AMfoRS team (Architectures and Methods for Resilient Systems)

Themes

System-level modeling, analysis and testing
Dependability of integrated systems and architectures
Dependability of emerging computing paradigms; In memory computing, Nanocrossbars
Secured integrated architectures
Multi-level dependability evaluations
Emerging technologies: Design, Test, Reliability and Security

Expertise

Scientific
System modeling, fault detection/tolerance design, modeling of variability and aging effects, methods for robustness analysis, hardware security, system-level testing

Fields of expertise
Dependability evaluation at various levels of the design flow, dependability improvement, hardware security, system-level testing, hybrid CMOS – Non-Volatile technology design

Know-how
Fault tolerance or detection, fault injection, cryptographic accelerators, hardware attacks and counter-measures for secured circuits, IEEE 1687-compliant testing

Industrial transfer
MAST - CAD-tool for system-level testing based on IEEE 1687 transferred to a company under incubation
IDSM functional continuous checking approach transferred to STMicroelectronics and Dolphin Integration
EARS – CAD-tool for digital circuit dependability analysis,
APP deposit IDDN.FR.001.530007.000.S.P.2016.000.10600

Research keywords

Dependability, Fault tolerance, Security, Aging, Robustness analysis, In-memory computing, Spiking Neural Networks, Emerging technologies

Contact

Lorena ANGHEL
GRENOBLE INP Professor
(+33) 4 76 57 46 96
lorena.anghel@univ-grenoble-alpes.fr

Paolo MAISTRI
CNRS Researcher
(+33) 4 76 57 49 88
paolo.maistri@ univ-grenoble-alpes.fr

System-Level Test Architectures and Standards

Keywords: Testing, IEEE 1687, System-level approaches, functional access

Members: M. Portolan

Contracts: MAST (Linksium)

1. Context and goals

The digital automated testing field has been living a huge evolution in recent years: chips are becoming more and more complex, System-on-Chips (SoCs) are now common and new approaches like 3D Chips/Stacked Dies are pushing design and performance boundaries even further. Testing these new components where traditional divisions between component, device and system are blurred is a formidable task, and the field is responding by generating new dedicated approaches. Most actors soon realized that such an effort is not feasible by one entity, so different standardizations were launched, some of which started bearing their first results. Of particular interest is the new IEEE 1687-2014 standard, which develops and extends over the widespread IEEE 1149.1 standard, commonly known as JTAG, by adding several disruptive innovations, among which of particular importance are:

- Dynamic-length scan chains
- Instrument usage through a JTAG chain

These two elements have the potential not only to completely change the landscape of automated testing, but also to bring the current EDA tool to the brink of breaking. Current tooling is in fact based on a structural analysis of a System Under Test (SUT), helped by precise hardware assumptions, like for instance the fixed length of the scan chains, which are used to achieve global optimization with reasonable effort. Once these assumptions fall, the complexity explosion itself might be enough to render the tool helpless.
As depicted in *Fig. 1*, traditionally Testing starts after the end of the Design phase, and its role is to separate the "Bad" circuits, which would malfunction, from the "Good" ones, which can be shipped to customers.

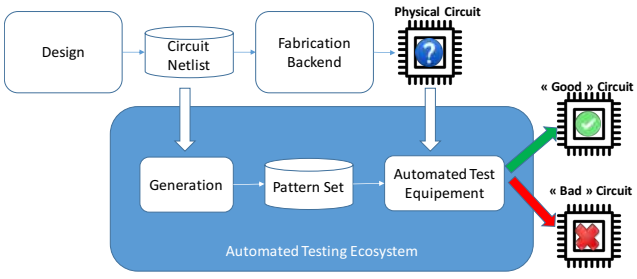


Fig. 1: The Automated Test Flow

The usual entry point of Testing is the Circuit Netlist obtained after Synthesis (and sometimes even after Routing), and the process is composed by two steps: Generation and Application.
During the Generation phase, the design is analyzed in order to obtain one or more sets of patterns, sometimes also called vectors. This process, usually called Automated Test Pattern Generation (ATPG), is based on a structural analysis of the synthetized netlist obtained at the end of the Design flow. Specialized Automated Test Equipment (ATE) machines are responsible for the Application of high-volume testing. The essential quality of an ATE is to be as fast as possible: factory testing is a crucial bottleneck because each fabricated circuit needs to undergo it. This means that the time spent of testing each specimen is immediately reverberated over the total cost because of the high number of fabricated circuits. A Tester must therefore be able to apply thousands of vectors in succession and identify failing chips. There is no time for complex decisions: even flow control is only used in order to minimize test time, usually by stopping at the first error. Debug and failure analysis can be done later, offline, thanks to log file or using specialized equipment.

We identified in this dichotomy the main limiting factor: while most actors are focusing effort on incremental evolutions of legacy solutions, we decided to take a long-term approach and analyze the whole Automated Test Flow in order to identify its core strength and its limiting factors. Work done in 2016 and 2017 allowed the development of a new software, the Manager for SoC Test (MAST), which implements several innovative and disruptive concepts. As specified in the previous Activity Report, 2018 was focused on the Incubation of MAST for the launch of a possible Start-Up, and on External communication.

2. Recent outcomes

This year we formalized a new Test Automation flow, depicted in *Fig. 2*.

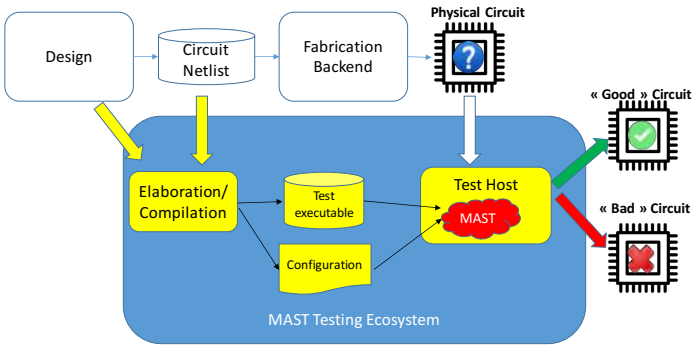


Fig. 2: MAST Testing Flow

The great difference with the legacy flow of *Fig. 1* lies in the Elaboration step and its communication with the Execution Backend, composed by MAST

being executed on the Test Host: instead of generating a full-fledged pattern set, the Elaboration provides an Executable file. It will be the role of the Host to generate patterns at the last possible moment, when all information is available. While algorithms are packed in the Executable file, the “Configuration” file depicted in the lower half of *Fig. 2* is responsible for holding topology information, to be used to build the DUT System Model: is can either be a standard language like ICL, a custom format to support specific topologies or a combination of both. It is an application of the Relocation principle, but using scan segments instead of variable addresses.

The incubation work (software development and scientific outreach) occupied most of 2018. This work resulted in a Patent application [1] and in the development of strong code base, protected through the French Agency for Software Protection (“Agence pour la Protection des Programmes”). In the meantime, we have been looking for academic cooperation to promote a widespread usage of MAST and of the new Automated Test Flow paradigm. The poster at the 2018 International Test Conference [2] marks the first common result with the Politecnico di Torino, Italy.

3. References

[1] M. Portolan, “Integrated Test Apparatus and Method”, FR 3066606 A1, Published: Nov 23, 2018

[2] M. Portolan, R. Cantoro, E. Sanchez, M. Reorda, “A Functional Approach to Test and Debug of IEEE 1687 Reconfigurable Networks”, 2018 International Test Conference, October 28-November 1, 2018, Phoenix, AZ, USA

Hardware/Software dependability analysis from High-Level descriptions

Keywords: dependability evaluation, fault injection, register criticality, high-level error models

Members: R. Leveugle, M. Portolan, K. Morin-Allory, J. Roux

Cooperations: LCIS, STMicroelectronics, Thalès Avionics, AEDvices consulting, LHC

Contracts: Safe-Air (AURA Region)

1. Context and goals

Significant effort has been targeted since more than fifteen years on developing efficient techniques to analyse, early in the design flow, the functional consequences of soft errors in digital circuits. The goal is to precisely identify the soft errors leading to unacceptable application disturbances, in spite of all the possible masking effects due to the circuit architecture or to the application characteristics. Most of the proposed techniques start from synthesizable RTL descriptions, already close to the final hardware in terms of cycle accuracy and in terms of memory cells identification. Embedded software is also taken into account in the case of systems based on microprocessors. Robustness evaluations may aim at (1) classifying the soft errors with respect to their functional impact, in order to compute derating factors on the application failure probability, (2) identifying error propagation paths, (3) identifying critical locations or registers, (4) ensure that a given set of behavioural properties always hold for a given set of soft errors (e.g., a given maximum multiplicity of erroneous bits). So-called fault injection techniques are used in most cases. For the latter case, the use of formal approaches had also been studied in order to avoid exhaustive fault injections, but with the classical limitations of formal techniques in terms of scalability and/or automation. New techniques have also been developed to improve or to avoid fault injections, without such limitations. Also, system-level properties (or modelling) are considered in order to refine the dependability evaluation.

2. Recent outcomes

The most recent fault injection platform developed in the team is based on emulation and takes advantage of the partial reconfiguration capabilities of Virtex 5 FPGAs. It was made freely available [1] and has been transferred in 2017 to STMicroelectronics. In 2018, work started to update this platform using more up-to-date FPGA fabrics and in particular Xilinx 7 structures. This work is on-going with the objective of a transfer for collaborative projects.

In parallel, a significant effort has been targeted towards new approaches, avoiding costly and time-consuming fault injections at least in some phases of the circuit development. It was demonstrated in 2016 that the dependability of digital circuits described at RT-level (including potential embedded software) can be accurately evaluated by leveraging the simulation environment developed for functional verification. Data lifetimes in circuit registers are analysed on the basis of a single functional simulation and the approach guarantees conservative criticality results that is not the case of statistical fault injection. Accuracy is similar to usual statistical fault injection conditions and allows identifying the most critical registers for a given application run. It is also faster than even emulation-based fault injections while not requiring specific equipment or skills. The analysis tool was protected in terms of intellectual property in 2016 [2] for valorization. The limitation is today that only intrinsic error masking can be identified; specific fault-tolerance mechanisms added to improve the intrinsic robustness cannot be taken into account in this version. Collaborative work with LIRMM laboratory was started [3]. Further work is discussed on this subject with several partners and also work is on-going about different applications of the methodology.

Current work also targets meaningful fault injections at higher level, i.e. on virtual platforms developed in SystemC TLM. The challenge here is to guarantee a good correlation between the robustness analysis on such a platform and the results obtained at RT-level. Another aspect of the work aims at refining the robustness analysis by taking into account system-level properties rather than only the correctness of the circuit boundaries. A previous study had shown that especially in cyber-physical systems circuit outputs may be incorrect during a large number of cycles and with large value discrepancies without significant impact on the global system behavior. Identifying such situations is important to avoid over-estimations of system-level failure rates.

3. References

[1] http://users-tima.imag.fr/amfors/leveugle/ATE-FIT5%20Page/ATE-FIT5.htm

[2] R. Leveugle, K. Chibani, M. Portolan, "EARS (Evaluation Avancée de Robustesse de Systèmes intégrés / Early Analysis of Robustness for integrated Systems)", APP deposit No. IDDN.FR.001.530007.000.S.P.2016.000.10600, December 30, 2016

[3] G. Di Natale*, M. Kooli*, A. Bosio*, M. Portolan, R. Leveugle, "Reliability of computing systems: from flip flops to variables", 23rd IEEE International Symposium on On-Line Testing and Robust System Design, Thessaloniki, Greece, July 3-5, 2017, pp. 196-198 (Invited paper)
* LIRMM, France

Design of secured crypto-processors and test access protections

Keywords: security, cryptographic systems, fault-based attacks, side-channel attacks, countermeasures, authentication, IEEE 1687

Members: P. Maistri, R. Leveugle, G. Di Natale, V. Reynaud

Cooperations: ENSMSE, LCIS, STMicroelectronics, Thalès, LIRMM, JTAG

Contracts: ALADDIn (IDEX Grenoble), HADES (Penta), SPICA (FUI)

1. Context and goals

A current trend for many products, and in particular for consumer and IoT products, is toward an increasing need of security (confidentiality, data integrity, and/or authentication). These services rely on cryptographic protocols and algorithms, which can be implemented in software or hardware according to the performance requirements, and to the cost constraints.

Many current secure implementations rely on specific hardware blocks to implement the main cryptographic functions. These functions can be tampered by various attacks, either active (fault-based attacks) or passive (side-channel attacks: computation time analysis, power analysis, observation of electromagnetic emissions…). So-called hardware attacks target the implementation rather than the algorithm itself and are today a significant threat for security, in addition to software- or network-based attacks.

The work done in the team aims at (1) better characterizing and modelling the effect and feasibility of attacks, in particular fault-based attacks by various means, and (2) propose innovative countermeasures (i.e., protections) against the different types of attacks. This section of the report is focused on a novel attack against a countermeasure developed in the team, together with an extension towards authentication and the possibility to secure test access in the context of complex devices.

Our countermeasures are mainly implemented at RT-Level, even when targeting low-level characteristics such as power consumption. A lot of work has been focused on the development and validation of robust re-usable cores (IPs) for cryptography. Previous and on-going studies cover asymmetric (mainly ECC), symmetric (AES and lightweight implementations), and homomorphic cryptosystems. New work is now dedicated to hierarchical access authorizations to embedded test and monitoring devices.

2. Recent outcomes

A project in 2017-2018 was targeted at validating a new second-order laser attack on AES crypto-processors protected by specific time redundancy (so-called DDR approach). The work was based on circuits manufactured in a previous project (ANR LIESSE) by our partner STMicroelectronics in 28 nm bulk and FDSOI technologies. It has been carried out in collaboration with our colleagues from ENSMSE in Gardanne, where extensive analyses have been conducted on the chips [1,2]. A few experimental campaigns were conducted: a first campaign aimed at validating the feasibility of the attack gave inconclusive results. A second campaign has been conducted in order to assess the correctness of the setup. The abnormal results obtained in this simplified set of experiments put into question the reliability of the circuits under attack. These partial results were shared with the international community at [3]. In the second part of the year, it was decided to use a new set of chips available from the previous project and thus new boards were set to fabrication. The boards have been partially completed, and will be soon object of full experimentation against fault and side channel attacks with our colleagues from ENSMSE (Gardanne) and LCIS (Valence).

Another project aims at protecting the access to embedded instruments in the framework of IEEE 1687 standard. In addition to protecting the scan test access, as in older circuits, it becomes necessary to selectively protect the access to specific monitors that can provide sensitive data to an attacker but need to be accessed on-line during the whole lifetime of the circuit. There is therefore a need for (hierarchical) authentication and a real concern about the security of such access when millions of chips can be disseminated worldwide. The main effort of the last year has been dedicated to proposing an improved approach for secure authenticated access to reconfigurable scan chains [4]. The solution is based on a hashing IP supporting mutual authentication between the device and the user. With respect to the state of the art, the proposed approach allows faster and cheaper authentication in large SoCs. The protocol is currently being evaluated from the point of view of overheads and performance, and it will be implemented on a FPGA prototyping board from the project partner JTAG Technologies.

3. References

[1] J.M. Dutertre, V. Beroulle, P. Candelier, S. De Castro, L.B. Faber, M.L. Flottes, P. Gendrier, D. Hély, R. Leveugle, P. Maistri, G. Di Natale, A. Papadimitriou, B. Rouzeyre, “Laser Fault Injection at the CMOS 28 nm Technology Node: an Analysis of the Fault Model, 14th Workshop on Fault Diagnosis and Tolerance in Cryptography” (FDTC'2018), Amsterdam, NETHERLANDS, 2018

[2] J.M. Dutertre, V. Beroulle, P. Candelier, S. De Castro, L.B. Faber, M.L. Flottes, P. Gendrier, D. Hély, R. Leveugle, P. Maistri, G. Di Natale, A. Papadimitriou, “Sensitivity to Laser Fault Injection: CMOS FD-SOI vs. CMOS bulk (Early Access)”, IEEE Transactions on Device and Materials Reliability, DOI: 10.1109/TDMR.2018.2886463, 2018

[3] P. Maistri, J.M. Dutertre, R. Leveugle, “Laser Attacks against DDR Redundancy”, Workshop on SecURity, REliAbiLity, test, prIvacy, Safety and Trust of Future Devices (SURREALIST'2018), Bremen, GERMANY, 2018

[4] V. Reynaud, P. Maistri, R. Leveugle, “Accès autorisé au réseau reconfigurable de test par ensemble de segments”, 13ème Colloque du GDR SoC/SiP, Paris, FRANCE, 2018

Digital System Failure Prediction face to Variability and Aging

Keywords: Reliability, Aging, Performance Monitors, Aging Monitors, DVFS, AVS, Machine Learning

Members: L. Anghel, M. Portolan, A. Sivadasan, R. Shah, K. Senthamarai Kannan

Cooperations: STMicroelectronics Crolles

Contracts: CIFRE contracts, HADES (Penta)

1. Context and goals

In complex SOC design manufactured in nanometric technologies, circuit functionality in all process corners and face to all kinds of variabilities need to be verified. Variabilities and wear-out degradation impact system performance, potentially resulting in timing and functional failures. Indeed, local and global variability, aging phenomena, such as NBTI and HCI, have become the most critical reliability issues. Hence, taking into account these phenomena during the during circuit and system design and validation steps are mandatory, especially for high reliable application such as automotive, health-care or other mixed-critical applications. In fact, these reliability threats can severely degrade performance and in the worst case they can provoke system failures. It is common for most of these applications to embed reliability and performance monitors.

The usage of in-situ monitors for error and pre-error detection allow decreasing the constraints imposed on the overall design. They are implemented together with adaptive voltage scaling (AVS) technique or Dynamic Voltage Frequency Scaling (DVFS) which are triggered by pre-errors in-situ timing monitors while adapting dynamically the frequency and the voltage according to the operating conditions and the application needs [1]. Therefore, the performance degradation can be compensated and the circuit's lifetime can be extended.

2. Recent outcomes

In the framework of a long term on going collaboration between TIMA and ST Microelectronics several aspects have been tackled and arrived to some maturity in 2018:

- Different pre-error monitors have been designed. They are based on Replica Path principle, but the design has been optimized to allow sensing local and global variability and aging degradations [3], [4].
- Variation and aging sensitivity estimation methodology was applied to a large digital circuit and results have been published in paper [4] and the correlation with CAD simulation and test data has been performed. Data were obtained with and without body-bias (AVS) techniques, with static and dynamic stress, which is crucial to understand the impact of the activity (workload) on the circuit degradation.
- Current design practices perform monitors insertion in near-critical paths, identified by classical techniques such as Static Timing Analysis (STA), with the assumption that they should be the first affected by transistor degradation. Even if promising this approach still has a serious weak point: phenomena such as NBTI and HCI are strictly correlated to circuit activity, generated by the executed workload. This means that: aging is not necessarily coherent with STA results, as near-critical paths may be seldom activated [4]. Experiments demonstrated how STA performed on aged circuits can deliver a significantly different set of near-critical paths depending on the workload.
- To tackle the above-mentioned issue, we decided to explore a new research direction consisting in find a near-optimal monitor placement strategy based on predictive aging modeling; Therefore, we prioritize data-driven automated learning approaches to model the specific near-unpredictable behavior of transistor aging. Our research activities focused on the development of Machine Learning algorithms for predicting circuit aging. This implied a deep bibliographic effort to identify the main physical aging phenomena and insert them in a ML-friendly model. This allowed us to couple a lightweight Machine Learning prediction framework with traditional, computationally intensive circuit simulations as validation. First applied on the older technology, this Proof-of-Concept was presented with success as a Poster [5], where the demonstrated its capability to reliably model path aging. The activity is progressing on two fronts: on the one hand port the model to the FDSOI 28nm technology, more prone to aging, and on the other hand, to apply the ML framework to the identification of an optimal set of monitor insertion points.

3. References

[1] A. Benhassain, F. Cacho, V. Huard, S. Mhira, L. Anghel*, C. Parthasarathy, A. Jain, A. Sivadasan, “Robustness of Timing In-Situ Monitors for AVS Management”, in Proc of International Reliability for Physics of Semiconductors, IRPS 2016

[2] A. Sivadasan, R. Shah, F. Cacho, L. Anghel, “NBTI aged cell rejuvenation with back biasing and resulting critical path reordering for digital circuits in 28nm FDSOI”, Design Automation and Test in Europe (DATE'2018), Dresden, GERMANY, March 19-23, 2018

[3] R. Shah, F. Cacho, L. Anghel, “Investigation of speed sensors accuracy for process and aging compensation”, IEEE International reliability Physics Symposium (IRPS'2018), San Francisco, USA, March 11-15, 2018

[4] R. Shah, F. Cacho, R. Lajmi, L. Anghel, “Aging Investigation of Digital Circuit using In Situ Monitors”, IEEE IIRW 2018

[5] K. Senthamarai Kannan, M. Portolan, L. Anghel, “Run-Time Aging Prediction Through Machine-Learning”, 2018 International Test Conference, October 28-November 1, 2018, Phoenix, AZ, USA

Emerging Technologies and Applications to Reliability, Test and Security

Dependable Spiking Neural Networks based on Emerging Technologies

Keywords: spiking neural networks (SNN), reliability and test of SNNs, spintronic and memristive synapses

Members: L. Anghel, G. Di Natale, E.I. Vatajelu

Cooperation: LEAT, CEA-LETI, CEA-SPINTEC

Project support: EU H2020 NANOxCOMP

1. Context and goals

In recent years, hardware-implemented Spiking Neural Networks (SNN) have shown to bring an improvement in efficiency compared to classical activity-based Artificial Neural Networks (ANN) while ensuring comparable classification performances. In order to get the maximum efficiency out of hardware implemented SNNs, functional modules (neuron-synapse) have to be designed in such a way that their input/output characteristics provide the learning and processing capability required by application. Additionally, the network connectivity has to allow for high integration with strong and reliable reconfiguration and adaptation characteristics.

For a robust and efficient hardware implementation of SNNs we have to jointly-consider the characteristics of the SNN itself (connectivity, neuronal activation function, learning rule and synaptic update) and the characteristic of the devices used to implement it (CMOS ON/OFF current and threshold voltage, conductivity modulation and current-compliance of the RRAM, etc.). Here we focus on a fully-connected SNN, that learns using the Spike Timing Dependent Plasticity (STDP) method with lateral inhibition, with integrate-and-fire neuron and resistive synapses (see figure). Fig. a) sketches the connectivity between two consecutive layers of such a network. The synapse and its control circuit are illustrated in Fig. b), while the functional structure of a neuron is illustrated in Fig. c).

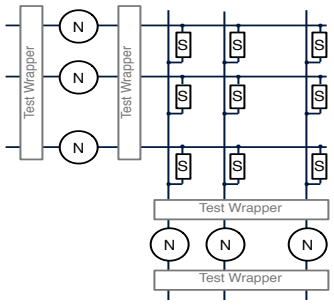
While many efficient solutions exist for testing traditional designs, to the best of our knowledge, there is no work so far dealing with the post-fabrication testing of a hardware-implemented SNN. **In this work we study main constraints of hardware implemented SNNs, possible faults and reliability issues, together with main challenges faced by post-fabrication test.** Circuits implementing SNNs have some major differences compared to classical circuits. Indeed, they resort to the combination of devices with both deterministic and stochastic behaviours and they include both digital and analogy elements. A test strategy suitable for SNNs should be able to test the correct operation of both neurons and synapses, it should be economical (in area, power and performance overhead), it should not depend on the training data nor on the context the network would be used. This last condition is important for general-purpose SNNs, which could be used in different scenarios. Indeed, it is possible for SNNs with online learning to be used for pattern classification within different contexts, as long as the problem space is equal or smaller than the input-neuron space.

2. Perspectives: SNN Testing

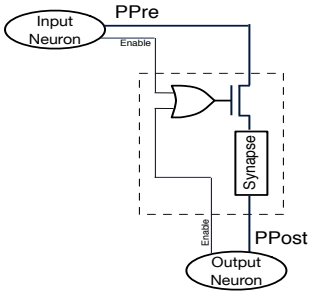
In future research, we will devise models and test strategies according to the specificity of the synaptic array. For the analog synapse implementation, a classical analog test could be used, while for the binary implementation with probabilistic programming an entirely new test strategy should be devised. The main challenge here is to identify a strategy to test for the randomness of the resistive device programing. This test has to be low cost and have no impact on the network behavior at runtime. A possibility is to start use similar strategy as the ones used for the testing quality of true random number generators. Regarding the neuron, it is in most cases an analog device, and classical analog test can be used. The main challenge here is to find a way to apply the test vectors and read the circuit response.

3. References

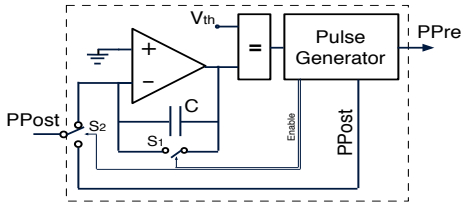
[1] L. Anghel, G. Di Natale, B. Miramond, E.I. Vatajelu, E. Vianello, “Neuromorphic Computing - From Robust Hardware Architectures to Testing Strategies”, 26th IFIP IEEE International Conference on Very Large Scale Integration (VLSI SOC 2018), Verona, ITALY, October 8-10, 2018



(a) Two-Layer Fully Connected SNN



(b) Synapse and control circuit



(c) Integrate and Fire Neuron

Emerging Technologies and Applications to Reliability, Test and Security

Security Primitives with Emerging Technologies

Keywords: True Random Number Generators, Physically Unclonable Functions, Security, Spin-Transfer-Torque Magnetic Tunnel Junction

Members: G. Di Natale, E.I. Vatajelu

1. Context and goals

Security primitives are low-level cryptographic algorithms used to build protocols for computer security systems. As an alternative to classical mathematical cryptography primitives, novel hardware security primitives are used today, such as Physically Unclonable Functions (PUFs) – for implementing low-cost authentication protocols or secure key generators, key storing, and one-way functions; and True Random Number Generators (TRNGs) – which provide random keys, device identification, and seeds for Pseudo Random Number Generators (PRNGs).

The Physically Unclonable Functions (PUFs) exploit intrinsic manufacturing variability introduced in a device during the fabrication process to generate a signature, unique to each single device. In order to guarantee its security, the generated secret key must be unique from device to device (unclonable), and, for a same device, it must be robust with respect to aging and environmental variations (reproducible).

True Random Number Generators (TRNGs) are used to generate random numbers from a physical process, rather than a computer program. They are implemented by taking advantage of thermal noise or other quantum phenomena and are expected to generate random bits with very high entropy and zero correlation. An on-chip TRNG design should occupy small area, give high bit rate, and have low power consumption, while assuring un-biased bit streams with high entropy per bit and low (no) correlation among them.

The continuous technology scaling has brought forth many challenges for today's ICs, such as increased variations in the fabrication process, increased sensitivity of transistors to operating conditions, increased power consumption, and so forth, pushing today's CMOS ICs to their limits. The severity of these issues and the ever-increasing need of smaller, more power efficient, and faster computers, have opened the field to the development of emerging technologies. The Spin-Transfer-Torque Magnetic Random-Access Memory (STT-MRAM) has emerged as a promising choice for embedded memories due to its reduced read/write latency and high CMOS integration capability. Moreover, the spin-based technology has been proven useful in various other applications like approximate computing. While many papers exist on the design and reliability analysis of emerging spin-transfer-torque magnetic devices, few studies exist for PUF and TRNG.

Concerning TRNGs, we have proposed a solution exploiting the intrinsic stochasticity characterizing the write operation [1]. Indeed, the write operation has a stochastic behaviour, i.e., the success of a write operation is a probabilistic phenomenon, due to the intrinsic thermal instability of all magnetic nanostructures. This instability can be manipulated to set the write probability to 50 %, thus generating random numbers. Compared to existing solution, in our work we provided a thorough analysis of the MTJ behaviour under variability and we have provided a solution dedicated to VLSI implementation of a CMOS-compatible MTJ-based TRNG. The solution exploits the post-processing power of XORs and does not require complex analog design to compensate the effects of variability and noise. This makes the design straightforward and guarantees the quality of the random numbers from the first generated bit.

2. Perspectives: Towards Reliable PUFs

In future research, we target the hardware implementation of reliable PUFs. We will build on our previous work [2-3] and we will propose solutions to identify and mask unreliable PUF responses with the purpose of assuring high reproducibility of the PUF response over large range of operation conditions at a minimum price.

3. References

[1] E.I. Vatajelu, G. Di Natale, “High-Entropy STT-MTJ-based TRNG, IEEE Transactions on VLSI”, DOI: 10.1109/TVLSI.2018.2879439

[2] E.I. Vatajelu, G. Di Natale, M. Barbareschi, L. Torres, M. Indaco, P. Prinetto, “STT-MRAM-Based PUF Architecture Exploiting Magnetic Tunnel Junction Fabrication-Induced Variability”. J. Emerg. Technol. Comput. Syst. 13, 1, Article 5 (May 2016), 21 pages. DOI: https://doi.org/10.1145/2790302

[3] E.I. Vatajelu, G. Di Natale, P. Prinetto, “Towards a highly reliable SRAM-based PUFs”. In Proceedings of the 2016 Conference on Design, Automation & Test in Europe (DATE'16). EDA Consortium, San Jose, CA, USA, 273-276

Emerging Computing: Design, Test and Dependability of In-Memory Computing Applications

Design for Testability and Reliability of In-Memory Computing Architectures

Keywords: in-memory computing, emerging technologies, design & reliability

Members: L. Anghel, E. I. Vatajelu

Cooperations: IMEP-LAHC, Aix-Marseille Université, SPINTEC-CEA

Today's computing systems are facing a plethora of issues related to architectural and technological limitations. Form an architectural perspective, three main problems have been identified as being the bottleneck for future system development and performance improvement, i.e., the memory wall, the power wall and the instruction-level parallelism wall. The memory wall refers to the ever-increasing gap between the speed of the Central Processing Unit (CPU) and that of the memory outside the CPU chip [1]. In recent years, processor speeds have increased significantly, while memory improvements have mostly been in density while transfer speeds remain mostly constant. As the speed gap increases, the CPU speed improvement becomes irrelevant since the overall computation speed is limited to the rate of transfer allowed by the memory. The problem is particularly acute in highly parallel systems, but occurs in platforms ranging from embedded systems to supercomputers, and is not limited to multiprocessors. The power wall refers to the peak power constraint of a system. With technology scaling, more and more devices can be packed on limited chip area which leads to considerable increase of the power density, despite the reduction of the supply voltage. This, in turn, causes an increase in the chip temperature and expensive cooling strategies need to be implemented. In addition to these architectural bottlenecks, the technology development is facing a scaling wall, i.e., a limited possibility of miniaturization of fabricated devices, diminished returns in performance and increased leakage power.

The main trend today is to speed up computing by using hardware accelerators, i.e., do computing-intensive jobs in hardware. Our work is focused on bringing computation kernels (which traditionally use large resources in SW) to HW and implement them by using the Computing In-Memory (CIM) paradigm. This means a change in the computing paradigm to data centric computing. The HW implementations take maximum advantage of specific emerging memory device characteristics.

Computing In-Memory (CIM) concept is based on merging memory elements and computational circuitry, to minimize the time and the energy needed to move data across the processor. Research efforts are directed to proving the effectiveness of emerging memory devices (spintronic and memristors) to achieve the in-memory computing ambition through the development of new low-cost architectures with high reliability and low power consumption [2].

In this context, **we perform a comprehensive study of emerging device integration in storage and computational structure leading to technology benchmarking in the context of the CIM paradigm.** We will address circuit design, test and reliability challenges related to memristive and spintronic technologies, from basic computation cells up to the CIM architecture.

Our current research and collaborations in relation with this topic are threefold: (i) to use enhanced compact models of emerging devices to perform failure analysis and define pertinent fault models; (ii) to investigate meaningful reliability threats affecting the CIM modules and architectures, and derive solutions for their mitigation (Design-for-Reliability), (iii) to establish design and test methodologies for these devices and architectures.

References

[1] C. Pancratov, J.M. Kurzer, K.A. Shaw, L. Matthew, "Why computer architecture matters: memory access", Computing in Science and Engineering, vol. 10, no. 4, pp. 71-75, 2008

[2] E.I. Vatajelu, L. Anghel, J.M. Portal, M. Bocquet, G. Prenat, "Resistive and Spintronic RAMs: Device, Simulation, and Applications", IEEE International On Line Testing (IOLTS'2018), Platja d'Aro, SPAIN, July 2-4, 2018

Emerging Computing: Design, Test and Dependability of In-Memory Computing Applications

Test and Reliability for Switching Nano-Crossbar Arrays

Keywords: in-memory computing, emerging technologies, design & reliability

Members: L. Anghel, E. I. Vatajelu

Cooperations: U Milano, KIT, ITU Turkey, IROC Technology

Project support: H2020 RISE NANOxCOMP

Nano-crossbar arrays have emerged as area- and power-efficient structures with an aim of achieving high performance computing beyond the limits of current CMOS. Due to the stochastic nature of nano-fabrication, nano arrays show different properties both in structural and physical device levels compared to conventional technologies. Mentioned factors introduce random characteristics that need to be carefully considered by synthesis process. For instance, a competent synthesis methodology must consider basic technology preference for switching elements, defect or fault rates of the given nano switching array and the variation values as well as their effects on performance metrics including power, delay, and area. Currently, computing is achieved with crosspoints behaving like switches, either as two-terminal or four-terminal.

A competent synthesis methodology must consider basic technology preference for switching elements, defect or fault rate of the given nano-crossbar and the variation values. The synthesis methodology used in this study comprehensively covers the all specified factors and provides optimization algorithms for each step of the synthesis. The main goal of this cooperation is the **development of a complete synthesis and optimization methodology for switching nano-crossbar arrays that leads to the design and construction of an emerging nanocomputer**.

We have investigated the possible defects and faults that can occur in these types of structures and have categorizing them as permanent and transient. In this context, defect tolerance basically means finding defect-free region or crosspoint which can be employed during logic synthesis. In our present study [1] we have performed a sensitivity analysis of crosspoints to these defects.

Our proposed method for defect tolerance utilizes the sensitivity analysis of crossbar to identify critical switches, and using mitigation factors to strengthen them. We have observed that the more restrictive an algorithm is in terms of area (results closer to optimal solution), the higher is the sensitivity of the output to cell defect. For instance, if an injected defect occurs in a multiple-choice cell, we have 2 option: (1) change the literal, if a different literal can be chosen to make the cell robust, (2) replace column (or row) with spare, if the literal is critical to the output.

In the study in [1] we have presented a synthesis methodology for crossbar arrays having crosspoints working as FET, diode/resistive/ memristive, or four-terminal switch based devices. We have analyzed logic synthesis-dependent defect/fault tolerance techniques, and have performed variation-area-power-delay performance optimization.

References

[1] M. Ceylan Morgul, O. Tunali, M. Altun, L. Frontini, V. Ciriani, El.I. Vatajelu, L. Anghel, C. Andras Moritz, M.R. Stan, D. Alexandrescu, “Integrated Synthesis Methodology for Crossbar Arrays”. In Proceedings of the 14th IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH'18). ACM, Athens, GR, 91-97

RISC-V based SoC Platform for Research Development and Education

Members: M. Benabdenbi, N. Ait Said

Keywords: RISC V, SoC design, Rocket Chip, prototyping, embedded system platform, processor, multi-core, education

1. Introduction

Designing modern System on a Chip (SoC) is based on the joint design of hardware and software (co-design). However many educational courses/labs worldwide target hardware design or software design only. Thus understanding the tight relationship between the two aspects is poorly addressed.

We are experimenting at Grenoble Institute of Technology a new set of training labs dedicated to Master 2 students and young researchers.

This set of training labs is based on the use of the free and open-source RISC-V processor architecture [1], and on the Rocket Chip [2] design platform, which is open source, recent, well maintained, and evolves with the standard RISC-V specification. RISC-V being open source and royalty free, its adoption is growing in the industry (Nvidia, Western Digital, Microsemi,...) as well as in academia (research and education). Rocket Chip allows the generation, simulation and FPGA mapping of highly configurable architectures, allowing students to deepen their knowledge and putting into practice the theoretical concepts acquired in courses and tutorials. The Rocket Chip SoC platform is also a good candidate for research development thanks to its hardware/software modularity and its tuning capabilities.

Here follows a short description of the hardware/software design environment and what has been developed for educational and develoment purpose, more details can be found in TIMA's website [3]. First research experiments using this platform are under development.

2. Architecture and prototyping platform

The proposed tutorial is based on the RISC-V instruction set architecture (ISA) belonging to the reduced-instruction set processor (RISC) family.

Although the RISC-V processor architecture is open-source, there are some commercial implementations such as Codixip's Codix-Bk processors. There are also open source implementations such as Rocket core [2], BOOM core (Berkley Out-of-Order Machine) [4], PULP, PULPino [5].

A technical study of a set of tools was carried out and led to the choice of the Rocket Chip platform as a SW/HW development platform, with the Rocket core (RISC-V based) as the main processor.

2.1. Design flow

The Rocket Chip platform consists of a SoC generator that can be used to generate different hardware architectures. Two interconnected design flows are available (see Fig.1):

1. A hardware design flow including hardware components, described either in CHISEL [6], a hardware description language (Constructing Hardware in a Scala Embedded Language), or in Verilog language. These resources can be compiled and synthesized for implementation on FPGA / ASIC with industrial tools, or translated into a hardware emulator (C Emulator or verilated emulator) executable on PC machines (x86). The result is an integrated hardware system that can operate on any RISC-V software code.

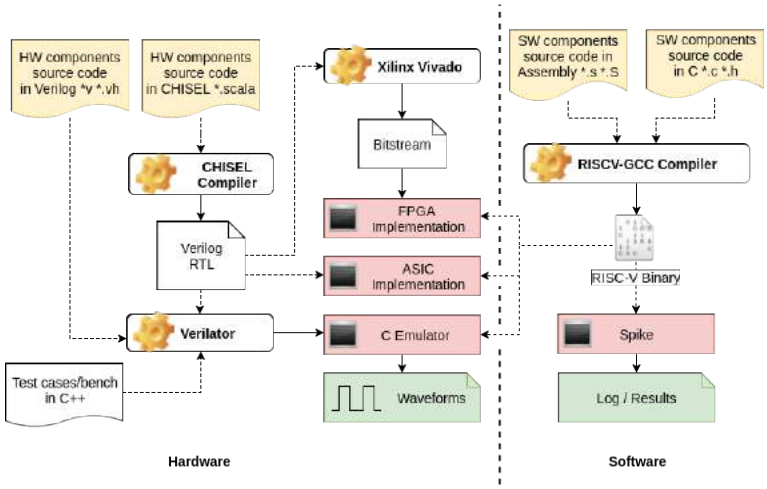


Figure 1: HW and SW design flow

2. A software design flow including software components. When compiled, assembled, and linked they lead to build a binary that will be run on the RISC-V target machines. These binaries can run as bare metal or they can rely on a real-time operating system such as FreeRTOS or run using an embedded Linux OS.

To execute the produced binary, an instruction set architecture simulator called Spike [7] is used. It is a simulation program that runs on a PC x86 machine, which verifies that the binary complies with the official specification. As Spike does not take into account some specificities of the hardware, a generated hardware emulator (C Emulator) is used to emulate the behaviour of the real hardware. It is a cycle accurate bit accurate (CABA) emulator that can be used to measure program run times in clock cycles, and allows the inspection of internal signals of the processor(s) and custom devices. The emulator allows the developers to test and debug their own devices written in either Verilog or CHISEL.

2.2. Rocket Chip Architecture

The Rocket Chip platform, depending on the chosen parameters, generates (see Fig. 2) different SoC architectures including:

1. Rocket tiles, composed of one or multiple Rocket cores with L1 cache data and instructions.
2. A system bus and peripherals: an open source TileLink bus that interconnects the various system components (standard or custom devices, interrupt controllers, tiles...)
3. A debugging interface: using Debug Transport Module (DTM) or JTAG for debugging and programming purposes.

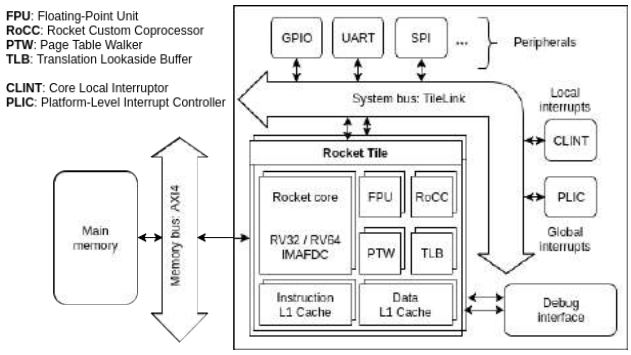


Figure 2: Architecture of the generated SoC

The hardware platform is highly configurable. One can configure different internal parameters: number of tiles, number of cores, sizes and policies of the caches, parameters of the MMU (PTW and TLB), activation or deactivation of the FPU, addition of dedicated hardware accelerators (RoCC). It is also possible to enhance the SoC by including peripherals and interrupt controllers.

3. Proposed Training

The developed training labs illustrate and deepen theoretical concepts seen in class lectures:

- Relationship between hardware and software
- Exceptions, interrupts and traps
- Multi-tasking, multi-processing, memory coherence
- Development and integration of custom devices

The main objectives are to become familiar with the software and hardware development environment: practical implementation of concepts and techniques such as cross-compilation, assembly, link editing, debugging, scripting, hardware design and emulation of different system architectures by tuning the architectural parameters. Events which can interrupt the execution of a program are also studied and implemented: exceptions, system calls and interruptions. Types of interrupts, sources, access to configuration registers, configuration of interrupt service routines (ISR) and change of context are particularly studied.

A basic multi-tasking application is implemented in assembly on a single-processor system in order to master the notions of time-sharing and context switching. Then, multiprocessor architectures are used to understand the mechanisms of memory coherence and the challenges that arise during the design of complex systems.

4. IV Rocket Chip for Research development

From a research point of view, in the AMForS group of TIMA we plan to use this platform to develop and validate methods and architectures to improve the reliability and robustness of SoCs:

- at system level (development of dedicated coprocessors such as Power Management Units, crypto processors, aging monitoring units, ...)
- at component level (enhancing the processor with fault tolerance features for example)

5. References

[1] RISC-V Foundation website: https://riscv.org/

[2] K. Asanović *et al.*, *The Rocket Chip Generator,* Technical Report No. UCB/EECS-2016-17, https://people.eecs.berkeley.edu/~krste/papers/EECS-2016-17.pdf

[3] LeaRnV: Understanding the hardware-software relationship using the RISC-V architecture and the Rocket Chip SoC platform, http://tima.univ-grenoble-alpes.fr/tima/fr/amfors/amforsoverview.html

[4] C. Celio, D. Patterson, K. Asanović, "The Berkeley Out-of-Order Machine (BOOM) Design Specification", University of California, Berkeley, December 4, 2016

[5] PULPino: An open-source microcontroller system based on RISC-V, https://pulp-platform.org

[6] J. Bachrach *et al.,* CHISEL: Constructing Hardware in a Scala Embedded Language, University of California, Berkeley, Design Automation Conference 2012, https://chisel.eecs.berkeley.edu/

[7] A. Waterman, Y. Lee et al. "Spike RISC-V ISA Simulator", website:https://github.com/riscv/riscv-isa-sim