

Reliability Considerations of ULP Scaled CMOS in Spacecraft Systems

Mark White, Ph.D., NASA Jet Propulsion Laboratory, California Institute of Technology

Kristen MacNeal, NASA Jet Propulsion Laboratory, California Institute of Technology

Mark Cooper, Ph.D., NASA Jet Propulsion Laboratory, California Institute of Technology

Key Words: Reliability, ultra-low power, scaled CMOS, spacecraft systems

SUMMARY & CONCLUSIONS

NASA, the aerospace community, and other high reliability (hi-rel) users of advanced microelectronic products face many challenges as technology continues to scale into the deep sub-micron region. 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 larger scaled devices. Higher performance, and lower operating and stand-by power characteristics of Ultra-Low Power (ULP) microelectronics are not only desirable, but also necessary to meet low power consumption design goals of critical spacecraft systems. The integration of these components in such systems, however, must be balanced with the overall risk tolerance of the project.

1 INTRODUCTION

ULP Microelectronics are devices that have a milli- or micro-watt power consumption and a supply voltage of about one volt or less. The total power consumption is the summation of the switching and leakage power consumed.

$$P_{total} = P_{ac} + P_{dc} \quad (1)$$

These devices are used in multiple markets, including portable wireless devices requiring extended battery life, biomedical devices, energy management circuits, memory technologies, and remote space-based sensors. ULP microelectronic architectures can significantly reduce the power required for digital processing, and they are important to achieve power design goals for spacecraft systems.

2 ULP MICROELECTRONICS

2.1 Goals

The use of ULP microelectronics increases the ability to achieve performance goals of advanced spacecraft systems. They help to achieve a reduction of overall system power, and offer greater functionality and enhanced performance. Lower power demands also allows major elements like solar arrays, spacecraft buses, batteries and heat sinks to do less work, therefore reducing mass and size.

2.2 Power Consumption

Overall power consumption is a top concern for devices used in space applications because of system power limitations. Power consumption is a major restriction in chip design due particularly to static leakage. Changes necessary to reduce power consumption may be at the circuit level, as well as with process parameters. Process changes to modify the threshold voltage reduce current drive and therefore, power consumption. Sub-threshold circuit operation is one approach to lowering power consumption. Total power dissipation is calculated with the sum of both dynamic (switching) power and the static (leakage) power dissipation.

$$P_D = (K * C_L * V_{dd}^2 * f) + (I_l * V_{dd}) \quad (2)$$

In this case, K is the switching factor, C_L is the average node capacitance loading of the circuit, V_{dd} is the supply voltage, f is the switching frequency, and I_l is the DC leakage current. P_{AC} dominates for circuits with high activity, while P_{DC} dominates for circuits with low activity.

2.3 ULP Benefits in Space Applications

ULP electronic architectures can significantly reduce the power required for digital processing as seen in Figure 1. Implementation of ULP digital microelectronics can lead to load power reductions of 20-40% or more, which result in either mass savings or increased power available for other functions [1]. The best percentage power reductions will be for relatively small, moderate bandwidth spacecraft with passive sensors. ULP technologies enables large amount of low power on-board digital processing such as image compression, as seen in Figure 1.

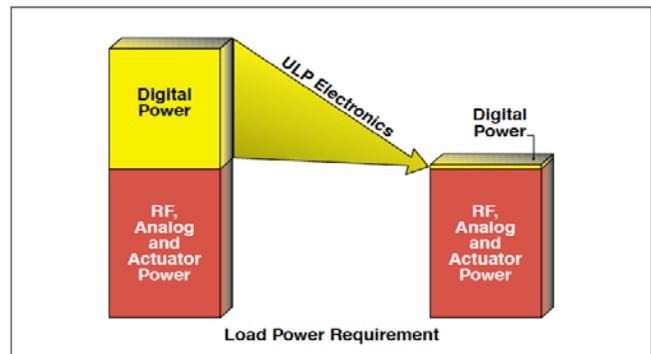


Figure 1. Improvements in digital processing load power requirements with ULP electronics. [1]

There are design trade-offs with ULP devices. As ULP microelectronics continues to scale, designs become more sensitive to temperature and statistical variations in threshold voltage. Process variation, particularly at low transistor voltages, results in timing delay spread and decreased noise margin. In addition, static leakage power is increased because transistors are mostly biased in the sub-threshold region. This leakage current is highly temperature dependent. Threshold voltage variation of more than 100mV is possible, which has impact on threshold sensitive circuits such as DRAM sense amps. To minimize the effect of these issues, more on-chip error correction is built-in, symmetric layouts for flip-flop circuits are employed, and internal supply voltages are more tightly controlled, with the use of on-chip voltage converters. As long as temperature and noise margin limitations can be accommodated, the benefits of ULP microelectronics are significant.

Low power systems have been implemented in spacecraft systems. For example, NASA's three Space Technology 5 (ST5) micro-satellites completed their planned 90-day mission in 2006. A lithium-ion power system for small satellites was used and it required a low-mass Li-Ion battery with triple junction solar cells. Another demonstration of ULP technology is the CMOS Ultra-Low Power Radiation Tolerant (CULPRit) technology [2]. This technology utilized a Reed-Solomon channel encoder that operated at 0.5 Volts (0.35um CMOS technology). It was designed to greatly reduce power consumption while achieving a radiation tolerance of 100 kRad total dose with single event upset (SEU) immunity of 40 MeV linear energy transfer (LET). Over 300 million telemetry frames were processed without a single failure.

Recently developed CubeSats are a class of research spacecraft called nanosatellites. NASA has selected 20 small satellites to fly as auxiliary cargo aboard rockets that have launched or are planned to launch in 2012. These cube-shaped satellites are approximately four inches long, have a volume of about one quart, and weigh 2.2 pounds or less. They are expected to conduct technology demonstrations, educational research, and science missions.

3 SCALING TRENDS AND BRANCHES

Scaling theory originally held the objective of enhancing important circuit characteristics by decreasing effective MOSFET transistor channel length with the constraint of constant electrical field strength. In this manner, reliability would not be excessively compromised, speed would increase, and power dissipation per function would decrease. In order to accomplish these performance improvements, CMOS transistor voltages, and consequently the power supply voltage, would need to decrease by the same scaling factor. However, for a variety of circuit design constraints, these decreases have not occurred at all technology nodes.

During the last several years, a wider range of scaling rules has been developed, recognizing that different tradeoffs are needed for various end-products. For example, DRAMs need lower standby current and lower gate leakage compared to

high-performance CMOS devices used in microprocessors and fast static RAMs. Consequently, thicker gates are needed for the internal pass transistors used in DRAMs compared to other CMOS processes. Table 1 shows general comparisons of the effects of scaling on these technologies [3].

Application	Primary Design Concerns	Power Supply Voltage	Gate Threshold Voltage	Reliability Issues
High performance	Switching speed	< 1V	0.3	Electromigration Contact integrity
Mainstream	Maximum functionality density	1 to 1.2V	0.4	Percent defect ratio of electrical test
Low power	Low total power dissipation	1.5 to 3.3V	0.25	Percent defect ratio of electrical test
Memory	Low leakage and low standby current	3.3V (internal back bias generator are often used to decrease leakage current)	0.5	Retention time for dynamic memories; early bit failures in large density SRAMs

Table 1. Scaling Trends for Different CMOS Applications.

Furthermore, modern CMOS technologies usually provide at least two different gate thicknesses, one for internal logic, which often operates at lower voltage with internally generated power, and one (or more) with thicker gate oxides for input/output (I/O) voltage that is compatible with 2.5 or 3.3 V external signals.

4 RELIABILITY CONSIDERATIONS

A recent publication from Intel [4] divides reliability issues associated with front-end processing into two categories: historical sources of variability, such as lithography, line roughness, and oxide thickness; and emerging sources of variability that are a direct result of scaling devices to very small dimensions. The main sources of variability are shown in Table 2. Most of the historical issues have been solved by clever improvements in processing. For example, the difficulty of coping with corner rounding and geometrical limitations associated with lithography has been overcome by designing devices in pairs for feature sizes of 65 nm and below.

Historical Process Variations	Emerging Process Variations
<ul style="list-style-type: none"> • Patterning proximity effects • Line edge and line width roughness • Surface roughness (polishing) • Variations in gate oxide thickness • Fixed charge and oxide traps 	<ul style="list-style-type: none"> • Random dopant fluctuations • Variations in implants and anneals • Variations associated with strain • Gate material granularity

Table 2. Effect of Processing Variations on Reliability.

Another key issue is that of device complexity. Large-scale devices contain extremely large numbers of transistors. Special techniques, including on-chip error correction and selective elimination of regions with defective devices that are identified during initial wafer probe tests, are used to improve manufacturing yield. Thus, there are aspects of the design that are quite different from older circuits. Complex circuit design, e.g., modern DRAMs, which are effectively small systems, must be taken into account from the standpoint of reliability.

4.1 Front-End Processing

The most important issues related to front-end processing are those involving the gate: random dopant fluctuations, discussed earlier, and the use of high-k dielectrics for applications that require the performance advantages of a thin gate oxide, but with lower gate leakage [3].

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 nm technology node), and strained crystalline structure (required by advances beyond 90 nm technology node). The impact to the VLSI user is that new failure mechanisms and concerns are expected beyond the 32 nm technology node, and must be considered in manufacturer selection, flight lot qualification, and VLSI screening [3].

Time-dependent dielectric breakdown (TDDB), contact integrity, and hot-carrier degradation are typically less important. However, we should note that hot-carrier degradation has a 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 the Moon or Mars.

Although there is considerable focus on mechanisms associated with front-end processing in, there is little that the end user can do to deal with them. 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 [3].

4.2 Back-End Processing

"Back-End" processing reliability issues include those related to metallization layers and vias, advanced packaging techniques, electrical testing, and end-user reliability testing.

There are many possible failure modes associated with metallization, particularly for processes that may use up to nine different metallization layers. Voids, grain boundaries, and thinning of metallization over non-planar regions are contributing factors, along with 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 [3].

Electromigration is also an important mechanism, particularly for regions such as 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. Recent information on via reliability from a 32 nm process shows that this reliability problem changes in character for highly scaled devices [4]. The 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 2 [5]. The figure illustrates that new reliability problems can be expected as devices are scaled below the 45 nm node [3].

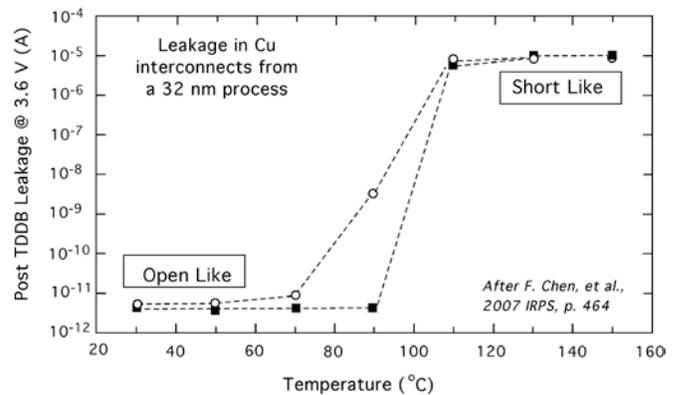


Figure 2. Metal-insulator transition in the low-k dielectric material used for interconnects in an advanced 32 nm process. The leakage takes place between the materials used in multi-level metallization regions.

5 DRAM RELIABILITY IMPACT

An important type of memory often used in flight computer avionics is Dynamic Random Access Memory (DRAM). DRAM is an important component of future low power spacecraft architectures and ULP digital processing avionics. DRAM is a type of volatile memory that needs to be periodically refreshed to retain its contents. A typical DRAM cell consists of a Metal Oxide Semiconductor Field Effect Transistor (MOSFET), that acts like a switch and a storage capacitor. Commercial scaled DRAM is a leading memory technology driver with market pressure to reduce cost per bit, as well as static and dynamic power consumption.

4.1 Types of Errors

With each new DRAM generation, there is a growing concern of increased DRAM faults or bit errors.

4.1.1 Hot Carrier Injection

An important type of intrinsic failure mechanism for DRAM is HCI, which has a negative activation energy,

meaning that the rate of reaction decreases with increasing temperature. S. Baeg, et al., showed a particular sensitivity to the AC-to-DC factor (ADF) of the DRAM. ADF sensitivity was reviewed under three parameters: device degradation, effective drain voltage in the word-line driving circuit, and access frequency [6]. It was found that ADF values decreased with increasing degradation, and that the AC lifetime was reduced with decreased ADF. The ADF was most influenced by reducing effective operating voltage. The timing margin in the word-line signal was more related to the change in ADF than the change in degradation. These failure issues were observed in numerous products with back-to-back accesses to one word-line circuit by system software [6].

There were five failures classified by Cisco Systems in their DRAM analyses of HCI [7]. The first type of failure occurred at the global word-line circuit, which was degraded due to a pumped voltage (V_{ccp}). This resulted in slower high transition of the word-line signal [7]. The second type of failure occurred mainly due to slower low transition of the word-line signal, which then affected the remainder of the operations. The third type was the same as the second failure; however, the HCI affected word-line driver increased threshold voltage. The fourth type of failure occurred at storage cells, which caused an increase in leakage from the affected cells. The last failure occurred at the voltage pumping circuit to generate the pumping voltage. As the transistor degraded, the V_{ccp} could not produce enough current, which caused an increase in V_{ccp} voltage. This last failure was under a DC type of HCI stress, and did not occur if the system stayed on [7].

4.1.2 Correctable and Uncorrectable Errors

There is some correlation between correctable and uncorrectable errors. If a correctable error occurs, the likelihood of an uncorrectable error increases [8]. The number of correctable errors also increases with operational life. B. Schroeder, et al. demonstrated that when looking at the errors per DRAM DIMM, there was no individual memory technology that performed better over the other (comparing DDR1 and DDR2). Increased temperature was correlated with higher correctable error rates but did not have a large effect on memory errors. Utilization also showed an increase in memory errors [8].

4.1.3 Non-recurring and Recurring Faults

DRAM experiences different types of faults and errors. 70% of errors are recurring, while the other 30% are non-recurring or transient. Non-recurring faults are also referred to as transient faults. They cause incorrect data to be read from a memory location until the location is overwritten with correct data. Once a transient fault has been repaired, the probability of errors to follow does not change [9]. Recurring faults cause memory locations to continuously return incorrect data. If a device experiences a recurring fault the likelihood of it experiencing another is increased [9].

4.2 Improvements in DRAM

4.2.1 Chip-Kill

Chip-kill is a more advanced version of ECC (error correcting code). It is known to result in 36x reduction in DRAM uncorrected error rate. It protects systems from multi-bit errors at any location on the memory chip. A hardware scrubber reads memory locations and corrects any correctable errors present. A controller also logs information about the locations of uncorrected errors. This information is helpful because it can show if a certain node is experiencing recurring faults and if that location experiences multiple errors [9].

4.2.2 Voltage Scaling

Self-dynamic voltage scaling is a method of optimizing the operating conditions without excessive power consumption in DRAM. It is accomplished by using a self-dynamic voltage scaling (SDVS) method that changes the internal supply voltage in response to the process skew and operation frequency. It allows for a greater timing margin in wider process variations, which helps reduce the power consumption. Two design techniques are used. The first is an adaptive bandwidth delay-locked loop (DLL). This allows for larger delays in times at low operating frequencies, and smaller delays at higher operating frequencies. The second technique is an adaptive clock gating. This reduces the switching power by more than 25% without changing the operating frequency or performance [10]. In order to overcome process variations, the DRAM changes the internal supply voltage. The SDVS can correct the performance of integrated circuits by changing the operating condition when it detects delay changes [10].

4.2.3 Self-Refresh

There is a proposed self-refresh scheme that automatically extends the self-refresh period by monitoring the number of failed cells using ECC for a 45nm embedded DRAM. The amount of self-refresh current is over three times the amount of leakage current. To reduce data-retention power consumption, self-refresh schemes, a temperature sensor, and ECC have been used. After tests, the increase and decrease in self-refresh period depend on the total number of failed cells in a refresh period. The self-refresh period increases when the number of failed cells is less than a given level, and the period decreases when the number is larger than this level. This is then repeated until the self-refresh ends. This scheme can reduce the power consumption by choosing an optimal self-refresh period regardless of process, voltage, and temperature variations. If failed bits are detected, the addresses are registered and the failed data are written back after they are corrected by ECC. If by the end of the self-refresh operation the number of failed cells does not exceed a specific level, the refresh period will be extended a fixed amount of time. At higher temperatures, the power consumption is larger than the conventional scheme. However, -10°C power consumption is reduced by 95%. At room temperature, power consumption is reduced by 86% [11].

4.2.4 Tiered Reliability Mechanism

Current chip-kill implementation schemes create numerous restrictions on system configuration, and suffer from wasted energy, wasted memory, reduced energy-efficiency, increased storage overheads, and increased circuit complexity. A new chip-kill implementation has been introduced by Udipi, et al., Localized and Tiered (LOT)-ECC, where multiple levels of localized and tiered error protections are employed [12]. LOT-ECC splits protection into multiple tiers so that it can detect errors, isolate their locations, and reconstruct them. The first layer of protection in LOT-ECC is local error detection (LED). The LED code checks every read operation to verify data fidelity. It also identifies the exact location of the failure. Layer two has Global Error Correction (GEC), which is used to aid in the recovery of lost data once layer one detects the error. The LED code is meant to pick up on single-bit and double-bit errors, row-failures, column failures, pin failures and chip failures. LOT-ECC captures all occurring faults while providing power and performance benefits. Compared to a modern commercial chip-kill implementation, LOT-ECC saves up to 44.8% memory power (static + dynamic) and reduces average memory access latency by up to 46.9%, in addition to a reduction in ECC circuit complexity, for a minor 14% additional storage overhead [12].

5 SPACECRAFT ENVIRONMENTS

5.1 Temperature Effects

Some applications require that parts be used well beyond the normal region where they are designed and tested by the manufacturer. For example, there are many cases where parts have to be used at low temperature, such as Mars surface missions, missions to cold bodies (such as asteroids), and in applications on conventional spacecraft outside the normal electronic enclosures, where lower temperatures are encountered.

The first issue that must be dealt with is that of determining whether the part can actually function properly at temperatures well below the normal design range. Laboratory characterization tests and modeling can be used to determine this. The second issue is that of packaging. A number of mechanisms are involved, including the mismatch of thermal expansion coefficients of leads, feed-throughs, the semiconductor die, and the package. Test and qualification methods need to be developed that take these mechanisms into account. It is important to distinguish between missions where the part is only cooled once, and cases where it must be frequently cycled between low and moderate temperatures (such as the Mars surface). Test and qualification methods are quite different for these two scenarios.

5.2 Soft Errors

Soft errors, or non-permanent failures, continue to be a concern for high-density RAM products, particularly those that are cosmic-ray neutron-induced soft errors. While single event upsets can usually be handled with error correction,

multiple-bit upsets are more difficult to deal with and for highly scaled devices, a single ion can produce more than 100 upsets [13]. Single event functional interrupts (SEFI's) generally have a lower probability of occurrence than ordinary upsets, but they can be difficult to detect; it is becoming increasingly difficult to detect SEFI's with highly scaled, high density products. Radiation hardening by design can reduce SEE effects, however, it often comes at a high cost and a much larger chip area.

Device scaling effects on soft errors are complex. The critical charge decreases with scaling, increasing the susceptibility, while the device area is lower with scaling, thereby decreasing susceptibility. The net effect is that the upset rate per-cell is not affected much, however, the upset rate per device increases because there are many more devices per chip.

ACKNOWLEDGEMENT

The work in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. Government sponsorship acknowledged.

REFERENCES

1. M. Freeman, "Impact of Ultra Low Power, Rad-Hard Electronics on Spacecraft," JHU/APL Final Report, (1999).
2. P. Yeh, G. Maki, "CMOS Ultra Low Power Radiation Tolerant (CULPRIT) Microelectronics," NASA GSFC/U. of Idaho. Presentation, (2003).
3. M. White, M. Cooper and A. Johnston, "Users Guide on Scaled CMOS Reliability," NASA JPL Pub. 11-12 11/11, (2011).
4. K. J. Kuhn, et al., "Process Technology Variations," *IEEE Trans. Elect. Dev.*, 58 (8), pp. 2197-2208 (2011).
5. F. Chen, et al., "Critical Ultra Low-k TDDDB Reliability Issues for Advanced CMOS Technologies," *IEEE Reliability Physics Symposium*, pp. 464-475 (2009).
6. S. Baeg, H. Nam, P. Chia, S. Wen, and R. Wong, "AC-DC factor sensitivity for DRAM components lifetime under hot-carrier injection," *Reliability Physics Symposium (IRPS), 2011 IEEE International*, pp.2D.1.1-2D.1.4, 10-14 April (2011).
7. S. Baeg, P. Chia, S.J. Wen, and R. Wong. "DRAM Failure Analyses under Hot-Carrier Injection." Cisco Systems Inc., (2011).
8. B. Schroeder, E. Pinheiro, and W.D. Weber. "DRAM Errors in the Wild: A Large-scale Field Study." *Communications of the ACM* 54.2 (2011): 100-07. *ACM Digital Library*. ACM, Inc.
9. V. Sridharan and D. Liberty, "A field study of DRAM errors," in *Silicon Errors in Logic System Effects*, (2012) Proceedings, Urbana-Champaign, IL. 6p.
10. H.W. Lee, et al., "A 1.6V 1.4Gb/s/pin consumer DRAM with self-dynamic voltage-scaling technique in 44nm CMOS technology," *Solid State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, vol., no., pp.502-504, 20-24 Feb. (2011).
11. S.S. Pyo, et al., "45nm low-power embedded pseudo-SRAM with ECC-based auto-adjusted self-refresh scheme," *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, vol., no., pp.2517-2520, 24-27 May (2009).
12. A.N. Udipi, et al., "LOT-ECC: Localized and Tiered Reliability Mechanisms for Commodity Memory Systems." University of Utah, (2012).
13. P. Dodd and L. Massengill, "Basic Mechanisms and Modeling of Single-Event Upset in Microelectronics", *IEEE Trans. on Nucl. Sci.*, 50 (3), (2003).

BIOGRAPHIES

Mark White, Ph.D.
Electrical Parts Engineering and Reliability
Jet Propulsion Laboratory
Pasadena, CA, 91109 USA

Email: Mark.white@jpl.nasa.gov

Dr. Mark White is presently Group Supervisor in the Electronic Parts Engineering Office at the Jet Propulsion Laboratory in Pasadena, CA. He holds Ph.D. and M.S. degrees in Reliability Engineering from the University of Maryland, as well as an MBA from University of Colorado and a B.S. in Electronics Engineering. Mark has Professional Certificates in both Engineering Management and Reliability Engineering. His research interests include physics-of-failure based microelectronics reliability, modeling and lifetime prediction, and advanced photonics. Before joining JPL in 2000, he was a senior member of the technical staff in the Electrical Parts Engineering Office at Lockheed Martin.

Kristen MacNeal
Electrical Parts Engineering and Reliability
Jet Propulsion Laboratory
Pasadena, CA, 91109 USA

e-mail: Kristen.M.Macneal@jpl.nasa.gov

Kristen MacNeal is currently attending University of California, Irvine and is intending to graduate June 2013. She is obtaining her degree in Biomedical Engineering. Her research at JPL involves low power microelectronics, specifically DRAM. She is also interested in robotics, medical electronics.

Mark S. Cooper, Ph.D.
Electrical Parts Engineering and Reliability
Jet Propulsion Laboratory
Pasadena, CA, 91109 USA

Email: Mark.S.Cooper@jpl.nasa.gov

Mark Cooper received PhD in Physics from University of California at Berkeley and advanced degrees from MIT in electrical engineering. He has worked at Aerospace Corporation, GTE Sylvania, Litton Guidance and Control Systems, before coming to Jet Propulsion Laboratory. At Litton he developed a program to utilize Plastic Encapsulated Microelectronics in military electronic systems equipment that has resulted in many systems with the same field failure rate as previously seen in similar systems using military standard parts. His current interests include evaluation and qualification of commercial technologies for space systems, derating photonic components, development of photonic system components for space applications.