Design for Electrical and Computer Engineers
Theory, Concepts, and Practice

Ralph M. Ford and Christopher S. Coulston

This document was prepared with LaTeX.

Contents

Chapter 1

System Reliability

are

> *Quality is never an accident. It is always the result of intelligent effort.—John Ruskin*

A typical design project in your academic career may never leave the confines of a laboratory. However, in industry, engineers develop systems that are used by the public at large, and issues beyond the functionality, such as reliability, safety, and maintainability become important factors in the success of the design. Over the past 20 years, industry has made a great shift to address reliability through the adoption of processes such as Quality Functional Deployment (QFD), Six-Sigma, and Robust Design. While othrefer chapters have addressed some elements of these processes, the objective of this chapter is to examine system reliability. Reliability attempts to answer the question of how long a system will operate without failing. Answering this question has inherent uncertainty and requires the use of probability and statistics. This chapter presents a review of basic probability theory and applies it to estimate the behavior of real-world devices. Reliability at the component and system levels is considered.

Learning Objectives

By the end of this chapter, the reader should:

- Have a familiarity with the basic principles of probability and understand how they apply to reliability theory.
- Understand the mathematical definition and meaning of failure rate, reliability, and mean time to failure.
- Understand how to determine the reliability of a component.
- Understand how to derate the power of electronic components for use under different operating temperatures.

- Understand how to determine the reliability of different system configurations.

DILBERT® by Scott Adams

Figure 1.1: Dogbert's Six Sigma Program. (Dilbert © United Feature Syndicate. Reprinted by permission.)

1.1 Probability Theory Review

Probability theory provides a formal framework to study chance events. It is a powerful tool for modeling engineering systems and is a requisite for reliability estimation. Although this section provides a review of some important concepts from probability, it is assumed that the reader is versed in the basics of probability theory.

In order to apply probability, some general definitions are examined first. An *experiment* is the process of measuring or quantifying the state of the world. The particular outcome of an experiment is an *event* (e_i),while the *event space* (E) is the set of all possible outcomes of the experiment. For example, consider an experiment where a six-sided die is rolled. The experiment is rolling the die and observing the outcome, the event is the particular outcome observed, and the event space for the experiment is the set $E = \{1, 2, 3, 3, 4, 5, 6\}$. The outcomes do not have to be numerical values. Another example experiment is tossing a coin, in which case the event space is $E = \{Heads, Tails\}$. Both are examples of a discrete event space because there are a finite number of experimental outcomes. In a discrete event space, the union of all the possible experimental outcomes defines the event space. If e_i is the i^{th} event in a discrete event space, then the event space is given by the union

$$E = \bigcup e_i \tag{1.1}$$

The probability of an event indicates how likely it is for an event to occur. This is quantified by the probability operator, P, that assigns to each event a real number between 0 and 1. The probability is the percentage of times that an event would occur if the experiment were repeated an infinite number of times (the Law of Large Numbers). Two of the three fundamental axioms on which probability theory is built are

$$P(e_i) \geq 0 \tag{1.2}$$

$$P(E) = 1 \tag{1.3}$$

The first axiom indicates that all probabilities are non-negative, while the second is a restatement of the event space definition—the outcome of an experiment must be an element of the event space. Armed with these definitions and axioms, some important concepts from probability are now examined.

Probability Density Functions

Not all event spaces are discrete as in the case of rolling a die or flipping a coin. Consider an experiment where the objective is to measure temperature. Clearly, such a measurement requires a variable having a continuous range of possible values. A random variable is defined as the outcome of an experiment that has a continuum of possible values. Random variables have a mathematical function known as the *probability density function* (PDF) associated with them, which when integrated, yields the probability of a range of events. A PDF is typically denoted as $p_X(x)$, where X takes values over the event space. Standard notation identifies random variables using upper case variables as the subscript for the PDF. The variable inside the parentheses is a lowercase dummy variable that does not have to match the random variable, but typically does. A question that the PDF allows us to ask is "*What is the probability that a random variable is in some range?*" Consider the case where the objective is to determine the probability that the random variable X lies between two values a and b. Written using the probability operator, this is indicated as $P(a \leq X \leq b)$. It is determined from the PDF as follows

$$P(a \leq X \leq b) = \int_a^b p_x(s)dx \tag{1.4}$$

Conceptually, this probability represents the area under the PDF between the two limits of integration as shown in Figure 1.2.

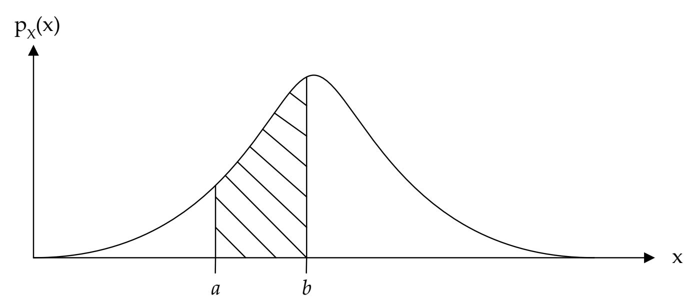


Figure 1.2: A probability density function. The area under the curve represents the probability that the random variable X lies in the interval $[a, b]$

.

Let's examine a few more important properties of probability density functions. The first, which is analogous to Equation 1.3, indicates that the probability of the event space occurring is equal to one. This is known as the normalization property and it is expressed as

$$\int_{-\infty}^{\infty} p_x(x)dx = 1 \tag{1.5}$$

Another interesting result is obtained by trying to determine the probability that a random variable takes on an exact value, for example $P(X = a)$. That is determined from the integral

$$P(X = a) = \int_a^a p_x(x)dx = 0 \tag{1.6}$$

This is a somewhat counterintuitive result—it indicates that the probability a random variable can take on a particular value is zero. Does this make any sense? Consider an experiment where the objective is to measure a voltage value for a random variable V. Now consider the question, "*What is the probability that the result of a voltage measurement equals* π *(the irrational number) volts?*" In practice, this question is impossible to answer because the precision required of the meter is infinite and contrary to its construction. So the mathematical and practical results are in harmony. There is a way around this dilemma, which is to determine the probability that the random variable is within a small range about the target value as follows

$$P(\pi < V < \pi + \Delta v) = \int_{\pi}^{\pi+\Delta v} p_V dv \approx p_V(\pi)\Delta v \tag{1.7}$$

This means that the probability a random variable is within a small range about a given value is approximated by the product of the PDF evaluated at the value and the size of the range.

Mean and Variance

Two useful and well-known statistics that are determined from the PDF are the mean (μ) and variance (σ^2). They are found from the PDF as follows

$$\mu_x = \int_{-\infty}^{\infty} x p_X(x)dx \tag{1.8}$$

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu)^2 p_X(x)dx \tag{1.9}$$

The *mean* is analogous to the center of mass of the PDF; it is also known as the average value. The *variance* is the average of the squared difference between the mean and the values of the PDF, where the squared term ensures that a positive difference is taken. The square root of the variance is known as the standard deviation σ.

Common Probability Density Functions

There are many PDFs available for describing the seemingly random variations in the behavior of observed systems and phenomena. In this section, three common PDFs (normal, exponential, and uniform) are presented.

The Normal Density

The most common density function encountered in the physical sciences and engineering is the normal density. Many population variations can be described by a normal density. For example, the resistance values of a large batch of 2.2k ohm resistors would likely follow a normal density. The normal density is defined as

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{1}{2}\frac{(x-\mu)^2}{\sigma}} \tag{1.10}$$

The mean, μ, and standard deviation, σ, are part of the definition of the PDF and used to alter the shape of the density to suit the particular need. The normal PDF is plotted in Figure 1.3. Varying μ allows the overall function to be shifted along the x-axis, while increasing σ spreads (or flattens) the function out. Calculating probabilities from the normal density can be done (although it takes a bit of work mathematically) so they are usually computed from something known as the Cumulative Distribution Function that is presented shortly.

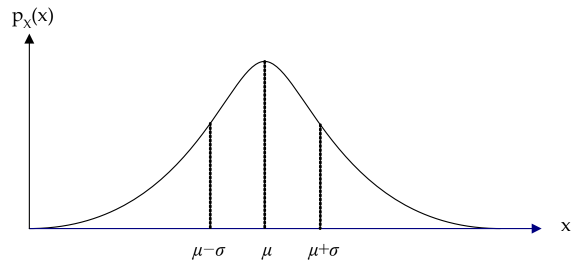


Figure 1.3: A normal density function with the mean (μ) and standard deviation (σ) shown.

The Uniform Density

The uniform density, plotted in Figure 1.3, models the outcome of an experiment where all outcomes are equally likely. Mathematically, the PDF for a uniform density is given by

$$p_x(x) = \frac{1}{b-a}, a \leq x \leq b, \tag{1.11}$$

where a, b are selected to meet the demands of a particular problem.

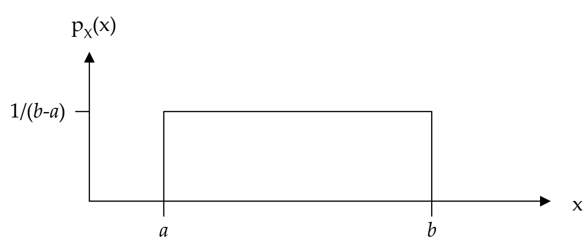


Figure 1.4: The uniform density on the interval $[a, b]$.

The Exponential Density

Exponential densities are often utilized to model time dependent functions, such as inter-arrival times between data packets in communication systems. As shown later, the exponential density also describes the behavior of component failures as a function of time. The mathematical description of an exponential density is

$$p_x(x) = \lambda e^{-\lambda x}, x \geq 0, \lambda \geq 0. \tag{1.12}$$

The PDF is characterized by the parameter λ which affects the shape of the curve as demonstrated in Figure 1.5.

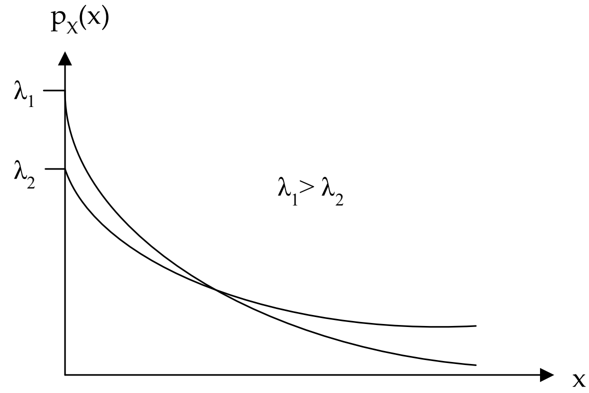


Figure 1.5: The exponential density for two different λ values.

Cumulative Distribution Functions

An important class of questions can be phrased as, "*What is the probability that a random variable X is less than value a?*" For example, the objective might be to determine the probability that an electronic component will malfunction within two years. Returning to the first question, it is clear that the goal is to determine the probability $P(X < a)$, which is found by integrating the PDF. This result is generalized by allowing the upper limit of integration to take on an arbitrary value that spans the range of the random variable. This produces a new function, known as the *Cumulative Distribution Function* (CDF), which is the integral function of the PDF and is defined as

$$\mathrm{CDF}(x) = \int_{-\infty}^{x} p_x(y)dy. \tag{1.13}$$

1.2 Reliability Prediction

Our main interest in the study of probability stems from the desire to quantify the reliability of a system. The following is a formal mathematical definition of ***reliability***.

> ***Definition***: *Reliability, $R(t)$, is the probability that a device is functioning properly (has not failed) at time t.*

In order to determine $R(t)$, it is necessary to first introduce some related mathematical entities and their meanings. The ***failure rate***, $\lambda(t)$, of a device is the expected number of failures per unit time. The failure rate is measured by operating a batch of devices for a given time interval and noting how many fail during that interval. A typical graph of failure rate versus time has the bathtub shape shown in Figure 1.6. The high initial failure rate is a result of manufacturing defects often referred to as infant mortality. Consequently, many manufactures will "burn-in" devices at the factory, so that if they fail, they do so before being sold. After the infant mortality phase, devices enter a phase of constant failure rate, where $\lambda(t) = \lambda$, known as the service life. Estimates for λ are determined empirically by testing a large number of components. They are usually expressed as a unit failure per a given number of hours, for example $\lambda = \text{failure}/10^6$ hours. After some period of time, devices start to wear-out and the failure rate increases. This usually happens as a result of mechanical wearing with age and use.

Properly designed electronic devices will not have a wear-out region, instead continuing on at a constant failure rate. This applies only to the electronic devices themselves, not necessarily to complete systems that will likely contain mechanical devices.

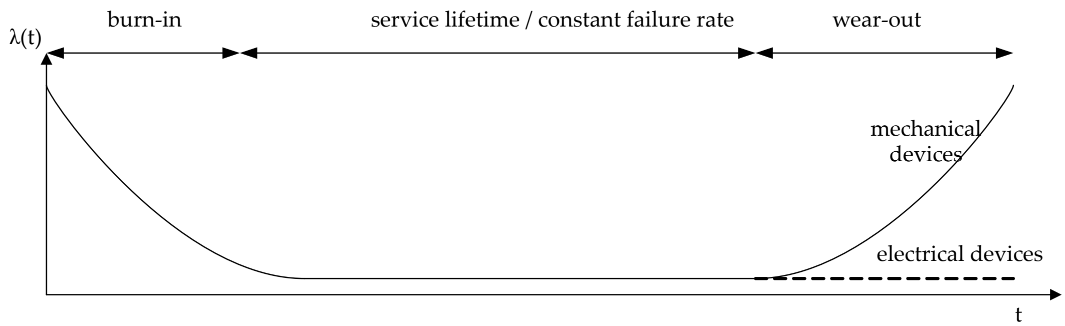


Figure 1.6: Failure rate as a function of time, also known as the bathtub curve.

In addition to failure rate, a PDF for the *failure time* of the device, $f_T(t)$ is defined, where the random variable is time T. This function allows the question to be asked "*What is the probability that a device will fail between time t_1 and t_2?*" It is important to note the difference between $\lambda(t)$ and $f_T(t)$. The failure rate tells us the average rate that a collection of identical devices will fail at a given time t, while $f_T(t)$ is a PDF used to determine the probability that a given device will fail within a specified time period. A CDF for $f_T(t)$ is determined as

$$F(t) = \int_0^t f_T(\tau)d\tau \tag{1.14}$$

$F(t)$ answers the question "*What is the probability that the device has failed by time t?*" and it is also known as the ***failure function***. Take a few seconds to go back and review the definition of $R(t)$. It is clear that $R(t)$ is directly related to $F(t)$ and is its complement. The relationship between the two is

$$R(t) = 1 - F(t) \tag{1.15}$$

Since $F(t)$ is a CDF, it increases monotonically from an initial value of 0 to a maximum value of 1 as time goes to ∞ as shown in Figure 1.7. Conversely, $R(t)$ starts at a value of 1 at time zero and decreases monotonically to a value of 0.

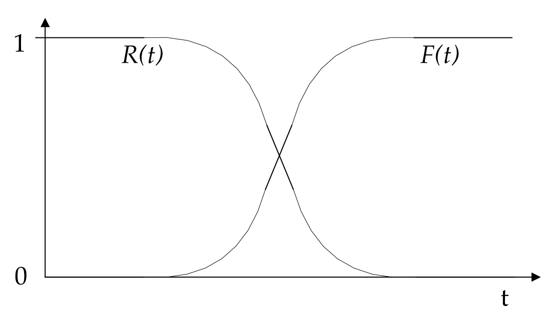


Figure 1.7: Example reliability and failure functions.

Since $\lambda(t)$ represents data that is measured empirically, it is useful to establish a relationship between $\lambda(t)$ and the ultimate goal of reliability, $R(t)$. To do so, a relationship between $\lambda(t)$, $R(t)$, and is established as follows. Consider a small period of time between t and $\Delta 4$, and determine the probability of device failure during this period. From the approximation developed in Equation 1.15, this probability is given by

$$P(\text{failure between} t \text{and} \Delta t) \approx f_T(t)\Delta t \tag{1.16}$$

How is this probability related to $R(t)$ and $\lambda(t)$? $R(t)$ provides the probability that the device is working at time t and $\lambda(t)$ gives the probability that the device will fail at time t. The product of $R(t), \lambda(t)$, and Δt gives the same probability of failure in Equation 1.16.

$$P(\text{failure betweent and} \Delta t) \approx R(t)\lambda(t)\Delta t \tag{1.17}$$

Setting Equations 1.16 and 1.18 provides the desired relationship between the three quantities

$$f_T(t) = R(t)\lambda(t), \tag{1.18}$$

that is fundamental in establishing the connection between $R(t)$ and $\lambda(t)$. However, the PDF $f_T(t)$ needs to be eliminated from Equation 1.18. This is accomplished through its relationship to the CDF $F(t)$, and thus $R(t)$, as follows

$$f_T(t) = \frac{d}{dt}F(t) = \frac{d}{dt}\big[1 - R(t)\big] = -\frac{d}{dt}R(t) \tag{1.19}$$

Equating this result with Equation 1.18 produces

$$-\frac{d}{dt}R(t) = R(t)\lambda(t) \Rightarrow \frac{-\frac{d}{dt}R(t)}{R(t)} = \lambda(t) \tag{1.20}$$

Integrating both sides gives

$$\int_0^t [\frac{-\frac{d}{dt}R(\tau)}{R(\tau)}]d\tau = \int_0^t \lambda(\tau)d\tau \Rightarrow -ln(R(t)) = \int_0^t \lambda(\tau)d\tau \tag{1.21}$$

and solving for $R(t)$ produces the final result for reliability as a function of $\lambda(t)$

$$R(t) = exp[-\int_0^t \lambda(\tau)d\tau] \tag{1.22}$$

During the service lifetime phase, the failure rate is constant, which simplifies to

$$R(t) = exp(-\lambda(t)) \tag{1.23}$$

This important result is now applied in Example 1.1.

Example 1.1 Transistor Reliability

Problem: Consider a transistor with a constant failure rate of $\lambda = 1/10^6$ hours. What is the probability that the transistor will be operable in 5 years?

Solution: This solution is found using the reliability function for a constant failure rate in Equation 1.23 as follows.

$$R(t) = exp(-\lambda t)$$
$$R(5\text{years}) = exp(-\frac{1}{10^6 \text{ hours}} * \frac{24 \text{ hours}}{\text{day}} * \frac{365 \text{ days}}{\text{year}} * 5 \text{ years}) =$$
$$= exp(-0.0438)$$
$$= 0.957$$
$$= 95.6\%$$

Mean Time to Failure

The ***mean time to failure*** (MTTF) is a quantity which answers the question, "*On average how long does it take for a device to fail?*" From its definition, it is apparent that the MTTF is the mean value of the random variable T (failure time). It is determined from the PDF and the definition of the mean in Equation 1.8 as follows

$$\text{MTTF} = \int_0^\infty t f_T(t) dt \tag{1.24}$$

At this point the form of the PDF for $f_T(t)$ is not known, but it can be found from Equation 1.20 since it is the negative derivative of $R(t)$. Assuming the form of $R(t)$ found in Equation 1.23 for a constant failure rate gives

$$f_T(t) = -\frac{d}{dt} R(t) = \lambda e^{-\lambda t} \tag{1.25}$$

This means that under the condition of a constant failure rate, the failure PDF follows an exponential density. The MTTF is found from $f_T(t)$ via integration by parts to be

$$\text{MTTF} = -\int_0^\infty t e^{-\lambda t} = \frac{1}{\lambda} \tag{1.26}$$

This makes intuitive sense because λ is the expected number of failures per unit time for a device. Consequently, the reciprocal of λ is the expected time between failures or MTTF. Let's consider a few examples.

Example 1.2 Transistor MTTF

Problem: Consider the transistor in Example 1.1. (a) Determine the MTTF, and (b) the reliability at the MTTF.

Solution: (a) From Equation 1.26 the

$$MTTF = \frac{1}{\lambda} = \frac{1}{1/10^6 \text{ hours}} = 10^6 \text{ hours}$$
$$= 114 \text{ years}$$

.

(b) From Equation 1.23 the reliability at 114 years is

$$R(t) = exp(-\lambda t)$$
$$R(114years) = exp(-\frac{10^6 \text{ hours}}{10^6 \text{ hours}}) = exp(-1) = 0.368$$
$$= 36.8\%$$

.

This is a bit counterintuitive. Although the average time between transistor failures is 114 years, an individual transistor has only a 36.8% chance of surviving to 114 years. It would seem logical that the reliability at 114 years should be 50% and that the transistor would have a 50-50 chance of failing. This would be true if $f_T(t)$ were symmetric about its mean, but that is not the case for the exponential density.

Example 1.3 Human lifespan estimation.

Problem: Data shows that for a 30 year old population, the failure (death) rate is constant with approximately 1.1 deaths per 1000 people per year. Given this data, estimate the MTTF of humans.

Solution: In order to find MTTF, λ is needed. From the information given it is

$$\lambda = \frac{(1.1/1000) \text{ failures}}{1 \text{ year}} = \frac{1.1 \text{ failure}}{10^3 \text{ years}} = \frac{1 \text{ failure}}{909 \text{ years}}$$

From this MTTF is computed as

$$MTTF = \frac{1}{\lambda} = 909 \text{ years!}$$

While great news for those of us seeking longevity, this calculation is clearly wrong since the upper limit on human lifespan is empirically known to be about 120 years. Why is this so? Serious problems arise if $R(t)$ is used in situations where the underlying assumption is invalid. The results in Equations 1.23 and 1.26 apply only if the failure rate is constant. Although that is nearly true for people in their 20s and 30s, it is not true as people age. People do wear out and the failure rate increases with age.

Failure Rate Estimates

The overriding objective of this chapter is to estimate the future behavior of devices that are used in electrical and computer systems. The particular behavior of interest is the state of a device's functionality—the reliability, is it working or has it failed? Equation 1.23 indicates that it is fairly straightforward to determine reliability, if the failure rate (λ) is known and is constant. One question to consider is what factors influence the failure rate of a device. Many of us probably have had experiences in the laboratory where we have caused devices to fail by subjecting them to conditions outside of the normal operating bounds, notably excessive current, power, or heat. In those cases the devices probably failed, or burned up, due to the operating conditions outside of the allowed bounds for the device. However, even when operated within the allowable norms of a device's operating conditions, variations in factors such as power, operating voltages, and temperatures impact λ.

The United States Military has kept copious records of device failures in the field and the conditions under which they operated. These records are synthesized in a handbook entitled Reliability Prediction of Electronic Equipment [MIL-HDBK-217F] that provides failure rates for various analog and digital components, along with adjustment factors to account for operating conditions, the environment, and device quality. Categories of devices included in the handbook are switches, fuses, diodes, optoelectronic devices, and microelectronic devices (op amps, logic devices, microcontrollers, microprocessors). The handbook was last published in 1991 and has been discontinued, but it is still widely accepted and used. Bellcore (subsequently Telcordia) has developed newer models [Tel96] based upon MIL-HDBK-217F that were updated to better predict the reliability of components. MIL-HDBK-217F is used here since it is freely available in the public domain.

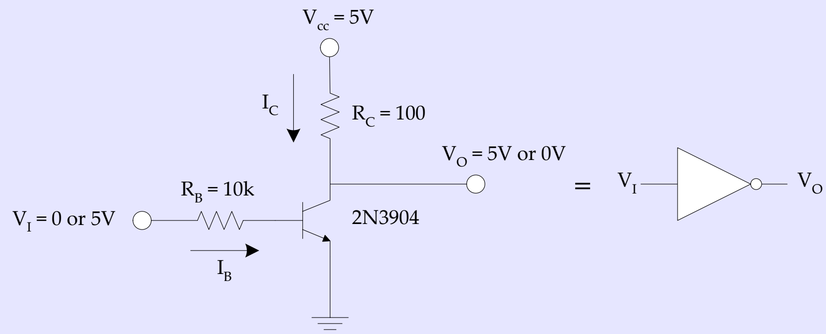
Failure rates for resistors, capacitors, transistors, and integrated circuits from MIL-HDBK-217F are included in Appendix **??**. For each device, a base failure rate is given, λ_b, and multiplied by a number of adjustment factors, denoted by the symbol π, to estimate the device failure rate λ. Each adjustment factor has a unique subscript and their values are found from tables or equations in the handbook.

For example, consider the low frequency field effect transistor in Appendix **??**. The overall failure rate is given by the equation $\lambda = \lambda_b \pi_t au \pi_A \pi_Q \pi_E$failures/$10^6$hours. λ_b is the base failure rate that is directly read from a table, π_τ is a temperature factor which is computed from an exponential equation (be careful to use the junction temperature as indicated in Appendix **??**), π_A is an application factor that depends upon how the device will be used, π_Q is a quality factor, and π_E is an environmental factor. The quality factor table lists some strange names and values from 0.7 to 8.0. The quality factor describes the level of burn-in and screening each device receives before leaving the factory. Joint Army/Navy (JAN, JANTX, and JANTXV) quality factors are the highest standard, and are usually required only for space vehicles. That individual attention to burn-in means that JAN parts are expensive and most JAN devices have passed their infant mortality phase before leaving the factory. When determining failure rate for a device from a table, it is common practice to always round parameters or values pessimistically so that the evaluation is a worst case analysis of its performance. That way the device should perform with a higher reliability when embedded into a system, hopefully causing only pleasant surprises in operation. Finally, the factor π_E is based upon the different operating environments that are identified in Appendix **??**. Example 1.4 demonstrates the application of the MIL-HDBK-217F standard for reliability estimation.

In summary, it is possible to estimate the reliability of devices if the failure rate is known and it is constant. The US Military and Telecordia handbooks provide guidance for estimating failure rates, and thus the component reliability. It must be kept in mind that they are estimates and not guaranteed to predict the exact performance. It is also apparent that this can become a rather time-consuming process if there are many components in a system, and thus the use of reliability software packages may be warranted.

Example 1.4 Reliability estimation using the MIL-HDBK 217F.

Problem: Consider the circuit below that contains a bipolar junction transistor (BJT). This electronic circuit is a simple digital logic inverter. When the input voltage V_I is 0, the BJT is off, no current flows through any branches of the device, and the output voltage, V_O, is 5V. When the input is 5V (high) the BJT goes into saturation due to the high base current (low R_B), producing a 50mA collector current, a large voltage drop across R_C, and an output voltage close to 0V. The average collector current for the two states is 25mA, producing an average power of 125mW (25mA*5V). Assume that the circuit is used in a missile launcher, the ambient temperature is 25°C, and that JANTX quality parts are used. Determine the MTTF and reliability for the 2N3904 BJT (a low power, low frequency BJT) in 20 years.



Solution: The objective of this problem is to determine the failure rate, from which the MTTF and reliability are estimated. From the MIL-HDBK-217F data in Appendix **??**, the failure rate is

$$\lambda = \lambda_b \pi_t au \pi_A \pi_R \pi_S \pi_Q \pi_E \frac{\text{failures}}{10^6 \text{hours}}$$

The base failure rate is given directly as $\lambda_b = 0.00074$. π_τ is the temperature factor and its value is determined from the relationship

$$\pi_\tau = exp[-2114[\frac{1}{T_j + 273} - \frac{1}{298}]]$$

where T_j is the junction temperature. As indicated in Appendix **??**, it is computed as

$$T_j = T_A + \Theta_{jA} P_D =$$

$$25°C\big(200\frac{°C}{W}\big)\big(125 * 10^{-3} W\big) = 50°C$$

The thermal resistance, $\Theta_J A$, is read from the 2N3904 datasheet (Appendix **??**) and will be examined in more detail shortly. The temperature factor is

$$\pi_\tau = exp[-2114[\frac{1}{50 + 273} - \frac{1}{298}]] = 1.73$$

π_A is an application factor (switched or linear amplification), and the value for the switched logic inverter is $\pi_A = 0.70$. π_R is a power rating factor that is computed based upon the maximum rated power dissipation of the 2N3904 (625mW from the component datasheet in Appendix **??**) as follows:

$$\pi_R = (P_R)^{0.37} = (0.625)^{0.37} = 0.84$$

π_S is a stress factor that is computed from the ratio of the maximum collector-emitter voltage over the maximum rated value of the device. In this circuit, the maximum value of V_{CE} is 5V (when the device is off and no current flows), while the maximum rated value of V_{CE} from the 2N3904 datasheet is 40V.

$V_S = \frac{\text{applied}V_{CE}}{\text{rated}V_{CE}} = \frac{5}{40} = 0.225$

π_Q is the quality factor, and based upon the fact that a JANTX part is $\pi_Q = 1.0$. π_E is the environmental factor, and for the missile launch application, $\pi_E = 32.0$.

All of this is brought together to compute the failure rate. The product of the adjustment factors is computed as

$$\lambda = \lambda_b \pi_t au \pi_A \pi_R \pi_S \pi_Q \pi_E \frac{\text{failures}}{10^6 \text{ hours}} =$$

$$= (\frac{7.4 * 10^{-4}}{10^6 \text{ hours}})(1.73)(0.7)(0.84)(0.066)(1.0)(32.0)$$

$$= \frac{1.59 * 10^{-3} \text{ failures}}{10^6 \text{ hours}}$$

This allows estimation of the MTTF and requested reliability

$$\text{MTTF} = \frac{1}{\lambda} = 6.3 * 10^8 \text{ hours} = 71,804 \text{years}.$$

$$R(20 \text{ years}) = exp(-\frac{20 \text{ years}}{71,804 \text{ years}}) = 0.9997$$

$$= 99.97\%$$

In conclusion, the BJT is estimated to be highly reliable in 20 years.

Thermal Management and Power Derating

One of the quantities computed in Example 1.4 was the junction temperature, T_J, which was computed from the power dissipated in the device and a quantity known as thermal resistance, θ. It is important to understand this in more detail, since it impacts the reliability of microelectronic devices. Furthermore, if the junction temperature exceeds a certain value, the device will fail. Thus, thermal management issues need to be taken into account. We start with a physical model in Figure 1.9 (a), which has a junction (the integrated circuit or device), enclosed by a case (the packaging of the device), surrounded by ambient environmental conditions. In part (b) a heat sink is included which aids in thermal transfer. Each element has associated with it a quantity known as thermal resistance that measures the ability of that particular element to transfer heat to another element. A result from heat transfer for electronics is that changes in temperature (ΔT) are proportional to the product of power dissipation (P_D) and the thermal resistance. This relationship is

$$\Delta T = P_D \Theta \tag{1.27}$$

It is similar to Ohm's Law where the change in temperature, power, and thermal resistance (units = °C/W) are analogous to voltage, current, and electrical resistance respectively. The total thermal resistance between two elements, such as between ambient and the junction, is the sum of all thermal resistances between them. In the case with no heat sink, this produces a junction to ambient resistance of $\Theta_{JA} = \Theta_{JC} + \Theta_{CA}$, while in the case with a heat sink, $\Theta_{JA} = \Theta_{JC} + \Theta_{CA} + \Theta_{SA}$. In the case of the heat sink, Θ_{CA} is replaced by the sum $\Theta_{CS} + \Theta_{SA}$, which has a lower combined thermal resistance and greater ability to dissipate heat.

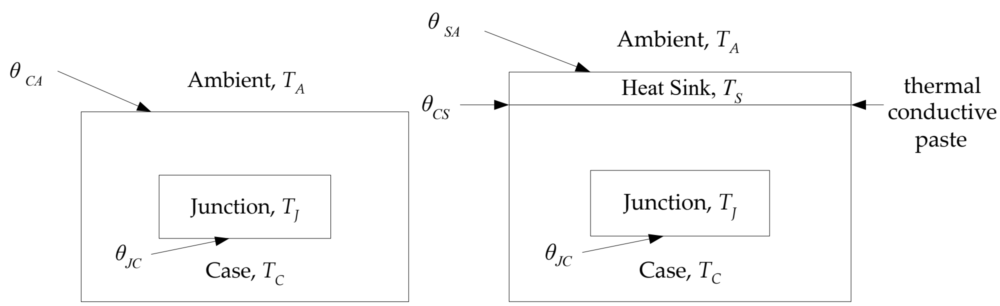


Figure 1.8: Physical model of microelectronic devices with thermal junctions and temperatures indicated. (a) Device inside casing. (b) Device inside casing with a heat sink added.

This Ohm's Law type of relationship means that thermal transfer can be modeled using familiar resistive circuits as shown in Figure 1.9. Based upon this circuit model, the temperature can be found at different points from the thermal resistance and power dissipation. Most importantly the junction temperature is found as

$$T_J = T_A + P_D\Theta_{JA} \tag{1.28}$$

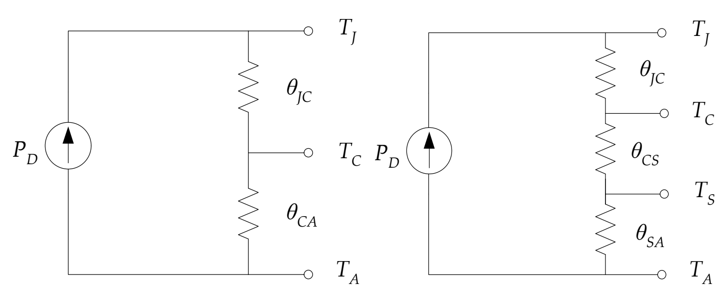


Figure 1.9: Resistive models for thermal transfer in microelectronic devices. (a) Device with no heat sink. (b) Device with a heat sink added.

Let's now apply these results. Manufacturer datasheets typically identify the absolute maximum power dissipation and note that the device should be derated if operated at ambient conditions above room temperature ($T_A = 25^{circ}C$). The datasheets also supply a maximum junction temperature for the device. It is clear from the resistive model that, for a fixed power dissipation, the junction temperature increases along with ambient temperature. If the maximum junction temperature is exceeded, the device will be destroyed. Another way to look at this is that as ambient temperature increases, the maximum amount of power a device can dissipate decreases. This decrease in maximum power dissipation is known as ***derating***. From Equation 1.28 the maximum power that can be dissipated in a device at a given ambient temperature is

$$P_{D,max} = \frac{T_{J,max} - T_A}{\Theta_{JA}} \tag{1.29}$$

From this relationship, a power derating curve is plotted in Figure 1.10 showing the maximum power versus ambient temperature. Example 1.5 demonstrates the application of this to the inverter in Example 1.4.

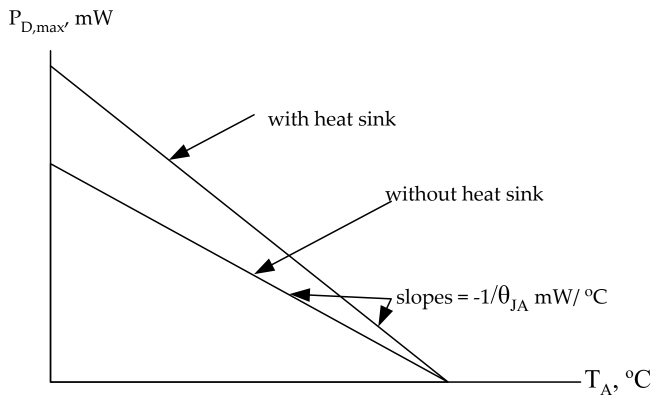


Figure 1.10: Typical power derating curves.

Example 1.5 Power Derating for the Inverter Circuit.

Problem: Assume the circuit in Example 1.4 is operating at an ambient temperature $T_A = 120°C$ and that no heat sink is used. (a) Determine the derated power and if the design is within the manufacturer's limits for power dissipation at this temperature and, (b) re-compute the reliability at 20 years based upon this elevated operating temperature.

Solution: (a) From the manufacturer datasheet in Appendix **??**, the 2N3904 BJT has a thermal resistance of $\Theta_{JA} = 200°C/W$ and a maximum junction temperature of 150°C. From this the maximum power is computed from Equation 1.10 as

$$P_{D,max} = \frac{(150 - 120)°C}{200°C/W} = 150mW$$

The derated, or maximum, power at this temperature is 150mW. Clearly, the 125mW of power dissipated as determined in Example 1.4 for the BJT is within this derated limit.

(b) To compute the failure rate the junction temperature and π_τ are re-computed.

$$T_J = T_A + P_D\Theta_{JA} = 120°C + (125 * 10^{-3}W)[\frac{200°C}{W}] = 145 \circ C$$

$$\pi_T = exp[-2114[\frac{1}{145+273} - \frac{1}{298}]] = 7.66$$

With this new value, the value of $\lambda = 7.00x10^{-3}/10^6 hours$, and the reliability is reduced slightly to 99.88%. Note, however, the junction temperature is quite high at 145°C and further increases in temperature would likely destroy the device.

Limits of Reliability Estimation

It must be kept in mind that the reliability estimates are just that, estimates, and there are limitations in their use. First, realize that the failure rate data comes from accelerated stress tests, where devices are put under stress beyond normal operating conditions, and from these the failure rates are estimated. (Nobody sits around waiting 20 years for the devices to fail!) The tests are based upon mathematical models for the failure rate and the device lifetime. Secondly, there are other factors that influence reliability that are not addressed by λ, such as the manufacturing processes used, the quality of manufacturing technologies, shock, and corrosion. Part of the value of reliability estimation is for comparative purposes when evaluating different design options. Applying these methods forces the designer to consider the operating conditions and factor them into the design.

1.3 System Reliability

The previous section focused on determining the reliability of a single device. It is natural to ask, "*How can the reliability of a system consisting of many devices be determined?*" In order to derive the overall reliability of a multi-component system, it is necessary to take into account the overall system structure.

Series Systems

Consider the inverter circuit in Example 1.4—failure of any one component in the circuit would lead to the failure of the overall system or circuit. Conceptually, a system in which the failure of a single component (or subsystem) leads to failure of the overall system is known as a ***series system***. Figure 1.11 shows a block diagram of a series system composed of boxes $S_1, S_2, \ldots S_n$ that represent the components, or the subsystems, of a larger system.

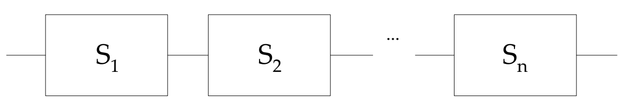


Figure 1.11: A series system consisting of components, or subsystems $S_1, S_2, \ldots S_n$.

To compute the overall reliability of a series system, $R_s(t)$, it is assumed that the failure of subsystems or components are independent events. The system is operable only if subsystems $S_1 and S_2 \ldots S_n$ are all simultaneously operating. Therefore, the probability of the overall system operating is given by the product of reliabilities for all of the subsystems as follows

$$R_s(t) = R_1(t)R_2(t)\ldots R_n(t) = \prod_{i=1}^{n} R_i(t) \tag{1.30}$$

It is important to remember that failures are assumed to be independent events, just as flipping a coin twice is considered two independent events. The overall system reliability is less than or equal to that of any single subsystem, since all reliability values are ≤ 1. Thus R_s decreases as the number of subsystems increases. Assuming a constant failure rate for all system components gives the following result for the overall system reliability

$$R_s(t) = e^{-\lambda_1 t} e^{-\lambda_2 t} \cdots e^{-\lambda_n t} = exp(-\sum_{i-1}^{n} \lambda_i t) \tag{1.31}$$

This leads to a series system failure rate and MTTF of

$$\lambda_s = \sum_{i=1}^{n} \lambda_i MTTF = \frac{1}{\lambda_s} \tag{1.32}$$

Example 1.6 revisits the inverter problem where the failure rates of all components are considered for system reliability estimation.

Example 1.6 Inverter circuit reliability.

Problem: For the system in Example 1.4 estimate (a) the overall system reliability in 20 years, and (b) the MTTF. Assume room temperature and that $\frac{1}{4}$ watt fixed composition resistors are used.

Solution: (a) Conceptually this is a series system—if any of the individual components fail, then the overall system will fail. That means that failure rates for the two resistors are needed in addition to the value previously computed for the transistor. They depend upon the power dissipated in each resistor, which is 125mW and 0.9mW for the collector and base resistors respectively. The failure rate for a fixed composition resistor from MIL-HDBK-217F is

$$\lambda_{resistor1} = \lambda_b \pi_R \pi_Q \pi_E \text{ failures}/10^6 \text{ hours}$$

For the collector resistor, R_C, the base failure rate is computed as

$$\begin{aligned}\lambda_b &= 4.5 * 10^{-9} exp\big[12\big[\frac{T+273}{343}\big]\big] exp\big[\frac{S}{0.6}\frac{T+273}{273}\big]\big] \\ &= 4.5 * 10^{-9} exp\big[12\big[\frac{25+273}{343}\big]\big] exp\big[\frac{0.125/0.25}{0.6}\frac{25+273}{273}\big]\big] \\ &= 3.77 * 10^{-4}\end{aligned}$$

The S term is the ratio of power dissipated to the maximum power rating. The values $\pi_R = 1.0$, $\pi_Q = 15.0$, and $\pi_E = 27.0$ are directly read from tables. Thus the overall failure rate for the collector resistor is

$$\lambda_{R_C} = 3.77 * 10^{-4}(1.0)(15.0)(27.0) = 1.53 * 10^{-1} \text{ failure}/10^6 \text{ hours}$$

The process for the base resistor, R_B, is similar, and results in

$$\lambda_{R_B} = 6.1 * 10^{-2} \text{ failure}/10^6 \text{ hours}$$

The total failure rate is given from Equation 1.32 as

$$\lambda_s = \lambda_{BJT} + \lambda_{R_C} + \lambda_{R_B} = 0.215 \text{ failures}/10^6 \text{ hours}$$

$$R_s(t) = exp(-\lambda_s t) = exp\Big(-\frac{0.215}{10^6 \text{ hours}} * \frac{24 \text{ hours}}{\text{day}} * \frac{365 \text{ days}}{\text{year}} * 20 \text{ years}\Big)$$

96.3%

Since resistors are pretty reliable devices, the overall system reliability decreases only a small amount relative to that of the BJT itself.

(b) The MTTF is given by $1/\lambda_s$ which in this case is 531 years.

Parallel Systems

From 1.30 it is clear that as more components are added to a series system, the reliability decreases. It is natural to ask if the reliability can be increased. The use of redundancy gives us a method to answer in the affirmative. A design has ***redundancy*** if it contains multiple modules performing the same function where a single module would suffice. By its very nature redundancy allows improperly functioning modules to be switched out of the system without affecting its behavior. With redundancy the overall system functions correctly when any one of the submodules is functioning. Figure 1.12 shows a simplified view of a ***parallel system*** with subsystems $S_1, S_2, \ldots S_n$.

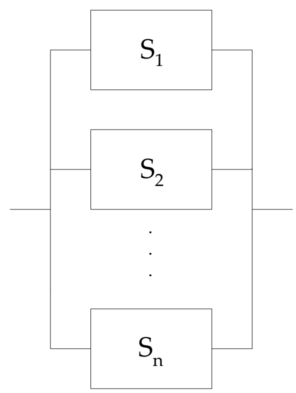


Figure 1.12: A parallel, or redundant, system consisting of subsystems $S_1, S_2, \ldots S_n$

In order to compute the reliability of a parallel system, note that a parallel system functions correctly when S_1 is functioning correctly, or S_2 is functioning correctly, $\ldots or S_n$ is functioning correctly. It would be nice if it were possible to write an equation stating that $R_s(t) = R_1(t) + R_2(t) + \ldots R_n(t)$ where + is the logical OR operator. Unfortunately, there is no direct way to realize the OR operation in probability theory. This is resolved by working with failure function, $F(t)$, instead. The probability that the system will fail by time $t, F_s(t)$, is equal to the probability that subsystem S_1 will fail and S_2 will fail and $\ldots S_n$ will fail. This probability is expressed mathematically as

$$F_s(T) = F_1(t)F_2(t)\ldots F_n(t) = \prod_{i=1}^{n} F_i(t) = \prod_{i=1}^{n}(1 - R_i(t)) \tag{1.33}$$

The overall system reliability of the parallel system is found from this as

$$R_s(t) = 1 - F_S(t) = 1 - \prod_{i=1}^{n} \left[1 - R_i(t)\right] \tag{1.34}$$

As more redundant components are added to a parallel system, additional $1 - R_i(t)$ terms are introduced into the product term. This decreases the value of the product, hence the overall system reliability is increased as more redundant systems are added.

In order for a parallel system to work, a mechanism must be in place to monitor each of the subsystems to make sure that they are operating correctly. Developing circuits to detect failures and control the switching between subsystems can be complex and is not considered here. Special care must be paid to the switching circuit itself, as malfunction of this circuit could lead to an overall system failure. An example of parallel system reliability is given in Example 1.7.

Example 1.7 Reliability of a Redundant Array of Independent Disks (RAID).

Problem: In a RAID, multiple hard drives are used to store the same data, thus achieving redundancy and increased reliability. One or more of the disks in the system can fail and the data can still be recovered. However, if all disks fail, then the data is lost. For this problem, assume that the individual disk drives have a failure rate of $\lambda = 10 failuresx/10^6 hours$. How many disks must the system have to achieve a reliability of 98% in 10 years?

Solution: The reliability of a parallel system with redundancy is given by Equation 1.34. Since all of the disks are identical, the expression simplifies to

$$R_s(t) = 1 - [1 - R_i(t)]^n$$

$$0.98 \leq 1 - \big[1 - exp\big(- \frac{10}{10^{\text{ hours}}} * \frac{\text{24 hours}}{\text{day}} * \frac{\text{365 days}}{\text{year}} * \text{10 years}\big)\big]^n$$

$$0.98 \leq 1 - [1 - 0.42]^n$$

$$0.02 \leq (0.58)^n$$

$$log(0.02) \leq nlog(0.58)$$

$$n \geq 7.2$$

In order to achieve this reliability, n=8 disks are required. The reliability of each individual disk is low at 42%, but using redundancy, the overall system reliability is quite high.

Combination Systems

Many real systems do not fit neatly into either parallel or series reliability models as shown in Figure 1.13. Rather, they may be a combination of the two, and such systems will be referred to as combination systems. One way to determine the reliability of a combination system is to utilize the results obtained for series and parallel systems in Equation 1.30 and 1.34. The system network is reduced by combining parallel subsystems into a single block, whose reliability is given by Equation 1.34, while series subsystems are reduced to a single block whose reliability is given by Equation 1.30. This is conceptually analogous to combining series and parallel resistances in electrical circuits. The network is continually reduced until only a single block remains whose reliability is known from all of the subsystem combinations.

To illustrate this, consider the system in Figure 1.13. To determine the reliability, start by combining the three parallel systems $S_2, S_3, and S_4$ whose reliability is determined by applica-

tion of Equation 1.34 to be $R_{s_{2-4}}(t) = 1-(1-R_2(t))(1-R_3(t))(1-R_4(t))$. The result is then combined with S_1 in series to give the overall system reliability,
$R_s(t) = R_1(t)\,[1-(1-R_2(t))(1-R_3(t))(1-R_4(t))]$.

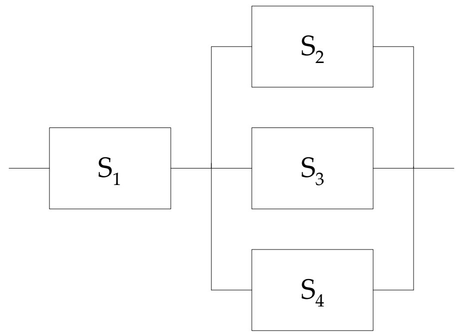
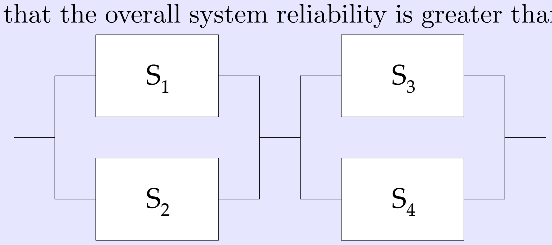


Figure 1.13: A combination series-parallel system. $S_2, S_3, and S_4$ are redundant parallel systems.

The chapter concludes with Example 1.8 that addresses combination system reliability.

Example 1.8 Combination system reliability.

Problem: Consider the system shown below with the following reliabilities at a fixed time $t, R_{s_1} = R_{s_2} = 80\%$. Determine the reliability that subsystems R_{s_3}, andR_{s_4} must have so that the overall system reliability is greater than 95%. You should assume $R_{s_3} = R_{s_4}$.



Solution: The parallel systems can combined into single systems whose reliabilities are

$$R_{s_{1,2}} = 1-(1-R_{s_1})(1-R_{s_2})$$

$$R_{s_{3,4}} = 1-(1-R_{s_3})(1-R_{s_4})$$

They are combined in series to give the overall system reliability

$$R_s = [1-(1-R_{s_1})(1-R_{s_2})] * [1-(1-R_{s_2})(1-R_{s_3})]$$

Substituting values and letting $R_{s_3} = R_{s_4}$ gives

$$0.95 = \left[1-0.2^2\right]\left[1-(1-R_{s_{3,4}})^2\right]$$

.

Solving for the reliabilities gives the final result

$$R_{s_3} = R_{s_4} = 0.90$$

This example demonstrates the power of redundant systems. $S_1 and S_2$ have somewhat low reliabilities relative to the overall system goal, but the reliability of the parallel combination of S_1 and S_2 is 96%. It requires a reliability for systems 3 and 4 of $R_{s_3} = R_{s_4} = 90\%$, while the combined reliability of systems 3 and 4 is 99%.

1.4 Summary and Further Reading

This chapter presented the basics of probability theory and methods for estimating the reliability of components and systems. Failure rate is an important quantity that is determined empirically and provides the rate of failure over the lifetime of a component or system. A mathematical definition of reliability was derived from this quantity, which takes a simple exponential form in the case of a constant failure rate. This was applied to estimate the reliability of single components, particularly using failure rates from MIL-HDBK-217F. Issues of thermal transfer and power derating were considered. Reliability estimation was extended to more realistic systems consisting of multiple components in series and parallel forms. The use of redundancy with parallel systems to increase the overall system reliability was addressed.

There are plenty of good textbooks available on the probability theory, if it is necessary to study probability theory further. The book Practical Reliability of Electronic Equipment and Products [Hna03] provides detailed coverage for electrical systems reliability. It includes factors not considered here such as thermal management on printed circuit boards, procurement practices, and electromagnetic interference. Two excellent articles that demonstrate the application of design for reliability and redundancy for an embedded system application are by George Novacek in Circuit Cellar magazine [Nov00, Nov01].