12

12 Derating and Uprating

Derating is the practice of limiting thermal, electrical, and mechanical "stresses" to levels below the manufacturer's specified ratings, to improve reliability. Derating allows added protection from anomalies unforeseen by the designer (e.g., transient loads and electrical surge).

12.1 Part Ratings

Ratings set by manufacturers of parts and subsystems on their environmental and operational limits affect the decision-making by the part users and equipment manufacturers. This section explains the ratings with the examples of electronic parts.

Part datasheets provide two types of ratings: *absolute maximum* ratings and *recommended operating conditions*. In general:

- *Absolute maximum ratings* (AMR) are provided as a limit for the "reliable" use of parts.
- *Recommended operating conditions* (ROC) are the conditions within which electrical functionality and specifications given with the part datasheet are guaranteed.

Intel (1995) considers the difference between absolute and maximum ratings as guidance to users on to how much variation from the recommended ratings can be tolerated without damage to the part. Motorola (*Boyle* v. *United Technologies Corp.* 1988) states that, when a part is operated between the ROC and AMR, it is not guaranteed to meet any electrical specifications on the datasheet, but the physical failure or adverse effects on reliability are not expected. Motorola notes margins of safety are added to the absolute maximum ratings to ensure the recommended operating conditions (*Boyle* v. *United Technologies Corp.* 1988).

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

12.1.1 Absolute Maximum Ratings

The absolute maximum ratings section in the datasheet includes limits on operational and environmental conditions, including power, power derating, supply and input voltages, operating temperature (e.g., ambient, case, and junction), and storage temperature. The IEC (IEC/PAS 62240 2001) defines absolute maximum ratings as "limiting values of operating and environmental conditions applicable to any electronic device of a specific type as defined by its published data, which should not be exceeded under the worst possible conditions. These values are chosen by the device manufacturer to provide acceptable serviceability of the device taking no responsibility for equipment variations, and the effects of changes in operating conditions due to variations in the characteristics of the device under consideration and all other electronic devices in the equipment."

The IEC (IEC/PAS 62240 2001) also states, "The equipment manufacturer should design so that, initially and throughout life, no absolute-maximum value for the intended service is exceeded for any device under the worst probable operating conditions with respect to supply voltage variation, equipment component variation, equipment control adjustment, load variations, signal variation, environmental conditions, and variation in characteristics of the device under consideration and of all other electronic devices in the equipment." In summary, companies that integrate electronic parts into products and systems are responsible for assuring that the AMR conditions are not exceeded.

Part manufacturers generally state that below the AMR but above the recommended conditions, the performance of the part is not guaranteed, but the useful life of the part will not be affected. That is, there are no reliability concerns below the AMR. Some manufacturers (e.g., Motorola) suggest that operating parameters within the recommended operating range are not guaranteed at or near the AMR, and there may be reliability concerns over the long term[1] (Lieberman 1998; Pfaffenberger and Patterson 1987; United States Department of Defense 1996). Motorola (Lycoudes 1995) also states that noise can push the environment beyond "destruct" limits when parts are operated near the absolute maximum ratings.

Philips notes, "The 'RATINGS' table (limiting values in accordance with the Absolute Maximum System—IEC 134) lists the maximum limits to which the device can be subjected without damage. This doesn't imply that the device will function at these extreme conditions, only that when these conditions are removed and the device operated within the recommended operating conditions, it will still be functional and its useful life won't have been shortened (Philips 1988).

12.1.2 Recommended Operating Conditions

Recommended operating conditions provided by part manufacturers include parameters such as voltage, temperature ranges, and input rise and fall time. Part manufacturers guarantee the electrical parameters (e.g., typical, minimum, and maximum) of the parts only when the parts are used within the recommended operating conditions.

[1]Some EIA/JEDEC documents refer to absolute maximum ratings as absolute maximum "continuous" ratings. In those documents, transient conditions under which these ratings may be exceeded are defined. For example, the JEDEC standard for description of low voltage TTL-Compatible CMOS logic devices [53], states that "Under transient conditions these rating [AMR] may be exceeded as defined in this specification."

Philips notes, "The recommended operating conditions table [in the Philips datasheet] lists the operating ambient temperature and the conditions under which the limits in the DC characteristics and AC characteristics will be met" (Philips 1988). Philips also states that "The table (of recommended operating conditions) should not be seen as a set of limits guaranteed by the manufacturer, but the conditions used to test the devices and guarantee that they will then meet the limits in the DC and AC characteristics table" (Solomon et al. 2000).

12.1.3 Factors Used to Determine Ratings

Factors used to determine the AMR and recommended operating conditions include (Rofail and Elmasry 1993):

- Margins determined through electrical testing and procedures and methods used to set specifications from device test characterization data
- Competitors' second source advantages and limits set to maintain parity with competitors' products
- Design rule limitations: physical dimensions of device elements
- Semiconductor fabrication process: manufacturing processes and conditions that affect temperature sensitivity of parameters.

12.2 Derating

The stress limits on the products are often determined through a combination of manufacturer[2]-specified stress limits and some "derating" criteria. Derating is the practice of using an electronic part in a narrower environmental and operating envelope than its manufacturer designated limits. The purpose of derating is to lower the (i.e., electrical, thermal, and mechanical) stresses acting on part. These lower stresses are expected to extend useful operating life where the failure mechanisms under consideration are wear out type. This practice is also expected to provide a safer operating condition by furnishing a "margin of safety" when the failure mechanisms are of overstress type.

The concept of derating is schematically shown in the Figure 12.1, Figure 12.2, and Figure 12.3. These concepts are based on the load strength interference relationship. Figure 12.1 shows the load distribution on a component and the strength distribution of the component in the same scale. There are potential reliability issues only if there is an intersection between the load and strength distribution.

12.2.1 How Is Derating Practiced?

The techniques for derating of electronic parts customarily comprise of the steps described later in the list. The exact methodology varies between organizations and

[2]The terms "manufacturer," "user," "OEM," and the like are used to identify different levels of entities in the electronics industry supply chain. It should be noted that under different circumstances, the same company can play the role of more than one such entity. Also (particularly in vertically integrated companies), the different division of the same organization can play the role of the separate entities.

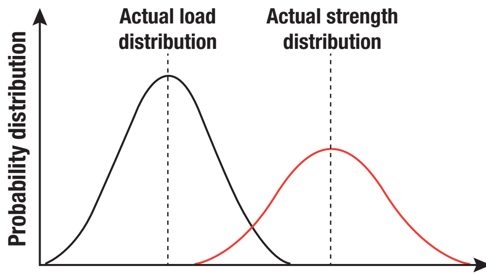

Figure 12.1 Load-strength interference.

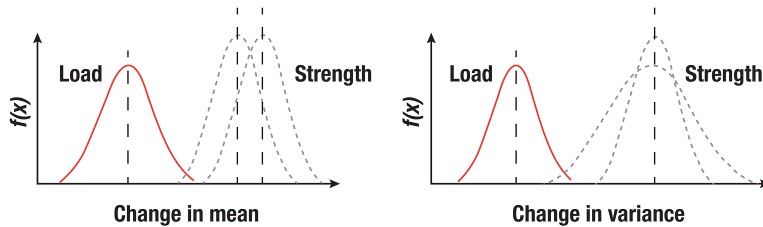

Figure 12.2 Influence of "quality" on failure probability.

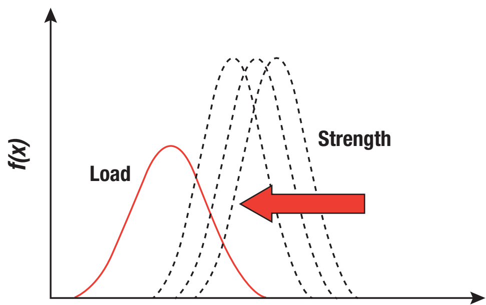

Figure 12.3 Influence of "durability" on failure probability.

the types of products under consideration. Sometimes, these procedures are presented as standards by acquisition agencies of dominant users. Larger original equipment and system manufacturers often develop and use own derating standards.

- The equipment or systems in which the parts are used are classified into categories according to their importance and criticality in the reliable functioning of the whole system. In military and space applications, this classification follows the criticality of the missions in which the systems are to be deployed. Rome Laboratories (Eskin et al. 1984) uses three derating levels, and so does Boeing Corporation (1996) and Naval Air Systems Command, Department of the Navy (1976).
- Sets of parts are identified as groups, which are believed to have similar responses to different environmental and operational stresses. It can be based on the material (Si, GaAs, and SiC), functional type of part (analog, digital, logic, and passive), technology (bipolar, field effect), types of packaging, and other considerations.
- For all such groups, stresses that possibly affect reliability of that group of parts are enumerated. Electrical and thermal stresses are used most often.

- For each type of part (or part class), derating guidelines are developed for the categories of equipments in which they are used for each such stress. These guidelines are usually one of the following types. Sometimes, a combination of these limits or coupled limits between a number of different parameters are used:
  - A percentage or fraction of the manufacturer specified limits
  - A maximum (or minimum) value
  - An absolute difference from the manufacturer limits.

The term "derating" by definition suggests a two-step process; first a "rated" stress value is determined from a part manufacturer's databook and then some reduced value is assigned. The "margin of safety," supposed to be provided by derating is the difference between the maximum allowable actual applied stress and the "demonstrated limits" of the part capabilities. The part capabilities as given by manufacturer specifications are taken as the "demonstrated limits." Sometimes, an upper limit on certain stress values is used irrespective of the actual manufacturer limit (some derated value of the manufacturer limit is used if that is lower than the derating limit). The propensity for the system design team inclines toward using conservative stresses at the expense of overall productivity. There are reasons to believe that the part manufacturers already provide safety margin while choosing the operating limits. When those values are derated by the users, it is effectively adding a second level of "safety margin."

The military standard, MIL-STD-975, was issued in 1976 and it remains the baseline for most derating guidelines in use by military and civilian users and manufacturers. Those guidelines have remained largely unchanged up to current versions, except for the addition of new part types and families. Although the intended purpose of derating was to provide a margin of safety, in reality, this has become a perceived guard against inadequate qualification and acceptance criteria. It is believed that the principal benefit of derating is not to extend the life of reliable parts but to protect against the presence of "maverick" parts and substandard lots. The design teams assume that the incoming parts will not fall within a narrow range of quality/performance characteristics. It is also assumed that these lower operational parameters will protect against random failures. Considering the arbitrary nature of such assumptions, a NASA Goddard Space Flight Center engineer recently noted that "It would be nice to be able to say the guidelines are all based on sound scientific and engineering principals and calculations but this does not seem to be so. Rather most are based on experience, best engineering estimate and a conservative philosophy. This should come as no surprise considering that many of the requirements are for simple quantities such as 50%."

12.2.1.1 Resistors The following is the typical resistor derating methodology.

- Maximum ambient temperature for use at full power (T_s)
- Maximum allowed operating temperature without power (T_{max})
- Generate "absolute maximum rating" curve (see Figure 12.4) 100% wattage and horizontal until T_s
  - Linearly connects T_s, 100% to $T_{\max}$, 0%

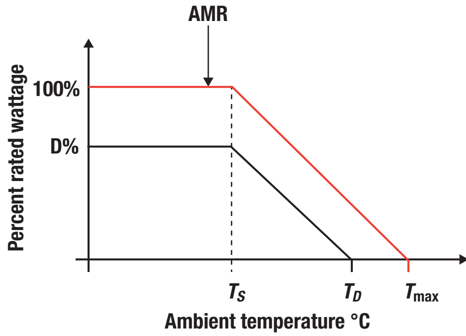

Figure 12.4 Generic resistor derating procedure.

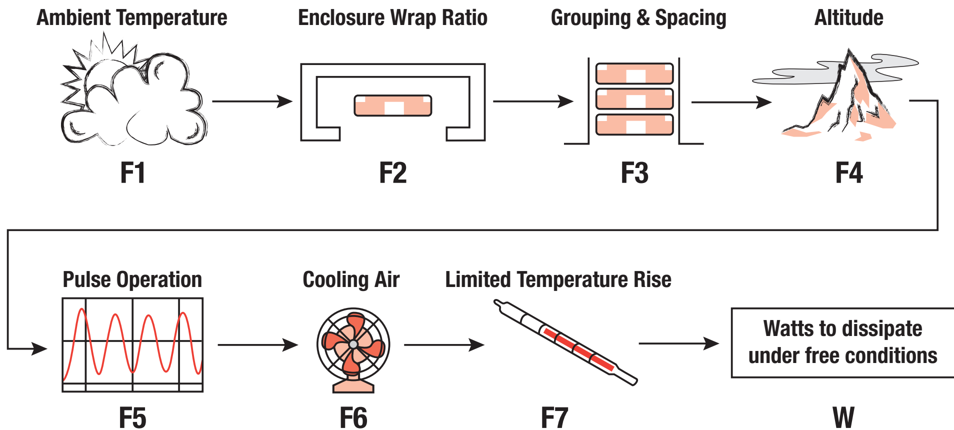

Figure 12.5 Ohmite's resistor recommendation.

- Generate "Derating Requirement" curve
  - Horizontal at % derated wattage until T_s
  - Linearly derated until maximum derated temperature, T_d

Selected resistor should have I2R rating of: W * F1 * F2 * F3 * F4 * F5 * F6 * F7 (see Figure 12.5). Note that the "F" values are multipliers, and some can be less than 1.

The maximum permissible operating temperature is a set amount. Any increase in the ambient temperature subtracts from the permissible temperature rise and therefore reduces the permissible power dissipation (Figure 12.6).

Enclosure limits the removal of heat by convection currents in the air and by radiation. The walls of the enclosure also introduce a thermal barrier between the air contacting the resistor and the outside cooling air. Hence, size, shape, orientation, amount of ventilating openings, wall thickness, material, and finish all affect the temperature rise of the enclosed resistor.

Figure 12.7 indicates for a particular set of conditions how the temperatures varied with the size of enclosure for a moderate size power resistor.

The temperature rise of a component is affected by the nearby presence of other heat-producing units, such as resistors, and electronic tubes. The curves (Figure 12.8) show the power rating for groups of resistors with various spacing between the closest points of the resistors, assuming operation at maximum permissible hot spot temperature.

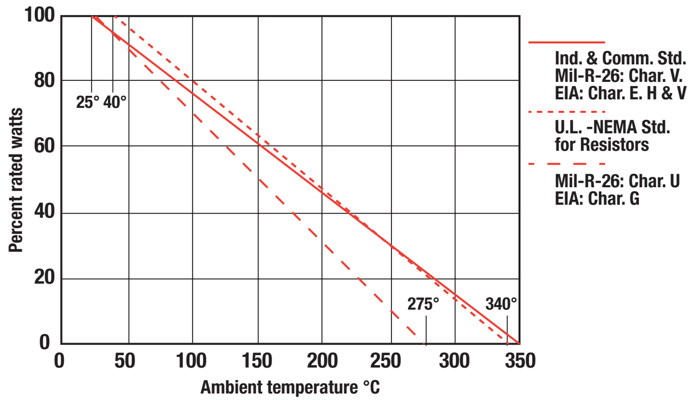

Figure 12.6 Derating of resistors for high ambient temperatures (Ohmite 2002).

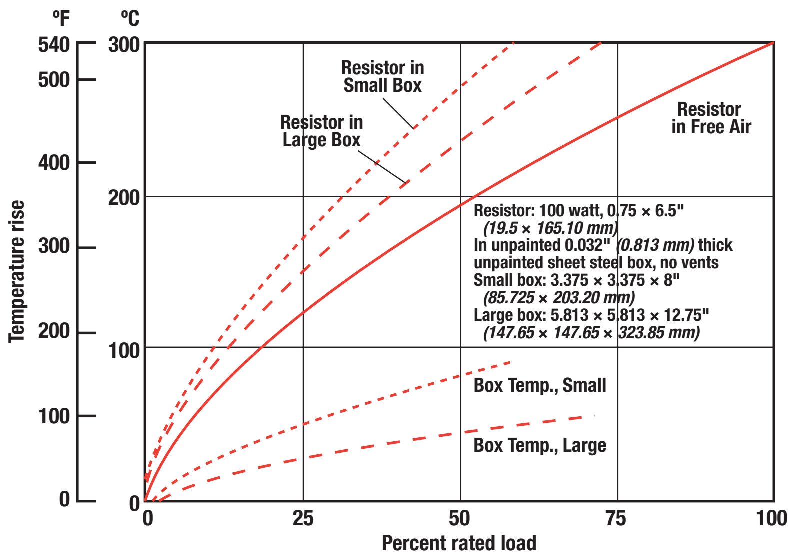

Figure 12.7 Example of effect of size of enclosure on temperature (Ohmite 2002).

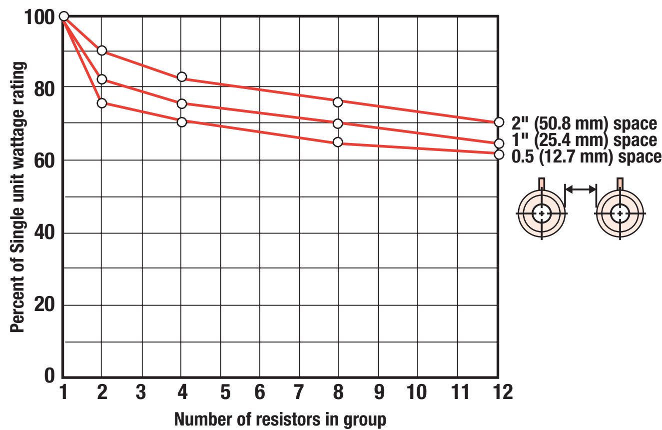

Figure 12.8 Derating of resistors to allow for grouping (Ohmite 2002).

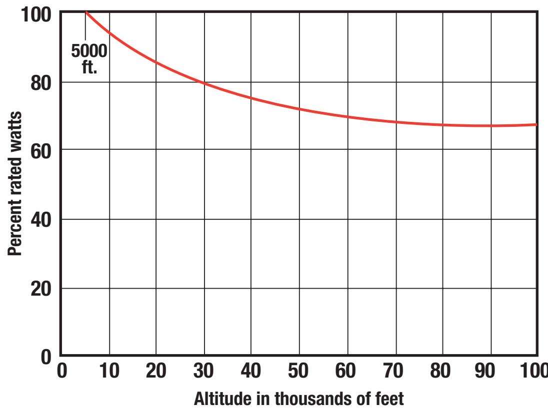

Figure 12.9 Altitude (Ohmite 2002).

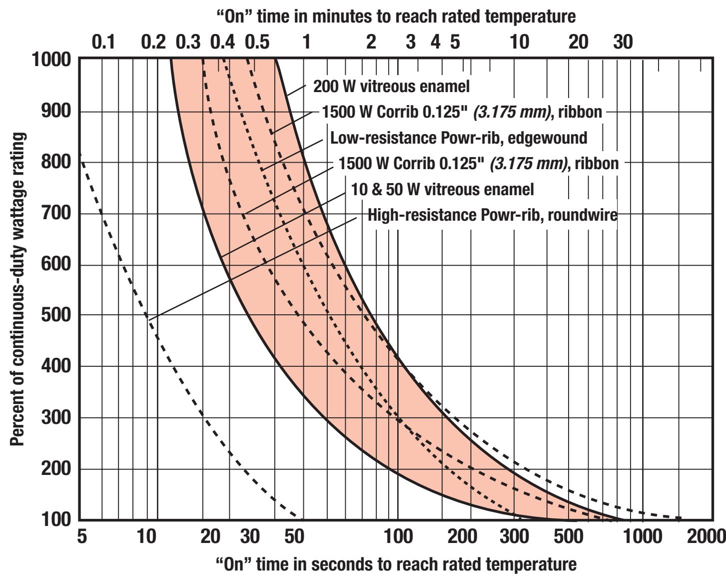

Figure 12.10 Time required for typical resistors to reach rated operating temperatures at various watt loads (Ohmite 2002).

The amount of heat which air will absorb varies with the density, and therefore with the altitude above sea level. At altitudes above 100,000 feet, the air is so rare that the resistor loses heat practically only by radiation (Figure 12.9).

Unlike the environmental factors, which result in reduction of the watt rating, pulse operation may permit higher power in the pulses than the continuous duty rating (Figure 12.10).

Resistors can be operated at higher-than-rated wattage when cooled. Forced circulation of air over a resistor removes more heat per unit time than natural convection does and therefore permits an increased watt dissipation. Liquid cooling and special conduction mountings also can increase the rating (Figure 12.11).

It is sometimes desirable to operate a resistor at a fraction of the Free Air Watt Rating in order to keep the temperature rise low. When it is desired to operate a

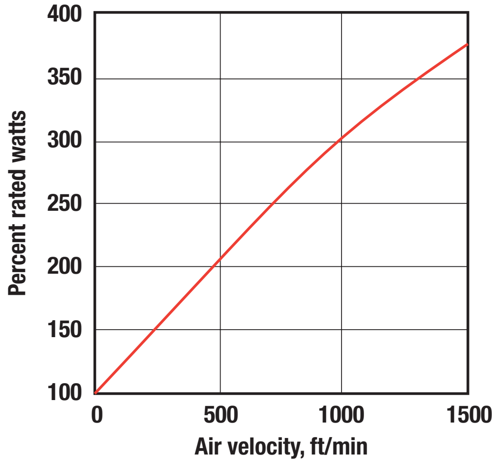

Figure 12.11 Cooling air.

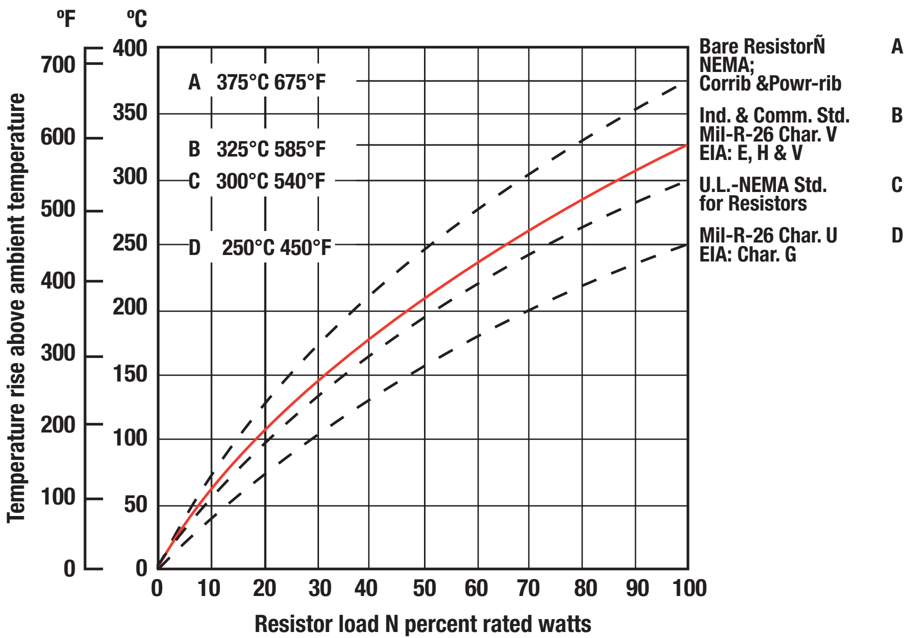

Figure 12.12 Limited temperature rise.

resistor at less than maximum temperature rise, the percent watts for a given rise can be read from the curve (Figure 12.12).

12.2.2 Limitations of the Derating Methodology

Thermal derating for semiconductor parts typically involves controlling the maximum steady-state temperature at some location in the system or part. For example, Naval Air Systems Command uses a set of thermal derating guidelines for electronic components, in which the junction temperature of TTL, CMOS, and linear amplifiers must be maintained 20°, 25°, or 30°C below of the manufacturers' rated junction

temperature depending on the criticality of the system. Besides trying to control the stresses directly resulting from them, limits on power, current, and voltages are also often used with a view to reduce the joule heating of the part. For example, NASA (Jet Propulsion Laboratory 1996) guidelines require an iterative process of alternately calculating junction temperature and reduction in power output until the junction temperature goes below its derated value. In fact, many military (Army, Air Force) and government (NASA) agencies, as well as industrial (Boeing) and consumer products (Philips, Siemens) manufacturers, either use or require use of some variations of these derating criteria. Unfortunately, there are several major shortcomings with this approach to derating which can render this process ineffective in achieving its purported goals.

12.2.2.1 Emphasis on Steady-State Temperature The use of temperature as the most common derating guideline follows the misplaced belief that steady-state temperature is the dominant cause of part failure, or deviation from rated performance. This outlook has been shaped by a belief system that dictates that "cooler is better" for electronics performance and reliability. There are serious doubts raised about this philosophy by many researchers (Evans 1988; Hakim 1990; Lall et al. 1997; Pecht 1996a; Pecht and Ramappan 1992a; Wong 1989). These articles clearly demonstrate the lack of faith in the steady-state temperature-based reliability prediction methodology. Although discredited by reliability community, statements such as "The reliability of a silicon chip is decreased by about 10% for every 2°C temperature rise" are still made by practitioners in the electronic packaging discipline (Yeh 1995).

Today's failure analysis reports rarely find equipment failures driven by part failure mechanisms dependent solely on steady-state temperature (Dasgupta et al. 1995; Hakim 1990; Lasance 1993). A significant portion of the failure mechanisms at the package-level is driven by thermal mismatch at mating surfaces of bimaterial interfaces (e.g., die fracture and wire bond fatigue) (Lall et al. 1995, 1997). Damage at such interfaces is actuated by thermal excursions due to environmental or power cycling. Just raising the steady-state temperature does not accelerate rate of failure at those interfaces, so lowering temperature is not going to increase reliability.

Temperature often manifests itself as a second order effect at bimaterial interfaces—as diffusion and intermetallic formation. Typically, diffusion is critical to formation of bimaterial joints at interfaces—at the same time, too much intermetallic often degrades the strength of an interface. The rate of intermetallic formation can be controlled by choosing appropriate mating metals, controlling surface roughness, and using barrier metals—such that interface strength after degradation is still greater than required to ensure field life.

Several chip-level failure mechanisms are also driven by localized temperature gradients. Temperature gradients can exist in the chip metallization, chip, substrate, and package case, due to variations in the conductivities of materials and defects such as voids or nonuniformities in metallizations. The locations of maximum temperature gradients in chip metallization are sites for mass transfer mechanisms, including electromigration. Current thermal derating practices do not provide any limits on these types of thermal stresses. Further, Pecht and Ramappan (1992a) found that the majority of electronic hardware failures over the past decade were not component failures at all, but were attributable to interconnects and connectors, system design, excessive environments, and improper user handling (Table 12.1). No amount of thermal derating is going to reduce the number of occurrences of these types of failures. Thus,

Table 12.1 Common failure mechanisms for microelectronics (Pecht and Ramappan 1992a)

Source of data	Year	The dominant causes of failure
Failure analysis for failure rate prediction methodology (Manno 1983)	1983	Metallization (52.8%); oxide/ dielectric (16.7%)
Westinghouse failure analysis memos (Westinghouse 1989)	1984–1987	Electrical overstress (40.3%)
Failure analysis based on failures experienced by end-user (Bloomer 1989)	1984–1988	Electrical overstress and electrostatic discharge (59%); wirebonds (15%)
Failure analysis based on Delco data (Delco 1988)	1988	Wirebonds (40.7%)
Failure analysis by power products division (Taylor 1990)	1988–1989	Electrical overstress damage (30.2%)
Failure analysis on CMOS (private correspondence)	1990	Package defects (22%)
Failure in vendor parts screened per MIL-STD-883	1990	Wire bonds (28%); test errors (19%)
Pareto ranking of failure causes per Texas Instruments study (Ghate 1991)	1991	Electrical overstress and electrostatic discharge (20%)

derating of individual parts without regard to the complete system can be unproductive and costly.

Performance improvement with lower temperature is another assumption made when one decides to lower the operating temperature for microelectronics. It is indeed true that certain performance parameters show significant improvement at lower temperature, but the temperature ranges at which those effects become prominent are generally well below even the derated temperature. In the normal operating range, it is well known that the performance parameters vary with temperature. For the system designer, it is important to understand the nature and extent of such variations. Deviations in electrical parameters from the required range can occur at both ends of the rated operating temperature range and using only lower-than-rated temperature is not going to provide improvement in performance. In addition, it was found by Hitachi that peak performance metrics of computers show a downward turn when there exists large temperature gradient within the system, even when the maximum temperature was maintained within rated limits.

It should be acknowledged that lower maximum temperature might also result in lower values of thermal gradients or lower temperature excursion, which might be beneficial for the system. Even in those situations, it is more advantageous and scientifically sound to directly find the limits on gradients and cycling rather than using lower temperature believing that to be a panacea. One can also employ better design techniques, which can ensure control on those thermal parameters even at relatively higher maximum temperature.

12.2.2.2 Is the "Right" Temperature Derated? There are many choices of the temperature, which might be controlled through derating in an electronic system. The logical choice of the temperature being derated should be the one at the location where temperature influences failure mechanisms and/or performance parameters. The electrical parameter variations are calculated based on the junction temperature of

semiconductor parts. For passive parts, temperatures at the winding, dielectric, coils, or other parts depending on the type of the part can influence the electrical characteristics. Depending on the packaging and attachment process of the passive parts to boards, temperatures at certain points can influence its reliability and/or electrical parameter. In the derating factor choices, the temperature most often derated is the junction temperature for the semiconductor parts; for other electrical parts, specification of operating temperature is also common. At close examination, it becomes difficult to identify the logic behind the choice of those temperatures.

The most common thermally driven (not necessarily by steady state temperature) driven reliability problems on the part or package are not at the junction; it is not even on the semiconductor. These problems occur at various parts of the package, including the traces, etallizations, and bond pads. As discussed in the previous section, other thermal stresses acting on areas other than the die affect the reliability more strongly. Lall et al. (1995) provide a thorough and detailed analysis of the temperature effects on part reliability with examples. They also show that how design modifications at different levels (from die to system) can suppress those failure mechanisms regardless of temperature rise. Attempts to improve the long-term reliability of a part through reduction of “junction” temperature cannot be justified from scientific standpoint.

This practice of derating of junction temperature would be defensible from an engineering standpoint if the temperatures within a part or package were constant or temperature differences were negligible. If there exists any significant temperature difference within a package, then it is important to identify the temperature pattern within such a package and to choose the valid temperature to derate. Junction temperature appears to be a reasonable compromise to use the junction temperature assuming that the temperature throughout the die, its metallization, and passivation is constant. Recent studies involving simultaneous solving of the electrical and thermal field equations (Fushinobu et al. 1995; Goodson et al. 1995) show that there are non-negligible difference in temperature between the different portions of the part. Fushinobu et al. (1995) found a hot spot in GaAs MESFET devices at the drain side of the channel. The temperature difference between the source and the drain can be in the same order of magnitude as the level of the derating of the junction temperature. In this situation, it will be futile to expect the derating to have any significant effect on the reliability.

The temperature discrepancies within the die, metallization, passivation, and other parts in close proximity of the die become more prominent in the Silicon-on-Insulator (SOI) devices. In SOI devices, the thermal resistance of the silicon dioxide is more than the bulk silicon in other devices. That makes the channel to substrate thermal resistance more than 10 times higher as compared with bulk devices. The value of that thermal resistance can be comparable to the thermal resistance of the package itself. The manifestation of this thermal resistance on the thermal map of the package is more complex than just a temperature difference. Due to all these variations related to junction temperature, Philips Semiconductors (1993) proposed that instead of the so-called junction temperature, a virtual temperature derived from the reliability data be used. This temperature would be analogous to electron temperature, which is used to describe the energy level of electrons.

As smaller and smaller (deep submicron level) devices are becoming more prevalent, another inadequacy of the emphasis on junction temperature usage becomes prominent. The transmission delay coming from the interconnects has become a major

contributor to the total delay time of a device (Nakayama 1993). The current derating practice does not take into account the effects of temperature changes on the electrical properties of the interconnects. Also, as many new specially developed and processed materials are being used in the parts, it becomes more important to identify the effects of temperature change on the electrical properties of those materials. Naem et al. (1994) describe the thermal effect on the resistivity of polycrystalline silicon titanium salicide, which is used in advanced CMOS parts to reduce sheet resistance. That study demonstrates the importance of taking into account the other seemingly nonthermal parameters in determining the thermal coefficient of resistivity (TCR). The effect of current density on TCR was very pronounced at higher current and temperature. The effects of manufacturing defects on the changes in resistivity also become more conspicuous at those conditions.

Another imprecisely defined and inadequately measured variable makes the calculation of junction temperature unreliable. That variable is the package thermal resistance; both case-to-junction and ambient-to-junction. The values used for these thermal resistances are critical when one attempts to control the junction temperature through derating. Unless this value is determined and used properly, the calculation for the required environmental temperature becomes ineffectual. Lasance (1993) provides a detailed account of the sources and levels of variability in the computation of junction to case and ambient thermal resistances. The nonlinear relationship of thermal resistance with power dissipation and temperature, and the dependence of this value to possible manufacturing defects makes it imprudent to use manufacturer-listed thermal resistance values under all circumstances.

We find that there are many variabilities associated with elusive junction temperature, in the methods of measuring and predicting that temperature, and the published data of package thermal resistances. With the target temperature at the junction being predicted imprecisely, it is difficult to determine what, if any, effect derating of that temperature is having on the reliability and performance of the part.

12.2.2.3 All Parts Are Not Equal A common but unfounded practice is to derate temperature for a technology (e.g., TTL and CMOS) or a packaging class (e.g., hybrids), based on the assumption that reliability of all such parts has similar dependence on temperature. While there are some 67 "major" semiconductor part groups (Ng 1996), each exhibiting fundamentally different characteristics, many derating guidelines use very similar (often identical) derating factors for a broad group of parts. For example, NASA sets an absolute limit on junction temperature for all digital microcircuits, and Naval Air Systems Command, Department of the Navy (1976) guidelines suggest identical thermal derating for all digital CMOS parts. However, there are many variations in the CMOS technology with respect to their temperature tolerance. For example, twin-tub CMOS technology is much more tolerant to latchup breakdown than bulk type CMOS parts; changes in doping profile can and do make many MOS parts resistant to hot carrier aging. Some newer derating guidelines (Boeing 1996) use more subgroups of parts while prescribing temperature limits. Still, due to the fast changes in part technology, it is difficult to derive deterministic derating guidelines for all new types of parts.

Even when the parts and packages are of similar technology in terms of material, scaling, and functionality, their performance parameters can be rather different over the same range of junction temperatures. A high speed benchmarking study (Maniwa and Jain 1996) found differences in the rate of change of propagation delay with

temperature for similar ASICs made by various major manufacturers. That fact, coupled with the variations in voltage and process derating, shows significant performance difference between manufacturers. If one is to apply derating to maintain performance parameters within the same level, that parameter should be different for different manufacturers. The current derating guidelines do not grant such options.

12.2.2.4 Derating Is Not without Cost Thermal derating of parts is not without cost. The reduction of temperature can require expensive, large, and heavy thermal structures or the addition of a cooling system. When active cooling is provided for a box as a reliability enhancement measure, that box will then be dependent on the reliable operation of the cooling system (Leonard 1991a). This reference provides a thorough critique of the present practice of using the temperature–reliability relationship. It also provides the readers the broad picture of the role of temperature in overall electronic part management. Reduction in junction temperature without active cooling can require expensive and time-consuming redesign with increase in size and weight of the system.

Possibly the most important cost of derating is in the sacrificed productivity of electronic system for the express purpose of reduction in junction temperature. One of the most common examples is the reduction of power dissipation in electronics parts, which in turn requires the reduction of speed of operation for MOS devices. If the reduction in temperature does not serve any positive purpose, then this loss of productivity is unjustifiable.

12.2.2.5 The Potential for Reduced Reliability and Performance Thermal derating guidelines have the potential for doing more damage than good, in that they give the design team a false sense of security about achieving increased reliability at lower temperatures. For example, lower temperatures may not necessarily increase reliability, since some failure mechanisms (e.g., hot electron) are inversely dependent on temperature. Other device failure mechanisms that show inverse dependence of temperature are ionic contamination when the devices are not operational, and reverse second breakdown (Lall et al. 1995).

Even when the low temperature itself is not the culprit causing failure, the process of achieving that lower temperature can have a serendipitous negative effect on the system. One such unintended consequence of actively lowering the operating temperature is the possible introduction of thermal cycling during the process of startup and shutdown of the system. This is observed in systems with unrealistically low constraints on the junction temperature—imposed by system designers with the perceived notion of improving reliability by lowering temperature. Boeing had to upwardly revise the maximum junction temperature for the Comanche light helicopter, because the initial low temperature limits were causing unique problems, such water deposit through condensation, besides initiating large thermal cycles at startup (Pecht 1996b).

It is generally assumed that the functionality of electronic devices improves at lower temperature. Although it is true for many parameters, there are situations in which performance parameters degrade at lower temperature. Huang et al. (1993) found larger degradation of current gain at lower temperatures while stress testing bipolar transistors. Some other phenomena, such as a large increase in leakage current in Poly-SOI MOSFET (Bhattacharya et al. 1994), logic swing loss in BiCMOS circuits (Rofail and Elmasry 1993), kink, and hysteresis (Kasley et al. 1993) occur at cryogenic temperatures. It should also be noted that for many performance parameters, there

are no good or bad values, only a range of acceptable values. If reduction in temperature pushes those parameters beyond acceptable range, then that reduction is detrimental. Threshold or starting voltage for MOS devices is a good example of such a parameter.

12.2.2.6 Interaction of Thermal and Nonthermal Stresses Thermal and nonthermal stresses do not act independent of each other in precipitating failure as is implicitly assumed in many derating guidelines. For example, temperature and current density accelerate electromigration. Lowering the current density reduces the dependence of equipment reliability on steady-state temperature—that is, benefitting in improved life due to lowered temperature in spite of the temperature dependence is often beyond the designed-for field life and is thus of little consequence. In case of electromigration, the site of failure is also dependent on both current density and temperature gradient. In general, part reliability, for mechanisms with a dominant dependence on more than one operating stress (temperature and nontemperature), complicated by dependence on magnitudes of manufacturing defects, can often be maximized more economically through methods other than lowering temperature.

The interaction of various thermal and nonthermal stresses modifies the dominant dependence of the failure mechanisms on one or more of the stresses. For example, temperature transients generated by the On/Off duty cycle (often expressed as a ratio of the on time to the total time ON/(ON+OFF)) modify the dependence of the metallization corrosion on steady-state temperature. At low duty cycle values, metallization corrosion has a dominant dependence on steady-state temperature due to the exponential acceleration of the corrosion chemical reaction rate. However, at higher values (in the neighborhood of 1.0), accompanied by the evaporation of the electrolyte, metallization corrosion has a dominant dependence on duty cycle and a mild dependence on steady-state temperature. Brombacher (1992) lists sets of stresses, which work in combination in affecting different failure mechanisms. Often, optimal sets of these parameters can be found that does not involve lowering of temperature.

12.2.2.7 Technology Improvements Are Not Reflected Recent developments of thermally tolerant active and passive part technologies are often not reflected even in the newer derating guidelines (Boeing 1996). McCluskey (McCluskey 1996; McCluskey et al. 1996) quotes other sources (Pecht 1994), which shows that the common thin-film resistors can operate at temperatures higher than 200°C. The temperature limits for resistors listed in the guidelines are much lower than that limit. McCluskey also reports development of capacitors that are reliable at temperatures much higher than the limits in most derating guidelines.

All semiconductor devices are derated for very similar maximum junction temperatures in derating guidelines. There are two basic flaws in adopting this type of guidelines. The first problem lies in the fact that these guidelines do not take into account the varied response of different semiconductor materials to temperature. The intrinsic temperature of silicon carbide is 1000°C as compared with 400°C for silicon. The second problem with this guideline is its conservative outlook, which does not take into account current developments in high temperature electronics (Bromstead and Baumann 1991; Dreike et al. 1994; McCluskey et al. 1996). These blanket limits imposed on all semiconductors by some of the derating guidelines is applicable to all types of devices irrespective of their individual technology and the architecture in

which those are used. This approach precludes the use of new technologies if the guidelines are strictly followed.

There remains another related issue, which needs to be clarified in this context. The main problem with the current derating methodology is in the approach taken, not with the exact numbers. The criticism of unjustified conservative values is not an attack on the exact values. Even if the capacities of the newer materials are reflected in the derating guidelines, unless the approach is changed, one can potentially continue to underutilize the capacities of the newer materials.

12.2.2.8 Summary of the Limitations of Thermal Derating Considering the preceding examples of various types of the shortcomings of the thermal derating process, we can conclude that this practice is deeply flawed. The process is not scientifically sound; besides, it can be costly, unproductive, and even harmful. The process also tends to put the users and manufacturers in a position of adversarial relationship instead of cooperation. In the worst case, it can become a blame allocation methodology. The process itself appears to be an isolated step in the design and use of microelectronics. The basic flaws in the process can briefly be summarized as the following:

- Overemphasizes steady-state temperature.
- The temperature the process tends to derate may not be the most important factor in determining performance and reliability.
- Does not recognize recent advances in semiconductor design and production methodology.
- Does not take into account the physical models developed to describe the failure mechanisms and performance parameters.
- Tries to safeguard individual parts without regard to their design, production and the actual use in circuitry.
- Groups parts together for similar derating in an arbitrary manner.
- Does not utilize the expertise of the manufacturers in determining the use profile of parts.

The process of component-by-component derating will make the system operate at a level that is presumed to be safe for the weakest link in the system. It is definitely a process where productivity is sacrificed in the quest of perceived safety. Even if the process does not harm reliability or performance, it is not wise to employ derating if it adds no value, reduces the functionality of systems, or increases the cost of operation.

12.2.3 How to Determine These Limits

We have seen in the previous sections that it might be necessary to find or verify the thermal limits on a device in certain conditions. Before any such task is undertaken, the user should verify through simulation/experimentation or both that the extended operational conditions or stricter performance requirements are truly necessary. Military electronic history is replete with cases of specifying unrealistically

harsher (Fink 1997) environment than necessary. As this task of determining new limits at the end of the user for complex circuitry can be expensive, this should be done only after verifying that there is reason to go beyond the manufacturer specified limits.

When one needs to set such limits, that methodology has to be based on scientific grounds. The salient features of this process are given in the next section.

The inputs required for deriving these stress limits are more comprehensive than the apparent simple choice of device type, and, in some cases, operating environment in the current derating methodologies. One needs to enumerate the desired mission life, device and system architecture (including material, geometry, and layout), required performance parameters, and the worst-case manufacturing defects for the particular manufacturer and device type. It also requires closer cooperation and open sharing of information between the manufacturers and users of devices.

The limiting values of steady-state temperature, temperature cycle magnitude, temperature gradient, and time-dependent temperature change, including nontemperature operating stresses are determined for a desired device mission life. Physics-of-failure are used to relate allowable operating stresses to design strengths through quantitative models for failure mechanisms. Failure models will be used to assess the impact of stress levels on the effective reliability of the component for a given load. The quantitative correlations outlined between stress levels and reliability will enable design teams and users to tailor the margin of safety more effectively to the level of criticality of the component, leading to better and more cost-effective utilization of the functional capacity of the component.

Theoretical, numerical, and experimental results on the effect of temperature on the electrical performance of devices are used to determine the thermal limits for maintaining electrical performance. Inclusion of the effect of temperature on the packaging and interconnect electrical properties will allow more accurate determination of such limits.

12.3 Uprating

Uprating is a process to assess the capability of an electronic part to meet the functional and performance requirements of an application in which the part is used outside the manufacturers' specified operating conditions. In an ideal world, there would not be a need for a book on uprating. The part manufacturers would supply parts with appropriate ratings, for all products and systems. One would not have to be concerned about using any part beyond its ratings. Also performance, cost, assembly, test, and obsolescence would not be factors of concern. However, the ideal world does not exist.

Electronic parts are most often commodity items. The profitability in this highly competitive market comes from economies of scale. Electronic manufacturers do not generally benefit from the creation of boutique parts, unless alternatives do not exist and a significant price premium can be charged. This has been exemplified by the decline of the U.S. military's qualified part list (QPL) and qualified manufacturer list (QML) procurement program; showing that niche parts with specific ratings, tests, screens, and documentation cannot be kept available by mandate.

This chapter provides the rationale for uprating and the role of uprating in the part selection and management process. Options are then given for situations where the ratings of a part are narrower than the application requirements.

Uprating is a process to assess the ability of a part to meet the functionality and performance requirements of the application in which the part is used outside the manufacturers' recommended operating range.[3]

Today's semiconductor parts are most often specified for use in the "commercial" 0–70°C, and to a lesser extent in the "industrial" −40 to 85°C, operating temperature range, thus satisfying the demands of the computer, telecommunications, and consumer electronics and their markets. There is also demand for parts rated beyond the "industrial" temperature range, primarily from the aerospace, military, oil and gas exploration, and automotive industries. However, the demand is often not large enough to attract and retain the interest of major semiconductor part manufacturers to make extended temperature range parts.

It is becoming increasingly difficult to procure parts that meet the engineering, economic, logistical, and technical integration requirements of product manufacturers, and are also rated for temperature ranges (Solomon et al. 2000). Yet there are products and applications that do require parts that can operate at temperatures beyond the industrial temperature range. It is desired that parts for these products incorporate technological advancements in the electronics industry in terms of performance, cost, size, and packaging styles.

For electronic parts, there is a limit of voltage, current, temperature, and power dissipation, called the absolute maximum ratings, beyond which the part may not be reliable. Thus, to operate in a reliable manner, the part must be operated within the absolute maximum rating.

There are also operational limits for parts, within which the part will satisfy the electrical functional and performance specifications given in the part datasheet. These ratings are generally narrower (within) than the absolute maximum ratings. A part may be used beyond the recommended operating rating but never beyond the absolute maximum rating.

Product manufacturers who perform system integration need to adapt their design so that the parts do not experience conditions beyond their absolute maximum ratings, even under the worst possible operating conditions (e.g., supply voltage variations, load variations, and signal variations) (IEC Standard 60134 1961). It is the responsibility of the parts selection and management team to establish that the electrical, mechanical, and functional performance of the part is suitable for the application.

Uprating is possible because there is often very little difference between parts having different recommended operating conditions in the datasheet. For example, Motorola notes (Lycoudes 1995) that "There is no manufacturing difference between PEMs (plastic encapsulated microcircuits) certified from 0 to 70°C and those certified from −55 to 125°C. The same devices, the same interconnects, and the same encapsulants are used. The only difference is the temperature at which the final electrical testing is done." In fact, many electronic parts manufacturers have used the same die for various

[3]Thermal uprating is a process to assess the ability of a part to meet the functionality and performance requirements of the application in which the part is used beyond the manufacturer-specified recommended operating temperature range. Upscreening is a term used to describe the practice of attempting to create a part equivalent to a higher quality level by additional screening of a part (e.g., screening a JANTXV part to JAN S requirements).

"temperature grades" of parts (commercial, industrial, automotive, and military). For example, Intel[4] (Intel 1990) stated in their military product data book: "there is no distinction between commercial product and military product in the wafer fabrication process. Thus, in this most important part of the VLSI manufacturing process, Intel's military products have the advantages of stability and control which derive from the larger volumes produced for commercial market. In the assembly, test and finish operations, Intel's military product flow differs slightly from the commercial process flow, mainly in additional inspection, test and finish operations."

Parts may also be uprateable for temperature because part manufacturers generally provide a margin between the recommended operating temperature specification of a part and the actual temperature range over which the part will operate. This margin helps maximize part yields, reduce or eliminate outgoing tests, and optimize sample testing and statistical process control (SPC). Sometimes, this margin can be exploited, and thus the part can be uprated.

12.3.1 Parts Selection and Management Process

Equipment manufacturers must have procedures in place for the selection and management of electronic parts used in their products. When uprating electronic parts, it is necessary to follow documented, controlled, and repeatable processes, which are integrated with the parts selection and management plans. The parts selection and management plan ensures the "right" parts for the application, taking into account performance requirements, assemblability, quality, reliability, and part obsolescence. The maintenance and support of some existing products require replacement parts to be available over the product life cycle. In the case of avionics, this period can be more than 10 years (Jackson et al. 1999a; Solomon et al. 2000). When companies stop producing avionics parts with wide recommended operating ranges, replacement parts become obsolete. One option is to use a "commercial" or "industrial" temperature range part as a substitute.

The performance assessment step of the parts selection management process assesses whether a part "will work" in its intended application. If the recommended operating condition in the datasheet of the part is outside the actual environment in the application, then options to mitigate this problem must be addressed.

12.3.2 Assessment for Uprateability

Uprating of parts can be expensive and time consuming, if there is no analysis of "promising" parts prior to the actual uprating process (Pecht 1996b). In other words, candidate parts should be accessed for their uprateability prior to conducting any uprating tests.

The best way to see if a part is uprateable is to obtain the simulation and characterization data from the part manufacturers. The data include product objective specifications, product and packaging roadmaps, device electrical simulation models, and temperature characterization data. Depending on the part manufacturer, some

[4]The Intel statement on the military and commercial parts shows that the practice of using the same die for various temperature ranges is common among manufacturers. In the mid-1990s, Intel stopped producing military temperature grade parts for business reasons.

of these data are available freely, while other may be available upon request, or in some cases, by signing a nondisclosure agreement.

Some datasheets or associated documents include electrical parameter data beyond the recommended operating temperature limits. These data may be useful in preassessing if a part can be uprated.

The effects of temperature (and other factors such as voltage and frequency) on different electrical parameters can be estimated using models available from part manufacturers. Often, the device electrical simulation models are made available to the public, although the models are often "sanitized" so that any proprietary information is masked (Micron Semiconductor 1998). Simulation models of devices can be used to calculate the effects of temperature variation on device parameters[5] (e.g., the BSIM3 model for short channel MOSFETs) (Foty 1997). Device simulations therefore can be used to estimate if the part will be uprateable, and what parameter changes may be expected at application operating conditions.

The cost of offering the models are minimal since they are developed during the design process. Circuit level models can be prepared in such a way that they do not reveal details of physical design. SPICE models are available from many companies, including Analog Devices, TI, and National Semiconductor. IBIS, VHDL, and VERILOG models are available from some companies such as Cypress Semiconductor and Intel. The model parameters can be examined to assess the effects of different factors on part electrical parameters over the target application conditions.

12.3.3 Methods of Uprating

Uprating is carried out after the part, the part manufacturer, and the distributors have been assessed (Jackson et al. 1999a, 1999b; Maniwa and Jain 1996) based on datasheets, application notes, and any other published data. Three methods for part uprating (see Figure 12.13) are overviewed in this chapter. The International Electrotechnical Commission and the Electronics Industry Association (IEC/PAS 62240 2001) accept these methods as industry best practices. Publications for the U.S. Department of Defense acknowledge these methods as effective and rigorous (Lasance 1993).

12.3.3.1 Parameter Conformance Parameter conformance is a process of uprating in which the part is tested to assess if its functionality and electrical parameters meet the manufacturer's recommended operating conditions over the target temperature range. Electrical testing is performed with the semiconductor manufacturer-specified test setups to assess compliance within the semiconductor manufacturer-specified parameter limits. The tests are of "go/no-go" type, and are generally performed at the upper and lower ends of the target application conditions. A margin may be added to the test, either in a range wider than the target application conditions or tighter electrical parameter limits for the test. The electrical parameter specifications in the datasheet are not modified by this method.

12.3.3.2 Parameter Recharacterization Parameter recharacterization is a process of uprating in which the part functionality is assessed and the electrical parameters are

[5]Different models provide different levels of details on the parameter estimates. Some examples of SPICE models are: Level 3, HSPICE, and BSIM3. Some SPICE versions allow the user to select the model to be used for transistor level analysis.

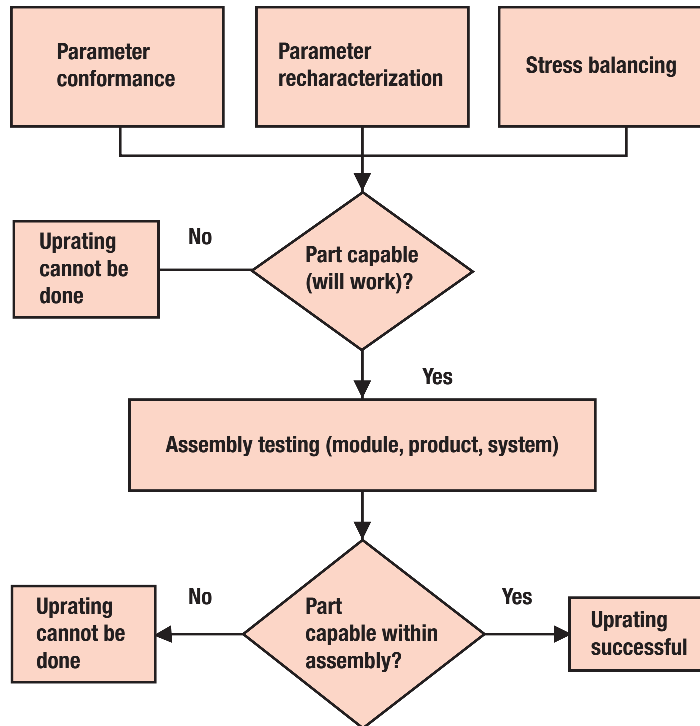

Figure 12.13 Approaches to thermal uprating of electronic parts.

characterized over the target application conditions, leading to a possible respecification of the manufacturer-specified datasheet parameter limits. The parameter recharacterization method of uprating seeks to mimic the part manufacturer's characterization process. The electrical parameter limits of parts rated for multiple temperature ranges are often obtained using the concept of parameter recharacterization (Pecht 1996b; Pendsé and Pecht 2000) and is shown in Figure 12.13. Electrical testing is followed by data analysis and margin estimation.

In parameter recharacterization, part electrical parameters are tested at several points in the target application conditions, the parameter values are recorded, and the parameter distributions are plotted. Figure 12.14 exemplifies the process. Here, propagation delay is on the horizontal axis and the population distribution on the vertical axis. Curve "1" is the distribution of the parameter at the manufacturer's specified temperature limit, and curve "2" is the distribution of the same parameter for the target application temperature limit.

The margin at the manufacturer-specified temperature range is the difference between the limit[6] of the distribution "1" and the specification limit on the parameter, PSPEC.[7] From distribution "2," at the target temperature, a new limit can be defined

[6]The limit may be chosen by the designers (e.g., 6-σ limit) as per the application yield and risk tolerances.

[7]Several factors influence the margin on a parameter, including the test equipment accuracy and confidence interval for the sample size. From distribution "2," at the target temperature, a new limit can be defined after adding a margin to it, and the modified parameter limit PNew can be obtained. One may chose to not modify the parameter limit, if the margin is still acceptable at the target temperature.

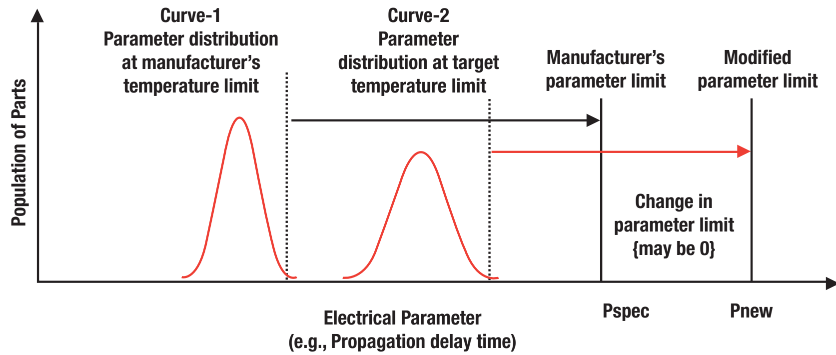

Figure 12.14 The parameter recharacterization method of uprating.

after adding a margin to it, and the modified parameter limit PNew can be obtained. One may chose to not modify the parameter limit, if the margin is still acceptable at the target temperature.

12.3.3.3 Stress Balancing Stress balancing is a process of thermal uprating in which at least one of the part's electrical parameters is kept below its maximum allowable limit to reduce heat generation, thereby allowing operation at a higher ambient temperature than that specified by the semiconductor part manufacturer (McCluskey 1996). The process assesses the possibility that the application may not need to use the full performance capability of the device, and that a power versus operating temperature trade-off for the part may be possible. For active electronic parts, the power temperature relation is:

$$T_J = T_A + P \cdot \theta_{JA}, \tag{12.1}$$

where T_J is the junction temperature, T_A is the ambient temperature, P is the power dissipation, and θ_{JA} is the junction-to-ambient thermal resistance. The performance of the part will generally depend upon the junction temperature. If the junction temperature is kept constant, then the temperature-dependent performance of the part should not change.

For a constant junction temperature, Equation 12.1 shows that higher ambient temperatures can be allowed if the power dissipation is reduced. However, the power dissipation of the part is often a function of some electrical parameters (e.g., operating voltage, and frequency), which will have to be changed. Thus, a trade-off can be made between increased ambient temperature and a change in some electrical parameter(s).[8]

The first step in stress balancing is to assess the electrical parameters, which can be used to change the power dissipation. The second step is to calculate the reduction in power dissipation required at the application temperature, by using the relationship given in Equation 12.1. The third step is to determine the changes in electrical parameters necessary to achieve the reduction in power dissipation. The fourth step is to conduct electrical tests to ensure the capability of the part to operate in the application environment with changed electrical parameters.

[8]Another option is to reduce the thermal resistance θ_{JA} of the part in the application, which may be achieved using heat sinks or providing cooling systems.

12.3.4 Continued Assurance

Part manufacturers provide product change notices (PCNs) for form, fit, and functional changes in their parts. However, change notices provided by the manufacturer do not necessarily reflect the changes in electrical performance that may occur beyond the recommended operating conditions in the datasheet. Thus, all changes need to be assessed by the part selection and management team for their possible effects on the use of parts beyond their manufacturer-specified recommending operating conditions in the datasheet. The changes in the parts that generally warrant a new uprating assessment include:

- Change in the temperature rating(s) of the part
- Change in power dissipation
- Changes in the thermal characteristics of the part, caused by changes in the package type, size or footprint, die size change, and materials
- Changes in the electrical specifications of the parts.

Semiconductor process changes (e.g., a die shrink, a new package, or an improvement in a wafer process) may or may not affect the part datasheet, but may affect the part performance beyond the recommended operating conditions. The changes in production sites may also result in changes in the uprateability of a part.

Specific protocols for reassessment of uprateability should be included in all uprating documents. For example, one may perform go/no-go tests on a sample of all incoming lots. Any failure of the part performance or deviation from the original lot (on which uprating was performed) will mean that uprating needs to be repeated.

Changes in the application of the part may also warrant reconsideration of the uprateability of the part. Changes in the environment, target temperature range, and system thermal design are factors that must be monitored. Electrical margin changes or part replacements (uprated or not uprated) may result in changes in system-level thermal interactions, and additional testing at the system level may be necessary.

12.4 Summary

Uprating is often possible because of the way electronic parts are designed and manufactured. The methods of uprating take into consideration the issues related to part testing and specification margins to assess the ability of the parts to operate over their target temperature range.

The uprating assessment of a part determines the electrical functional capability of parts in their target application conditions. This determines whether a part "can work" in a given environment. However, to determine if a part "won't fail" in the application environment, the reliability of the part needs to be determined for the application. The methods of determination of reliability can vary and may include assessment of manufacturers' qualification test results, additional tests performed by the equipment manufacturers, and virtual qualification.

Problems

12.1 Select any nonelectronic item of everyday use and identify the recommended operating conditions and absolute maximum ratings from its documentation (e.g., datasheet, product specification, and web literature). List both the ratings clearly with source.

12.2 List all the environmental and operational parameters for which you may be able to uprate a computer. You will need to refer to the specifications and the ratings.

12.3 Select a mechanical design item (e.g., gear, beam, shaft, and engine).

(a) List three stresses related to that items which can be derated to improve the reliability of the item.
(b) For each listed stress level listed in (a), discuss how you would derate those stresses to improve reliability of the mechanical item.
(c) Relate the stress level listed in (b) to the definition of reliability and explain what part of reliability definition is addressed in the derating plan.
  (i) Within specified performance limits: A product must function within certain tolerances in order to be reliable.
  (ii) For a specified period of time: A product has a useful life during which it is expected to function within specifications.
  (iii) Under the life-cycle application conditions: Reliability is dependent on the products life-cycle operational and environmental conditions.

12.4 Can you utilize the derating factors provided by a part manufacturer to design an accelerated test? Under what conditions can you make use of that information? Use an example from today's lecture notes to explain.

12.5 Find a datasheet on a nonelectrical product specifying AMR.

12.6 Find a datasheet on a nonelectrical product specifying the recommended operating conditions. Why are these not AMR?