Scaled CMOS Reliability and Considerations for Spacecraft Systems:

Bottom-up and Top-down Perspectives

Mark White
Jet Propulsion Laboratory, California Institute of Technology
Pasadena, California USA 91109
(818) 393-4173, mark.white@jpl.nasa.gov

Abstract*—The recently launched Mars Science Laboratory (MSL) flagship mission, named Curiosity, is the most complex rover ever built by NASA and is scheduled to touch down on the red planet in August, 2012 in Gale Crater. The rover and its instruments will have to endure the harsh environments of the surface of Mars to fulfill its main science objectives. Such complex systems require reliable microelectronic components coupled with adequate component and system-level design margins. Reliability aspects of these elements of the spacecraft system are presented from bottom-up and top-down perspectives.

Keywords - Scaled CMOS; component reliability; derating; thermal margin; spacecraft systems.

I. INTRODUCTION

Complex NASA flagship missions, including Spirit and Opportunity - the two Mars Exploration Rovers that landed on Mars in 2004 - and a much larger rover, Mars Science Laboratory (MSL) named Curiosity, scheduled to land later this year, were built to withstand the extreme radiation environments of space and the harsh Mars surface environments, including severe temperature extremes and extended thermal cycles, for long periods of time. The surface temperature in the Gale Crater, where Curiosity will land and maneuver, may vary from about -94°C to +28°C. An artist's depiction of Curiosity in the descent and landing stage is shown in Figure 1.

Figure 1. Artist's depiction of Curiosity in the descent and landing stage [1].

The research in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA).

The radioisotope power system on the MSL rover produces electricity from the heat of plutonium's radioactive decay. Although the power gradually decays, the Multi-Mission Radioisotope Thermoelectric Generator (MMRTG) has a minimum lifetime of 14 years. In Curiosity's planned two-year mission on Mars, the MMRTG will allow the rover to operate without the limitations of solar cells that provided power for the two MER missions. Curiosity contains ten scientific instruments and a sample acquisition, processing, and distribution system [1]. With ever increasing complexities of evolving spacecraft systems, such as those inherent to Curiosity, the reliability of these systems and components is essential to meeting the primary mission objectives.

This paper discusses some of the component-level reliability challenges from a bottom-up perspective, and how spacecraft systems must also consider reliability from a top-down perspective, especially with emerging technologies.

II. COMPONENT-LEVEL RELIABILITY: BOTTOM-UP PERSPECTIVE

A. *Scaled CMOS Reliability*

Curiosity is controlled by a module called the Rover Compute Element (RCE) buried deep within the Rover for maximum protection against the environmental conditions on Mars' surface. There are two RAD750 radiation-hardened CPUs (A-side normally on, B-side normally off but on standby should problems occur with the A-side) capable of up to 400 MIPS with 256MB of on-board DRAM, 2GB of Flash Memory - both with error detection and correction - and 256KB of EEPROM. While this amount of computing power and memory may only be roughly equivalent to that of a standard home computer, the critical components that make up this vital module, as with the other avionics modules, must have demonstrated reliability to fulfill the mission duration and environmental conditions. The microelectronics used in such systems, especially for critical applications, are often several generations behind leading edge technologies for several reasons: older technologies are generally more robust from a design standpoint; they operate at lower power densities; they have demonstrated and proven reliability; they have met rigid qualification programs on qualified production lines; they have process parameters optimized over a longer production history and therefore have narrower statistical process variation; they

have met and demonstrated optimized yield and quality control targets; and they have the benefit of many more parts and lots produced than on a new technology production run, thereby improving confidence on the process controls and overall reliability of the product.

Emerging technologies, including aggressively scaled complementary metal-oxide semiconductor (CMOS), bring new reliability challenges, new fabrication approaches, new failure mechanisms, more advanced packaging schemes, more potential failure points due to increased complexities and interconnect levels, and have limited production history. These aspects must be considered when assessing the reliability of such components in spacecraft systems.

Note that, due to the direct applicability and relevance to reliability issues with CMOS scaling, portions of the following discussion through Section D are extracted from, and may be found in greater detail in [2].

Decreasing the feature size of CMOS devices not only allows more components to be placed on a single chip, but it increases performance by allowing faster switching (or clock) speeds, with reduced power compared to devices with larger feature size. Some general scaling trends for CMOS are shown in Figure 2; the values are taken from [3]. The horizontal scale is metal-oxide-semiconductor field-effect transistor (MOSFET) channel length, not feature size. For advanced devices, channel length is approximately 65% of the feature size used in processing.

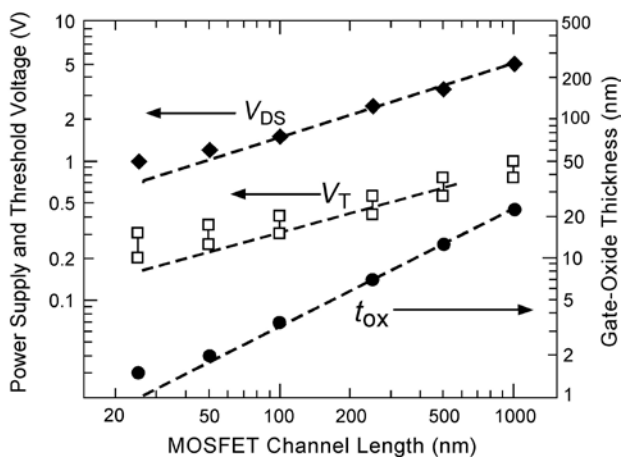

Figure 2. CMOS Scaling trends for power supply voltage, gate threshold voltage, and channel length [3].

Scaling theory originally had an objective of enhancing important circuit characteristics by decreasing effective MOSFET transistor channel length using constant electrical field strength as a guide. In this manner, reliability would not be excessively compromised, speed would increase, and power dissipation per function would decrease. In order to accomplish this, however, power supply voltage (and other important voltages within the integrated circuit) would need to decrease by the same scaling factor. This scaling scenario was effective for feature sizes on the order of 1 micron (with power supply voltages between 3.3 and 5 V), but fails when the supply voltage becomes comparable to the bandgap voltage. Newer scaling laws are more complicated, taking power dissipation and switching speed into account.

In addition, limiting power supply voltage (V_{DS}) decrease with smaller feature size, improves the inter-element spacing within individual transistors (the depletion width of a reverse-biased junction depends on voltage, and a lower voltage is required for reduced lateral spacing).

V_{DS} for advanced devices levels off at about 1 V, partly because of the need to maintain sufficiently large logic signals to provide noise margins compatible with other integrated circuits (which often operate at higher voltages), thus ensuring adequate design margin. Gate threshold voltage decreases with feature size, as shown in Figure 2. Although it is possible to reduce threshold voltage to about 0.25 V, higher values are required to be consistent with noise margin requirements as well as circuit requirements at higher temperatures. Most space systems require that digital and linear integrated circuits operate at junction temperatures below approximately 110°C, or 40°C below the manufacturer's maximum rating [4,5,6]. Power dissipation is a major concern for all integrated circuits. It often leads to higher junction temperature, which may result in sharply decreased reliability and lifetime. Package considerations, including thermal resistance, also have an impact.

B. "Front-End" Processing Reliability Issues

The most important issues related to front-end processing are those involving the gate: random dopant fluctuations (which affect threshold voltage), and the use of high-k dielectrics for applications that require the performance advantages of a thin gate oxide, but with lower gate leakage.

Significant process changes have become necessary to continue Moore's law scaling; examples include the use of hafnium oxide metal gate transistors (required by advances beyond the 65 nanometer technology node), and strained crystalline structure (required by advances beyond 90 nanometer technology node). The impact to the VLSI user is that new failure mechanisms and concerns are expected beyond the 32 nanometer technology node, and must be considered in manufacturer selection, flight lot qualification, and VLSI screening.

Time-dependent dielectric breakdown (TDDB), contact integrity, and hot-carrier degradation are typically less important. Note that hot-carrier degradation has negative activation energy, causing it to be more severe at low temperature. It may be of considerable importance for applications requiring extended operation at low temperature, such as surface exploration missions on Mars.

Although there is considerable focus in the literature on mechanisms associated with front-end processing, there is little that the end user can do to deal with them. Packaged devices do not provide direct access to internal transistors, limiting the ability to examine most of the mechanisms associated with front-end processing. For example, hot-carrier degradation is usually investigated with test transistors, applying much higher voltages to the drain and gate. Most packaged devices incorporate overvoltage protection that limits the maximum

voltage. There are other cases, such as DRAMs, where the internal voltage used for access transistors is derived internally; it is unaffected by the external power supply voltage. Most manufacturers investigate these issues thoroughly, and ensure that their design rules and processing technology provide adequate reliability margins. Therefore, the main emphasis should be on other aspects of reliability, particularly those that are unique to NASA space applications.

C. "Back-End" Processing Reliability Issues - Metallization

There are many possible failure modes associated with metallization. Some devices may use up to nine different metallization layers. Voids, grain boundaries, and thinning of metallization over non-planar regions may degrade reliability, along with non-uniformity of the multitude of vias that are required to make connections between the different metallization levels. These mechanisms are somewhat intimidating, because a part can still function properly with localized defects or geometrical deficiencies. There is no obvious way to detect such defects in finished devices. Changes in their characteristics during extended operating periods can result in catastrophic failure.

Electromigration is also an important mechanism, particularly for regions of the chips which include clock drivers and I/O circuits where higher currents are required. However, this is expected to be less important for devices used in space applications due to derating requirements that reduce the average current (note that dynamic current in CMOS scales directly with operating frequency).

Another issue is electromigration from vias in copper interconnects. Low-k dielectrics are used in more advanced processes, and metal from the contacts can migrate within the dielectric materials. This process is quite different from CMOS processes with larger feature sizes, which do not use the new insulator materials [7]. Recent information on via reliability from a 32 nm process shows that this reliability problem changes in character for highly scaled devices. That study showed that only some of the vias actually failed, but that the ratio of those that failed—the number of fatal defects—increased relative to the total number of vias, as the area of the via was reduced. The study also showed that if the power through the vias was too high, a transition from insulator to conductor could take place, even at much lower temperatures than the typical temperature required for such transitions. This is shown in Figure 3. It illustrates that new reliability problems can be expected as devices are scaled below the 45 nm node.

D. "Back-End" Processing Reliability Issues – Packaging

A number of package types are used in advanced VLSI/CMOS. Conservatism of the space community notwithstanding, advanced VLSI devices may not be available in traditional military packages. These military packages have a large database and heritage in space applications, and therefore, have lower risk in future space missions. VLSI parts in dual-side flatpacks and dual-inline packages (DIPs) are now becoming rare. Four-sided flatpacks (quad flatpacks) using hermetic (ceramic) technology are also becoming rare.

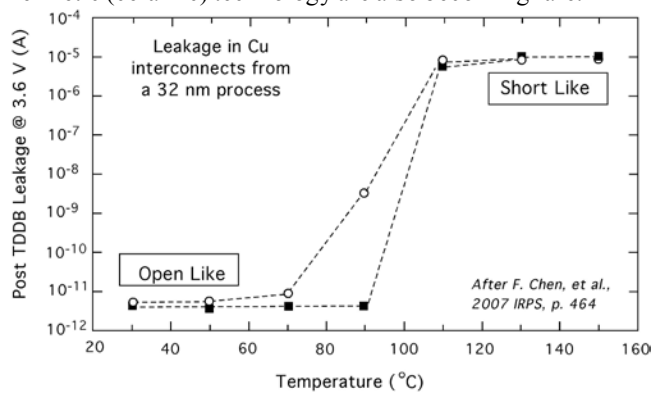

Figure 3. Metal-insulator transition in the low-k dielectric material used for interconnects in an advanced 32 nm process [7]. The leakage takes place between the materials used in multi-level metallization regions.

Commercial technology is driving toward a more I/O efficient usage of pinout. Area arrays (ball grid arrays, pin grid arrays, and column arrays) are becoming the industry standard. Since the commercial marketplace is driven by high volume and low cost, advanced technology devices may only be available in non-hermetic plastic packages, which may be built in large batches at lower cost and with more uniformity.

Reliability may be adversely affected since the plastic package is hygroscopic and therefore will allow moisture to transport through the plastic into the die region (including the bond wires). For this reason, the plastic must be optimized to minimize transport of chemically active (and deleterious) ingredients such as chlorides and ionic contaminants (sodium and potassium).

There is also more stress on the die within the package due to mismatch in the thermal expansion coefficient of the die and package. That not only limits maximum and minimum temperatures, but may also affect screening methods such as the temperature cycling traditionally used in burn-in and qualification. Temperature cycling is particularly important for Mars surface missions, which are exposed to (Martian) daily temperature extremes, unlike most other space missions.

Ball grid arrays are used for large-scale devices with high pin count. In such packages, the semiconductor chip is "flipped" and placed over a carrier with an array of solder balls. The balls are aligned with contact regions on the inverted chip. Hermetic packages are not available. More stress occurs at the corner regions of these packages compared to the center, which can cause cracking as well as intermittent or open contacts.

Column grids, which use small columns of material to establish contact between the package and device contacts, have also been developed. Column grid arrays provide a broader contact area compared to ball grid arrays, and are preferred for space products. However, they are only available for some devices.

Ball and column grid arrays provide an interesting conflict regarding testing and screening. One method to evaluate their reliability is to subject them to a series of thermal cycles, using

the number of such cycles before failure as a metric. Although this is probably a good way to evaluate these devices for a Mars surface application where daily thermal cycles occur, it may be ineffective for more conventional space applications where only small thermal cycles take place.

A different approach is to evaluate the thickness of intermetallic growth in contacts (at a constant temperature), which is a more likely failure mechanism for conventional space applications. Recent results for three different ball grid designs are shown in Figure 4 [8]. The results fit a diffusion model, and provide a better way to evaluate this particular failure mechanism compared to deep thermal cycling.

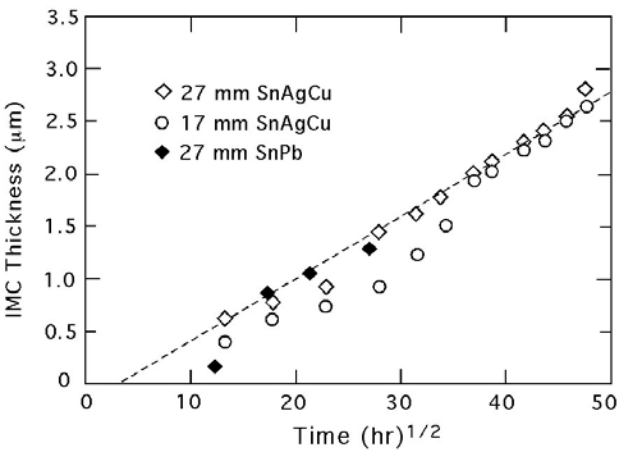

Figure 4. Increase in the buildup of intermetallic growth compounds in column grid arrays [8]. The results fit a diffusion model, and are more directly applicable to failure modes expected in conventional space applications where deep thermal cycling is not expected.

E. Deratin g

Reliability can be improved by limiting junction temperatures and other key operating stress characteristics of the device. As a component ages, the environment and operational stress can cause physical and material property changes that influence part performance, e.g., electromigration, time-dependent dielectric breakdown, friction-related wear, and package cracking. Such changes are often dependent on operating temperature, temperature extremes and rate of temperature change, however, other factors such as device manufacturing variation, device aging characteristics, electrical stress, timing, humidity, vibration, shock, radiation, and electromagnetic interference can also affect component performance depending on device type, construction and materials used. Component performance degradation is often accelerated when a part is used outside of its designed operating conditions or near its maximum rated parameters for long periods of time. Derating prevents small changes in operating characteristics from creating large increases in failure rate. The derating factor needed depends on the tolerance of the design to variation in operating parameters over the lifetime of the device in the expected operating environment. The full impact of environmental and operational variation on different device types can be difficult to quantify over time, thus device construction, material characteristics, physics-of-failure, manufacturing variability, design and performance margin, stress testing, aging characteristics, and application experience all play a role in determining acceptable derating margins. Limiting the junction or channel temperature and the electrical stress (power dissipation, output current) have proven to extend operating life and enhance reliability for most device types.

Power dissipation density is increasing in the more aggressively scaled VLSI CMOS technologies, as internal hot spots during ordinary operation are becoming more frequent. Furthermore, variations in thermal resistance in more advanced packaging (particularly plastic packaging, which is notorious for being a poor heat conductor) is expected to exacerbate this problem. Derating voltage is considered a risky strategy for advanced VLSI parts. Typically, several voltages are generated internally in such chips and many different types of transistors of varying geometries are used to gain performance advantage. Reducing the voltage even to the lower half of the recommended operating range may result in the unintended consequence of peculiar functionality in complex devices. Such oddities may not become obvious in all circuits until the mission is launched. Another approach is to derate the maximum operating frequency (historically 80%), but this too could lead to performance issues depending on the architectures with some advanced technologies, e.g., FPGAs and other complex VLSI devices. Certainly, derating frequency to 80% of f_{max} would be device and system tolerant dependent, but the active power dissipation would thereby also be reduced to 80% uniformly within the complex VLSI device.

III. SYSTEM-LEVEL RELIABILITY: TOP-DOWN CONSIDERATIONS

A. Design Margin and Reliability Analyses

System-level reliability is often determined by the amount of design margin in the constituent elements of the system and the use of different redundancy schemes. For spacecraft systems, the following analyses play a key role in the overall reliability assessment. They include the overall system-level thermal analysis, where boundary conditions of the thermal control surface are defined considering the mission environment; the electronic parts stress analysis (EPSA), which uses the maximum power dissipated in each component and the thermal resistance from junction-to-case; the worst-case analysis (WCA), where worst-case electrical parameter extremes are determined from the maximum and minimum predicted temperatures for all phases of the mission; the single event effects analysis (SEEA), to identify the severity of an SEE on the system for a given mission environment; and the failure modes, effects and criticality analysis (FMECA), a bottom-up analytical method performed either at the functional or component level to determine the probability of failure modes against the severity of their consequences. These combined analyses and adequate system design thermal margin improve overall system reliability.

B. Thermal Margin Stack-up

The level of conservatism in the system design thermal margin must be considered in aggregate. Beginning at the semiconductor junction level, to the component case and circuit board assembly, to the protoflight operating conditions and

predicted allowable in-flight system operational conditions; margin is cumulatively stacked-up to comprise a robust system design to improve reliability and reduce uncertainty. An illustration of an example of integrated thermal margins of the different elements is represented in Figure 5.

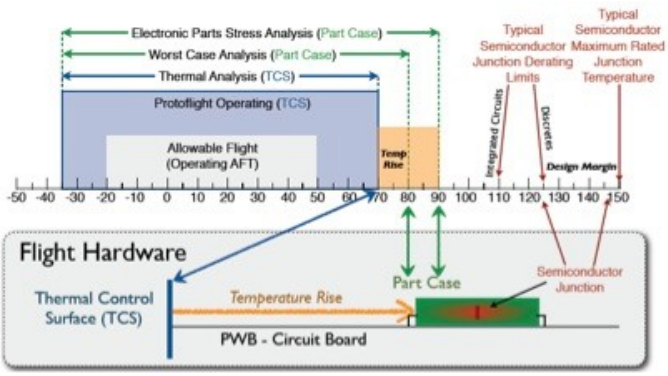

Figure 5. Illustration of integrated thermal margins [9].

C. *System Design, Thermal Management and Reliability*

System reliability can largely be improved through optimized thermal management and a deep understanding of the anticipated operating environment. This may be accomplished through in-depth knowledge and understanding of the operating stresses on the components; the material properties and their behavior over temperature; the potential failure mechanisms and physics-of-failure; by maintaining adequate process controls at all levels; by analyzing past performance (screening, qualification, and lot-to-lot variability); by understanding the performance degradation at steady state temperatures and/or extended thermal cycles; and by designing in adequate thermal margins to account for uncertainty.

IV. CONCLUSIONS

The new NASA flagship mission, Curiosity, is a transitional mission, using relatively few advanced microelectronic devices compared to missions with less robust reliability requirements. The main reason for this is the established reliability of older devices, produced on radiation-hardened process lines. However, advanced devices were used in specific cases, particularly memories, requiring new approaches for testing and qualification because they are fabricated on commercial processing lines.

New space missions will increasingly rely on more advanced technologies, partly due to cost, but also because of system requirements for higher performance, particularly in instruments and high-speed processing. This paper has discussed component-level reliability challenges with scaled CMOS in spacecraft systems from a bottom-up perspective. Some of the fundamental "Front-end" and "Back-end" processing reliability issues with more aggressively scaled parts have been discussed. Metallization concerns from increasing layers of metal, increasing numbers of vias and interconnects, electromigration from high current densities, and high junction temperatures from high power densities will remain a challenge with aggressively scaled devices.

More advanced packaging configurations (flip-chip, column grid array and ball grid arrays with very high pin counts) as compared to traditional dual side flat packs and dual inline packages, necessitate a reevaluation of traditional screening and qualification approaches for space applications. Although temperature cycling will remain a critical issue for Mars surface exploration missions, different approaches for reliability – such as the buildup of intermetallic growth compounds – may be more appropriate for conventional space missions that are exposed to only a limited number of temperature cycles.

Effective thermal management from the system-level to the component-level is a key element in the overall design of reliable systems. Both perspectives (top-down and bottom-up) play a large role in robust, reliable system design. Thermal management in space systems must consider a wide range of issues, including thermal loading of many different components on conduction cooled boards, radiation degradation of components, which may cause standby currents to increase, and the frequent temperature cycling of some systems, such as MSL. Conservative design practices are helpful, but they must be supplemented by radiation and reliability data for the wide range of microelectronic devices that are used on modern spacecraft.

ACKNOWLEDGMENT

The author would like to acknowledge M. Cooper and A. Johnston for their review comments.

REFERENCES

[1] Mars Science Laboratory (MSL) fact sheet 2011. Available on line at: http://mars.jpl.nasa.gov/msl/.

[2] M. White, M. Cooper, and A. Johnston, "Users Guide on Scaled CMOS Reliability," NASA JPL Pub 11-12 11/11, pp. 7-9 (2011).

[3] Y. Taur, et al., "CMOS Scaling into the Nanometer Region," *Proc. of the IEEE*, 85, pp. 486–504 (1997).

[4] M. White, M. Cooper, and Y. Chen, "Impact of Junction Temperature on Microelectronic Device Reliability and Considerations for Space Applications," *IEEE International Intergrated Reliability Workshop*, pp. 133-136 (2003).

[5] M. White, JPL Guideline: JPL Derating Guidelines, Rev. H, D-8545 pp. 15-16 (2010).

[6] K. Sahu, EEE-INST-002 Instructions for EEE Parts Selection, Screening, Qualification, and Derating, NASA/TP-2003-212242, Add. 1, p. M3-15 (2008).

[7] F. Chen, et al., "Critical Ultra Low-k TDDB Reliability Issues for Advanced CMOS Technologies," *IEEE Reliability Physics Symposium*, pp. 464–475 (2009).

[8] P. Lall, et al., "Leading Indicators of Failure for Prognostication of Leaded and Lead-Free Electronics in Harsh Environments," *IEEE Trans. Comp. and Packaging Technology*, 32(1), pp. 135–144 (2009).

[9] J. Klohoker, JPL Guideline: Reliability Analyses for Flight Hardware in Design, Rev. 4, D-5703 p. 45 (2010).