13

13 Reliability Estimation Techniques

Reliability is defined as "the ability of a product to function properly within specified performance limits for a specified period of time, under the life-cycle application conditions." Reliability estimation techniques include methods to evaluate system reliability throughout the product life cycle. The major components of reliability estimation techniques are the test program and the analysis of data from the tests. Test programs are developed throughout the life cycle, that is, design, development, production, and service, to ensure that reliability goals are met at different stages in the product life cycle. Data from test programs during each stage are acquired and processed to evaluate the reliability of a product at each stage in the life cycle.

The purpose of reliability demonstration testing is to determine whether the product has met a certain reliability requirement with a stated confidence level prior to shipping. Tests should be designed to obtain maximum information from the minimum number of tests in the shortest time. To achieve this, various statistical techniques are employed. A major problem in the design of adequate tests is simulating the real-world environment. During its lifetime, a product is subjected to many environmental factors, such as temperature, vibration, shock, and rough handling. These stresses may be encountered singly, simultaneously, or sequentially, and there are many other random factors.

13.1 Tests during the Product Life Cycle

Various tests are carried out at different stages in a product's life cycle to ensure that the product is reliable and robust. The different stages in a product's life cycle and tests suggested to be carried out in each stage are listed in Table 13.1.

13.1.1 Concept Design and Prototype

The purpose of the tests in the concept design and prototype stage is to verify breadboard design and functionality. Tests are conducted to determine the need for parts,

Reliability Engineering, First Edition. Kailash C. Kapur and Michael Pecht.

Table 13.1 Stages in a product's life cycle and suitable tests in each stage

Phase description	Suitable tests
Concept design and prototype	Engineering verification test
Performance validation to design specification	Design verification test
Design maturity validation	Highly accelerated life test, accelerated stress test, stress plus life test (STRIFE), accelerated life test
Design and manufacturing process validation	Design maturity test, firmware maturity test, process maturity test
Preproduction low volume manufacturing	Burn-in, reliability demonstration test
High volume production	Highly accelerated stress screen, environmental stress screen, reliability demonstration test, ongoing reliability test, accelerated life test
Feedback from field data	Fleet/field monitoring/surveillance

materials, and component evaluation or qualification to meet system performance and other reliability design criteria.

13.1.2 Performance Validation to Design Specification

In this stage, tests are carried out to verify the functional adequacy of the design and the product performance. Tests are used to corroborate preliminary predictions. Failure modes and effects analysis is carried out to disclose high-risk areas and reliability problems in the proposed design.

13.1.3 Design Maturity Validation

Tests are carried to evaluate a design under environmental conditions, to verify the compatibility of subsystem interfaces, and to review the design. The design margins and robustness of the product are tested and quantified at this stage. The results from these tests will assist in developing a better design with minimal reliability issues.

13.1.4 Design and Manufacturing Process Validation

Design acceptance tests are used to demonstrate that the design meets the required levels of reliability. A reliability demonstration test is considered mandatory for design acceptance. Tests conducted at this stage include the design maturity test (DMT), firmware maturity test (FMT), and process maturity test (PMT). The purpose of DMT is to show that the product design is mature and frozen and is ready for production. The integration of hardware and software components is tested in the FMT. The process control is demonstrated in PMTs.

13.1.5 Preproduction Low Volume Manufacturing

Product screening—the process of separating products with defects from those without defects—is carried out at the preproduction stage to reduce infancy defects due to process, manufacturing and workmanship. Screening tests in the preproduction stage assist in reducing the infant mortality of a product. Burn-in is one of the mostly commonly used product screening tests. A reliability demonstration test may also be carried out to demonstrate the reliability of low volume manufacturing.

13.1.6 High Volume Production

Tests are carried out to determine the acceptability of individual products in order to ensure production control and critical interfaces, parts, and material quality. Screening at this stage does not improve the overall reliability of the product or change the yield; rather, it allows companies to ship products without defects to customers. Commonly conducted screens include the highly accelerated stress screen and environmental stress screen. To demonstrate the long-term reliability at high volume production, reliability demonstration tests are also carried out.

13.1.7 Feedback from Field Data

Tests and evaluation programs are carried out during field use of the product for continued assessment of reliability and quality. The data from field use are utilized to improve the design and reliability of the next version or generation of a product.

13.2 Reliability Estimation

Product reliability can be estimated from the test data using parametric or nonparametric techniques. In parametric estimation, the distribution of the test data should be known or assumed. Parametric techniques provide an inaccurate estimation if the assumptions are incorrect. The parameters are the constants that describe the distribution. Nonparametric estimates do not assume that the data belong to a given probability distribution. Generally, nonparametric estimates make fewer assumptions than parametric estimates. Nonparametric estimates apply only to a specific test interval and cannot be extrapolated. In this chapter, parametric estimates with underlying binomial, exponential, and Weibull distributions are described. The frequently used parametric estimates include: (1) point estimate: a single valued estimate of a parameter/reliability measure; (2) interval estimate: an estimate of an interval that is believed to contain the true value of the parameter; and (3) distribution estimate: an estimate of the parameters of a reliability distribution.

Data are often collected from a sample that is representative of the population to estimate the parameters of the entire population. For instance, the time to failure of light bulbs manufactured in a lot may be assessed to estimate the longevity of all the light bulbs manufactured by the company. Another example is the periodic sampling of manufactured goods to estimate the defect rate of the total population. Sample acceptance testing can also be conducted at the receipt of goods in order to assess and estimate the ability of the entire lot to meet specifications. The confidence interval is a measure of the uncertainty associated with making a generalization about the population based on a sample. These concepts are given later on in this chapter.

13.3 Product Qualification and Testing

A successful accelerated stress test (AST) program meets customer requirements, lowers life-cycle costs, and reduces the time to market for a product. A physics-of-failure (PoF)-based qualification methodology has been developed. The fundamental concepts of the PoF approach and a set of guidelines to design, plan, conduct, and implement a successful accelerated stress test are discussed below.

The inputs to the qualification methodology include hardware configuration and life-cycle environment. The output from the methodology is a *durability assessment*, where the accelerated stress test results are correlated to field life estimates through quantitative damage metrics and acceleration transforms. The PoF-based qualification methodology is a five-step approach, as shown in Figure 13.1. Step 1, *virtual qualification*, identifies the potential failures under the life-cycle loading inputs and their respective failure sites, modes, and mechanisms. Virtual qualification is a PoF-based process to assess the life expectancy under anticipated life-cycle loading conditions. The tasks in step 2, *accelerated test planning and development*, are to design test specimens, set up the experiment, determine accelerating loads, collect data using sensors, monitor data responses, and devise data postprocessing schemes. In step 3, *specimen characterization*, the test specimens' responses to test loads are determined, and the overstress limits of the test specimen are identified to precipitate the failure mechanisms of interest. *Accelerated life testing* (ALT) in step 4 evaluates the intrinsic product vulnerability to applied loads due to wear-out mechanisms. *Virtual testing* is conducted in step 5 on a limited sample size using test load levels scaled back from the destruct level profiles. The steps in the qualification methodology are explained in detail in the following sections.

13.3.1 Input to PoF Qualification Methodology

The inputs to the PoF qualification methodology may be broadly classified as hardware configuration and life-cycle loads. The following sections provide a description of the input parameters.

13.3.1.1 Hardware Configuration The product architecture and material properties must be identified and documented. First, the architecture of the product and the primary performance features of the product are studied. A database of a product's configuration (including physical dimensions, functionality, and constitutive elements) and layout (including electrical traces, routing, and connectivity) will assist in the development of effective analysis and verification procedures. A PoF-based approach cannot be applied to a black box configuration. On the contrary, specific information and detailed understanding of the hardware is required. A comprehensive documentation of the parts list, part dimensions, and primary performance features (electrical, mechanical, etc.) will be useful for PoF analysis.

Second, PoF methodology accentuates the understanding of material behavior and therefore requires that all the intrinsic material properties be documented. Material models must be characterized over the loading conditions experienced by the product. For instance, the fracture toughness of a material must be characterized over a range of temperatures or loading rates, depending on the loading conditions. The fracture toughness of a material to be used in a thermal management system must be

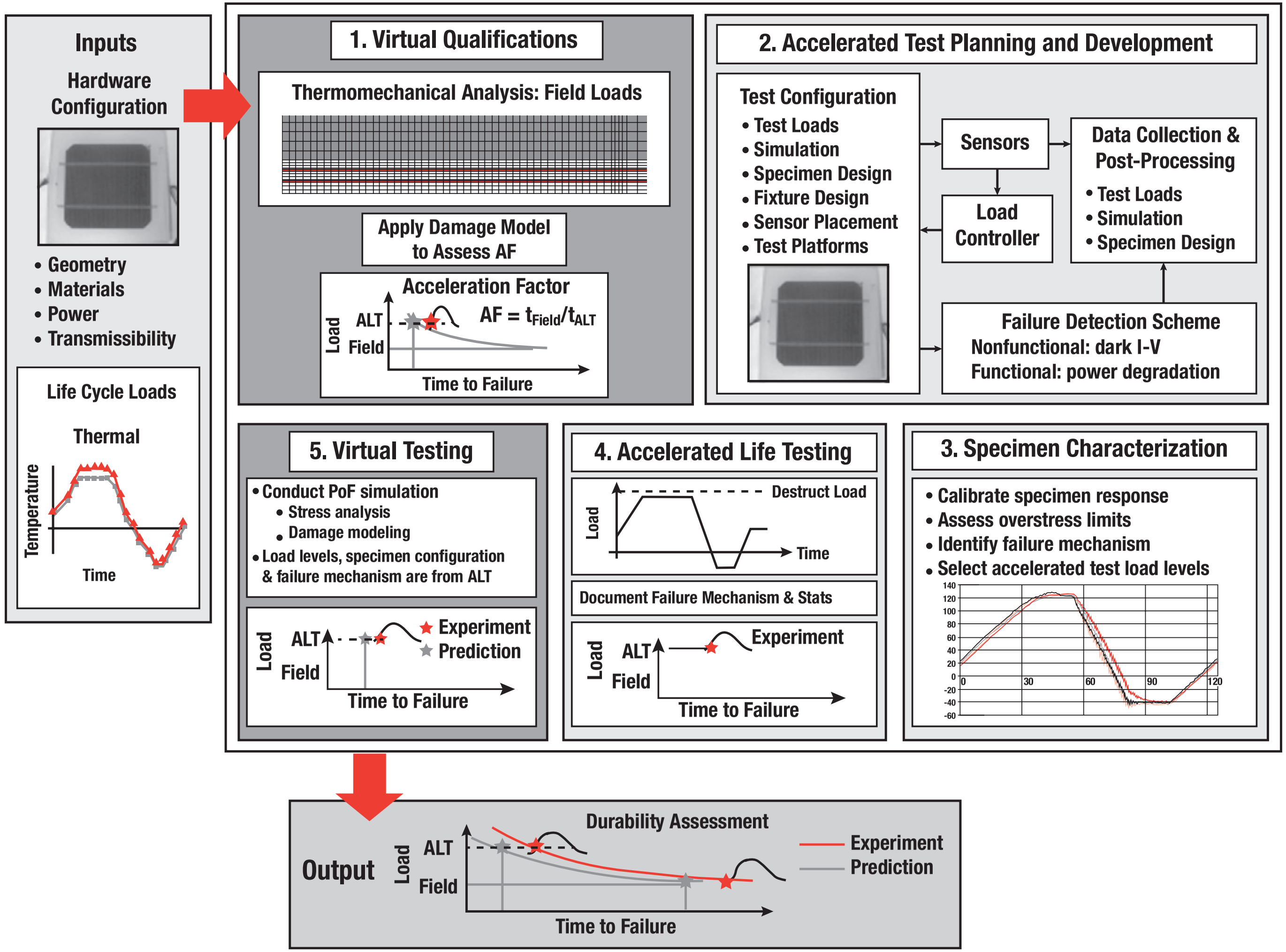

Figure 13.1 Example flowchart of PoF methodology.

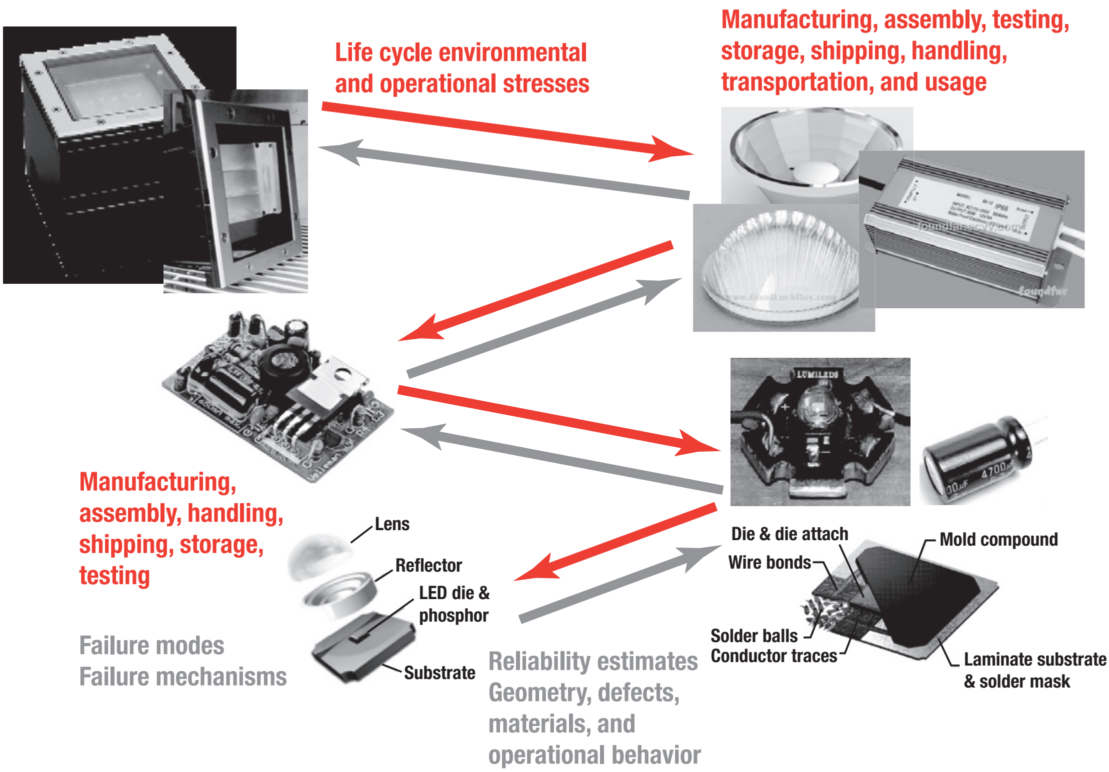

Figure 13.2 PoF requires information about and understanding of the entire supply chain.

characterized over a range of temperatures. On the other hand, if the same material were to be used in a moving system, the fracture toughness of the material must be characterized over a range of loading rates. The properties of common materials are obtained from material handbooks or literature. If the material properties are not readily available, then coupons of the materials are constructed to measure the appropriate material property. Additionally, the stress limits of the materials used in the hardware will assist in determining the design limits of the product. The PoF methodology is a detail-oriented process that requires detailed inputs for the materials used and how they are put together. All the materials and material properties that go into each and every single part of a product down to every IC must be available. An illustration of the breakdown of materials used in an LED is shown in Figure 13.2. Understanding the degradation mode and root-cause failure mechanisms that will be triggered under life-cycle stresses enables engineers to select appropriate materials to be used in the hardware. Product life can be calculated using analytical and PoF models and, based on the estimated product life, appropriate materials can be selected.

An LED device is an optoelectronic device that consists of various electronics and optical materials packaged in a mold material. Although the LED manufacturer builds the final product, the materials are provided by various suppliers in the supply chain. Hence, the manufacturer has to rely on the suppliers for information on failure modes and mechanisms. The manufacturers rely on engineers at each level in the supply chain to conduct physics-of-failure modeling. Each engineer in that supply chain contributes their own piece to the puzzle. One of the major benefits of the PoF approach is that the material properties can be cataloged long before a particular product enters the design cycle and development cycle. Therefore, designers and engineers do not have to wait to build a product and test it to determine the reliability model constants. In other words, before building any product, engineers can estimate the expected life using existing PoF models. However, the results from PoF models have an associated uncertainty due to the uncertainty in the inputs. This associated uncertainty can be orders of magnitude different from the actual life. If a company estimates that a product will last 5 years, it could really be half of that, or it could be year 10. Unfortunately, this is all within the band of uncertainty, and it is very difficult to design for finite life in this region. As a result, companies try to overdesign in this region. For example, if a company has to design for vibration for 5 years, that company would design the product to last 25 years to eliminate the possibility of failures within 5 years.

Electronics manufacturers must take into account the variability introduced into the material properties by the manufacturing and assembly processes. For instance, to measure the properties of the copper used in electronics, fatigue tests are carried out on copper coupons in a laboratory. These material properties are tested without building any electronics. Unfortunately, the building of electronics affects those properties depending on the assembly process of the copper in the printed circuit card. The damage accumulated due to manufacturing processes and manufacturing defects can cause variability in these model constants.

Appropriate PoF models should be selected to estimate the damage accumulated during use and predict the anticipated life under applied loading conditions. The selection of PoF models is based on the type of material used and the loading conditions. For instance, since solder is a viscoplastic material solder, it has to be modeled with a viscoplastic model, which is a combination of elastic deformation, plastic deformation, and time-dependent or rate-dependent creep deformation. Figure 13.3 shows the evolution of solder microstructure under fatigue loading until crack formation. The amount of damage accumulation and the corresponding microstructure evolution are functions of the temperature cycling range and rate.

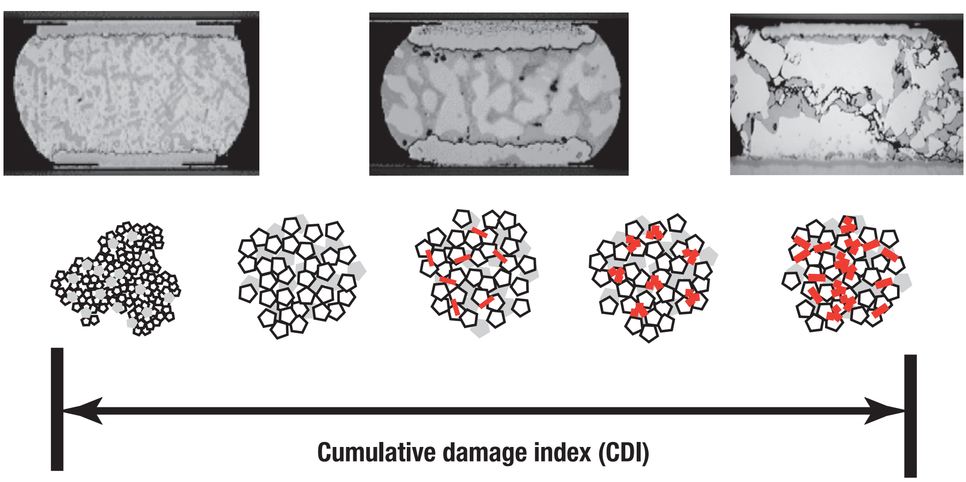

Figure 13.3 Physics-of-failure assessment: failure mechanism identification.

In order to qualify different materials in a product, the engineer has to go through the entire physics of failure process. The engineer needs the dimensions and material properties of the chip, solder, board, underfill, and so on. Finite element simulation is carried out at different loading conditions to find the stress distributions in the critical solder joints.

Life-Cycle Environment The second part of the input is to identify the operational use environments and document field failures history.

Operational Use Environment Environmental environment defines the loading conditions experienced by the product during use. There is a distinction between a load on a system and a stress. Loads are the boundary conditions on a system and refer to what the environment is doing to the system. The temperature, electrical usage, voltage, current, temperature, humidity, and mechanical vibration that a product is subjected to are all examples of loads. Information on loads does not necessarily provide the failure mechanism at a failure site. For that, understanding of the stresses causing the failure at the failure site is required. Stress is the intensity of an applied load at a failure site. Intensity will be defined differently depending on which stress type and which failure mechanism is in view. In order to monitor degradation, many parameters must be measured. One must know the usage pattern, because if the usage pattern changes over the life of the product, then the degradation rate also changes. Thus, the environmental conditions must be monitored, including humidity, temperature, mechanical acceleration, cyclical change of temperature, and rate of change of temperature. To relate these functional parameters to the environment, pattern recognition and correlation estimation are carried out. It is advantageous to know which environmental conditions have the highest impact on degradation. In addition to the applied loading conditions, diurnal environmental changes can also cause stresses. For instance, temperature changes from day to night, or from indoors to outdoors, result in stresses in portable electronics.

A product experiences loads even before field use. These loads include stresses during manufacturing, testing, transportation, handling, and storage. Stresses prior to field use can result in product failure before reaching the customer, known as “dead on arrival.” The rate of change and duration of exposure of the loads prior to field use are important determinants of the stress magnitudes induced in a product. In addition to the operational loading conditions, a designer should adequately understand the storage conditions of the product. The shelf life, storage and transportation environments, and rework criteria should be understood clearly to adequately design the product for long-term storage. A well-designed, rugged product should survive the loads applied during operation, handling, and storage. Any accelerated life durability tests should take into account all the environmental conditions, rework environments, and workmanship factors that a product is expected to encounter.

Field Failures History Understanding the history of prior field failures and preliminary failure analysis results is useful for identifying the dominant failure sites, modes, and mechanisms. Every time a product is changed in favor of a new technology, the company must utilize relevant information from the previous product’s history to identify potential failure mechanisms. From a PoF perspective, if a company effectively utilizes the previous product’s failure history, the more likely it is that the new product will be reliable. The knowledge and model constants of the previous product

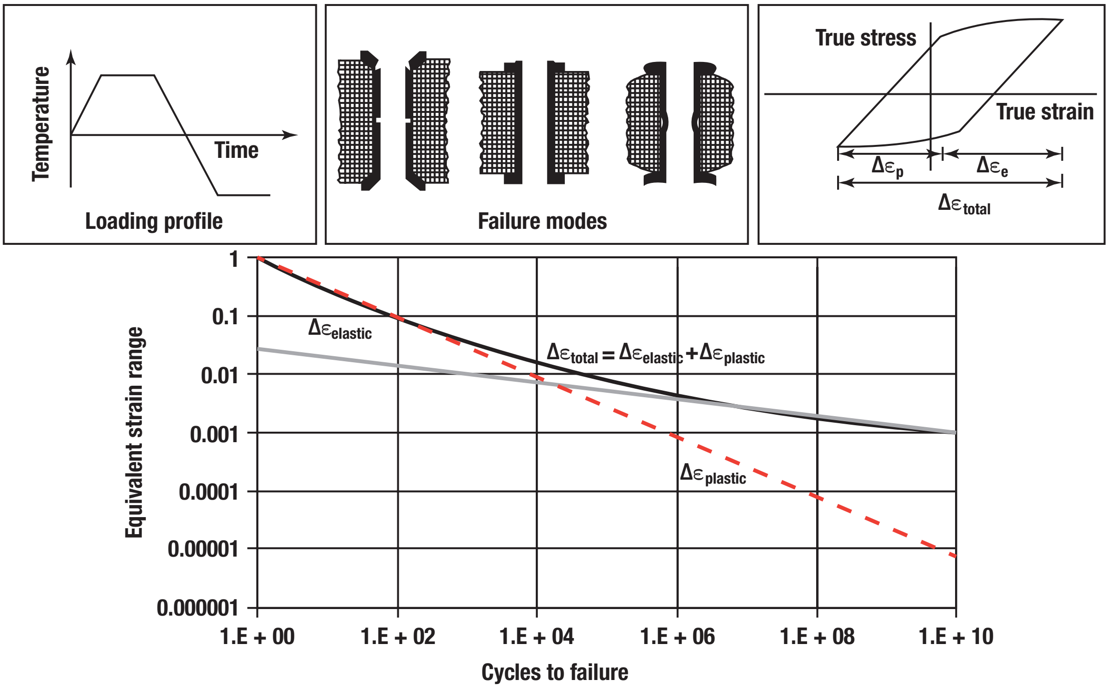

Figure 13.4 Plated through hole (PTH) low cycle fatigue in printed wiring boards (PWBs).

can be extrapolated for the new product. The field life is related to the test results by acceleration factors. Based on the observed modes in the field failure history, the test program can be appropriately tailored to identify and precipitate the weak links. For example, if the field failure history reports interconnect failures from creep-fatigue interactions as the predominant failure mechanism, temperature and vibration loads may be applied during testing to precipitate fatigue failures at the interconnects.

Fatigue-induced fracture due to cyclic loading is one of the most common failure mechanisms observed in electronic products. Figure 13.4 is a thermal cycling loading profile applied to a printed circuit assembly and the corresponding stress-strain history. The hysteresis loop for cyclic loading consists of elastic loading, which causes low cycle fatigue, and plastic loading, which causes high cycle fatigue. In the strain range versus cycles to failure plot, the steeper slope is the low cycle fatigue region, whereas the shallower slope represents the high cycle fatigue region. The actual fatigue data follow the superposition of these two regions in a log-log scale, represented by the solid black curve. If thermal cycling were conducted at different strain levels, the fatigue data would fall along the black curve with some scatter around it.

13.3.2 Accelerated Stress Test Planning and Development

The design of test loads, choice of test vehicle, and identification of issues related to test setup must be carried out prior to the accelerated testing of a product. This constitutes step 2 of the PoF approach to accelerated product qualification.

13.3.2.1 Test Loads and Test Matrix The first step in accelerated stress test planning is to determine the accelerated stress test matrix based on the dominant environmental loads and failure mechanisms. For example, if creep and mechanical fatigue are identified as the dominant failure mechanisms, the potential test loads are thermal and vibrational loads, respectively. Designing the test matrix depends on the dominant test loads and program objectives. For example, in a case study where the primary interest is to explore the interaction effects between thermal and vibrational loads, the test matrix can consist of several load cases involving simultaneous and sequential applications of repetitive shock vibration (RSV) and temperature cycling (TC). To conduct physics-of-failure modeling, quantitative information about the design of the product and its hardware configuration is imperative. The geometries, materials, and overall life-cycle loading of the product are also required. Overall life-cycle loading refers to the combination of stresses experienced by the product over its entire life cycle, including diurnal environmental loading.

13.3.2.2 Test Vehicle An appropriate test specimen should be designed to reduce the test time. For example, if the focus of a test is to understand surface-mount interconnect failures, a nonfunctional circuit card assembly is an appropriate test vehicle instead of testing an entire functional electronic assembly. A nonfunctional test specimen enables acceleration of the stresses beyond the operating limit of the product and is limited by the intrinsic material destruct limits. Accelerating the applied stresses results in considerable test time reduction.

13.3.2.3 Test Setup Test fixtures should be designed to have appropriate transmissibility. The main components of a test setup are selecting test platforms, identifying sensor monitoring schemes, designing fixtures, identifying failure detection, monitoring procedures, and postprocessing.

13.3.2.4 Selecting the Test Platform Selection of the test platform is driven by the test loads. For example, electro dynamic shakers or repetitive shock chambers can be used for vibration loading. On the other hand, a repetitive shock chamber is more appropriate for simultaneous application of multiple loading, such as temperature and vibration.

13.3.2.5 Stress Monitoring Scheme

Sensor Type The selection of sensor type is dependent on the type of stresses being monitored. For example, accelerometers and strain gauges are used to monitor and control vibrational loads, thermocouples are used to monitor temperatures, and capacitive gauges are used to sense moisture.

Optimal Sensor Location An engineering analysis has to be performed to determine the optimal quantity and location of the sensors. For example, the number of vibration sensors and their strategic placement on the test vehicle is based on preliminary modal analysis of electronic assemblies.

Fixturing Issues The test fixture should be designed and built such that the applied loads (thermal, mechanical, electrical, and chemical) are transmitted to the test vehicle with minimum loss. In addition, the applied loads should be effectively transmitted to potential weak links of the product.

Failure Monitoring and Detection Schemes If the test specimens are functional, dedicated electrical monitoring is required for failure detection, and the stress limit for the applied loads can be the operating limit or the destruct limit. If the test specimens are nonfunctional (daisy-chained), event detectors can be used to detect transient electrical opens, and the maximum stresses applied are limited by the destruct limit. For the verification of electrical failures, spectrum analyzers, oscilloscopes, circuit tracers, and signal conditioners may be used.

Data Acquisition and Postprocessing Schemes The chosen test platform needs to be adequately supported by control systems and data acquisition systems. For example, a test setup for combined temperature and vibration testing of daisy-chained components at the circuit card assembly level requires a test chamber capable of applying the stresses simultaneously. In addition, sensors to monitor, control, and collect data, event detectors for failure detection, and spectral, cycle-counting, and wavelet algorithms to postprocess the data are also required. Based on the stress-monitoring schemes, commercially available or custom-made software may be used for postprocessing schemes. For example, for vibration testing, commercially available software is equipped with time-domain, frequency-domain, and fatigue analysis tools. Testing of circuit card assemblies under a random vibration environment requires collection of data that, upon further postprocessing, can be used to compute the associated damage.

13.3.3 Specimen Characterization

Specimen characterization includes overstress tests to determine destruct limits of the specimen, tests to characterize the response of the specimen to the entire range of loads anticipated in the accelerated test, failure analysis, design of accelerated test profiles, and PoF assessment of expected time to failure under accelerated test loads.

13.3.3.1 Overstress Tests The objective of an accelerated test is to reduce test time by accelerating the degradation that occurs during field conditions. However, the accelerated testing should not precipitate failure mechanisms that may not occur in field or at unintended locations by stressing the product beyond its stress limit. Therefore, to efficiently design an accelerated test profile, the stress limits of a product should be known. Overstress tests are conducted to determine the stress limits, in particular the destruct limits, of a product. For example, to stimulate interconnect fatigue failures in flip-chip packages, the maximum temperature to be applied in the accelerated test may be limited by the thermal properties of the underfill material. The stress limits (specification limits, design limits, and operating limits) obtained from step 1 can be used in load profile design to determine the destruct limits for the product. The general criterion for profile design is to precipitate failures by overstress mechanisms.

13.3.3.2 Specimen Response Characterization under Anticipated Accelerated Test Loads The specimen response should be calibrated over the entire accelerated test load range. The characterization should also include the interaction effects between the applied environmental loads that otherwise would have been overlooked. Furthermore, results from the characterization serve as a verification of the PoF stress analysis conducted in step 1. Note: If the load profiles of overstress tests and ALT are significantly different, additional specimen characterization is required. Otherwise, the data collected from destruct tests can be used to characterize the specimen response over the entire accelerated load range.

13.3.3.3 Failure Analysis Failure analysis is conducted to verify that the dominant failure modes observed in overstress tests are indeed due to overstress and not wearout. Quantifying the overstress data enables verification of preliminary PoF models. Overstress tests, in conjunction with accelerated tests, are useful for determining the spatial location of acceleration transforms estimated in step 1.

13.3.3.4 Accelerated Life Test Profiles Accelerated test profiles should be designed such that wearout failure mechanisms encountered during field use are precipitated. Failure data from overstress tests are used to compute the necessary stress levels required to excite wearout failure mechanisms during ALT. Time-scaling techniques based on the PoF model of the relevant failure mechanism are utlized. For example, in vibration testing of assemblies, time-scaling techniques based on Steinberg's criteria can be used.

13.3.3.5 PoF Assessment of Expected Failures under Accelerated Test Loads Physics of failure is a multistep process. First, the stresses at various key sites of the product are determined from stress analysis. Mechanical, electrical, or chemical stress analyses are generally employed. Second, a finite element analysis (FEA) or some simple closed-form model is used, or a prototype may be built to measure the impact of these stresses on the product under study.

Third, PoF models are employed to assess the reliability under accelerated loading conditions. The quantitative parameters of the loading stresses are the inputs to the PoF failure models. PoF assessment is carried out for all the potential failure mechanisms. For each potential failure mechanism, the corresponding stresses are provided as inputs to the appropriate physics of failure model to obtain life data. The life results from each PoF model are then analyzed to determine the dominant failure mechanisms likely to occur the earliest. To determine the reliability of the system as a whole requires much more complex calculation.

PoF methods generally identify dominant failure mechanisms and treat each failure mechanism individually. However, failures can occur due to interactions between different degradation modes. To aggregate all the degradation mode information to the system level requires reliability tools, such as plot diagrams. The input for each block in a plot diagram comes from the individual physics of failure models. The plot diagram provides information about the system-level reliability.

Sensitivity studies are then carried out to determine the outcome when the environment or the design of the product is altered. Sensitivity studies enable reliability engineers to understand the behavior of a product under different environments, thereby making the product more robust. Pareto ranking of potential sources of failure can be used to determine the most dominant failure modes. To improve the reliability of a product, the stress and life margins may be increased. A quicker and more efficient method would be risk mitigation. To mitigate risks, a product can be ruggedized or the stresses can be managed by auxiliary systems. For example, if there is excessive vibration in a system, shock absorbers may be added. If there are excessive thermal stresses, active cooling devices may be implemented. Risk mitigation solutions can be implemented based on the margins.

Based on the initial stress analysis and damage modeling, preliminary estimates of in-service life are obtained. This involves conducting stress analysis, assessing the failure modes, and estimating the acceleration transforms. Accelerated stress tests in conjunction with overstress tests are used to determine the spatial location of the acceleration transform whose functional form has been predetermined from preliminary PoF assessment.

13.3.4 Accelerated Life Tests

Accelerated life tests evaluate the intrinsic vulnerability of a product to applied loads due to wear-out failure mechanisms. Successful implementation of accelerated wear-out test strategies requires: failure mechanisms generated in ALT to be the same as those observed and identified in the preliminary PoF assessment; and extrapolation of results from accelerated tests to field life conditions using acceleration transforms as reliability predictors to enable proactive product design. The primary tasks in this step include implementation of ALT and verification of observed failure modes.

13.3.4.1 ALT Implementation, Data Acquisition, and Failure Analysis Figure 13.5 shows how the time to failure changes with the change in stress level. The time to failure and stress level are plotted on the x-axis and z-axis, respectively. The time to failure is depicted as a distribution, as the ALT results are based on a set of tested samples. As the stress level increases, the mean value of the time to failure decreases, thereby shortening the test duration.

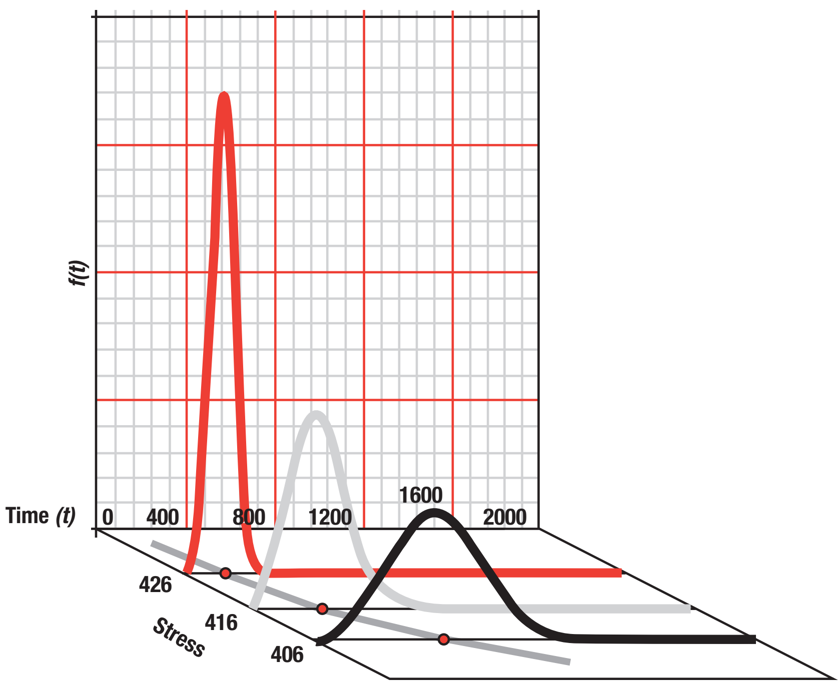

Figure 13.5 Accelerated testing for PoF model calibration.

The shape of the distribution changes at different stress levels. As the width of the distribution decreases, the peak goes up, since the area under the curve is constant. Each curve represents the probability density function at that particular stress level. The concept of accelerated testing is that if a product has to last in the field for 5 years, then, based on this graph, the reliability engineer can estimate the stress levels to precipitate failures in a shorter duration. Thus, with the help of acceleration factors, it can be verified in a short time period that a product indeed has a 5-year life for that failure mechanism. Acceleration transformations need to be developed for each failure mechanism. Typically, accelerated testing is a five-step process, two of which involve physical testing. The rest require PoF modeling, because without PoF modeling, the outcome of the test cannot be quantified. There should be two sets of inputs: the hardware configuration and the life-cycle usage conditions (loading). The output is a time-to-failure assessment or reliability assessment.

The reliability of a product must be designed based on the application conditions. The anticipated savings in the life-cycle and replacement costs as a result of having a reliable product should be the primary motivation for designing a reliable product. For example, if a thousand light bulbs in an auditorium that need to be replaced once per year were replaced with long-life LEDs, the cost savings would be significant. Understanding the life-cycle loading conditions is necessary for designing a reliable product for a specified period of time. For instance, a customer may request an LED manufacturer for a product that lasts for 50,000 hours. To ensure that the product will last 50,000 hours, the LED manufacturer should design its hardware configuration such that accelerated testing is carried out based on the life-cycle stresses experienced by the product.

If a company is in the middle of the life-cycle chain, then it must process information in both directions. The company in the middle of the chain affects the product through the stresses it puts the product through in manufacturing, testing, shipping, handling, storage, and distribution. The LED manufacturer thus needs to know not only what the customer will do with the product, but what the company in the middle of the chain is going to do with it.

13.3.5 Virtual Testing

Virtual testing (VT) is similar to virtual qualification (VQ). However, VQ is for the field configuration (determined in the input phase), whereas VT is for the accelerated test configuration. If we obtain the life-cycle loading conditions, understand hardware configuration, and implement a physics of failure model, then we can develop a virtual assessment methodology of the system's behavior, degradation, and failure. In real time, prognostic and health management (PHM) techniques can track the health of the system and predict anomalies and failures ahead of time. If PHM technologies are unavailable, we can keep updating our assessment by continuous life-cycle monitoring. The goal is to have an instantaneous assessment that continuously updates the remaining useful life of the product.

Physics of failure models update the assessment in real time to determine the remaining useful life. VT assesses time to failure under accelerated life-cycle testing (ALT) loads for the potential failure mechanisms by simulation. Based on a comparison with VQ results, acceleration factors are assessed. PoF methods are used to assess anomalies. PoF methods are also be used to assess the fraction of life used at any

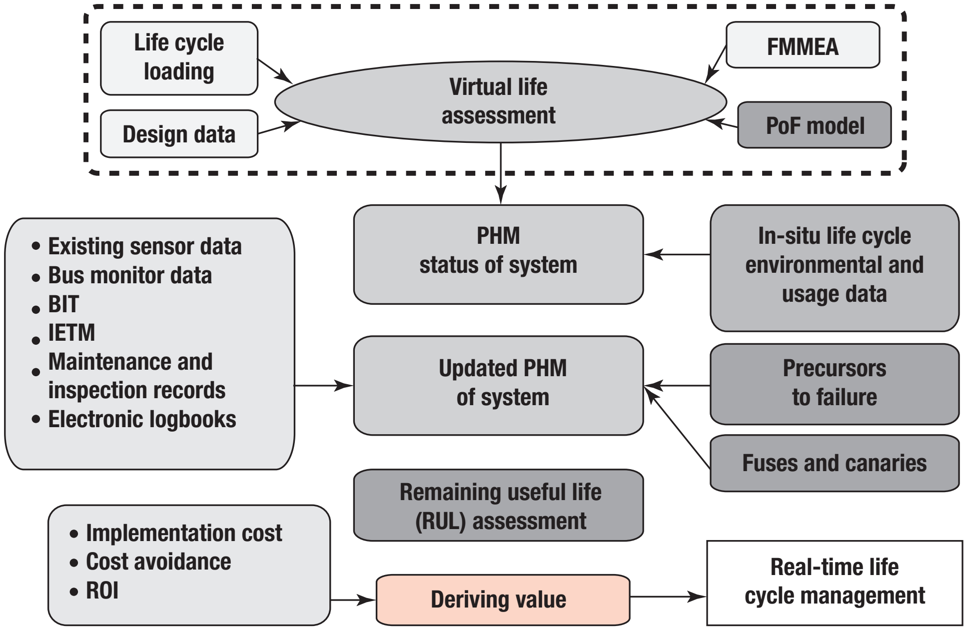

Figure 13.6 PHM hybrid approach.

point of time. However, these are assessments, not measured values. They are results from simulation. The purpose of PHM-based VT is to estimate the health of a system in real time and determine the remaining life (see Figure 13.6). VQ consists of two steps: stress and damage analyses.

13.3.5.1 Stress Analysis Simulations of all dominant load cases during the accelerated test conditions of the specimen are conducted—for example, thermal, thermomechanical, vibration/shock, hygromechanical, diffusion, radiation, and electromagnetic. Potential failure sites are identified through simulations. Physics of failure models are used to determine the prognostic distances. At every stage, by monitoring the actual usage and feeding that back into the PoF model, the virtual life assessment is periodically updated. One can also use PoF to design fuses and identify the best precursors to failure.

13.3.5.2 Damage Analysis After identification of potential failure sites, PoF models for each stress type are applied.

13.3.6 Virtual Qualification

Depending on the specific case, the VQ step is conducted first or last. VQ is similar to VT, except that VQ is for field configuration, while VT is for accelerated test configuration.

The goal of VQ is to determine the reliability risks under combined life-cycle loads by identifying the potential weak links and dominant failure mechanisms. The output of VQ is a ranking of the potential weak links under the expected life-cycle load combinations. PoF is used to identify the potential failures under life-cycle loads. The stress and damage analysis steps in VQ are similar to that in VT, except that the life-cycle loading experienced during operation is simulated.

It is necessary to identify and prioritize the failure mechanisms that could potentially be activated in the product during its life-cycle operation. Depending on the relative severity of the identified failure mechanisms, the predominant stresses of the respective failure mechanisms can be used as parameters for the AST. Accelerated tests should represent actual operating conditions without introducing extraneous failure mechanisms or nonrepresentative physical and/or material behavior. For example, temperature, a parameter that accelerates corrosion-induced failures, may also accelerate ionic contamination failures. The AST should be designed such that there is no possibility of the failure mechanism shifting from corrosion to ionic contamination during the test. Based on the initial stress analysis and damage modeling, preliminary estimates of in-service life are obtained. This involves conducting stress analysis, assessing the failure modes, and estimating the acceleration transforms. Preliminary PoF assessment determines the functional form of a product's acceleration transform for varying qualities that result from manufacturing variabilities. The exact spatial location of the acceleration transform is determined through systematic implementation of exploratory and accelerated stress tests.

13.3.7 Output

An acceleration factor (AF) is estimated by using the results from virtual qualification (step 1) and virtual testing (step 5). The durability of the product is then estimated by extrapolating ALT results using the acceleration factors estimated from simulation. The assumption is that the virtual test and virtual qualification results have the same interrelationship as the product durability under accelerated conditions and life-cycle load conditions. Assuming the relationship holds true, the acceleration factor is estimated as:

$$\mathrm{AF} = t_{field} / t_{\mathrm{ALT}},$$

where t_{field} is the predicted time to failure under life-cycle conditions (result from virtual qualification), and t_{ALT} is the predicted time to failure under accelerated test conditions (result from virtual testing).

PoF models are periodically verified and updated with the results from accelerated testing.

Life assessment provides scientific methods to interpret and extrapolate ALT results to field life estimates through quantitative acceleration factors. The primary features of life assessment techniques include: the ability to correlate, verify, and update PoF model predictions with the observed ALT results; and the ability to forecast product reliability under field conditions and thereby make design tradeoffs to enhance reliability. The time-to-failure under life-cycle loads is obtained by multiplying the acceleration factor with the experimental time-to-failure. Life assessment techniques shift the emphasis from end-of-line testing to PoF-based test methodologies, thereby improving product life-cycle with proactive design and process techniques.

13.4 Case Study: System-in-Package Drop Test Qualification

To demonstrate the PoF qualification methodology, a case study to quantify the reliability of a land grid array (LGA), laminate-based RF (LBRF) system-in-package (SiP) component under drop loading is presented. The physics-of-failure approach (Figure 13.7), which uses the results of accelerated tests and simulations of field and accelerated loadings to estimate the expected lifetime under life-cycle loading, is shown.

The PoF approach to qualification of SiP with accelerated stress testing (AST) under drop test loading is shown in Figure 13.7. As explained in the previous section, the PoF approach consists of five phases. The input information depends on the specific SiP package that is selected for PoF qualification. The life-cycle loading depends on the end use and application. However, it is difficult for the SiP manufacturer to identify all possible use environments. The manufacturer often chooses to qualify a product for a set of standard environments and provides acceleration transforms that different end users can then use to extrapolate the results to their respective end-use environments. The types of loads expected for a SiP are temperature, humidity, vibration, and drop.

13.4.1 Step 1: Accelerated Test Planning and Development

The design and fabrication of a test specimen and experimental setup to subject the SiP package to accelerated stress environments is carried out in this step. The test board is a standard Nokia mechanical shock test board with 12 LGA SiP components laid out in rows of two each (Figure 13.9). The connector harness for daisy-chain monitoring is on the edge of the test board. Each component has a daisy-chain network connecting all of the perimeter bumps of the LGA such that, if one interconnect fails, the network reports failure. Five of the 10 boards are populated on one side, and five on the other side. The test board is fixed to the drop platform from the four mounting holes at each corner of the board.

The accelerated loads to be used in the qualification program are then selected based on the loading expected in use environments. The types of accelerated tests planned for SiP packages based on the input loads are thermal cycling, thermal shock, temperature/humidity/bias, moisture cycling, vibration, and drop. In this case study, the qualification of a SiP package under drop loads is demonstrated.

The accelerated test plan for drop testing of LGA SiP specifies dropping the test vehicle to an acceleration of 1500 G, which corresponds to just over 1 m height. The drop is guided in the out-of-plane direction with the component side down.

The center four components (rows 3 and 4) are expected to experience the maximum loading under drop loading (see Figure 13.8). The loading experienced is dependent on the stiffness of the board material. However, even if the stiffness at the boundaries were infinite, the center of the board would experience the maximum loading. Rows equidistant from the centerline of the board are expected to experience the same loading conditions due to the symmetry of the board. (It will be shown that this is not, in fact, a valid assumption.) The continuity is monitored in real time with high-speed DAQ.

Hardware and software are required to monitor the response of test vehicles to the applied loading and for data acquisition and postprocessing. Hardware to monitor

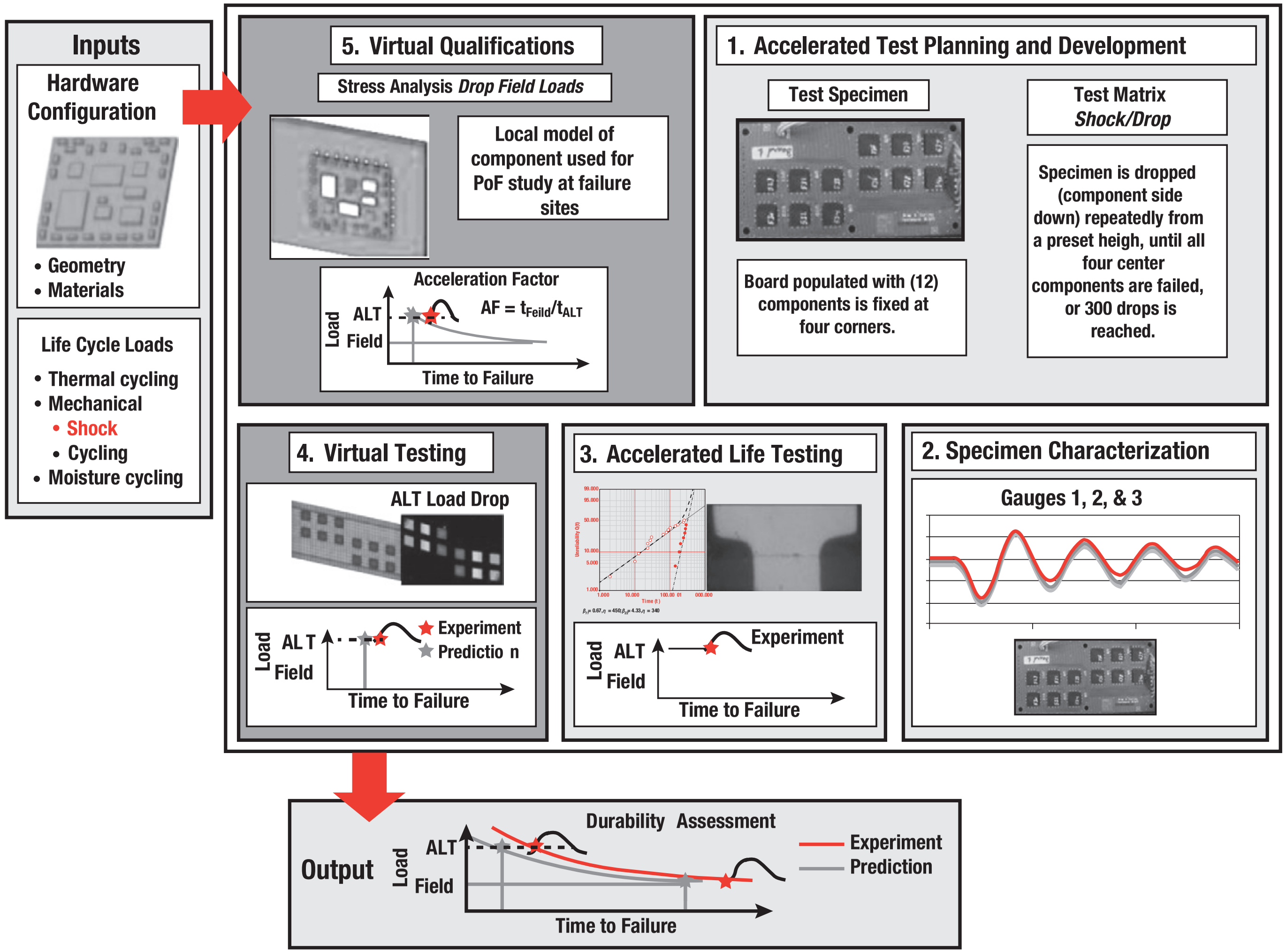

Figure 13.7 Physics-of-failure flowchart for SiP qualification.

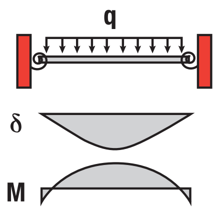

Figure 13.8 Loading, deflection, and moment diagram (2D) for the test vehicle.

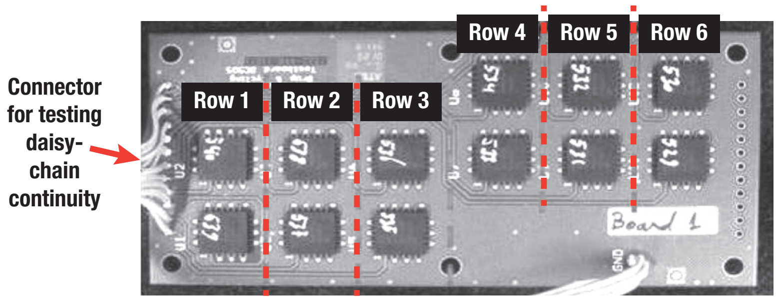

Figure 13.9 Drop test vehicle showing the six classifications of component. The rows of components that are expected to see similar loading conditions are 1 and 6; 2 and 5; and 3 and 4.

response includes sensors such as thermocouples, moisture sensors, accelerometers, strain gages, and electrical parametric sensors.

Hardware designed and implemented for real-time failure detection during accelerated stress testing is also used. Examples include built-in tests for real-time functional checks, event detectors for checking the status of interconnects in daisy-chained mechanical specimens, and data loggers that detect increases in resistance due to interconnect degradation. Methods and algorithms are identified for data postprocessing to quantify the specimen response and damage.

13.4.2 Step 2: Specimen Characterization

The purpose of specimen characterization is to understand how the specimen is physically responding to accelerated loading and to collect data to calibrate simulation-assisted models for acceleration factors. The test specimen's response to test loads is explored by sensor placement. Examples include modal characterization for vibration or drop tests, temperature mapping for thermal tests, or moisture absorption mapping for humidity tests.

To characterize the specimen's response to mechanical drop, specimens attached with a strain gauge and an accelerometer were dropped from a drop tower. This step documented actual strain and acceleration histories at strategic locations on the board during a drop event. The gauge locations can be seen in Figure 13.10. Gauges 4 and 5 are oriented in the direction of the shorter edge of the board, whereas all other gauges are in the direction of the longer edge of the board.

Readings were recorded at four different drop heights, from 0.25 to 1 m. The strain histories for gauges 1, 2, and 3 are shown in Figure 13.11. The acceleration histories of the test vehicle, for each drop height, are shown in Figure 13.12. Gauges 2 and 3, which correspond to components U5 and U8, respectively, gave the most consistent data and were used to extrapolate a majority of the durability data. As seen in Figure

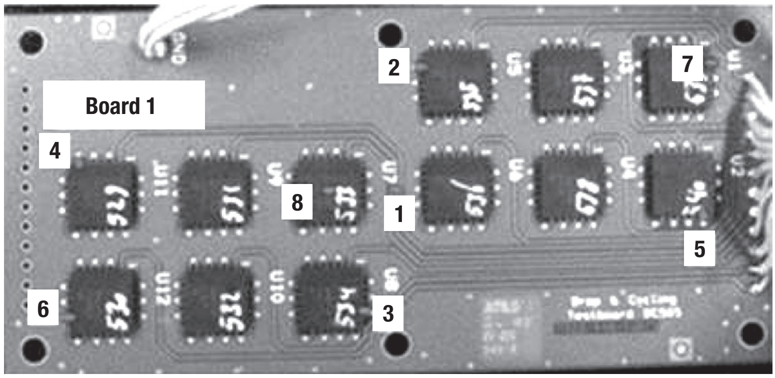

Figure 13.10 Gauge placement: gauges 1–7 are on the reverse side of the board shown here.

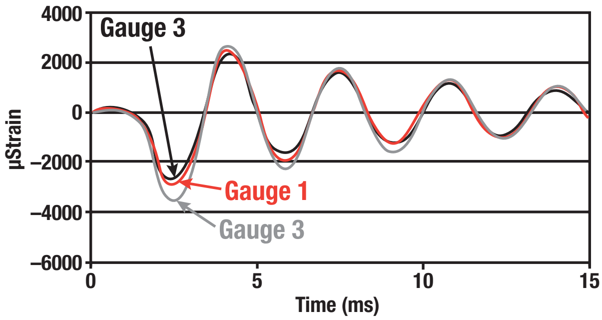

Figure 13.11 Examples of strain gauge time histories: 0.75-m drop. Gauge 3 output is larger than gauge 2 output. Gauge 1 is not the mean of 2 and 3.

13.13, gauges 2 and 3 did not experience the same loading profile, as expected. The ratio of the two strain ranges is listed in Figure 13.13 and is used later when the relations between drops-to-failure and various experimental data are calculated (strain range, acceleration, drop height, etc.). The strain range ratio between gauges 2 and 3 is $\Delta\varepsilon_3/\Delta\varepsilon_2 = 6567/5220 = 1.26$ (in microstrain).

The purpose of AST is to accelerate relevant wear-out failure mechanisms to quickly qualify the product for its life cycle. The overstress limits of the test specimen are identified at this stage to avoid the possibility of inadvertently accelerating the loading levels beyond the overstress limits of the test vehicle. These limits can often be assessed by examining material datasheets. For instance, temperature limits can be identified by looking at the phase transition limits (e.g., glass transition temperatures, recrystallization temperatures, and melting temperatures) of the materials used in the test article. When the datasheets cannot provide relevant data, step-stress testing techniques such as HALT™ can be used to systematically explore the overstress limits of the product. The safe load limits for the accelerated wearout test can then be selected by suitably scaling back from the overstress limits.

13.4.3 Step 3: Accelerated Life Testing

In this step, the accelerated stress regimes designed in step 2 are applied long enough to cause accelerated wear-out failures in the test specimen. Tests are carried out on a statistically significant sample size, and the time to failure of each sample is recorded. Failure data are grouped by failure mechanisms, and the statistics of failure

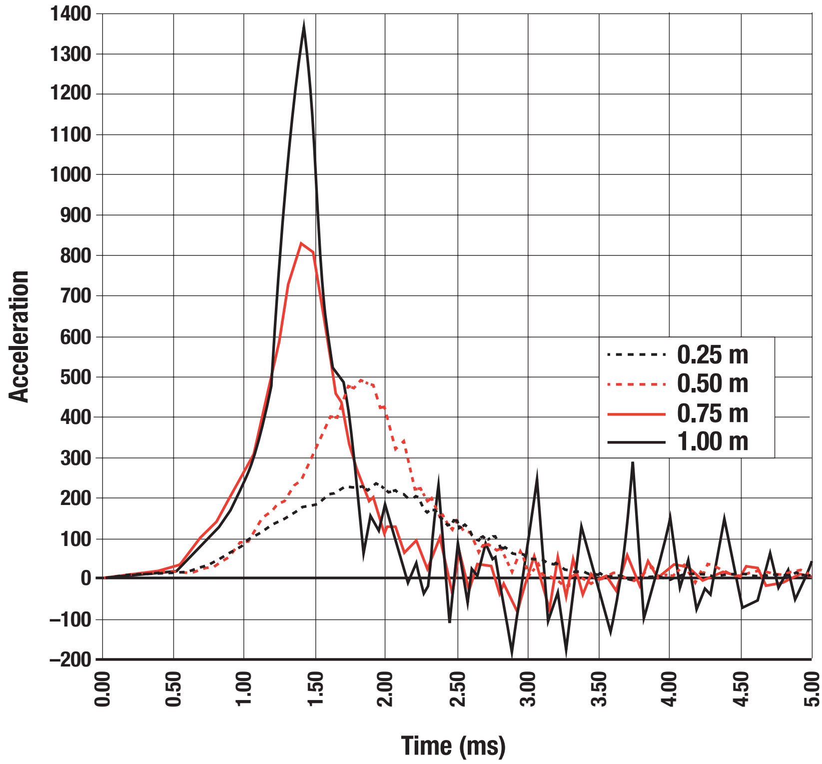

Figure 13.12 Filtered acceleration history data for each of the four drop heights.

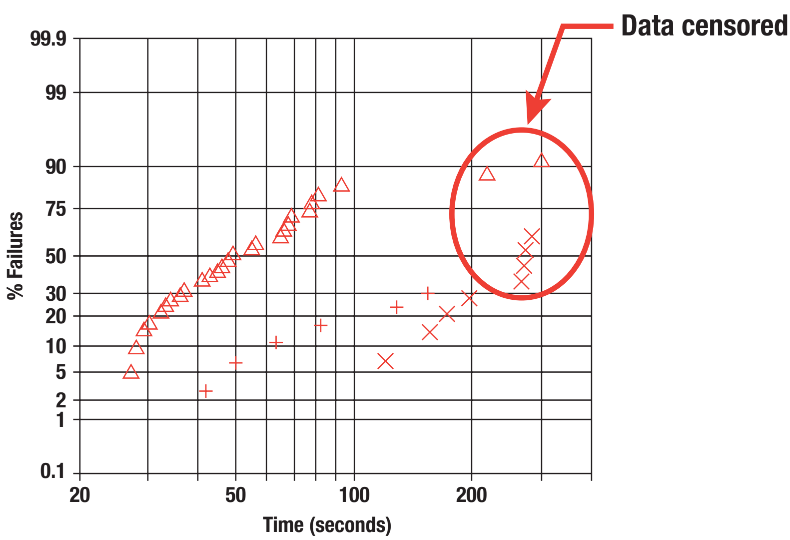

Figure 13.13 Lognormal plot of the three groups: rows 3 and 4 (Δ), rows 1 and 6 (+), and rows 2 and 5 (X). The data that seems to follow a separate failure mechanism is censored.

distributions are assessed for similar failure mechanisms. These failure distributions are later extrapolated using PoF simulations to field loading conditions. Failure analysis is conducted on the failed specimens to determine the failure sites and mechanisms.

Resource limitations may sometimes force early termination of the test after a fixed period, even if no failures are generated. In this case, PoF simulations for potential failure mechanisms can be used to assess the acceleration factors for each mechanism. Thus, the test duration can be extrapolated to the equivalent field duration for each potential failure mechanism.

Drop testing was conducted for 10 test boards, each with 12 components. For the first and last drop tests, the board was dropped until one specimen registered failure. The remaining eight boards were dropped until either all center four components registered failures, or 300 drops, whichever occurred first. The raw data from drop test is reported in Table 13.2.

In Table 13.3, it appears that the high end of the data follow a different pattern and therefore belong to a different failure mechanism. These data are censored from this qualification.

Table 13.4 shows the ratios of drops to failure for the two components that correspond to strain gauges 2 and 3 from the specimen characterization section. These data will be used later to calculate the durability data.

Failure analysis is conducted after termination of the test to identify the failure sites and mechanisms. Nondestructive failure analysis techniques should be employed initially to extract as much information about the failure sites and mechanisms as possible, followed by destructive failure analysis techniques. Visual, optical microscopy, and X-ray inspection methods are commonly employed nondestructive techniques. In the case of SiP, X-ray spectroscopy was utilized to identify the potential failure sites, as shown in the left image in Figure 13.14. Although extensive voiding was observed in the X-ray images, no cracks were seen.

After X-ray inspection, the SiP specimens were subjected to a "dye and pry" technique. In a standard dye and pry process, a failed component attached to the board is soaked in an indelible dye. In the case of a mechanically shocked/cycled specimen, the board is also flexed while soaking. The dye is expected to seep into cracks that exist on the interconnects. After the dye dries, the component is pried off the board. The presence of dye on the interconnects of the pried sample confirms the presence of an existing crack due to drop testing. In this study, the dye and pry technique was modified by milling up through the board close to the component but stopping just before reaching it. The purpose of this modified approach was to compensate for an LGA's extremely low stand-off height, which makes it very difficult to pry off without damaging the perimeter.

Typical results from a dye and pry analysis are shown on the right side of Figure 13.14. The presence of dye shows that there are edge delaminations in the large center thermal pad, but no fractures in the peripheral solder joints. Five interconnects had fracture through the solder, and the rest had copper traces pulled off near the component during the prying process. However, since no dye was observed in these fractures, the fracture is assumed to be an artifact of the prying.

Another commonly used destructive failure analysis technique is to cross-section the sample in order to identify the failure sites under optical or scanning electron microscopy. However, no fracture sites that would result in electrical failure were observed in the failed SiP.

Table 13.2 Overall drop test data of 10 boards

Side	U1–U12	U13–U25	U13–U25	U1–U12	U13–U25	U1–U12	U13–U25	U1–U12	U13–U25	U1–U12
Board No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Device No.										
1				42		63		50		
2				0	82					
3		0		284		0				
4			196	0		274		271		
5		68	219		78		56	300	33	
6	45	49	69	43	28	41	30	48	41	46
7		32	32	55	37	66	28	65	29	
8		36	29	93	27	77	34	81	27	
9			172					267		
10						120		156		
11										
12				128		154				

Boards 2–9 were dropped until failure of the center devices (U5–U8). The first and last boards were dropped until the first failure.
The data of the three symmetrical groups (rows 3 and 4, 2 and 5, and 1 and 6) are plotted in a lognormal plot of percent failed versus drops to failure.

Table 13.3 Reliability data for the three groups: number of drops until 50% fail (N50), 95% confidence level (95% CL), and shape parameter (σ)

Location	N50	95% CL	σ
Row 1, 6	270	183–414	0.806
Row 2, 5	394	250–656	0.806
Row 3, 4	56	43–73	0.806

Table 13.4 Drops-to-failure ratios for the components in the U5 and U8 positions for side 1 and side 2

	U1–U12	U13–U25
N_{f5}/N_{f8}	$300/83.67 = 3.6$	$90.8/30.6 = 3.0$

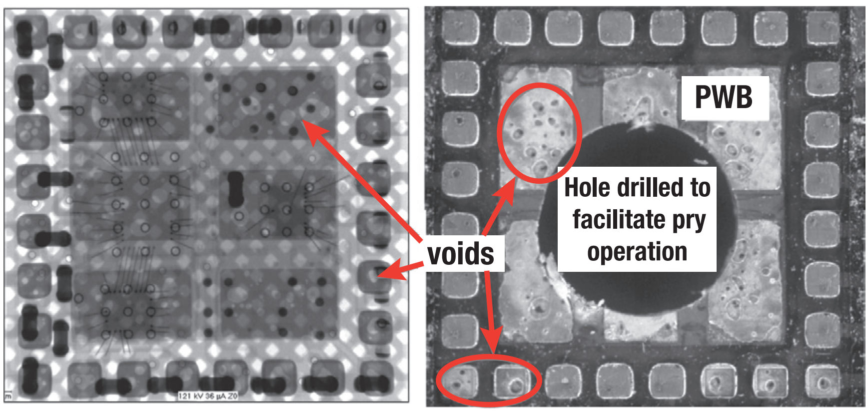

Figure 13.14 X-ray (left) and “dye and pry” images of a typical component.

The failure site and mechanism were identified through a process of desoldering the interconnects, polishing the surface of the board, and examining the copper traces on the board. All desoldered components were found to have failures: in the copper traces at the neck after transition from the pad area, and on the side of the components that is parallel to the short edge of the board (shown in Figure 13.15). Figure 13.16 is an elevation view of the crack that confirms that the cracks were induced by fatigue due to the bending of the board.

13.4.4 Step 4: Virtual Testing

Virtual testing is conducted with PoF simulations to identify the potential failures under the accelerated stress testing (AST). First, the stresses are assessed at the failure sites observed in step 3 under the applied accelerated loads. A combination of experimental measurements on the test vehicle and/or modeling of the test vehicle response to the accelerated loading are carried out.

Second, damage accumulation rates are assessed using appropriate damage models for each wearout failure mechanism. The output of this step is a prediction of mean time to failure for accelerated test loading. The test results from step 3 are used to calibrate the model, if necessary.

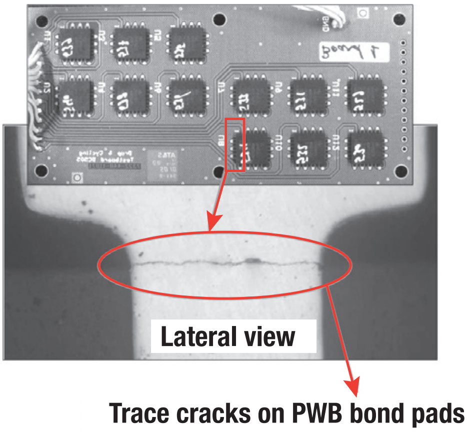

Figure 13.15 Failures were finally found after desoldering the components from the board and polishing the surface of the board. A faint line can be seen in the FR4 on either side of the copper crack.

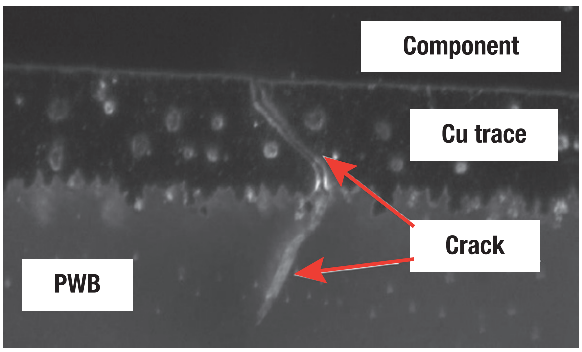

Figure 13.16 Elevation view of the crack, confirming that the failure mode was bending as opposed to pad lift-off.

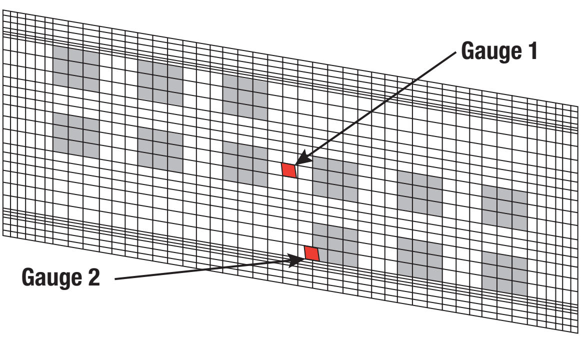

Figure 13.17 Modal analysis to correlate flexural strain to failure data 2D four-node "shell" elements. PWA is modeled with uniform thickness. Footprints of SiPs are given different properties: density and stiffness are increased to represent PWB + SiP; boundary conditions: four corners fixed.

13.4.5 Global FEA

A modal analysis using a 2D representative model using shell elements was carried out in finite element modeling (FEM), as shown in Figure 13.17. Boundary conditions, geometries of the board/components, and material properties of the board/components were inputs to the model.

Using the specimen characterization data, the FEA model can be correlated to drop test conditions. Data can also be collected from the areas in the model where the strain gauges were located and compared.

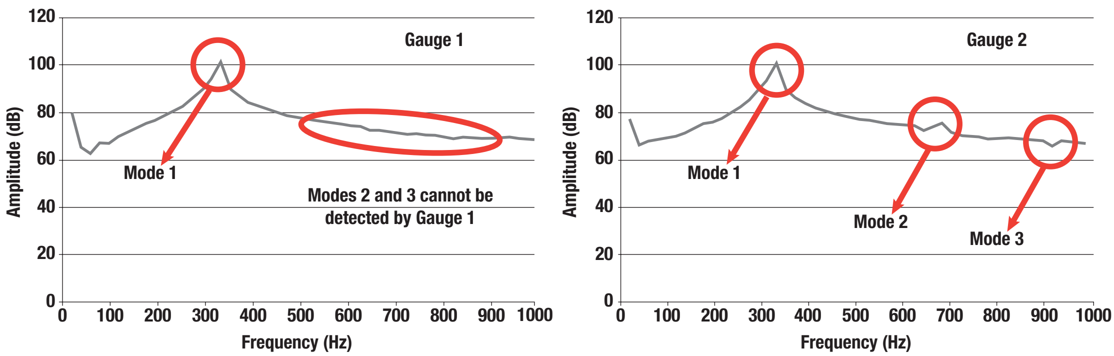

Figure 13.18 FFT of the measured strain histories from the specimen characterization phase. The chart on the left is from gauge 1 and clearly shows influence from only mode 1. The chart on the right is from gauge 2 and shows influence from the first three modes.

Table 13.5 Comparison chart of the measured modal frequencies versus those calculated by the finite element model

Mode frequency comparison	Measured (Hz)	FEA (Hz)	% Difference
Mode 1	332	350	5
Mode 2	703	730	4
Mode 3	918	1000	9

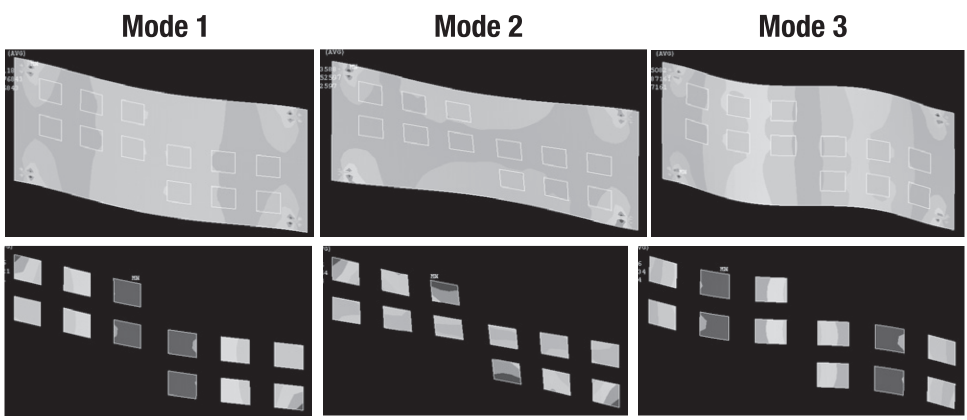

Figure 13.19 Bending strain contours of the first three mode shapes of the entire model (top row) and just the component footprints (bottom row).

Figure 13.18 shows fast Fourier transforms (FFTs) of the strain histories gathered in the specimen characterization phase. Table 13.5 shows the comparison of the first three modes observed in the characterization and calculated by the FEM.

13.4.6 Strain Distributions Due to Modal Contributions

FEM analysis shows why gauge 1 and gauge 2 provided different data. Figure 13.19 shows that mode 2 is a twisting mode. Since the major contributions to the

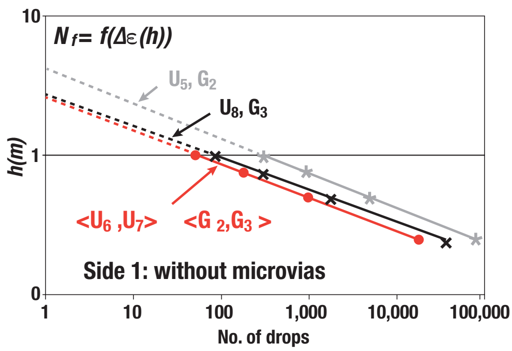

Figure 13.20 Number of drops to failure as a function of drop height for U8, U5, and the average of U6 and U7.

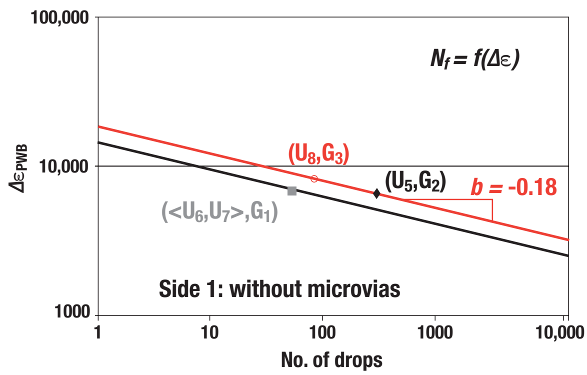

Figure 13.21 Number of drops to failure as a function of the strain-range measured on the PWB for gauges 2 and 3 (U8 and U5).

deformation seem to be from modes 1 and 2, by superimposing mode 2 on mode 1, the deformations are subtracted at gauge 1 and added at gauge 2.

13.4.7 Acceleration Curves

Acceleration curves are created by combining the strain gauge data from the gauges placed near the critical solder joints of the critical components. Acceleration curves are created as functions of parameters such as drop height (Figure 13.20) and PWB strain range (Figure 13.21). The results from acceleration curves can be extrapolated to different drop heights or different strain ranges for this specimen.

13.4.8 Local FEA

A local model of one component and the surrounding PWB was developed in FEA (Figure 13.22). The purpose of a local FEA model was to better understand the stress and strain fields in the component during bending and to correlate the observed PWB strain at the gauges to the solder strain experienced in the critical solder joint. Even though the failure site was observed in the copper trace, this FEA transform can used to predict the strain at the solder interconnects.

The local model was constrained at the bent edge, in the plane of the board, to simulate the stretch in the PWB that also happens during the drop event. The strain field predicted in the FEA (seen in Figure 13.23) was used to develop the strain transfer function shown in Figure 13.24. Using the strain transfer function and the PWB

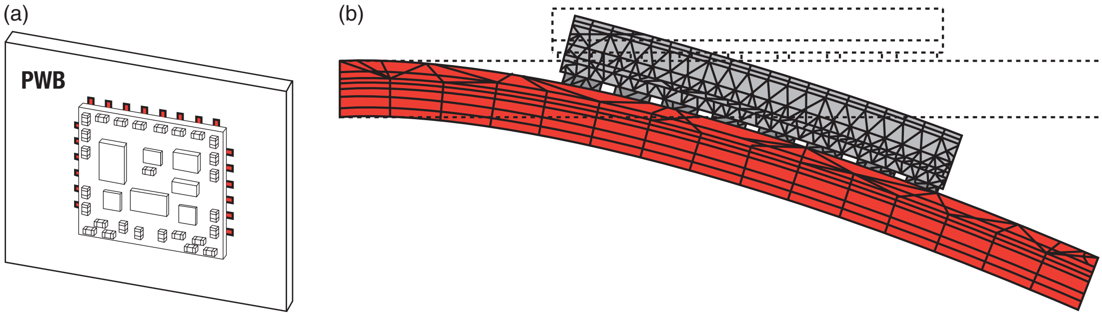

Figure 13.22 3D local FEA model of U8 component: (a) top view, decapsulated; and (b) side view, deflected.

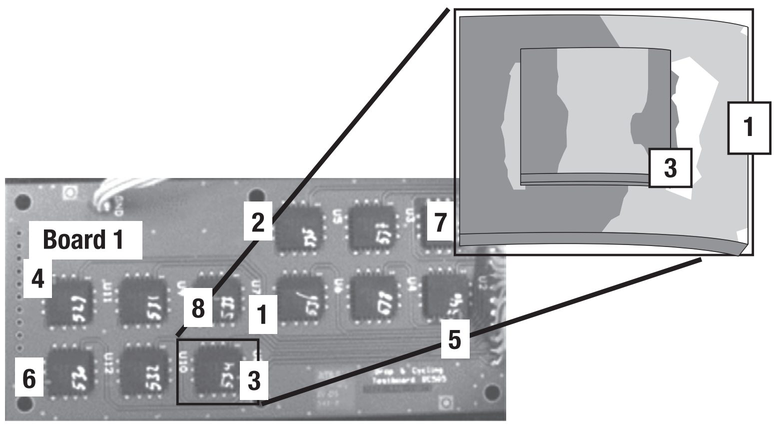

Figure 13.23 Schematic of the location of the local model with respect to the test board and the corresponding strain fields developed in component U8 during drop.

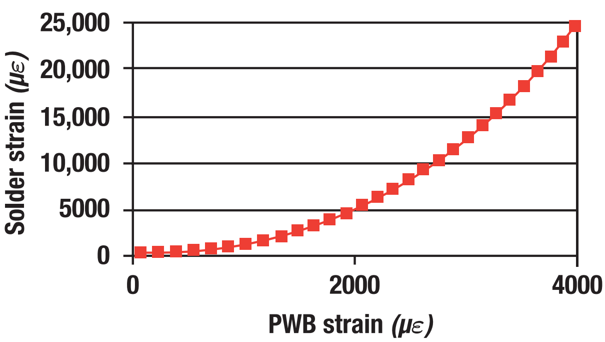

Figure 13.24 Strain transform from measured PWB strain to solder strain experienced by the critical solder joint, as predicted by FEA.

strain measured from the strain gauge, the bending strain experienced by the critical solder joint is estimated.

13.4.9 Step 5: Virtual Qualification

In this step, the time to failure is assessed for the same failure mechanisms addressed in step 4, but now under life-cycle loading conditions. The ratio of the estimated time

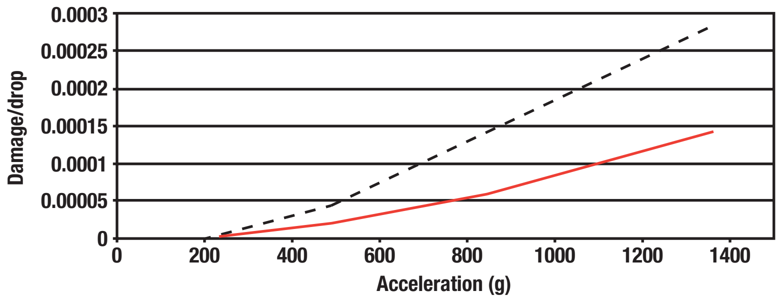

Figure 13.25 Damage curves for components U5 and U8 as a function of the drop acceleration.

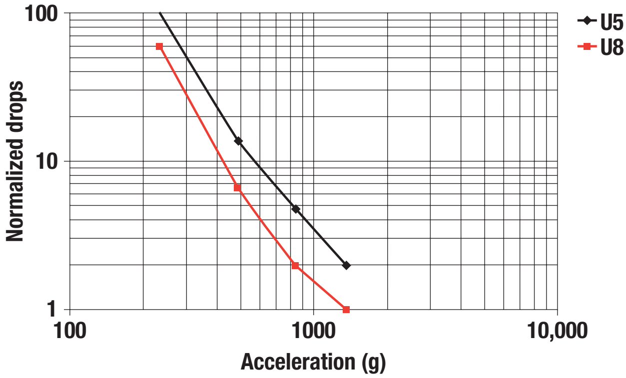

Figure 13.26 Acceleration curves for solder failure: normalized number of drops to failure versus drop acceleration.

to failure in steps 4 and 5 provides the acceleration factor between the test environment and the field environment.

The simulations to obtain acceleration factors for the SiP include stress analysis: thermal (steady-state/transient), mechanical (static/dynamic), thermomechanical, moisture absorption, and hygromechanical; and damage models, including fatigue, fracture, creep, corrosion. The final output of the process is the durability assessment of the product, which is obtained by extrapolating the accelerated test results to field conditions through the use of acceleration factors.

The durability data can be transformed to a function of drop acceleration, as shown in Figure 13.25. This is a more general metric to use, as the actual acceleration experienced by the impact event can change, for any given height, based on the drop conditions. The total strain range experienced by the gauge for each of the first four pulses during the impact was measured. This includes initial impact and clatter afterwards. The damage for each pulse was calculated and related to the acceleration that caused the strain range.

13.4.10 PoF Acceleration Curves

Finally, the acceleration curves can be calculated, as shown in Figure 13.26.

13.4.11 Summary of the Methodology for Qualification

The five-step PoF approach was illustrated for the qualification of a SiP-type package in a drop-loading environment. Although failure did not occur at the solder interconnects, valuable insights were gained from the PoF study. It was observed that the increased robustness of the package type resulted in the transfer of weak points to the copper traces in the test board. The insights from this PoF study will allow for board redesign.

It is good practice to state the results of all engineering analyses with the degree of certainty (or uncertainty) associated with it, for example, confidence intervals. Similar to the way confidence intervals around estimates can be used to estimate unknown distribution parameters, confidence intervals around a regression line can be used to estimate the uncertainties associated with regression relationships.

Specifics about the type of confidence interval used are also imperative while reporting the failure data. The confidence level, one- or two-sided interval, sample size, how the samples were chosen, and the methods of analysis are some information to be included.

Under some circumstances, estimation and visualization of the confidence interval may not be possible. For example, a very small sample size is likely to produce a very wide confidence interval that has no practical use. In such cases, data visualization techniques are used to display the complete results without making any statistical claim to facilitate making judgments on the data.

13.5 Basic Statistical Concepts

A population is a set of data collected from all the members of a group. A sample is a set of data collected from a portion of the population. Since it is not possible or even advisable to measure the whole population (e.g., the act of measurement could damage the samples and make them unusable), data obtained from a sample are used to make estimates about the population. Figure 13.27 describes a schematic of estimating population parameters from a sample. To obtain the population parameters from a sample, the population and sample must be created from the same process.

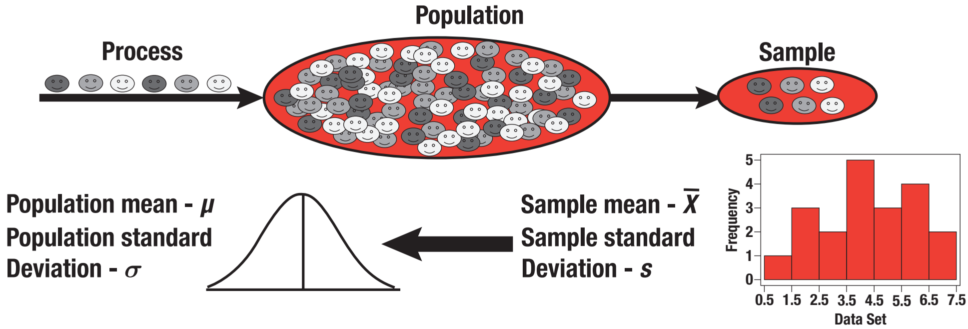

Figure 13.27 Schematic of estimation of population parameters from sample parameters.

An underlying random variable of interest is denoted by X. The variables X_1, $X_2, \ldots, X_n$ are random samples of size n from the population represented by X, if they are all independent and have the same probability distribution based on the random variable X. The observed data, $X_1, X_2, \ldots, X_n$, is also referred to as a random sample. A statistic is a point estimate derived from the observed data and is defined as $\hat{\Theta} = g(X_1, X_2, \ldots\ldots, X_n)$. Some examples of a statistic are mean

$$\bar{X} = \sum_{i=1}^{n} X_i$$

and variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

13.5.1 Confidence Interval

A confidence interval is an interval estimate computed from a given data sample that includes the actual value of the parameter with a degree of certainty. The width of the confidence interval is an indication of the uncertainty about the actual parameter. The confidence interval puts a boundary around these point estimates and provides the likelihood that the population parameters are within those boundaries.

Inferential statistics is used to draw inferences about a population from a sample. Statistics from a sample include measures of location, such as mean, median, and mode, and measures of variability, such as variance, standard deviation, range, or interquartile range.

Standard deviation of a set of measurements is not the same as confidence interval. Standard deviation is a measure of the dispersion of a measurement. In general, the greater the standard deviation is, the wider is the confidence interval on the mean value of that measurement. However, there is more to the statistics of a set of measurements than standard deviation.

When the probability of θ being in the interval between l and u is given by $P(l \le \theta \le u) = 1 - \alpha$, where $0 \le \alpha \le 1$, the interval $l \le \theta \le u$ is called a $100 \times (1 - \alpha)$ percent confidence interval. In this definition, l is the lower confidence limit, u is the upper confidence limit, and $(1 - \alpha)$ is called the confidence level, usually given as a percentage.

A confidence interval can be either one or two sided. A two-sided (or two-tailed) confidence interval specifies both a lower and upper bound on the interval estimate of the parameter. A one-sided (or one-tailed) confidence interval specifies only a lower or upper bound on the interval estimate of the parameter. A lower one-sided $100(1 - \alpha)$ percent confidence interval is given by $l \le \theta$, where l is chosen so that $P(l \le \theta) = 1 - \alpha$. Conversely, an upper one-sided $100(1 - \alpha)$ percent confidence interval is given by $\theta \le u$, where u is chosen so that $P(\theta \le u) = 1 - \alpha$.

13.5.2 Interpretation of the Confidence Level

The common perception is that the confidence level is the probability of a parameter being within the confidence interval. Although this assumption is intuitive and gives

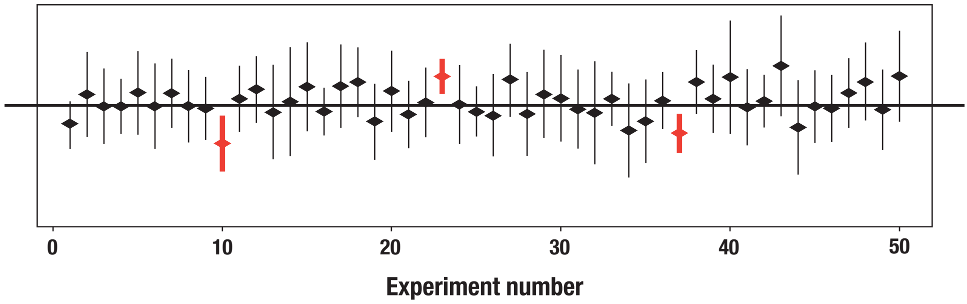

Figure 13.28 Conceptualization of confidence interval.

a measure of understanding, the conceptual definition of confidence interval is more subtle. One engineering statistics textbook (Montgomery and Runger 2007, p. 262) states the nuance in the following way: "in practice, we obtain only one random sample and calculate one confidence interval. Since this interval either will or will not contain the true value of θ, it is not reasonable to attach a probability level to this specific event. The appropriate statement would be that the observed interval [*l*, *u*] brackets the true value of θ with confidence level 100(1 − α). This statement has a frequency implication; that is, we don't know if the statement is true for a specific sample, but the method used to obtain the interval [*l*, *u*] yields correct statements 100(1 − α) percent of times."

Figure 13.28 shows fifty confidence intervals on the mean computed from samples taken from a population at a confidence level of 95%. The solid line represents the true mean calculated from the whole population. We expect that 95% of all possible samples taken from the population would produce a confidence interval that includes the true value of the parameter being estimated, and only 5% of all samples would yield a confidence interval that would not include the true value of the parameter. The simulated case shows that three (approximately 5%) of the confidence intervals do not contain the true mean.

With a fixed sample size, the higher the confidence level is, the larger the width of the interval will be. A confidence interval estimated at a 100% confidence level will always contain the actual value of the unknown parameter, but the interval will stretch from $-\infty$ to $+\infty$. However, such a large confidence interval provides little insight. For example, we can say with a very high confidence level that the age of all students in a reliability class is between 1 and 150 years, but that does not provide any useful information.

Selection of the confidence level is part of the engineering risk analysis process. For example, with a confidence interval analysis, the expected worst cases on warranty returns over a period can be estimated. An estimate can then be made of the spare parts to stock based on the point estimate of a 95% or 99% confidence level (or any other chosen value) of the expected warranty return. The decision will depend on the balance between the cost of storing the spares versus the cost of delay in repair time due to the unavailability of spares. In many engineering situations, the industry practices or customer contracts may require the use of a specific confidence level—frequently, values of 90% or 95% are quoted.

13.5.3 Relationship between Confidence Interval and Sample Size

The value of the confidence intervals depends on the measurements for each sample. As long as the measurements made on the samples are from the same population, an increase in sample size will reduce the width of the confidence interval, provided that the confidence level is kept constant. However, when conducting an experiment or gathering data from the field, data may come from multiple populations; in those cases, a large sample size may actually increase the confidence interval. For example, in the manufacturing of baseball bats, the hardness values of samples taken from the production line can be recorded. If the production parameters are all under control, then increasing the number of samples that come from the same population will narrow the confidence interval. However, if for some period the production parameters are out of control, the hardness values for samples taken during those times will differ. Therefore, increasing the sample size by including samples from the "out of control" population will increase the confidence interval.

13.6 Confidence Interval for Normal Distribution

Concepts of the confidence interval are often illustrated using the normal distribution, partly because it is a symmetric distribution described by two parameters. In a population with normal distribution, there is a direct relation between confidence interval and sample size.

This section describes the calculation of confidence intervals for three cases: confidence interval on an unknown mean with known variance, confidence interval on an unknown mean with an unknown variance, and confidence interval on differences between two population means with a known variance.

13.6.1 Unknown Mean with a Known Variance for Normal Distribution

Consider a population with an unknown mean, μ, and a known variance, σ^2. The variance may be known from past experience or prior data, such as physical processes that create the population or the control charts. For this population, random samples of size n yield a sample mean of $\bar{X}$. The $100(1-\alpha)$ percent confidence interval for the population mean is given by:

$$\bar{X} - \frac{Z_{\alpha/2}\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{Z_{\alpha/2}\sigma}{\sqrt{n}}, \tag{13.1}$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution. Correspondingly, to obtain the one-sided confidence intervals, Z_α replaces $Z_{\alpha/2}$; setting $l = -\infty$, and $u = +\infty$, in the two cases, respectively, the one-sided confidence intervals are given by:

$$\mu \leq u = \bar{X} + \frac{Z_\alpha \sigma}{\sqrt{n}} \tag{13.2}$$

and

$$\bar{X} - \frac{Z_\alpha \sigma}{\sqrt{n}} = l \leq \mu. \tag{13.3}$$

When using a sample mean, $\bar{X}$, to estimate the actual but unknown mean, μ, the "error" is $E = |X - \mu|$. With a confidence level of $100(1 - \alpha)$, for a two-sided interval, the error is within the precision of estimation given by:

$$E \leq \frac{Z_{\alpha/2}\sigma}{\sqrt{n}}. \tag{13.4}$$

Therefore, we can choose a sample size, n, that allows $100(1 - \alpha)$ percent confidence that an error will not exceed a specified amount, E.

$$n = \left[\frac{Z_{\alpha/2}\sigma}{E}\right]^2, \tag{13.5}$$

where n is rounded up to the next integer.

Example 13.1

Consider measuring the propagation delay of a digital electronic part. You want to have a 99.0% confidence level that the measured mean propagation delay is within 0.15 ns of the real mean propagation delay. What sample size do you need to choose, knowing that the standard deviation of the propagation delay is 0.35 ns?

Solution:
Using Equation 13.5, the value of n is found to be 37.

$$n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2 = \left(\frac{Z_{.005} \times 0.35}{0.15}\right)^2 = \left(\frac{2.58 \times 0.35}{0.15}\right)^2 \approx 37.$$

In this application, α is 0.01 and $\alpha/2$ is 0.005. From the standard normal table, $Z_{0.005} = 2.58$.

13.6.2 Unknown Mean with an Unknown Variance for Normal Distribution

The t-distribution is used to develop the confidence interval in this case. Assuming the population to be normal, the sample variance, S^2, is used to estimate the population variance, σ^2, which is not known. Then,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}, \tag{13.6}$$

has t-distribution with $n - 1$ degrees of freedom.

Suppose a population has an unknown variance, σ^2. A random sample of size n yields a sample mean, $\bar{X}$, a sample variance, S^2, and as an upper $\alpha/2$ percentage point of the t-distribution with $(n - 1)$ degrees of freedom. The two-sided $100(1 - \alpha)$ percent confidence interval in this case is given by:

$$\bar{X} - \frac{t_{\alpha/2,n-1}s}{\sqrt{n}} \le \mu \le \bar{X} + \frac{t_{\alpha/2,n-1}s}{\sqrt{n}}. \tag{13.7}$$

Example 13.2

The tensile strength of a synthetic fiber used to manufacture seatbelts is an important characteristic in predicting the reliability of the product. From past experience, the tensile strength can be assumed to be normally distributed. Sixteen samples were randomly selected and tested from a batch of fibers. The sample's mean tensile strength was found to be 49.86 psi, and the sample's standard deviation was found to be 1.66 psi. Determine an appropriate interval to estimate the batch mean tensile strength.

Solution:
Since we are only concerned with tensile strengths that are too low, a one-sided confidence interval on the batch mean, μ, is appropriate. Since the population (batch) variance is unknown and the sample size fairly small, a confidence interval based on the t-distribution is necessary. A one-sided, 99% confidence interval for the batch mean μ is:

$$\bar{X} - \frac{t_{\alpha,n-1}S}{\sqrt{n}} \le \mu \Rightarrow 49.86 - \frac{(1.753)1.66}{\sqrt{16}} \le \mu \Rightarrow 49.13 \le \mu.$$

13.6.3 Differences in Two Population Means with Variances Known

A confidence interval for the difference between means of two normal distributions specifies a range of values within which the difference between the means of the two populations $(\mu_1 - \mu_2)$ may lie. A random sample, n_1, from the first population, with a known standard deviation of σ_1, yields a sample mean of X_1. Similarly, a random sample, n_2, from the second population, with a known standard deviation of σ_2, yields a sample mean of X_2. Then, a two-sided $100(1 - \alpha)$ percent confidence interval for the difference between the means is given by:

$$\bar{X}_1 - \bar{X}_2 - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le (\mu_1 - \mu_2) \le \bar{X}_1 - \bar{X}_2 + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}, \tag{13.8}$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of the standard normal distribution.

Example 13.3

Tensile strength tests are performed on two different types of aluminum wires used for wire bonding power electronic devices. The results of the tests are given in the following table:

Type	Sample size, *n*	Sample mean tensile strength (kg/mm^2)	Known population standard deviation (kg/mm^2)
1	15	86.5	1.1
2	18	79.6	1.4

What are the limits on the 90% confidence interval on the difference in mean strength ($\mu_1 - \mu_2$) of the two aluminum wires?

Solution:

$$\begin{aligned} l &= \bar{X}_1 - \bar{X}_2 - Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \\ &= 86.5 - 79.6 - 1.645\sqrt{\frac{(1.1)^2}{15}+\frac{(1.4)^2}{18}} = (6.9 - 0.716) \\ &= 6.184\ \text{kg/mm}^2. \end{aligned}$$

Also

$$\begin{aligned} u &= \bar{X}_1 - \bar{X}_2 + Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \\ &= 86.5 - 79.6 + 1.645\sqrt{\frac{(1.1)^2}{15}+\frac{(1.4)^2}{18}} = (6.9 + 0.716) \\ &= 7.616\ \text{kg/mm}^2. \end{aligned}$$

13.7 Confidence Intervals for Proportions

In engineering applications, the outgoing quality of a product is often estimated based on testing a sample of the parts. If $\hat{p}$ is the proportion of observations in a random sample of size n that belongs to a class of interest (e.g., defects), then an approximate $100(1 - \alpha)$ percent confidence interval on the proportion, p, of the population that belongs to this class is:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \tag{13.9}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentage point of a standard normal distribution. This relationship holds true when the proportion is not too close to either 0 or 1 and the sample size n is large.

Example 13.4

An inspector randomly selects 200 boards from the process line and finds 5 defective boards. Calculate the 90% confidence interval for the proportion of good boards from the process line.

Solution:
Use Equation 13.9:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$\frac{195}{200} - 1.64\sqrt{\frac{0.975(0.025)}{200}} \leq p \leq \frac{195}{200} + 1.64\sqrt{\frac{0.975(0.025)}{200}}$$
$$0.957 \leq p \leq 0.993.$$

The result implies that the total population is likely (90% probability) to have a proportion of good boards between 0.997 and 0.993. Note that no assumption is made regarding what the total population is.

13.8 Reliability Estimation and Confidence Limits for Success–Failure Testing

Success–failure testing describes a situation where a product (component, subsystem) is subjected to a test for a specified length of time, T_0 (or cycles, stress reversals, miles, etc.). The product either survives to time T_0 (i.e., it survives the test) or fails prior to time T_0.

Testing of this type can frequently be found in engineering laboratories where a test "bogy" has been established and new designs are tested against this bogy. The bogy will specify a set number of cycles in a certain test environment and at predetermined stress levels.

The probability model for this testing situation is the following binomial distribution, which gives the probability that the number of successes is y out of n items tested:

$$P(y) = \binom{n}{y} R^y (1-R)^{n-y}, \quad y = 0, 1, \ldots, n, \tag{13.10}$$

where

n = the number of items tested
R = the probability of surviving the test for the product
y = the number of survivors out of n,

and

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}, \quad y = 0, 1, \ldots, n. \tag{13.11}$$

The value R is the reliability, which is the probability of surviving the test. The minimum variance unbiased estimator of R is

$$\hat{R} = \frac{y}{n}. \tag{13.12}$$

The 100(1 − α) percent lower confidence limit on the reliability R is calculated by

$$R_L = \frac{y}{y + (n - y + 1) F_{\alpha, 2(n-y+1), 2y}}, \tag{13.13}$$

where $F_{\alpha,2(n-y+1),2y}$ is obtained from the F tables. Here again, n is the number of items tested and y is the number of survivors.

The 100(1 − α) percent upper confidence limit on R is given by

$$R_U = \frac{(y+1) \times F_{\alpha,2(y+1),2(n-y)}}{(n-y) + (y+1) F_{\alpha,2(y+1),2(n-y)}}. \tag{13.14}$$

The F tables that are usually available are somewhat limited in terms of degrees of freedom. Therefore, it is convenient to have an approximation for the lower confidence limit that uses the standard normal distribution. The lower confidence on reliability can be approximated by:

$$R_L = \frac{y - 1}{n + Z_\alpha \sqrt{\dfrac{n \times (n - y + 1)}{(y - 2)}}}, \tag{13.15}$$

where

Z_α = the standard normal variable, as given in Table 13.6

y = the number of successes

n = the sample size.

It should be noted that Z is the standard normal variable. Values given in Table 13.6 can be read from cumulative distribution tables for standard normal variables given in Appendix C.

Table 13.6 Standard normal variables

Confidence level (1 − α)	Z_α
95	1.645
90	1.281
80	0.841
75	0.678
70	0.525
50	0.000

Example 13.5

A weapon system has completed a test schedule. The test is equivalent to 60 missions. Dividing the test schedule up into 60 missions results in seven failed missions. Estimate the mission reliability.

Solution:
In this case, the number of successes (y) is $y = 60 - 7 = 53$ successful missions out of $n = 60$ missions. Then the point estimate for mission reliability is

$$\hat{R}_m = \frac{53}{60} = 0.883.$$

Let us now find a 75% lower confidence limit. The exact lower 75% limit is found by using an F value of

$$\mathrm{F}_{0.25,16,106} = 1.24.$$

Substituting this into the confidence limit equation gives

$$R_L = \frac{53}{53 + (8 \times 1.24)} = 0.842.$$

The 75% lower confidence limit on mission reliability is $0.842 \leq R_m$. If the normal approximation was used, the lower limit's value would be

$$R_L = \frac{52}{60 + 0.675 \times \sqrt{\dfrac{60 \times (60 - 53 + 1)}{51}}} = 0.838.$$

As can be seen, this approximation provides limits that are reasonably close to the exact values.

Example 13.6

Gas turbine engines are subjected to a 10-hour burn-in test after assembly. Out of 30 engines produced in a month, one engine failed to pass the test.

(a) Find a 95% lower confidence limit on engine reliability relative to this test using the exact equations with the F-distribution.

Solution:

$$\alpha = 0.05, n = 30, y = 29$$

$$R_L = \frac{y}{y + (n - y + 1) F_{\alpha,2(n-y+1),2y}}$$

$$= \frac{29}{29 + 2 \times F_{0.05,4,58}} = \frac{29}{29 + 2 \times 2.538} = 0.851039.$$

(b) Find a 95% upper confidence limit on engine reliability using the above test results.

Solution:

$$\alpha = 0.05, \quad n = 30, \quad y = 29$$

$$R_U = \frac{(y+1)\times F_{\alpha,2(y+1),2(n-y)}}{(n-y)+(y+1)F_{\alpha,2(y+1),2(n-y)}}$$

$$= \frac{30F_{0.05,60,2}}{1+30\times F_{0.05,60,2}} = \frac{30\times 19.48}{1+30\times 19.48} = 0.998292.$$

Example 13.7

The customer wants to demonstrate a reliability of 98% relative to this test with 96% confidence using success testing. What sample size should the test engineer use (with no failures) to demonstrate the customer's reliability requirements?

Solution:

$$n = \frac{\ln(1-C)}{\ln R} = \frac{\ln 0.04}{\ln 0.98} = 159.3289 \approx 160.$$

13.8.1 Success Testing

Sometimes in receiving inspection and engineering test labs a no-failure ($r = 0$ or $y = n$) test is specified. The goal is usually to ensure that a reliability level has been achieved at a specified confidence level. A special adaptation of the confidence limit formula can be derived for this situation. For the special case where $r = 0$ (i.e., no failures), the lower 100 $(1 - \alpha)$ percent confidence limit on the reliability is given by:

$$R_L = a^{1/n} = (1-C)^{1/n}. \tag{13.16}$$

where α is the level of significance and n is the sample size (i.e. number of units placed on test).

If $C = (1 - \alpha)$, the desired confidence level (0.80, 0.90, etc.), then the necessary sample size to demonstrate a desired lower limit on reliability level, R_L, is

$$n = \frac{\ln(1-C)}{\ln R_L}. \tag{13.17}$$

For example, if $R_L = 0.80$ is to be demonstrated with a 90 percent confidence level,

$$n = \frac{\ln(0.10)}{\ln(0.80)} = 11. \tag{13.18}$$

Thus, 11 items must be tested, with no failures. This is frequently referred to as success testing.

13.9 Reliability Estimation and Confidence Limits for Exponential Distribution

Two types of tests are typically considered:

1. *Type 1 Censored Test.* The items are tested for a specified time, T, and then the testing is stopped.
2. *Type 2 Censored Test.* The test time is not specified, but the testing is stopped when a desired number of items fail.

Let us consider the situation when n items are being tested and the test is stopped as soon as r failures are observed ($r \leq n$). This is type 2 censoring, with nonreplacement of items. Let the observed failure times be, in order of magnitude,

$$0 = t_0 =< t_1 < t_2 < \cdots < t_{r-1} < t_r. \tag{13.19}$$

Then, making the transformation,

$$u_i = \begin{cases} nt_1, & \text{when } i = 0 \\ (n-i)(t_{i-1} - t_i) & \text{when } i = 1, 2, \ldots, r-1. \end{cases} \tag{13.20}$$

The (u_i, i = 0, . . . , r − 1) are independently and identically distributed with the common density function,

$$\left(\frac{1}{\theta}\right)e^{-u/\theta}. \tag{13.21}$$

The total time on test is given by

$$\begin{aligned} V(t_r) &= \text{total time on test} \\ &= \sum_{i=0}^{r-1} u_i \\ &= \sum_{i=0}^{r-1} t_i + (n-r)t_r. \end{aligned} \tag{13.22}$$

Then

$$\hat{\theta} = \frac{V(t_r)}{r} = \frac{1}{r}\left[\sum_{i=1}^{r} t_i + (n-r)t_r\right], \tag{13.23}$$

is the minimum variance unbiased estimator of θ. Since

$$V(t_r) = \sum_{1=0}^{r-1} u_i, \tag{13.24}$$

and the $\{u_i\}$ are independently distributed with a common exponential density function, it follows that $V(t_r)$ has a gamma distribution with parameters (θ, r). Hence,

$$2V(t_r)/\theta = 2\hat{\theta}r/\theta, \tag{13.25}$$

is distributed as χ^2_{2r}.

The $100(1-\alpha)\%$ confidence limits on θ are given by:

$$P\left[\chi^2_{1-(\alpha/2),2r} \leq \frac{2\hat{\theta}r}{\theta} < \chi^2_{\alpha/2,2r}\right] = 1-\alpha \tag{13.26}$$

or

$$\frac{2\hat{\theta}r}{\chi^2_{\alpha/2,2r}} \leq \theta \leq \frac{2\hat{\theta}r}{\chi^2_{1-(\alpha/2),2r}}. \tag{13.27}$$

Life testing procedures are often used in a quality control context to detect the deviations of θ below some desired levels, such as θ_0. For a significance level of α, the probability of accepting H_0 is

$$P_\alpha = P\left(\frac{2r\hat{\theta}}{\theta_0} \leq \chi^2_{\alpha,2r} \middle| \theta = \theta_0\right) = 1-\alpha. \tag{13.28}$$

The expected time to complete the test is given by

$$E(t_r) = \theta \sum_{i=1}^{r} \frac{1}{n-i+1}. \tag{13.29}$$

Let

$\theta_0 =$ desired reliability goal for mean time between failures (MTBF)
$1-\alpha =$ probability of accepting items with a true MTBF of θ_0
$\theta_1 =$ alternative MTBF ($\theta_1 < \theta_0$)
$\beta =$ probability of accepting items with a true MTBF of θ_1.

With this information, reliability testing consists of putting n items on test and stopping the test when the number of failures is given by the smallest integer satisfying:

$$\frac{2T}{\chi^2_{\alpha/2,2(r+1)}} \leq \theta \leq \frac{2T}{\chi^2_{1-(\alpha/2),2r}}. \tag{13.30}$$

Thus, when we know θ_0, θ_1, α, and β, we can compute the necessary value for r.

For the *type 1* censored test, where r failures are observed on an interval of total test time, $V(t_r) = T$, the $100(1 - \alpha)$ percent confidence limits on θ are given by a modification of Equation 13.30:

$$\frac{2T}{\chi^2_{\alpha/2,2(r+1)}} \leq \theta \leq \frac{2T}{\chi^2_{1-(\alpha/2),2r}}. \tag{13.31}$$

Example 13.8

Sixteen thousand device-hours (total time on test) are accumulated in a failure-terminated test, with four failures.

(a) What are the upper and lower one-sided 90% confidence limits on MTBF?

(b) What are the one-sided 90% confidence limits on reliability for a 100-hour period?

Solution:
For this problem, we have

$$T = 16{,}000 \text{ hours}$$
$$C = 1-\alpha = 0.90; \quad \alpha = 0.10; \quad \alpha/2 = 0.05; \quad 1-\alpha/2 = 0.95$$
$$r = 4.$$

Therefore,

$$\text{MTBF}(l) = \frac{2(16{,}000)}{\chi^2_{0.10;8}} = \frac{32{,}000}{13.362} = 2395 \text{ hours}$$
$$\text{MTBF}(u) = \frac{2(16{,}000)}{\chi^2_{0.90;8}} = \frac{32{,}000}{3.490} = 9195 \text{ hours.}$$

If the lower and upper 0.90 confidence limits on the MTBF for the item are 2395 and 9195 hours, the lower and upper 0.90 confidence limits on its reliability for any 100-hour interval are:

$$R(l) = e^{\frac{-100}{2{,}395}} = e^{-0.0417} = 0.9591$$
$$R(u) = e^{\frac{-100}{9{,}195}} = e^{-0.0109} = 0.9891.$$

Example 13.9

Twenty-one thousand device-hours (total time on test) are accumulated in a time-terminated test, with seven failures. What are the upper and lower one-sided limits on MTBF with 0.99 confidence?

Solution:
Here

$$T = 21{,}000 \text{ hours}$$
$$C = 1-\alpha = 0.99; \quad \alpha = 0.01; \quad \alpha/2 = 0.005; \quad 1-\alpha/2 = 0.995$$
$$r = 7.$$

Therefore,

$$\text{MTBF}(l) = \frac{2(21{,}000)}{\chi^2_{0.01;16}} = \frac{42{,}000}{32.000} = 1313 \text{ hours}$$
$$\text{MTBF}(u) = \frac{2(21{,}000)}{\chi^2_{0.99;14}} = \frac{42{,}000}{4.660} = 9013 \text{ hours.}$$

Example 13.10

Ten automotive air conditioning switches were cycled and observed for failure. Testing was suspended when the fourth failure occurred. Failed switches were not replaced. The failures occurred at the following cycles: 8900, 11,500, 19,200, and 29,300.

The assumption for this problem is that time to failure for the switch follows an exponential distribution with parameter θ.

Solution:

(a) Find the point estimator for the mean life (θ) of the switches.

This is a failure-truncated test. Hence the point estimator for θ is found by using Equation 13.23 and is

$$\hat{\theta} = \frac{8900 + 11{,}500 + 19{,}200 + 29{,}300 + (10-4)29{,}300}{4}$$
$$= \frac{244{,}700}{4} = 61{,}175.$$

(b) Find the 90% two-sided confidence limits on θ.

$$\chi^2_{0.05,8} = 15.507 \quad \chi^2_{0.95,8} = 2.733$$
$$\frac{2 \times 244{,}700}{15.507} \le \theta \le \frac{2 \times 244{,}700}{2.733}$$
$$31{,}559.9 \le \theta \le 179{,}071.$$

(c) The warranty for these switches is for 3000 cycles. Find the 95% one-sided upper confidence limit on the percent failures during the warranty period.

$$R_L(3000) = e^{-\frac{3000}{31{,}559.9}} = 0.9093$$

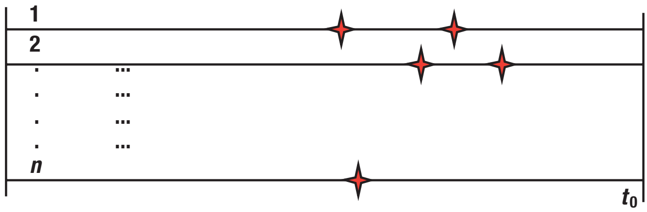

Figure 13.27 Time-truncated testing. Failure points are denoted by ✦.

Table 13.7 Values of parameter γ and dF for confidence limit calculations on MTBF

Type of Test	MTBF (*l*) γ	MTBF (*l*) dF	MTBF (*u*) γ	MTBF (*u*) dF
Two-sided failure terminated	$\alpha/2$	$2r$	$1-\alpha/2$	$2r$
One-sided failure terminated	α	$2r$	$1-\alpha$	$2r$
Two-sided time terminated	$\alpha/2$	$2r+2$	$1-\alpha/2$	$2r$
One-sided time terminated	α	$2r+2$	$1-\alpha$	$2r$
No failures observed	α	2	–	–

Note: *r* is the number of failures observed.

or

$$F_U(3000) = 0.0907.$$

We are 95% confident that for 3000 cycles, the percent of failures is less than 9.07%.

Figure 13.27 illustrates a time-truncated situation in which there are n test stands. Items are replaced by new items on the test stand when they fail and testing is stopped on every test stand at time t_0. Thus the total time on test is nt_0.

The confidence limits for MTBF assuming an exponential distribution can be summarized by:

$$\mathrm{MTBF} = \frac{2T}{\chi^2_{r;dF}}, \tag{13.32}$$

where T is the total time on test, and the values for the parameter γ and dF (degrees of freedom) for the χ^2 distribution can be obtained for different testing conditions from Table 13.7.

A common situation occurs when an estimate of the MTBF and the confidence interval around it is of interest, but no failures have occurred. You can still calculate a lower one-sided confidence limit, which is a conservative value for MTBF. Of course, there is no upper confidence limit. The lower confidence limit on MTBF is given by

$$\frac{2T}{\chi^2_{\alpha,2}} \leq \theta. \tag{13.33}$$

13.10 Summary

The purpose of reliability estimation, demonstration, and testing is to determine whether a product has met certain reliability requirements with a stated statistical confidence level. Various tests are done throughout the life cycle of the product and are discussed. A five-step physics of failure approach was presented for the qualification of a SiP-type package in a drop-loading environment. Basic statistical concepts for estimation and confidence intervals are covered. Confidence intervals for both normal and binomial distributions are presented with examples. Finally, reliability estimation and confidence limits when the time to failure follows exponential distribution are discussed.

Problems

13.1 To get a 95% confidence interval on mean thermal conductivity, with an error less than 0.10 Btu/hr-ft-°F, what is the desired sample size? Assume $\sigma = 0.30$ Btu/hr-ft-°F at 100°F and 550 W.

13.2 An inspector found 10 defective keyboards from a sample of 300. Calculate the 95% confidence interval for the proportion of good units. What would be the 95% confidence interval for the proportion of bad units?

13.3 Gas turbine engines are subjected to a 10-hour burn-in test after assembly. Out of 40 engines produced in a month, three engines failed to pass the test. Develop a 95% two-sided symmetrical (both lower and upper) confidence limits on the engine reliability relative to this test using the F distribution.

13.4 The following data represent kilometers to failure for a set of vehicles:

43,000	27,200	10,600	12,400
27,000	4,100	200,000	18,200
68,000	40,500	109,000	14,200
46,000	2600	2400	24,500

(a) Estimate the MTBF.

(b) Set a 90% lower confidence limit on the 10% failure kilometer or B10 life.

(c) With 90% confidence, find the 2400 km lower limit on reliability.

13.5 For a test vehicle, major electrical failures occurred at the following kilometers:

63	17,393	23,128
114	18,707	24,145
14,820	19,179	33,832
16,105	22,642	34,345

The vehicle was driven a total of 36,000 kilometers.

(a) Estimate the MTBF.
(b) Determine the 90 percent two-sided confidence interval for the MTBF.
(c) Estimate the reliability function.
(d) Determine the 95 percent lower confidence limit for the 1,200 kilometer reliability.
(e) With 90 percent confidence estimate the kilometer at which 10 percent of the population will fail.

13.6 Twelve disk drives for computers were cycled and observed for failure. Testing was suspended when the third failure occurred. Failed disk drives were not replaced. The failures occurred at the following hours: 791; 909; 1522. The assumption for this problem is that time to failure for the disk drives follows an exponential distribution with parameter θ.

(a) Find the point estimate for the mean life θ of the disk drives.
(b) Find the 80% two-sided confidence limits on θ.
(c) The warranty for these disk drives is for 5000 hours. Find the 90% one-sided upper confidence limit on the percent failure during the warranty period.

13.7 What is accelerated testing? What is the purpose of doing accelerated testing? Explain with examples.

13.8 What is a qualification test? Can qualification tests reduce the over-stress failure of products?

13.9 What is HALT? Can HALT results be used to predict product reliability? Explain.

13.10 Explain with examples the steps in determining qualification testing conditions.

13.11 Describe how accelerated testing conditions and the accelerating factor are determined.

13.12 Discuss the different test data that can be used to assess reliability of parts. Which of these types of data is most appropriate for making reliability assessment?