5 Product Development

To ensure acceptable product reliability, an organization must follow certain practices during the product development process. These practices impact reliability through the selection of parts (materials), product design, manufacturing, assembly, shipping and handling, operation, maintenance, and repair. Best practices for reliability, listed below and described in this book, dictate that the organization should:

- Define realistic product reliability requirements determined by factors including the targeted life cycle application conditions and performance expectations. The product requirements should consider the customer's needs and the manufacturer's capability to meet those needs.
- Define the product life-cycle conditions by assessing relevant manufacturing, assembly, storage, handling, shipping, operating, and maintenance conditions.
- Ensure that the supply-chain participants have the capability to produce the parts (materials) and services necessary to meet the final reliability objectives.
- Select parts (materials) that have sufficient quality and are capable of delivering the expected performance and reliability in the application.
- Identify the potential failure modes, failure sites, and failure mechanisms by which the product can be expected to fail.
- Design to the process capability (i.e., the quality level that can be controlled in manufacturing and assembly), considering the potential failure modes, failure sites, and failure mechanisms, obtained from the physics-of-failure analysis, and the life-cycle profile.
- Qualify the product to verify the reliability of the product in the expected life-cycle conditions. Qualification encompasses all activities that ensure that the nominal design and manufacturing specifications will meet or exceed the reliability goals.

Reliability Engineering, First Edition. Kailash C. Kapur and Michael Pecht.
© 2014 John Wiley & Sons, Inc. Published 2014 by John Wiley & Sons, Inc.

- Ensure that all manufacturing and assembly processes are capable of producing the product within the statistical process window required by the design. Variability in material properties and manufacturing processes will impact the product's reliability, so characteristics of the process must be identified, measured, and monitored.
- Manage the life-cycle usage of the product using closed-loop, root-cause monitoring procedures.

5.1 Product Requirements and Constraints

Various reasons justify the creation, modification, or upgrade of a product. For example, as discussed in Chapter 4, a company may want to address a perceived market need or open new markets. In some cases, a company may need to develop new products to remain competitive in a key market or to maintain market share and customer confidence. In other cases, a company may want to satisfy specific strategic customers, demonstrate experience with a new technology or methodology, or improve the maintainability of an existing product. In addition, product updates are often developed to reduce the life-cycle costs of an existing product.

To make reliable products, suppliers and customers throughout the supply chain must cooperate. The IEEE 1332 (IEEE Std. 1332–1998) addresses this cooperation through the three reliability objectives discussed in the previous chapter. First, the supplier must understand the customer's requirements and product needs in order to generate a comprehensive design specification. Second, the supplier must employ appropriate engineering activities so that the resulting product satisfies the customer's reliability requirements. Finally, the supplier must assure the customer that the reliability requirements and product needs have been satisfied.

Initially, requirements are formulated into a requirements document, where they are prioritized. The specific people involved in prioritization and approval will vary with the organization and the product. For example, for safety-critical products, safety, reliability, and legal representatives may all provide guidance.

As we have noted, once a set of requirements has been completed, the product engineering function creates a response to the requirements in the form of a specification. The specification states the requirements that must be met; the schedule for meeting the requirements; the identification of those who will perform the work; and the identification of potential risks. Differences in the requirements document and the preliminary specification become the topic of trade-off analyses.

After product requirements are defined and the design process begins, there should be an assessment of the product's requirements against the actual product design. As the product's design becomes increasingly detailed, it becomes more important to track the product's characteristics in relation to the original requirements. The rationale for making changes should be documented. The completeness with which requirement tracking is performed can significantly reduce future product redesign costs. Planned redesigns or design refreshes through technology monitoring and use of roadmaps ensure that the company is able to market new products or redesigned versions of old products in a timely, effective manner to retain its customer base and ensure continued profits.

5.2 Product Life Cycle Conditions

The life cycle conditions of the product influence decisions regarding product design and development, materials and parts selection, qualification, product safety, warranty, and product support (maintenance). The phases in a product's life cycle include manufacturing and assembly, testing, rework, storage, transportation and handling, operation[1] (modes of operation, on-off cycles, etc.), and repair and maintenance.

During each phase of its life cycle, a product will experience various environmental and usage loads. These loads may be thermal (steady-state temperature, temperature ranges, temperature cycles, and temperature gradients); mechanical (pressure levels, pressure gradients, vibrations, shock loads, and acoustic levels); chemical (aggressive or inert environments, ozone, pollution humidity levels, contamination, and fuel spills); environmental (radiation, electromagnetic interference, and altitude); electrical loading conditions (power, power surge, current, voltage, and voltage spikes); or the extent and rate of product degradation, among others. Reliability depends upon the nature, magnitude, and duration of exposure to such loads.

Defining and characterizing life-cycle loads is often an uncertain element of the overall design-for-reliability process. The challenge occurs because products can experience completely different application conditions depending on the application location, the product utilization or nonutilization profile, the duration of utilization, and maintenance and servicing conditions. For example, typically all desktop computers are designed for home or office environments. However, the operational profile of each unit may be completely different depending on user behavior. Some users may shut down the computer after it is used each time; others may shut down only once at the end of the day; still others may keep their computers powered all the time. Furthermore, one user may keep the computer by a sunny window, while another may keep the computer near an air conditioner; thus, the temperature profile experienced by each product, and hence its degradation due to thermal loads, would be different.

Four methods are used to estimate product life-cycle loads: market surveys and standards, similarity analysis, field trial and service records, and in situ monitoring. Market surveys and standards provide a very coarse and often inaccurate estimate of the environmental loads possible in various field applications. The environmental profiles available from these sources are typically classified according to industry type, such as military, consumer, telecommunications, automotive, and commercial avionics.

Similarity analysis is a technique for estimating environmental loads when sufficient field histories for similar products are available. Before using data on existing products for proposed designs, the characteristic differences in design and application use for the comparison products need to be reviewed. For example, electronics inside a washing machine in a commercial laundry are expected to experience a wider distribution of loads and use conditions (due to a larger number of users) and higher usage rates than a home washing machine. As another example, it has been found that some Asians use a dishwasher to wash vegetables, in addition to eating utensils. These dishwashers experience higher usage rates than those used only for washing dishes.

[1] Operational conditions are sometimes referred to as the life-cycle application conditions.

Field trial records provide estimates of the environmental profiles experienced by the product. The data depend on the durations and conditions of the trials, and can be extrapolated to estimate actual environmental conditions. Service records provide information on the maintenance, replacement, or servicing performed. These data can give an idea of the life-cycle environmental and usage conditions that lead to servicing or failure.

Environmental and usage conditions experienced by the product over its life cycle can be monitored in situ (Vichare et al. 2004). These data are often collected using sensors, either mounted externally or integrated with the product and supported by telemetry systems. Load distributions should be developed from data obtained by monitoring products used by different customers, ideally from various geographical locations where the product is used. The data should be collected over a sufficient period to provide an estimate of the loads and their variation over time. In situ monitoring provides the most accurate account of load histories and is most valuable in design-for-reliability (DFR) and product reliability assessment.

5.3 Reliability Capability

The selection of a supply chain is often based on factors that do not explicitly address reliability, such as technical capabilities, production capacity, geographic location, support facilities, and financial and contractual factors. A selection process that takes into account the ability of suppliers to meet reliability objectives during manufacturing, testing, and support can improve the reliability of the final product throughout its life cycle and provide valuable competitive advantages.

Reliability capability is a measure of the practices within an organization that contribute to the reliability of the final product and the effectiveness of these practices in meeting the reliability requirements of customers. Reliability capability assessment is the act of quantifying the effectiveness of reliability activities, using a metric called reliability capability maturity. From a reliability perspective, maturity indicates whether the key reliability practices employed by an organization are well understood, supported by documentation and training, applied to all products throughout the organization, and continually monitored and improved.

5.4 Parts and Materials Selection

A parts and materials selection and management methodology helps a company to make risk-informed decisions concerning the incorporation of parts and materials into a product. The part assessment process is shown in Figure 5.1. Key elements of part assessment include performance, quality, reliability, and ease of assembly.

The goal of performance assessment is to evaluate the part's ability to meet the performance requirements (structural, mechanical, electrical, thermal, biological, etc.) of the product. In general, there is often a minimum and a maximum limit beyond which the part will not function properly, at least in terms of the datasheet specifications. These limits, or ratings, are often called the recommended operating conditions.

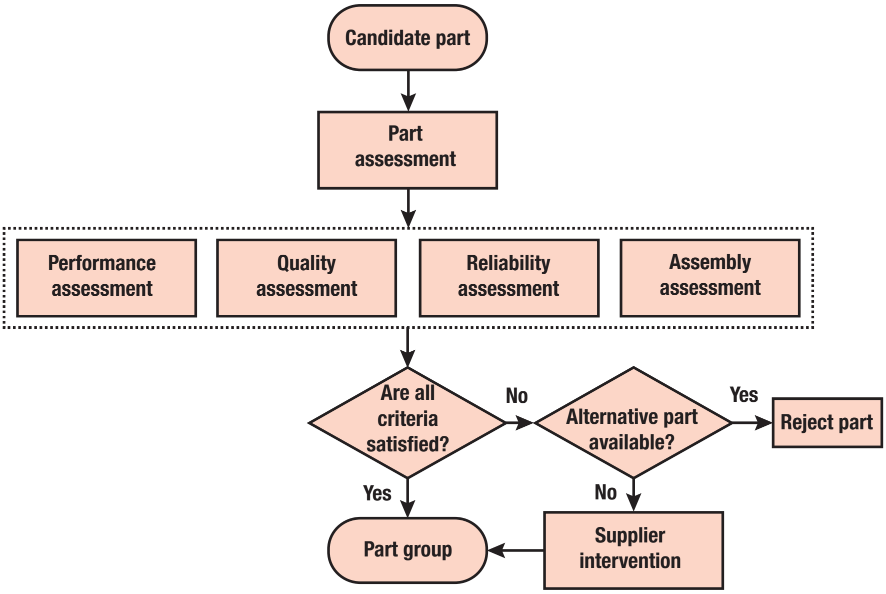


Figure 5.1 Part assessment process.

Quality is evaluated by outgoing quality and process capability metrics. Reliability assessment results provide information about the ability of a part to meet the required performance specifications in its targeted life-cycle application for a specified period of time. Reliability is evaluated through part qualification and reliability test results.

A part is acceptable from an assembly viewpoint if it is compatible with the downstream assembly equipment and processes. Assembly guidelines should be followed to prevent damage and deterioration of the part during the assembly process. Examples include a recommended temperature profile, cleaning agents, adhesives, moisture sensitivity, and electrical protection. As new technologies emerge and products become more complex, assembly guidelines become more necessary to ensuring the targeted quality and reliability of the parts and the product.

5.5 Human Factors and Reliability

All systems are of, by, and for humans. Human factors therefore are critical in the system design process and must be weighed against safety, reliability, maintainability, and other system parameters in order to affect trade-offs that increase system effectiveness. Human interaction with a system includes:

- Design and production of systems
- Operators and repairers of systems
- Operators and repairers as decision elements.

The human machine interface consists of such aspects as allocation of functions (human vs. machine), automation, accessibility, human tasks, stress characteristics, and both the information presented to the operator or repairer and the reliability of interfaces and decisions based on such information. Both human and machine elements of a system can fail, and their failures have varying effects on the system's performance. Some human errors cause total system failure or increase the risk of such failure. Human factors exert a strong influence on the design and ultimate reliability of a system (Kirwan 1994).

Both reliability and human factors are concerned with predicting, measuring, and improving system effectiveness. When the human machine interface is complex, the possibility of human error increases, resulting in an increase in the probability of system failure. An interesting facet of relationship among human factors, reliability, and maintainability is that the system's reliability and maintainability depends on the detection and correction of system malfunctions. This task is generally performed by people. Thus, the system performance can be enhanced or degraded depending on the human response. The quantification of human reliability characteristics and the development of a methodology for quantifying human performance, error prediction, control, and measurement are given in many sources (Gertman and Blackman 1994; Meister 1996).

The reliability of a system is affected by the allocation of system functions to humans, machines, or both. Favorable human characteristics include the ability to:

1. Detect certain forms of energy.
2. Be sensitive to a wide variety of stimuli within a restricted range.
3. Detect signals and patterns in high noise environments.
4. Store large amounts of information for long periods and remember relevant facts.
5. Learn from experience.
6. Use judgment.
7. Improvise and adopt flexible procedures.
8. Arrive at new and completely different solutions to problems;
9. Handle low probability or unexpected events.
10. Perform fine manipulations.
11. Reason instinctively.

Characteristics tending to favor machines are:

1. Computing capacity
2. Performance of routine, repetitive, and precise tasks
3. Quick response to control signals
4. Ability to exert large amounts of force smoothly and precisely
5. Ability to store and recall large amounts of data
6. Ability to reason deductively
7. Insensitivity to extraneous factors
8. Ability to handle highly complex operations that involve doing several things at once.

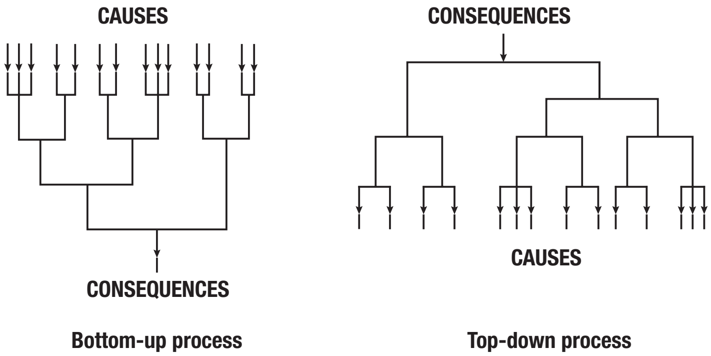


Figure 5.2 Bottom-up versus top-down methods.

5.6 Deductive versus Inductive Methods

Deduction comprises reasoning from the general to the specific. In a deductive system analysis, it is postulated that the system itself has failed in a certain way, and an attempt is made to find out what modes of system or subsystem (component) behavior contribute to this failure. These methods are also called top-down. One of the very popular and useful deductive methods is fault tree analysis (FTA), which is covered in Section 5.8.

Induction involves reasoning from individual cases to a general conclusion. In this case, a particular fault or initiating condition is postulated and an attempt to ascertain the effect of that fault or condition on system operation is made. These methods are also called bottom-up. The reliability block diagram (RBD) is one example of an inductive method that is covered in Chapter 17. Another very popular and useful method is failure modes, effects and criticality analysis (FMECA), which is discussed in the next section. Figure 5.2 shows the difference between backward versus forward methods. The arrows indicate the direction of these tree-like graphs.

In general, both deductive and inductive approaches must be employed to get a complete set of failure/fault/accident sequences. The deductive approach has the benefit of focusing the analysis on the undesired event, while the inductive approach is useful in assuring that the analysis is broad enough to encompass all possible scenarios.

5.7 Failure Modes, Effects, and Criticality Analysis

Failure modes, effects, and criticality analysis (FMECA) is a design evaluation procedure used to identify all conceivable and potential failure modes and to determine the effect of each failure mode on system performance. Criticality analysis in FMECA helps to develop priorities for continuous improvement. This procedure is accomplished by formal documentation, which serves (1) to standardize the procedure, (2) as a means of historical documentation, and (3) as a basis for future improvement.

Correct usage of the FMECA process will result in two improvements:

1. An improvement in the reliability of the product through the anticipation of problems and the institution of corrections prior to going into production.
2. An improvement in the validity of the analytical method itself through strict documentation that illuminates the rationale for every step.

Failure modes and effects analysis is an iterative, systematic, documented process performed to identify basic failure/faults at the part level and determine their effects at higher levels of assembly. Criticality analysis in FMECA helps to develop priorities for continuous improvement. The analysis can be performed utilizing either actual failure modes from field data or hypothesized failure modes derived from design analysis, reliability prediction activities, and experience with how parts fail.

In their most complete form, failure modes are identified at the part level, which is usually the lowest level of direct concern to the designer of the product or process. In addition to providing insight into failure cause-and-effect relationships, FMECA provides a disciplined method for proceeding part by part through the system to assess failure consequences.

Failure modes are analytically induced into each component, and failure effects are evaluated and noted, including the severity and frequency (or probability) of occurrence. As the first mode is listed, the corresponding effect on performance at the next higher level of assembly is determined. The resulting failure effect becomes, in essence, the failure mode that impacts the next higher level. Iteration of this process results in establishing the ultimate effects at the system level.

The analysis of all failure modes usually reveals that each effect or symptom at the system level is caused by several different failure modes at the lowest level. This relationship to the end effect provides the basis for grouping the lower-level failure modes.

Using this approach, probabilities for the occurrence of system failure can be calculated, based on the probability of occurrence of the lower-level failure modes. Based on these probabilities and a severity factor assigned to the various system effects, a criticality number can be calculated. Criticality numerics also provide the basis for corrective action priorities, engineering changes, and resolution of problems in the field.

The procedure consists of a sequence of logical steps, starting with the analysis of lower level subsystems or components. The analysis assumes a failure point of view and identifies all potential modes of failure, along with the causative agent, termed the "failure mechanism." The effect of each failure mode is then traced up to the systems level (MIL_STD_1629 (SHIPS)).

As mentioned before, a criticality rating is developed for each failure mode and its resulting effect. The rating is based on the probability of occurrence, severity, and detectability. For failures scoring a high rating, design changes to reduce criticality are recommended. This procedure is aimed at providing a more reliable design.

A failure mode is the manner in which a failure can occur—that is, the way in which the products fails to perform its intended design function, or performs the function but fails to meet its objectives. For example, failure modes of a cell phone include a button that doesn't cause a number to register, or a microphone that doesn't pick up your voice.

Sometimes, the failure modes are intentionally accentuated so that the user of the product will become aware of the existence of a problem. For example, a bad-smelling substance is sometimes added to natural gas to indicate the existence of a leak. Another example is the grinding noise when the brake pads wear out on a car.

Failure mechanisms are the processes by which a specific combination of physical, electrical, chemical, and mechanical stresses induces failures. For example, fracture, fatigue, and corrosion are failure mechanisms.

The purpose of failure modes, mechanisms, and effects analysis (FMMEA) is to identify potential failure mechanisms and models for all the potential failures modes of a product, and then to prioritize failure mechanisms for efficient product development. FMMEA is based on understanding (1) the relationships between product requirements and the physical characteristics of the product (and their variation in the production process), and (2) the interactions of product materials with loads (stresses under application conditions) and their influence on the product's susceptibility to failure with respect to the use conditions.

5.8 Fault Tree Analysis

FTA is a method for system safety and reliability analysis (*Fault Tree Handbook* 2002). The concept was originated by Bell Telephone Laboratories as a technique to evaluate the safety of the Minuteman Launch Control System. Many reliability techniques are inductive and are concerned primarily with ensuring that hardware will accomplish its intended functions. FTA is a detailed deductive analysis that usually requires considerable information about the system. Concerned with ensuring that all critical aspects of a system are identified and controlled, it is a graphical representation of the Boolean logic associated with the development of a particular system failure (consequence), called the "top event," into basic failures (causes), and called "primary events." These top events can be broad, all encompassing events, such as "the release of radioactivity from a nuclear power plant" or "the inadvertent launch of an ICBM missile," or they can be specific events, such as "failure to insert control rods" or "energizing power available to ordnance ignition line."

FTA is of value for:

- Providing options for qualitative and quantitative reliability analysis
- Helping the analyst to understand system failures deductively
- Pointing out the aspects of a system that are important with Respect to the failure of interest
- Providing the analyst an insight into system behavior.

A fault tree is a model that graphically and logically represents the various combinations of possible events, both fault and normal, that occur in a system and lead to the top event. The term "event" denotes a dynamic change of state that occurs in a system element. A fault event is an abnormal system state. A normal event is an event that is expected to occur. System elements include hardware, software, and human and environmental factors. (Details about the construction of fault trees can be found in the reference mentioned at the beginning of this section.)

FTA is a deductive methodology to determine the potential causes of failures and to estimate the failure probabilities. FTA addresses system design aspects and potential failures, tracks down system failures deductively, describes system functions and behaviors graphically, focuses on one error at a time, and provides qualitative and quantitative reliability analyses. The purpose of a fault tree is to show the sets of events—particularly the primary failures—that will cause the top event in a system.

FTA provides critical information that can be used to prioritize the importance of the contributors to the undesired event. The contributing importances provided by FTA vividly show the causes that are dominant and that should be the focus of any safety or reliability activity.

More formal risk–benefit approaches can also be used to optimally allocate resources to minimize both resource expenditures and the probability of occurrence of the undesired event. These risk–benefit approaches are useful for allocating resource expenditures, such as safety upgrades to complex systems like the Space Shuttle.

FTA can be applied to both an existing system and a system that is being designed. When it is applied to a system being designed for which specific data do not exist, FTA can provide an estimate of the failure probability and the important contributors to failure, using generic data to bracket the design components or concepts. FTA can also be used as an important element in the development of a performance-based design.

When applied to an existing system, FTA can be used to identify weaknesses and to evaluate possible upgrades. It can also be used to monitor and predict behavior. Furthermore, FTA can be used to diagnose causes and potential corrective measures for an observed system failure.

The approaches and tools to obtain this information and apply it in decision-making are important topics. FTA can be simply described as an analytical technique, through which (1) an undesired state of the system is specified (usually a state that is critical from a safety or reliability standpoint), and (2) the system is then analyzed in the context of its environment and operation to find all the realistic ways in which the undesired top event can occur.

The fault tree itself is a graphic model of the various parallel and sequential combinations of faults that will result in the occurrence of the predefined undesired event. The faults can be events associated with component hardware failures, human errors, software errors, or any other pertinent factors that can lead to the undesired event. A fault tree thus depicts the logical interrelationships of basic events that lead to the top event of the fault tree.

A fault tree is composed of a complex of entities known as “gates” that serve to permit or inhibit the passage of fault logic up the tree. The gates show the relationships of events needed for the occurrence of a “higher” event. The “higher” event is the output of the gate; the “lower” events are the “inputs” to the gate. The gate symbol denotes the type of relationship of the input event required for the output event.

The qualitative evaluations basically transform the FT logic into logically equivalent forms that provide more focused information. The principal qualitative results that are obtained are the minimal cut sets (MCSs) of the top event. A cut set is a combination of basic events that can cause the top event. An MCS is the smallest combination of basic events that result in the top event. The basic events are the bottom events of the fault tree. Hence, the MCSs relate the top event directly to the basic event causes. The set of MCSs for the top event represent all the ways that the basic events can cause the top event.

A more descriptive name for a MCS may be “minimal failure set.” The set of MCSs can be obtained not only for the top event, but for any of the intermediate events (e.g., gate events) in the FT. A significant amount of information can be obtained from the structure of MCSs. Any MCS with one basic event identifies a single failure or single event that alone can cause the top event to occur. These single failures are often weak links and are the focus of upgrade and prevention actions. Examples of such single failures are a single human error or component failure that can cause a system failure.

An MCS having events with identical characteristics indicates a susceptibility to implicit dependent failure, or a common cause that can negate a redundancy. An example is an MCS of failures of identical valves. A single manufacturing defect or single environmental sensitivity can cause all the valves to simultaneously fail.

Failures can be classified in several ways (e.g., hardware faults or human error, or one of many possible hardware faults: early, random, or aging; primary, secondary or command; passive or active). More information on this classification is given in *Fault Tree Handbook with Aerospace Applications* (2002).

The quantitative evaluations of a FT consist of the determination of top event probabilities and basic event importances. Uncertainties in any quantified result can also be determined. Fault trees are typically quantified by calculating the probability of each MCS and by summing all the cut set probabilities. The cut sets are then sorted by probability. The cut sets that contribute significantly to the top event probability are called the dominant cut sets. While the probability of the top event is a primary focus in the analysis, the probability of any intermediate event in the fault tree can also be determined.

Different types of probabilities can be calculated for different applications. In addition to a constant probability value that is typically calculated, time-related probabilities can be calculated to provide the probability distribution of the time of first occurrence of the top event. Top event frequencies, failure or occurrence rates, and availabilities can also be calculated. These characteristics are particularly applicable if the top event is a system failure.

In addition to the identification of dominant cut sets, the importances of the events in the FT are among the most useful information that can be obtained from FT quantification. Quantified importances allow actions and resources to be prioritized according to the importances of the events causing the top event. The importance of the basic events, the intermediate events, and the MCSs can be determined.

Different importance measures can be calculated for different applications. One measure is the contribution of each event to the top event probability. Another is the decrease in the top event probability if the event were prevented from occurring. A third measure is the increase in the top event probability if the event were assured to occur. These importance measures are used in prioritization, prevention activities, upgrade activities, and maintenance and repair activities. Thus, substantial rich qualitative and quantitative information can be obtained from a FT.

5.8.1 Role of FTA in Decision-Making

FTA has numerous uses in enhancing product reliability:

- To understand the logic leading to the top event
- To prioritize the contributors leading to the top event
- As a proactive tool to prevent the top event
- To monitor the performance of the system
- To minimize and optimize resources

- To assist in designing a system
- As a diagnostic tool to identify and correct causes of the top event.

5.8.2 Steps of Fault Tree Analysis

A successful FTA requires the following steps be carried out:

1. Identify the objective for the FTA.
2. Define the top event of the FT.
3. Define the scope of the FTA.
4. Define the resolution of the FTA.
5. Define ground rules for the FTA.
6. Construct the FT.
7. Evaluate the FT.
8. Interpret and present the results.

5.8.3 Basic Paradigms for the Construction of Fault Trees

The basic paradigm in constructing a fault tree is to “think small,” or more accurately, “think myopically.” For each event that is analyzed, the *necessary and sufficient immediate events* (i.e., the most closely related events) that result in the event must be identified. The key phrase is “the necessary and sufficient immediate events.” The analysis does not jump to the basic causes of the event. Instead, a small step is taken and the immediate events that result in the event are identified. This taking small steps backwards assures that all of the relationships and primary causes will be uncovered. It also provides the analyst with insight into the relationships that are necessary and sufficient for the occurrence of the top event of the fault tree. This backward stepping ends as the basic causes are identified, which constitute the *resolution* of the analysis.

5.8.4 Definition of the Top Event

Some guidelines for the definition of the top event include the following:

1. To define the top event, define the criteria for the occurrence of the event. For a system failure, first define the system success criteria.
2. Assure that the top event is consistent with the problem to be solved and the objectives of the analysis.
3. If unsure of the top event, define alternative definitions that cover the top event and assess the applicability of each one.

5.8.5 Faults versus Failures

A distinction is made here between the rather specific word “failure” and the more general word “fault.” As an example of the distinction, consider a relay. If the relay closes properly when a voltage is applied across its terminals, this is a relay "success." If, however, the relay fails to close under these circumstances, this is a relay "failure." Another possibility is that the relay closes at the wrong time due to the improper functioning of some upstream component. This is clearly not a relay failure; however, untimely relay operation may well cause the entire circuit to enter into an unsatisfactory state. An occurrence like this is referred to here as a "fault." Generally speaking, all failures are faults but not all faults are failures. Failures are basic abnormal occurrences, whereas faults are "higher order" or more general events.

There are three phases in FTA. The first step is to develop a logic block diagram or a fault tree using elements of the fault tree. This phase requires complete system definition and understanding of its operation. Every possible cause and effect of each failure condition should be investigated and related to the top event. The second step is to apply Boolean algebra to the logic diagram and develop algebraic relationships between events. If possible, simplify the expressions using Boolean algebra. The third step is to apply probabilistic methods to determine the probabilities of each intermediate event and the top event. The probability of occurrence of each event has to be known; that is, the reliability of each component or subsystem for every possible failure mode has to be considered.

The graphical symbols used to construct the fault tree fall into two categories: gate symbols and event symbols. The basic gate symbols are AND, OR, *k*-out-of-*n* voting gate, priority AND, exclusive OR, and inhibit gate. The basic event symbols are basic event, undeveloped event, conditional event, trigger event, resultant event, transfer-in and transfer-out event (Kececioglu 1991; Lewis 1996; Rao 1992). Quantitative evaluation of the fault tree includes calculation of the probability of the occurrence of the top event. This is based on the Boolean expressions for the interaction of the tree events. Figure 5.3 shows the commonly used symbols in creating a fault tree. For the quantitative analysis, the basic Boolean relations are shown in Table 5.1.

In engineering analysis, the symbol for $\cup$ is $+$ and the symbol for $\cap$ is $\bullet$. Using the engineering symbols, for an application of the use of these rules, consider the simplification of the expression

$$(A+B)\bullet(A+C)\bullet(D+B)\bullet(D+C).$$

Applying the distributive law to $(A + B) \bullet (A + C)$ results in

$$(A+B)\bullet(A+C)=A+(B\bullet C).$$

Likewise,

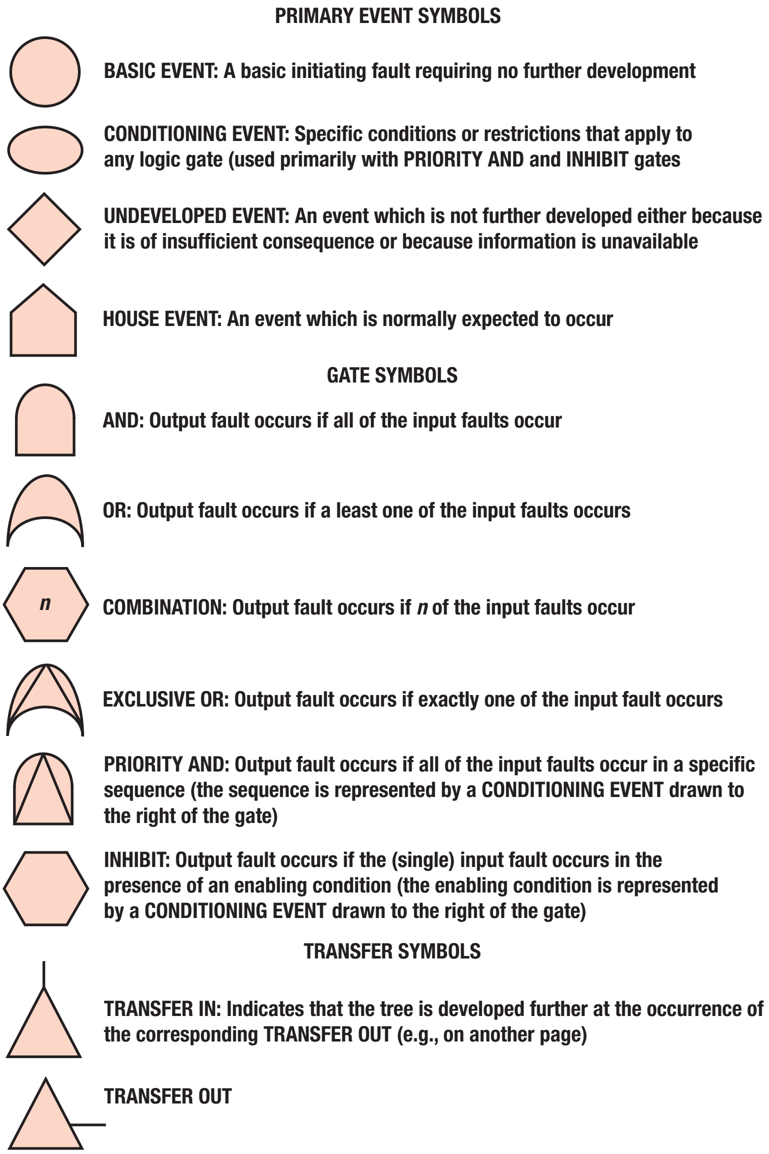
$$(D+B)\bullet(D+C)=D+(B\bullet C).$$

An intermediate result produced is

$$(A+B)\bullet(A+C)\bullet(D+B)\bullet(D+C)=(A+B\bullet C)\bullet(D+B\bullet C).$$

Letting E represent the event B $\bullet$ C results in

$$(A+B\bullet C)\bullet(D+B\bullet C)=(A+E)\bullet(D+E)=(E+A)\bullet(E+D).$$



PRIMARY EVENT SYMBOLS

BASIC EVENT: A basic initiating fault requiring no further development

CONDITIONING EVENT: Specific conditions or restrictions that apply to any logic gate (used primarily with PRIORITY AND and INHIBIT gates

UNDEVELOPED EVENT: An event which is not further developed either because it is of insufficient consequence or because information is unavailable

HOUSE EVENT: An event which is normally expected to occur

GATE SYMBOLS

AND: Output fault occurs if all of the input faults occur

OR: Output fault occurs if a least one of the input faults occurs

COMBINATION: Output fault occurs if *n* of the input faults occur

EXCLUSIVE OR: Output fault occurs if exactly one of the input fault occurs

PRIORITY AND: Output fault occurs if all of the input faults occur in a specific sequence (the sequence is represented by a CONDITIONING EVENT drawn to the right of the gate)

INHIBIT: Output fault occurs if the (single) input fault occurs in the presence of an enabling condition (the enabling condition is represented by a CONDITIONING EVENT drawn to the right of the gate)

TRANSFER SYMBOLS

TRANSFER IN: Indicates that the tree is developed further at the occurrence of the corresponding TRANSFER OUT (e.g., on another page)

TRANSFER OUT

Figure 5.3 Fault tree symbols: events and gates.

Another application of distributive law yields

$$(E+A)\bullet(E+D)=E+A\bullet D=B\bullet C+A\bullet D.$$

Therefore, the final result is

$$(A+B)\bullet(A+C)\bullet(D+B)\bullet(D+C)=B\bullet C+A\bullet D.$$

The original expression has been substantially simplified for purposes of evaluation. This idea can be applied to simplify fault trees.

Table 5.1 Rules of Boolean algebra

Mathematical symbolism	Designation
$X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$	Distributive law
$X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$	
$X \cup X = X$	Idempotent law
$X \cap X = X$	
$X \cup (X \cap Y) = X$	Law of absorption
$X \cap (X \cup Y) = X$	
$(X \cap Y)' = X' \cup Y'$	DeMorgan's theorem
$(X \cup Y)' = X' \cap Y'$	
$X \cup (X' \cap Y) = X \cup Y$	Useful result
$X' \cap (X \cup Y') = X' \cap Y'$	

Example 5.1

Reliability Block Diagram for Blackout (see Figure 5.4)
Blackout happens if both the off-site power and the emergency power fail. The emergency power fails if either the voltage monitor or the diesel generator fails. The voltage monitor signals the diesel generator to start when the offsite voltage falls below a threshold level. The fault tree for the blackout event is shown in Figure 5.5.

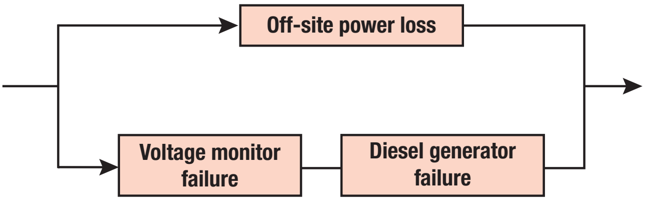


Figure 5.4 Emergency power system.

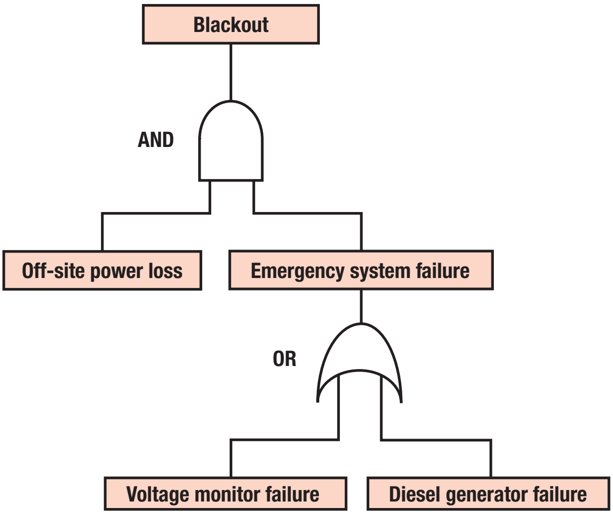
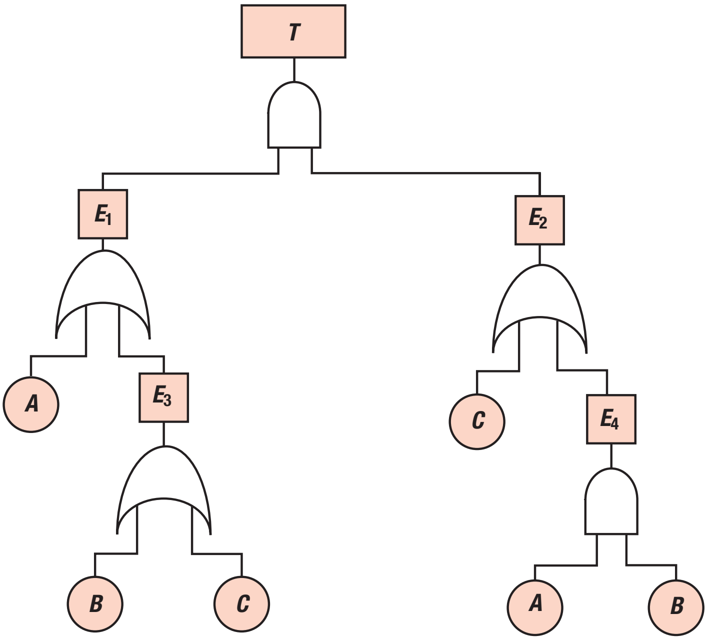


Figure 5.5 Fault tree for blackout.

Example 5.2

Analyze the following fault tree:



Top-Down Evaluation

1. $T = E_1 \cap E_2$
2. $E_1 = A \cup E_3$; $E_2 = C \cup E_4$
3. $E_3 = B \cup C$; $E_4 = A \cap B$
4. $T = (A \cup E_3) \cap (C \cup E_4) = [A \cup (B \cup C)] \cap [C \cup (A \cap B)]$.

Bottom-Up Evaluation

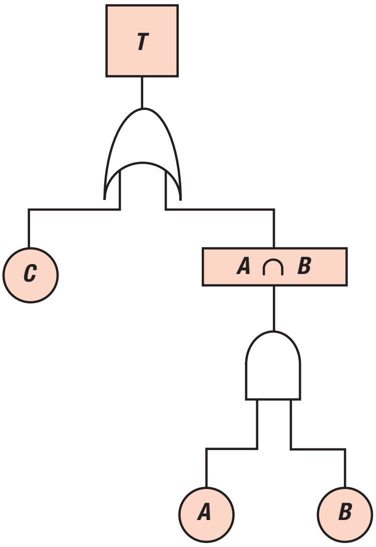
1. $E_3 = B \cup C$; $E_4 = A \cap B$
2. $E_1 = A \cup E_3$; $E_2 = C \cup E_4$
3. $E_1 = A \cup (B \cup C)$
4. $E_2 = C \cup (A \cap B)$
5. $T = E_1 \cap E_2 = [A \cup (B \cup C)] \cap [C \cup (A \cap B)]$.

Either evaluation direction can be used for FTA.

- Associative law: $A \cup (B \cup C) = (A \cup B) \cup C$
- Commutative law: $(A \cup B) \cup C = C \cup (A \cup B)$
  Thus $T = [C \cup (A \cup B)] \cap [C \cup (A \cap B)]$
- Distributive law: $T = C \cup [(A \cup B) \cap (A \cap B)]$
  $A \cap B = B \cap A$

- Associative law: $T = C \cup [(A \cup B) \cap B \cap A]$
- Absorption law: $(A \cup B) \cap B = B$
- $T = C \cup (B \cap A)$.

Hence, the tree can be reduced to show T occurs only when C or both A and B occur:



One of the main purposes of representing a fault tree in terms of Boolean equations is that these equations can then be used to determine the fault tree's associated MCSs and minimal path sets. Once the MCSs are obtained, the quantification of the fault tree is more or less straightforward. The minimal path sets are essentially the complements of the MCSs and define the "success modes" by which the top event will not occur. The minimal path sets are often not obtained in a fault tree evaluation; however, they can be useful in particular problems.

5.8.6 Minimal Cut Sets

By definition, a MCS is a combination (intersection) of primary events sufficient for the top event. The combination is a "minimal" combination in that all the failures are needed for the top event to occur; if one of the failures in the cut set does not occur, then the top event will not occur (by this combination).

Any fault tree will consist of a finite number of MCSs that are unique for that top event. One-component MCSs, if there are any, represent those single failures that will cause the top event to occur. Two-component MCSs represent the two failures that together will cause the top event to occur. For an n-component MCS, all n components in the cut set must fail in order for the top event to occur.

The MCS expression for the top event can be written in the general form,

$$T = M_1 + M_2 + \cdots + M_{\mathrm{k}},$$

where T is the top event, and $M_{i=1,2,\ldots k}$ are the MCSs. Each MCS consists of a combination of specific component failures, and hence the general n-component minimal cut can be expressed as

$$M_i = X_1 \bullet X_2 \bullet \cdots \bullet X_n,$$

where X_1, X_2, and so on, are basic component failures in the tree. An example of a top event expression, as shown in Example 5.2, is

$$T = A + B \bullet C,$$

where A, B, and C are component failures. This top event has a one-component MCS (A) and a two-component MCS ($B \bullet C$). The MCSs are unique for a top event and are independent of the different equivalent forms the same fault tree may have.

To determine the MCSs of a fault tree, the tree is first translated to its equivalent Boolean equations. A variety of algorithms exist to translate the Boolean equations into cut sets. Two of the most common are the "top-down" or "bottom-up" substitution methods to solve for the top event. The methods are straightforward and involve substituting and expanding Boolean expressions. The distributive law and the law of absorption are used to remove the redundancies.

5.9 Physics of Failure

Once the parts (materials), load conditions, and possible failure risks based on the FMMEA have been identified, the design guidelines based on physics-of-failure models aid in making design trade-offs, and can also be used to develop tests, screens, and derating[2] factors. Tests based on physics-of-failure models can be planned to measure specific quantities, to detect the presence of unexpected flaws, and to detect manufacturing or maintenance problems. Screens can be planned to precipitate failures in "weak" products while not deteriorating the design life of the shipped product. Derating or safety factors can be determined to lower the stresses for the dominant failure mechanisms.

5.9.1 Stress Margins

Products should be designed to operate satisfactorily, with margins (the design margins) at the extremes of the stated recommended operating ranges (the specification limits). These ranges must be included in the procurement requirement or specifications.

Figure 5.6 schematically represents the hierarchy of product load (stress) limits and margins. The specification limits are set by the manufacturer to limit the conditions of customer use. The design margins correspond to the load (stress) condition that

[2] Derating is the practice of subjecting parts to lower electrical or mechanical stresses than they can withstand to increase the life expectancy of the part.

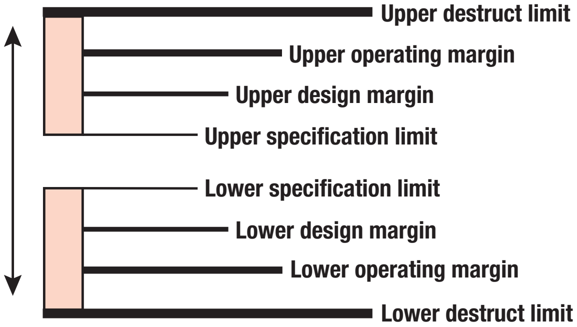


Figure 5.6 Load (stress) limits and margins.

the product is designed to survive without field failures. That is, the operating margin is the expected load (stress) that may lead to a recoverable failure. The destruct margin is the expected load (stress) that may lead to permanent (overstress) failure.

Statistical analysis and worst-case analysis should be used to assess the effects of product parameter variations. In statistical analysis, a functional relationship is established between the output characteristics of the product and its parameters. In worst-case analysis, the effect of the product outputs is evaluated on the basis of end-of-life performance values.

5.9.2 Model Analysis of Failure Mechanisms

Model analysis of failure mechanisms is based on computer-aided simulation. Model analysis can assist in identifying and ranking the dominant failure mechanisms associated with the product under life-cycle loads, determining the acceleration factor for a given set of accelerated test parameters, and determining the time to failure corresponding to the identified failure mechanisms.

Each failure model comprises a load analysis model and a damage assessment model. The output is a ranking of different failure mechanisms, based on the time to failure. The load model captures the product architecture, while the damage model depends on a material's response to the applied loads. Model analysis of failure mechanisms can be used to optimize the product design so that the minimum time to failure of the product is greater than its desired life. Although the data obtained from model analysis of failure mechanisms cannot fully replace those obtained from physical tests, they can increase the efficiency of tests by indicating the potential failure modes and mechanisms that can be expected.

It should be remembered that the accuracy of modality results depends on the accuracy of the process inputs—that is, the product geometry and material properties, the life-cycle loads, the failure models used (e.g., constants in the failure model), the analysis domain, and discretization approach (spatial and temporal). Hence, to obtain a reliable prediction, the variability in the inputs should be specified using distribution functions, and the validity of the failure models should be tested by conducting the appropriate tests.

5.9.3 Derating

To ensure that the product remains within the predetermined margins shown in Figure 5.6, derating can be used. Derating is the practice of limiting loads (e.g., thermal, electrical, or mechanical) to improve reliability. Derating can provide added protection from anomalies unforeseen by the designer (e.g., transient loads or electrical surges). For example, manufacturers of electronic parts often specify limits for supply voltage, output current, power dissipation, junction temperature, and frequency. The product design team may decide to ensure that the operational condition for a particular load, such as temperature, is always below the rated level. The load reduction is expected to extend the useful operating life, when the failure mechanisms under consideration are wearout type. This practice is also expected to provide a safer operating condition by furnishing a margin of safety when the failure mechanisms are of the overstress type.

As inherently suggested by the term “derating,” the methodology involves a two-step process: “rated” load values are first determined, and then a reduced value is assigned. The margin of safety that the process of derating provides is the difference between the maximum allowable actual applied load and the demonstrated limits of the product.

In order to be effective, derating must target the appropriate, critical load parameters, based on models of the relevant failure mechanisms. Once the failure models for the critical failure mechanisms have been identified, using, for example, FMMEA, the impact of derating on the effective reliability of the product for a given load can be determined. The goal should be to determine the “safe” operating envelope for the product and then operate within that envelope.

5.9.4 Protective Architectures

The objective of protective architectures is to enable some form of action, after an initial failure or malfunction, to prevent additional or secondary failures. Protective techniques include the use of fuses and circuit breakers, self-sensing structures, and adjustment structures that correct for parametric shifts.

In designs where safety is an issue, it is generally desirable to incorporate some means of preventing a product from failing or from causing further damage when it fails. Fuses and circuit breakers are used to sense excessive current or voltage spikes and disconnect power from the electronic products. Similarly, thermostats can be used to sense critical temperature-limiting conditions, and to power-off the product until the temperature returns to normal. Self-checking circuitry can also be incorporated to sense abnormal conditions and restore them to normal, or to activate circuitry that will compensate for the malfunction.

In some instances, it may be desirable to permit partial operation of the product after a part failure, possibly with degraded performance, rather than completely power off the product. For example, in shutting down a failed circuit whose function is to provide precise trimming adjustment within a deadband of another control product, acceptable performance may be achieved, under emergency conditions, with the deadband control product alone.

Protective architectures must be designed considering the impact of maintenance. For example, if a fuse protecting a circuit is replaced, the following questions need to be answered: What is the impact when the product is reenergized? What protective architectures are appropriate for postrepair operations? What maintenance guidance must be documented and followed when fail-safe protective architectures have or have not been included?

5.9.5 Redundancy

The purpose of redundancy is to enable the product to operate successfully even though one or more of its parts fail. A design team often finds that redundancy is the quickest way to improve product reliability if there is insufficient time to explore alternatives. It can be the most cost-effective solution, or perhaps the only solution, if the reliability requirement is beyond the state of the art.

A redundant design typically adds size, weight, and cost. When not properly implemented, redundancy can also provide a false sense of reliability. If a failure cause can affect all the redundant elements of a product at the same time, then the benefits of redundancy will be lost. Also, failures of sensing and switching circuitry or software can result in failure even in the presence of redundancy.

5.9.6 Prognostics

A product's health is the extent of deviation or degradation from its expected normal physical and performance operating condition (Vichare et al. 2004). Knowledge of a product's health can be used to detect and isolate faults or failures (diagnostics) and to predict an impending failure based on current conditions (prognostics). Thus, by determining the advent of failure based on actual life-cycle conditions, procedures can be developed to mitigate and manage potential failures and maintain the product.

Prognostics can be designed into a product by (1) installing built-in fuses and canary structures that will fail faster than the actual product when subjected to life-cycle conditions (Mishra and Pecht 2002); (2) sensing parameters that are precursors to failure, such as defects or performance degradation (Pecht et al. 2001); (3) sensing the life-cycle environmental and operational loads that influence the system's health, and processing the measured data using physics-of-failure models to estimate remaining useful life (Mishra et al. 2002; Ramakrishnan and Pecht 2003).

5.10 Design Review

The design review, a formal and documented review of a system design, should be conducted by a committee of senior company personnel who are experienced in various pertinent aspects of product design, reliability, manufacturing, materials, stress analysis, human factors, safety, logistics, maintenance, liability, and so on. The design review spans all phases of product development from conception to production and can be extended over the useful life of the product. In each phase, previous work is updated and the review is based on current information.

A mature design requires trade-offs between many conflicting factors, such as performance, manufacturability, reliability, safety and maintainability. These trade-offs depend heavily on experienced judgment and require continuous communication among experienced reviewers. The design review committee approach has been found to be extremely beneficial to this process. The committee adopts the system's point of view and considers all conceivable phases of design and system use, to ensure that the best trade-offs have been made for the particular situation.

A complete design review procedure must be multiphased in order to follow the design cycle until the system is released for production. A typical example of a review

Table 5.2 Design review committee

	Review phase			
Member	**1**	**2**	**3**	**Responsibility**
Chairperson	x	x	x	Ensure that review is conducted efficiently. Issue major reports and monitor follow-up.
Customer rep.	x	x	x	Ensure that the customer's viewpoint is adequately presented (especially at the design trade-off stage).
Design engineer (of this product)	x	x	x	Prepare and present initial design with calculations and supporting data.
Design engineer (not of this product)	x	x	x	Review and verify adequacy of design.
Reliability engineer	x	x	x	Evaluate design for maximum reliability consistent with system goals.
Manufacturing engineer		x	x	Ensure manufacturability at reasonable cost. Check for tooling adequacy and assembly problems.
Materials engineer		x		Ensure optimum material usage considering application and environment.
Stress analyst		x		Review and verify stress calculations.
Quality control engineer		x	x	Review tolerancing problems, manufacturing capability, inspection strategies, and testing problems.
Human factors engineer		x		Ensure adequate consideration of human operator. Identify potential human-induced problems.
Safety engineer		x		Ensure safety of operating and auxiliary personnel.
Maintainability engineer		x	x	Analyze for ease of maintenance repair and field servicing problems.
Logistics engineer		x	x	Evaluate and specify logistical support. Identify logistics problems.

committee, including personnel and their responsibilities, is shown in Table 5.2. Here, the review process has been subdivided into three phases, and each phase is an update of detailed analysis based on the latest knowledge.

Ultimately, the design engineer has the responsibility for investigating and incorporating the ideas and suggestions posed by the design review committee. The committee's chairperson is responsible for adequately reporting all suggestions by way of a formal and documented summary. The design engineer then can accept or reject various points in the summary; however, he or she must formally report back to the committee, stating reasons for the actions taken.

Considerably more thought and detail than the basic philosophy presented here must go into developing the management structure and procedures for conduct in order to have a successful review procedure. The review procedure must consider not only reliability, but all important factors to ensure that a mature design will result from the design effort.

5.11 Qualification

Qualification tests are conducted to identify and assess potential failures that could arise during the use of a product. Qualification tests should be performed during initial product development, and also after any significant design or manufacturing changes to an existing product.

In some cases, the target application, and therefore the use conditions, of the product may not be known. For example, a part or an assembly may be developed for sale to the open market for incorporation into many different types of products. In such cases, standard qualification tests are often employed. However, passing these tests does not mean that the product will be reliable in the actual targeted application. As a result, it is generally not sufficient to rely on qualification tests conducted on the parts (materials) of a product to determine or ensure the reliability of the final product in the targeted application.

Most often, there is insufficient time to test products for their complete targeted application life under actual operating conditions. Therefore, accelerated (qualification) tests are often employed. Accelerated testing is based on the premise that a product will exhibit the same failure mechanisms and modes in a short time under high-load conditions as it would exhibit in a longer time under actual life-cycle load conditions. The purpose is to decrease the total time and cost required to obtain reliability information for the product under study.

Accelerated tests can be divided into two categories: qualitative tests and quantitative tests. Qualitative tests generally overstress the products to determine the load conditions that will cause overstress or early wearout failures. Such tests may target a single load condition, such as shock, temperature extremes, or electrical overstress, or some combination of these. The results of the tests include failure mode information, but qualitative tests are not generally appropriate to estimate time to failure in the application.

Quantitative tests target wearout failure mechanisms, in which failures occur as a result of cumulative load conditions. These tests make analysis possible to quantitatively extrapolate from the accelerated environment to the usage environment with some reasonable degree of assurance.

The easiest form of accelerated life testing is continuous-use acceleration. The objective of this approach is to compress useful life into the shortest time possible. This approach assumes that the product is not used continuously, and that, when the product is not used, there are no loads (stresses) on it. For example, most washing machines are used for 10 hours per week on average. If a washing machine was continuously operated, the acceleration factor[3] would be (24)(7)/10 = 16.8. Thus, if the warranty or design life of the product was 5 years, the product should be tested for 5/16.8 = 0.3 years, or 106 days.

Continuous-use acceleration is not very effective with high-usage products, or with products that have a long expected life. Under such circumstances, accelerated testing is conducted to measure the performance of the product at loads (stresses) that are more severe than would normally be encountered, in order to accelerate the damage accumulation rate in a reduced time period. The goal of such testing is to accelerate time-dependent failure mechanisms and the damage accumulation rate to reduce the time to failure. Based on the data from accelerated tests, the time to failure in the targeted use conditions can be extrapolated.

Accelerated testing begins by identifying all the significant overstress and wearout failure mechanisms from the failure modes, mechanisms, and effects analysis

[3] The acceleration factor is defined as the ratio of the life of the product under normal use conditions to that under an accelerated condition.

(FMMEA). The load parameters that cause the failure mechanisms are selected as the acceleration parameters, and are commonly called accelerated loads. Typical accelerated loads include thermal loads, such as temperature, temperature cycling, and rates of temperature change; chemical loads, such as humidity, corrosives, acid, solvents, and salt; electrical loads, such as voltage or power; and mechanical loads, such as vibration, mechanical load cycles, strain cycles, and shock/impulses. Accelerated tests may require a combination of these loads. Interpretation of the results for combined loads requires a quantitative understanding of their relative interactions.

Failure due to a particular mechanism can be induced by several acceleration parameters. For example, corrosion can be accelerated by both temperature and humidity, and creep can be accelerated by both mechanical stress and temperature. Furthermore, a single accelerated load can induce failure by several mechanisms. For example, temperature can accelerate wearout damage accumulation of many failure mechanisms, such as corrosion, electrochemical migration, and creep. Failure mechanisms that dominate under usual operating conditions may lose their dominance as the load is elevated. For example, high-power electronics can generate temperatures that evaporate moisture. Conversely, failure mechanisms that are dormant under normal use conditions may contribute to device failure under accelerated conditions. Thus, accelerated tests require careful planning if they are to accelerate the actual usage environments and operating conditions without introducing extraneous failure mechanisms or nonrepresentative physical or material behaviors.

Once the failure mechanisms are identified, it is necessary to select the appropriate acceleration load; to determine the test procedures and the load levels; to determine the test method, such as constant load acceleration or step-load acceleration; to perform the tests; and to interpret the test data, which includes extrapolating the accelerated test results to normal operating conditions. The test results provide failure information to assess the product reliability, to improve the product design, and to plan warranties and support.

5.12 Manufacture and Assembly

Improper manufacturing and assembly can introduce defects, flaws, and residual stresses that act as potential failure sites or stress enhancers (or raisers) later in the life of the product. The effect of manufacturing variability on time to failure is depicted in Figure 5.7.

A shift in the mean or increase in the standard deviation of key parameters during manufacturing can result in early failure due to a decrease in the strength of the product. Generally, qualification procedures are required to ensure that the normal product is reliable. In some cases, lot-to-lot screening is required to ensure that the variability of assembly and manufacturing-related parameters are within specified tolerances. Here, screening ensures the quality of the product by precipitating latent defects before they reach the final customer.

5.12.1 Manufacturability

The design team must understand material limits and manufacturing process capabilities to construct products that promote produceability and reduce the occurrence of

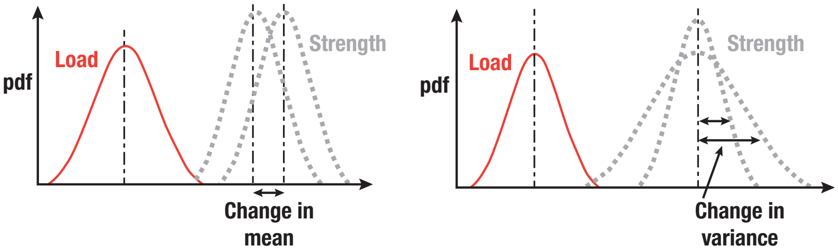


Figure 5.7 Influence of quality on failure probability.

defects. The team must also have clear definitions of the threshold for acceptable quality and of what constitutes nonconformance. Products with quality nonconformances should not be accepted.

A defect is any outcome of a process that impairs or has the potential to impair the performance of the product at any time. A defect may arise during a single process or may be the result of a sequence of processes. The yield of a process is the fraction of products that are acceptable for use in a subsequent process sequence or product life cycle. The cumulative yield of the process is approximately determined by multiplying the individual yields of each of the individual process steps. The source of defects is not always apparent, because defects resulting from a process can go undetected until the product reaches some downstream point in the process.

It is often possible to simplify processes to reduce the probability of workmanship defects. As processes become more sophisticated, however, process monitoring and control are necessary to ensure a defect-free product. The bounds that specify whether the process is within tolerance limits, often referred to as the process window, are defined in terms of the independent variables to be controlled within the process and the effects of the process on the product. The goal is to understand the effect of each process variable on each product parameter to formulate control limits for the process—that is, the condition in which the defect rate begins to have a potential for causing failure. In defining the process window, the upper and lower limits of each process variable beyond which defects might occur must be determined. Manufacturing processes must be contained in the process window by defect testing, analysis of the causes of defects, and elimination of defects by process control, such as using closed-loop corrective action systems. Establishing an effective feedback path to report process-related defect data is critical. Once this is accomplished and the process window is determined, the process window itself becomes a feedback system for the process operator.

Several process parameters may interact to produce a different defect than would have resulted from an individual parameter acting independently. This complex case may require that the interaction of various process parameters be evaluated by a design of experiments.

In some cases, a defect cannot be detected until late in the process sequence. Thus, a defect can cause rejection, rework, or failure of the product after considerable value has been added to it. This cost can reduce return on investment by adding to hidden factory costs. All critical processes require special attention for defect elimination by process control.

5.12.2 Process Verification Testing

Process verification testing is often called screening. Screening involves 100% auditing of all manufactured products to detect or precipitate defects. The aim of this step is to preempt potential quality problems before they reach the field. Thus, screening can aid in reducing warranty returns and increase customer goodwill. In principle, screening should not be required if parts (materials) are selected properly and if processes are well-controlled.

Some products exhibit a multimodal probability density function for failures, with peaks during the early period of their service life due to the use of faulty materials, poorly controlled manufacturing and assembly technologies, or mishandling. This type of early-life failure is often called infant mortality. Properly applied screening techniques can successfully detect or precipitate these failures, eliminating or reducing their occurrence in field use. Screening should only be considered for use during the early stages of production, if at all, and only when products are expected to exhibit infant mortality field failures. Screening will be ineffective and costly if there is only one main peak in the failure probability density function. Further, failures arising due to unanticipated events, such as lightning or earthquakes, may be impossible to cost-effectively screen.

Since screening is conducted on a 100% basis, it is important to develop screens that do not harm good products. The best screens, therefore, are nondestructive evaluation techniques, such as microscopic visual exams, X-rays, acoustic scans, nuclear magnetic resonance, electronic paramagnetic resonance, and so on. Stress screening involves the application of loads, possibly above the rated operational limits. If stress screens are unavoidable, overstress tests are preferred over accelerated wearout tests, since the latter are more likely to consume some useful life of good products. If damage to good products is unavoidable during stress screening, then quantitative estimates of the screening damage, based on failure mechanism models, must be developed to allow the design team to account for this loss of usable life. The appropriate stress levels for screening must be tailored to the specific product. As in qualification testing, quantitative models of failure mechanisms can aid in determining screen parameters.

A stress screen need not necessarily simulate the field environment, or even utilize the same failure mechanism as the one likely to be triggered by this defect in field conditions. Instead, a screen should exploit the most convenient and effective failure mechanism to stimulate the defects that can show up in the field as infant mortality. This requires an awareness of the possible defects that may occur in the product and familiarity with the associated failure mechanisms.

Any commitment to stress screening must include the necessary funding and staff to determine the root cause and appropriate corrective actions for all failed units. The type of stress screening chosen should be determined by the design, manufacturing, and quality teams. Although a stress screen may be necessary during the early stages of production, stress screening carries substantial penalties in capital, operating expense, and cycle time, and its benefits diminish as a product approaches maturity. If many products fail in a properly designed screen test, the design is probably faulty, or a revision of the manufacturing process may be required. If the number of failures in a screen is small, the processes are likely to be within tolerances and the observed faults may be beyond the resources of the design and production process.

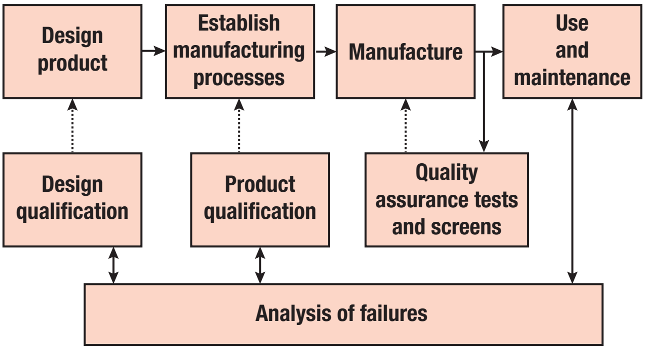


Figure 5.8 Reliability management using a closed-loop process.

5.13 Analysis, Product Failure, and Root Causes

Product reliability needs to be ensured using a closed-loop process that provides feedback to design and manufacturing in each stage of the product life cycle. Data obtained from manufacturing, assembly, storage, shipping periodic maintenance, and use and health monitoring methods can be used to aid future design plans, tests, and perform timely maintenance for sustaining the product and preventing catastrophic failures. Figure 5.8 depicts the closed-loop process for managing the reliability of a product over the complete life cycle.

The objective of closed-loop monitoring is to analyze all failures throughout the product life cycle to identify the root cause of failure. The root cause is the most basic casual factor or factors that, if corrected or removed, will prevent recurrence of the situation. The purpose of determining the root cause(s) is to fix the problem at its most basic source so it does not occur again, even in other products, as opposed to merely fixing a failure symptom.

Correctly identifying root causes during design, manufacturing, and use, followed by taking appropriate corrective actions, results in fewer field returns, major cost savings, and customer goodwill. The lessons learned from each failure analysis need to be documented, and appropriate actions need to be taken to update the design, manufacturing process, and maintenance actions.

After products are developed, resources must be applied for supply chain management, obsolescence assessment, manufacturing and assembly feedback, manufacturer warranties management, and field failure and root-cause analysis. The risks associated with the product fall into two categories:

- *Managed Risks.* Risks that the product development team chooses to proactively manage by creating a management plan and performing a prescribed monitoring regime of the field performance, manufacturer, and manufacturability
- *Unmanaged Risks.* Risks that the product development team chooses not to proactively manage.

If risk management is considered necessary, a plan should be prepared. The plan should contain details about how the product is monitored (data collection), and how the results of the monitoring feed back into various product development processes. The feasibility, effort, and cost involved in management processes must be considered.

5.14 Summary

The development of a reliable product is not a matter of chance; rather, it is a rational consequence of conscious, systematic, and rigorous efforts conducted throughout the entire life cycle of the product. Meeting the targeted product reliability can only be assured through robust product designs, capable processes that are known to be within tolerances, and qualified parts (materials) from vendors whose processes are also capable and within tolerances. Quantitative understanding and modeling of all relevant failure mechanisms can guide design, manufacturing, and the planning of test specifications.

When utilized early in the concept stage of a product's development, reliability analysis serves as an aid to determine feasibility and risk. In the design stage of product development, reliability analysis involves the selection of parts (materials), design trade-offs, design tolerances, manufacturing processes and tolerances, assembly techniques, shipping and handling methods, and maintenance and maintainability guidelines. Engineering concepts such as strength, fatigue, fracture, creep, tolerances, corrosion, and aging play a role in these design analyses. Physics-of-failure concepts, coupled with mechanistic and probabilistic techniques, are used to assess the potential problems and trade-offs and to take corrective actions.

Problems

5.1 Production lots and vendor sources for parts that comprise a design are subject to change, and variability in parts characteristics is likely to occur during the fielded life of a product. How does this affect design decisions that impact reliability?

5.2 Discuss the relationship between manufacturing process control and stress margins. How does this affect qualification? What are the implications for product reliability?

5.3 List five characteristic life-cycle loads for a computer keyboard. Describe how the product design could address these in order to ensure reliability.

5.4 Explain how the globalization of the supply chain could affect the parts selection and management process for a product used for critical military applications.

5.5 Explain the distinction between FMEA and FMMEA and how this is significant for design for reliability. For example, how would an FMMEA affect product qualification testing?

5.6 Explain how the intended application for a product would affect the decision on whether to incorporate redundancy into its design. Include in your answer a discussion of the relevant constraints related to product definition.

5.7 Discuss the concept of design for manufacturability, and how it can lead to improvement of product reliability. Provide a specific example.

5.8 What are the advantages and disadvantages of virtual qualification as compared with accelerated testing? How can these be combined in a qualification program to reduce the overall product design cycle time?

5.9 For a top-level event *T*, the following MCSs were identified: ABC, BDC, AE, ADF, and BEF. Draw a fault tree for the top event of these MCS.

5.10 Using the rules of Boolean algebra, show that

$$[(A \bullet B)+(A \bullet B')+(A' \bullet B')] = A' \bullet B.$$

5.11 Using the rules of Boolean algebra, show that

$$(A' \bullet B \bullet C') \bullet (A \bullet B' \bullet C')' = C + [(A' \bullet B') + (A + B)].$$

5.12 Using the rules of Boolean algebra, show that

$$[(X \bullet Y)+(A \bullet B \bullet C)] \bullet [(X \bullet Y)+(A' + B' + C')] = X \bullet Y.$$