
Sean Keller, Alain J. Martin, Chris Moore

California Institute of Technology
Pasadena, CA 91125, USA &
Situs Logic
Pasadena, CA 91106, USA
{sean,alain,cc}@async.caltech.edu

Abstract—This paper describes DD1, an asynchronous radiation-hard 8-bit AVR® microcontroller (MCU) implemented in TSMC 40LP, a low-power bulk 40nm CMOS process. Designed for extreme reliability, DD1 uses quasi-delay-insensitive (QDI) asynchronous logic and contains full-custom radiation-hard memories and logic cells. The chip was found fully functional on first silicon over a range of operating voltages from near-threshold (500mV) to above the nominal $V_{DD}$ (1.1V). It qualifies as both ultra-low power (<100µW/MHz) and radiation-hard by design. At 550mV the MCU operates at 1MIPS with a power consumption of 18µW/µW/MIPS. At 1.1V it runs at 20MIPS consuming 75µW/MIPS (1.5mW total). After extensive testing, it was found to be total-dose and latch-up immune and has an upset immunity of 2E-6 SEE/device-day (CREME96 geosynchronous near-earth orbit).

I. INTRODUCTION

In order to build energy efficient digital CMOS circuits, the supply voltage must be reduced to near the device threshold voltage (NTV) [1]–[9]. Problematically, due to random parameter variation, supply scaling reduces circuit robustness. Moreover, the effects of parameter variation worsen as device dimensions diminish, further reducing robustness [10], [11]. Furthermore, dimensional scaling also increases the probability that high energy particle strikes (i.e., radiation) cause functional failures [12]. As devices shrink, it is increasingly difficult to build energy efficient reliable systems, and QDI logic is an excellent approach to tackling this problem because of its almost complete independence of delay [13], [14]. This paper presents a full-custom QDI microcontroller designed specifically for high reliability at low operating voltages in a modern CMOS process.

In order to mitigate the effects of soft errors induced by radiation, the MCU (called DD1) uses duplicated logic cells (see [15]), and duplicated SRAM cells (see [16]) for memories and registers. In order to facilitate voltage scaling in the face of parameter variation, the logic cells utilize at most four FETs in series in both the pull-down and pull-up networks (see [10] for an in depth discussion of transistor stack design for low-voltage operation). Nonetheless, parameter variation still has a significant impact on robustness and reliability. As $V_{DD}$ is reduced DD1 eventually fails to function as a result of parameter variation; two failures modes occur: (1) gates and memories either fail to change state or do so erroneously, or (2) timing assumptions are violated [11]. In QDI circuits, timing violations are only possible at isochronic forks, and the probability of these failures can be mitigated by increasing the length of the adversary paths [13], [14]. DD1 utilizes adversary paths with no fewer than five gates in series.

The remainder of this paper is organized as follows. Section II introduces the design of DD1. Section III discusses the approach used to increase MCU robustness against the effects of radiation. Section IV discusses physical device testing and the results of real silicon measurements. Section V presents a deep dive into low-power QDI circuit robustness to parameter variation and estimates the probability of failure due to both gate switching errors (due to parameter variation and noise) and timing failures due to adversary path violations. Section VI discusses related work and Section VII concludes the paper.

II. DESIGN OVERVIEW

The DD1 architecture consists of four main components: core, instruction memory (IMEM), data memory (DMEM), and input/output (IO). The DD1 core decomposition includes four main QDI subcomponents: register file, peunit, decode, exec unit (see Figure 1 for details). QDI technology is characterized by a fine-grain decomposition of the pipeline into small modules communicating by four-phase handshake protocols. The typical target module for decomposition is the precharge half buffer or PCHB (see [13]). Standard techniques for manual and semi-automated decomposition were used to generate the core subcomponent PCHB decompositions.

Figure 2 depicts three different instructions under execution. Figure 2a depicts an ALU instruction under execution; note that the decode block has been optimized to determine the absence of branching far more quickly than the full decode can be completed. Figure 2b depicts the execution of a branch instruction; here the execution block has been optimized to quickly produce the branch decision. Figure 2c depicts the execution of the uncommon return instruction, which is slow to both decode and execute. The payoff for this delay is lower energy and latency for more common instructions.
DD1 implements the AVR® reduced core instruction set architecture [17]. DD1 has an 8-bit-wide datapath with 16-bit instructions (a few are 32-bits). The fabricated prototype has two on-die memories (SRAM)—a 2KB instruction memory and a 256B data memory—and a 32x8-bit general-purpose register file. The IOs consist of 16 general-purpose IOs and a timer unit. DD1 achieves both ultra-low power and radiation hardness by combining the use of asynchronous QDI technology (including novel radhard by design techniques) with advanced bulk CMOS technology. Cell sizing for high noise margins, body biasing, and a virtual $V_{DD}$ (for memories) were used to achieve reliable low voltage and ultra-low power operation.

The chip uses five different power supply rails that can be adjusted to modulate speed and power of individual subcomponents. One rail supplies power to the register file, another rail to the rest of the core. The other three rails are used to body-bias the PFETs. One rail controls the register file N-well potential, another the instruction and data memories, and a third the core.

III. RADIATION ROBUSTNESS

Vulnerability to radiation effects is a growing problem as feature size and operating voltage decrease and operating frequency increases. An energetic particle hitting a chip may contribute to 3 types of malfunctions: total ionizing dose (TID), single event latchup (SEL), or single-event upset (SEU). In total ionizing dose effects, the accumulation of charges in the dielectric may change the device threshold voltage or create a conducting path between the source and drain of a transistor. In a single-event latchup, a spurious current created by a particle hit triggers latchup. In a single-event upset, a particle strike causes a temporary bit flip; if the erroneous bit is latched or stored in memory, a permanent error follows. (See, e.g., [18].)

Many applications require both robustness to radiation and low power. But these two requirements conflict: robustness benefits from large devices and a high $V_{DD}$; power efficiency necessitates the opposite. Hence, there are few, if any, radiation tolerant devices in advanced (40nm and smaller) technology. The MCU described in this paper makes use of the most advanced bulk CMOS technology available at the start of the project. Combined with the inherent energy efficiency of QDI logic (no clock network or PLL, automatic shut-off of idle parts, simple voltage scaling), the small feature size and the chosen high $V_t$ help achieve ultra-low power in spite of the overhead of any radiation-hard-by-design approach. On the other hand, the small critical charge ($Q_{crit}$) of this technology, and the use of bulk CMOS make the design more vulnerable to SEEs. The challenge was to rely on the efficiency of the QDI-specific radiation-hard-by-design method, and to utilize new statistical methods to achieve high noise margins in the presence of significant random parameter variation.

The QDI solution for SEU detection and correction used in this project is very different from that of clocked logic. It relies on a “persistence” property unique to QDI: when a change of input values of an operator (i.e. a QDI gate) causes a change of value on the output of that operator, the input values will persist until a feedback path causes the inputs to change. Both detection and correction of an SEU are realized by duplicating gates at an appropriate level of granularity, as shown conceptually in Figure 3 [15]. Multiple upsets that would affect both gates of the duplication are avoided by physical separation of the two halves of the duplicated logic. Lastly, latchup is highly unlikely to occur when operating near-
threshold [3]; however, since latchup can cause catastrophic failure, N-well regions and P-well regions are isolated by walls of well diffusion, contacts, and metal.

Fig. 3: Conceptual representation of the duplication scheme for SEU detection and correction in QDI random logic. The two rectangles represent 2 identical components with inputs and outputs duplicated. The two circles marked C are Muller-C elements. The actual implementation is more complex.

The two SRAMs and the register file are protected against SEUs by the Dual Interlocked Storage Cell (DICE) implementation of the bit cell [16]. The IMEM is partitioned into 8 banks of 128 16-bit words each. Access to different banks can overlap by pipelining the request through a distribution tree. At 600mV, simulation results show a dynamic energy usage of 0.43pJ per read at the bank level, and a leakage power of 42nW for a single bank. The read latency is 17.3ns.

The DICE SRAM is robust against multiple time-separated upsets, since errors are self-corrected rapidly upon their occurrence. A DICE cell is protected against simultaneous multiple upsets by splitting the cell so that any number of bit upsets can happen in one half without affecting the other half or permanently changing the state of the cell.²

IV. SILICON TEST RESULTS

After successful tapeout (Figure 4 depicts an actual photograph of a die), DD1 was subjected to a variety of performance and reliability tests. The performance measurements in Figures 5 and 6 correspond to a voltage sweep of all packaged chips from 540mV to 1.2V with an N-well voltage bias of 300mV. The minimum energy point is at 550mV at 1MIPS and 18pJ/instruction, which confirms the energy optimality of NTV operation (see [6], [19]). The maximum performance point is 1.2V at 27 MIPS and 80pJ/instruction.

Cobalt-60 radiation testing was performed at AFRL for TID, and heavy-ion testing at the Berkeley BASE cyclotron for single event effects (SEE) and latchup. DD1 is TID immune up to and including 1Mrad(Si). (Time constraints prevented further testing.) The changes in operating frequency and active power consumption due to TID were small. Leakage current increased but total leakage remained negligible at under 10uA after 1Mrad(Si) exposure and did not affect functionality.

For heavy ion testing, the DUT ran an encryption/decryption program over random data. As the program exercised all aspects of the chip, the generated cross-section represents errors/device at full utilization. Fluences³ of up to 1e8 were used to generate data, and recorded data points consist of tens of upsets. The chips are latchup immune up to and including an effective linear energy transfer (LET) of 196 MeV/(mg/cm²), and qualify as radhard with 2e-6 SEE/device-day.

Fig. 4: Die photo of the prototype chip in 40nm CMOS. Area 1 is the logic; Area 2: the register file; Area 3: the DMEM; Area 4: the IMEM. The pad-frame is 2.4mm x 2.4mm. The core area is 0.37mm².

Fig. 5: Measured plot of energy-per-instruction vs. cycle time for 22 chips across a voltage range from 540mV to 1.2V.

²No means of physical separation can protect against all upsets; e.g., simultaneous particle strikes on each half of a cell can cause a failure.

³The fluence is the number of radiative particles to strike the chip during a test window.


V. ROBUSTNESS ANALYSIS

Two goals—minimize power and maximize reliability—drove the design and implementation of DD1. In order to achieve these goals and to maximize the likelihood of functional first-silicon, three conservative (and critical) design assumptions were made early in the project:

- the minimum energy operating point falls between 600mV and 700mV,
- if all adversary paths contain at least five gates in series, timing will not be violated, and
- by using combinational gates throughout (with the exception of SRAM) the chip will be robust.

The primary goal of this section is to analyze and quantify these assumptions in the TSMC40LP technology as applied to the MCU detailed in this paper. The organization of this section is as follows. Section V-A discuss parameter variation in modern CMOS processes. In Section V-B, the near-threshold model (developed in [4]) is used to determine the minimum-energy operating point as a function of activity factor. Section V-C uses the near-threshold statistical delay model from [4] along with the adversary path timing assumption (see [14]) to estimate the probability of a timing failure in the DD1 core. Finally, the DD1-core robustness is estimated in Section V-D (see [11]), and the probability of functional failure is compared to the probability of timing failure.

A. Parameter Variation

In modern CMOS technologies, device parameters such as channel length, oxide thickness, dopant concentration, etc. can have significant deviations from their nominal values due to process-induced and intrinsic parameter fluctuations [20]. Process variability can be considered a global, predictable, and gradual skew in device characteristics introduced by the complexity of manufacturing chips [21] (e.g., from thermal gradients during fabrication [22]). Intrinsic parameter fluctuations are truly statistical in nature and cause significant deviations from device to device within a chip. Intrinsic variations can be attributed to atomistic effects (e.g., random dopant fluctuation (RDF)) and device structure variations (e.g., line edge roughness (LER)) [20], [21], [23]. There are a number of different ways to characterize and partition these effects, and the approach used in this paper is to consider a global component wherein all devices on a chip are affected in the same way, and a local component wherein each device on a chip has a number of statistical parameters drawn from distributions with mean values set by the global skew. This style of partitioning variation is not as accurate as a full combined statistical model, but it is a good, albeit slightly pessimistic approximation [21].

Considering variation in terms of a global and a local component simplifies statistical analysis, and global variation can be almost completely mitigated by way of body biasing the various subsystems of DD1 (see Section II). All variation analysis is performed at the TT (typical-typical) corner, because the vast majority of die can be biased to the TT-corner (all tested die could). For circuits operating subthreshold, the local component of variation is dominated by RDF and is accurately modeled by normally distributed uncorrelated device threshold ($V_t$) variation [24]. Near-threshold, local variation does exhibit some degree of spatial correlation, and at the process-nominal $V_{DD}$ spatial correlation is significant and cannot be ignored. This increase in the spatial correlation of local variation as a function of $V_{DD}$ can be attributed to the fact that channel-length variation has little effect on devices operating subthreshold but becomes the dominant effect at approximately twice the threshold voltage [24]. Channel length variation is spatially correlated between devices within some radius, and is straightforward to model [21], [24], [25]. Given that DD1 was designed and optimized for near-threshold operation, local parameter variation is treated as random and uncorrelated. Furthermore, SPICE simulations, along with foundry-provided statistically-extracted BSIM4 models, are used throughout this paper as a basis for correctness; these models are considered accurate over the entire device operating range [26].

B. DD1 Minimum Energy Operating Point

From the delay model derivation in [4], the FO4 delay of a minimum size inverter in TSMC40LP can be calculated and plotted as a function of $V_{DD}$ (as depicted in Figure 7). Compared to BSIM4 SPICE simulation the mean absolute error is 7.2% and the maximum absolute error is 18%.

From [4], the leakage current ($I_{off}$), of minimum-size devices in TSMC40LP can be plotted and compared to BSIM4 SPICE simulations as depicted in Figure 8. The NFET $I_{off}$ mean absolute error is 4.1% and the maximum absolute error is 13%; the PFET $I_{off}$ mean absolute error is 4.0% and the maximum absolute error is 1.6%.

Finally, using the energy model derived in [4], the minimum energy operating point for DD1 can be estimated. DD1 operates at approximately 3,000 FO4 delays per cycle (this actually

Fig. 6: Measured plot of frequency (in MIPS) vs. $V_{DD}$ for 22 chips across a voltage range from 540mV to 1.2V.
exceeds the performance requirements\(^4\), so a corresponding path length of 3,000 FO4s is assumed; i.e., \(L_{dp} = N_i = 3,000\). In TSMC40LP the dynamic switching capacitance for an FO4 chain of inverters can be estimated as \(C_{dyn} \approx 1.2 \alpha F \cdot L_{dp}\), where \(L_{dp}\) is the path length. With this and the parameters

\(^4\)DD1 was designed for minimum energy operation regardless of the delay cost. In order to reduce dynamic energy, the sizing cost function was designed to heavily penalize device up sizing; this of course yielded much larger output to input capacitances for some gates than that of a design optimized by way of logical-effort analysis.

The rigorous definition and proofs from [14] can be reduced to a simple statement. When designing QDI circuits using Martin Synthesis [13], only one type of timing assumption is made: every isochronic fork must satisfy the constraint that the isochronic branch of the fork transitions faster than the corresponding adversary path. The DD1-core contains approximately 20K isochronic forks with adversary paths consisting

---

**Figure 7**: Inverter FO4 delay approximation, plotted for entire \(V_{DD}\) range against a BSIM4 SPICE simulation of TSMC40LP (25°C, TT-Corner) for minimum-size inverter driving an FO4 load. Fit from 135mV to 750mV, yielding \(V_{1} = 515\text{mV}, n = 1.60\) (where \(n\) is the subthreshold slope factor), and \(\frac{C_{sub}}{I_P} = 0.869 \frac{\text{nF}}{\text{V}}\).

**Figure 8**: Off-current plotted for entire \(V_{DD}\) range against a BSIM4 SPICE simulation of TSMC40LP (25°C, TT-Corner) for minimum-size devices with \(V_{t} = 515\text{mV}, n = 1.60\). Fit from 135mV to 750mV, resulting in \(\eta = 0.110\) (where \(\eta\) is the drain-induced barrier lowering factor), NFET \(I_0 = 552\text{nA}\), and PFET \(I_0 = 190\text{nA}\).

**Figure 9**: DD1 minimum-energy operating voltage vs. activity factor (\(\alpha\)) (25°C, TT-Corner).

**Figure 10**: Depiction of the length-five simplified adversary path timing assumption. The delay on the isochronic branch is labeled as \(t_{di}\), and the adversary path delay is labeled as \(t_{da}\).
of a variety of gates; however, there are at least five gates on every adversary path. Most of the DD1 adversary paths contain seven or more gates, and approximately 100 adversary paths contain only five gates in sequence. One simple approximation that provides an upper bound on the probability of failure is to consider an isochronic fork with a single FO4 inverter delay on the isochronic branch \((t_{da})\), and either five or seven FO4 inverter delays on the adversary path \((t_{da})\) as depicted in Figure 10; the assumption that \( t_{di} < t_{da} \) is referred to throughout as the simplified adversary path timing assumption (SAPTA). It is reasonable to assume that the probability of an adversary path timing failure in DD1 is strictly less than the probability that either 100 instances of length-five SAPTA fail or 20K instances of length-seven SAPTA fail (i.e., the probability of the union of these two events). The exact timing assumption is difficult to specify, because the transition on the isochronic branch does not cause a subsequent gate to switch. The isochronic branch of the fork must tie or cut a subsequent gate [27], thus preventing this gate from switching erroneously due to transitions on the adversary path. This section models this tie or cut time as the shortest propagation delay, an FO1, and further justification and experimental confirmation is left as future work.

![TSMC40LP NFET Vt distribution](image)

**Fig. 11**: TSMC40LP NFET \( V_t \) distribution (25°C, TT-Corner).

The TSMC40LP process is a low-power process with \( V_t = 515 \text{mV} \) at 25°C in the TT-Corner. As shown in Figure 11, \( V_t \) is a normally distributed RV with mean, \( \mu_{V_t} = 515 \text{mV} \), and standard deviation, \( \sigma_{V_t} = 34 \text{mV} \), (computed from statistical BSIM4 models using the methods from [24]). From [4], \( t_{di} \) and \( t_{da} \) can be modeled as log-normal RVs with expected values and variances calculated using the methods from [4], where \( L_{dp} = 1 \) for \( t_{di} \) and \( L_{dp} = 5 \) or 7 for \( t_{da} \) (see Figure 12). Assuming independence, the probability that \( t_{di} > t_{da} \) can be calculated by integration of the joint PDF. That is,

\[
P(FAIL(SAPTA)) = P(t_{da} < t_{di}) = 
\int_{-\infty}^{\infty} \int_{-\infty}^{y} f_{t_{da}}(x)f_{t_{di}}(y)\,dx\,dy,
\]

where \( f_{t_{da}}(x) \) and \( f_{t_{di}}(y) \) are the density functions for \( t_{di} \) and \( t_{da} \) respectively. For a log-normal RV the PDF follows from a change of variables on the normal PDF. That is, if \( Z \) is a log-normally distributed RV, then the corresponding PDF, \( f(Z) \), is given by

\[
f(Z) = \frac{1}{Z\sigma Z\sqrt{2\pi}} e^{-\frac{(ln Z - \mu_Z)^2}{2\sigma_Z^2}},
\]

where

\[
\mu_Z = ln(E[Z]) - \frac{\sigma_Z^2}{2},
\]

\[
\sigma^2_Z = ln \left( 1 + \frac{VAR[Z]}{(E[Z])^2} \right).
\]

From [4]

\[
E[t_{di}(V_t)] = \int_{-\infty}^{\infty} C_{load}\frac{V_{DD}}{I_F \sigma_V \sqrt{2\pi}} V_t e^{\frac{-k(V_{DD} - V_t)^2}{\sigma_V^2}} dV_t,
\]

\[
VAR[t_{di}(V_t)] = \int_{-\infty}^{\infty} C_{load}^2 \frac{V^2_{DD}}{I_F^2 \sigma_V \sqrt{2\pi}} V_t e^{\frac{-2k(V_{DD} - V_t)^2}{\sigma_V^2}} dV_t - (E[t_{di}(V_t)])^2,
\]

\[
E[t_{da}(V_t)] = L_{dp}(da) \cdot E[t_{di}(V_t)],
\]

\[
VAR[t_{da}(V_t)] = L_{dp}(da) \cdot VAR[t_{di}(V_t)].
\]

As such, \( P(FAIL(SAPTA)) \) can be directly computed by way of numerical integration. Figure 13 shows the estimated DD1 SAPTA failure probability vs. \( V_{DD} \). Over the minimum
energy operating range ~ 400mV – 800mV (see Figure 9), the DD1 timing failure probability ranges from 74\% to astronomically unlikely. At 500mV (just slightly below the threshold voltage) the estimated probability of failure is approximately one in ten-thousand, or put another way, the expected yield loss is 0.01\%—two to three orders of magnitude less than manufacturing defect yield loss. Of course, ensuring reliable timing is of little consequence if there is a high probability that gates simply fail to switch or are easily corrupted by noise.

D. DD1 Combinational Gate Robustness Estimate

It is possible to use the work developed in [4] to accurately compute the probability of SNM-based failures in DD1. At the time of the writing of this paper, the tools needed to perform this analysis are not complete, so this is left as future work. However, it is possible to derive a back-of-the-envelope robustness estimate. DD1 contains full-custom radiation-hardened memories with a separate supply voltage from the core, so the memories are not considered in this analysis. The DD1-core also contains a few full-custom registers, but these are biased separately and also removed from this analysis. The DD1-core contains ~16K standard cells. All of these cells (with the exception of the arbiter) are constructed using CMOS combinational logic, and the cell topology can be represented as ~120K inverter equivalent pairs. Using the least robust gate pair (NAND3, NOR3)$^3$, an accurate upper bound on the probability of failure can be calculated. That is, for the purpose of calculating robustness, the DD1-core can be represented by chains of 120K minimum-size alternating NAND3 and NOR3 gates, and the probability of failure can be computed using the chain-failure probability equation developed in [11].

Figure 14 plots DD1 robustness vs. $V_{DD}$. Timing failure probability is calculated assuming 100 independent length-five and 20K length-seven SAPTA instances in TSMC40LP (25\°C, TT-Corner). The parameters are taken from Figures 7, 8, and 11. Robustness failures are calculated by way of the methods detailed in [11] with $\delta = -0.013$, and 120K (NAND3, NOR3) pairs in TSMC40LP (25\°C, TT-Corner).

VI. RELATED WORK

Few labs have designed and fabricated radiation hardened devices in modern processes; but the design and test of a radiation-hard PowerPC device in a 0.25\um technology is discussed in [28]. The majority of modern radiation-hardened devices are designed and then implemented using programmable logic devices; e.g., see [29]. It is difficult to draw conclusions about the power, performance, and radiation robustness benefits of such an architecture without a tapeout. Radiation robustness of SRAM cells has been explored in modern (65nm) technologies in [30]. There are several commercially available radiation-hardened microcontrollers. Table I shows performance comparisons with 2 radhard MCUs from Aeroflex and Intersil, which have similar radiation robustness but much higher power figures, and 2 regular ultra-low power MCUs, the Atmel 24A and TI MSP430.
TABLE I: The DD1 is compared to two similar commercial radhard microcontrollers by Aeroflex and Intersil, and two regular low-power designs by TI and Atmel.

<table>
<thead>
<tr>
<th>Component</th>
<th>Memory (Kb)</th>
<th>Active Power (mW)</th>
<th>Subthreshold SRAM Yield</th>
<th>TID Threshold</th>
<th>Rad Hardness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aeroflex</td>
<td>256</td>
<td>20</td>
<td>1.00</td>
<td>Immune</td>
<td>N/A</td>
</tr>
<tr>
<td>TI</td>
<td>256</td>
<td>3.6</td>
<td>1.10</td>
<td>Immune</td>
<td>Unspecified</td>
</tr>
<tr>
<td>Atmel</td>
<td>256</td>
<td>1.10</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

VII. CONCLUSION

This paper described the design, analysis, and test results of a QDI radiation-hard 8-bit microcontroller taped out in a low-power common 40nm CMOS process. Radiation-hardness was achieved by way of duplicated memories and logic along with full-custom layout and design. The chip was found fully functional and radhard on first silicon over a range of operating voltages from near-threshold (500mV) to above the nominal VDD (1.1V).

ACKNOWLEDGMENT

We are grateful to James Lyke and Millay Morgan of AFRL for their support through an SBIR Phase I/II grant from the Air Force Research Laboratories, and David Alexander and Alonzo Vera for their help with the radiation tests. The research described in this paper is in part supported by a grant from the National Science Foundation.

REFERENCES