Source: STANDARD HANDBOOK OF ELECTRONIC ENGINEERING

CHAPTER 3.3
SEMICONDUCTOR RELIABILITY

Clarence M. Bailey, Jr., Richard L. Doyle

INTRODUCTION

Electronic systems are designed to operate for a specified period, which is determined by customer requirements, as well as by cost and performance. Since electronic systems consist largely of electronic devices, the system reliability is mostly dependent on the reliability of the individual devices in the system application environment. Semiconductor devices, particularly integrated circuits, considering their fundamental and complex functions, are at the heart of most modern electronic systems.

It is not possible to predict the lifetime or degradation rate (reliability) of any *individual* semiconductor device. However, it is possible to treat populations of such devices probabilistically, and thereby predict average lifetimes. This probabilistic approach is discussed here in terms of different failure rate models, which can be used to describe and predict reliability (on the average) for semiconductor devices. Failure rate λ for all electronic components including semiconductors is considered constant when used in the various reliability databases described in "Part-Failure Modeling," p. 3.10, since these databases use the exponential reliability distribution. However, these databases do not account for failure rate change ($\dot{\lambda}$). $\dot{\lambda}$ is a first-order derivative of failure rate and exhibits a sloped straight line on a linear failure rate chart. When this slope is positive, failure rate is increasing (wearout); when this slope is negative, failure rate is decreasing (infant mortality). The slope is zero when the failure rate is constant, which is generally considered the useful life of the part. However, all three regions are becoming more important as the result of more complex ICs, less than 100 micron trace widths, and lower operating and gate voltages.

This section accounts for all three regions of the failure rate curve by using a more complex reliability distribution such as the Weibull distribution. For the more complex distributions, the hazard rate $\lambda(t)$ is used to represent the instantaneous rate of failure for units of a population that have survived to time t. Electronic device reliability is usually described in terms of a particular hazard rate model, and particular parameter values, given that model. It is noted that the model also applies to electronic passive devices as well; and, in the final analysis, to electronic systems composed of such semiconductor and passive devices. The discussion here is confined, however, to semiconductors, and the model is referred to as the device hazard rate model.

DEVICE HAZARD RATE MODEL

General

Historically, hazard rates have been modeled in terms of the traditional bathtub curve. Such a curve has three regions that are associated with infant mortality, steady-state operation, and wearout. Infant mortality is characterized by an initially high, but rapidly decreasing hazard rate. These early failures come from a small fraction of the population that can be considered weak. The defects in these weak units are usually not immediately fatal

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

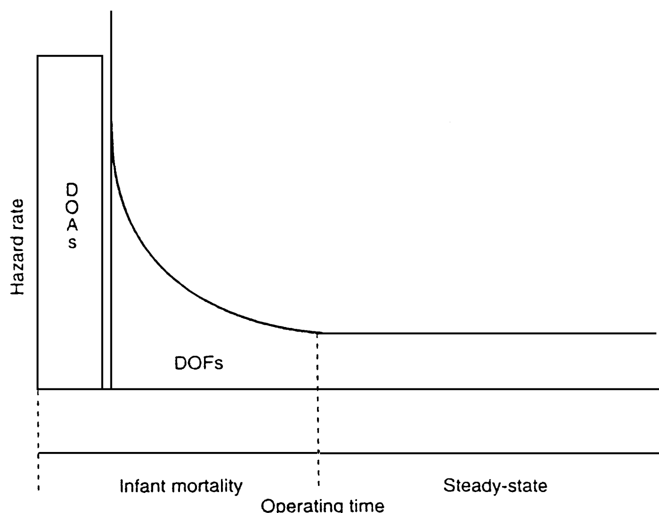

FIGURE 3.3.1 Conceptual reliability model.

but cause failure within a short period of time. After the majority of the weak units fail, operation moves into the steady-state region, in which failures occur at a much slower rate. The steady-state region is therefore characterized by a constant or slowly changing hazard rate. Wearout occurs when the hazard rate rises and the remaining units fail. For most semiconductor devices in normal environments the wearout period is far enough away to have little or no impact on the reliability of the device through normal equipment service life. However, as ICs continue to develop in complexity and physical parameters become smaller, the wearout region is starting to infringe on the useful life of the part. This may have to be addressed in the near future, but at the present time we are only concerned with the infant mortality and steady-state regions of the curve. The wearout region will have to be fit with a separate distribution in the future.

Figure 3.3.1 shows a conceptual reliability model for devices. The failures that occur very early in life are called dead-on-arrivals (DOAs), which are a part of infant mortality and are represented by the vertical box shown in Fig. 3.3.1. Most frequently they occur at first circuit board test. They may have been good when shipped but were found to have failed at various levels of equipment assembly and test. They are sometimes found at first equipment turn-on after shipment to the field. DOAs cannot be related to operating time. A device can test as satisfactory, be assembled into equipment, and then fail to work before the equipment has been in operation. The rate of such failures may in some cases be time-dependent, resulting from the same failure mechanisms as found later in infant mortality. On the other hand, others seem to be event-dependent, owing to handling during equipment manufacture and test. Although their existence is recognized, an accurate quantitative estimate of failure owing to DOAs is not possible; and we do not attempt to include them in our reliability model. Fortunately, most DOAs are found during the equipment manufacturing process. The failures that occur during operation are called "device operating failures" (DOFs). DOFs, with the exception of DOAs, encompass the infant mortality as well as the steady-state failures. Wearout is not included in this conceptual model because, as stated previously, it is not expected to occur during service life.

Failure Time Distributions

The model uses two statistical distributions that are useful in modeling device failures. The *exponential distribution* is characterized by a constant failure rate and is used to describe the steady-state hazard rate beyond the infant mortality region of device life. Pertinent functions for the exponential distribution are listed as follows and are illustrated in Fig. 3.3.2.

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

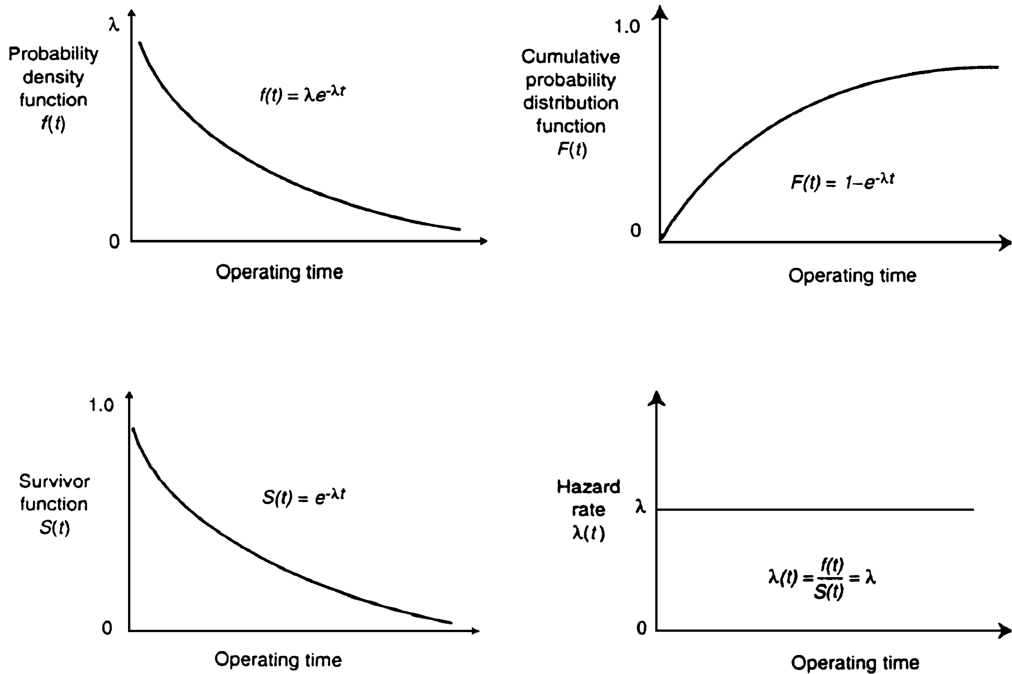

FIGURE 3.3.2 Exponential distribution.

- The probability of failure within some time interval 0 t_0 t the probability density function, is

$$f(t) = \lambda e^{-\lambda t} \qquad (t \geq 0, \lambda > 0)$$

- The probability that a device fails at or before a time t, the cumulative distribution function, is

$$F(t) = 1 - e^{-\lambda t}$$

- The probability of surviving to time t, the survivor function, is

$$S(t) = e^{-\lambda t}$$

- The hazard rate is constant:

$$\lambda(t) = \frac{f(t)}{S(t)} = \lambda$$

The Weibull distribution is especially used for modeling infant mortality failures. For the Weibull distribution, the hazard rate varies as a power of device age. The pertinent functions of the Weibull distribution are listed as follows and are illustrated in Fig. 3.3.3.

- The probability density function is

$$f(t) = \lambda_1 t^{-\alpha} e \frac{-\lambda_1 t^{1-\alpha}}{1-\alpha} \qquad (t \geq 0, \lambda_1 > 0, \alpha < 1)$$

where $\lambda_1 > 0$ is the scale parameter of the distribution hazard rate, and α is the shape parameter of the distribution.

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

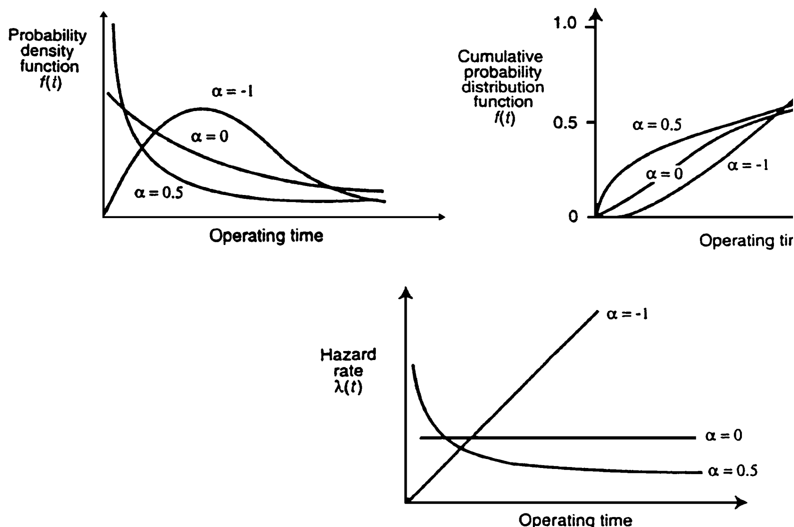

FIGURE 3.3.3 Weibull distribution.

- The cumulative distribution is

$$F(t)=1-e\frac{-\lambda_1 t^{1-\lambda}}{1-\alpha}$$

- The hazard rate is

$$\lambda(t) = \lambda_1 t^{-\alpha},\ \alpha < 1,\ t > 0$$

- The failure rate change ($\dot{\lambda}$) is

$$\dot{\lambda}(t) = -\ \alpha\lambda,\ t^{-(\alpha+1)},\ \alpha < 1,\ t > 0$$

- The survivor function is

$$S(t) = e\frac{-\lambda_1 t^{1-\alpha}}{1-\alpha}$$

When $0 < \alpha < 1$, the hazard test rate decreases with time ($\dot{\lambda}$ is negative). Therefore a positive α is used to model infant mortality. If $\alpha < 0$ the Weibull increases with device age ($\dot{\lambda}$ is positive) and device wearout may be modeled for this range of the shape parameter. If $\alpha = 0$, the hazard rate is constant ($\dot{\lambda}$ is zero) showing that the exponential distribution is a special case of the Weibull distribution. Since the shape of the Weibull distribution changes with α, it is called the shape parameter.

Specific Hazard Rate Model

Figure 3.3.4 shows the specific hazard rate model used to characterize semiconductor device reliability as well as other electronic devices and systems. The hazard rate in the infant mortality region is modeled by a Weibull hazard rate that decreases with time. In the steady-state region, the reliability is characterized by the exponential distribution where the failure rate is constant. A feature of the Weibull is that the hazard rate is a straight

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

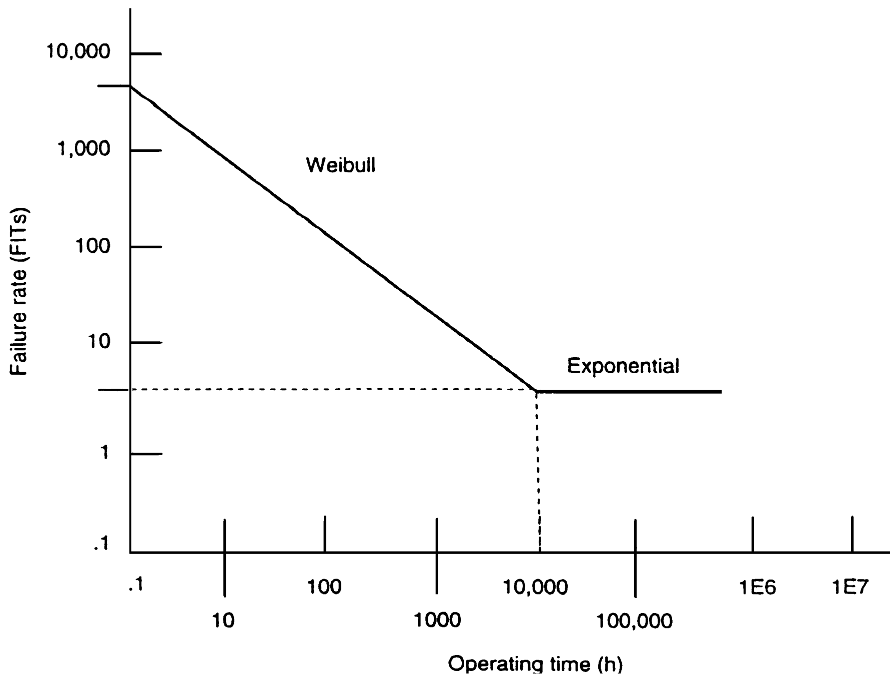

FIGURE 3.3.4 A hazard rate model for electronic components is the basis for the electronic equipment and systems model.

line when plotted on log-log scales as in Fig. 3.3.4. In such a plot, the slope is $-\alpha$, with $0 \le \alpha < 1$, and the intercept at $t = 1$ h is λ_1, and hazard rate in infant mortality is described by

$$\lambda(t) = \lambda_1 t^{-\alpha},\ 0 \le \alpha < 1,\ 0 < t < t_c \tag{1}$$

Beyond some time t_c, assumed to be 10^4 h (slightly over the approximately one year of infant mortality), the hazard rate is assumed to remain constant; that is

$$\lambda(t) = \lambda_L, \quad \text{for } t \ge t_c = 10^4 \text{ h}$$

The hazard rate unit is the FIT (or one failure in 10^9 device hours). The hazard rate shown in Fig. 3.3.4 is typical but does not necessarily correspond to any particular device. As background for this model we note that there are two distinct, but different sources for information on semiconductor device reliability: accelerated life tests, and factory- and field-monitored performance. The former provide information about reliability in the very long term, the latter primarily gives information in the short term. Accelerated test conditions are required to determine the main lifetime distribution of the main population of the devices. For semiconductor devices this is usually well described by the lognormal distribution, and, based on such accelerated tests, we can relate the distribution to normal use conditions. Specifically, the maximum hazard rate, at use conditions, can be determined from accelerated life test data. Figure 3.3.5 shows a possible relationship between accelerated stress test results and the hazard rate model. In contrast, short-term hazard rates can be directly measured from field studies of no more than 2 to 3 years' duration. In those studies, a plot of the logarithm of the observed hazard rate versus the logarithm of time is usually found to fit a straight line and can be modeled by the Weibull distribution. Beyond the infant mortality period, such field studies

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

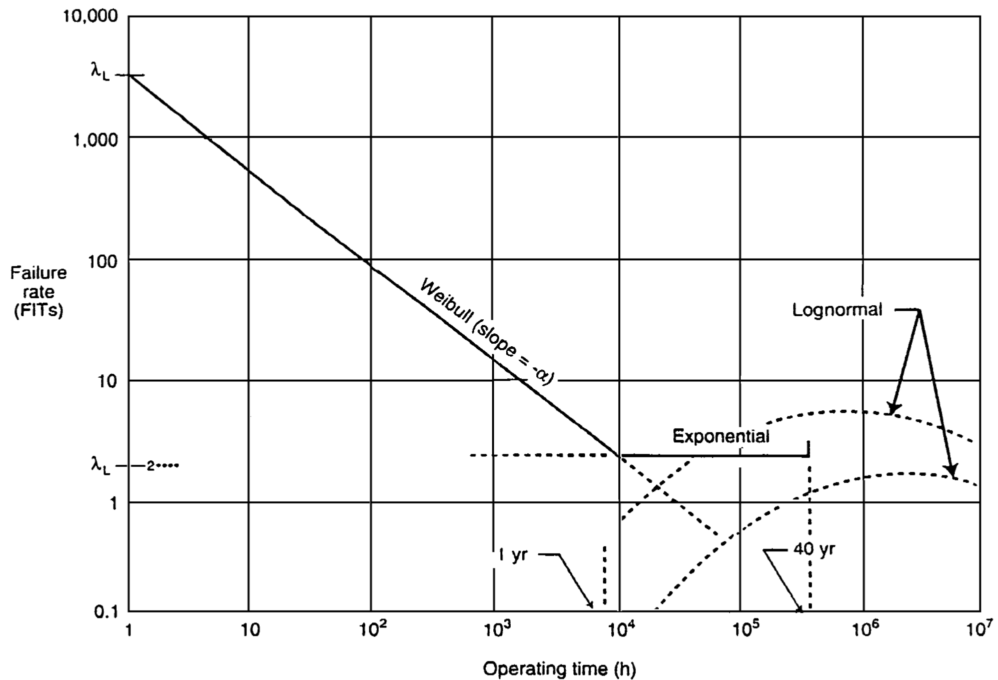

FIGURE 3.3.5 This device failure-rate model is a combination of a Weibull and an exponential distribution. Possible relationships between accelerated-stress results and the model are shown by the two-dashed lognormal curves.

have not contributed much significant information about the hazard rate, except that it is believed that the hazard rate continues to fall or levels off. Hence, we adopt the exponential distribution, with its constant failure rate, beyond 10,000 h—a conservative approach. The 10,000 h crossover point is arbitrary but reasonable. Beyond that point, the hazard rate is changing very little, and the constant failure rate model should be adequate.

Accelerated Life Model

Most modern semiconductor devices are so reliable that at normal use conditions only a very few failures are encountered. Many months of operation of large populations of devices might be needed to acquire statistically significant information on the hazard rate of those devices—at, in many cases, large costs. Another method, requiring less time, quantities of devices, and costs is accelerated testing. In an accelerated life test, a number of devices are subjected to failure-causing stresses that are at levels above what those devices would experience in use conditions. This type of accelerated aging test allows a distribution of failure times to be obtained, albeit at more stressful conditions than use conditions. In order to use this method, we must have a relationship between the distributions of the failure times at accelerated aging conditions to the distribution of failure times at use conditions, called an "accelerated life model." Not only does this accelerated life model allow us to "qualify" or prove-in devices, but it also allows us to develop accelerated stress tests for screening out some infant mortality failures; and, more importantly, relate hazard rate data acquired at one use condition to another different, use condition. Concerning the latter, it is noted that Table 3.3.3 provides possible hazard rate estimates at a reference temperature of 40°C.

We usually characterize an accelerated life model by a linear relationship between failure times at different conditions. If we designate t_1 as the failure time of a device at use conditions, t_2 as the failure time of the device

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

at accelerated (more stressful) conditions, A as an accelerated factor, and $\lambda_1(t)$ and $\lambda_2(t)$ are the hazard rates at use and more stressful conditions, respectively, then

$$\lambda_1(t) = \frac{1}{A}\lambda_2(t/A) \tag{2}$$

The acceleration factor in the preceding equation may be a result of several stress variables. The stress variable of most importance and significance for semiconductor and other electronic devices as well as electronic systems is temperature. For this we use the well-known Arrhenius relationship. Arrhenius fits the temperature dependence of a rate constant k, independent of time, to the general form:

$$k = k_0 e^{-E_n/k_B T} \tag{3}$$

where T = absolute temperature (degree, Kelvin)
E_a = activation energy
k_B = Boltzmann constant
k_0 = constant

We can derive from this general form, for two different temperatures T_1 and T_2, an acceleration factor owing temperature:

$$A_T = e^{(E_a/k_B)[(1/T_1)-(1/T_2)]} \tag{4}$$

If the activation energy for the process leading to device failure is known, then the acceleration factor for comparing reliability at two different temperatures can be calculated from the preceding equation. The Boltzmann constant is 8.6×10^{-5} eV/K. (°Kelvin is °Celsius + 273.) The activation energy eV is in electron volts.

INFANT MORTALITY CONSIDERATIONS

Infant Mortality Device Defects

These stem from a variety of device design, manufacturing, handling, and application related defects. Even though the failure mechanisms may vary from product to product and lot to lot, infant mortality can still be modeled adequately. Infant mortality failures are generally caused by manufacturing defects—including such defects as oxide pinholes, photoresist or etching defects, conductive debris on the chip, contaminants, scratches, weak bonds, and partially cracked chips and ceramics. Some infant mortality defects result from surface inversion problems, likely because of gross contamination or gross passivation defects. Many defects result from packaging problems. Some can be attributed to workmanship and manufacturing variations. Such variations can be reduced by changes in design or fabrication techniques, or in handling of the devices during manufacture. One important factor is a result of voltage spikes from *electrostatic discharge* (ESD), particularly at the circuit board manufacturing level. Some devices such as CMOS are more susceptible to this mechanism. Other infant mortality defects are inherent in design rules and constraints, or in the practical limitations of the manufacturing process and material control.

Effect of Temperature on Infant Mortality Hazard Rates

Studies of the infant mortality period show a low activation energy for the failure mechanisms contributing to semiconductor (and other electronic device) failures during infant mortality. These studies indicate an effective activation energy may be in the range of 0.37 to 0.42 eV, indicating that a single energy activation of 0.4 eV is a reasonable estimate for establishing time-temperature tradeoffs in infant mortality. Based on this value of 0.4 eV

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

activation energy, the temperature acceleration factor, A_T, can easily be calculated from Eq. (4). It cannot, however, be used directly as a multiplier for the hazard rate at the desired operating temperature. (That procedure is only correct when the hazard rate is constant as in the exponential distribution.) Instead, it can be shown that when the distribution is Weibull, as in the infant mortality period, the multiplier for the hazard rate is $(A_{IM})^{1-\alpha}$. The hazard rate is then given by

$$\lambda(t) = (A_{IM})^{1-\alpha} \lambda_1 t^{1-\alpha} \tag{5}$$

Effect of Operating Voltage on Infant Mortality Hazard Rates

The dielectric breakdown of an oxide field, in say metal oxide semiconductor (MOS) devices, has been shown to be accelerated by an electric field. Failure analysis of such devices shows that about 30 percent of infant mortality failures are because of oxide-related failures, which are a function of applied voltage and of oxide thickness. Further investigations have established a voltage-dependent acceleration factor, which can be applied to device burn-in, where a voltage stress in excess of operating voltage is applied. Namely, the acceleration factor owing voltage stress is

$$A_V = e^{|C/t_{ax}(V_1 - V_2)|} \tag{6}$$

where C is the voltage acceleration factor in angstroms per volt, t_{ax} is the oxide thickness in angstroms, V_1 is the stress voltage in volts, and V_2 is the operating voltage in volts. A conservative estimate of C is 290 A°/V.

Effect of Temperature Cycling on Infant Mortality Hazard Rates

Mechanical defects such as weak wire bonds, poor pad adhesion, and partially cracked chips on ceramics constitute a significant portion of infant mortality failures. These failure mechanisms involve either plastic deformation or crack propagation, which can be caused by repeated stress in alternate directions resulting in fatigue failure. Fatigue failure can be increased indirectly by increasing the range of ΔT. There is much evidence that temperature cycling of devices results in decreasing hazard rates, and the decrease is a function of the number of cycles. However, there appears to be no general form for an acceleration factor. More than likely, the acceleration factor is a primary function of the materials and geometries of the specific device construction, and would have to be determined by experimentation. Temperature cycling is the major methodology for a technique known as *environmental stress screening* (ESS) or *sampling testing* (EST). The commonly used technique of step-stress testing is the most effective means for determining what kinds and levels of temperature cycling would be most effective for ESS or EST of the particular semiconductor device design type.

Infant Mortality Screening

The term *screening* here refers to the use of some environmental stress as a screen for reducing infant mortality defects. Involved are such accelerated stresses as temperature, temperature cycling, combined temperature and bias, and voltage (as discussed previously for CMOS devices). The selection of the proper stress depends largely on the semiconductor design and the nature and degree of the failure mechanisms contributing to the majority of the infant mortality defects.

Temperature cycling, discussed previously, followed by gross functional and continuity testing is very effective in screening out many mechanical defects. In addition to the defects mentioned before, it will accelerate failures caused by intermetallics in gold-aluminum wirebond systems, if temperature cycling follows a high temperature bake. Temperature cycling is an effective screen for poor seals of hermetic devices; and, in plastic encapsulated devices, temperature cycling accelerates defects caused by fatigue stressing of wire bonds because of plastic material surrounding the lead and bond area.

Screening with temperature alone—called "stabilization bake" or "temperature storage"—is an effective technique for certain failure mechanisms. It involves storing devices at an elevated temperature for a specific

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

period of time. It is used to find diffusion and chemical effects, as well as material deterioration. It accelerates failures caused by oxide and gross contamination defects. It will accelerate the formation of intermetallics in a gold-aluminum wirebond system; if followed by temperature cycling, bond failures are likely to occur sooner. One effective method, referred to commonly as *burn-in*, is discussed in the following section.

Burn-in

Burn-in is an effective means for screening out defects contributing to infant mortality. Burn-in combines electrical stresses with temperature and time, and can be characterized as either static or dynamic. In static burn-in, a dc bias is applied to the device at an elevated temperature. The bias is applied in such a way as to reverse bias as many of the device junctions as possible. In dynamic burn-in the devices are operated so as to exercise the device circuit by simulating actual system operation. Static burn-in is only performed at the device level. Dynamic burn-in offers the option of use at both the device and the system level. When performed at the device level, dynamic system operation is simulated at a high temperature within the capabilities of the device. When performed at the system level, the system is operated at an elevated temperature. Since parts of the system are relatively limited in terms of temperature capability, the system burn-in is generally limited to lower temperatures than device burn-in. Static burn-in appears to be more effective for defects resulting from corrosion and contamination. Dynamic burn-in does not appear to be as effective for these problems. On the other hand, dynamic burn-in provides more access and complete exercise of internal device elements. It appears to be a much more effective screen for more complex, higher scale of integration devices. The choices of the type of burn-in (static or dynamic) and the specific stress conditions therefore depend on the device technologies, complexities, and predominant failure mechanisms, as well as the reliability requirements.

We concentrate here on a dynamic device burn-in based on the assumption of the Weibull distribution for infant mortality. This model assumes that the hazard rates of semiconductor devices are monotonically decreasing. Operation during device or system burn-in produces a certain amount of aging, and will result in a reduced hazard rate during subsequent aging. Subsequent operation begins at the reduced hazard rate and continues to decrease with additional operating time. The effect of burn-in is a function of the burn-in temperature and the subsequent operating temperature. The effective operating time is

$$t_{\text{eff}} = A_T A_V t_{bi}$$

when t_{bi} is the burn-in time at T_{bi}, the ambient burn-in temperature. (A_V is the voltage acceleration factor if performed as a part of the burn-in.) The Arrhenius equation is used to find the acceleration factor, A_{bi}, for burn-in compared to normal operation at the device hazard rate reference temperature. The hazard rate at the reference ambient temperature is

$$\lambda(t) = \lambda_1 (t_{\text{eff}} + t)^{-\alpha}$$

where $t = 0$ corresponds to the start of the device age after the burn-in. Figure 3.3.6 shows that the effect of burn-in is a decrease in the early life hazard rate. It is noted that the modeled burn-in assumes that all infant mortality defects would appear in the equivalent time of burn-in. This is a simplified effect. The stress is typically temperature, but temperature alone may not necessarily eliminate failures that might result from other stresses such as temperature cycling. Experience may well show a higher hazard rate after burn-in than predicted, followed by a more rapid decrease in hazard rate. If the temperature of subsequent operation is not at the device hazard rate reference temperature, then calculation of the hazard rate after burn-in is only slightly more complicated. The hazard rate after burn-in is then

$$\lambda_a(t) = (A_{op})^{1-\alpha} \lambda_1 [t + t_{bi}(A_{bi}/A_{op})]^{-\alpha} \tag{7}$$

where A_{op} is the acceleration factor for the operating temperature relative to the reference temperature, and

$$A_{bi} = A_T A_V$$

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

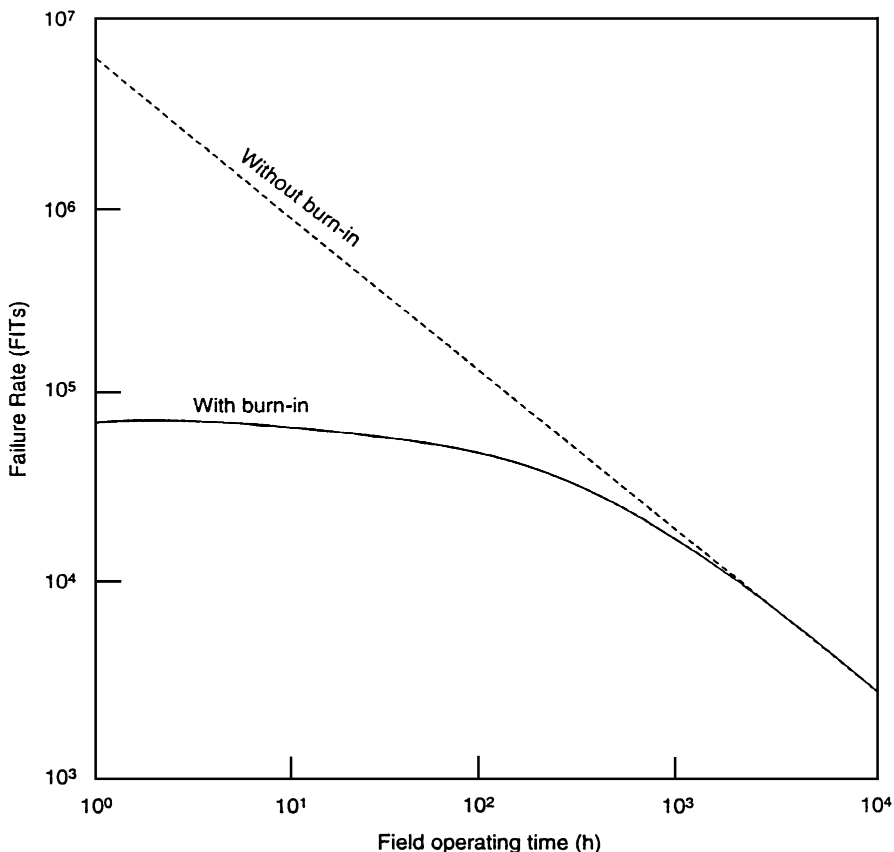

FIGURE 3.3.6 Effect of burn-in.

Infant Mortality Example

Question 1. If a device has an infant mortality hazard rate, at a reference and use temperature of 40°C, characterized by λ_1 = 15775 FITs, and α = 0.7, and if no burn-in is performed, what percentage of the devices will fail in a year (8760 h)?

Solution. The expected percent of failures, with no replacement, is given by

$$\frac{\overline{N}}{n} = F(t) = 1 - e^{(\lambda_1/1-\alpha)t^{1-\alpha}} \tag{8}$$

where $\overline{N}$ = expected number of failures
n = number of devices
$F(t)$ = Weibull cumulative distribution function

Then

$$F(t) = 1 - e^{-[10^{-9}(15775)/(1-0.7)](8760)^{1-0.7}} = 0.0008 = 0.08\,\text{percent}$$

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

Question 2. What percentage of such devices will fail in a year if they have been burned in for 10 h at 125°C?

Solution. The hazard rate after burn-in is

$$\lambda(t) = \lambda_1(t_{\text{eff}} + t)^{-\alpha} \tag{9}$$

If $\int_0^t \lambda(t')dt' << 1$, then we may use an approximate equation for dropout:

$$\text{percent failing} = 10^{-7}\int_0^{8760} \lambda_1(t_{\text{eff}} + t')^{-\alpha}\,dt' = 10^{-7}\lambda_1 \int_{t_{\text{eff}}}^{t_{\text{eff}}+8760} t^{-\alpha}d(t)$$

$$\text{percent failing} = 10^{-7}\frac{15775}{1-0.7}\left[(t_{\text{eff}} + 8760)^{1-0.7} - (t_{\text{eff}})^{1-0.7}\right]$$

$$\text{percent failing} = 0.00525\left[(t_{\text{eff}} + 8760)^{0.3} - (t_{\text{eff}})^{0.3}\right]$$

Using $E_a = 0.4$ eV, the acceleration factor, A_T, owing to temperature calculates to be approximately = 24. If, in addition, the device is an MOS device with gate oxide thickness of 250 Å, nominally operated at 5.5 V but is burned in at 7.5 V, the equation for voltage acceleration gives a voltage acceleration factor, A_V, approximately = 10. So the total burn-in acceleration factor = (24)(10) = 240, and the t_{eff} = 2400 h. The percentage failing then in 1 year after burn-in is

$$\text{percent failing} = 0.00525[(2400 + 8760)^{0.3} - (2400)^{0.3}] = 0.03 \text{ percent}$$

STEADY-STATE (LONG-TERM) CONSIDERATIONS

Long-term device reliability is defined as the reliability in the postinfant mortality period of device life, also referred to as the steady-state period of life. The hazard rate during this period is modeled as an exponential, constant failure rate. Since experience and theory indicate that, during the steady-state period, the hazard rate is near-constant and slowly changing, the assumption of the constant failure rate is felt to be a reasonable approximation. A knowledge of the contributing failure modes and mechanisms, and how to control them, is necessary to interpret and estimate long-term hazard rates. The discussion of the long-term reliability of semiconductor devices then includes a discussion of the steady-state failure mechanisms and the accelerating stresses.

Steady-State (Long-Term) Failure Mechanisms

Table 3.3.1 is a summary of many of the failure mechanisms in silicon devices. This table lists the processes leading to failure, appropriate accelerating stresses, and the ranges of activation energies. It is important to understand the accelerating stresses, as related to the specific failure mechanism, if one is to use accelerated test data to control and predict long-term hazard rates. Based on the known activation energies for the device failure mechanisms in steady-state, high temperature and other accelerated stress tests can be conducted to determine the effects of the accelerated stress tests on the device failure distribution and lifetime. Hazard rates at lower (use) stress can be extrapolated from the stress test data based on the assumed activation energy. In this activity, the lognormal distribution is found to be extremely valuable. There is much available information on the use of the lognormal distribution in analyzing accelerated testing data. The interested reader is urged to investigate the literature as needed.

Although it is difficult to list all failure mechanisms, it is possible to group some mechanisms together according to the accelerated stresses which affect them:

(*a*) Chemical reaction in contact areas, where contact metals react with the semiconductor material, and growth of intermetallic materials at the bonds of dissimilar metals such as gold and aluminum—elevated temperature, no electrical bias.

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

TABLE 3.3.1 Time-Dependent Failure Mechanisms in Silicon Devices

Device association	Failure mechanism	Relevant factors	Accelerating factors	Acceleration (apparent E_a)
Silicon-oxide and silicon-silicon oxide interface	Surface charge accumulation	Mobile ions V, T	T	$eV = 1.0 - 1.05$
	Dielectric breakdown	E, T	E	
	Charge injection	E, T, Q_{ss}	E, T	$eV = 1.3$ (slow trapping)
Metallization	Electromigration	T, j, A, gradients of T and j, grain size	T, j	$eV = 0.5 - 1.2$ j to j^4 dependence
	Corrosion (chemical, galvanic, electrolytic)	contamination H, V, T	H, V, T	Strong H effect $eV = 0.3 - 0.6$ (for electrolysis); V may have thresholds
	Contact degradation	T, metals, impurities	Varied	
Bonds and other mechanical interfaces	Intermetallic growth	T, impurities, bond strength	T	A1-Au $eV = 1.0 - 1.05$
	Fatigue	Bond strength temperature cycling	T extremes in cycling	
Metal penetration	Aluminum penetration into silicon	T, j, A	T, j	$eV = 1.4 - 1.6$
Hermeticity	Seal leaks	Pressure, differential atmosphere	Pressure	

V = voltage, T = temperature, E = electric field, j = current density, A = area, H = humidity, Q_{ss} = interfacial fixed charge.

(*b*) Surface inversion, surface charge movement, and dielectric breakdown—elevated temperature and voltage.
(*c*) Weak wire bonds, mismatches of seal materials in hermetic packages, and mismatches of thermal expansion of the chip and its package—temperature cycling.
(*d*) Corrosion of materials, such as electrolytic corrosion of aluminum or gold producing opens (aluminum) or dendritic growth causing shorts (gold)—temperature, humidity, and voltage aided by some contamination.
(*e*) Electromigration of metallization stripes resulting in opens at one end of the stripe—elevated temperature and high current density.
(*f*) Soft errors in MOS DRAM devices, caused by α-particles—the errors are temporary depending largely on the memory cell size or the charge being stored, and the incidence of random α-particles. Accelerated test by a particle source.

Determining Semiconductor Failure Rate

The following is the standard method of determining semiconductor failure rate based on stress testing samples of the integrated circuit of interest.

Example. Determine the constant failure rate in FITs (failures per billion hours) at 90 percent confidence at a junction temperature of 75°C. Use the following test results: Use an activation energy of 1.0 eV. One lot of 40 pieces tested at 150°C for 1500 with no failures. The other lot of 44 pieces tested at 150°C for

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

2500 h with no failures. Assume that the 150°C was junction temperature and not a die temperature or environmental chamber temperature.

This section explains how to calculate the failure rate using the summary test data. The failure rate resulting from these data is an average, or estimate, of the typical expected failure rate for this type of product and process technology. This calculation is made for the upper 90 percent confidence limit for the failure rate estimate using Chi-square statistics. The following formula predicts the maximum failure rate or worst-case condition:

$$\lambda(\text{max}) = \frac{\chi^2(1-a)}{2t} \tag{10}$$

with dof = $2(r + 1)$

where χ^2 = Chi-square distribution value
r = number of failures
dof = degrees of freedom
t = device hours
a = statistical error expected in estimate.

For 90 percent confidence, $a = 0.1$ or $1 - a = 0.9$, $(1 - a)$ can be interpreted to mean that we can state with statistical confidence of $1 - a$: (i.e., 90 percent) that the actual failure rate is equal to or less than the calculated maximum λ(max) failure rate.

To get the total number of device-hours, multiply the total samples tested (survivors) times the total hours of test.

The test results represent different operating and test conditions, therefore, temperature corrections are required for the data. These data require a correction using the Arrhenius equation for an activation energy of 1.0 eV. The acceleration factor is 369.6 for converting test data from 150 to 75°C.

To find the maximum failure rate at normal operating temperature (30°C) and junction temperature (75°C) with 90 percent confidence, using Eq. (10) we have

$\chi^2 = 4.61$*
$r = 0$
dof = 2
$t = (40 \times 1500 + 44 \times 2500\ \text{h})$ unit hours = 3,240,026 h
$a = 0.1$

For 90 percent confidence, $a = 0.1$ for Q failures, since $1 - a = 0.9$ equals 90 percent for P success.

Steady-State (Long-Term) Example

Question. If a VLSI semiconductor device has a long-term hazard rate, λ_L, of 40 FITs at 40°C, what percent of such devices will fail per year, after infant mortality, when operated at a temperature of 50°C? The activation energy is 0.4 eV.

Solution. The acceleration factor for operation at 50°C from a reference temperature of 40°C is

$$A_T = e^{(E_a/k_B)[(1/T_1)-(1/T_2)]} = e^{(0.4/8.6\times10^{-5})[(1/273+40)\,-\,(1/273+50)]} = 1.58 \tag{11}$$

*For a total of 0 failures, $\chi^2 = 4.61$ with 90 percent confidence. Values of χ^2 can be found in a number of statistical tables or using Excel.
$\lambda(\text{max}) = 4.61/[2 \times (40 \times 1500 + 44 \times 2500\ \text{h})]$
$\lambda(\text{max}) = 13.559 \times 10 - 6$ failures per hour or
$\lambda_{max}(150°\text{C}) = 13.559$ failures per million hours = 13,559 FITs
$\lambda_{max}(75°\text{C}) = 13{,}559$ FITs/369.6 = 36.7 FITs

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

Since, in the steady-state, constant failure rate period, the acceleration factor is A_T times the hazard rate at the reference temperature, the hazard rate at the operating temperature is approximately = 63 FITs. The percent failure per year is

$$\frac{\overline{N}}{n} = 1 - F(t) = 1 - e^{-(63\times 10^{-9})8760}$$
$$= 0.00055$$
$$= 0.055 \text{ percent}$$

Crossover Time from Infant Mortality to Steady State

Under the reference operating conditions, the crossover time is that time when infant mortality ends and the long term starts, and they have the same value of hazard rate at 10,000 h. However, for increased temperature, the constant hazard rate of the model increases more than the infant mortality hazard rate, leading to an apparent discontinuity in the model at 10,000 h. But the infant mortality region of the model is only intended to model failures that occur at a rate greater than the long-term rate. Therefore the transition time may be taken to be that time when the infant mortality hazard rate becomes equal to the steady-state hazard rate. Then for operation at accelerated conditions, the transition should occur before 10,000 h.

SEMICONDUCTOR DEVICE HAZARD RATE DATA

General

Semiconductor device hazard rate estimates are presented in this section. These estimates are based on experience, from various sources, with the various devices in communications systems in the field and in factory testing, as well as in accelerated testing. The hazard rates presented are believed to be typical of product from device suppliers who incorporate good quality and reliability programs as a part of their normal manufacturing procedures. The estimates are believed to be well within the current state of the art, and good suppliers should have little or no difficulty in producing devices that meet or exceed these hazard rate estimates.

Reference Operating Temperature

Reference conditions, for the reference operating temperature (case temperature) of 40°C, are similar to those for central office type telephone equipment; that is, the environment is air conditioned, humidity is not a significant factor, and the room ambient temperature is about 25°C. Implicit in this is the assumption that operation of the device in equipment causes its internal ambient to rise 15°C above a 25°C room ambient. If the internal ambient temperature is above 40°C, the hazard rates will be higher than the tabulated numbers.

Hazard Rate Multipliers

The hazard rates of semiconductor devices are affected by the environmental application of the devices, as well as temperature. Table 3.3.2 gives an example of the environmental application factors developed to relate the hazard rates to the environment in which the device is being used. Typically a reliability analysis will be performed using a standard failure rate datebase. Then one should use the failure rate multipliers specified by the database. These reliability databases are described on p. 3.10, “Part-Failure Modeling,” and include: MIL-HDBK-217, Telcordia/Bellcore SR-332, CNET’s RDF 2000, British Telecom’s database, Reliability Analysis Center’s PRISM, and IEEE STD 493.

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

TABLE 3.3.2 Environmental Application Factor (E), Example

Environment	E
Permanent structures, environmentally controlled	1.0
Ground shelters, not temperature controlled	1.1
Manholes, poles	1.5
Vehicular mounted	8.0

TABLE 3.3.3 Possible Semiconductor Device Hazard Rate Data

Device class	Expected hazard rate (in FITs) λ_L	α
Silicon diode		
A. General purpose	4	0.6
B. Microwave	4	0.6
C. Rectifiers	5	0.6
D. Surge protector	5	0.6
E. Switching	5	0.6
F. Varactors	6	0.75
G. Varistors	2	0.75
Integrated circuits		
A. Digital		
1. Bipolar		
1–100 gates	5	0.4
101–1000 gates	15	0.4
1001–5000 gates	25	0.4
2. CMOS		
1–100 gates	5	0.6
101–1000 gates	12	0.6
1001–10,000 gates	25	0.6
ASIC	40	0.6
3. NMOS		
1–100 gates	2	0.7
101–1000 gates	10	0.7
1001–10K gates	20	0.7
B. Linear		
≤ 100 transistors	10	0.6
101–300 transistors	15	0.6
301–1000 transistors.	30	0.6
C. Memory—PROM, EPROM, EEPROM		
1. Bipolar		
16K bits	20	0.75
128K bits	50	0.75
2. NMOS		
8K–256K bits	75	0.6
1M bits	90	0.6

(Continued)

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

TABLE 3.3.3 Possible Semiconductor Device Hazard Rate Data (*Continued*)

Device class	Expected hazard rate (in FITs) λ_L	α
3. CMOS		
8K–256K bits	70	0.6
512K bits	80	0.6
IM bits	90	0.6
D. Memory—RAM and ROM		
1. Bipolar		
<64K bits	10	0.7
64K bits	20	0.7
256K bits	40	0.7
2. CMOS		
<256K bits	2	0.6
256K bits	4	0.6
1M bits	10	0.6
4M bits	30	0.6
3. NMOS		
<64K bits	6	0.6
64K bits	8	0.6
256K bits	10	0.6
E. Microprocessors		
1. Bipolar		
4 bits	50	0.65
2. CMOS		
4 bits	80	0.7
8 bits	80	0.7
32 bits	80	0.7
3. NMOS		
8 bit	80	0.6
16 bit	80	0.6
F. Digital signal processors		
1. CMOS		
8 bits	25	0.6
16 bits	30	0.6
32 bits	50	0.6
Opto-electronics		
A. Alphanumeric displays		
1. LCD/LED		
1 Character	20	0.6
8 Characters	60	0.6
16 Characters	110	0.6
32 Characters	200	0.6
B. LEDs	10	0.6
C. Opto-isolators	20	0.6
Transistors		
A. FET	20	0.6
B. Microwave	2	0.6
C. NPN, PNP		
up to 6 W	5	0.6
>6 W	20	0.6

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.

SEMICONDUCTOR RELIABILITY

Activation Energies

As stated previously an activation energy of 0.4 eV is used for all devices in the infant-mortality period. The acceleration factor, A_{IM}, thus calculated is applied to the hazard rate by the factor, $(A_{IM})^{1-\lambda}$. An activation energy of 0.4 eV may also be used for devices in the steady-state region. This is an "effective activation energy" based on the variety of different failure mechanisms. In the steady state case, the acceleration factor, A_T, thus calculated, is applied directly to the long-term hazard rate.

Hazard Rate Table

If you are using a standard failure rate database, you should use the failure rates of the parts as specified by the database. These reliability databases are described in "Part-Failure Modeling," p. 3.10. You may want to apply the following techniques to the infant mortality portion of the system. Table 3.3.3 presents the hazard rate estimates by device type, and generally by scale of integration, at the reference operating temperature of 40°C. The long-term hazard rate is presented in terms of the parameter λ_L in FITs. The Weibull slope during infant mortality is listed in terms of the parameter α. The infant mortality hazard rate λ_1 is not listed but may be calculated from the relationship

$$\lambda_1 = \lambda_L\,(10{,}000)^{\alpha}$$

As stated before, the activation energy which applies during infant mortality and possibly during steady state is 0.4 eV.

BIBLIOGRAPHY

"Reliability Stress Screening of Electronic Hardware," IEC 60300-3-7 (1999–05).

Amerasekera, E. A., and F. N. Najm, "Failure Mechanisms in Semiconductor Devices," 2nd ed., Wiley Europe, 1997.

Chien, W. T-K., and C. H. J. Huang, "Practical 'Building-in' Approaches for Semiconductor Manufacturing," *IEEE Trans. on Reliability*, December, 2002.

Downloaded from Digital Engineering Library @ McGraw-Hill (www.digitalengineeringlibrary.com)
Copyright © 2004 The McGraw-Hill Companies. All rights reserved.
Any use is subject to the Terms of Use as given at the website.