High-reliability Fault Tolerant Digital Systems in Nanometric Technologies: Characterization and Design Methodologies

C. Bolchini, A. Miele, C. Sandionigi
Politecnico di Milano
Dip. Elettronica e Informazione
Milano - Italy
{lastname}@elet.polimi.it

M. Ottavi, S. Pontarelli, A. Salsano
Università di Roma, "Tor Vergata"
Dip. Ingegneria Elettronica
Roma - Italy
{lastname}@ing.uniroma2.it

C. Metra, M. Omaña, D. Rossi
Università di Bologna
DEIS
Bologna - Italy
{firstname.lastname}@unibo.it

M. Sonza Reorda, L. Sterpone, M. Violante
Politecnico di Torino
Dip. Automatica e Informatica
Torino - Italy
{firstname.lastname}@polito.it

S. Gerardin, M. Bagatin, A. Paccagnella
Università di Padova
Dip. Ingegneria dell'Informazione
Padova - Italy
{firstname.lastname}@dei.unipd.it

I. Project goal

While the shrinking of minimum dimensions of integrated circuits till tenths of nanometers allows the integration of millions of gates on the single chip, it also implies the growth of the importance of effects that could reduce the reliability of circuits. In particular, the reduced integration step, the reduced supply voltage that lowers the noise immunity, the growing power needs, the eventual integration of both digital and analog circuits on the same chip and the highly growing of radiation sensitivity [1], [2], [3] require an accurate evaluation of possible reliability reduction for the occurrence of:

- permanent faults due to the aging of device materials [4], the interruptions of metal interconnections due to electromigration [5] or the crack of the insulation oxide of transistor [6];
- transient faults, known as Single Event Effects (SEE), which are much more likely than in the past due to the reduced transistors' sizes [3]: in particular new technology devices are more prone to crosstalk in the interconnects and to radiation effects.

These latter effects due to ionizing particles of cosmic origin, alpha and neutrons, were deeply studied in space [7], but were also observed at ground level [8] and at flight altitude [9]. These faults usually cause the change of a stored bit (*soft error* – SE) in a memory cell and are modeled as bit-flip. We can foresee that next generation circuits will be much more prone to such effects even at sea level together with other effects, nowadays considered only in space application, such as Multiple Bit Upset [10].

To reach the goal of acceptable values of the yield at foundry and of availability and reliability in the field, techniques for achieving fault tolerant properties need to be defined and adopted not only in restricted, safety-critical application environments, but also in more traditional scenarios. Moreover, if we consider the current trend towards new families of components, the *Systems-on-Programmable-Chips* (SoPCs), consisting in hardwired microprocessors embedded in a Field Programmable Gate Array (FPGA) module, these issues become even more relevant, being these devices particularly susceptible to these effects, because of the high integration and the presence of large on-chip memories.

This paper reports the main contribution of a project devoted to the definition of techniques to design and evaluate fault tolerant systems implemented using the SoPC paradigm, suitable for mission- and safety-critical application environments. In particular, the effort of the five involved research units has been devoted to address some of the main issues related to the specific technological aspects introduced by these flexible platforms. The overall target of the research is the development of a design methodology for highly reliable systems realized on reconfigurable platforms based on a System-on-Programmable Chip (SoPC), as discussed in the next section.

II. Main Contributions

Given the complex architecture, each one of the unit focused on a part of the problem, based on the expertise, contributing to the overall design and analysis methodology. In the following, we report a brief description of the investigated issues and achieved results.

A. Definition of suitable fault models for SoPC – Università di Bologna

One of the main issues to be addressed in the definition of methods and techniques to harden systems implemented on SoPC is the identification of a set of fault models, for

978-1-4673-3044-2/12/$31.00 © 2012 IEEE

both permanent and transient faults, to be used as a starting point for the selection/development of proper fault tolerant techniques able to guarantee the desired level of reliability. These same models are also used for the validation phase, here performed via fault-injection, to evaluate the effectiveness of the selected/developed fault-tolerant techniques.

Within the project, we considered the following faults:

- resistive bridgings and opens in local and global interconnects;
- transient faults due to energetic particles hitting the considered circuit;
- inductive and capacitive crosstalk among interconnects, considering both the interconnects within each block composing the SoPC and the interconnects among the different blocks;
- signal integrity issues due to power supply noise (mainly due to the simultaneous switching of high capacitive bus line drivers) and to degradation phenomena (mainly due to Negative Biased Temperature Instability, NBTI);
- problems due to clock faults, considering both the issues within the blocks composing the SoPC, and the ones in the communication among the different blocks.

The preliminary definition of the fault models have been used as a reference for the development of suitable hardening technique, to recover from the effects of these possible faults.

The main results of this part of the research have been published in [11], [12], [13], [14], [15].

B. On-line detection of permanent faults processors – Politecnico di Torino

With respect to the adopted reference architecture, one of the main issues of the project was to address fault-related aspects in the processor(s) constituting the SoPC architecture. More specific, the research in this direction focused on the definition of new techniques for on-line detection of permanent faults in state-of-the-art processors. These techniques are increasingly required in practice, since there is a growing need for guaranteeing a sufficient level of reliability in processor-based systems used for safety-critical applications. Standards and regulations often explicitly enforce these requirements, as it happens in the automotive area (through the ISO 26262 standard) or avionics (through the DO-254 standard).

Fault detection can clearly be achieved resorting to Design for Testability techniques: however, this solution requires having the possibility to modify the adopted circuit structures, which is not always the case. In other situations, the system company simply assembles different off-the-shelf devices developing boards and the related software. A similar scheme is used when Systems-on-Chips are adopted: IP cores take the place of devices in this case.

In these cases, a common solution to achieve on-line fault detection may be based on periodically running suitable test programs, in charge of exciting possible faults and making their effects visible, e.g., by looking to specific memory variables. A further advantage of this approach lies in the fact that the test is performed at the same speed of normal operations, thus offering a better defect coverage than other solutions. Such an approach has been proposed a long time ago [16], but it is now increasingly popular because of the availability of effective techniques for the development of effective test programs, able to detect faults not only within the processor, but also in any component within the system (memories, peripherals, chunks of logic) [17]. This problem has been tackled by following two paths: in the former, the architecture of todaysÕs pipelined processors has been considered, and solutions have been devised for the generation of suitable test programs for the most critical components within them. For example, effective algorithms have been devised for the test of Branch Prediction Units (based both on the Branch History Table [18] and the Branch Target Buffer architecture) and for caches in both single- and multi-processor systems. Case studies proving the effectiveness of these techniques on real devices have also been reported [19]. The second avenue of attack concerned the Very Long Instruction Word processors (VLIW), that are now commonly used in several signal processing applications, especially when power and performance constraints must be considered at a time. For VLIW processors it is possible to adopt most of the on-line test techniques already available for other processors: however, their specific architecture requires special solutions for some components (e.g., the register file [20]). Moreover, their regular architecture allows automating the generation of effective test code, once the architecture of the processor is known, thus making functional test practical for real applications [21], [22].

The proposed approaches can be exploited to harden the part of the system related to the processor, in order to provide fault detection/tolerance properties allowing the overall system to react to the occurrence of problems.

C. Error Detection and Correction in Content Addressable Memories – Università di Roma, "Tor Vergata"

The trend of embedding more complex microprocessors in SoPC has been confirmed in these years by the announcement of novel devices of Xilinx and Altera with embedded Intel or ARM microprocessors. With the increase of silicon area devoted to the implementation of these microprocessors, also the area devoted to implement Content Associative Memories (CAM) [23] such as cache and Translation Lookaside Buffers (TLB) is increased. These structures play an important role in modern microprocessor [24], and the percentage of chip area devoted to their implementation is increasing, and consequently is more likely that an error occurs in these structures. This appears to be an important research topic, since standard error correction code cannot be applied to these memories. Therefore, during this research project, we developed some methods able to detect and correct errors in CAM.

A Content Addressable Memory (CAM) is an SRAM based memory that can be accessed in parallel to search for a given search word, providing the address of the matching data as a result . In cache structures, the CAM checks if a memory address is present in the cache, and if so, it returns the cache line where the memory content is stored. In TLBs, the

CAM provides the translation between the virtual and physical memory addresses.Therefore new approaches to mitigate SEU effects in CAM can be seen as enabling technologies to use large CAMs in complex systems while ensuring high levels of reliability. Two methods are proposed, not requiring any modification to a CAM's internal structure and therefore can be easily applied at system level.

The first method combines the use of parity bits with a duplication and comparison scheme to detect and correct errors in CAM [25]. The use of a parity encoded CAM avoids the occurrence of pseudo-HIT[1]. Therefore, when the two CAMs provide a discordant HIT/MISS signal, since no false HITs occurs, we can detect the occurrence of an error on the CAM that provided the MISS signal. When both CAMs give a MISS signal, because of the assumption of single fault applies, then neither of the CAMs is a HIT and the overall status is MISS.

The second method uses a hash-based probabilistic structure called "Bloom filter" to check the correctness of CAM results [26]. In a Bloom filter when data has to be stored (or queried) it is hashed with multiple hash functions, and at the output of each hash a corresponding memory location is written (read). Since Bloom filter uses standard memories to store its data, we assume that the filter memory is protected by Error Correcting Codes. Bloom filters permit to efficiently store and query the presence of data in a set. But, while a CAM suffers from SEU induced errors, the probabilistic nature of Bloom filters is characterized by the so-called false-positive effect, as a consequence. However, by combining the use of a Bloom filter with a CAM, the complementary limitations of these modules can be compensated.

D. Design methodologies for implementing hardened systems onto programmable fabric – Politecnico di Milano

The two preceding contributions focused on specific elements of the adopted reference architecture; here we report the research focused on the reconfigurable fabric constituting the SoPC, used to implement programmable functionalities. In particular, the attention has been devoted to the definition of a methodology that, by exploiting strategies for the detection and possibly mitigation of fault effects, and by combining them with partial dynamic reconfiguration, is able to identify one (or more) interesting architectural solution satisfying both functional and reliability-aware requirements. More in detail, two are the elements that have been defined within the project:

- definition of a design methodology for the realization of systems able to mitigate SE effects in the programmable part of SoPC platforms, by exploiting both classical and new fault detection and tolerance techniques, together with reconfiguration features [27], [28], [29]; and
- definition of a suitable controller architecture for the management of the reconfiguration phase used to recover from the detected soft errors [30], [31].

The methodology allows the design of a hardened system onto the reconfigurable portion of the SoPC, supporting the possibility to recover after the occurrence and detection of a soft error. In particular, the reconfiguration controller is in charge of monitoring the error signals produced made available by the application of hardening techniques, and when a problem is detected, a reconfiguration is triggered to recover by re-programming the board with the original configuration.

The approach identifies a set of possible hardening techniques, such as duplication with comparison, triple modular redundancy, error detecting/correcting codes, and applies them to the part of the system implemented on the programmable portion of the device. In order to be as general and as flexible as possible, different techniques can be applied to different parts of the system. Therefore, a design space exploration is performed, estimating costs and benefits of the different hardening solutions, to identify the most promising solution.

E. Sensitivity of Floating Gate Memories to Ionizing Radiation – Università di Padova

The definition of the a specific fault model is one of the main aspect of any analysis and hardening methodology. Based on the preliminary results stemming from the research reported in Section II-A, we devoted our attention to the analysis of the sensitivity to ionizing radiations, to support the adopted fault models, and to offer a validation approach to the proposed hardening methodologies previously presented. In fact, Floating gate (FG) memories are a key component of today's SoPC and are used for both configuration/code and non-volatile data storage. Scaling has been profoundly altering the sensitivity of FG cells, once considered radiation immune. We then analyzed the behavior of scaled FG cells under several sources of ionizing radiation. In particular we investigated these aspects:

- threshold voltage (V_{th}) shifts and upsets induced by heavy ions (HI),
- effects induced by alphas, protons and atmospheric neutrons, and
- retention errors.

Using both experiments and simulations we analyzed the generation of tails in the V_{th} distributions of FG arrays after HI irradiation, distinguishing two types of events [32]: i) large events, occurring when an ion goes through the FG, responsible for the secondary peak in the V_{th} distributions; ii) small events, due to ions passing by FGs, related to the transition zone. Small events show characteristics similar to V_{th} shifts induced by total dose. The best agreement between experimental data and simulations is obtained when energy deposition in the FG is considered, with strong implications for error rate calculations. To further identify the sensitive volume (SV), we also studied the angular dependence of HI induced errors and V_{th} shifts [33]. We showed that the cross section for cells belonging to the HI induced secondary peak at high LET for NOR and at all the tested LETs for NAND is consistent with a SV with the same width and length as the FG, and considerably thicker than the tunnel oxide.

[1] A pseudo-hit occurs when a corrupted memory content corresponds to another content. If this content is searched, the CAM gives as response the location where the error occurred

We then focused on three types of particles with low ionizing power: protons, neutrons and alphas. The corruption of FG bits due to high-energy protons was analyzed in 41-nm Single Level Cell (SLC) NAND Flash memories [34]. Proton-induced upsets at low doses are not negligible due to a combination of direct and indirect ionization effects. Variability of energy deposition in the SV, the sequence of direct and indirect ionizing events, as well as the V_{th} and electric field reduction associated with each event were included in a model of proton-induced upsets. We then showed that starting from a feature size of 50 nm, Multi Level Cell (MLC) Flash memories are sensitive to alpha particles, whereas SLC devices do not show any sensitivity down to a feature size of 34 nm. We also investigated atmospheric neutron effects [35]. Charge loss is shown to occur especially at the highest program levels, causing raw bit errors in MLC NAND. A rapid increase in sensitivity for decreasing feature size was observed, as a result that charge used to store a bit decreases with each new generation, whereas the ion track remains constant. In spite of this, only in some conditions do the neutron and alpha particle threat appear to be of some significance, but nowhere large enough to defy current mandatory ECC requirements in NAND devices.

Finally, the retention of FG cells was studied up to one year after HI exposure [36]. Retention errors showed up less than 3 hours after reprogramming the devices irradiated with high LET ions. The cross section for retention errors follows a Weibull curve: compared to HI upsets, retention errors have a threshold LET more than 10 times higher and a saturation cross section about two orders of magnitude smaller. Modeling shows that, for the 65-nm MLC NOR Flash devices considered, just two traps in the tunnel layer are enough to generate a retention error, explaining the weak dependence on the incidence angle.

III. Conclusions and future work

The goal of defining a set of tools and methodologies to design and analyze complex systems implemented on SoPC is an ambitious one, because of the several aspects that need be taken into consideration. In this project, several of the main issues have been tackled and methodologies have been proposed to support the designer in the implementation of a hardened system on the adopted platform. These approaches, although developed within the project scenario, have been validated and evaluated independently, but not integrated within a single framework; such an integration and an overall validation of the complete methodology are considered future work.

Acknowledgements

This work has been partially funded by MIUR-PRIN project #2008K4P7X9_002.

C. Sandionigi is now with CEA-LIST, France.

M. Ottavi is funded by the Italian Ministry for University and Research; Program “Incentivazione alla mobilità di studiosi stranieri e italiani residenti all’estero”, D.M. n.96, 23.04.2001

References

[1] R. C. Aitken, “Nanometer technology effects on fault models for ics testing,” *IEEE Computer*, vol. 32, no. 11, pp. 46– 51, 1999.

[2] J. Ziegler and H. Puchner, “SER- History, Trends an Challenges,” 2004.

[3] “International technology roadmap for semiconductors,” 2010. [Online]. Available: http://public.itrs.net/

[4] C. Constantinescu, “Trends and challenges in vlsi circuit reliability,” *IEEE Micro*, vol. 23, no. 4, pp. 14–19, 2003.

[5] P.-C. Li and T. K. Young, “Electromigration: the time bomb in deep-submicron ics,” *IEEE Spectrum*, vol. 33, no. 9, pp. 75–78, 1996.

[6] F. W. Sextong, “Destructive single-event effects in semiconductor devices and ics,” *IEEE Trans. Nuclear Science*, vol. 50, no. 3-2, pp. 603–621, 2003.

[7] A. H. Johnston, “Radiation effects in advanced microelectronics technologies,” *IEEE Trans. Nuclear Science*, vol. 45, no. 3, pp. 1339–1354, 1998.

[8] E. Normand, “Single event upset at ground level,” *IEEE Trans. Nuclear Science*, vol. 43, no. 6, pp. 2742–2750, 1996.

[9] ——, “Single-event effects in avionics,” *IEEE Trans. Nuclear Science*, vol. 43, no. 2(1), pp. 461–474, 1996.

[10] S. Buchner, A. B. Campbell, T. Meehan, K. Clark, D. McMorrow, C. Dyer, C. Sanderson, C. Comber, and S. Kuboyama, “Investigation of single-ion multiple-bit upsets in memories on board a space experiment,” *IEEE Trans. Nuclear Science*, vol. 47, no. 3, pp. 705–711, 2000.

[11] D. Rossi, M. Omana, and C. Metra, “Transient fault and soft error on-die monitor,” in *Proc. IEEE Int. Symp. Defect and Fault Tolerance in VLSI Systems*, 2010, pp. 391–398.

[12] M. Omana, D. Giaffreda, C. Metra, T. Mak, S. Tam, and A. Rahman, “On-die ring oscillator based measurement scheme for process parameter variations and clock jitter,” in *Proc. IEEE Int. Symp. Defect and Fault Tolerance in VLSI Systems*, 2010, pp. 265–272.

[13] D. Rossi, N. Timoncini, M. Spica, and C. Metra, “Error correcting code analysis for cache memory high reliability and performancy,” in *Proc. IEEE Design, Automation & Test in Europe Conference & Exhibition*, 2011, pp. 1–6.

[14] D. Rossi, M. Omana, C. Metra, and A. Paccagnella, “Impact of aging phenomena on soft error susceptibility,” in *Proc. IEEE Int. Symp. Defect and Fault Tolerance in VLSI Systems*, 2011, pp. 18–24.

[15] M. Omana, D. Rossi, N. Bosio, and C. Metra, “Low cost nbti degradation detection and masking approaches,” *to appear in IEEE Trans. Computers*, 2012.

[16] S. M. Thatte and J. A. Abraham, “Test generation for microprocessors,” *IEEE Trans. Computers*, vol. 29, no. 6, pp. 429–441, 1980.

[17] M. Psarakis, D. Gizopoulos, E. Sanchez, and M. S. Reorda, “Microprocessor software-based self-testing,” *IEEE Design & Test of Computers*, vol. 27, no. 3, pp. 4–19, 2010.

[18] E. Sanchez, M. S. Reorda, and A. Tonda, “On the functional test of branch prediction units based on branch history table,” in *Proc. IFIP/IFEE Int. Conf. Very Large Scale Integration and SoC*, 2011, pp. 278–283.

[19] P. Bernardi, L. Ciganda, M. de Carvalho, M. Grosso, J. Lagos-Benites, E. Sanchez, M. S. Reorda, and O. Ballan, “On-Line Software-Based Self-Test of the Address Calculation Unit in RISC Processors,” in *Proc. IEEE European Test Symposium*, 2012.

[20] D. Sabena, M. S. Reorda, and L. Sterpone, “A new SBST algorithm for testing the register file of VLIW processors,” in *Proc. IEEE Int. Conf. Design, Automation & Test in Europe*, 2012, pp. 412–417.

[21] ——, “On the optimized generation of software-based self-test programs for VLIW processors,” in *Proc. IEEE Int. Conf. on Very Large Integration*, 2012.

[22] ——, “On the development of Software-Based Self-Test methods for VLIW processors,” in *Proc. IEEE Int. Symp. Defect and Fault Tolerance in VLSI Systems*, 2012.

[23] K. Pagiamtzis and A. Sheikholeslami, “Content addressable memory (cam) circuits and architectures: A tutorial and survey,” *IEEE Journal of Solid-State Circuits*, vol. 41, no. 3, pp. 712–727, 2006.

[24] D. A. Patterson and J. L. Hennessy, *Computer Architecture: A Quantitative Approach*, M. Kaufmann, Ed., 2003.

[25] S. Pontarelli, M. Ottavi, and A. Salsano, “Error detection and correction in content addressable memories,” in *Proc. IEEE Int Symp. Defect and Fault Tolerance in VLSI Systems*, 2010, pp. 420–428.

[26] S. Pontarelli and M. Ottavi, “Error detection and correction in content addressable memories by using bloom filters,” *IEEE Trans. Computers, to appear*.

[27] C. Bolchini, A. Miele, C. Sandionigi, N. Battezzati, L. Sterpone, and M. Violante, “An integrated flow for the design of hardened circuits on sram-based fpgas,” in *Proc. IEEE European Test Symposium*, 2010, pp. 214–219.

[28] C. Bolchini, A. Miele, and C. Sandionigi, “A novel design methodology for implementing reliability-aware systems on sram-based fpgas,” *IEEE Trans. Computers*, vol. 60, no. 12, pp. 1744–1758, December 2011.

[29] ——, “Automated resource-aware floorplanning of reconfigurable areas in partially-reconfigurable fpga systems,” in *Proc. IEEE Int. Conf. Field Programmable Logic and Applications*, 2011, pp. 532–538.

[30] C. Bolchini, L. Fossati, D. M. Codinachs, A. Miele, and C. Sandionigi, “A reliable reconfiguration controller for fault-tolerant embedded systems on multi-FPGA platforms,” in *Proc. IEEE Symp. Defect and Fault Tolerance in VLSI Systems*, 2010, pp. 191–199.

[31] C. Bolchini, C. Sandionigi, L. Fossati, and D. M. Codinachs, “A reliable fault classifier for dependable systems on sram-based fpgas,” in *Proc. IEEE Int. On-Line Testing Symposium*, 2011, pp. 92–97.

[32] S. Gerardin, M. Bagatin, A. Paccagnella, G. Cellere, A. Visconti, M. Bonanomi, A. Hjalmarsson, and A. Prokofiev, “Heavy-ion induced threshold voltage tails in floating gate arrays,” *IEEE Transactions on Nuclear Science*, vol. 57, pp. 3199–3205, 2010.

[33] S. Gerardin, M. Bagatin, A. Paccagnella, A. Visconti, M. Bonanomi, and S. Beltrami, “Angular dependence of heavy-ion induced errors in floating gate memories,” *IEEE Trans. Nuclear Science*, vol. 58, pp. 2621–2627, 2011.

[34] S. Gerardin, M. Bagatin, A. Paccagnella, J. R. Schwank, M. R. Shaneyfelt, and E. W. Blackmore, “Proton-Induced Upsets in 41-nm NAND Floating Gate Cells,” *IEEE Trans. Nuclear Science*, In press.

[35] S. Gerardin, M. Bagatin, A. Ferrario, A. Paccagnella, A. Visconti, S. Beltrami, C. Andreani, G. Gorini, and C. Frost, “Neutron-induced upsets in nand floating gate memories,” *IEEE Trans. Device and Materials Reliability*, vol. 12, pp. 437–444, 2012.

[36] M. Bagatin, S. Gerardin, and A. Paccagnella, “Retention Errors in 65 nm Floating Gate Cells after Exposure to Heavy Ions,” in *Proc. IEEE Nuclear Space Radiation Effects Conference*, 2012.