. Because coated architectural glass is relatively new, test procedures that simulate failure mechanisms in actual environments are evolving as manufacturers accumulate field data. Information of this type is generally proprietary and not published. Ellipsometry[11, 12] and ATR[14] (discussed above) appear to be techniques by which surface property changes could be quantified for measuring the rate at which surfaces weather in nonliquid environments.

The high-temperature, high-humidity, cyclic-temperature tests of actual components by the microelectronics industry represent typical procedures used to develop and evaluate improved systems before placing them in actual service.[33, 34] The types of failures encountered in these "accelerated" tests have been found to occur in service, so the tests are useful in predicting whether or not a unit design will fail, assuming that it will not fail by another mechanism.

The industry uses Weibull plots to analyze the data.[35, 36] This technology is significant because the electrical characteristics of the system are similar to those of silicon solar cells (semiconductor material, interconnect corrosion and fatigue, encapsulation purposes, etc.). Glass applied by fusion of powders is a prime silicon-device encapsulant where thermal-fatigue, electrical-stability, and moisture-resistance characteristics are critical.[37]

Manufacturers of insulating glass have also developed information relevant to the weatherability of solar-array encapsulant systems. For example, a set of tests described in Canadian Government Specification Board Specification 12-GP-8 has been used to qualify manufacturers and compare systems, but there has been no direct correlation between any one type of test and field-failure experiences. However, the use of the tests has resulted in improved product quality and a decreased field-failure rate.[38] One U.S. manufacturer has reported that conclusive correlations between accelerated and field performances were not possible[39], whereas another (using organic sealants) reported that field performances correlated with combined UV-exposure (500 hr) and temperature-cycling (20 to 120 F) test results[40]. The expression:

$$F(x) = 1 - \exp(-x/\theta) \quad ,$$

where $F(x)$ = cumulative failure probability, x = failure time, and θ = mean failure time, was found to describe the failure probability of field units after 17 years of experience.[40] The conclusions from a review of the development of test methods used by insulating-glass manufacturers[38] were that none of the methods can help to predict durability in actual service, but that, combined with field experience, they can help to predict poor field performance[41]. ASTM Committee E-6 is preparing a draft of proposed practices for accelerated testing of insulating glass, which is not yet available.

Discussion and Summary

This review of aging characteristics and/or weatherability data for glasses did not reveal quantitative information on degradation rates of glass under normal ambient conditions. The review indicates, in part, the proven durability of the material. Although weathering is not usually a problem with modern-day glasses, it can occur in specialty glasses that have not been developed with this consideration in mind. Moisture and temperature are usually the two most important factors affecting weatherability. Although they may not visibly affect the transmittance or appearance of glass, they can drastically reduce its long-term strength under stress by a phenomenon known as static fatigue (or stress corrosion). If the glass is not exposed to a high mechanical stress, aging in a moist environment may increase or decrease the strength slightly. Solar-array encapsulant systems will probably be designed so that residual stresses in the system are minor, but environmental stresses such as wind and snow loads may require that static fatigue characteristics be considered in the design.

In most encapsulation systems using glass today, the glass is used in conjunction with some organic sealant material (e.g., pottant, film, adhesive). Consequently, degradation of the glass and sealant interface would be a more likely cause of failure than degradation of either bulk material. This type of degradation phenomenon is known to be enhanced by moisture.[42]

Test procedures which can quantitatively and nondestructively measure degradation rates at interfaces and/ or surfaces would be highly useful in this respect.

Although over-stressed tests on powdered or bulk glass are commonly used in the laboratory to evaluate durability, quantitative measurements are usually made only on materials tested in liquid media, where the media can be analyzed. Experience has shown a general correlation between weatherability and resistance to water, but weatherability must be verified by exposure of bulk specimens to high humidities. Unfortunately, techniques for quantitatively measuring surface-degradation characteristics are not well-developed, and visual observations are usually employed. For predicting the life of a system, degradation rates must be known, and procedures by which surface (and/or interface) degradation can be quantitatively measured must be developed.

Degradation Effects and Tests for Polymeric Materials

Table 1 lists important weather and other environmental parameters that can affect the service life of polymeric materials. Perhaps the most important of these is ultraviolet (UV) radiation. Since radiation below a wavelength of 0.29 μm is filtered out by the earth's ozone layer, UV radiation between 0.29 and 0.40 μm is of the greatest concern.[43] This radiation is in the range (70 to 100 kcal) of the dissociation energy of the chemical bonds found in polymers.[44] In addition to bond cleavage, UV radiation can radicalize double bonds, excite electrons, and activate other chemicals which will promote degradative chemical reactions in the polymer molecule.[45] The principal reaction results are chain scission, cross-linking, and/or depolymerization (unzipping). The specific reaction varies from polymer to polymer.[46] Chain scission reduces the polymer's molecular weight, makes it more susceptible to vapor diffusion, and weakens the macrostructure. Crosslinking entails an increase in localized density, which creates internal stresses and the development of micro-cracks that embrittle the material, reduce light transmittance, and increase the surface area open to chemical attack.

While UV radiation is always less than 10 percent of the total spectral radiation, its intensity varies markedly from season to season and from hour to hour. Cloud cover sharply reduces UV intensity; intensity at certain wavelengths varies more than at others.[47] Commercial polymers exhibit absorbance peaks in the UV range at which they are particularly sensitive to degradation reactions.[43, 47] Some peaks are listed in Table 2. As a result of these observations, the total integrated intensity of sunlight, which is a commonly measured index of solar-radiation intensity, does not adequately represent actinic UV radiation.

Oxidation is one of the major reactions promoted by UV radiation. Oxygen concentration at the surface often is essentially constant, and the diffusion rate through polymers is low. And, except in thin films, the UV light does not penetrate very far into the material. There-fore, oxidation occurs at the surface of the polymer.[48, 49] The specific effects of oxi-dation reactions vary, although they often involve discoloration and embrittlement.

Water has two effects on a polymer: it can react with the polymer to weaken it, or it can leach out stabilizing chemicals or clean a degraded surface, thereby exposing a new surface for attack.[50] The various forms in which water can contact a surface greatly complicate the analysis of its weathering effect. While oxygen concentration is constant, the amount of water vapor in the air changes continually. Water in the form of liquid

TABLE 1. WEATHER AND OTHER ENVIRONMENTAL PARAMETERS AFFECTING POLYMER DETERIORATION

Parameter	Description
UV Radiation	0.29 to 0.40-μm wavelength range
Oxygen	O_2 or O_3
Water	Rain, snow, dew, frost, fog, and humidity
Temperature	Absolute value, thermal gradient, and thermal fluctuations
Chemical Pollutants	SO_2, NO_2, CO, and NaCl
Particle Bombardment	Dust, sand, hail, bird droppings, and insects

TABLE 2. ACTIVATION-SPECTRA MAXIMA IN ULTRAVIOLET REGION FOR SEVERAL POLYMERS

Polymer	Activation-Spectrum Peak, A
Polyesters (various formulations)	3250
Polystyrene	3185
Polyethylene	3000
Polypropylene	3700
Polyvinyl chloride	3200
Polyvinyl acetate	< 2800
Polycarbonate	2850-3050 and 3300-3600
Cellulose acetate butyrate	2950-2980
Styrene acrylonitrile	2900-3250

23

moisture also varies in occurrence, thickness of coverage, and duration of coverage. The effects of rain are less well known than the effects of dew or humidity, since rain often contains atmospheric contaminants. Moreover, reactions between water and the polymer are often activated by UV radiation.[46]

The maximum temperature that an exposed polymer can be expected to reach in a module has been estimated in one analysis to be 77 C.[51] Such a temperature is not sufficient to cause thermal breakdown of the more stable polymers.[46] High temperatures do increase degradation reaction rates, however. This fact is particularly pertinent since the highest UV-radiation intensity occurs in the summer months when the temperature is also highest.[47] Thermal gradients and thermal fluctuations also affect material degradation. These factors result in stresses being set up in the material. Stressed polymers not only chemically react more rapidly[52], but, also increase the tendency of the polymer to delaminate if it is laminated to a substrate with a different thermal expansion coefficient[53].

The effects of chemical pollutants are less well known. Studies only recently have been initiated to evaluate these effects.[54] Because of the variable nature of weather and polymers, it has been difficult to separate the effects of atmospheric pollutants from the major factors in polymer deterioration discussed previously.

Nonchemical degradation must also be considered since the array encapsulant must remain transparent to solar radiation. Dust, sand, hail, bird droppings, and insects are the principal "nonchemical" deteriorating factors. Polymer surfaces often become charged and, therefore, attract dust. These parameters have received little study and are often unmeasured.

Natural Weathering Tests

Natural weathering tests often are used for end-use service testing. Samples are usually mounted at a 45-degree angle facing south. A suitable property is then measured periodically to determine material changes. This technique has many limitations. First, it is unreasonable to test a material expected to last 20 years in this manner. Second, samples have often been mechanically unstressed during the test while they may be stressed in use. Third, the mounting angle affects the results since a 45-degree angle will not always receive the maximum solar energy. The 45-degree angle also fixes the moisture runoff condition. These conditions may or may not duplicate use conditions. However, the most serious limitation in many natural weathering tests has been that the weather parameters influencing the samples' deterioration have not always been recorded. For example, the UV-radiation intensity impinging on a test sample (which is the most important single paramenter) the sample temperature, thermal fluctuations and gradients, and the duration of moisture on the sample are often not measured. Also, airborne chemical pollutants have usually been ignored in data analyses because atmospheric-pollution data are often lumped into a single quantity (i.e., air-quality index) which does not delineate between specific chemical pollutants.

There are still further limitations to the natural weathering tests. In tests lasting 2 years or less, the time of the year at which the test begins strongly affects both the results and the analysis.[55] A sample tested for 18 months may encompass one summer or two. A sample enduring two summers has undergone a much more severe test than the sample tested through one summer. The physical property tested is sensitive to the type of degradation that has occurred. Surface-oxidized samples, for example, may exhibit little loss in tensile strength,

yet show a much more brittle behavior and a reduced elongation prior to failure.[56, 57] The appropriate property for measurement is one that is sensitive to, but also indicative of, the end use of the material. Solar-cell encapsulant materials that retain their tensile strength yet lose their transparency are obviously failures. Ideally, the property should be tested by a non-destructive technique to eliminate the large number of samples needed to reduce sample-to-sample variation and still permit testing at various times. Finally, outdoor weather tests are valid only in the area in which the tests are conducted. Reference 58 discusses how different failure mechanisms occur in outdoor weathering in the Panama Canal Zone, a small geographical area compared with the continental United States.

Many of the above limitations can be easily overcome. Mounting the samples equatorially so that they always receive the maximum amount of radiation eliminates the mounting-angle dilemma. Although such mounting is not always consistent with end-use exposure, it maximizes the degradation rate and enables the results to be correlated with UV-radiation intensity. UV-radiation intensity impinging on the samples must be recorded. Such a measurement provides a better exposure parameter than days, sun hours, or total solar radiation. UV radiation includes both direct solar radiation and scattered skylight. If the samples are to be mechanically stressed in use, they can be stressed in the weathering test. The level of stress applied to the test sample should be the same as that in the prototype, thus avoiding a change in the failure mechanism.

Sample temperature, thermal gradients, and thermal fluctuations should also be recorded, as should humidity and the duration of surface moisture on a sample. Duration of surface moisture is more important than total rainfall. A heavy rainfall during the night that drains off the sample by morning will have much less effect on the degradation of a polymer than dew that remains and is exposed to sunlight. Airborne pollutants will also have to be monitored.

Oxygen level, while quite constant at a specific locale, does vary with altitude. For that reason, the barometric pressure should be recorded at each testing locale.

As mentioned previously, sample testing ideally should be nondestructive. Measuring changes in IR reflectance, UV absorptance, gloss, light transmission, ESR, and chemiluminescence, as discussed later, enable chemical changes in the sample to be monitored. Attempts can then be made to relate these changes to the macroscopic property of interest.

Unfortunately, the above procedures do not remove the limitations of geographical differences, nor do they abbreviate the testing time span. Equatorial mounting of the samples with focused mirrors (designated EMMA, Equatorial Mount with Mirrors for Acceleration, by Desert Sunshine Exposure Tests, Inc.) has been utilized to accelerate degradation by concentrating the solar radiation on the sample by a factor of eight.[47] There has not been a simultaneous eight-fold increase in degradation rate, however.[47] Correlation is lacking because UV intensity does not have a simple multiplicative effect on degradation. Various degradation reactions occur, often simultaneously, often competitively, and sometimes alternately. With several reactions possible, increasing only one parameter will not increase the overall degradation rate proportionately. Even more serious is the fact that the prime degradation mechanism may change. Schafer shows how the degradation mechanism of PVC changes when UV intensity is increased above a certain level.[59]

Artificial Weathering Techniques

Since it is seldom practical to test a material for 20 years, efforts have been made to develop accelerated testing methods which can be used to predict actual service life. In general, these efforts have been disappointing. The major difficulty is that "weather" cannot be accelerated.[46] In accelerated aging tests, one of the degrading forces (usually UV radiation) is increased above normal levels in an attempt to accelerate the rate of degradation of the material. EMMA and EMMAQUA* are devices designed to accomplish this. The intensity of UV radiation is increased by a factor of eight in these devices. In laboratory units, artificial sources are used to generate UV radiation. Figure 1 shows the UV energy distribution of several common UV sources in comparison to solar radiation. The Xenon arc most closely matches solar radiation, although differences are still evident. Many of the reported accelerated tests used a carbon arc as the radiation source. The carbon arc is highly deficient in UV radiation between 0.30 and 0.34 μm, a range particularly damaging to polymers. The intensity of radiation from the various sources also fluctuates. Radiation from the carbon arc fluctuates in a highly erractic manner.[60] The sunshine carbon arc, which more closely approximates solar radiation than the enclosed carbon arc, consists of three filaments which burn sequentially. They were found to radiate at different intensities.[60] Commercial Xenon-arc lamps produce steady radiation which gradually decreases with time.[60]

Because of the variability in radiation frequency between the artificial UV sources and the sun, it is difficult to equate impinging artificial UV radiation with solar radiation. Usually, the test samples are located so that the incident radiation most closely approximates the integrated UV energy of the sun at some average condition (e.g., noon in the month of June). Attempts at increasing the radiation intensity to achieve acceleration will further distort the differences between the solar radiation and the artificial radiation. Therefore, tests are usually run continuously to accelerate the exposure time to UV radiation. Even this tactic is not ideal in that the cyclic nature of UV solar radiation is not simulated. Some accelerated-weathering devices have a night cycle, but this does not simulate the natural cyclic daytime intensity of solar UV radiation. The "night" cycles last only several minutes out of every hour.

In most accelerated testing devices, deionized, distilled water is sprayed on the samples periodically. Some devices even maintain a constant humidity. All maintain a constant temperature, and very few introduce any of the other parameters listed in Table 1. Clearly, artificial weathering does not simulate natural weathering. The only major accelerating parameter is UV exposure. Because natural degradation is a complex combination of many factors, it is not surprising that artificial weathering tests do not correlate with natural weathering tests. When one considers that outdoor exposure tests in Cleveland, Ohio, cannot be correlated with similar tests in Miami, Florida, because of the difference in climate, it is not surprising that laboratory tests, which simulate neither weather situation, cannot be correlated to either one.

The major use of laboratory weathering tests has been to rank materials in order of degradation stability and to evaluate the effectiveness of stabilizers and pigments.[61]

---

* EMMAQUA is an EMMA device which periodically sprays water on the samples.
  (Desert Sunshine Exposure Tests, Inc., Phoenix, Airzona).
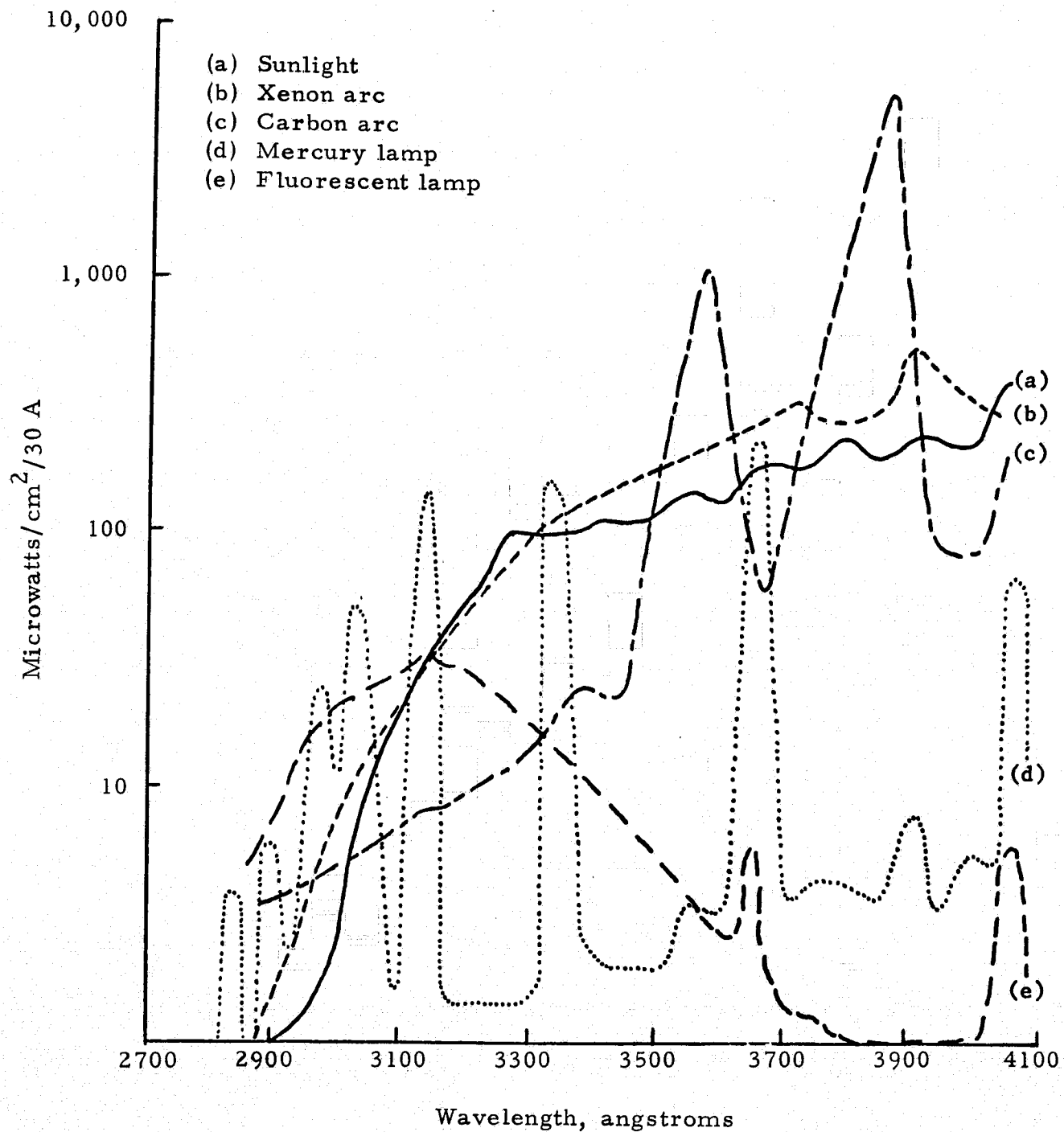
FIGURE 1. UV DISTRIBUTION OF VARIOUS UV SOURCES

(a) Sunlight
(b) Xenon arc
(c) Carbon arc
(d) Mercury lamp
(e) Fluorescent lamp

Discussion and Summary

One of the major difficulties with current weathering tests is that insufficient information is recorded about the test environment. Useful weathering data can be obtained only if the total incident UV-radiation energy, atmospheric humidity, duration of surface moisture, sample temperature, thermal gradients in the sample, and type and concentration of atmospheric chemical species are recorded continuously during the exposure period. Recording of this necessary information is a prerequisite of any worthwhile exposure test, accelerated or otherwise.

One of the most successful and relatively simple accelerated-testing schemes involves reducing the failure times of the test samples by increasing the stress level of the most influential deteriorative parameter.[62, 63] Stress level is then plotted against the logarithm of failure time. If the failure times span several decades, a reasonably confident extrapolation can be made of one order of magnitude. If the stress level at the extrapolated time is less than the expected use stress level, the sample can be expected to survive to the extrapolated time in normal use. This type of experiment is quite versitile. The failure time is determined by the experimenter as the time at which a specified reduction in some key property is reached. The exact failure mechanism need not necessarily be known, only that it is the same throughout the test. Since the degradation mechanism may change as the result of the higher-than-normal stress, the test must be capable of indicating when a change in failure mechanism occurs. Changes in failure mechanism are observed as discontinuities in the slope of the stress-log failure time curve. This has been shown by Schafer for the case of PVC subjected to increasing UV intensity.[59] Above a certain level of UV radiation, the failure mechanism changes. There are serious limitations to this approach, however, the most important being the requirement of maintaining the same failure mechanism at every stress level. For the case of a solar encapsulant, the extrapolated time is 20 years ($\sim$6.3 x 10^8 seconds). To extrapolate with reasonable confidence to this time, data must be available at 2 years, 73 days ($\sim$6.3 x 10^6 seconds), 7.3 days, and 17.5 hours ($\sim$6.3 x 10^4 seconds). The other limitations to this accelerated-testing approach are the assumptions that one parameter is responsible for degradation and that one physical property can always be used to determine failure.

An alternative approach that avoids some of the difficulties with the accelerated stress/failure time approach is to measure some extremely sensitive parameter under normal exposure conditions. Techniques involving attenuated total reflection (ATR) of the infrared spectrum [44, 48, 49, 64, 65], UV absorption spectra [64, 66], electron spin resonance (ESR) spectra[67], and chemiluminescence [68, 69] can detect very small chemical changes in a material long before any macroscopic property changes occur. Changes in chemical structure due to degradation reactions that have been measured by these techniques have been correlated with property deterioration of polymers. ATR of polyethylene IR-absorption spectra enables measurement of the increase in carbonyl bonds formed during photooxidation.[25] The amount of carbonyl formed can be related to total UV-energy exposure and the reduction of the material's elongation to break.[70] Since this technique can be applied to outdoor exposure testing, data can be taken at 17.5 hours, 7.3 days, 73 days, and 2 years to generate a log-time extrapolation curve. The testing method is nondestructive, allowing the same samples to be tested at each time interval. More frequent measurements can also be made to determine the effect of parameters other than UV exposure. The ideal situation is one in which the sample is tested continuously providing an excellent match between weather conditions and degradation. Using this technique, more realistic testing could occur weekly or monthly.

This approach has not as yet been applied to long-time predictive testing. The measurement techniques have been developed only recently and the testing procedure is most likely expensive. A limitation is that the key failure property cannot always be related to some chemical-degradation process, such as the effect of hail on polymers. Nonchemical degradation must be treated in some other manner. Weathering tests will have to be regional until the effect of each weather parameter can be fit into some kinetic-rate-of-degradation model.

A final caution is that raw materials themselves are sometimes altered slightly, occasionally substantially, over the years. Since many of the photo-induced reactions are with end groups, residues, and additives, data taken on a polymer may not remain valid after these changes are made. Suppliers do not usually announce such changes. Grinsfelder reports weathering data performed in factorial experiments where the largest significant variable was the resin supplier.[71]

MAJOR EXPERIMENTAL AGING TESTS AND MODELS

Literature data on aging of polymeric materials are quite extensive. The types of aging studies found can be divided into two almost mutually exclusive classes according to the goals of the study. The great bulk of literature on aging of polymeric materials provides purely descriptive data, for example, material properties before and after aging for a given time under a given set of conditions. The other class includes studies that attempt to describe aging behavior by mathematical modeling. The latter type of information is discussed in this section of the report. The most pertinent findings, both with respect to the data illustrated and to the methodology brought to bear on their mathematical representation are summarized in this section.

Aging Studies and Models

NBS-MCA Study

One of the most ambitious studies of natural and accelerated weathering to date was the joint industry-government program undertaken by the National Bureau of Standards and Manufacturing Chemists Association.[72, 73] In this program, 20 plastic materials, including six generic plastic species, were exposed beginning in 1966 in Miami, Florida; Phoenix, Arizona; and Washington, D. C. Both clear films and white films of various thicknesses were exposed. Table 3 describes the materials exposed and lists the properties measured. The program was designed to last for 10 years, but was almost aborted for lack of funds after only 6 years. Property measurements were made initially at 3-month intervals and later in the program at 1-year intervals. Since the stated goal of the study was to correlate accelerated and natural aging data, weatherometer testing was also done on the same materials.

In the NBS data-analysis approach, the property-versus-time data were fitted to the following equation derived from the Weibull probability density function:

$$P = b_1 \exp\left\{-\left[\left(\frac{t + b_2}{b_3}\right)^{b_4}\right]\right\} + b_5 , \qquad (3)$$

where P is the property level at time, t, and b_1 through b_5 are fitted parameters. The five fitted parameters were claimed to have physical significance as follows:

b_1 — is related to maximum property level
b_2 — is related to pre- or post-aging
b_3 — is related to characteristic life defined as the time required for 63-percent property degradation
b_4 — is related to the presence or absence of an initial induction period
b_5 — is the asymptotic property level at infinite time.

Because of these relationships, the b_i's were called the exposure parameters.

The next step in the data analysis consisted of fitting the important exposure parameters, or meaningful functions of them (such as characteristic life) to a linear equation in meteorological variables as follows:

30

$$b_i = C_i + C_{iL}(L) + C_{iU}(U) + C_{iH}(H) + C_{iR}(R) + C_{iT}(T) \qquad (4)$$

where L is total radiation in langleys, U is UV radiation (Coblentz langleys), H is relative humidity, R is inches of rainfall, T is air temperature, and the C_i's were parameters fitted by stepwise regression. Thus, from the basic meteorological variables L, U, H, R, and T, it should be possible to predict the values of b_i, which in turn could be used to predict property levels or rate of property change at time, t.

The complete data analysis was made only for elongation at break and only for 3 years of data. Therefore, a judgement as to the success or failure of this approach is rather difficult. From the limited data analysis, however, successful results appear to have been obtained for some of the material-site combinations, whereas gross deficiencies were apparent for other combinations.

TABLE 3. MATERIALS AND PROPERTIES EVALUATED IN NBS-MCA STUDY ON WEATHERING OF PLASTICS

Number of Compositions	Materials Evaluated		Thickness, mils
	Generic Class	Color	
Two	Polyethylene	Translucent	1 and 60
One	Polymethylmethacrylate	Clear	60
One	Polyvinylfluoride	Clear	1
One	Polyethylene terephthalate	Clear	5
One	Crosslinked polyester	Clear	60
Fourteen	Various PVC materials	Clear and white	9, 10, and 60

Properties Evaluated	
Color	Tensile properties
Haze	Flexural properties
Glass	Electrical properties
Surface roughness	UV spectra

Leikina and Tatevos'yan Study

In this accelerated-aging study[74], a central-composite, response-surface experimental design[75] was employed to study the effects of temperature, X_1 (30 to 65 C), and radiation intensity, X_2 (150 to 300 W/m^2), on tensile strength, elongation to break, and IR absorbtion (1720-1780 cm^{-1}) of 0.1-mm polyethylene film. This type of design allowed each response to be fitted to a second-degree polynomial equation in X_1 and X_2 as follows:

$$P = b_0 + b_1 X_1 + b_2 X_2 + b_{11} X_1^2 + b_{22} X_2^2 + b_{12} X_1 X_2 , \qquad (5)$$

where P is the fitted property and the b_i's are parameters fitted via multiple-linear-regression

analysis. Statistical–significance testing indicated that the equations fitted within experimental error. From the fitted equations, interpolations, possible within the limits of the design, enabled property predictions for conditions not actually evaluated. Since such fitted equations are purely empirical, however, extrapolation outside of the design limits results in predictions with a high level of uncertainty.

Leikina, Tatevos'yan, Kuznetsova, and Melkumov Study

In this study[76], the effects on polyethylene and polyvinyl chloride of four variables — irradiation intensity (350–700 w/m^2), specimen surface temperature (20 to 75 C), fraction of time exposed to water during test cycle (1/6 to 1/2), and total test time (170 to 340 hr) — were investigated. The investigators employed a Soviet DKSTV-6000 weatherometer with a xenon lamp. The experimental design employed was a full 2^4 factorial in which all combinations of the two levels of the four variables were examined. This design permitted the fitting of the following empirical equation to the same properties as in the previous study:

$$P = b_o + \sum_{i=1}^{4} b_i X_i + \sum_{j=1}^{4} \sum_{i=1}^{4} b_{ij} X_i X_j \quad , \tag{6}$$

where, again, P is the property level and the b_i's are fitted parameters. Again, the fitted equations were useful for interpolation, but not for extrapolation.

This study and the previous one are significant because the experimental designs enabled interactions between different independent variables to be elucidated. Synergistic effects go undetected in classical one–variable–at–a–time experimentation, but may be the most important information obtained from an experimental investigation. For example, for polyethylene tensile strength, the interaction between exposure time and temperature turned out to be the most significant effect.

Kamal Study

Kamal's study[77] is similar to the two Russian studies just cited, but is somewhat more rigorous even though it predates them. Using a xenon–arc weatherometer, he set up a program consisting of 16 sets of fixed weatherometer conditions (e.g., fixed temperature, fixed length of wet time, and fixed fraction of total time sample was wet). For each set of conditions, polysytrene, polyvinyl chloride, and crosslinked–polyester samples were exposed for 400 hr, and properties of interest, including tensile strength, color change, and UV absorption, were measured at 100-hr intervals. Property-versus-time data were fitted to the equation:

$$\log P = b_o + b_1(t-250) \quad , \tag{7}$$

where P is the property level at time, t, and b_o and b_1 are fitted parameters. Clearly, b_o is the log of the property level after 250 hr, and b_1 is the logarithmic rate of change of the property with time. Kamal referred to b_o and b_1 as exposure parameters and fitted each of them to a quadratic equation in the weatherometer variables as follows:

$$b_o \text{ or } b_1 = C_o + C_1 X_1 + C_2 X_2 + C_3 X_3 + C_{11} X_1^2 + C_{22} X_2^2$$
$$+ C_{33} X_3^2 + C_{12} X_1 X_2 + C_{13} X_1 X_3 + C_{23} X_2 X_3 \quad , \tag{8}$$

where X_1 = temperature, X_2 = length of wet cycle, and X_3 = percent of time in each cycle during which the sample was wet.

As part of this study, samples of the same materials were exposed outdoors with concurrent estimates of temperature, wetness, and UV energy conditions. From information in Equation (8), outdoor property values were predicted as a function of time and compared with the actual exposure data. Some of the results obtained in this manner were surprisingly good. Figure 2 shows the results for tensile strength and color change of polysytrene for up to 52 weeks. With the exception of the 52-week color-change value, the predictions shown by the solid line matched the actual data points well.

Natural Rubber Producers Research Association (NRPRA) Study

In this study[78], K. D. Thomas and R. Sinnott predicted room-temperature modulus changes in both polyacrylonitrile and polychloroprene elastomer systems for up to 5 years from accelerated heat-aging data at 100 to 150 C. Assuming first-order degradation kinetics, they fitted the equation:

$$\log \frac{P_t}{P_o} = kt \ , \tag{9}$$

were P_t and P_o are property levels at time, t, and time zero, respectively, and k is the first-order rate constant. They then related k to temperature, T, using the Arrhenius equation:

$$k = A \exp (-B/RT) \ , \tag{10}$$

where A and B are fitted constants (B is the activation energy for the process causing failure). These equations were then used to predict P_t at room temperature for several properties for up to 5 years. The results are shown in Table 4 for 100 percent modulus along with actual property levels from concurrent outdoor testing. As can be seen the results are quite good, especially in view of the fact that extrapolations were made over wide temperature and time intervals.

TABLE 4. RUBBER AND PLASTICS RESEARCH ASSOCIATION
(RAPRA) 5-YEAR PREDICTION RESULTS FOR
MODULUS[78]

| Time at Room Temperature, years | Change in Modulus, percent | | | |
| | Nitrile Rubber | | Neoprene Rubber | |
	Predicted	Observed (RAPRA)	Predicted	Observed (RAPRA)
1	6	5	4	16
2	12	11	8	19
3	19	17	12	20
4	26	22	16.3	22
5	34	27	21	23
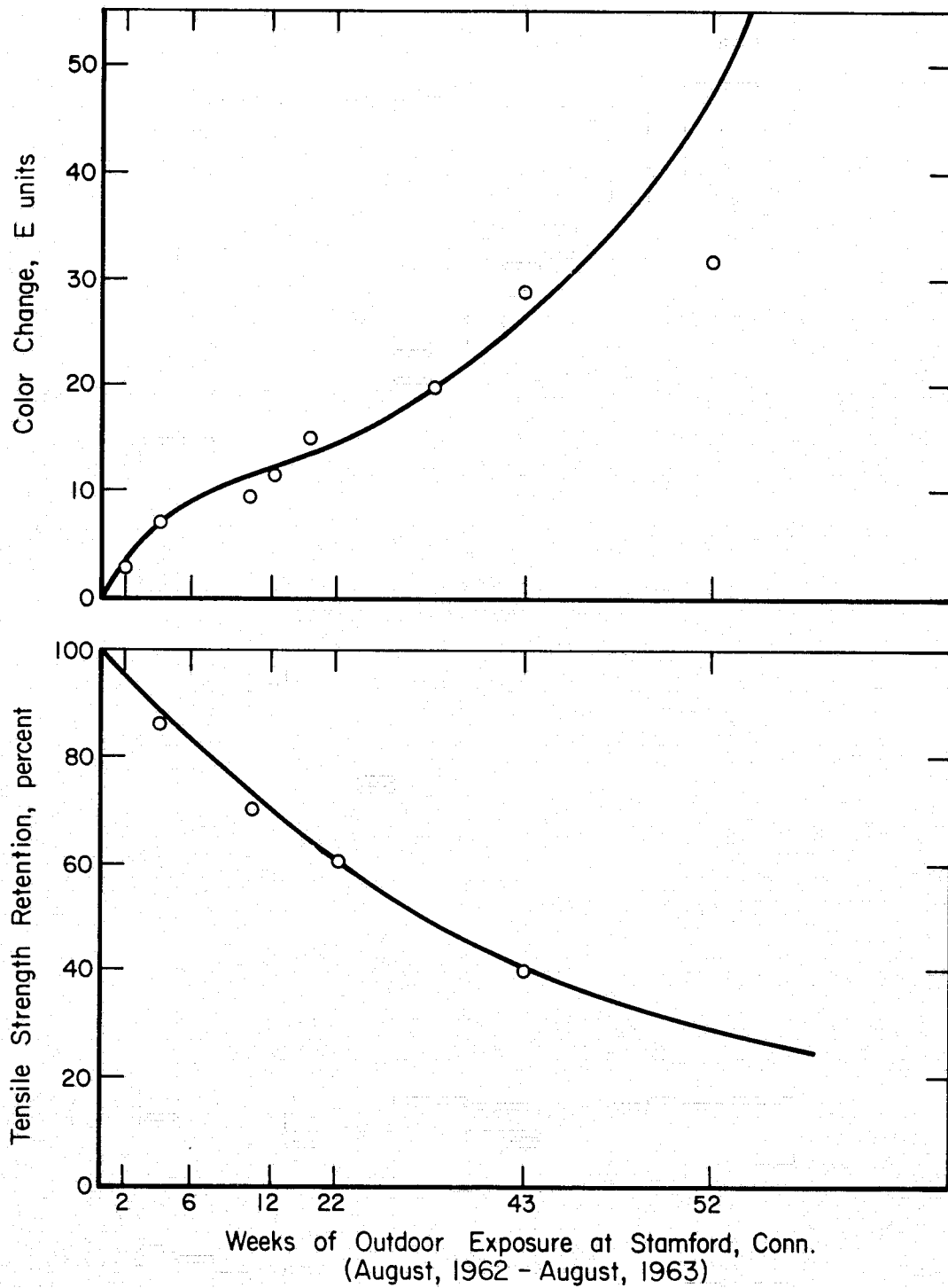
FIGURE 2. OUTDOOR EXPOSURE PERFORMANCE OF POLYSTYRENE
VERSUS VALUES PREDICTED FROM XENON-ARC
WEATHEROMETER[10]

Lockheed Study

The purpose of this study[79] was to predict service life of a propellant. In the accelerated–testing plan, cartons of the propellant were aged under nitrogen at 30, 70, 86, 115, and 145 F and at 0 and 5 percent strain for periods up to 120 weeks. Gel content, degree of swell, creep compliance, dilation, and crack propagation were measured at intervals during the testing period. The property–versus–time data were fitted to one of the following equations:

$$P = P_0 + k \log t \quad, \text{or} \tag{11}$$

$$P = P_0 + k_1 \log t + k_2 \log^2 t \quad, \tag{12}$$

where P and P_0 are property levels at time t and 1 week, respectively, and k, k_1, and k_2 are fitted rate constants. The linear equation usually was adequate. Where it was not, the quadratic equation was employed. The rate constants k, k_1, and k_2 were related to temperature by the Arrhenius Equation (10). The compliance equations were extrapolated over time and temperature to arrive at a predicted 1 sigma (σ) service life interval of 4.6 to 11 years. Good agreement with 7- to 8-year modulus data was claimed.

Hill Air Force Base Study

In this study[80], several properties of various components are being monitored at storage conditions to "detect changes which could reduce service life estimates". The components include various potting compounds, adhesives, spiralloy, and pressure seals. Properties include tensile strength, elongation, hardness, and lap–shear strength. Breakway torque and leak rate are measured on the seals. Data are presently available for up to 8 years for the adhesives and potting compounds and 13 years for the seals. Measurement intervals for all but the seals are approximately 1 year except for the first year where 1- and 6-month measurements were made for some properties. The number of replications at each data point vary from three to ten. Pressure seals are tested every other month with no replication.

Reported data treatment consists simply of fitting a linear relationship to the data as follows:

$$P = P_0 + b_1 t \quad, \tag{13}$$

where P is property level at time t, P_0 is initial property level, and b_1 is the fitted slope. The slope is then tested for statistical significance. No attempts are made to predict service life or to evaluate alternative models.

Since the information in the report is rather sparse, the Hill Air Force Base was contacted for some of the details related above and to determine the availability of raw data not included in the report. The data are stored on computer tapes and can be made available.

Critique of Literature Aging Studies

In critiquing the literature on aging studies*, the following characteristics desirable for both accelerated- and abbreviated-test plans are considered:

(1) Replication

(2) Sufficient data points for model building

(3) Sufficient number of properties measured

(4) Characterization and/or randomization over materials studied.

Replication is necessary if the signal and noise content of the data are to be separated. Without replication, it is impossible to determine to what extent the model is being fitted to noise rather than to signal.

Without a sufficient number of data points, model adequacy may be hard to assess and difficulties may be encountered in comparing alternative models. For example, if the NBS five-parameter model is fitted to only six data points, only one degree of freedom is available to statistically test the model. The resultant test is not very powerful.

All properties which can conceivably lead to failure should be measured for obvious reasons. It appears to be common practice to measure only properties which are known a priori to give the desired results so that a particular methodology can be illustrated.

Finally, the materials studied should be characterized completely so that, to put it quite simply, one knows what one is predicting the life of. This is especially true for polymers where a generic polymer type can encompass a wide range of molecular weights, molecular-weight distributions, degrees of branching, amounts of residual monomer, etc. Furthermore, if the materials tested are selected (within a generic polymer type) by some random process, predicted lifetimes and life distributions will be applicable to a much wider range of materials. It is desirable to have both characterization and randomization in an aging study. If this is not practical, it is essential that one or the other be done.

The literature studies discussed above are compared in Table 5 according to these four criteria. The inadequacies of even the best existing aging studies with regard to the goals of this program are quite apparent. Characterization of and/or randomization over materials is nonexistent unless it is done and not mentioned in the reports. Replication, sufficient data points, and sufficient properties are also common deficiencies. The test plan to be developed must be free of these deficiencies. Another glaring deficiency of published outdoor-aging studies is the failure to monitor properly variables that can contribute to failure, such as UV intensity, relative humidity, average temperature, temperature extremes, pollutant levels, and percent wet time.

---

*This critique relates only to the goals of this study and is not intended to reflect on the overall quality of the individual studies.

TABLE 5. CRITIQUE OF AGING STUDIES OF POLYMERIC MATERIALS REPORTED IN THE LITERATURE FOR SPECIFIC APPLICABILITY TO THE EVALUATIONS FOR THIS PROGRAM

Investigator	Reference	Replication	Sufficient Data Points	Sufficient Properties	Characterization or Randomization Over Materials
NBS–MCA	(72,73)	No	Yes-no	Yes	No
RAPRA	(78)	No	No	No	No
Kamal	(77)	No	Yes	No	No
Leikina, Tatevos'yan	(74,75)	Yes	Yes	Yes-no	No
Leikina, Tatevos'yan et al.	(76)	No	Yes	Yes-no	No
Lockheed	(79)	Yes	Yes	Yes	No
Hill AFB	(80)	Yes	Yes	Yes	No

Critique of Property–Versus–Time Models

The following property–versus–time models have been employed in studies discussed above:

$$P = P_0 + b_1 t \qquad (13)$$

$$P = P_0 \exp(-kt) \text{ or } \ln P = \ln P_0 - kt \qquad (9, 7)$$

$$P = P_0 + k \log t \qquad (11)$$

$$P = P_0 + k_1 \log t + k_2 \log^2 t \qquad (12)$$

$$P = b_1 \exp\left\{-\left[\left(\frac{t + b_2}{b_3}\right)^{b_4}\right]\right\} + b_5 \qquad (3)$$

None of the mathematical models used in the referenced studies are of the mechanistic type – that is, they are generally not based on prestated fundamental considerations. This fact is significant. In any predictive aging test, the ideal goal is to achieve a model built as nearly as possible on fundamental physical principles. In those rare cases where considerable aging data are available on a particular material or device under normal–stress conditions, the search for an appropriate model is made somewhat easier. Moreover, one can consider in the experimental design the degree of "statistical bias" against one or more models. On the other hand, where such normal–stress data are not available, the choice of realistic physical models at the test design stage is made difficult, especially where the allowed period of the accelerated test is restricted. Under restrictive time limits, iterative full–factorial test designs,

although they possess desirable features, often have to be ruled out. A significant number of full-factorial experiments are never completed because of the extended time required to explore a complex model with many variables.

In summary, the best published aging studies found to date exhibit serious deficiencies relative to the goals of this program. As a result, building and verifying an aging model from available literature data is not possible. What can be done is to consider the various types of deficiencies and design a better aging test to overcome the deficiencies. In spite of the shortcomings of the studies, they indicate that an accelerated aging study, if properly designed, can predict the change in at least some properties under actual outdoor-aging conditions. The various property-versus-time models found in the literature thus provide useful background information for this study.

PAST STATISTICAL APPROACHES TO ACCELERATED TESTING

The statistical approach to accelerated testing is well summarized in Chapter 9 of the 1974 textbook by Mann, Schafer, and Singpurwalla.[81] Among the 30 references at the end of the chapter dealing with accelerated testing, approximately half are directly related to accelerated testing, and half provide background information in statistics and reliability. The emphasis of the following review is placed on optimizing elements of the experimental design of predictive aging tests such as stress levels and allocation of test specimens among the stress levels, not on a general review of reliability experiments. Suggested avenues of approach for improving the optimal procedures are given at the end of this section.

In statistical terms, the problem of accelerated testing is expressed in Reference (81) as follows. Let $f(t; \theta)$ denote the probability density function of t, the time-to-failure of a system, with θ denoting a vector of parameters. It is assumed that the numerical values of the parameters may depend on the stress levels, but the functional form for $f(t; \theta)$ does not. In this relation, the accelerated test problem is viewed as one of statistical estimation: how should the parameters be estimated using data obtained from accelerated life tests conducted at higher-than-normal stress levels in order to predict the parameter values (and hence the specific probability density function) that would hold under a normal stress condition?

Specific estimation procedures are summarized for the case in which the times-to-failure are assumed to be exponentially distributed at a given stress level:

$$f(t; a_i) = a_i \exp(-a_i t), \quad a_i > 0, \quad t \geqslant 0 ,$$

where a_i denotes the reciprocal mean time to failure at each of K stress levels S_i, $i = 1, \ldots,$ K. Several postulated relationships between the parameter a_1 and stress level S_i are treated:

Power Rule[82]: $\quad a_i = A/S_i{}^B, \quad A > 0$

Arrhenius Model[83]: $\quad a_i = \exp[A - (B/S_i)]$

Eyring Model[84]: $\quad a_i = AT_i \, [\exp(-B/kT_i)]$

$$\bullet \, \exp[CS_i + (DS_i/kT_i)] ,$$

where A, B, C, D denote parameters to be estimated, and in the Eyring model T_i denotes a thermal stress, S_i denotes a non-thermal stress, and k denotes Boltzmann's constant. These models have found application to accelerated testing of electronic component parts.

In general, the resulting estimation procedures are non-linear, and require that attention be directed toward re-parameterizing the models to achieve asymptotic independence under large-sample theory. Large-sample theory is concerned directly with the question of how many stress levels are required to obtain good estimation. For the power rule and the Arrhenius model, this question is examined using maximum relative likelihood functions for each parameter to be estimated. If these functions are found to be approximately normally distributed, then inferences concerning mean time to failure under normal-stress conditions are statistically justified. Based on computer simulated data, approximate normal distributions were obtained for the maximum relative likelihood functions for parameters A and B of both the power rule

and the Arrhenius model for K as small as 10. For these data, as few as five stress levels provided approximate normality for the Arrhenius parameters; the B parameter of the power rule, however, showed some asymmetry with only five stress levels.

These estimation procedures are also applied to the case in which the exponential distribution of times to failure has a positive location parameter γ:

$$f(t; a_i; \gamma_i) = a_i \exp[-a_i(t-\gamma_i)], \ a_i > 0, \ \gamma_i > 0, \ t \geqslant \gamma_i \quad ,$$

where t must be greater than γ_i, and γ_i is assumed to depend on the stress level S_i as follows:

$$\gamma_i = \alpha - \beta S_i \quad ,$$

with α and β parameters that must be estimated. Again, the results are highly complex, with heavy reliance on asymptotic results. A numerical example is presented based on simulated data for the Arrhenius model using five stress levels and 15 specimens per stress level, with the test continued at each stress level until nine failures occurred. The unbiased estimations of the failure rates at each stress level were found to be within 10 percent of the known values used in the simulations, except for the lowest stress level.

The parametric results described above require strong assumptions. When such assumptions cannot be made, the statistical approach typically seeks non-parametric methods that involve weaker assumptions. One non-parametric method for accelerated testing is presented in Reference (81). However, according to the authors, this method requires the existence of life data in the unaccelerated mode. Such data usually do not exist. Morever, the objective of accelerated life tests is to eliminate the need for obtaining such data.

Other statistical approaches include an accelerated test that exposes a system to a stress that increases continuously over time until failure occurs. The work of Allen[85] is relevant to this case, and his procedures are applied to the power rule model in Reference (81). It is shown, for example, that a Weibull hazard function results from a stress that increases linearly over time.

The actual design of accelerated life tests is treated in Reference (81) only in terms of Zelen's work[86], which emphasizes the desirability of using factorial designs when several different kinds of stresses are important. The tractability of the analysis is improved if the stresses do not interact. In practice, however, interactions among stresses tend to be the rule rather than the exception. The effects of most non-thermal stresses are usually altered by changing the operating temperature. Thus, temperature usually interacts with every non-thermal stress, and thereby greatly complicates the conceptual design of an accelerated test.

Number and Spacing of Stress Levels

In 1964 Hoel and Levine[87] solved a problem of determining at what points data should be taken and what proportion of the observations should be taken at each point in order to minimize the variance of the predicted value of a polynomial regression curve at a specified point extrapolated beyond the range of the observations. This work appears to offer a theoretical basis for improving the precision of a prediction at an extrapolated point. As a numerical example of the kind of results obtained, the authors present a case in which 52 observations

are to be taken at four points in the interval (-1, 1). A conventional approach would require an equal allocation of 13 measurements to be taken at four equally spaced X-values denoted by $X_i = -1, -1/3, 1/3$, and 1. For a maximum precision extrapolation at $X = 2$, however, Hoel and Levine show that an optimal allocation of the 52 measurements would require that 5, 12, 20, and 15 measurements be taken at the points: $X_i = -1, -1/2, 1/2, 1$, respectively. Note that the optimal extrapolation procedure requires both unequal spacing for the X-values with an unequal allocation of measurements assigned to those X-values. In general, the optimal allocation and optimal spacing are functions of the distance of the extrapolated point from the measurement interval. Thus, if optimal extrapolation were required for $X = 4$, a different allocation and spacing would be obtained. In general, the optimal spacing is given by the Chebychev points:

$$X_i = \cos i\pi/K, \; i = 0, 1, ..., K \qquad , \qquad (14)$$

for $K + 1$ observation points in the interval (-1, 1). The associated optimum number of measurements to be taken at X_i is an integer approximation to the value of np_i, where n denotes the total number of measurements to be taken, and the proportions p_i are given by

$$p_i = | L_i(x) | \Big/ \sum_{i=0}^{K} | L_i(x) | , i=0, 1, ..., K \qquad , \qquad (15)$$

where $L_i(x)$ denotes a Lagrange polynomial given by

$$L_i(x) = \frac{(x-x_0) \cdots (x-x_{i-1})(x-x_{i+1}) \cdots (x-x_K)}{(x_i-x_0) \cdots (x_i-x_{i-1})(x_i-x_{i+1}) \cdots (x_i-x_K)} \qquad . \qquad (16)$$

The Langrange polynomials are used to provide an exact polynomial fit $\hat{y}(x)$ to each average value, say $\bar{y}_i$, obtained from np_i measurements taken at the associated value of x_i. The fitted polynomial equation is given by the following expression:

$$\hat{y}(x) = \sum_{i=0}^{K} L_i(x)\bar{y}_i \qquad . \qquad (17)$$

The minimal variance of the predicted value of $\hat{y}$ corresponding to the point x is given by the expression

$$V[\hat{y}(x)] = (\sum_{i=0}^{K} | L_i(x) |)^2 (\sigma^2/n) \qquad , \qquad (18)$$

where n denotes the total number of measurements to be taken and σ^2 denotes the variance assumed to hold for the y-measurements taken at each x_i.

For the numerical example, the specific Lagrange polynomials are found to be given by the following expressions:

$$L_0(x) = -(2/3)\,(x-1)\,[x^2-(1/4)] \quad ,$$

$$L_1(x) = (4/3)\,[x-(1/2)](x^2-1) \quad ,$$

$$L_2(x) = -(4/3)\,[x+(1/2)]\,(x^2-1) \quad ,$$

$$L_3(x) = (2/3)\,(x+1)\,[x^2-(1/4)] \quad .$$

It may be noted that $L_0(x)$, $L_1(x)$, $L_2(x)$, and $L_3(x)$ are constructed in such a way that the respective polynomials take the values $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, and $(0, 0, 0, 1)$ at the x-values $(-1, -1/2, 1/2, 1)$. It is this property that results in an exact polynomial fit to the average values $\bar{y}_i$. Evaluation of these polynomials at $x = 2$ yields the following absolute values:

$|L_i(2)| = (5/2, 6, 10, 15/2)$, so that $\Sigma\,|L_i(2)| = 26$ and the p_i-values are seen to be $(5/52, 12/52, 20/52, 15/52)$. The variance of the predicted value at $x = 2$ is then found to be given by

$$V[\hat{y}(2)] = (26)^2(\sigma^2/52) = 13\,\sigma^2 \quad .$$

The variance of the predicted value, based on equal allocation and spacing, is obtained by first constructing new Lagrange polynomials for the non-optimal spacing. The above procedure then yields approximately $20\,\sigma^2$ for the resulting variance. The non-optimal allocation and spacing thus yields a variance that exceeds the minimum variance by more than 50 percent. This reduction in variance yields a corresponding reduction in confidence intervals at the extrapolated point.

In the context of accelerated testing, the preceding numerical example would be equivalent to coding four stress levels so that the highest stress level has a coded value of -1, the next highest stress has a coded value of $-1/2$, etc. On the basis of these coded values, the normal stress condition would correspond to $x = 2$. The optimal allocation would require that five modules be tested at the stress level corresponding to $x = -1$, 12 modules be tested at $x = -1/2$, etc. The test data would then provide an average degradation rate, say $\bar{y}_i$, experimentally determined for each of the four stress levels. The third-degree Lagrange polynomial would then be fitted to the four average degradation rates. Finally, the fitted polynomial would be evaluated at the extrapolated point $x = 2$ in order to obtain the predicted minimum variance (maximum precision) degradation rate for normal conditions.

Mann[88] made important generalizations of these procedures and applied them to accelerated testing by considering measurements assumed to have a two-parameter Weibull distribution at each stress level:

$$F(t) = \begin{cases} 1 - \exp(-(t/\delta)^{1/b}), & t \geq 0 \\ 0, & \text{otherwise;} \end{cases}$$

with the parameters δ and b restricted to positive values. The shape parameter, b, is assumed to be constant over all stress levels, whereas the scale parameter, δ, is assumed to depend on stress level. In addition, it is assumed that the logarithm of the failure time is a polynomial function of known degree, K, of reciprocal stress $1/\sigma$.

Following Mann's notation the optimal spacing is given by

$$1/\sigma_i = \left\{-\cos\ [(K-i)\pi/K]\right\}\ (\frac{1}{c} - \frac{1}{d})/2 + (\frac{1}{c} + \frac{1}{d})/2,\ i = 0,\ 1,\ ...,\ K\ ,$$

where K + 1 stress levels σ_i are to be assigned between minimum and maximum stresses denoted by c and d, respectively. This result corresponds to the Chebychev points given by Equation (14). The optimal proportions of the n measurements to be allocated to each stress level are found to be

$$p_i^+ = p_i\ [1-0.43\ (K + 1)/n] + \frac{0.43}{n}\ ,$$

where p_i are given by Equation (15), and the variance of the predicted value is approximately given by

$$V^+[\hat{y}(x)] = V[\hat{y}(x)]\ /\ [1-0.43\ (K + 1)/n]\ ,$$

where $V[\hat{y}(x)]$ is given by Equation (18). The Lagrange polynomials given by Equation (16) are also used by Mann to obtain a polynomial fit to the data. However, the arithmetic mean value $\bar{y}_i$, shown in Equation (17), is replaced by a more general weighted mean. The weights for this mean are obtained from the elements of the inverse of an associated covariance matrix.

Little and Jebe[89] have also applied the results of Hoel and Levine to problems of extrapolation in fatigue testing. Their formulations have included unequal variances at different stress levels and constraints on experimental material, time, and cost. A three-parameter model frequently used in fatigue testing is assumed:

$$N = a(S-S^*)^\beta \tag{19}$$

where N denotes the fatigue life in cycles, S represents an alternating stress amplitude, and α, β, and S^* are parameters. A linear form in α and β is obtained by taking logarithms. If S^* is assumed known, the model is seen to be identical to the power rule model discussed as a model for accelerated testing in Reference (81). The following results are obtained for a case in which only two stress levels are to be tested with either time or cost, t, constrained by the relation:

$$t = a/s^\beta\ ,\ \beta > 0\ , \tag{20}$$

where α and β denote parameters, and s = (S-S*). It is also assumed that the variances are stress dependent as shown by the relation:

$$\sigma^2 = 1/s^\delta\ ,\ \delta > 0\ , \tag{21}$$

where δ denotes a parameter. Under these assumptions the stress levels selected for testing must satisfy the relation:

$$[(\beta/2) + \delta]\ \ell n\ (s_2/s_1) = 1.2784\ , \tag{22}$$

where $s_1 < s_2$, and the numbers of specimens, K_1 and K_2 tested at s_1 and s_2, respectively, must satisfy the relation:

$$(K_1/K_2) \, [\ln(s_1/s_0)/\ln(s_2/s_0)] = (s_2/s_1) \, (-\beta/2)-\delta \, , \qquad (23)$$

where $s_0 < s_1 < s_2$ and s_0 denotes the low stress level at which optimal extrapolation is desired.

A particularly interesting finding is the result that under some conditions the precision obtained by extrapolation exceeds the precision that would be obtained by direct testing under the low–stress condition. This results from the decreasing variance among measurements taken at high stress levels together with the fact that long test times at low stress permit relatively few specimens to be tested because of cost constraints.

As a numerical example, let the high stress measure be 100 so that $s_2 = 100$. Suppose that the cost of experimentation is inversely proportional to the fourth power of the stress measure, $t = a/s^4$, so that increasing the stress measure by a factor of 2 will yield a decrease cost factor of $1/16$. Finally, suppose $\delta = 0$ so that the variability of the measurements is independent of the stress level. Substitution of these values into Equation (22) shows that $2\ln(100/s_1) = 1.2784$. It follows that $s_1 = 52.8$, and the low–stress measure for experimentation is equal to 52.8. Equation (23) then yields

$$(K_1/K_2) \, (\ln(100/s_0) \, / \, \ln(52.8/s_0)] = (100/52.8)^{-3} \qquad .$$

Suppose it is desired to extrapolate to a normal–stress measure of $s^0 = 20$. The preceding relation then shows that

$$\begin{aligned} K_2/K_1 &= [\ln(100/20) \, / \, \ln(52.8/20)] \, (100/52.8)^3 \\ &= 11.26 \qquad . \end{aligned}$$

Thus, for every specimen tested at the lower stress level $s_1 = 52.8$, the optimal design requires that approximately 11 specimens be tested at the higher stress level, $s_2 = 100$, in order to minimize the prediction variance at a stress level corresponding to $s = 20$. The authors indicate that the variance of the extrapolated prediction is only 17.8 percent as large as the variance would have been if the specimens had been allocated equally at s_1 and s_2.

Blight[90] considers the problem of minimizing the cost of experimentation when the number of specimens that can be tested at any given time is restricted. He determines how many failed specimens should be replaced during a life test that is continued until a specified number of failures have occurred. It is assumed that the tests to failure are exponentially distributed, so that the results also apply to a Weibull distribution with a known shape parameter.

A partial replacement procedure is defined as follows. Suppose that n specimens are put on test, and that the first r that fail are replaced but subsequent failures are not replaced. The experiment ends as soon as the rth failure occurs, so that $k + 1 \leqslant r \leqslant n + k$. The expected cost of the experiment is found to be given by

$$c(n, k) = (n + k)C_2 + [(k/n) + \sum_{j=1}^{r-k} (n-j+1)^{-1}]\theta C_1 \, , \qquad (24)$$

where C_1 denotes the cost of running the experiment per unit time, C_2 denotes the cost of putting a specimen on life test, and θ denotes the mean time to failure parameter of the exponential distribution: $\theta^{-1} \exp(-t/\theta)$, $t > 0$. Finally, it is assumed that N denotes the maximum number

44

of specimens that can be simultaneously tested. It follows that n and k must satisfy the following inequalities: $0 \leq k \leq r-1$, $1 \leq n \leq N$, $n + k \geq r$. It is important to note that the precision of the estimation procedure depends on r but not on n or k. Thus, r can be chosen as part of the experimental design.

The optimal values of n and k are summarized by the following results:

(1) If $r \leq 1 + N^2/(\alpha + N)$, then n*, the optimum number of specimens that should be tested, is given by the integer part of the following expresssion:

$$n^* = (1/2) [r + (r^2 + 4\alpha r)^{1/2}] \quad , \tag{25}$$

where $\alpha = C_1\theta/C_2$, and for this case no replacements of failed specimens should be made, so that k* = 0.

(2) If $r > N^2/(\alpha + N)$, then n* = N and the optimum number of failed specimens k* that should be replaced is given by the integral part of the following expression:

$$k^* = [r - N^2/(\alpha + N)] \quad . \tag{26}$$

Examination of this expression shows that every specimen should be replaced if $N(N-1) < a$.

As a numerical example suppose that a maximum of 12 specimens can be simultaneously tested. Suppose further that $C_1 = 1$, $C_2 = 100$, $\theta = 1000$, and that testing will cease when 10 failures have occurred. With these values it is seen that $N^2/(\alpha + N) = (12)^2/(10 + 12) = 6.55$ and this is less than r = 10. Thus, the second case described above holds and it follows that 12 specimens should be put on test and that the optimal number of replacements is given by k* = (10 – 6.55) = 3. The expected cost for this life test procedure is obtained from Equation (24):

$$C(12, 3) = (15) (100) + [(3/12) + \sum_{j=1}^{7} (13-j)^{-1}] (1000) = 2570 \quad .$$

For comparison, the expected cost for no replacement procedure is given by C(12, 0) = 2603, and for a complete replacement procedure is given by C(12, 9) = 2933.

As indicated by Equation (26) the optimal number of replacements k* depends on $\alpha = C_1\theta/C_2$ which, in turn, depends on θ the mean time to failure. Thus, k* cannot be computed until the mean time to failure is known or assumed.

Izenman and Rinott[91] propose two sequential procedures to circumvent the fact that prior to experimentation θ is typically unknown. In these procedures it is decided before experimentation that at least the first m failed specimens will be replaced. More than m replacements may or may not be made depending on the outcome of the sequential decision rule given below. As in Blight's development, the experiment is assumed to end when r failures have occurred. The magnitude of r determines the precision of the statistical inference procedures involved in confidence interval estimations or in testing statistical hypotheses concerning θ. To achieve the required precision, it may be necessary to choose r > N so that at least r – N replacements must be made. This means that if r > N, then m must be chosen so that $m \geq r-N$; if r < N, then $m \geq 0$.

After m specimens have been successively failed and replaced, the experiment continues until the $(m+1)^{st}$ failure occurs. It must next be decided whether or not the $(m+1)^{st}$ failure should be replaced. To make this decision, the maximum likelihood estimator of θ is computed by dividing the total accumulated lifetimes of all N specimens by $(m+1)$. With this value of $\hat{\theta}$ it is possible to compute $\alpha = C_1\hat{\theta}/C_2$ and substitute the result into Equation (26) to yield an estimate of the optimal number of replacements. If the current number of replacements is less than, or equal to, the optimal number, $\hat{k}_{m+1}$, then the $(m+1)^{st}$ failure is replaced; otherwise no further replacements are made. In either case the experiment is continued until a total of r failures have occurred.

A variation of this procedure requires that when $(m+1)$ failures have occurred and m replacements have been made, the number of additional replacements is computed once and for all to be

$$\hat{k}_b = \max(m, \hat{k}_{m+1})$$

A numerical example given by the authors is extracted from Bell Laboratory field data for lifetimes of a semiconductor device. A total of 15 modules are tested simultaneously with each module consisting of 24 sixteen-device packs for a total of 384 semiconductor devices. To reduce cost, a pack of 16 devices is replaced whenever a device within the pack fails, instead of replacing an entire module. For the example, r was set equal to 20 so that r–N = 5. Thus, m was taken equal to five and the first five failures were replaced with certainty; decisions regarding subsequent failures were made after each failure occurred. The cost structure was represented by $C_1 = 2$ and $C_2 = 100$. The first five failures occurred at times 0, 72, 120, 312, and 336. The sixth failure occurred at time 528. The mean time between failures was then estimated by $\hat{\theta} = (15)(528)/6 = 1320$, and the optimal number of replacements was obtained from Equation (26): $k^* = \{20 - 225/[(0.02)(1320) + 15]\} = (14.57)$, so that the optimal number of replacements was estimated to be 14. Thus, the sixth failure was replaced and experimentation was continued. At the seventh failure, k^* was found to be 15, and again replacement was made. The process continued until 16 failures occurred. At that time, the computed optimal replacement number remained at 15 so that the 16th failure was not replaced. The experiment was continued without further replacements until 20 failures occurred. The final value of $\hat{\theta}$ was found to be 2202 hours at a corresponding cost of $C = 2(3/20) + 100(15 + 15) = 9240$.

A number of computer simulation runs suggest that the sequential procedures are nearly as efficient as Blight's procedure which requires that θ be known.

Application of Statistical Approach to Experimental Design

The results given in the preceding section are illustrative of the kinds of concepts and results obtained from a statistical approach to the design of accelerated tests. Under specified assumption, these results are primarily related to various "local" aspects of the design. For example, given a Weibull distribution of times to failure, how should the parameters be estimated? Given an exponential distribution of times to failure with a mean time to failure related to a constant thermal stress by an Arrhenius model, how should the data be analyzed? Although these are important and useful results, there are many shortcomings between them and the needed "global" design that would specify the number and spacing of stress levels,

number of modules to be tested at each stress level, required precision of measurements, etc. The aim of this section is to better identify these shortcomings in order to make recommendations on how they may be reduced.

Determination of Number of Test Stress Levels

In Reference (92), a minimum of five stress levels is recommended for accelerated testing. Ideally, those stress levels would range from the lowest stress consistent with constraints on the duration of the test to the highest stress consistent with the requirement that the dominant failure mode at this stress is the same as for the normal-stress level. All five stress levels are viewed as higher than normal with the same dominant failure mode. The fundamental hazard of accelerated testing stems from the possibility that the dominant failure mode obtained at the higher-than-normal stress levels will be different from the dominant failure mode that holds at the normal-stress condition. The maximum stress that can be used is usually unknown prior to experimentation. Thus, a guess is necessary, and the data must subsequently be analyzed to determine whether the guess was correct. That is, the structure of the data must be examined specifically to determine whether or not a change has occurred in the dominant degradation mechanism. Only one failure mechanism is wanted at all five stress levels. The rate of degradation of this failure mechanism is expected to increase as the severity of the stress increases, but no change is desired for the mechanism itself.

The examination to determine whether or not a change in mechanism has occurred is frequently aided by plotting the degradation rate against a suitable measure of stress. Linear plots may serve as evidence that no change in failure mechanism has occurred. For example, if the logarithm of the observed degradation rate is plotted versus reciprocal temperature and is found to be linear, then an Arrhenius-type relation holds. This, in turn, suggests that over the tested range of stress levels, no change in mechanism has occurred; only the rate of degradation has changed.

The reasons for choosing five stress levels can be elaborated as follows. Note that one higher-than-normal stress level is not sufficient because one stress level gives rise to a single point on a plot of degradation rate versus stress. Such a point may be represented by the coordinates $(\mathring{q}_1, S_1)$, where $\mathring{q}_1$ denotes an average time rate of degradation of quality at a stress level of magnitude S_1. It is desired to extrapolate from this point to obtain the coordinates of the point corresponding to normal stress conditions, say $(\mathring{q}_o, S_o)$. The magnitude of S_o can be stated, but the associated magnitude of $\mathring{q}_o$ cannot. In fact, the purpose of conducting the accelerated test is to obtain an estimate of $\mathring{q}_o$ by testing at higher-than-normal stress levels. Because extrapolation from a single point is not possible, it follows that at least two stress levels must be tested. However, two points permit only a straight line extrapolation and provide no basis for testing the linearity of the relation between $\mathring{q}$ and S over the range of stress levels used in the test. Thus, to test for linearity over the stress range, at least three points are needed. This means at least three stress levels are required. Now, if the three points are not found to yield a linear relation, even under reasonable continuous transformations, then it may be assumed that a change in mechanism has occurred. That is, the highest stress level may be too high. This result, in turn, threatens the interpretation of the intermediate stress. Should the results of the intermediate stress be associated with the high-stress condition, the low-stress condition, or with a transition state in which two competing failure mechanisms are equally dominant? To answer these questions successfully requires that two more stress levels be included, giving a minimum of five stress levels. An extrapolation to the normal-stress condition can then be accomplished provided that at least the lower two stress levels appear to

involve the same failure mechanism. This approach provides some protection against setting the high stress levels too high. The two highest stress levels, for example, may be found to be associated with the wrong failure mechanism, and the intermediate stress may represent a transition condition. Even in this case, an extrapolation could be carried out to the normal-stress condition. If more favorable results were obtained from this experimentation, with the lower three, four, or all five stress levels showing the same failure mechanism, then the extrapolations could be made with increasing confidence, especially if the lowest test stress is not much higher than the normal-stress condition.

Some additional support for the use of a minimum of five stress levels has been given by Mann, Schafer, and Singpurwalla.[81] As noted previously, the estimation procedures for the Arrhenius and Eyring models showed for some simulated data that five stress levels yielded symmetric forms for the maximum relative likelihood functions. An example for the power-rule model showed that a minimum of ten stress levels may be needed to estimate one of the model parameters. In Hoel and Levine[87], the degree of the extrapolating polynomial is assumed to be known. In the context of accelerated testing, this is equivalent to assuming that the number of stress levels that should be tested is also known. Thus, the Hoel and Levine approach provides a direct basis for determining the optimum extrapolation procedure, given the degree of the regression polynomial; the approach does not provide a basis for determining the optimum degree. For this reason, the results of Mann[88] obtained for the Weibull distribution do not include the determination of the optimum number of stress levels. Similarly, the results of Little and Jebe[89] appear to be based on the assumption that no change in the failure mechanism occurs so that only two stress levels are required.

In summary, these fragmentary results appear to be generally consistent with a recommendation of at least five stress levels. However, no direct development of the optimum number of stress levels has been found in the literature.

Determination of Number of Test Modules

McCallum, Thomas, and Waite[92] recommend at least 15 modules be tested at each stress level, with two randomly selected modules removed at five equally spaced times during the expected duration of the test. Under this procedure, five modules would be expected to complete the experiment. One of the randomly selected cells would be subjected to a destructive tear-down test in which the most precise and sophisticated measurements would be used to identify all changes in material or performance properties. These diagnostic measurements would aim to reveal precursors of failure and give the earliest possible evidence of failure mechanisms and their associated degradation rates. The second module removed from test would be available for special purpose tests. For example, a module may be removed from test and then replaced on test after changes have been made in design, fabrication, or maintenance. To test sequential effects of different stress levels, a module may be removed from the test at one stress level and be placed on test at a different stress level, etc.

The primary reasons for requiring the number of modules to fall between 15, at the beginning of the experiment, and five, at the end of the experiment, is to achieve approximate normality for the average degradation rates. With fewer than five observations, the distribution of such averages cannot be deduced directly from the data, nor can it be inferred by appeal to the central-limit theorem.

In their treatment of accelerated testing, Mann, Schafer, and Singpurwalla[81] recognize the importance of the number of stress levels and the number of modules tested at each stress level. In particular, the shape of the maximum relative likelihood function depends on these experimental design parameters. It is suggested that the desired symmetry of the likelihood function may be achieved by manipulation of the parameters, but the suggestion is not implemented.

It is interesting that the extrapolation procedures of Hoel and Levine[87], and the resulting applications made by Mann[88], and Little and Jebe[89] determine the optimal allocation of M modules among K stress levels. However, the procedures do not explicitly require any minimum number of modules at each stress level. Presumably, M must be chosen to be sufficiently large so that the optimal allocation results in at least one module on test at each stress level. Also, it should be noted that the precision of the extrapolation depends on σ^2, the variance of measurements at each stress level. In practice, this variance is usually not known, and consequently, must be obtained from the data. This suggests that the optimal allocation should be further constrained to provide a reasonably good estimate of the variance. If the variance is allowed to differ from one stress level to another, then sufficient modules must be tested to provide an acceptable estimate of the variance for each stress level using only the data determined at that stress level.

The partial replacement schemes of Blight[90] and Izenman and Rinott[91] give the optimal number of replacements of failed modules in order to reduce the expected experimental costs. However, the number of items that can be tested at each stress level is assumed to be an input to the optimization; it is not an output. Thus, again, no direct calculation of the required number of modules is given.

RESULTS AND TECHNICAL DISCUSSION — PART II: DEVELOPMENT OF ACCELERATED TEST METHODOLOGY

RESULTS AND TECHNICAL DISCUSSION —
PART II: DEVELOPMENT OF ACCELERATED
TEST METHODOLOGY

Methodologies used in the past for designing aging prediction tests have encompassed several concepts and philosophies. In many cases involving the predictive behavior of single materials in a given environment, the experimental design of the aging tests is based on an assumed mathematical model derived from relevant physical processes. Iterative experiments are designed and implemented to estimate the parameters in the model. As noted in Part I of the Technical Discussion, this approach has been successful, at least in some degree, and has statistical appeal. Unfortunately, in many cases the experiments were never finished primarily because of excessive time and costs.

In this study, the device in question is not a single material; the photovoltaic module, whatever its "final" design and configuration, has many interfaces, many materials, and is intended to operate in widely different environments. Consequently, designing aging experiments around an assumed single aging process is not recommended. Instead, the recommended approach involves the conceptional treatment of a full factorial design, but relies heavily on engineering judgment to prune the factorial design to fewer tests. Further conceptional iterations adjust the final design to upgrade the statistical quality to the extent practical. In particular, the number of levels at each stress to be tested experimentally must be sufficiently large to protect against developing an incorrect (biased) model from the experimental results. Thus, the goal was to develop a methodology for designing a reliable test that can be carried out within practical constraints of time and costs.

It must be emphasized that implementation of the recommended methodology, as in any methodology for predictive accelerated tests, is not a routine task. Substantial contributions are required from statisticians, material scientists, and testing engineers.

The methodology developed in this study is presented in three sections. The first covers selection of stresses, number of stress levels, and the allocation of modules among the stress levels; the second discusses instrumentation of the design, and the third presents data analysis methods. Following the third section, the steps involved are summarized. In that part of the methodology described in the first section (stress levels, etc.), an example is used to illustrate the recommended procedure. An illustrative module is proposed and its materials identified. Obviously, the experimental design is material specific, so the test design would change with any module design change. Nevertheless, the chosen example for the module design has many features believed to be representative of modules that may be fabricated in the near future.

In any full example of a test design, the instrumentation of the design (properties measured and with what instrumentation) would be integrated into the selection of stresses, levels, and specimen allocation. In fact, this point is introduced at appropriate places in the first section of the methodology development. Such an integration requires sensitivity and cost information on the available instruments. Program time and cost constraints precluded obtaining this information in the present study. Additionally, expending such an effort for a hypothetical module design (even though the design is reasonable) is probably not justifiable. Therefore, the test design example was carried through for only the first part of the experimental design. The separate section on instrumentation treats the subjects of precision/sensitivity and costs in selecting the instrumentation.

DESIGN OF ACCELERATED-TEST
STRESS LEVELS AND SPECIMEN ALLOCATIONS

The review of past approaches to accelerated testing shows that especially detailed statistical investigations have been made by Mann[88] and Singpurwalla[93] for selected models related to the Weibull distribution and to the Arrhenius and Eyring relations. Most investigations restrict attention to the case of a single stress. This restriction often precludes extrapolation to normal-stress conditions, especially for photovoltaic modules that are exposed to a variety of stresses associated with temperature cycling, relative humidity, ultraviolet radiation, and pollutants such as sulfur dioxide. Because accelerated testing involves extrapolations from overstress to normal-stress conditions, the optimal extrapolation procedures of Hoel and Levine[87] appear to be especially relevant to the design of accelerated tests. In the present development, this procedure is extended to the case of several simultaneous stresses. To date, an overall methodology for developing an optimized accelerated test program for several stresses has not been found in the literature. The present report extends the state of the art, and identifies problem areas that remain to be solved.

Overview of the Design Approach

The methodology developed in this study for the design of accelerated tests is first described in general terms and then is applied to a specific example of a module. The methodology is carried out in several stages. The first stage involves the extraction and quantification of engineering judgment relevant to a particular degradation mechanism which is expected to occur under normal operating conditions. The proposed procedure is based on explicit consideration of all combinations of stress levels that could be used in the test program. However, not all combinations would be tested. An explicit procedure based on a complete factorial experimental design is proposed. The procedure is intended to aid in quantifying engineering judgment and to provide explicit documentation for the judgmental processes that, in turn, yield the stress combinations to be used for testing. Because of the judgmental basis of this procedure, several iterations may be required in order to obtain a satisfactory set of test conditions. The selected test conditions are represented in the form of a hierarchical tree in which the stress combinations to be tested are associated with the terminal points of the branches of the tree. In summary, the first stage requires:

(1) Conceptual examination of combinations of those stresses believed to be relevant to the normal operating conditions and potentially usable in accelerated testing

(2) Selection of a subset of these stresses for actual testing

(3) Representation of the selected test conditions as a hierarchical tree.

In addition to identifying the higher-than-normal stress levels to be used in the accelerated testing, the user must also numerically code these stress levels so that +1 and -1 correspond to the highest and lowest stress levels for each type of stress, such as thermal or mechanical. The same coding procedure is then applied to obtain a corresponding quantitative representation of the normal-stress conditions.

The second stage of the design process consists of representing the hierarchical tree by means of a multivariate Lagrange polynomial that is a generalization of that used by Hoel and Levine, as discussed in Part I of the discussion. This representation shows how the experimental design is quantitatively dependent on the normal-stress condition. The algebraic representation is then used to optimize the selection of the number of accelerated stress conditions, the spacing between the accelerated test conditions, and the number of modules to be tested at each stress combination. The total number of modules to be tested and the degree of the polynomial representing the hierarchical tree can be numerically explored to obtain the most appropriate final design.

The proposed procedure extends and makes practical the use of research results described earlier. The factorial basis for experimental designs of Zelen[86] is used in the first stage to extract and quantify recent engineering information. The concepts associated with the extrapolation procedures of Hoel and Levine[87], used subsequently for accelerated test design by Mann[88], and Little and Jebe[89], are extended and applied to accelerated test design by several different kinds of stresses. The innovation that underlies the proposed approach consists of using a factorial layout to arrive at selected test conditions represented by a hierarchical tree, which, in turn, can be represented by a multivariate Lagrange polynomial that can be optimally extrapolated by procedures that are generalizations of those given by Hoel and Levine.[87]

Extrapolation and Quantification
of Engineering Judgment

For the first stage of the proposed methodology, engineering input is essential. Even fragmented engineering knowledge can be quite effective in improving the validity and reducing the costs of accelerated testing. The outline below suggests activities needed to provide input to the proposed approach.

- Identify and quantify all environmental conditions believed to affect the expected life of the module when operated under intended-use conditions. If possible, identify anticipated repair and maintenance procedures and their associated schedules over the expected life of the module. If field operations in different geographic locations are anticipated, then the environmental conditions for these locations must be summarized. Ideally, these summaries will be based on historical data obtained at the geographic locations of interest. These conditions will constitute the normal stress conditions.

- Construct a table that lists every module element that may be a source of degradation of performance over time. Consider each element to be operated under normal operation conditions for the expected lifetime of the module. Partition the module into mutually exclusive geometric elements, such as volumes, surfaces, interfaces, interconnects, etc.

- Identify, by appropriate tables and drawings, all specific design characteristics, materials, fabrication processes, etc., believed relevant to the operational life of the modules to be tested.

- For each element of the module, identify all thermal, mechanical, chemical, electrical, and radiation stresses that may contribute to degradation of performance when the module is operated under normal-stress conditions. For each of these generalized stresses, it may prove useful to identify a corresponding generalized strain such that the stress-strain product has units of work per unit volume, area, or distance.

- For each generalized stress (or strain), show in a table the possible measurement techniques that could be used to measure associated changes in the performance characteristics of the module.

- Examine the combinations of environmental/operating conditions together with their associated generalized stresses and strains and attempt to identify a single degradation process that is most likely to be responsible for the end-of-life condition of the module. This degradation process is called the dominant failure process.

- To insure that the hypothesized failure process is the dominant one, identify all generalized stresses (thermal, mechanical, chemical, electrical, radiation, etc.) that could affect the degradation rate associated with the dominant failure process. Quantify each of these stresses so that −1 and +1 denote the lowest and highest levels, respectively, that could be used in an accelerated test. That is, the stress level coded as −1 should be the lowest over-stress level that would be expected to yield measurable degradation during the time period allowed for the accelerated test; the stress level associated with +1 should be the highest stress level consistent with the requirement that anticipated module failures at the highest stress level would be due to the same failure mechanism expected to occur under the normal-stress conditions. The initial assignment of codes is to be done for each stress separately; subsequent examinations of combinations of stresses may cause some codes to be changed in order to comply with the above requirements for the combinations of stresses as well as for single stresses.

Development of the Basic Elements of the Test Design
With An Example

The procedure outlined above was applied to a specific example using a module design and fabrication materials representative of solar-cell modules under current development. It must be emphasized, however, that the example is intended to illustrate the methodology for designing accelerated tests. Any change in the module design or fabrication materials would be likely to change the dominant failure mechanisms and to yield a substantial change in the associated accelerated tests.

Figure 3 shows the solar-cell module chosen for illustration, together with the assumed fabrication materials. This design is a modification of one taken from Reference (94).

Identification of Module Elements
and Associated Failure Modes

Table 6 lists module elements, such as the glass cover, adhesives, interconnects, etc. For each design element, the table also lists possible failure modes and the corresponding generalized stresses. The indicated stresses are believed to be those that affect the degradation rates of the associated failure modes. The table shows, for example, that the rate of transparency loss of the glass cover is expected to be affected by stresses associated with ultraviolet radiation, abrasion, and chemical attack. Solar radiation is the source of the ultraviolet stress; wind, dust, and cleaning may be sources of abrasion; and pollution, salt, etc., may be sources of chemical attack. Stresses not associated with a loss of transparency include temperature cycling, mechanical loads, relative humidity, and electrical fields.

- Interconnect
- Barrier (fluorocarbon)
- Back cover (silicone)
- Cell
- Lead
- Adhesive (silicone)
- Front cover (soda-lime glass)
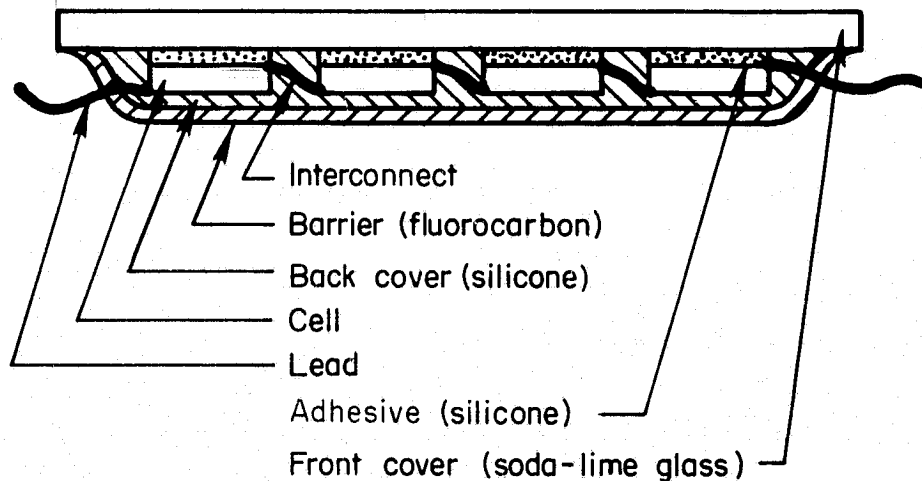
**FIGURE 3. SCHEMATIC CROSS SECTION OF MODULE EXAMPLE
DISCUSSED IN TEXT (Not to scale)**

The last column of Table 6 shows an assessment of the relative importance of each failure mode. The assessments are expressed as numerical codes between 1 and 4, with 1 denoting the least important failure mode. These ratings were obtained by group consensus of five individuals representing scientific disciplines associated with polymers, glass, ceramics, electronics, and physics. The assignment of 4 to a failure mode indicates that it is a potentially dominant failure mode that could be frequently observed under normal operating (field) conditions. Such failure modes could potentially prevent a 20-year life for the solar-cell module. Accelerated tests might be justified to determine whether or not this is the case.

**Identification of the Dominant
Failure Modes**

For long life, the most troublesome design elements listed in Table 6 are expected to involve the interfaces among the silicone, FEP fluorocarbon, and silver/copper leads where the electrical current is delivered from the module. The anticipated failure modes for these interfaces are assumed to be the dominant failure modes under normal (field) operating conditions. These failure modes, shown under design element 13 in Table 6, include water-vapor transmission, delamination, increased resistivity, and shorting. All of these modes are expected to have increased degradation rates under higher-than-normal stress levels for ultraviolet radiation, SO_2 exposure, thermal cycling, and relative humidity. Mechanical stress is associated with water-vapor transmission and delamination, and electrical stresses are associated with delamination, increased resistance, and shorting.

**Initial Identification of Possible
Test Conditions**

Table 7 lists 16 possible combinations of high and low stress levels for each of four stresses. The high and low levels are indicated by H and L, respectively. The first combination shown in the table indicates a test in which the lowest stress levels for ultraviolet radiation (UV), temperature cycling (T), relative humidity (RH), and sulfur dioxide (SO_2) are tested simultaneously.

TABLE 6. ENUMERATION AND ASSESSMENT OF POTENTIAL FAILURE MODES ASSOCIATED WITH MODEL MODULE DESIGN ELEMENTS AND ENVIRONMENTAL STRESSES[a]

Design Element/Failure Mode	Stress[b] UV	Abrasion	Chemical Attack	Temperature	Mechanical Stress	Relative Humidity	Electrical	Assessment of Potential Failure Mode[c]
1. Transparent cover (soda-lime glass)								
a. Loss of transmittance	X	X	X					2
b. Loss of mechanical strength		X	X	X	X	X		1
2. Glass/adhesive interface								
a. Delamination	X			X	X	X		4
b. Loss of transmittance	X			X		X		3
3. Bulk adhesive (silicone)								
a. Loss of transmittance	X			X	X	X		3
b. Loss of mechanical properties	X		X	X	X	X		3
4. Adhesive/cell interface								
a. Loss of transmittance	X		X	X	X	X		3
b. Delamination	X		X	X	X	X		4
5. Adhesive/metallization interface								
a. Bond degradation	X		X	X	X	X		3
b. Increase in series resistance	X		X	X	X	X	X	3
6. Adhesive/interconnect interface								
a. Delamination	X		X	X	X	X	X	1
b. Increase in interconnect resistance			X	X	X	X	X	3
7. Cell/interconnect interface								
a. Loss of electrical contact			X	X	X	X	X	3
8. Cell/bottom–cover interface								
a. Delamination			X	X	X	X	X	1
b. Increase in series resistance			X	X	X	X	X	3
9. Bottom–cover interconnect interface								
a. Delamination			X	X	X	X	X	1
b. Increase in series resistance			X	X	X	X	X	1
10. Bulk back cover (silicone)								
a. Increase in water vapor transmission	X		X	X	X	X		2
b. Degradation of mechanical properties	X		X	X	X	X		2
11. Back–cover/barrier interface								
a. Delamination	X		X	X	X	X		3
b. Decrease in barrier properties	X		X	X	X	X		3
12. Barrier (fluorocarbon)								
a. Increase in water vapor transmission	X	X	X	X	X	X		3
b. Degradation of mechanical properties	X	X	X	X	X	X	X	3
13. Silicone/barrier/lead interface								
a. Increase in water vapor transmission	X	X	X	X	X	X		4
b. Delamination	X	X	X	X	X	X	X	4
c. Decrease in withstand voltage	X	X	X	X	X	X	X	4
14. Glass/barrier interface								
a. Increase in water vapor transmission	X		X	X	X	X		4
b. Delamination	X	X	X	X	X	X		4
15. Glass/back–cover interface								
a. Increase in water vapor transmission	X		X	X	X	X		2
b. Delamination	X		X	X	X	X		2

(a) See Figure 3 for schematic of model module.
(b) An "X" indicates the stress was judged to be effective in causing major degradation.
(c) The severity assessment is based on a scale of 1 to 4; a value of 4 represents the failure modes judged most likely to occur.

TABLE 7. EXPECTED SEVERITY RATINGS FOR A COMPLETE FACTORIAL DESIGN INVOLVING 16 POSSIBLE ACCELERATED TESTS

Combination	Stress Label (UV, T, RH, SO_2)[a]	Initial Rating[b]	Adjustment Factors[c]				Final Rating[h]
			(d)	(e)	(f)	(g)	
1	(L L L L)	4					4
2	(L L L H)	5					5
3	(L L H L)	9			2.0		18
4	(L L H H)	10			2.0		20
5	(L H L L)	13					13
6	(L H L H)	14		2.25			32
7	(L H H L)	18					18
8	(L H H H)	19		2.25		1.5	64
9	(H L L L)	7					7
10	(H L L H)	8					8
11	(H L H L)	12			2.0		24
12	(H L H H)	13			2.0		26
13	(H H L L)	16	1.5				24
14	(H H L H)	17	1.5	2.25			57
15	(H H H L)	21	1.5				32
16	(H H H H)	22	1.5	2.25		1.5	111

(a) See text for definitions of low (L) and high (H) stress condition for ultraviolet radiation (UV), temperature cycling (T), relative humidity (RH), and sulfur dioxide (SO_2).

(b) Initial ratings are assigned without consideration of interactions.

(c) Initial ratings are multiplied by these factors to account for various interactions believed to be important. A blank entry indicates an adjustment factor equal to 1.0.

(d) Accounts for interaction between high UV and high T.

(e) Accounts for interaction between high T and high SO_2.

(f) Accounts for interaction between low T and high RH.

(g) Accounts for interaction between high T, high RH, and high SO_2.

(h) Obtained by multiplying initial rating by each interaction adjustment factor.

With these low levels of stress, measurable degradation for design element 13 should occur within the duration of the accelerated test. The last test combination shown in the table represents the simultaneous application of the high levels of all four stresses to the test modules. These high levels are to be chosen such that the dominant failure modes for element 13 are the same as those for normal (field) operations.

The assumed high and low levels for each of these stresses are:

- Ultraviolet Radiation (UV)
  - The low level of UV radiation consists of the continuous (24 hours/day) application of UV radiation at a flux equal to the historical instantaneous maximum associated with a given geographic location.
  - The high level of UV radiation consists of the continuous application of ten times the flux used in the low UV stress level.

- Temperature Cycling (T)
  - The low level of thermal stress consists of temperature cycling between –40 C and 100 C at a frequency of six cycles per day.
  - The high level of thermal stress consists of temperature cycling between –40 C and 150 C at a frequency of six cycles per day.

- Relative Humidity (RH)
  - The low level for the stress due to relative humidity is 50 percent RH at 100 C.
  - The high level for the stress due to relative humidity is 95 percent RH at a temperature of 100 C.

- Sulfur Dioxide (SO_2)
  - The low level of SO_2 consists of continuous (24 hours/day) exposure at a concentration equal to the historical maximum concentration associated with a given geographic location.
  - The high level of SO_2 consists of a continuous exposure at a concentration equal to ten times that used in the low stress level.

The third column of Table 7 shows numerical ratings assigned to each of the test combinations listed in Column 2. Each rating is intended to reflect the relative severity of the associated test condition, consistent with available engineering knowledge and historical experience. Because such information is extremely limited, considerable subjectivity is required to assign these ratings. It is shown subsequently how the assigned ratings for each of the 16 possible test conditions can be further examined to produce further insight into their general validity. If inconsistencies are detected, then the individual ratings are to be re-examined and adjusted until an acceptable assignment of ratings is obtained. The entire procedure may require several iterations.

The ratings shown were obtained by focusing attention on the relative effects that the stresses would be expected to have on design element 13 under the simplifying assumption that each stress acts independently of the other stresses. After these ratings were obtained, they were then adjusted to better account for the expected interactions among the four stresses. The adjustments were made by multiplying the initial ratings by the factors shown in columns 4 through 8 to obtain the final ratings shown in Column 9.

58

These adjustments involve the following possible interactions:

- High ultraviolet with high temperature cycling

- High-temperature cycling with high concentrations of sulfur dioxide

- Low-temperature cycling with high relative humidity

- High levels of temperature, relative humidity, and sulfur dioxide.

Main Effects and Interactions
Among the Test Variables

Table 8 shows a Yates analysis of the assigned severity ratings given in the last column of Table 7. These are standard calculations customarily used to analyze data obtained from complete factorial experiments.[95] The only liberty taken in these calculations involves the fact that the severity ratings analyzed by this technique represent subjective engineering assessments rather than objective measurements. Such an analysis is useful, however, to indicate which combinations of the test variables are most necessary to include in the experimental design for the accelerated tests. The results may also be used to check the assignments of the ratings given in Table 7.

The analysis of a complete factorial experiment yields quantitative measures of the "main" effects of the variables, and "interaction" effects among all pairs of variables, triples of variables, etc. These are shown in the last column of Table 8. The second entry of the last column shows that the main effect of sulfur dioxide (SO_2) is 23. Reading down the column, it is seen that the main effects of relative humidity (RH), temperature (T), and ultraviolet radiation (UV) are given by 20, 30, and 14, respectively. A comparison of these numbers shows that the largest main effect is due to temperature cycling.

The concept of the main effect of temperature, for example, may be stated as follows:

- The main effect of temperature is given by the difference between the average severity for all test conditions at high temperatures and the average severity for all test conditions at low temperatures.

By means of the severity ratings shown in Table 7, this definition may be used to compute the main effect of temperature as follows:

- Average severity at high thermal stress:

$$(13 + 32 + 18 + 64 + 24 + 57 + 32 + 111)/8 = 44$$

- Average severity at low thermal stress:

$$(4 + 5 + 18 + 20 + 7 + 8 + 24 + 26)/8 = 14.$$

The difference between those two averages is seen to be 30 in agreement with the main effect of the thermal stress shown in the last column of Table 8. The main effects of SO_2, RH, and UV may be similarly verified.

TABLE 8. YATES COMPUTATION OF MAIN EFFECTS AND INTERACTIONS FOR COMPLETE FACTORIAL DESIGN INVOLVING 16 POSSIBLE ACCELERATED TESTS

Final Rating[a]	(b)	(c)	(d)	(e)	(f)	(g)	Main Effects and Interactions
4	9	47	174	463	16	29	Mean
5	38	127	289	183	8	23	SO_2
18	45	65	68	163	8	20	RH
20	82	224	115	75	8	9	RH x SO_2
13	15	3	66	239	8	30[h]	T
32	50	65	97	171	8	21	T x SO_2
18	81	3	28	35	8	4	T x RH
64	143	112	47	71	8	9	T x RH x SO_2
7	1	29	80	115	8	14	UV
8	2	37	159	47	8	6	UV x SO_2
24	19	35	62	31	8	4	UV X RH
26	46	62	109	19	8	2	UV x RH x SO_2
24	1	1	8	79	8	10	UV x T
57	2	27	27	47	8	6	UV x T x SO_2
32	33	1	26	19	8	2	UV x T x RH
111	79	46	45	19	8	2	UV x T x RH x SO_2

(a) Source: Table 7.
(b) First 8 entries in this column are obtained by adding successive pairs of entries in column (1); e.g., 4 + 5 = 9; 18 + 20 = 38, etc. Last 8 entries are obtained by taking successive differences in column (1); e.g., -4 + 5 = 1; -18 + 20 = 2; etc.
(c) The procedure used to obtain column (2) is repeated to obtain column (3); e.g., 9 + 38 = 47; 45 + 82 = 127; -9 + 38 = 29; -45 + 82 = 37, etc.
(d) Column (4) is obtained from column (3) using the same procedure.
(e) Column (5) is obtained from column (4) using the same procedure.
(f) Divisors for entries of column (5).
(g) Obtained by dividing entries of column (5) by entries in column (6). These magnitudes give quantitative estimates of the main effects and interactions listed in the last column. The main effects are given by single labels: SO_2, RH, T, and UV; interactions are given by combinations of labels; e.g., RH x SO_2 denotes the interaction between relative humidity and sulfur dioxide.
(h) This is the largest magnitude for a main effect and is used as the first splitting variable in the hierarchical tree representation (Figure 4).

The largest two-variable interaction is shown to be 21 in Table 8 and is associated with the interaction between temperature and sulfur dioxide, T x SO$_2$. This measures the extent to which the effect of SO$_2$ differs depending on whether the temperature stress of the test condition is at a high level or low level. The definition of the T x SO$_2$ interaction can be expressed as follows:

- The interaction effect between temperature and sulfur dioxide is equal to the difference between the (conditional) main effect of SO$_2$ at the high-temperature condition and the (conditional) main effect of SO$_2$ at the low-temperature condition.

By means of the severity ratings shown in Table 7, this definition may be used to compute the T x SO$_2$ interaction as follows:

- Main effect of SO$_2$ at high T:

$$(32 + 64 + 57 + 111)/8 - (13 + 18 + 24 + 32)/8 = 22$$

- Main effect of SO$_2$ at low T:

$$(5 + 20 + 8 + 26)/8 - (4 + 18 + 7 + 24)/8 = 1.$$

The difference between these two conditional main effects is seen to be equal to 21, in agreement with the interaction measure shown in Table 8 for T x SO$_2$. The measures for the remaining interactions shown in Table 8 can be similarly computed. However, for subsequent purposes, the primary result of interest is the fact that the largest main effect is associated with temperature.

Hierarchical Representation of
Main Effects and Interactions

Figure 4 is a hierarchical representation of the main effects and interactions. The horizontal scale at the top of the tree represents severity rating. The average severity rating for the 16 possible test conditions shown in Table 7 is found to be 29 and this is shown by the box labeled 29 at the top of the tree. The preceding calculation of the main effect of temperature shows that average severity at the high-temperature stress is 44 and the average severity at the low-temperature stress is 14. These averages are shown on the tree by the boxes labeled 44 and 14, respectively. In a similar manner the conditional main effect for UV, RH, and SO$_2$ are computed under the condition that the temperature stress is high. These conditional main effects are found to be 24, 25, and 44 for UV, RH, and SO$_2$, respectively. Because SO$_2$ yields the largest main effect at the high-temperature stress, it is shown in a tree as a splitting variable for the high-temperature stress. The tree shows that the average severity is 22 over all possible test conditions with temperature at high stress and sulfur dioxide at low stress; high-temperature stress and high-sulfur-dioxide stress yields an average severity of 66.

The construction of the hierarchical representation continues in this manner. At each stage of development of the tree, it is determined what test variable has the largest main effect, given the conditions associated with previous splits of the tree. As shown in Figure 4, the splitting process can continue until the terminal boxes represent the 16 possible test conditions shown in Table 7.

Severity Rating

```
0   10  20  30  40  50  60  70  80  90  100  110  120
```

Example: This terminal box represents an accelerated test involving high stress levels for temperature, T, sulfur dioxide, SO_2, and relative humidity, RH, and a low ultraviolet stress UV. The rating of 64 indicates the expected severity of this test combination on a scale from 0 to 120.

L = T = H

L = SO₂ = H

L = RH = H

L = UV = H

RH

UV

RH

UV   UV

SO₂   SO₂

- Numbers in boxes denote expected severity ratings
- Test combinations are represented by 16 terminal boxes
- Labels T, SO_2, RH, and UV denote stresses associated with temperature cycling, sulfur dioxide, relative humidity, and ultraviolet radiation
- Right and left branches correspond to high and low levels of stress, respectively
- Source of data: Table 7

FIGURE 4. HIERARCHICAL TREE REPRESENTATION OF EXPECTED SEVERITY RATINGS FOR 16 POSSIBLE ACCELERATED TEST CONDITIONS

Interpretation of the Hierarchical Tree

Several important observations can be made based on the graphical representation provided by the hierarchical tree:

- Those test combinations having high (low) average severities are displayed toward the right (left) side of the tree.

- The conditional main effect of a variable is proportional to the horizontal distance between the two groups formed by that variable at a split. For example, the large horizontal distance for the split on SO_2 at the high-temperature stress shows that the level of SO_2 has a dramatic effect on average severity, varying between 22 and 66.

- The sequence of variables down a branch of the tree gives an indication of the relative importance of the successive variables in determining average severity. For example, the most important variable is the temperature stress because the first split involves temperature. At the high-temperature condition, the next most important variable is seen to be SO_2; whereas, at the low-temperature stress the next most important variable is RH. The sequences of variables down each branch of the tree are quite different from branch to branch.

62

- Each box in the hierarchical tree represents a particular subset of the possible tests defined in Table 7. For example, the box labeled with an average severity of 66 represents the four tests below it in the tree. These tests are characterized by having high stress levels for T and SO_2, but having both high and low stress levels for RH and UV.

By examination of this graphical representation, it is possible to reassess the validity of the assigned severity ratings shown in Table 7, and make adjustments where necessary to achieve a tree that is believed to correctly reflect current engineering knowledge. Once an acceptable tree is obtained, it is possible to consider the question of which combination of test variables should be included in the design for an accelerated test.

Final Selection of Test Conditions

From a statistical perspective, it would be desirable to test all 16 possible test conditions shown in Table 7. Such a test would permit the unambiguous quantitative assessment of the main effects and interactions for all the test variables as they relate to any measured dependent degradation measure. The calculations would be analogous to those outlined above for the assigned severity ratings. From a practical perspective, however, it is frequently too costly to test all possible combinations of all levels of all relevant stress variables. Moreover, in practice, some test conditions obtained by taking combinations of test levels for stress variables yield infeasible tests associated with phase changes, etc. Consequently, it is desirable to reduce the number of stress combinations to be tested to a minimum. The following procedure is aimed at accomplishing this. In general, the price paid for this reduction is the inability to analyze the resulting data in the usual manner associated with a complete factorial design. Only partial information is obtained on main effects and interactions. In some instances, a fractional factorial design may be used, but again, some of the stress combinations required for statistical "balance" frequently appear to represent costly "niceties". From an engineering perspective, more informative tests that do not fit into a statistical design would be performed.

Under the assumption that all 16 possible stress combinations cannot be tested, it is argued below that nevertheless, the 16 tests should be performed conceptually as outlined above, and the hypothesized severities should be represented in the graphical display of the hierarchical tree. This tree can then be used as a basis for the selection of which stress combinations should be tested.

To show how this may be done, first consider the less severe stress combinations shown in Figure 4. At the low stress levels of temperature and relative humidity, the average severity is shown to be six. The imposition of the UV and SO_2 stress levels under these conditions are expected to yield changes in severity values ranging between four and eight. Such small changes in severity suggest that the *variations* in UV and SO_2 at these temperature and relative humidity conditions should not be included in the test program. Instead of varying the UV and SO_2 levels under these conditions, each may be set at a fixed level. In this way, the tree representation is "pruned" so that the left-most terminal box is 6, corresponding to the low stress levels for temperature and relative humidity. Similar pruning can be continued, with care being taken to retain those tests expected to yield essential engineering information.

Figure 5 shows a pruned tree in which only five test combinations remain. These combinations provide information for each stress variable under those conditions that are most severe.
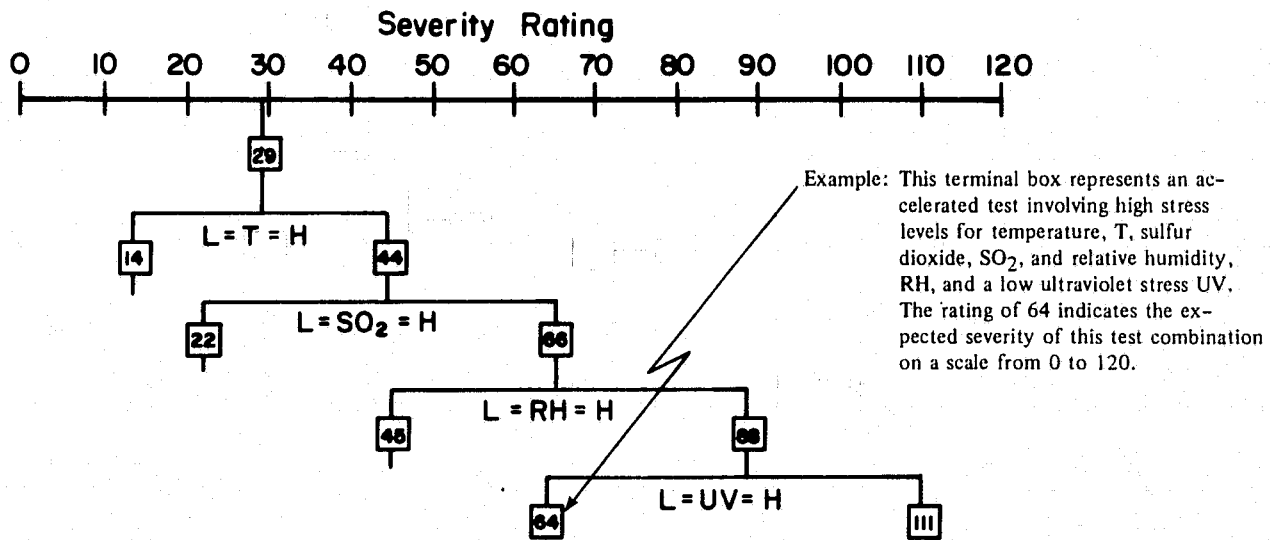
Severity Rating

Example: This terminal box represents an accelerated test involving high stress levels for temperature, T, sulfur dioxide, SO_2, and relative humidity, RH, and a low ultraviolet stress UV. The rating of 64 indicates the expected severity of this test combination on a scale from 0 to 120.

- Numbers in boxes denote expected severity ratings
- Test combinations are represented by 16 terminal boxes
- Labels T, SO_2, RH, and UV denote stresses associated with temperature cycling, sulfur dioxide, relative humidity, and ultraviolet radiation
- Right and left branches correspond to high and low levels of stress, respectively
- Source of data: Table 7

FIGURE 5. HIERARCHICAL TREE REPRESENTATION OF FINAL SELECTED TEST CONDITIONS

The five test conditions represented by the pruned tree are:

Test Condition	(T,	SO_2,	RH,	UV)	Degradation Measure
1	(-1	z	z	z)	y_1
2	(1	-1	z	z)	y_2
3	(1	1	-1	z)	y_3
4	(1	1	1	-1)	y_4
5	(1	1	1	1)	y_5

where the low and high stress levels are denoted by -1 and 1, respectively, and the z's may be chosen to be either -1 or 1. As shown later, different choices will permit these engineering tests to be embedded into different experimental designs having improved statistical properties. The corresponding dependent measures of degradation at each test condition are denoted by $y_1, \ldots, y_5$. These measures could be associated with delamination rates over time, or delamination amounts after a given test exposure, for example.

Algebraic Representation of the Hierarchical Tree

The five tests selected from the 16 possible tests are now regarded as the "points" to which a polynomial response surface is to be fitted. Each point represents a test condition. A generalized

64

Lagrangian polynomial is fitted to these points. For this example, the Lagrangian polynomial is linear for each of the four stresses, and may be written as follows:

$$y(x_T, x_{SO_2}, x_{RH}, x_{UV}) = (1/2)(1 - x_T)y_1$$

$$+ (1/4)(1 + x_T)(1 - x_{SO_2})y_2$$

$$+ (1/8)(1 + x_T)(1 + x_{SO_2})(1 - x_{RH})y_3$$

$$+ (1/16)(1 + x_T)(1 + x_{SO_2})(1 + x_{RH})(1 - x_{UV})y_4$$

$$+ (1/16)(1 + x_T)(1 + x_{SO_2})(1 + x_{RH})(1 + x_{UV})y_5, \quad (27)$$

where x_T, x_{SO_2}, x_{RH}, and x_{UV} denote the coded stress levels for temperature, sulfur dioxide, relative humidity, and ultraviolet radiation, respectively. The y-values in this equation are obtained from the experimental results. The test point $(1, 1, -1, z)$ associated with test condition 3, for example, corresponds to $x_T = 1$, $x_{SO_2} = 1$, $x_{RH} = -1$, and $x_{UV} = z$. Substitution of these values into the above equation shows that $y(1, 1, -1, z) = y_3$. Similar substitutions of the coordinates for the other test conditions show that the above polynomial yields an exact fit to each response measure $y_1, \ldots, y_5$ at each test point. This response surface may then be used to interpolate or extrapolate to obtain estimated response values at points different from those used in the experimental testing. To do this, it is only necessary to represent the point at which the response value is desired by the appropriate coordinates $(x_T, x_{SO_2}, x_{RH}, x_{UV})$, and to substitute these coordinates into Equation (27). In the accelerated test setting, the response surface is to be extrapolated to the point that corresponds to the normal stress condition at the geographic location of interest.

Representation of Stress Levels by a Numerical Scale

In order to extrapolate Equation (27), the coordinates of the point representing the normal stress conditions must be numerically defined. To be consistent with the preceding methodology, it is necessary that for each kind of stress (temperature, relative humidity, etc.) a numerical scale must be constructed so that +1 and -1 represent the highest and lowest stress levels to be used in the accelerated tests. On such a numerical scale, additional accelerated stress levels for each stress must be between -1 and +1, with the normal stress level assuming a scale value, x, that is less than -1.

The construction of a suitable numerical scale for each kind of stress is difficult. The primary difficulty stems from the fact that many experimental variables are multidimensional so that a suitable one-dimensional scale may not exist. In the case of a temperature cycle, for example, the amplitude, frequency, and mean temperature level may best be represented as three components of a vector. If these three components cannot be suitably represented by a single numerical measure, then it may be necessary to test each component as a separate experimental variable with its own numerical scale. This tends to increase the size and complexity of the accelerated test, and would be expected to improve the validity of the necessary extrapolations to the normal stress levels. Stated another way, accelerated testing requires quantitative extrapolations from test results obtained at high stress levels to stress levels corresponding to normal-stress conditions. Of necessity, numerical scales are required as a basis for such quantitative extrapolations. It follows that the validity of the extrapolations depends on the validity of the numerical scales so that careful construction of these scales is required.

It should also be noted that the proposed methodology applies equally well to the several components of a multidimensional stress. Each component is simply tested as a separate kind of stress.

It is advantageous to give explicit consideration to the problem of defining an appropriate numerical scale for each stress or stress component in the *design* of an accelerated test. It is all too easy to first run several over–stress tests and then find that there is no quantitative scale that will permit defensible extrapolations to normal conditions. The numerical scales used below for the example application are oversimplified, and are presented primarily to outline the methodology.

Scale for Temperature Cycling. For illustrative purposes, the range of a temperature cycle is taken to be the component of the thermal stress that is numerically scaled. The low–thermal–stress condition is given by cycling the temperature between –40 and 100 C to give a range of 140 Celsius degrees. The high–thermal–stress condition involves temperature cycling between –40 and 150 C to give a range of 190 Celsius degrees. A linear scale based on these values is given by $x_T = (T-170)/20$. When T takes the values of 150 and 190 C, it is seen that $x_T = -1$ and 1, and corresponds to the low and high stress condition, respectively. When T is set equal to the range of temperature that is expected under normal conditions, then x_T takes the corresponding scale value that is needed for extrapolation. Several possibilities are available for interpretation of the "range of temperature that is expected under normal conditions". As examples, such a range could be based on any of the following:

- Average daily temperature range over a specified number of years

- Average daily temperature range during the month of July over a specified number of years

- Average daily temperature range during the month of December over a specified number of years

- Average annual maximum temperature minus average annual minimum temperature over a specified number of years.

In general, such information can be obtained from published data and will vary according to geographic location. For illustrative purposes, the data in Table 9 were obtained from "Terrestrial Service Environments for Selected Geographic Locations"[3]:

The last column of the table shows the computed scale value for each geographic location. These values are seen to vary between –6.9 for Miami, Florida, and –4.3 for Fairbanks, Alaska. The average value of x_T across these 9 locations is found to be –5.7.

Scale for Relative Humidity. A scale for relative humidity is obtained as follows. The low and high stress levels are taken to be 50 and 95 percent at 100 C, and are assigned scale values of –1, and 1, respectively. The most frequently occurring combination (mode) of relative humidity and temperature is obtained from historical data[3] for the nine geographic locations considered above, as shown in Table 10.

TABLE 9. SELECTED TEMPERATURE DATA FOR SPECIFIC GEOGRAPHIC AREAS FROM REFERENCE (3)

| Geographic Location | Air Temperature, C | | Range, T | Scale |
	Maximum[a]	Minimum[a]		Value[b], x_T
Albuquerque, New Mexico	40.6	-26.1	66.7	-5.2
Bismarck, North Dakota	42.2	-40.6	82.8	-4.4
Boston, Massachusetts	36.1	-20.0	56.1	-5.7
Brownsville, Texas	37.2	- 2.2	39.4	-6.5
Cleveland, Ohio	34.4	-25.0	59.4	-5.5
Fairbanks, Alaska	34.4	-50.0	84.4	-4.3
Los Angeles, California	37.8	1.1	36.7	-6.7
Miami, Florida	34.4	1.7	32.7	-6.9
Phoenix, Arizona	46.7	- 5.0	51.7	-5.9

(a) Based on years 1965-1974.
(b) $x_T = (T-150)/20$.

TABLE 10. RELATIVE HUMIDITY AND TEMPERATURE RANGES FOR SELECTED GEOGRAPHIC AREAS

Geographic Location	Most Frequent Ranges for Relative Humidity and Air Temperature, (RH,T)	Midpoints (RH,T)
Albuquerque	(0,29), (4,20)	(15,12)
Bismarck	(50,69), (4,20)	(60,12)
Boston	(50,69), (4,20)	(60,12)
Brownsville	(50,69), (20,30)	(60,26)
Cleveland	(50,69), (4,20)	(60,12)
Fairbanks	(50,69), (4,20)	(60,12)
Los Angeles	(80,89), (4,20)	(85,12)
Miami	(50,69), (20,32)	(60,26)
Phoenix	(0,29), (20,32)	(15,26)

The scale for relative humidity should account for the general inverse relation between relative humidity and air temperature. For purposes of this example, this relation was not explicitly determined. Instead, the code value for relative humidity was arbitrarily set equal to a value of -3. This assignment reflects the fact that the field stress associated with relative humidity is expected to be appreciably smaller than the test stress that is associated with a scale value of -1. A more realistic application of the methodology would combine the relation between relative humidity and temperature with geographic data of the kind shown in Table 10 to obtain a proper scaled value for relative humidity.

Scale for Sulfur Dioxide. A scale for sulfur dioxide is obtained by taking the low and high stress levels to be 0.1 and 0.3 parts per million, respectively. Using a linear scale, it follows that $x_{SO_2} = (SO_2 - 0.2)/0.1$. Average values of SO_2 for Cleveland, Boston, and Phoenix are found to be 0.03, 0.02, and 0.01 ppm with resulting computed scale-values of -1.7, -1.8, -1.9, respectively.[95] Alternative scale-values could, of course, be based on maximum annual historical levels, maximum levels in the month of July, etc., provided such data can be obtained for the geographic locations of interest.

Scale for Ultraviolet Radiation. A scale for ultraviolet radiation is obtained as follows. It is assumed that 5 percent of the solar radiation is in the ultraviolet range. For a maximum solar flux of 100 ly/hr this yields a maximum UV flux of 5 ly/hr. A continuous exposure at this flux level over a 20-year period would yield a total cumulative UV exposure approximately equal to (5 ly/hr) (24 hr/day) 365 days/yr) = 876,000 ly. To achieve this 20-year cumulative total in an accelerated test conducted over a 2-year period, a UV flux of 50 ly/hr is required. This flux is taken to be the low-stress condition and is assigned a scale value of -1. A UV flux equal to 100 ly/hr for a period of 2 years is taken to be the high-stress level with a scale value of 1.0 assigned to the cumulative total of 1,752,000 ly.

The scale value for a particular geographic location is then obtained by estimating the cumulative total UV radiation expected in 20 years. The associated scale value is then given by $x_{UV} = (UV - 1.314 \times 10^6)/0.548 \times 10^6$. If the actual cumulative total is estimated to be 25 percent of the maximum, it would follow that $UV = (0.25) (876,000) = 219,000$ ly, and x_{UV} would be equal to -2.50. Because UV data are generally lacking, it may be necessary to assume a particular location as a reference value and then make corrections to adjust the data for latitude and altitude to obtain estimates of UV radiation for other geographic locations.

Coordinates of the Point Representing Normal-Stress Conditions

The preceding discussion of the scale construction for each stress indicates that the point representing the normal-stress level for each geographic location is given by the scale values: $(x_T, x_{RH}, x_{SO_2}, x_{UV})$. These coordinates vary from one geographic location to another. In a detailed study, these coordinates should be computed and examined to determine whether certain geographic locations may be grouped so that one accelerated test design may be applied to several locations. Such studies have not been carried out in this effort.

To complete the calculations for the illustrative example, the following coordinates are taken as the point which represents normal-stress conditions:

$$(x_T, x_{SO_2}, x_{RH}, x_{UV}) = (-6, -2, -3, -2.5).$$

Optimum Allocation of Modules to Test Conditions

The preceding results are now used to compute the optimum allocation of modules to the five accelerated test conditions. The computations are based on relations analogous to those of Hoel and Levine[87]. The relations are obtained under the assumption of equal variances for

each test condition. Alternative assumptions and possible generalizations are outlined in a discussion of the results.

The Lagrange polynomial for the five test conditions may be written as follows:

$$y\,(x_T, x_{SO_2}, x_{RH}, x_{UV}) = L_1 y_1 + L_2 y_2 + L_3 y_3 + L_4 y_4 + L_5 y_5,$$

where

$$L_1 = (1/2)\,(1 - x_T),$$
$$L_2 = (1/4)\,(1 + x_T)\,(1 - x_{SO_2}),$$
$$L_3 = (1/8)\,(1 + x_T)\,(1 + x_{SO_2})\,(1 - x_{RH}),$$
$$L_4 = (1/16)\,(1 + x_T)\,(1 + x_{SO_2})\,(1 + x_{RH})\,(1 - x_{UV}),$$

and

$$L_5 = (1/16)\,(1 + x_T)\,(1 + x_{SO_2})\,(1 + x_{RH})\,(1 + x_{UV}).$$

Evaluation of these L's at the extrapolation point given by $(x_T, x_{SO_2}, x_{RH}, x_{UV}) = (-6, -2, -3, -2.5)$ yields the following results:

$$L_1 = (1/2)\,(7) = 7/2,$$
$$L_2 = (1/4)\,(-5)\,(3) = -15/4,$$
$$L_3 = (1/8)\,(-5)\,(-1)\,(4) = 20/8,$$
$$L_4 = (1/16)\,(-5)\,(-1)\,(-2)\,(3.5) = -35/16,$$

and

$$L_5 = (1/16)\,(-5)\,(-1)\,(-2)\,(-1.5) = 15/16.$$

It follows that

$$S = \sum_{i=1}^{5} |L_i| = 206/16$$

and the optimal proportion of the modules to be tested at each test condition is given by

$$p_1 = |L_1|\,/S = 56/206,$$
$$p_2 = |L_2|\,/S = 60/206,$$
$$p_3 = |L_3|\,/S = 40/206,$$
$$p_4 = |L_4|\,/S = 35/206,$$

and

$$p_5 = |L_5|\,/S = 15/206.$$

Thus, if 206 modules are to be tested, then (56, 60, 40, 35, 15) modules should be allocated to tests 1 through 5, respectively. If 68 modules are to be tested, then the approximate optimum

69

allocation would be given by (18, 20, 13, 12, 5). If n denotes the number of modules to be tested, then the approximate optimum allocation is given by $(np_1, \ldots, np_5)$. In general, it is recommended that five be the minimum number of modules tested at a given condition.

Improving the Statistical Properties
of the Engineering Design

The engineering input is not likely to yield an experimental design with acceptable statistical properties. For this reason, it will generally be desirable to supplement the engineering design with test combinations selected to improve its statistical properties. Because, each additional test will increase the cost of the accelerated test, the improvement in design must be balanced against the increased cost.

In general, a supplemented engineering design will be improved provided the additional tests permit a more definitive analysis of the data. In statistical terms, improved analysis means that better estimates can be made of the main effects and lower-order interactions among the test variables. Exactly which main effects and interactions are of most importance is partly indicated by the engineering design itself. However, as more tests are added, further engineering input must be used to guide the process so that both the engineering properties and the statistical properties of the resulting design will be improved by the supplementary tests.

Need for Statistical Input

Thousands of experimental designs have appeared in the statistical literature. A statistician familiar with this literature can be expected to give good advice concerning exactly what tests would appropriately supplement an engineering design. Even moderate familiarity with the statistical design of experiments may be sufficient to yield marked improvements in the engineering design with the addition of a few tests.

Because the proposed methodology begins with a complete factorial design, the literature associated with fractional factorial designs is particularly relevant to determining what supplementary tests should be added to an engineering design. A useful review of fractions of asymmetrical factorial arrangements, irregular fraction plans, and sequences of fractional factorial plans is given by Addelman[96].

Estimation of Parameters

Explicit formulas for the model parameter can be obtained by simple rearrangement of the Lagrange polynomial representation of the hierarchical tree. For example, Equation (27) is expressed as a weighted combination of the dependent variables, $y_1, \ldots, y_5$. The weights involve algebraic functions of the independent variables x_T, x_{SO_2}, x_{RH}, and x_{UV}. The equation may be rearranged to find the coefficients of each independent variable and their products. The coefficient of x_{SO_2}, for example, is given as:

$$(1/16)\,(-4y_2 + 2y_3 + y_4 + y_5)\,x_{SO_2}.$$

This same coefficient is also found for the product $x_T x_{SO_2}$. Thus, in the rearranged equation the following term is obtained:

$$(1/16) \, (-4y_2 + 2y_3 + y_4 + y_5) x_{SO_2} \, (1 + x_T).$$

Comparison with the statistical model then shows that the parameter C is estimated by the expression:

$$\hat{C} = (1/16) \, (-4y_2 + 2y_3 + y_4 + y_5).$$

In a similar manner, the estimating equations for each parameter in the model can be obtained. For the first iteration design, for example, the remaining parameters are given as follows:

$$\hat{A} = (1/16) \, (8y_1 + 4y_2 + 2y_3 + y_4 + y_5),$$

$$\hat{B} = (1/16) \, (-8y_1 + 4y_2 + 2y_3 + y_4 + y_5),$$

$$\hat{D} = (1/16) \, (-2y_3 + y_4 + y_5),$$

and

$$\hat{E} = (1/16) \, (-y_4 + y_5).$$

In general, each parameter is estimated by a linear combination of the y values. In contrast to preferred statistical designs, these linear combinations may not be mutually orthogonal.

Confounding of Effects

As noted previously for the first iteration design, identical coefficients are obtained for x_{SO_2} and $x_T x_{SO_2}$. In general, this means that the design is not capable of obtaining separate estimates for the main effect of SO_2 and the interaction between T and SO_2. For this reason, these effects are said to be confounded. The design is only able to yield an estimate of the main effect of SO_2 at the high level of T. More tests would be required to obtain separate estimates of these two effects. Based on the engineering input, separate estimates are not desired. However, it is prudent to examine all such confounded terms to assure that their exclusion from the test program is justified.

The following list shows the confounded effects for the first iteration design:

SO_2, T x SO_2

RH, T x RH, SO_2 x RH, SO_2 x RH x T

UV, T x UV, SO_2 x UV, T x SO_2 x RH, RH x UV, T x RH x UV,

SO_2 x RH x UV, T x SO_2 x RH x UV.

Because a complete factorial design would require 16 tests and provide 16 parameter estimates, it is clear that a subset of five of these tests can provide only five parameter estimates with the remaining 11 parameters associated with confounded terms. The procedure described above permits an explicit identification of what confounding will occur.

71

Some Specific Statistical Designs

The following paragraphs outline some specific statistical designs that can be obtained by supplementing the engineering design shown in Figure 5. The five tests associated with the engineering design were obtained by selecting a subset of a 2^4 complete factorial involving 16 tests. Thus, 11 tests were eliminated because of engineering considerations. It is clear that if these 11 tests were re-included in the design, then a standard statistical design, a 2^4 complete factorial, would be obtained. It is not so clear whether any standard statistical designs exist that can be obtained by adding fewer than 11 supplementary tests to the five engineering tests. Thus, the statistical design problem consists of identifying those standard statistical designs that require a minimum number of tests and include the engineering tests as a subset.

Table 11 summarizes five specific designs that require fewer than 16 tests. Each design includes the five tests associated with the engineering design.

Design A consists of nine tests. The first eight tests constitute a 1/2 replication of a 2^4 complete factorial. Such a design is usually denoted as a 2^{4-1} fractional factorial. The five engineering tests are indicated by asterisks. With the exception of Test 9, the listing shows that four of the engineering tests are included in the 2^{4-1} fractional factorial.

The first eight tests of Design B consist of the 1/2 replicate complementary to Design A, plus one additional test (Test 9). A reduction to eight tests is shown for Design C, which consists of two quarter replicates. The first four tests constitute a quarter replicate, or 2^{4-2} fractional factorial, from Design B; the second four tests constitute a quarter replicate from Design A.

Design D consists of only six tests. According to Addelman[98], the first five of these tests represent the "most efficient" nonorthogonal five tests obtainable as a resolution III subset of a 2^4 factorial design. A resolution III design permits the estimation of all main effects when all two-factor and higher order interactions are absent. The original "resolution-R plan" was introduced by Box and Hunter[97] and was subsequently generalized by Webb[98]. In general, a higher resolution permits improved data analysis. For example, a resolution IV design is one in which no main effect is confounded with any other main effect or two-factor interaction, but two-factor interactions are confounded with each other; a resolution V design is one in which no main effect or two-factor interaction is confounded with any other main effect or two-factor interaction, but two-factor interactions are confounded with three-factor interactions. In general, an R-resolution design is one for which no K-factor effect is confounded with any other effect containing less than (R-K) factors.

A variety of useful results have been obtained for 2^n factorial designs of resolution III and V. For example, the minimum number of tests to yield resolution III and V are given by (n + 1) and $(n^2 + n + 2)/2$, respectively. The minimum number of runs for resolution IV designs has been examined by Margolin[99].

Design E in Table 11 consists of 11 tests. This design is an example of a one-factor-at-a-time design, and is ideally suited for sequential testing over time in which the tests are to be run in order from Test 1 through Test 11. This type of design was examined by Webb[98] as a "contractible permutation-invariant design, and by Rechtschaffner[100] as a "saturated" fractional factorial. The design is contractible in that the sequence of tests can be stopped after specified blocks of tests have been run; the design is saturated in that no degrees of freedom are available for estimation of error.

TABLE 11. STATISTICAL DESIGNS THAT INCLUDE THE FIVE
TESTS ASSOCIATED WITH THE ENGINEERING
DESIGN

Design	Test Number[a]	Design Matrix[b] (T, SO_2, RH, UV)	Comments
A	1	-1 -1 -1 -1	The first eight tests constitute a 2^{4-1}
	2*	1 1 -1 -1	factorial design.
	3	1 -1 1 -1	
	4	-1 1 1 -1	
	5*	1 -1 -1 1	
	6	-1 1 -1 1	
	7*	-1 -1 1 1	
	8*	1 1 1 1	
	9*	1 1 1 -1	
B	1	1 -1 -1 -1	The first eight tests constitute a 2^{4-1}
	2	-1 1 -1 -1	factorial design.
	3	-1 -1 1 -1	
	4	-1 -1 -1 1	
	5*	1 1 1 -1	
	6*	1 1 -1 1	
	7*	1 -1 1 1	
	8*	-1 1 1 1	
	9*	1 1 1 1	
C	1*	1 -1 -1 -1	Tests 1 through 4 constitute a 2^{4-2}
	2	-1 -1 -1 1	factorial design from design B above;
	3*	1 1 1 -1	tests 5 through 8 constitute a 2^{4-2}
	4	-1 1 1 1	factorial design from design A above.
	5*	-1 -1 -1 -1	
	6*	1 1 -1 -1	
	7	-1 -1 1 1	
	8*	1 1 1 1	
D	1	-1 -1 -1 -1	Tests 1 through 5 constitute the most
	2*	-1 1 1 1	efficient nonorthogonal five tests in a
	3*	1 -1 1 1	2^4 factorial having resolution III.
	4*	1 1 -1 1	
	5*	1 1 1 -1	
	6*	1 1 1 1	
E	1*	1 1 1 1	Tests 1 through 11 constitute a one-
	2*	-1 1 1 1	factor-at-a-time design that may be
	3*	1 -1 1 1	run sequentially.
	4*	1 1 -1 1	
	5*	1 1 1 -1	
	6	-1 -1 1 1	
	7	-1 1 -1 1	
	8	-1 1 1 -1	
	9	1 -1 -1 1	
	10	1 -1 1 -1	
	11	1 1 -1 -1	

(a) The five tests based on engineering inputs are labeled with an asterisk. These five tests are represented by the ordered symbols: (-1, Z, Z, Z), (1, -1, Z, Z), (1, 1, -1, Z), (1, 1, 1, -1), and (1, 1, 1, 1), where the low and high stress levels for T, SO_2, RH, and UV are denoted by -1 and 1, respectively. The values for Z may be chosen to be either -1 or 1. Different choices for the values of Z yield different representations for the five engineering tests, and thereby permit the five tests to be included as subsets of different standard experimental designs.

(b) Each row of the design matrix represents a single test. Row 2 of Design A represents a test in which temperature (T) and sulfur dioxide (SO_2) are set at high levels (+1), and relative humidity (RH) and ultraviolet exposure (UV) are set at low levels (-1).

73

It is noteworthy that the five engineering tests comprise the first five tests of Design E. The sequential ordering of the tests for this design indicates that after the first two tests are run, an estimate of the main effect of temperature can be obtained by comparing the results obtained from Test 1 and Test 3. After Test 3 is run the main effect of SO_2 can be obtained by comparing the results obtained from Test 1 and Test 3. After the five engineering tests have been run, the four main effects of T, SO_2, RH, and UV are obtained by comparing results with those of Test 1. After Test 6 is run, the two-factor interaction between T and SO_2 can be obtained using the results of Tests 1, 2, 3, and 6. In combination with Tests 1, 2, and 4, Test 7 provides an estimate of the interaction between temperature and relative humidity, T x RH. In a similar manner, the last four tests provide estimates of all remaining two-factor interactions: T x UV, SO_2 x RH, SO_2 x UV, and RH x UV, respectively.

The designs shown in Table 11 indicate that, for the example under consideration, a variety of standard statistical designs exist which require between 6 and 11 tests. These designs all include the five engineering tests discussed for this example. In other instances, however, the engineering tests may not be easily identified as a subset of some standard statistical design. In such a case, a non-standard design may be identified using a procedure of Kennard and Stone.[101] In their procedure, given that certain tests must be included in the design, the next test to be added is that test which is located at the greatest "distance" from the existing tests. To implement this procedure, it is necessary to define a "distance" metric between test conditions. For a large number of test variables, Kennard and Stone have developed a computer code to identify the supplementary tests.

Many other concepts have also been developed under the statistical design of experiments. The concept of "bias" is important, particularly in connection with minimization of variance. Whether it is better to minimize bias, B, variance, V, or the sum (V + B) has been studied in the literature. The difficulties of designs that focus on the minimization of bias are discussed by Stigler[102], and a restricted version of D-optimality is proposed to overcome these difficulties. The concept of D-optimality is associated with minimization of the generalized variance, and identifies those experimental designs that have a maximum determinant for their associated information matrix. The D-optimal designs were introduced by Kiefer[103]; a review of the practical use of such designs is given by Nalimov, Golikova, and Mikeshina[104].

The Use of Standard Experimental Designs for Accelerated Testing

The preceding discussion indicates the general desirability of augmenting the engineering tests to improve the statistical properties of the experimental design. Some general procedures and some specific examples are offered to identify desirable supplementary tests. It might be concluded from these examples that, once a suitable statistical design has been identified, a routine implementation of the design will yield useable results. Such a conclusion is likely to be grossly in error. The purpose of the following commentary is to better identify some strengths and weaknesses of using standard statistical designs as a basis for accelerated testing.

Standard experimental designs are most effectively implemented when the following conditions hold:

(1) A single "yield" is obtained at the end of a test for each experimental condition.

(2) The test conditions can be chosen to "span the space" so that interpolation among test results can be accomplished.

(3) All combinations of the test variables can be tested.

(4) The range of each variable is restricted to a relatively small region around the mean, so that a low-degree Taylor series expansion at the mean gives a good fit to the true response surface.

Each of these conditions is severely violated in accelerated testing. First, a single dependent variable (yield) usually does not suffice for an accelerated test. In general, the dependent variables constitute a wide variety of measurable properties, which are mechanical, chemical, electrical, optical, etc. Any property, whose magnitude may change over time and thereby degrades performance below acceptable levels, is a dependent variable that must be studied in accelerated testing. Some of these dependent variables may not be known prior to testing. Moreover, because accelerated testing focuses on time rates of degradation, the data must be taken throughout the test period. This means that each dependent variable is associated with a series of measurements taken over time. The internal structure of these time series forms the primary basis for the data analysis. Is the degradation rate for a particular property constant over time, or does it vary with time? Can the early portions of a time series be used to predict the later portion of the same series? Can early portions of the time series associated with the higher stresses for some dependent variables be used to predict the later portions of the time series associated with the lower stresses for the same, or different, dependent variables? Can the structure of each time series for each dependent variable be related to the environmental and/or operating conditions? How should the environmental and operating conditions be represented numerically (scaled) in order to better identify and quantify these relationships? These kinds of considerations show the oversimplification involved in condition 1 given above.

Because accelerated testing is concerned with extrapolation from higher-than-normal stress conditions to normal stress conditions, it is clear the condition 2 is violated. A 2-year accelerated test that is intended to yield valid extrapolations for a 20-year period is seen to require a tenfold extrapolation. Even the most optimistic practitioners are likely to express reservations concerning the validity of extrapolations of such a magnitude.

Considering condition 3, the most obvious restriction on testing all combinations of levels of the test variables is excessive cost. However, other restrictions are also commonplace. Certain combinations of levels of the test variables may be prevented by the physics of the device; other combinations may be deemed unnecessary because of prior engineering information. The exclusion of particular test combinations frequently tends to degrade the statistical design. Such exclusions may occur in accelerated testing even after the experiment is under way because some test conditions may ultimately be found to change the dominant failure mode from that associated with normal operating conditions.

To achieve maximum acceleration in an accelerated test it is desirable to operate the devices under the highest possible stress that does not change the dominant mode of failure associated with normal operating conditions. If these high stresses are not near normal stress levels, then a low-degree expansion of Taylor's series may not yield a good fit to the response surface and condition 4 is violated.

The preceding commentary is intended to indicate that considerations related to standard experimental designs in accelerated testing are important for initial structuring of an acceptable

design. The general concepts associated with the design of experiments are useful to help organize and develop an initial plan for experimentation. However, the design itself is not likely to provide much guidance for the required detailed analyses of the resulting time series, the scaling of environmental and operating conditions, and the predictions "within" and "between" the time series obtained at different stress levels. Such analyses are discussed in more detail in later sections of this report.

A Consistency Check on Engineering
Input Information

Because the engineering input is obtained from a complete factorial, a simple check can be used to test the consistency of this information. The check is accomplished by making a separate hierarchical tree for each half replicate of the complete factorial. The resulting trees are then compared with each other and with the tree representing the original complete factorial.

Figures 6 and 7 show the trees that result from the two half replicates, labeled A and B. Half-replicate A consists of the eight test combinations labeled 1, 4, 6, 7, 10, 11, 13, and 16 in Table 7; half-replicate B consists of the remaining eight test combinations. The final ratings shown in the last column of Table 7 were separately analyzed for each of these half replicates in order to obtain the trees shown in the figures. The trees show excellent consistency. The variable associated with the first split is that variable having the largest main effect. In both half-replicate trees, this variable is temperature. The average severity ratings at low and high temperatures are 14 and 46 for half-replicate A, and 14 and 42 for half-replicate B. These averages may be compared with 14 and 44 shown in Figure 4. In agreement with the complete factorial tree of Figure 4, the half-replicate trees show that the largest conditional main effect at high temperature is associated with sulfur dioxide, whereas at low temperature the largest main effect is associated with relative humidity. Because the half-replicate trees are based on only eight test conditions, their remaining portions cannot separately identify average severities for the remaining variables. However, comparisons with the results shown in Figure 4 show good agreement for the general range of its variables. The largest discrepancy is associated with the most severe condition shown to be 111 for half-replicate A and 64 for half-replicate B. This difference simply results from the fact that 111 and 64 are the largest severity ratings contained in their respective half replicates.

These results show that the engineering input information is largely self-consistent. That is, essentially the same information would have been obtained from a subjective evaluation of the severity of only 8 of the 16 possible test conditions. The eight test conditions for either half replicate would have identified essentially the same tree structure. If major differences are found between the half-replicate trees, the engineering input should be reexamined in an effort to improve the consistency. It should be noted, however, that just because the engineering input is consistent, does not mean that it is correct. No statistical procedures can determine the validity of the engineering input; only self-consistency can be assessed.

Finally, when a large number of variables must be considered in the experimental design, it may be desirable to restrict attention to a half replicate, or even a quarter replicate, in order to keep the number of test conditions that must be subjectively evaluated reasonably small.

Number of Stress Levels

The calculations made previously illustrating the basic elements of the design are based on the assumption that each stress is represented by a low level and a high level. This simplifying
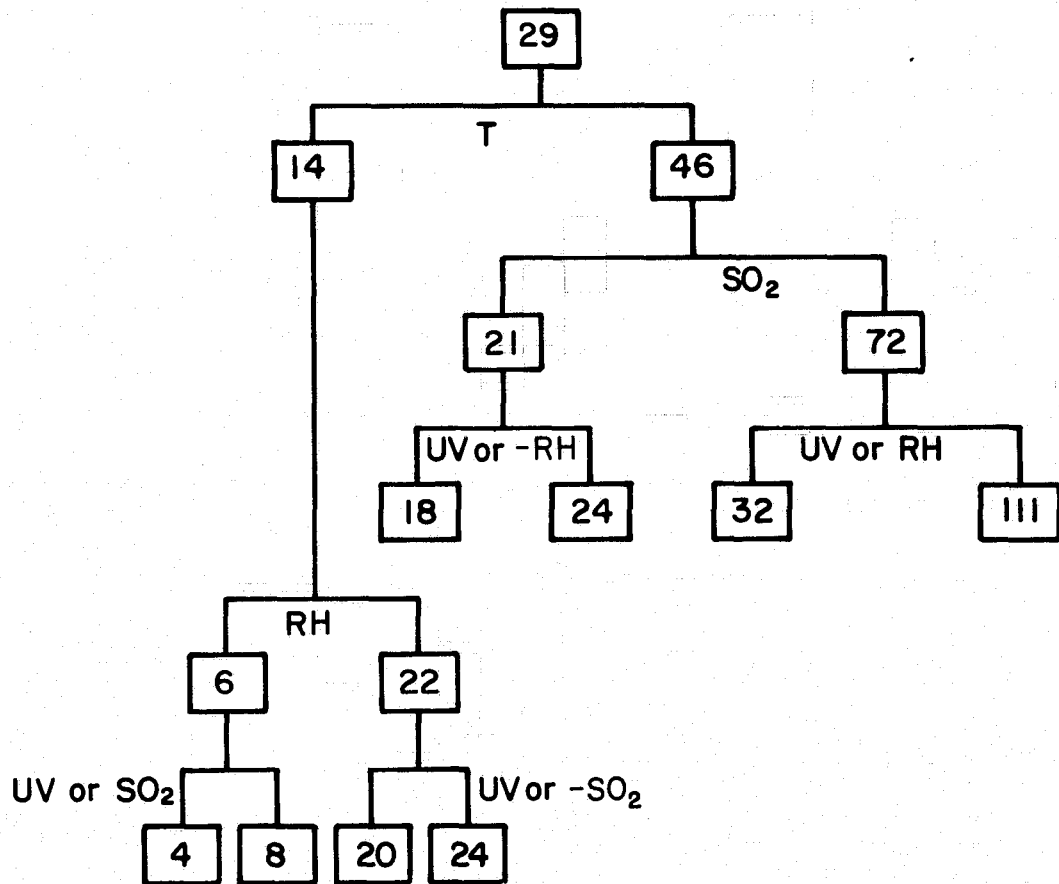
FIGURE 6. HIERARCHICAL TREE REPRESENTATION BASED ON HALF–REPLICATE A
OF ENGINEERING INPUT INFORMATION
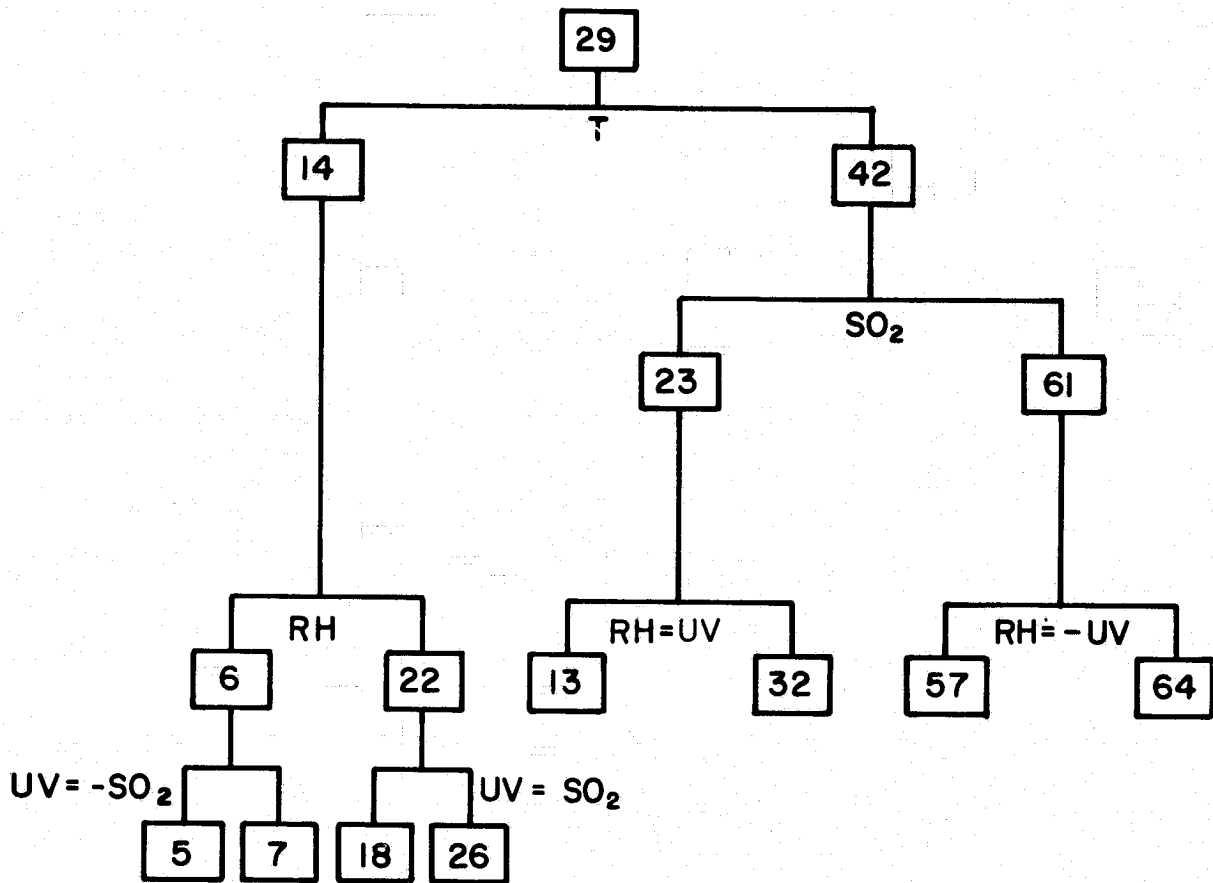
FIGURE 7. HIERARCHICAL TREE REPRESENTATION BASED ON HALF–REPLICATE B
OF ENGINEERING INPUT INFORMATION

assumption is made primarily to minimize the number of combinations of stress levels that must be considered in the first iteration of the methodology. Thus, the 16 stress combinations listed in Table 7 represent the minimum number of combinations that can be obtained from the low (L) and high (H) stress levels for the four stresses UV, T, RH, and SO_2.

For many stresses, the use of only two levels would not be recommended. For example, the generalized stress associated with temperature cycling should probably be represented by more than two levels because degradation rates are not expected to be linearly dependent on temperature. If three levels (low, intermediate, high) are used for temperature, then quadratic effects could also be identified; four levels could identify cubic effects, etc. Clearly, the use of many levels for a given stress would permit the identification of high-order (polynomial) nonlinearities. However, practical considerations typically impose severe limitations on the number of stress levels, and result in the question — What is the minimum number of stress levels that should be used?

As discussed previously, statistical considerations have sometimes suggested the use of a minimum of five stress levels.[81,92] In some examples, this number yields approximately symmetric forms for the maximum relative likelihood functions. However, it must also be noted that it is assumed that the dominant failure mode does not change over these five stress levels. If, in fact, the dominant failure mechanism changes, for example, at the highest stress condition, then the data for the high stress condition cannot be used in the extrapolation process to estimate degradation rates that would occur at normal stress levels. That is, for purposes of data analysis, the high temperature stress level is "lost", with the result that only four valid stress levels remain. This suggests the desirability of including even more than five stress levels in the experimental design in order to better assure that five valid stress levels will be available for the final analyses.

In summary, a minimum of five stress levels is recommended for each stress that is expected to exhibit nonlinear effects or is expected to interact strongly with other kinds of stresses. If five stress levels are used, special care should be taken to assure that the dominant failure mode for the module is the same at all five stress levels.

Number of Test Modules

The preceding calculations determine what proportion of the total number of test modules should be allocated to each test condition in order to maximize precision at an extrapolated point. The procedure determines the proportions, not the total number of modules. To achieve some statistical validity, the recommended total number of modules can be obtained by computing: $N = 5/p_{min}$, where p_{min} denotes the smallest computed proportion obtained for any test condition in the experimental design. This procedure results in testing a minimum of five modules at some test conditions, with proportionally larger numbers of modules tested at the remaining conditions. Thus, any averages that are computed in the course of the data analysis will be based on sample sizes of at least five. Under rather general conditions, such averages may be expected to be approximately normally distributed.[105] This, in turn, permits the averages to be analyzed using a wide variety of standard statistical techniques.

The absolute minimum number of modules can be computed using $N = 1/p_{min}$, with the result that some test conditions will involve only one module. This total number of modules is not recommended because of the minimal statistical validity that will result for those test conditions involving fewer than five modules.

Spacing Between Stress Levels

As noted previously, the extrapolation procedure of Hoel and Levine[87] indicates that the spacing between stress levels for a given stress are given by $x_i = -\cos (i/K)$, $i = 0, 1, \ldots, K$ for $K + 1$ stress levels located in the interval $(-1, 1)$. For the case of two stress levels, say low and high, it is seen that $K = 1$, $x_0 = -\cos (0) = -1$, and $x_1 = -\cos (1) = 1$. Thus, the low and high stress levels correspond to $x = -1$ and $x = 1$, respectively. For the case of three stress levels, say low, intermediate, and high, it follows that $K = 2$, $x_0 = -\cos (0) = -1$, $x_1 = -\cos (\pi/2) = 0$, and $x_2 = -\cos (1) = 1$. Thus, the three stress levels should be equally spaced at $x = -1, 0$, and 1. For the case of four or more stress levels, unequal spacings are indicated. For example, for four stress levels, it is found that the stress levels should be located at $x = -1, -1/2, 1/2$, and 1.

The extrapolation procedure of Hoel and Levine indicates that the optimum *proportion* of modules that should be tested at each stress level depends on the point at which the precision of the extrapolation is to be maximized. It is interesting to note, however, that the optimal *spacing* among the stress levels does not depend on the extrapolation point. The cosine formula shows that the optimal spacing depends only on the number of stress levels to be used in the experimental design.

Final Engineering Design of Test Conditions

Figure 8 shows the final engineering design for the example module resulting from the present effort. This hierarchical tree differs from that given in Figure 4 in several respects. The number of temperature cycling levels is increased from two to three. In general, five levels of temperature are recommended. However, for the module–design example considered here, the fabrication materials, especially the glass cover, are considered to be relatively insensitive to the temperature range involved in the test. The introduction of a third level for temperature per-mits some assessment of quadratic relationships. Figure 8 also shows that two levels of relative humidity are included in the test design for each level of temperature. The effect of SO_2 and UV are tested at the two higher temperature levels. The high and low levels of the variables correspond to the right and left branches, and are coded as 1 and -1, respectively. The actual test conditions corresponding to the codes are taken to be those previously defined for the example module. The intermediate level of the temperature cycle is coded as 0, and corresponds to a temperature cycle between -40 and 130 C to give a temperature range of 170 C.

Lagrange polynomials are fitted to this design in order to estimate the number of modules required to optimize the extrapolation to normal conditions. The normal conditions are given by the coded values: $(x_T, x_{SO_2}, x_{RH}, x_{UV}) = (-6, -2, -3, -2.5)$, as developed earlier in this report. The computed numbers of test modules for each test condition are shown by the termi-nal box numbers in the tree. These numbers have been rounded to the nearest five modules for each test condition. The number of modules are computed so that the minimum number of modules tested at any condition is five.
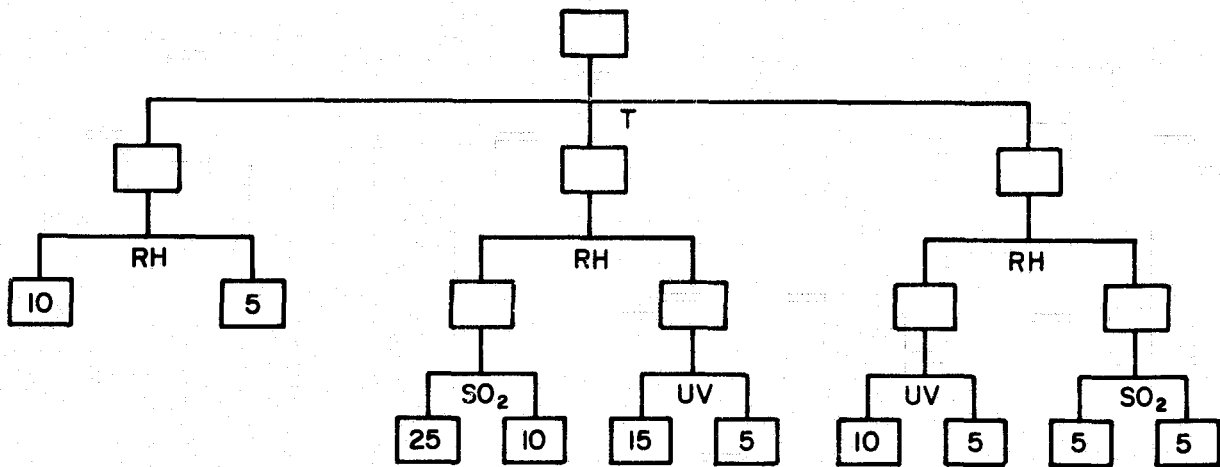
FIGURE 8. FINAL ENGINEERING DESIGN INVOLVING THREE
LEVELS OF TEMPERATURE CYCLING

(Numbers indicate the number of modules to be tested
at each test condition)

The engineering test conditions defined by Figure 8 are:

Test Condition	T, SO_2, RH, UV)
1	$(-1\ z\ -1\ 1)$
2	$(-1\ z\ 1\ 1)$
3	$(0\ -1\ -1\ z)$
4	$(0\ 1\ -1\ z)$
5	$(0\ z\ 1\ -1)$
6	$(0\ z\ 1\ 1)$
7	$(1\ z\ -1\ -1)$
8	$(1\ z\ -1\ 1)$
9	$(1\ -1\ 1\ z)$
10	$(1\ 1\ 1\ z)$

The values of the z's may be chosen to be -1, 0, or 1. Different choices will permit the ten engineering tests to be imbedded into different experimental designs having improved statistical properties. The development of alternative statistical designs is carried out as illustrated earlier for the initial engineering design.

Required Parametric Studies

The recommended procedure for designing an accelerated test involves a number of stages. Several iterations are to be expected among those stages in order to arrive at the best compromise design that is consistent with time and cost constraints. A number of design parameters

can be explicitly considered within the recommended methodology. Specifically, it is recommended that parametric studies be made to determine the effects on the total number of tests and the total number of modules that result from variations in the following design parameters:

- The number of stress levels for each stress

- The location of the extrapolation point

- The number and type of supplementary tests added to achieve improved statistical design.

Such parametric studies may be carried out in the manner suggested by the preceding examples.

Selection of the Final Design

The final design should represent the best possible compromise between engineering and statistical considerations that is consistent with time and cost constraints. In general, the final design is expected to emerge from the parametric studies. If several designs are competitive, other considerations such as the required instrumentation, may be used to make a final choice.

The procedure recommended above can also provide useful documentation for those designs that are not selected. The strengths and weaknesses of each alternative design, reasons for the consideration of the design, reasons for rejection, etc., all serve to better document the factual basis for the final selection. Because of the enormous costs associated with accelerated testing, the desirability of such documentation is obvious.

A Needed Refinement

The Hoel and Levine extrapolation procedure as outlined above involves an important simplifying assumption. It is assumed that the variance, σ^2, of the dependent variable does not change with stress level. If the hierarchical tree is used as a basis for analysis, the validity of this assumption must be challenged. For accelerated testing, equal variances are not likely to be associated with different stress levels. It is possible, for example, that tests conducted at high stress levels may yield smaller variances for measured values simply because degradations and failures occur in short time intervals during which time the test conditions are well controlled. The tests conducted at low stress conditions over time periods of many months may yield increased variability among measured results simply because of the inability to control test conditions precisely over long time periods. It thus appears desirable to assume that each test condition is associated with its own variance. An initial examination of the multi-variable case suggests that a generalization could be made for the design methodology outlined above for accelerated testing. The generalization would change the optimal allocation of modules among the test conditions. The optimal allocation of modules at a given stress condition would be partially dependent on the magnitude of the variance of any dependent variable measured at the stress condition. This variance, in turn, would depend on the precision of measurement. With this refinement, it would then be possible to determine trade-off cost relationships between instrument precision and the desired precision at the extrapolated point representing normal stress conditions.

As described earlier, the analysis of accelerated test data requires two distinct stages. The first stage consists of fitting a model to the values of a dependent variable measured at specified times. The fitted model will have parameters. Ideally, the same model will be found suitable for each test, so that differences between test conditions correspond to differences between corresponding parameter values. The second stage of analysis consists of determining how the parameter values depend on the various stresses. This stage is difficult because it usually is not known which stresses are important; which parameters are related to one kind of stress but not another; whether some stress measure should be transformed, or combined with other stress measures; etc. As a particular example, consider the stress associated with a temperature cycle. In combination with SO_2, temperature should be scaled according to $1/T$ to reflect the fact that a chemical reaction is probably involved. In combination with RH, however, the relevant measure for degradation may consist of the number of freeze-thaw cycles that occur below the dew point temperature. In this case, a "counting" scale, rather than $1/T$, would be appropriate for temperature. In combination with UV, a proper temperature scale could be different from either of these.

These arguments suggest that the scaling of each stress measure is not obtained by the simple assignment of –1 for the lowest stress level and +1 for the highest stress level. In order to be satisfactory for purposes of data analysis, these codes must correspond to mathematical scale transformations of a real variable. If such is the case, then the hierarchical tree and its model may serve as a basis for data analysis.

<div align="center">

Summary of Procedures Used in
the Design Methodology

</div>

A proposed methodology for the design of accelerated tests is described by an example applied to the photovoltaic module shown in Figure 3. An examination of the module design and fabrication materials serves to identify the expected failure modes. The stresses believed to affect the rates of degradation for each failure mode are shown in Table 6. Further examination shows that the expected dominant failure mode under normal stresses involves the interfaces associated with silicone, FEP fluorocarbon, and silver/copper leads. Water–vapor transmission, delamination, and increased resistance or shorting constitute the failure modes. A complete factorial experimental design shown in Table 7 is conceptually analyzed in order to estimate the relative severity of 16 possible test conditions including high and low stress conditions for stresses associated with temperature, sulfur dioxide, relative humidity, and ultraviolet radiation. The severity of each test combination is estimated in a systematic manner to account for the separate effects of important interactions of the stresses. The severity ratings are then analyzed in Table 8 using Yates' method to compute the expected main effects and interactions of all test variables. Successive Yates analyses are used to obtain a graphical representation of the expected conditional main effects. This representation is shown in Figure 4 as a hierarchical tree, which is used as a basis to reduce the number of tests from 16 to 5, as shown by the terminal cells of the pruned tree in Figure 5. Before applying the extrapolation procedure of Hoel and Levine, the pruned tree is represented algebraically by means of Lagrangian polynomials. Numerical scales for temperature cycling, relative humidity, sulfur dioxide, and UV radiation are constructed in order to obtain coordinates for the extrapolation point associated with normal operating conditions at various geographic locations. Finally, a generalized version of Hoel and Levine's extrapolation procedure is used to identify the proportions of the total number of test modules that should be tested at each of the five tests in order to maximize the precision of the extrapolations to normal stress conditions using test results obtained from accelerated testing.

The recommended procedure then calls for consideration of various alternative designs, such as those shown in Figures 6 and 7. The ideal number of stress levels for a given stress is taken to be five. The minimum number of modules to be tested at any given stress combination is also taken to be five. The spacing between stress levels is taken to correspond to the Chebychev points in the interval –1 to 1. Statistical assessments related to the underlying model, main effects, interactions, confounding, etc., are also included to determine whether supplementary test conditions are justified. These quantitative studies serve as a basis for parametric studies which, in turn, lead to the selection of the final design (see Figure 8 for the example).

TESTING AND INSTRUMENTATION TO MONITOR DEGRADATION OF MODEL MODULE, MODULE INTERFACES, AND MODULE MATERIALS

After the experimental design is completed to the extent that sample size, environmental factors, stress levels, etc. are chosen, the emphasis should turn to instrumenting the test design. In the illustrative example, the various predicted modes of failure have been identified in Table 6. One should find for each mode of failure a method or methods of test that will enable measurement of properties that are precursors to failure. Primary emphasis will be on failure modes with the highest ranks in Table 6, since these have been judged, a priori, to have the highest probability of failure. It is important, however, that all conceivable sources of failure be considered in the accelerated testing program to insure against erroneous *a priori* judgments. It becomes quite evident from a consideration of Table 6 that not all of the properties leading to failure can be monitored directly on the assembled module. To follow such degradative changes, it will be desirable to expose sub-systems (i.e., interfaces and materials) along with the assembled modules in the accelerated test. For example, for the delamination of the glass/barrier interface, several specimens of the fluorocarbon barrier adhered to glass would be exposed along with the assembled modules during the accelerated or abbreviated test. Following preselected exposure times, samples would be removed from exposure and tested. Individual materials would be treated in the same manner. In some cases, the same specimens can test both materials and interfaces. For example, after the fluorocarbon/glass adhesive strength is measured, transmittance could be measured on the glass part of the test specimen. The need to economize on test samples will, of course, depend upon the space available for exposure testing. In all cases, the deliverable output power of the module — the ultimate quality factor — should be measured under standard conditions. For most if not all modes of failure, it is anticipated that more than one test method will be capable of following the degradative process which leads to failure. One therefore requires some means of discrimination among candidate test methods. The most appropriate criteria for such discrimination are precision and cost.

Discriminations Among Test Methods

Since precision and accuracy are sometimes confused, a clear distinction needs to be made before discussing precision. In short, precision is a measure of closeness of repeated measurements of a property, whereas accuracy is a measure of how close a test measurement is to some known reference value. These concepts are illustrated in Figure 9. Here, the two bell-shaped curves represent the spread of test values for Testing Methods 1 and 2 with means and standard deviations μ_1 and μ_2, and σ_1 and σ_2, respectively.

The true reference value is indicated on the abscissa. In this example, Test Method 2 has a smaller standard deviation than Test Method 1 and is therefore more precise. On the other hand, the population mean, μ_2, for Method 2 is farther from the true value than the mean Tests 1, μ_1. Therefore, one says that Method 1 is more accurate than Method 2. Thus, accuracy and precision are two independent concepts. In this study, precision, not accuracy, is of primary concern since the interest is in property changes. To illustrate this, suppose that a test is inaccurate by a factor δ, i.e., instead of getting the true value the measuring instrument gives $\eta + \delta$. Then for two η measurements, at times t_1 and t_2, the rate of property change would be $[(\eta_2 + \delta) - (\eta_1 + \delta)]/(t_2 - t_1) = (\eta_2 - \eta_1)/(t_2 - t_1)$. The inaccuracies cancel out in measuring the rate of change.
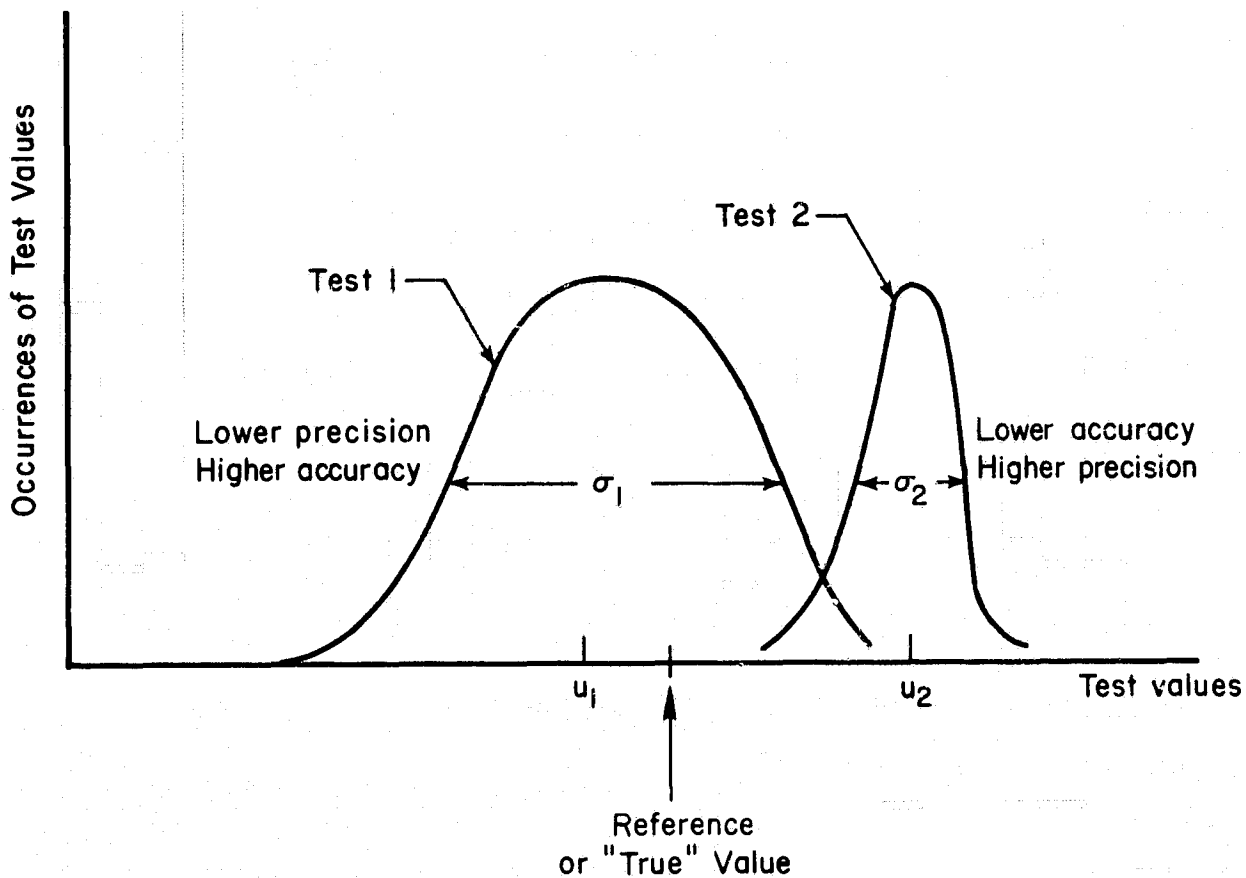
FIGURE 9. ILLUSTRATION OF ACCURACY AND PRECISION

Measures of Precision

As described above, precision is a measure of the spread (or dispersion) of data values around their mean value. An obvious measure of dispersion is the mean deviation, M, where

$$M = \frac{\sum\limits_{i=1}^{N} |x_i - \overline{x}|}{N} . \tag{28}$$

N is the number of data points, x is the data mean, and x_i is the value of an individual measurement. This measure is difficult to manipulate mathematically and is, therefore, seldom used in statistical computations. A better measure of dispersion, which is most commonly used, is the standard deviation, S, where

$$S = \left[\frac{\sum\limits_{i=1}^{N} (X_i - \overline{x})^2}{N - 1} \right]^{1/2} \tag{29}$$

86

Inspection of Equation (29) reveals that the standard deviation carries the same units as the data. This is usually a desirable property for any measure of dispersion. For comparing different testing methods, however, the standard deviation is unsatisfactory. For example, suppose one is using two tests with modulus expressed in psi and elongation at break expressed in percent to determine the degree of vulcanization of an elastomeric compound. The modulus shows a standard deviation of roughly 70 psi and elongation a deviation of approximately 20 percent. Clearly, this is insufficient information to favor one test over the other because one cannot usefully compare psi and percentage units.

One logical alternative would be the use of the coefficient of variation, CV, where

$$CV = \frac{S}{\bar{x}} . \tag{30}$$

Clearly, the coefficient of variation is unitless, and as such would be expected to allow comparison of different test precisions. This is true in some cases, but not generally.

Precision of Selected Tests

Perhaps the largest repository of information on test accuracy and precision is the American Society for Testing and Materials published test methods. The methods for which the coefficient of variation are given or calculable from the given data are shown in Table 12, along with their respective coefficients of variation. This table reveals a wide range of precision, and underscores the need for a systematic methodology for test discrimination. As noted previously, comparison of coefficients of variation is usually invalid for comparing different tests, and such statistics are of limited value. The problem of test discrimination, however, has not gone unattended. Mandel [106,107] recognized the problem and defined a relatively sensitive statistic for test discrimination.

Mandel's Sensitivity Ratio

Suppose that two test methods, M and N, both measure the same property, say Q. If M is a function of Q and N is a function of Q, it follows that M is functionally related to N. Suppose that the standard deviation of test M is σ_M and that of test N is σ_N. Mandel has shown that the two tests can be compared by a quantity called the relative sensitivity of test M with respect to test N, RS(M/N), where

$$RS(M/N) = \frac{|dM/dN|}{\sigma_M/\sigma_N} \tag{31}$$

Note that the relative sensitivity is unitless and favors test M when RS(M/N)>1 and test N when RS(M/N)<1. This concept is illustrated in Figure 10 where the differentials are shown as ΔM and ΔN. RS(M/N) is not necessarily constant, but will vary from point to point depending upon the shape of the sensitivity curve and whether or not the standard deviations are constant. Thus, one cannot say that test M is "better" than test N unless RS(M/N)>1 for all values of M and N. In the special case where dM/dN, σ_M, and σ_N are constant, the relative sensitivity will, of course, also be constant.

87

TABLE 12. COEFFICIENT OF VARIATION FOR VARIOUS ASTM TESTS ON PLASTIC MATERIALS

Test	Coefficient of Variation[a], percent
Density	<1
Linear Dimensional Change	<1
Dilute Solution Viscosity	1
IR Absorbance of CH_3 on Polyethylene Type II	2
Coefficient of Linear Thermal Expansion	2
Water Vapor Transmission	2.5
Barcol Hardness	1-3
IR Absorbance of CH_3 on Polyethylene Type III	7
High-Speed Tensile at Break	4-8
Tensile Impact Energy	6-12
IR Absorbance of CH_3 on Polyethylene Type I	15
High-Speed Elongation at Break	15-18

(a) Coefficient of variation is standard deviation expressed as percent of mean.
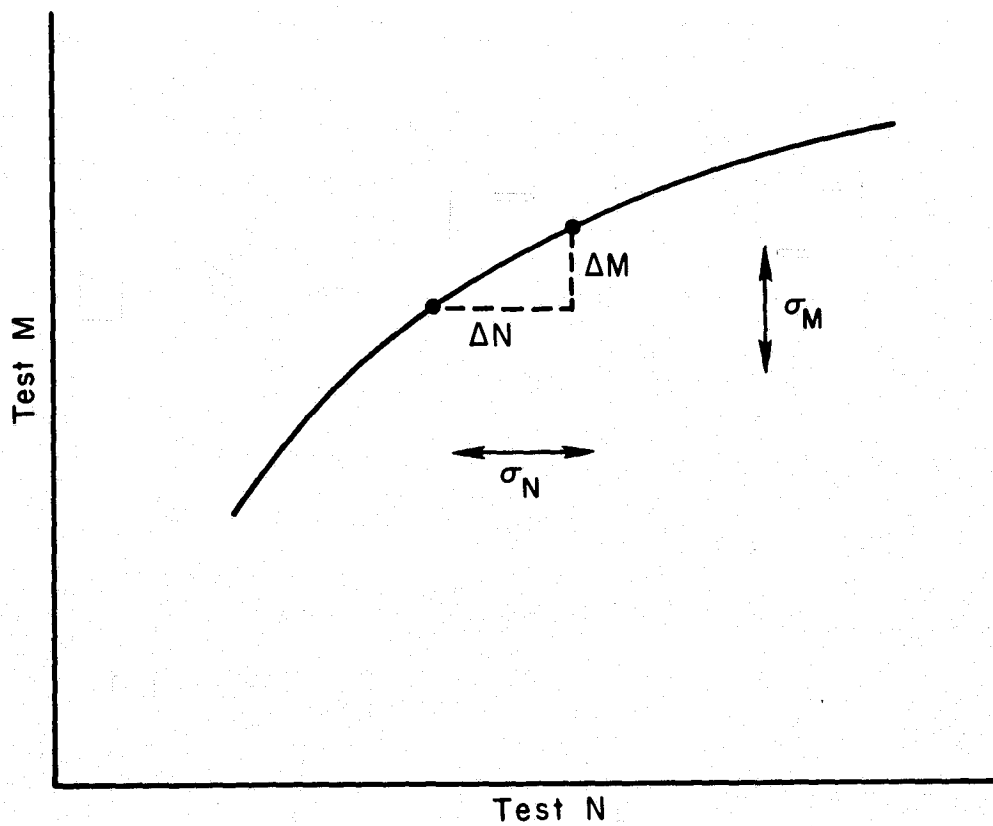
FIGURE 10. ILLUSTRATION OF RELATIVE SENSITIVITY
OF TWO TESTS, M AND N

As an example, suppose that one has the properties, P_1 and P_2, from two different test methods, each of which can follow the degradation of, say, a polymer film as a function of exposure time, t. Furthermore, suppose that P_1 and P_2 have equal standard deviations and decrease with exposure time, t, according to the following relationships,

$$P_1(t) = P_{10} + b_1 t + b_{11} t^2 \tag{32}$$

$$P_2(t) = P_{20} + b_2 t \quad , \tag{32a}$$

where P_{10} and P_{20} are initial property levels for tests P_1 and P_2, respectively, $P_1(t)$ and $P_2(t)$ are property levels at time t for tests P_1 and P_2, and b_1, b_{11}, and b_2 are constants. Solving Equation (32a) for t and substituting in Equation (32), one gets

$$P_1(t) = P_{10} + b_1 \left[\frac{P_2(t) - P_{20}}{b_2} \right] + b_{11} \left[\frac{P_2 - P_{20}}{b_2} \right]^2 \quad . \tag{33}$$

Then the relative sensitivity of test P_1 with respect to test P_2, $RS(P_1/P_2)$, is as follows:

$$RS(P_1/P_2) = \frac{dP_1}{dP_2} = \gamma_0 + \gamma_1 P_2(t) \quad , \tag{34}$$

where

$$\gamma_0 = \frac{b_1}{b_2} - 2P_{20} \left(\frac{b_{11}}{b_2^2} \right) \tag{35}$$

and

$$\gamma_1 = 2 \left(\frac{b_{11}}{b_2^2} \right) \quad . \tag{36}$$

As a specific example, suppose one has

$$P_1(t) = 100 - 5t - 0.5t^2 \qquad 0 \leqslant t \leqslant 10 \tag{37}$$

and

$$P_2(t) = 100 - 10t \qquad 0 \leqslant t \leqslant 10 \quad . \tag{38}$$

These relationships are shown in Figure 11.

Then $\gamma_0 = 1.5$ and $\gamma_1 = -0.01$, and one has

$$RS(P_1/P_2) = 1.5 - 0.01 P_2 \quad . \tag{39}$$

This relationship is shown in Figure 12, and indicates that at $P_2 = 50$ (or t = 5) the two tests are equally sensitive. At $P_2 < 50$ (t < 5), P_1 is more sensitive, and at $P_2 > 50$ (t > 5), P_2 is more sensitive. The results are similar if $\sigma_1 \neq \sigma_2$, in which case γ_0 and γ_1 are both multiplied by the constant σ_2/σ_1.
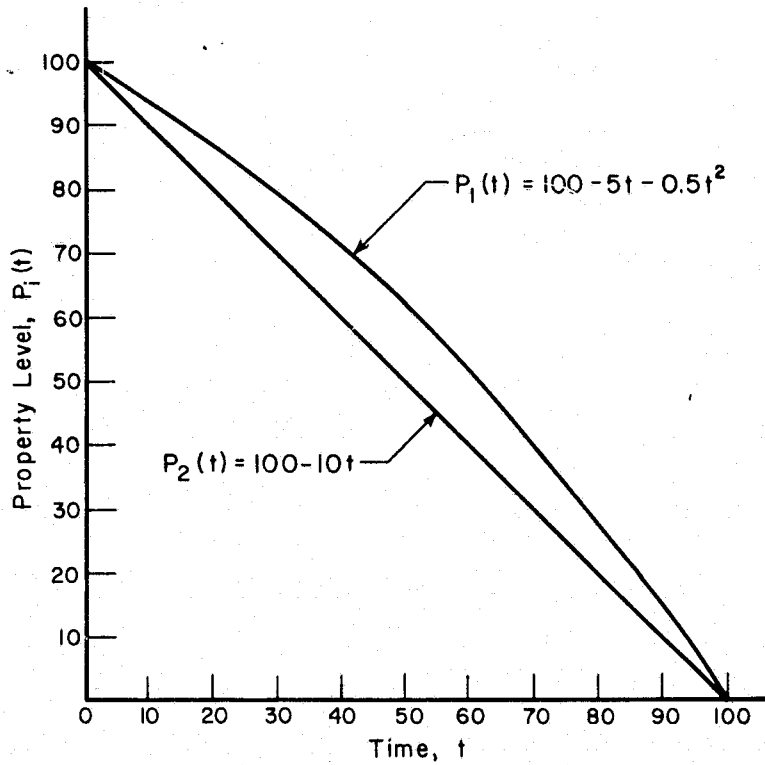
FIGURE 11. PLOT OF PROPERTIES P_1 AND P_2 AS FUNCTIONS OF TIME

In Figure 11, the curves are labeled:

$P_1(t) = 100 - 5t - 0.5t^2$

$P_2(t) = 100 - 10t$
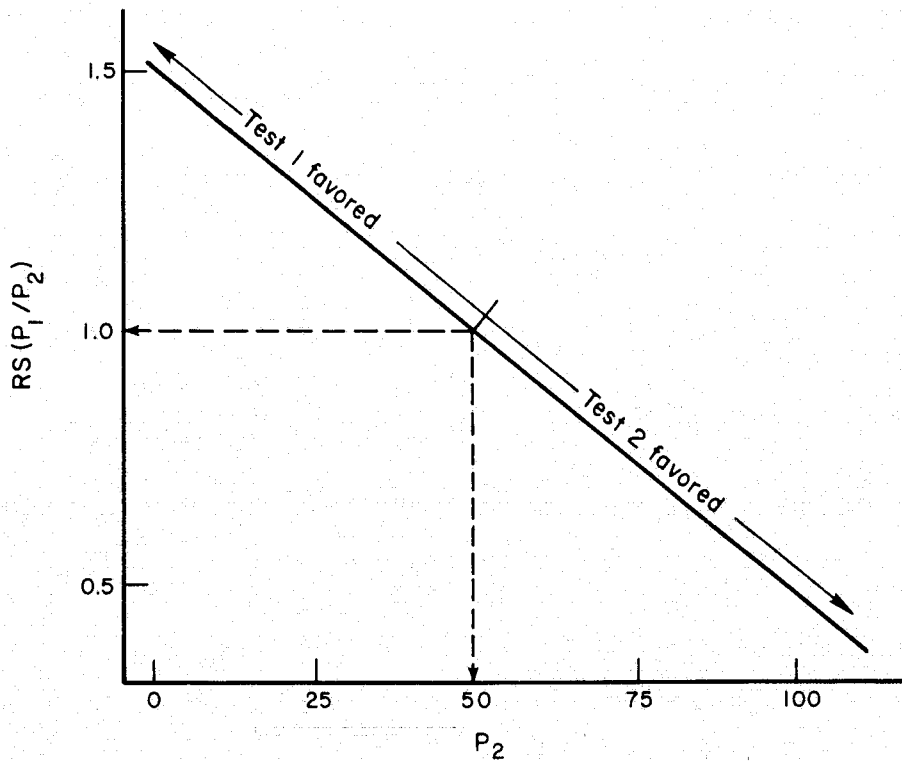
FIGURE 12. RELATIVE SENSITIVITY AS A FUNCTION OF TEST P_2 FOR HYPOTHETICAL EXAMPLES

Transformations of Scale

For any scheme for test-method discrimination to be completely valid, transformations of scale should not affect the discrimination results. For example, suppose one has determined that elongation at break is twice as sensitive a measure of the degree of vulcanization of rubber as tensile strength. If one modifies the tensile test to give results on a logarithmic scale, is the sensitivity ratio still 2 to 1? Mandel has shown that the relative sensitivity is unchanged by a transformation of scale for one or both tests. This result is important enough to repeat its deviation here.

For tests M and N, the relative sensitivity is given by Equation (30). Now suppose that one transforms the scale of M as follows:

$$M^* = f(M) \qquad (40)$$

where M^* is the transformed version of M. Then the relative sensitivity of M with respect to N is

$$RS(M^*/N) = \frac{|dM/dN|}{\sigma_{M^*}/\sigma_N} \qquad (41)$$

Now since M is a function of N, then so is M^*. Then by differentiating (40) with respect to N one obtains

$$\frac{dM^*}{dN} = \frac{df(M)}{dN} = \frac{df(M)}{dM} = \frac{dM}{dN} \qquad (42)$$

Now by the law of propagation of errors, one can write

$$\sigma_{M^*} = \left| \frac{df(M)}{dM} \right| \sigma_M \qquad (43)$$

Then by substituting Equations (42) and (43) into Equation (41), one gets

$$RS(M^*/N) = \frac{\dfrac{df(M)}{d(M)} |dM/dN|}{\dfrac{df(M)}{dM}(\sigma_M/\sigma_N)} = \frac{|dM/dN|}{\sigma_M/\sigma_N} = RS(M/N)$$

Thus, changing the scale of M does not change the relative sensitivity.

Sensitivity and Coefficient of Variation

Coefficients of variation are often used to measure precision and are often referenced in the literature. Thus, it would be conventient if this statistic could somehow be used for

discrimination among test methods. Mandel has shown that this is possible in cases where test values are related by one of the following relationships

$$M = kN \qquad\qquad (44a)$$

$$M = k/N \qquad\qquad (44b)$$

where M and N are data from the respective tests and k is a constant. This is a rather restrictive condition but may be satisfied in some practical cases.

Inferences Concerning the Sensitivity Ratio

The discussion of the sensitivity ratio thus far has dealt with population parameters, σ_M, σ_N, and slope, dM/dN. All of these parameters must be estimated from actual data, and are, therefore, subject to random fluctuations. A means is required to determine if a calculated sensitivity ratio is significantly greater than unity. Unfortunately, an exact solution is not possible, and it is necessary to resort to an approximation. First, it is assumed that the slope dM/dN, estimated by, say, K, is sufficiently precise that it can be ignored as a source of variation for the sensitivity ratio. This is often a reasonable assumption. Since the sensitivity ratio would normally be estimated from data from a well-designed experiment, the experimenter can usually minimize the variability of K. The variation in the sensitivity ratio is then assumed to be a result only of the variation in estimates of σ_N and σ_M, denoted S_N and S_M. It is known that the quantity $(S_N^2/\sigma_N^2)/(S_M^2/\sigma_M^2)$ follows an $F_{i,j}$ distribution, where i and j are the number of degrees of freedom associated with S_M and S_N, respectively. Now it can be shown that the quantity $[|K| (S_N/S_M)]/\sqrt{F}$ represents a lower confidence limit for the sensitivity ratio $|K| (\sigma_N/\sigma_M)$. If this lower limit exceeds unity, it is concluded that, at the confidence level chosen, M is more sensitive than N. If it does not exceed unity, it must be concluded that the data are consonant with the hypothesis of equal sensitivities. In most instances, however, a choice has to be made between two tests, and it would be sensible to select the one with the higher calculated sensitivity, regardless of the lack of statistical confidence at the desired level.

Sensitivity, Sample Size, and Cost

In discriminating among candidate test methods, cost as well as precision must be considered. For example, if σ_i is the standard deviation for a single measurement from test i, the standard deviation for a *mean* of N observations is $\sigma_i/\sqrt{N}$. This illustrates that the precision of any test method can be increased to any level desired simply by increasing the number of replicates, N. It is therefore necessary to consider testing costs and sample size in any test-discrimination methodology. These factors can be easily incorporated into the relative sensitivity concept.

Mandel has shown that if RS(M/N) = k, a single measurement by method M has the same precision as the mean of k^2 measurement by Method N. Furthermore, he has shown that if C_M is the cost of a single measurement for method M, and C_N is the cost for method N, then the ratio of costs, C(M/N), to achieve equal precision is

$$C(M/N) = \frac{C_M}{K^2 C_N} \quad \text{or} \quad \frac{C_M}{[RS(M/N)]^2 C_N} \quad . \qquad (45)$$

Some examples are given in Table 13. In example 1, for equal costs and a sensitivity ratio of 1, the cost ratio is 1. If the cost of test M is doubled with no change in sensitivity ratio, the cost ratio is also doubled. In example 5, if test N costs twice as much as test M, but test M is twice as sensitive as test N, then one would have to spend 8 times as much for test N to get the same precision as in test M. One is therefore able to evaluate various testing methods not only on the basis of precision but also on the basis of cost.

<div align="center">

TABLE 13. EXAMPLE OF COST RATIO AT EQUAL PRECISION AS A FUNCTION OF TEST-ING COSTS AND SENSITIVITY RATIO

</div>

	Examples				
	1	2	3	4	5
C_M	1	2	1	2	1
C_N	1	1	1	1	2
RS(M/N)	1	1	2	2	2
C(M/N)	1	2	1/4	1/2	1/8

Statistical Experimental Design for Determining Sensitivity Ratio for Two Test Methods That Measure Property Degradation

The goal of an experiment to determine sensitivity ratio for two tests is the determination of the regression equation of test M results on those of test N, as well as the standard deviations of each test with as much precision as possible and at a reasonable cost. Without prior knowledge of the form of the mathematical relationship between the two tests, at least four levels of the property being measured should be evaluated, preferably five or six. These points should be spread over as wide a property range as possible so that comparisons will be valid at any test level(s) likely to be found in the future.

In accelerated testing, one wants to discriminate among tests which can follow rate of degradation. A reasonable design would be the following in which degradation of a plastic sheet is used as an example, along with measurements from two nondestructive tests. Suppose one desires six levels at which to evaluate the two tests. At each level, a minimum of three replicates are required; so one requires a minimum of 3 x 6 = 18 samples of the plastic sheet. Preferably, these should be taken from random areas of the material to be fully representative. Three samples are randomly selected from the 18, and the remainder are put in equivalent positions in a Weatherometer. At the end of each of five weathering periods, three of the samples are randomly withdrawn from the chamber. The total weathering period should be long enough to allow nearly complete degradation of the plastic. If this time is unknown, more samples will be required so that sample-withdrawal times can be determined in a sequential manner.

After all 15 (or more) samples have been withdrawn from the Weatherometer, they should be combined with the original three samples and tested in random order by both methods. For each sample, a coin should be flipped to determine which of the two tests to run first. This allows for possible changes in the sample due to testing.

The first step in analyzing the data is to calculate the means and standard deviations for each group of replicates for each test. Bartlett's test[108] for homogeneity of variance is then applied. If heterogeneity is not found to be significant, the standard deviations for each test are pooled for a single estimate for each test, i.e., S_M and S_N. A regression analysis is then run for the means of test M onto those of test N. First, a linear relationship should be tried as follows:

$$M = b_0 + b_1 N \quad , \tag{46}$$

where b_0 and b_1 are fitted intercept and slope, respectively. If fit is not adequate as judged by calculation of an F ratio for lack of fit[109], a quadratic equation should be fitted as follows:

$$M = b_0 + b_1 N + b_{11} N^2 \tag{47}$$

and evaluated for lack of fit. The process is continued until adequate fit is found. Once an adequate equation is obtained, the derivative of M with respect to N is taken, and the relative sensitivity calculated from Equation (31). If Equation (46) gives an adequate fit, RS(M/N) will be a constant; otherwise it will be a function of N. The design and analysis are easily extended to cover any number of tests if the RS values are all constant. If they are not, linear and non-linear programming methods are indicated.

Example of Test Selection for Model Module

From Table 6, the material and interface properties which must be measured to monitor model photovoltaic module degradation are (1) light transmittance, (2) mechanical strength, (3) delamination, (4) water vapor transmission, (5) electrical resistance, and (6) voltage endurance (i.e., withstand voltage). One test method and proper instrumentation for measuring for light transmittance or transparency are given in ASTM-D1746-70. The most applicable measures of mechanical strength include stress-strain behavior in tension, ASTM-D-638 for plastics and ASTM-D-416 for elastomers; brittleness temperature for elastomers and plastics, ASTM-D-746; and impact resistance, D3099. Delamination or adhesive strength is measured with a lap-shear specimen as discussed in ASTM-D816. Water vapor transmission measurement is discussed in ASTM E96. An alternative method of measurement employing a film pouch or "pillow pack" was developed at Battelle and is discussed in the final report of Study 3 of the Encapsulation Task of the LSSA contract. Electrical resistance measurement is discussed in ASTM-D1458 and ASTM-D229. Finally, withstand voltage is measured as in ASTM-D2275. These do not constitute an exhaustive listing of possible test methods for each property but are merely given as examples. In cases where testing on actual modules is not feasible, testing is done on subsystems constructed in the same manner as the modules.

Aside from measurement of properties leading to failure modes, it will be possible in some cases to measure precursors to changes in the property of interest. For example, loss of transmittance in a plastic film can result from oxidation. The chemical changes involved in such oxidation can be measured, in many materials, before oxidation results in a measurable decrease in light transmittance. The precursor oxidation products, for example, can be measured by chemiluminescence and attenuated total reflectance infrared spectroscopy (i.e., ATRIR). When such precursors to property degradation exist and can be measured, it is usually highly desirable to do so since earlier prediction of failure time is often possible.

As the above discussion illustrates, multiple test methods for a single property are the rule rather than the exception. As a result, the test discrimination methodology discussed in the previous section will play a major role in test selection.

DATA ANALYSIS PROCEDURES

The data obtained from accelerated life tests are costly because of the long test times and the highly precise measurements required to detect small degradations. The data are complex because of the severity and variety of the imposed test conditions. In contrast to testing materials, additional complexity results from the testing of entire modules, or devices that involve combinations of materials. Moreover, accelerated test data are always somewhat controversial because of the real possibility that the normal stress failure mode has changed at one or more of the overstress conditions. The data analysis procedures must properly account for such changes in failure mode. The prediction of normal-stress performance from data obtained at overstress conditions necessarily involves extrapolation. All extrapolations are hazardous, and those associated with accelerated testing particularly so because of the possibility of a change in failure mechanisms and the general difficulty of defining "normal-stress conditions". Because of the complexity of these problems, no single method of data analysis is recommended. Instead, it is recommended that the widest possible variety of methods be used to analyze the data.

Generalized Approaches to Data Analysis

Partial characterizations of the various methods of data analysis are suggested by the labels conceptual, statistical, and empirical. The relevance of such methods for analysis of accelerated test data is discussed in Reference (92). Consequently, only the main features of these methods are outlined below.

Conceptual Approach to Data Analysis

It is recommended that efforts be made to analyze the data using models derived from physical and chemical laws that may be associated with the degradation processes. Of particular importance are the reaction-rate laws, such as the Arrhenius or Eyring relations.[81] More generally, an effort should be made to construct conceptual models based on physical relationships that may account for observed performance, and then use these models to correlate the data. If successful, this conceptual approach to the analysis of accelerated test data tends to yield an improved understanding of the degradation processes together with better defined limits for the validity of extrapolations. The successful conceptual analysis is the ideal analysis. Unfortunately, such analyses may prove unsuccessful, particularly for accelerated testing of complex modules in early stages of development when the appropriate concepts have not been identified.

Statistical Approach to Data Analysis

The data obtained from accelerated life tests frequently have large "noise" components. This noise results from inherent variability among the tested modules, imprecision in the measurement procedures, difficulties in maintaining strict control over the test environments for long periods of time, etc. In a general sense, statistical methods are designed to identify and quantify the "signal" and "noise" components of data. Consequently, a second approach to the analysis of accelerated life test data should be based on the use of statistical methods. A wide variety of such methods

exists. Parametric methods include those based on least-squares regression and response surfaces, and typically require relatively strong assumptions concerning the structure of the noise component of the data (normal, log-normal, etc.). Non-parametric methods are based on weaker assumptions, but tend to be less powerful and require more data than the parametric methods. An outline of the statistical approaches believed to be most useful for accelerated testing is given subsequently in this section of the report.

Empirical Approach to Data Analysis

Because of the initial lack of understanding that frequently accompanies accelerated testing, it is important to document all kinds of information that at times may appear to be irrelevant to the test program. Such information includes manufacturer, test number, instrument I.D. numbers, technicians, dates, sequences of test measurements, pre-conditioning or burn-in procedures, etc. All too often such empirical information contains important clues that are needed to account for the relationships found to occur in accelerated test data. A portion of the analysis of accelerated life test data should always include such an empirical approach.

Objectives of the Data Analysis

The initial objectives of the data analysis involve two kinds of extrapolations: (1) extrapolations over time to predict future degradations of performance within each test condition, and (2) extrapolations from higher stress test conditions to predict future degradations of performance at lower stress test conditions. Both kinds of extrapolations are intended to be performed several times during the course of the accelerated test program. These extrapolation procedures permit successive modifications to be made in the fitted models as increasing amounts of data become available. Moreover, the documentation of the predictions and subsequent verifications provide a "track record" that indicates the predictive ability of the fitted models. Based on the observed tract records, the final objective of the data analysis consists of making an overall assessment of the ability of the resulting models to predict degradation under normal stress conditions.

Examples of the kinds of models that may be used for extrapolations within stress levels are given later. Special attention is directed towards selection of the best among several alternative models. An example of generating a track record is also included.

Examples of extrapolations from higher stress conditions to lower stress conditions are not included. Although the central problem of accelerated testing involves extrapolation to normal stress conditions using data obtained at higher stress conditions, this kind of extrapolation is not well understood. For this reason, it is recommended that explicit extrapolations be made among the stress levels included in the test program. Clearly, if extrapolation procedures cannot be validated among the over-stress conditions included in the accelerated test, then they cannot be assumed to be valid when extrapolated from the over-stress test conditions to the normal (untested) stress conditions.

Extrapolations Among Stress Conditions

The principle approaches to extrapolations among stress conditions, from primitive to advanced,

are summarized below. In each case the extrapolations must be preceded by fitting models to the data obtained within each stress condition. The models fitted to the data within each stress condition should contain the smallest number of parameters (fitted constants) that will yield valid predictions over time for each stress condition. If the degradation mechanisms are the same for each stress condition, it may be assumed that a single model may be used for all stress conditions. The difference between performance at one stress condition relative to another stress condition is then reflected by corresponding differences among the numerical values of the fitted parameters. In order to identify a model suitable for use across all stress conditions, the empirical, statistical, and conceptual approaches outlined previously are recommended.

In the simplest case, only one parameter, a constant degradation rate, may be required to fit a straight-line model to the data at each over-stress condition. Extrapolations among stress conditions can then be based on "acceleration factors" obtained by taking the ratios of the degradation rates at the higher stress conditions to degradation rates at a lower stress condition. In such a case, for example, it may be found that 1 hour of operation at the highest over-stress condition yields the same amount of degradation as 10 hours of operation at the lowest over-stress condition. This would indicate an acceleration factor of ten between two stress conditions. In such a simple case, no explicit relationships are required to "explain" the acceleration factors in terms of temperature, ultraviolet radiation, or other environmental factors. The relationships are simply statistical correlations that may be used to make predictions at lower stress conditions but do not indicate specific functional dependencies on the environmental factors. Extrapolations based on simple statistical correlations are extremely hazardous and are considered to be unsatisfactory. The inclusion of a statistical design in the test program may reduce the hazard. Such designs are required in order to better identify the contributions of the various environmental factors to the degradation process.

The recommended procedure for obtaining a model based on engineering input is described earlier in this report. In general, the first iteration of this procedure is expected to yield an experimental design which is statistically deficient. As noted earlier, if time and funds permit, the experimental design may be augmented to remove the statistical deficiencies. The resulting design will yield a conventional statistical design such that the relationships between the parameter values and the environmental factors will typically be expressed in terms of main effects and interactions. To the extent that the statistical deficiencies are not removed, then limited, and possibly ambiguous, information will be obtained for these relationships. In many instances, even if a conventional statistical design is used, the design may become severely "degraded" because the data obtained at the higher stress conditions may be associated with a change in failure mode and be invalid as a basis for extrapolation to normal stress conditions.

Selection and Evaluation of Models

Clearly, the first step in applying the model discrimination methodology is to formulate a set of candidate models. For extrapolation over stresses in accelerated testing, one would, of course, want to be sure that the set included models that have been used successfully in the past. Some of these are:

Arrhenius Equation

$$k = \exp \left(A - \frac{B}{T} \right) , \qquad (48)$$

Generalized Eyring Equation

$$k = AT \left[\exp\left(\frac{-B}{kT}\right) \right]\left[\exp\left(CV + \frac{DV}{kT} \right) \right] , \qquad (49)$$

Inverse Power Law

$$\ln k = \frac{A}{V^B} , \qquad (50)$$

where T is temperature, V is a generalized stress variable (a nonthermal stress for the Eyring equation), k is a rate-of-degradation parameter, and all other terms are fitted constants.

In fitting property, P, versus time, t, data, some candidate models might be kinetic models for various orders. For example, for first-order degradation, one has

$$P_t = P_0 \exp(-kt) , \qquad (51)$$

where P is property level at time t, P_0 is the initial property level, and k is the first-order rate constant. Of course, here again, empirical and mechanistic models which have been applied in the literature should be included, examples of which are[110, 111, 112]:

$$P = P_0 + bt \qquad (52)$$

$$P = P_0 + k \log t \qquad (53)$$

$$P = P_0 + k_1 \log t + k_2 \log^2 t \qquad (54)$$

$$P = b_1 \exp\left[-\left(\left(\frac{t+b_2}{b_3} \right)^{b_4} \right) \right] + b_5 \qquad (55)$$

where P_0 and P are property levels at time zero and t, respectively, and all k's and b's are fitted constants. Aside from candidates from the literature, low-order polynominals should be evaluated as yardsticks by which candidate mechanistic models can be evaluated. Various metrics should be evaluated for the polynominals by applying transformations to independent and/or dependent variables. A useful transformation family for this purpose was proposed by Box and Cox[113] as follows:

$$T = \begin{cases} (Y + \lambda_2)^{\lambda_1} & (\lambda_1 \neq 0) \\ \ln (Y + \lambda_2) & (\lambda_1 = 0, Y > \lambda_2) \end{cases} , \qquad (56)$$

where Y represents the dependent or independent variable, T is the corresponding transformed variable, and λ_1 and λ_2 are variable parameters. Variable transformation often results in better fitting and/or more meaningful models.[114] Furthermore, when two or more stresses are involved, variable transformation can result in the elimination of interactions between stresses. Box and Cox [113] point out that such linearization often tends to make extrapolation more valid than for the unlinearized model. The same argument applies to second-order effects (i.e., V^2) in a single stress where a curvilinear relationship is transformed into a linear relationship that is more amenable to extrapolation.

Once a set of models is selected and fitted to the data, standard statistical tests can be employed to eliminate the poorest models. [115] For example, for linear regression models, a nonsignificant regression F ratio, a significant lack of fit F ratio, and/or an abnormal pattern in the residuals would disqualify the poorest models. Discriminating among the remaining models is more difficult and can be accomplished by establishing a prediction track record as discussed previously. With the exception of the prediction track record method and residual analysis, screening models for nonlinear regression models is done in a different way as discussed in Reference 115.

Study With Artificial Data

To illustrate and evaluate the model discrimination methodology described above, artificial data were generated from the "true" equation

$$P = 850 - 40 t \quad , \tag{57}$$

where P is property level at time, t, in years. Property levels, P_i, were calculated for various times and duplicates were taken for each value of t up to 2 years. Normally distributed error with mean zero and variance, $\sigma^2 = 1$ was added to each P_i. The resultant "data" are shown in Table 14. This model was chosen to illustrate the problem of predicting 20-year life from data collected for 2 years. The results, however, are applicable to other extrapolation problems, including extrapolation of property level or degradation rate over one or more stresses. The candidate models are listed in Table 15 and are simply transformed versions of Equation (57). As can be seen, Model 20 is the "true" model.

All of the equation forms shown in Table 15 were fitted to data shown in Table 14. Five regression analyses were run for each model for data covering 0.4 year (N = 8 data points), 0.8 year (N = 16), 1.2 years (N = 24), 1.6 years (N = 32), and 2 years (N = 40), respectively. After running each set of 20 regression analyses for each N, the fitted equations were judged according to two criteria: (1) F ratio for lack of fit and (2) nonrandom residual patterns. All models exhibited F ratios for regression significant at the 5 percent level of significance. The results for the first discrimination phase are shown in Table 16. The earliest eight data points covering a 0.4-year period resulted in no models being rejected. In other words, all 20 fitted models were consonant with the earliest eight data points. After 16 data points had been collected, the procedure was repeated and resulted in elimination of Models 1 and 2. After 40 data points, all but four models had been rejected: Models 5, 10, 15, and 20. These gave 20-year predictions of 416, 228, 315, and 47, respectively, where the 'true 20-year value was 50. It is interesting, and perhaps distressing, that three equations that exhibited excellent fit to the data gave extremely poor predictions! At this point, if the poorest 20-year prediction were acceptable, no further discrimination would be necessary. Suppose, however, that further discrimination was desirable; then one must look at the predictive track records for the four remaining models.

These are shown in Table 17. A matrix was constructed for each model. The row indices represent the time interval over which the data were collected. The column indices represent times for which predictions were made. The means of the two data values for each prediction time are shown at the bottom of the table. The first entry in the prediction matrix for Model 5, i.e., 834.9, is then the predicted value of P at 0.4 year made from data up to 0.2 year. The number in row three, column four, i.e., 789.1, is the P predicted from data for time, t = 1.6 year from data covering 0.8 year, etc. The 20-year predictions for each time interval are shown in the column beside the matrix. Since the quantities of real interest are the differences between predicted values and actual data values (i.e., residuals), residual matrices were constructed for each matrix of Table 17

TABLE 14. ARTIFICIAL 20-YEAR PROPERTY DATA

Time, years	Actual Property Level, P_i	Artificial Data, Actual Property Level + Error $\overline{P_i + \epsilon}$	$\overline{P_i + \epsilon}$
0.1	846	846.6	845.2
0.2	842	843.3	841.1
0.3	838	838.9	838.0
0.4	834	833.0	832.9
0.5	830	830.1	830.6
0.6	826	824.9	825.9
0.7	822	823.5	822.4
0.8	818	819.6	816.8
0.9	814	814.1	815.7
1.0	810	809.6	808.2
1.1	806	805.7	806.4
1.2	802	804.3	802.2
1.3	798	797.3	797.9
1.4	794	795.4	793.5
1.5	790	790.2	788.1
1.6	786	785.5	785.2
1.7	782	783.4	782.3
1.8	778	777.4	776.6
1.9	774	775.6	773.6
2.0	770	770.0	768.7

TABLE 15. CANDIDATE MODELS

Model No.		Model
1	$1/P$	$= b_0 + b_1/(t+1)$
2		$= b_0 + b_1\sqrt{t}$
3		$= b_0 + b_1 \ln(t+0.5)$
4		$= b_0 + b_1/\sqrt{t+1}$
5		$= b_0 + b_1 t$
6	$\sqrt{P}$	$= b_0 + b_1/(t+1)$
7		$= b_0 + b_1\sqrt{t}$
8		$= b_0 + b_1 \ln(t+0.5)$
9		$= b_0 + b_1/\sqrt{t+1}$
10		$= b_0 + b_1 t$
11	$\ln P$	$= b_0 + b_1/(t+1)$
12		$= b_0 + b_1\sqrt{t}$
13		$= b_0 + b_1 \ln(t+0.5)$
14		$= b_0 + b_1\sqrt{t+1}$
15		$= b_0 + b_1 t$
16	P	$= b_0 + b_1/(t+1)$
17		$= b_0 + b_1\sqrt{t}$
18		$= b_0 + b_1 \ln(t+0.5)$
19		$= b_0 + b_1/\sqrt{t+1}$
20		$= b_0 + b_1 t$

Note: b_0 = fitted intercept.

b_1 = fitted slope.

102

TABLE 16. RESULTS OF MODEL DISCRIMINATION FOR VARIOUS SUBSETS OF DATA COVERING A 2-YEAR PERIOD

No. of Data Points	8	16	24	32	40
Time, years	0.4	0.8	1.2	1.6	2.0
Model					
1	X	*	*	**	**
2	X	X	*	**	**
3	X	X	*	**	**
4	X	X	*	**	**
5	X	X	X	X	X
6	X	X	**	**	**
7	X	X	**	**	**
8	X	X	**	**	**
9	X	X	**	**	**
10	X	X	X	X	X
11	X	X	**	**	**
12	X	*	**	**	**
13	X	X	**	**	**
14	X	X	**	**	**
15	X	X	X	X	X
16	X	X	**	**	**
17	X	X	**	**	**
18	X	X	**	**	**
19	X	X	**	**	**
20	X	X	X	X	X

Key: X — Model retained.
  * — Model disqualified due to significant lack of fit F ratio.
  ** — Model disqualified due to abnormal residual pattern.

TABLE 17. PREDICTION MATRICES FOR MODELS 5, 10, 15, AND 20

		Times For Which Predictions Were Made, years					
		0.4	0.8	1.2	1.6	2.0	20
Model 20	0.2	834.8	820.0	805.2	790.4	775.6	109.6
("True Model")	0.4		816.445	779.405	782.36	765.325	-1.5
	0.8			802.416	786.61	770.806	59.6
	1.2				786.57	770.76	59.26
	1.6					769.80	46.4
	0.2	834.82	820.17	805.64	791.25	777.0	269
	0.4		816.60	799.88	783.33	766.96	209
Model 10	0.8			802.66	787.21	771.91	239
	1.2				786.95	771.57	237
	1.6					770.34	228
	0.2	834.77	820.265	806.0	792.0	778.2	472
	0.4		816.730	800.334	784.425	768.496	308
Model 15	0.8			802.89	787.786	772.964	329
	1.2				787.31	772.34	326
	1.6					770.863	318
	0.2	834.9	820.6	806.9	793.6	780.7	451
	0.4		817.0	801.2	786.0	771.3	419
Model 5	0.8			803.5	789.1	775.1	432
	1.2				787.9	773.7	427
	1.6					771.7	420
Mean of Two Duplicate Data Values		832.95	818.2	803.25	785.35	769.35	50[a]

Note: The leftmost data column (0.2, 0.4, 0.8, 1.2, 1.6) is labeled vertically "Years of Data Upon Which Prediction is Based".

(a) True 20-year value with no error.

104

and are shown in Table 18. The residual matrices correspond to the prediction matrices. For example, the first element in the matrix for Model 5, i.e., -1.95, represents the difference between the mean of the two actual data values at t = 0.4, 832.95 and the corresponding predicted value, 834.9. The same correspondence applies to all other matrix entries.

Cumulative Standard Error of Prediction (CSEP)

At each measurement time — 0.4, 0.8, 1.2, 1.6, and 2.0 years — one can compute a cumulative standard error of prediction as follows:

$$CSEP = \left\{ \left[\sum_{i=1}^{c} (r_i)^2 \right] \Big/ c \right\}^{1/2} \qquad , \qquad (58)$$

where c is the number of residuals calculated up to the most recent time for which residuals were calculated, and r_i is the ith residual. For example, for Model 5 at t = 0.4, one has c = 1, r_i = -1.95, so

$$CSEP = [(-1.95)^2/1]^{1/2} = 1.95 \quad .$$

Likewise, at t = 0.8, one has c = 3, r_1 = -1.95, r_2 = -2.4, and r_3 = 1.2. Then

$$CSEP = \left[\frac{(-1.95)^2 + (-2.4)^2 + (1.2)^2}{3} \right]^{1/2} = 1.92 \quad .$$

The CSEP values are given below each residual matrix for each of the four models in Table 18. Plots of CSEP versus time are shown in Figure 13, where Model 17 was included to illustrate the poor predictive power of a noncandidate model discarded in the initial model screening. At the end of 1.6 years, Models 10, 15, and 20 remain candidates since they all are close together on the CSEP scale. At the end of the 2-year period, however, Model 20, the true model, emerged as the best predictive model.

Regression F Ratio as a Model Discrimination Statistic

When a regression equation is fitted to data, three criteria are most commonly applied to judge its adequacy: (1) regression F ratio, (2) lack of fit F ratio, and (3) examination of residuals (i.e., actual–predicted values). If the regression F ratio is significant at a preselected significance level, the lack of fit F ratio is not significant at a preselected significance level, and the residuals do not exhibit abnormal behavior, it was concluded that the fitted equation is an adequate representation of the data. Put another way, one says that the data are consonant with the fitted equation. Notice that nothing is said about how useful the equation can be expected to be for extrapolation or to what degree the fitted equation is of correct functional form to be a mechanistic model. For example, suppose that two fitted equations exhibit acceptable residual behavior and nonsignificant lack of fit F ratio and significant regression F ratios of, say, 10 and 1000. Instinctively, one feels that the equation with an F ratio of 1000 is more likely to be the true model than the one with an F of 10. However, current statistical practice generally would not admit to such a conclusion. Recent work by Box and Wetz[116], Draper and Smith[115], Derringer[117], and Suich and Derringer[118], however, indicate that the higher the regression F ratio, the better the prediction accuracy of the fitted equation. Put another way, the higher the regression F ratio, the "closer" the fitted equation approaches the "true" equation. Clearly then, when extrapolation is the goal, a large regression F ratio should give more assurance than a smaller but statistically

		Times For Which Predictions Were Made, years				
		0.4	0.8	1.2	1.6	2.0
	0.2	-1.85	-1.8	-1.95	-5.05	-6.25
	0.4		1.755	3.845	2.99	4.025
Model 20	0.8			0.834	-1.26	-1.456
("True Model")	1.2				-1.22	-1.41
	1.6					-0.451
		1.85	1.80	2.2	2.58	2.50 ←— CSEP[a]
	0.2	-1.87	-1.97	-2.39	-5.90	-7.65
	0.4		1.6	3.37	2.02	2.39
Model 10	0.8			0.59	-1.86	-2.56
	1.2				-1.6	-2.22
	1.6					-0.99
		1.87	1.82	2.14	2.69	3.15 ←— CSEP
	0.2	-1.82	-2.065	-2.75	-6.65	-8.85
	0.4		1.47	2.916	0.925	0.854
Model 15	0.8			0.36	-2.436	-3.614
	1.2				-1.96	-2.99
	1.6					-1.513
		1.82	1.80	2.08	2.84	3.50 ←— CSEP
	0.2	-1.95	-2.4	-3.65	-8.25	-11.35
	0.4		1.2	2.05	-0.65	-1.95
Model 5	0.8			-0.25	-3.75	-5.75
	1.2				-2.55	-4.35
	1.6					-2.35
		1.95	1.92	2.18	3.43	4.53 ←— CSEP

Years of Data Upon Which Prediction is Based

(a) Cumulative standard error of prediction.
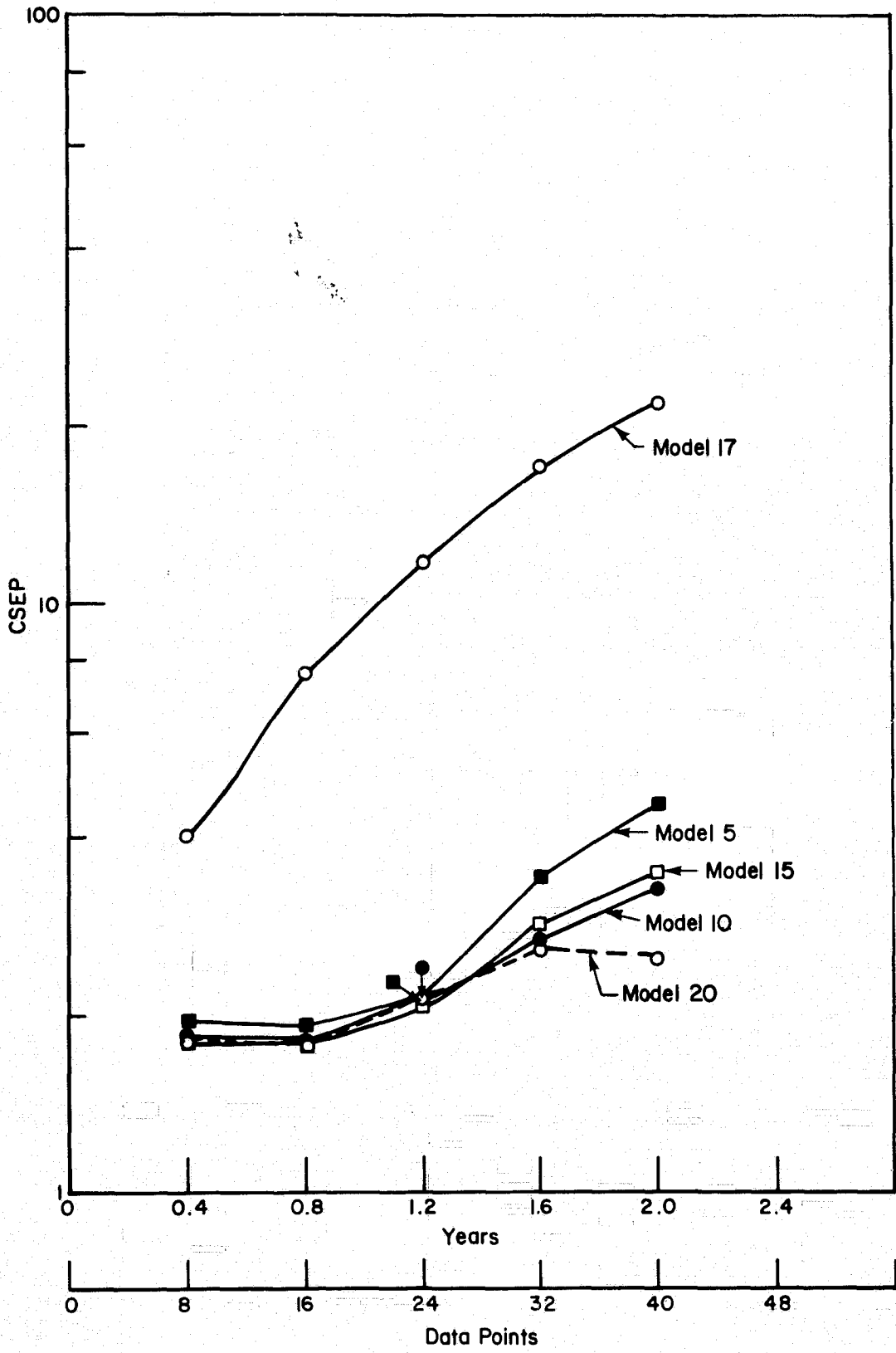
FIGURE 13. CUMULATIVE STANDARD ERROR OF PREDICTION VERSUS
TIME FOR MODELS 5, 10, 15, 17, AND 20

107

significant F ratio. As an illustration of this, consider Table 19, where predictions from Models 17, 5, 10, 15, and 20 fitted to 1.6 years of data (N = 32) were made for 3 through 10 years. As can be seen, the models can be ranked in their order of increasing prediction accuracy as 17 <5 <15 <10 <20, which is the same ranking as their respective regression F ratios, i.e., 920 (model 17) <7654 (Model 5) <8889 (Model 15) <9420 (Model 10) <9847 (Model 20. This strongly suggests that the size of the regression F ratio is an indicator of predictive accuracy.

Regression F ratios for Models 1, 5, 10, 15, and 20 are given in Table 20 as a function of time and number of data points covering a 7.5-year period. Model 1 was representative of the poorest fitting equations eliminated in the first discrimination phase. After 2 years of data were accumulated, all five models exhibited very large F ratios for regression, Model 20 exhibiting the largest in agreement with the CSEP plots. Aside from the magnitude of the F ratio statistic discussed above, its rate of change with the number of data points is quite interesting. The F ratio for the true model increased continuously with N at a very high rate. For the other three models, the F value increased to a point and then began to decrease. This behavior is in agreement with statistical theory. For the linear case, the regression F ratio is equal to MSR/s^2 if the model is of correct functional form, where MSR is the mean square for regression and the residual mean square, s^2, is an estimate of the true experimental error, σ^2. The expected value of MSR is $\sigma^2 + b \Sigma (x_i - \bar{x})^2$, where b is the slope of the true linear relationships, and $\Sigma (x_i - \bar{x})^2$ is the sum of squared deviations of independent variable settings from their mean. As N increased, s^2 approached the constant, σ^2, whereas MSR increased continuously as did the resulting F ratio. If, however, the model being fitted is not of correct functional form, the expected value of residual mean square is equal to $\sigma^2 + (\Sigma B_i^2)/(N-2)$, where B_i is a measure of the discrepancy between the fitted and the true model at each point. Instead of approaching σ^2, in this case, the residual mean square increases continuously at a rate dependent upon the size of discrepancies, B_i. The regression F ratio then increases at a lower rate for an incorrect model and will generally peak and begin to decrease as the B_i becomes larger as data are collected for longer times.

The F-regression versus N values in Table 20 are plotted in Figure 14 for Models 1, 5, 10, 15, and 20. Model 1 was included as representative of the models rejected by conventional statistical measures. Following the curve for increasing N, the first model to be eliminated was Model 1 which peaked out at N $\cong$ 18. Models 5, 15, and 10 were eliminated in that order so that eventually only Model 20, the true model, remained in the set.

In summary then, in the early stages of the model-discrimination process, conventional statistical tests were employed to weed out the poorest predictor equations leaving only those (Models 5, 10, 15, 20) which were consonant with the data. These models were then assessed with plots of CSEP and regression F ratio versus N (or time). Both methods succeeded in selecting Model 20 as the "best" equation in roughly the same amount of time ($\approx$2 years). The discrimination for F, however, does not show up well on the logarithmic scale used in Figure 14. Figure 15 shows plots of the true model data points without error (i.e., column P_i in Table 14) fitted to Models 5, 10, 15, and 20. The resulting equations were:

$$P = 850 - 40 \ t \ \text{(Model 20)} \tag{59}$$
$$\sqrt{P} = 29.16136 \ t \ \text{(Model 15)} \tag{60}$$
$$\ln P = 6.74616 - 0.04953 \ t \ \text{(Model 10)} \tag{61}$$
$$P^{-1} = 0.0011742 + 0.61358 \times 10^{-4} \ t \ \text{(Model 5)} \tag{62}$$

As figure 15 shows, there is very little difference among these four equations when plotted as P versus t in the first 2-year period. In view of this, it is quite encouraging that the correct model was isolated with the proposed methods.

108

TABLE 19. PREDICTIONS FOR YEARS 3 TO 10 FOR VARIOUS MODELS FITTED TO DATA FOR 1.6 YEARS

Time, years	Predicted Value For The Property (P) for Given Model Numbers					True Value of P
	17	5	10	15	20	
3	760	736	732	734	730	730
4	743	704	694	699	689	690
5	727	675	658	665	649	650
6	713	648	622	633	609	610
7	700	161	587	603	569	570
8	688	600	554	574	529	530
9	677	579	521	546	489	490
10	666	559	490	520	448	450

TABLE 20. REGRESSION F RATIO VERSUS N AND TIME FOR MODELS 5, 10, 15, AND 20

Time, years	Data Points, N	F Regression			
		Model 5	Model 10	Model 15	Model 20
0.2	4	8	8	8	8
0.4	8	187	189	189	190
0.8	16	1298	1291	1295	1286
1.2	24	3819	3887	3882	3872
1.6	32	7654	9423	8889	9847
2.0	40	13206	18338	16828	19304
3.0	45	16432	34618	29223	34766
4.0	50	14239	69597	42134	70690
5.0	55	10013	83341	34544	136930
7.5	60	4451	76393	19257	273551
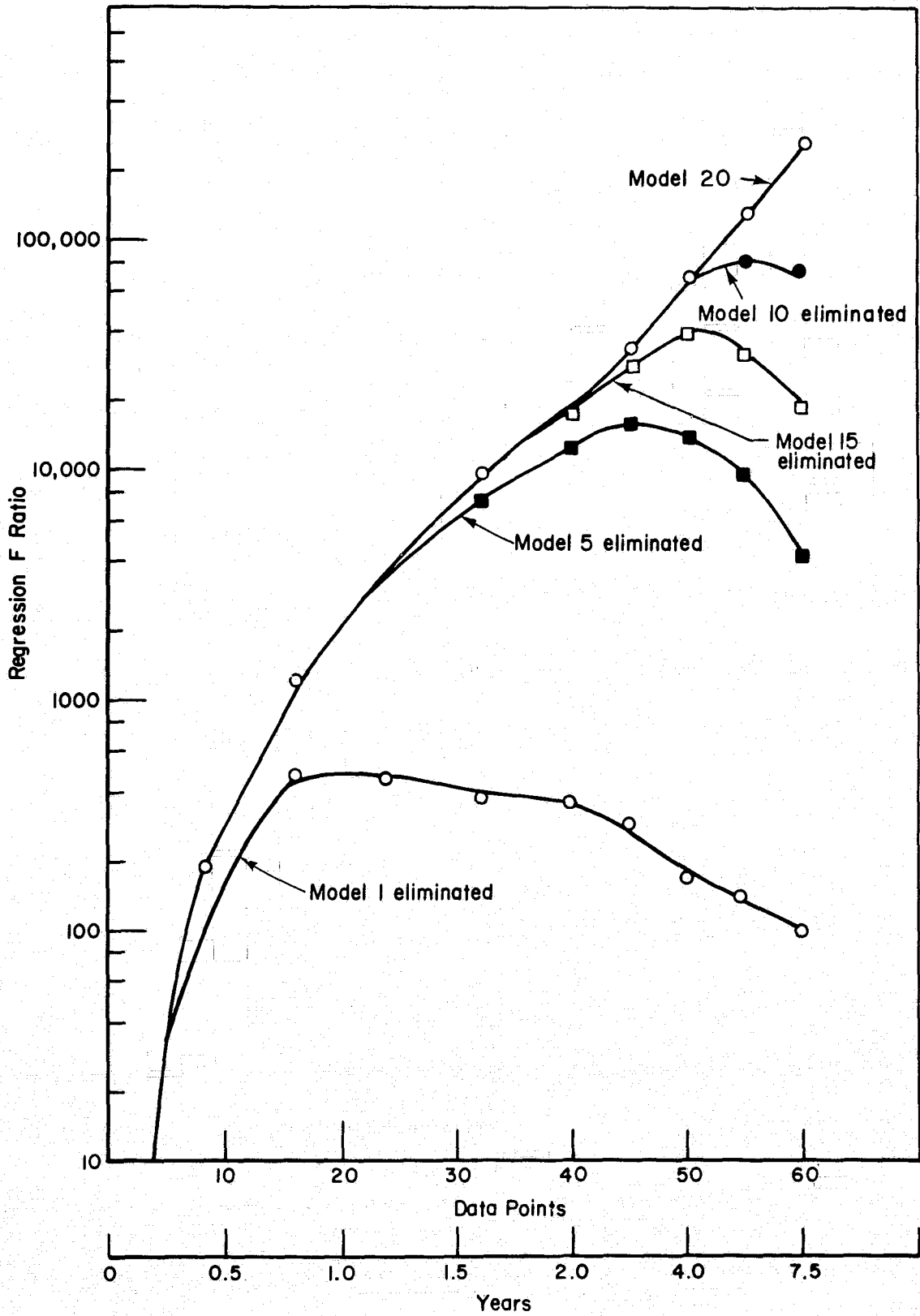
FIGURE 14. REGRESSION F RATIO VERSUS NUMBER OF DATA
POINTS FOR MODELS 1, 5, 10, 15, AND 20

110
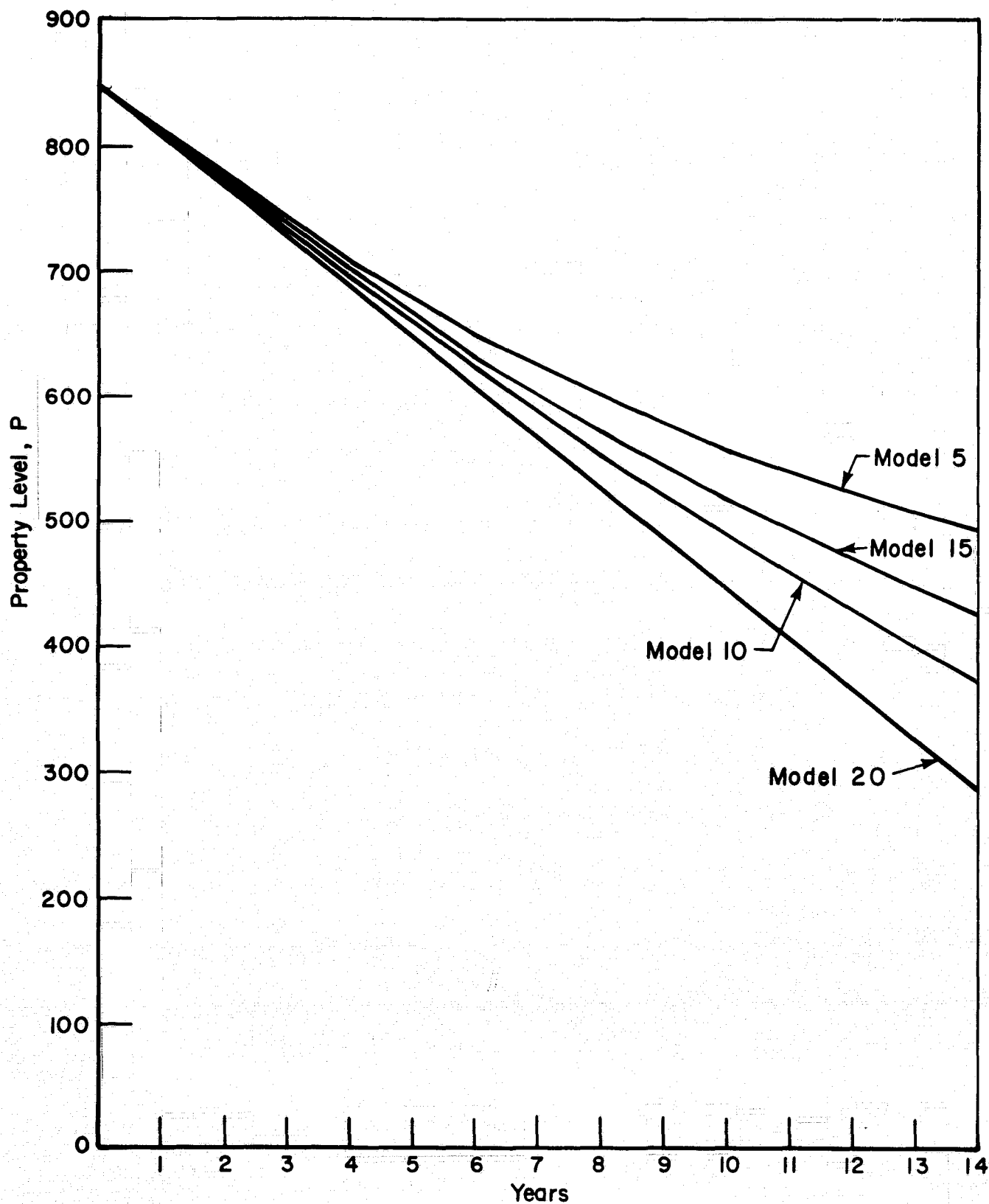
FIGURE 15. PLOT OF PROPERTY LEVEL (P) VERSUS TIME FOR MODELS
5, 10, 15, AND 20 FITTED TO EQ. (57) WITHOUT ERROR

111

In a study by Welch[119], service life of an epoxy resin was studied as a function of two stresses — humidity and temperature — at levels higher than use conditions. The actual data are given in Table 21 rearranged in real-time order to simulate the order in which data were obtained. Regression analyses were run for 8, 10, 12, 14, 15, 16, 17, 18, and 19 data points. Each regression analysis included all points available up to that particular time. The models to which each data set was fitted were as follows:

$$\ell n\ t = b_0 + b_1\ H + b_2\ C \qquad \text{Model A}$$
$$\ell n\ t = b_0 + b_1\ \ell n\ H + b_2\ C \qquad \text{Model B}$$
$$\ell n\ t = b_0 + b_1\ H + b_2/(C + 273) \qquad \text{Model C}$$
$$\ell n\ t = b_0 + b_1\ \ell n\ H + b_2/(C + 273) \qquad \text{Model D}$$
$$\ell n\ t = b_0 + b_1\ \ell n\ H + b_{11}\ \ell n^2\ H +$$
$$b_2/(C + 273) + b_{22}/(C + 273)^2 +$$
$$b_{12}\left[\frac{\ell n\ H}{(C + 273)}\right] \qquad \text{Model E}$$

Model D is derived from kinetic considerations including an Arrhenius relationship between rate of degradation and reciprocal temperature. This is the relationship which Welch used although he treated the data graphically in two phases instead of using a straightforward fitting of Model D. Model E is Model D with second-order terms added to account for any deviations from the assumed semimechanistic relationship. Models A, B, and C were modifications of Model D. The resultant plots for F regression and CSEP versus number of data points (i.e., time) are shown in Figures 16 and 17.

TABLE 21. WELCH[119] DATA ON SERVICE LIFE OF EPOXY RESIN

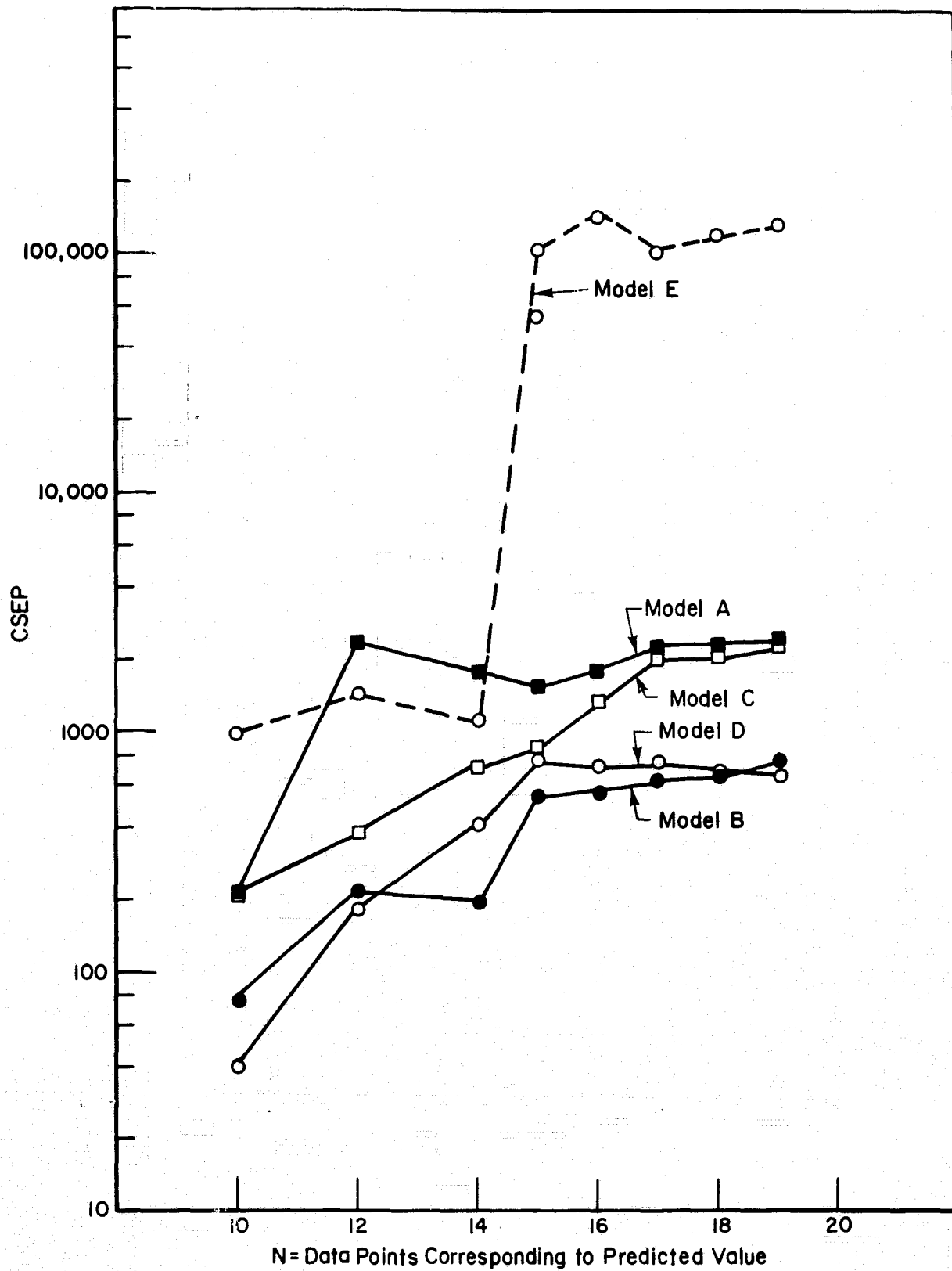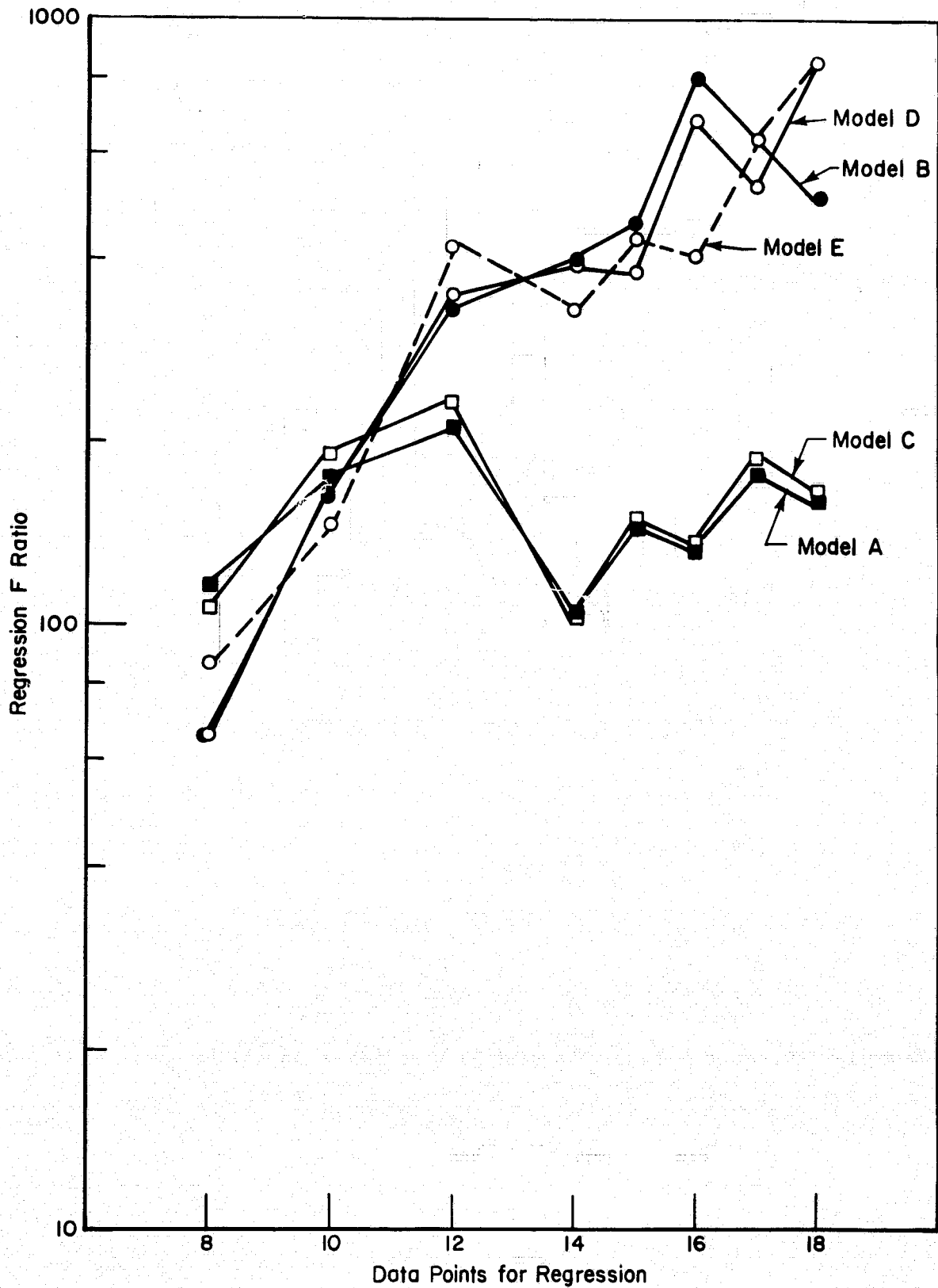| Temperature, C	$[H_2O]$, moles/ liter x 10^2	Time to Failure, hours
95	2.62	54
95	2.26	73
95	2.04	89
95	1.57	112
82	1.65	140
82	1.42	220
82	1.30	230
95	0.523	295
82	1.00	330
66	0.879	690
66	0.750	910
66	0.695	1080
66	0.575	1260
82	0.331	1460
49	0.416	2920
49	0.355	4240
66	0.167	4550
49	0.329	4880
49	0.299	5800

FIGURE 16. CUMULATIVE STANDARD ERROR OF PREDICTION VERSUS
N FOR WELCH DATA AND MODELS A, B, C, D, AND E

113

FIGURE 17. REGRESSION F RATIO VERSUS N FOR WELCH
DATA AND MODELS A, B, C, D, AND E

114

The F plot indicated that Models B, D, and E would be the best candidates for prediction purposes since F generally increased with N as opposed to the relatively flat behavior with Models A and C. The CSEP plot, however, eliminated Model E early in the course of data collection, and at the end of the program, only Models B and D remained candidates. Model D is the semimechanistic model used by Welch for prediction, whereas Model B is Model D with a linear term for temperature replacing reciprocal temperature. Ideally, additional experiments would have been run to discriminate between these two models. In the absence of additional data, a minimax procedure might be used, the prediction would be made from both models and the least favorable result accepted. In any case, the example is in general agreement with the findings from the simulated study with the exception of the plots for Model E. This model gave a high F value but showed poor predictability (high CSEP). The poor predictability is not unexpected since it is generally acknowledged that the higher the order of the polynomial, the poorer the extrapolation accuracy. It appears, therefore, that the F criterion is good for comparison of models of the same order, whereas the CSEP procedure is independent of the order of the polynomial. Of course, further simulations would be required to verify this behavior.

In summary, the proposed data analysis methodology consists of formulating a list of candidate models for property–versus–time data, degradation–versus–stress data, etc. Through various discrimination techniques, the number of viable models is narrowed down as more and more data are obtained. For *all* potentially viable models, an extrapolation is made over time, stress, or other variable. At any point in the model–discrimination procedure, if predicted values made from all remaining models are acceptable, a decision can be made based on this minimax criterion and further experimentation discontinued.

Of course, this methodology is based upon the assumption that the "true" model is in the set of candidates. Actually, the assumption is not as restrictive as it appears. In all life testing that includes an extrapolation step, a successful extrapolation depends, whether stated or not, upon the model being correct. The criterion used to judge this model adequacy is typically how well four or five data points fit a straight line, such as the Arrhenius relationship, a criterion which, this work has shown, is quite hazardous. At least, in the method proposed in this study, great care is taken to give the "true" model a chance to emerge. To make sure the "true" model is an element of the candidate set, the candidate set can be made infinite in a sense. For example, at each data–analysis juncture in the procedure, the models remaining in the candidate set can be put through transformations from the family of transformations shown in Equation (56) using dependent and/or independent variable(s). If an index of fit such as regression F ratio is plotted as a function of λ_1 and/or λ_2, the resulting function is continuous and can be maximized. At each data–analysis junction in the methodology, the F statistic (or other statistic indicator of model adequacy such as CSEP applied retroactively) can be maximized over a domain of λ_1, λ_2 values. This is, in a sense, equivalent to looking at an infinite set of models each time the data are analyzed, and retaining the best one(s), thus maximizing the probability that the true model will be found.

Data Analysis When Degradation Mechanism
Changes Over Time

A major concern with most accelerated life testing is how to extrapolate in instances where the degradation mechanism and the resultant rate of degradation change over time. If the rate of degradation changes discontinuously at some point in time for property–versus–time models or at some stress level for property–versus–stress models, an inaccurate extrapolation will result unless only data collected after the mechanism change are used. For example, in predicting a 20–year property level

from 2 years of data, one has no hope of an accurate extrapolation if the degradation mechanism changes discontinuously after 2 years. From chemical kinetic considerations, it is more reasonable to assume, however, that if the degradation rate changes as a function of time or stress level, it will do so in a continuous manner. This is still a problem when it has been decided a priori to use a particular model (such as the Arrhenius relationship) for extrapolation, since the extrapolation will be from a linear equation fitted to data generated by a nonlinear process. The many literature examples of data plotted as ℓn k versus $1/T$ that show a hint of curvature support the existence of this problem.

In the extrapolation methodology described in this report, a continuous degradation mechanism change with time or stress presents no barrier because the model will change continuously to reflect the change in mechanism. A weighted regression analysis that weighs data points inversely with their age would probably lead to even better results in such situations.

Alternative Procedure

Another procedure for model discrimination, although not employed in the present study, should be mentioned since it is receiving considerable attention in the statistical literature. The procedure was developed by Box and Hill and is based upon the concept of entropy, as defined by Shannon[120], in relation to information theory. Shannon defined entropy, S, as

$$S = -\sum_{i=1}^{k} \pi_i \, \ell n \, \pi_i \, ,$$

where π_i is the probability associated with a symbol i, in this instance a model i. Without repeating here the mathematical details, the resultant discrimination proceeds as follows. A set of candidate models (M_1, M_2, . . . M_i, . . . M_k) is proposed and each is assigned a prior probability, π_i, where, of course, $\sum_{i=1}^{k} \pi = 1.0$. When no reason exists to favor any model(s) over the rest, equal probabilities, $\frac{1}{k}$, are assigned to each model. After a few initial experiments, the probabilities for each model are updated. Then the experimentation and probability updating are continued until one model emerges as clearly the most probable. The example from Reference (121), shown in Table 22, illustrates the evolutionary process. Here n is the experiment number, x_1 and x_2 are independent variable settings, y is the response, and the π_i's are the model probabilities for four models. In this instance, Model 2 evolved as the most probable. In a modification of this method, Hill, Hunter, and Wichern[122] added to the entropy-based model discrimination a criterion of the optimal setting of the independent variables for the next experiment resulting in a joint experimental-design and model-discriminating methodology.

Although these entropy-based methods are rather complex mathematically, they could be programmed on a computer for routine use given sufficient resources to do so. It is, of course, not known at this time whether they would be superior to the track-record or regression F-ratio-based methods. These entropy-based discrimination methods would be a potentially valuable subject for future development.

TABLE 22. EXAMPLE FROM BOX AND HILL AS DISCUSSED IN TEXT

n	x_1	x_2	y	π_1	π_2	π_3	π_4
0	(Prior Probabilities →)			0.25	0.25	0.25	0.25
1	25	575	0.396				
2	25	475	0.723				
3	125	475	0.422				
4	125	575	0.130	0.0069	0.4290	0.5008	0.0633
5	125	600	0.098	0.0019	0.5602	0.4291	0.0088
6	125	600	0.056	0.0018	0.8639	0.1339	0.0004
7	50	450	0.797	0.0021	0.9736	0.0243	0.0000
8	100	600	0.033	0.0032	0.9956	0.0012	0.0000

SUMMARY OF RECOMMENDED PROCEDURES FOR DESIGN
OF ACCELERATED LIFE-PREDICTION TESTS

The procedures involved in designing accelerated tests are complex. To help bring together the various steps involved in the developed methodology, and their sequence, Table 23 identifies in a synoptic form the recommended steps as described in this report. The second column furnishes comments on, and results of, the application of the methodology developed in this study to the selected module design and materials (Figure 3) which were examined as an example in the study. The last column of Table 23 identifies the page in the report where the various steps are treated.

TABLE 23. SUMMARY OF RECOMMENDED PROCEDURES FOR DESIGN OF ACCELERATED LIFE-PREDICTION TESTS

Step Number	General Procedure	As Applied to Selected Example of Module Design	Text Page Number
1	Choose module design and materials.	See Figure 3 for design; the example modules employ a glass front cover, a silicone adhesive and backcover and a fluorocarbon moisture barrier.	55
2	By examination of material surfaces, volumes, interfaces, interconnects, etc., identify the expected failure modes.	See Table 6; seven failure modes are identified, and are associated with delamination, lead corrosion, and electrical breakdown at lead exits.	56
3	For each failure mode, identify the generalized stresses believed to affect the rates of degradation.	See Table 3; temperature cycling, UV radiation, relative humidity and SO_2 are identified as critical stresses.	31
4	Choose that failure mode that is expected to be the dominant failure mode under normal stress conditions.	The potential failure modes are assessed on a scale from 1 to 4, with 4 representing failure modes judged most likely to occur. The ratings are listed in the last column of Table 6.	56
5	For each stress associated with the dominant failure mode, identify the highest stress level that would not be expected to change the dominant failure mode.	The highest values associated with each stress to be included in the test are as follows: temperature, 150C; UV radiation, ten times flux occurring under normal stress conditions; relative humidity, 95 percent at T = 100C; SO_2, continuous exposure to concentration equal to ten times maximum concentration associated with normal stress. Temperature is to be cycled between -40C and 150C at a frequency of six cycles per day. These high stress levels are assigned coded values of +1.	55
6	For each stress level associated with the dominant failure mode, identify the lowest (over stress) level that would be expected to yield measurable degradations within the allowed time and cost constraints.	The lowest values of stresses are as follows: temperature, -40C; UV radiation, maximum flux occurring under normal stress conditions; relative humidity, 50 percent at 100C; SO_2, maximum concentration associated with normal stress. Note that the lowest level of temperature stress involves cycling from -40C to 100C at a frequency of 6 cycles per day. These low stress levels are assigned coded values of -1.	55
7	Form a complete factorial design consisting of all possible combinations of the high and low stress levels for each stress associated with the dominant failure mode.	The sixteen possible combinations of the high (+1) and low (-1) levels of T, UV, RH, SO_2 are expressed as a 2^4 complete factorial design, as shown in Column 2 of Table 7.	57
8	Assume that each combination of stress levels represents an accelerated test condition and estimate the relative severity of each test.	Engineering judgments are made of the relative severity for each of the sixteen combinations of the stress levels. The resulting severity ratings are shown in the last column of Table 7.	57
9	Analyze the severity ratings to determine the main effects and interactions of all stresses.	The main effects and interactions for T, UV, RH, and SO_2 are computed, using the engineering input information. The results are shown in Column 7 of Table 8.	60
10	Generate a hierarchical tree representation of the severity ratings of the complete factorial experiment. For each successive branch (split) in the tree, use that stress that is found to yield the largest main effect conditioned on the levels of the stresses involved in the previous splits.	The hierarchical tree is shown in Figure 4. The largest main effect is shown to be temperature at the top of the tree. Successive high-severity splits are associated with SO_2, RH, and UV, respectively.	62
11	Assume that the terminal conditions of each branch of the tree represent a test condition, and successively eliminate those low-severity test conditions which are judged to be of nearly equal severity.	Eleven low-severity test conditions are eliminated from the hierarchical tree shown in Figure 4. The resulting tree consists of five tests, as shown by the terminal boxes in Figure 5.	64
12	Define the algebraic model that represents the hierarchical tree by means of Lagrangian polynomials.	The algebraic model for the hierarchical tree, shown in Figure 5, is given by Equation (27)(p. 65).	41.65
13	Define a numerical scale for each stress and determine the coordinates of the extrapolated point that represents the normal stress condition. The scale values of -1 and +1 must correspond respectively, to the lowest and highest over-stress test level for each stress.	The coordinates for the normal stress conditions are based on historical data, and are given by $(x_T, x_{SO_2}, x_{RH}, x_{UV}) = (-6, -2, -3, -2.5)$.	68
14	Identify statistical deficiencies in the resulting experimental design, and augment the design by adding test conditions to remove as many of these deficiencies as possible within the time and cost constraints of the resulting tests.	The five test conditions derived from the engineering judgments show excessive confounding among main effects and interactions. The five tests are augmented to remove some of these deficiencies, as shown in Table 11 (p. 73).	73

(continued on next page)

TABLE 23. (Continued)

Step Number	General Procedure	As Applied to Selected Example of Module Design	Text Page Number
15	Consider a variety of alternative designs that include more stress levels for each stress and are consistent with engineering input and statistical requirements.	For the example module, an engineering design involving three levels of temperature cycling is shown in Figure 8 (p. 81).	81
16	Make parametric trade-off studies among these designs in order to better identify the best experimental design.	Further iterations between the engineering designs and the augmented statistical designs would serve as a basis for parametric studies involving the number of stress levels for each stress, the number of modules for each test, and the degree of confounding among main effects and interactions. Two iterations are reported in this effort, as shown in Figure 5 (p. 64) and Figure 8 (p. 81).	81
17	Apply the extrapolation procedure of Hoel and Levine to determine the optimal allocation of modules among the test conditions.	The optimal allocation of modules among the test conditions are given for the two iterations of the engineering design as indicated on p. 69 and p. 81.	69.81
18	Consider a minimum of five modules at any test condition and provide for a minimum of five stress levels for each important stress, such as temperature.	Because the example module involves a combination of stable materials at rather high temperatures, it is considered unnecessary to use five levels of temperature. For RH, SO_2, and UV, only two stress levels are shown in the example because two are sufficient to illustrate the methodology. Actual accelerated test designs would be expected to involve five stress levels for each stress except where previous experience or the statistical design shows it to be unnecessary.	80
19	Select instrumentation appropriate for each test condition, taking into account the accuracy, precision, calibration requirements, maintenance, and costs.	Among these topics only precision, accuracy, and test sensitivity are treated in the text. Integrating the sensitivity requirements into the engineering is a needed refinement (see page 82), especially when the sensitivities change from one stress condition to another.	85
20	Select appropriate sequence of measurements and data acquisition procedures.	This step is not treated in this report.	
21	Use empirical, statistical, and conceptual methods to analyze the data that results from the implemented accelerated test.	These approaches are outlined in the present report. More detailed expositions appear in the references.	96
22	During the test program make predictions of the expected degradations both within and between stress conditions in order to generate a "track record" for predicting degradation at lower stress conditions using data obtained from the higher stress conditions.	A variety of procedures are illustrated using artificial data covering a two-year period with the objective of making 20-year predictions.	100

CONCLUSIONS AND RECOMMENDATIONS

As the first major part of this program, selected subjects relevant to accelerated testing in cases where prediction of service life was of significant concern were reviewed and analyzed. This review suggests that few such accelerated tests have been totally successful. The reasons for the lack of success range over a large spectrum, from inadequate experimental design to a decline in interest and support. As the second major part of this program, an improved methodology for designing appropriate tests was developed. Criteria were that (1) the tests should have an acceleration factor as large as possible, (2) the methodology should be readily applicable, and (3) the methodology should draw upon the acceptable features of previous work. As a consequence of these criteria and the review and analysis, the methodology developed puts heavy emphasis on applying engineering judgments to the problem, in addition to statistical design considerations, at the beginning of the test design. A detailed summary of specific steps involved in the test design with the improved methodology is presented in the previous section (Table 23). This methodology extends the state of the art in accelerated testing in the following areas:

(1) The factorial basis for experimental designs is used to extract and quantify engineering inputs to the design problems. This extends the initial work of Zelen.[94]

(2) The engineering input is analyzed and represented graphically in the form of a hierarchical tree. This innovation permits a "global" examination of the "local" engineering inputs regarding the severity of various combinations of stress levels.

(3) The graphical tree may be "pruned" to eliminate test conditions that are judged to be unwarranted, because of cost for example. This step permits the resulting test design, based on engineering inputs, to be fully documented with reasons for inclusion or exclusion of each test condition.

(4) The resulting pruned hierarchical tree is represented algebraically in terms of Lagrangian polynomials, providing the basis for the application of a generalized version of the optimum extrapolation procedure to normal stress levels introduced by Hoel and Levine.

(5) The Hoel and Levine extrapolation procedure is generalized to apply to a hierarchical tree structure.

(6) The algebraic model that represents the hierarchical tree provides a basis for explicit identification of possible statistical deficiencies in the experimental design that result from engineering input alone. This innovation serves as a basis for removing selected statistical deficiencies by the addition of specific tests to the experimental design.

(7) The extrapolation procedure is designed to maximize precision at a "point" which corresponds to a normal–stress condition. In the accelerated test context, this advancement indicates how data for normal–stress conditions affect the resulting experimental design (how an accelerated test for Bismarck, North Dakota would differ from an accelerated test for Miami, Florida, for example).

In general, the innovative developments presented in this report are believed to provide a basis for making parametric trade-off studies to aid in the selection of a final test design. These developments also serve to identify some important problem areas that need to be better resolved to finalize the optimization of an experimental design for accelerated testing. These are:

(1) Numerical scales that properly represent the levels of each stress in the over-stress and normal stress condition are difficult to construct. These difficulties are typified by a stress associated with temperature, for example, which is ideally scaled as $1/T$ in some contexts, but scaled simply as T in other contexts.

(2) The fact that many stresses are best represented as vectors of several components is troublesome. A temperature cycle has components associated with frequency, amplitude, and mean value, for example.

(3) The extrapolation procedure of Hoel and Levine assumes that the precision of measurement does not depend on stress level. This is not expected to be the case for accelerated testing. The extrapolation procedure should be generalized so that measurement precision can more properly be taken into account in the experimental design.

(4) Explicit constraints associated with time, costs, and precision of measurement are not, but should be, included in the design problem for accelerated testing in as direct a way as desired.

Out of the problems that still remain come two principal recommendations. The first one involves further effort toward the translation of stress components used in accelerated tests, such as temperature cycling, into the conventionally measured environmental parameters characteristic of a particular geographic (weather) site, such as average temperature, temperature extremes, etc. This translation is involved with scaling factors and extrapolation, as mentioned previously.

Clearly, how an environmental parameter scales depends upon the principal failure modes of a module, and the expected failure modes depend upon the module design. Currently, many module designs are being considered by the technical community. An accelerated test design will depend, therefore, on the module design, consistent with the maximum engineering input suggested in the methodology described in this report. To make the methodology more meaningful for a wide range of designs, it is recommended that the principal elements of a module design be considered separately with respect to possible modes of failure. The environmental parameters that affect degradation of these elements should be elucidated through whatever engineering input can be brought to bear on the problem, with emphasis on developing appropriate scaling factors. Through this procedure, differentiation among sites can be accomplished to the extent that the engineering input is technically sound. In such a manner, more meaningful accelerated tests can be designed for any module design.

Elements of the module design comprise such items as interfaces between dissimilar materials, corrosion of the cell metallization, and optical transmission of the top cover of the module. The influence — and scaling factors — of the environmental parameters depend upon the material choices for each design element. To be most useful then, the engineering input must be made for several selected designs and material choices.

The second recommendation involves the instrumentation of the test design. The methodology developed in this program delineates the influence of precision and costs associated with the

instrumentation selected for measuring property changes. Optimization of these precision/cost considerations was not attempted in the example of the test design given in this study because further development of this aspect of the design methodology is first needed, as noted previously. Moreover, more attention than was possible in this program needs to be given to the selection of the instrumentation, particularly in regard to what instrumentation is appropriate to measure property changes, in situ, associated with failure modes. In several failure modes, delamination for example, indirect measurements will be required to indicate precursors of failure. It is recommended, therefore, that a study be initiated which will (1) reveal which indirect measurements might be used to serve as precursors of failure modes, and (2) identify what instrumentation for this purpose is available now or can be made available with reasonable effort. Sensitivity, precision, cost, and applicability to in situ measurements are among the characteristics of the instrumentation that need to be known.

REFERENCES

(1) "Project Quarterly Report for the Period July 1976–September 1976", Low–Cost Silicon Solar Array Project, Jet Propulsion Laboratory Report No. ERDA/JPL-1012-77/1.

(2) Carmichael, D. C., Gaines, G. B., Sliemers, F. A., Kistler, C. W., and Igou, R. D., "Review of World Experience and Properties of Materials for Encapsulation of Photovoltaic Arrays", Battelle's Columbus Laboratories Report No. ERDA/JPL-954328-76/4 (July 31, 1976).

(3) Thomas, R. E., and Carmichael, D. C., "Terrestrial Service Environments for Selected Geographic Locations", Battelle's Columbus Laboratories Report No. ERDA/JPL-954328-76/5 (June 24, 1976).

(4) Carmichael, D. C., Gaines, G. B., Sliemers, F. A., and Brockway, M. C., "Evaluation of Encapsulation Methods and Materials for Low–Cost Long–Life Silicon Photovoltaic Arrays", Battelle's Columbus Laboratories Report No. ERDA/JPL-954328-77/2 (to be issued).

(5) Adams, P. B., "Glass Containers for Ultrapure Solutions" in Corning Research, 1973 (reprinted from *Ultrapurity*, Marcel Dekker, Inc., 1972), pp 119–177.

(6) Doremus, R. H., "Interdiffusion of Hydrogen and Alkali Ions in a Glass Surface", J. Non-Crystalline Solids, 19, 137–144 (1975).

(7) Walters, H. U., and Adams, P. B., "Effects of Humidity on the Weathering of Glass", J. Non-Crystalline Solids, 19, 183–199 (1975).

(8) Anderson, P. R., Bacon, F. R., and Byrum, B. W., "Effect of Surface Treatments on the Chemical Durability and Surface Composition of Soda Lime Glass Bottles", J. Non-Crystalline Solids, 19, 251–262 (1975).

(9) Sieger, J. S., "Chemical Characteristics of Float Glass Surfaces", J. Non-Crystalline Solids, 19, 213–220 (1975).

(10) Simpson, H. E., "Measuring the Surface Durability of Glass", Bull. Amer. Ceram. Soc., 30 (2), 41–45 (1951).

(11) Hench, L. L., "Characterization of Glass", Chapter 8 in *Characterization of Materials in Research, Ceramics and Polymers*, J. E. Burke and V. Weiss (Eds.), Syracuse University Press (1975).

(12) Tsuchihashi, S., Konishi, A., and Kawamoto, Y., "Chemical Attack of Optical Glass Surface by Formic Acid Vapor", J. Non-Crystalline Solids, 19, 221–239 (1975).

(13) Pantano, C. G., Jr., Dove, D. B., and Onada, G. Y., "Glass Surface Analysis by Auger Emission Spectroscopy", J. Non-Crystalline Solids, 19, 41–53 (1975).

(14) Berstein, V. A., and Nikitin, V. V., "Glass Surface Study by Means of Multiple ATR Spectroscopy", in *Proceedings of the Tenth International Congress on Glass*, Tokyo, Japan (July, 1974), pp 105–110.

(15) Mould, R. E., "Strength and Static Fatigue of Abraded Glass Under Controlled Ambient Conditions: IV, Effect of Surrounding Medium", J. Amer. Ceram. Soc., 44 (10), 481-491 (1961).

(16) Weiderhorn, S. M., "Environmental Stress Corrosion Cracking of Glass", National Bureau of Standards Report 10-565 (April, 1971), AD 722692.

(17) ASTM Test Designation C158-72, "Flexure Testing of Glass".

(18) Wilson, R. E., Perry, H. A., and Koubeck, F. J., "Salt Water Resistance of Surface-Compression Strengthened Glass", Naval Ordnance Laboratory Report NOLTR 71-56 (October 1971).

(19) Wiederhorn, S. M., Evans, A. G., Fuller, E. R., and Johnson, H., "Application of Fracture Mechanics to Space Shuttle Windows", J. Amer. Ceram. Soc., 57 (7), 319-323 (1974).

(20) Wiederhorn, S. M., "Crack Growth as an Interpretation of Static Fatigue", J. Non-Crystalline Solids, 19, 169-181 (1975).

(21) Weyl, W. A., Colored Glasses, Soc. Glass Technology (1951), pp 497-514.

(22) Garg, H. P., "Effect of Dirt on Transparent Covers in Flat-Plate Solar Energy Collectors", Solar Energy, 15, 299-302 (1974).

(23) Newton, R. G., "Problems Arising from the Weathering of Poorly Durable Glass", in Proceedings of the Tenth International Congress on Glass, Tokyo, Japan (July, 1974), pp 49-54.

(24) Ministry of Tech., London (Eng.), Materials Div., "The Effect of Storage on the Strength of Annealed and Toughened Clear White Plate Glass", Report No. D., Mat. 153 (September, 1968), 35 pp.

(25) Corney, N. S., Ministry of Defense, London (Eng), Directorate of Research Mat., "The Evaluation of VHR 2A and VHR 3A Glasses (Glaverbel-Mecaniver S.A.) for Aerospace Applications", Report No. DR-Mat-198 (March, 1975), 33 pp.

(26) MIL-HDBK-17, "Plastics for Flight Vehicle, Part II. Transparent Glazing Materials".

(27) ASTM Test Designation C225-73, "Resistance of Glass Containers to Chemical Attack".

(28) Mendel, J. E., and Ross, W. A., "Chemical Durability of Glasses Containing Radioactive Fission Product Waste", BNWL-SA-4854, Paper presented at Annual Meeting of American Ceramic Society (April, 1974).

(29) Simmons, J. H., Mills, S. A., and Howell, B. F., "Effect of Phase Separation on the Physical and Chemical Properties of Glasses − Density and Chemical Durability", Nat'l Bur. of Standards, Washington, D. C., Inst. for Mat'l Res., Final Tech. Report No. NBSIR-74-510 (July 16, 1974).

(30) Morey, G. W., The Properties of Glass, Second Edition, Reinhold (1954).

(31) Hagy, H. E., "Physical Properties of Glass", Section 17 in *The Handbook of Glass Manufacture*, Volume II, F. V. Tooley (Ed.), Books for Industry Incorporated, and The Glass Industry Magazine (1974).

(32) ASTM Test Designation C 724-72T, "Acid Resistance of Ceramic Decorations on Architectural Type Glass".

(33) Fitch, W. T., Motorola, 13th Annual Reliability Phys. Sym. (April, 1975), pp 58-69.

(34) Khajezadeh, H., and Rose, A. S., RCA, 13th Annual Reliability Phys. Sym. (April, 1975), pp 87-92.

(35) Brodeur, J. E., Fairchild Semiconductor, 13th Annual Reliability Phys. Sym. (April, 1975), pp 255-256.

(36) Lycoudes, N., Motorola, 13th Annual Reliability Phys. Sym. (April, 1975), pp 257-259.

(37) Perri, J. A., "Glass Encapsulation", *Microelectronic Technology*, S. L. Marshall (Ed.), Boston Tech. Publishers, Inc. (1967), pp 215-217.

(38) Solvason, K. R., and Wilson, A. G., "The Development of Evaluation Procedures for Factory-Sealed Double Glazing in Canada", Proceedings of a Seminar on the Durability of Insulating Glass, National Bureau of Standards, Bldg. Sci. Series Report, BSS 20 (February, 1970), pp 11-16.

(39) Gwyn, J. D., "Manufacturers Test Methods: Correlation with Field Experience; Expected Field Life", Proceedings of a Seminar on the Durability of Insulating Glass, National Bureau of Standards, Bldg. Sci. Series Report, BSS 20 (February, 1970), pp 35-36.

(40) Mayzoni, J. R., and Bowser, G. H., "Test Methods and Field Experience with Double-Glazed Units", Proceedings of a Seminar on the Durability of Insulating Glass, National Bureau of Standards, Bldg. Sci. Series Report, BSS 20 (February, 1970), pp 37-42.

(41) Robinson, H. E., "An Exploratory Study of Laboratory Test Methods and Standards for Factory-Sealed Double-Glazed Window Units", Proceedings of a Seminar on the Durability of Insulating Glass, National Bureau of Standards, Bldg. Sci. Series Report, BSS 20 (February, 1970), pp 49-58.

(42) Jemian, W. A., Wilcox, R. C., and Hsu, A.D.T., "Stability of Glass Fiber Reinforced Composites", Auburn University Final Report, Contract DAAH01-73-C-0466 (November 25, 1974) (AD/A-003499/1st).

(43) Lock, L. M., Frank, G. C., and Stevens, J. R., "A Study of Some Factors Affecting the Photodegradation of Textile Yarns. Part I: The Spectral Distributions of Sun-Sky and Xenon-Arc Radiation", Textile Res. Jour., 43, 483-488 (August, 1973).

(44) Titus, J. B., "The Weatherability of Polystyrenes and Related Copolymers and Terpolymers", Plastic Report 38 (July, 1969), 56 pp.

(45) Warnick, W. V., "Predicting the Exterior Durability of Polymer Coatings for Metal Substrates", ACS Organic Coatings and Plastic Chemistry Division Preprints, 33 (1), 1-11 (April, 1973).

(46) Kamal, M. R., and Saxon, R., "Recent Developments in the Analysis and Prediction of the Weatherability of Plastics", Appl. Poly. Symposia, 4, 1–28 (1967).

(47) Hirt, R. C., and Searle, N. Z., "Energy Characteristics of Outdoor and Indoor Exposure Sources and Their Relation to the Weatherability of Plastics", Appl. Poly. Symposia, 4, 61–83 (1967).

(48) Baum, B., "The Weathering Degradation of Polyolefins", Poly. Engr. and Sci., 14 (3), 106–211 (1974).

(49) Blais, P., Carlsson, D. J., and Wiles, D. M., "Surface Changes During Polypropylene Photo-Oxidation: A Study by Infrared Spectroscopy and Electron Microscopy", Jour. Poly. Sci.: Part A-1, 10, 1077–1092 (1972).

(50) Szabo, E., and Lally, R. E., "World-Wide Weathering of Poly (Vinyl Chloride)", Poly. Engr. and Sci., 15 (4), 277–280 (April, 1975).

(51) Garner, B. L., and Papillo, P. J., "Accelerated Outdoor Exposure Testing in Evaluation of Ultraviolet Light Stabilizers for Plastics", Ind. Eng. Chem., Prod. Res. and Dev., 1 (4), 249–257 (December, 1962).

(52) Pinkerton, D. M., "Mechanisms of the Weathering of Polymers", Proc. Royal Australian Chem. Inst. (January, 1972), pp 33–39.

(53) Greenberg, S. A., McCargo, M., and Palmer, W. L., "Investigation of FEP TEFLON as a Cover for Silicon Solar Cells", NASA CR-72970 (August, 1971), 69 pp.

(54) Jellinek, H.H.G., "Chain Scission of Polymers by Small Concentrations (1 to 5 ppm) of Sulfur Dioxide and Nitrogen Dioxide, Respectively, in Presence of Air and Near Ultra-violet Radiation", Jour. Air Poll. Cont. Assoc., 20 (10), 672–674 (October, 1970).

(55) Newland, G. C., and Tamblyn, J. W., "Actinometry of Sunlight at Kingsport, Tennessee", Appl. Poly. Symposia, 4, 119–129 (1967).

(56) Martinovich, R. J., and Hill, R. J., "Practical Approach to the Study of Polyolefin Weatherability", Appl. Poly. Symposia, 4, 141–154 (1967).

(57) Quakenbos, H. M., and Samuels, H., "Practical Problems in Predicting Weathering Performance", Appl. Poly. Symposia, 4, 155–165 (1967).

(58) Portig, W. H., Bryan, J. C., and Dobbins, D. A., "Determination of Optimum Tropic Storage and Exposure Sites. Phase II: Patterns and Predictions of Tropic Materials Deterioration", USATTC Rept. No. 7405001 (1974).

(59) Schafer, V., "Accelerated Light Exposure in the Xenotest: Results, Limits, and Comparison With Conventional Equipment", Appl. Poly. Symposia, 4, 111–118 (1967).

(60) Clark, J. E., and Harrison, C. W., "Accelerated Weathering of Polymers, Radiation", Appl. Poly. Symposia, 4, 97–110 (1967).

(61) Jordan, J. M., McIllroy, R. E., and Pearce, E. M., "Accelerated Aging of Poly(Methyl Methacrylate) Copolymers and Homopolymers", Appl. Poly. Symposia, 4, 205–218 (1967).

(62) Lloyd, P.F.V., "Stress Rupture Testing of Thermoplastic Pipe", Plastics (British), 39–42 (August, 1964).

(63) Kuhlman, H. W., and Wolter, Fritz, "Investigation of Short–Term Ring Specimen Stress Rupture Testing", presented to the USASI B31.8 Committee on Gas Distribution Piping Systems, November 8, 1967, Las Vegas, Nevada.

(64) Gray, V. E., and Cadoff, B. C., "Survey of Techniques for Evaluating Effects of Weathering on Plastics", Appl. Poly. Symposia, 4, 85–95 (1967).

(65) Isakson, K. E., "Use of Infrared Specular Reflectance in Study of Ultraviolet Degradation of Polymer Films", Jour. Paint Tech., 44 (573), 41–62 (October, 1972).

(66) Mullen, P. A., and Searle, N. Z., "The Ultraviolet Activation Spectrum of Polycarbonate", Jour. Appl. Poly. Sci., 14, 765–776 (1970).

(67) Hama, Y., Hosono, K., Furui, Y., and Shinohara, K., "ESR Study of Free Radicals Produced in Polyethylene Irradiated by Ultraviolet Light", Jour. Poly. Sci.: Part A-1, 9, 1411–1419 (1971).

(68) Nathan, R. A., Mendenhall, G. D., Hassel, J. A., and Wallace, J. D., "Measurement of Ultraweak Chemiluminescence as an Analytical Tool", Industrial Research, 17, 62–67 (December, 1975).

(69) George, G. A., "The Phosphorescence Spectrum and Photodegradation of Polystyrene Films", Jour. Appl. Poly. Sci., 18, 419–426 (1974).

(70) Kamal, M. R., "Cause and Effect in the Weathering Plastics", Poly. Engr. and Sci., 10 (2), 108–121 (1970).

(71) Grinsfelder, H., "Analysis of Plastic Weathering Results", Appl. Poly. Symposia, 4, 245–262 (1967).

(72) National Bureau of Standards and Manufacturing Chemists Association, "Outdoor Performance of Plastics I through X", 1968–1972.

(73) Clark, J. E., "Correlation of Accelerated and Outdoor Weathering Tests of Plastics", NBS Report (1967).

(74) Leikina, M. E., and Tatevos'yan, G. O., "Planning an Experiment for Studying the Light Aging of Polyethylene Film", Soviet Plastics, No. 7, 73 (1970).

(75) Box, G.E.P., and Wilson, K. B., Journal of Royal Statistical Society (B), 13, 1-55 (1951).

(76) Leikina, M. E., Tatevos'yan, G. O., Kuznetsova, I. B., and Melkumov, A. N., "Use of Mathematical Experiment Planning Method to Investigate the Influence of Artificial Weather Factors on Certain Properties of Polymer Films", Soviet Plastics, No. 2, 41 (1971).

(77) Kamal, M. R., "Effect of Variables in Artificial Weathering on the Degradation of Selected Plastics", Polymer Engr. and Sci., 6, 333 (1966).

(78) Thomas, D. K., and Sinnott, R., "Prediction of Shelf Storage Life from Accelerated Heat Aging Tests", Journal of IRI, p 163 (1969).

(79) Myers, G. E., "Chemical Structural Aging Effects", MANCP Report No. 297 (1974).

(80) Dalaba, E. M., "Minuteman Stage III Component Report", AFRPL-TR-74-58 (1974).

(81) Mann, N. R., Schafer, R. E., and Singpurwalla, N. D., *Methods for Statistical Analysis of Reliability and Life Data*, John Wiley & Sons, Inc., New York (1974).

(82) Levenbach, G. J., "Accelerated Life Testing of Capacitors", IRE Transactions, PGROC (10), 9-20 (1957).

(83) Thomas, R. E., "When is a Life Test Truly Accelerated?", Electronic Design, pp 64-70 (1964).

(84) Thomas, R. E., and Gorton, H. C., "Research Towards a Physics of Aging in Electronics", Proceedings of the Second Annual Symposium on the Physics of Failure in Electronics.

(85) Allen, W. R., "Inferences from Tests with Continuously Increasing Stress", Journal of the Operation Research Society of America, pp 302-312, (1959).

(86) Zelen, M., "Factorial Experiment in Life Testing", Technometrics, 1 (3), 269-288 (1959).

(87) Hoel, P. G., and Levine, A., "Optical Spacing and Weighting in Polynomial Prediction", Am. Math. Stat., 35, 1553-1560 (1964).

(88) Mann, N. R., "Design of Over-Stress Life-Test Experiments When Failure Times Have the Two-Parameter Weibull Distribution", Technometrics, 14, 437-451 (1972).

(89) Little, R. E., and Jebe, E. H., "A Note on the Gain in Precision for Optimal Allocation in Regression as Applied to Extrapolations in S-N Fatigue Testing", Technometrics, 11, 389-391 (1969).

(90) Blight, B. J. N., "On the Most Economical Choice of a Life Testing Procedure for Exponentially Distributed Data", Technometrics, 14, 613-618 (1972).

(91) Izenman, A. J., and Rinott, Y., "Sequential Designs for Replacements in Exponential Life Test" (submitted for publication).

(92) McCallum, J., Thomas, R. E., and Waite, J. H., "Accelerated Testing of Space Batteries", NASA SP-323, Scientific and Technical Information Office, National Aeronautics and Space Administration, Washington, D.C., 212 pp, (1973).

(93) Singpurwalla, N. D., "Inference from Accelerated Life Tests Using Arrhenius Type Re-Parameterizations", Technometrics, 15, 289-299 (1973)

(94) Carmichael, D. C., Gaines, G. B., Sliemers, F. A., and Kistler, C. W., "Materials for Encapsulation Systems for Terrestrial Photovoltaic Arrays", 12th IEEE Photovoltaic Specialists Conference, November, 1976.

(95) Davies, O. L., *The Design and Analysis of Industrial Experiments*, Oliver and Boyd, London, Hafner Pub. Co., New York (1960).

(96) Addleman, S., "Recent Developments in the Design of Factorial Experiments", Journal of the American Statistical Association, 67, 103–111 (1972).

(97) Box, G. E. P., and Hunter, J. S., "The 2^{k-p} Fractional Factorial Designs", Technometrics, 3, Part I, p 311–351; Part II, p 449–458.

(98) Webb, S. R., "Design, Testing and Estimation in Complex Experimentation; Expansible and Contractible Factorial Designs and the Application of Linear Programming to Combinatorial Problems", Aerospace Research Laboratories, Report No. 65–116, Part 1 (June, 1965).

(99) Margolin, B. H., "Results on Factorial Designs of Resolution IV for the 2^n and $2^n 3^n$ Series", Technometrics, 11, 431–444 (1969).

(100) Rechtschaffner, R. L., "Saturated Fractions of 2^n and 3^n Factorial Designs", Technometrics, 9, 569–576 (1967).

(101) Kennard, R. W., and Stone, L. A., "Computer-Aided Design of Experiments", Technometrics, 11, 137–148 (1969).

(102) Stigler, S. M., "Optimal Experimental Design for Polynomial Regression", Journal of the American Statistical Association, 66, 311–318 (1971).

(103) Kiefer, J., "Optimum Experimental Designs", Journal of the Royal Statistical Society (B), 21, 272–307 (1959).

(104) Nalimov, V. V., Golikova, T. I., and Mikeshina, N. G., "On Practical Use of the Concept of D-Optionality", Technometrics, 12, 799–812 (1970).

(105) Cramer, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton, New Jersey (1946).

(106) Mandel, John, *The Statistical Analysis of Experimental Data*, Interscience, Chapter 14 (1964).

(107) Mandel, J., and Stiehler, R. D., "Precision Measurement and Calibration — Statistical Concepts and Procedures", NBS Special Publication 300, Vol. 1, p 179.

(108) Dixon, N. J., and Massey, F. J., *Introduction to Statistical Analysis*, Third Edition, McGraw-Hill, p 308 (1967).

(109) Draper, N. R., and Smith, H., *Applied Regression Analysis*, John Wiley and Sons, Inc., New York, New York, p 26 (1966).

(110) Dalaba, E. M., "Minuteman Stage III Component Report", AFRPL-TR-58 (1974).

(111) Myers, G. E., "Chemical Structural Aging Effects", MANCP Report No. 297 (1974).

(112) Clark, J. E., "Correlation of Accelerated and Outdoor Weathering Tests of Plastics", NBS Report (1967).

(113) Box, G.E.P., and Cox, D. R., "An Analysis of Tranformations", Journal of the Royal Statistical Society, Series B, 26 (2), 211 (1964).

(114) Derringer, G. C., "An Empirical Model for Viscosity of Filled and Plasticized Elastome Compounds", Journal of Applied Polymer Science, 18, 1083 (1974).

(115) Draper, N. R., and Smith, H., "Applied Regression Analysis", John Wiley and Sons, Inc., New York (1966).

(116) Box, G.E.P., and Wetz, J. M., "Criteria for Judging Adequacy of Estimation by an Approximating Response Function", Technical Report No. 9, Department of Statistics, University of Wisconsin, Madison, Wisconsin (1973).

(117) Derringer, G. C., "A Monte Carlo Study of Regression Model Adequacy", Master of Science Thesis at University of Akron, Akron, Ohio (1975).

(118) Suich, R., and Derringer, G. C., "Is the Regression Equation Adequate? – One Criterion", Technometrics (May, 1977).

(119) Welch, G. L., "Using Accelerated Hydrolytic Reversion Data to Predict Service Life of Elastomeric Potting Compounds", Technical Report AFML-TR-70-297 (February, 1971).

(120) Shannon, C. E., "A Mathematical Theory of Communication", Bell System Technical Journal, 27, 379-423 and 623-656 (1948).

(121) Box, G.E.P., and Hill, W. J., "Discriminations Among Mechanistic Models", Technometrics, (9) 1, 57 (1967).

(122) Hill, W. J., Hunter, W. G., and Wickern, D. W., "A Joint Design Criterion for the Dual Problem of Model Discrimination and Parameter Estimation", Technometrics, (10) 1, 145 (1968).