AMfoRS team (Architectures and Methods for Resilient Systems)

Themes

Multi-level specification and verification of hardware/software on-chip architectures
System-level modeling, analysis and testing
Dependability of integrated systems and architectures
Secured integrated architectures
Multi-level dependability evaluations
Emerging architectures

Expertise

Scientific
Systems modeling, requirements formalization, temporal logics, automatic proofs, fault detection/tolerance, modeling of aging effects, methods for robustness analysis, hardware security, system-level testing

Fields of expertise
Correctness verification and dependability evaluation at various levels of the design flow, formal methods, dependability improvement, hardware security, system-level testing

Know-how
Analysis and formalization of requirements, formal and semi-formal verification, assertion-based design, fault tolerance or detection, fault injection, cryptographic accelerators, hardware attacks and counter-measures for secured circuits, IEEE 1687-compliant testing

Industrial transfer
Transfer of HORUS technology in EDA tools for mixed systems by Dolphin Integration
Transfer of ISIS technology in EDA tools by Dolphin Integration
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

Specification and verification of complex systems, assertion-based verification, correct-by-construction design, dependability, fault tolerance, security, aging, robustness analysis

Contact

Régis LEVEUGLE
GRENOBLE INP Professor
(+33) 4 76 57 46 86
regis.leveugle@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.

Similarly, structural test implies a non-interactive environment where input vectors are applied to the SUT and the output are compared with pre-computed expected vectors: the complete flow is depicted in Figure 1

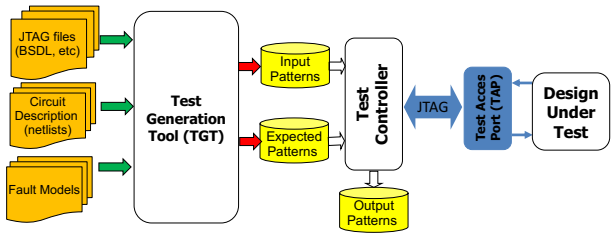

Figure 1: Traditional JTAG test environment

This flow is adapted for structural testing, but it is extremely inefficient for instrument usage, which implies the possibility of changing the test data and execution flow based on the actual outputs. Also, a given SUT might embed tens or even hundreds of instruments: an efficient handling of their concurrent execution must be provided. None of the current EDA tool is doing this.

2. Recent outcomes

The work has been focused on developing an innovative backend able to provide the features needed for the evolution of automated testing while maintaining a compatibility with IEEE standards and EDA tools. The goal is the development of a complementary solution to the current approaches, focusing on test execution rather than on test generation as it has been the case up until now.

The final proposal for a new Functional Execution Flow is based on State of the Art software and Operating System (OS) engineering and is depicted in Figure 2.

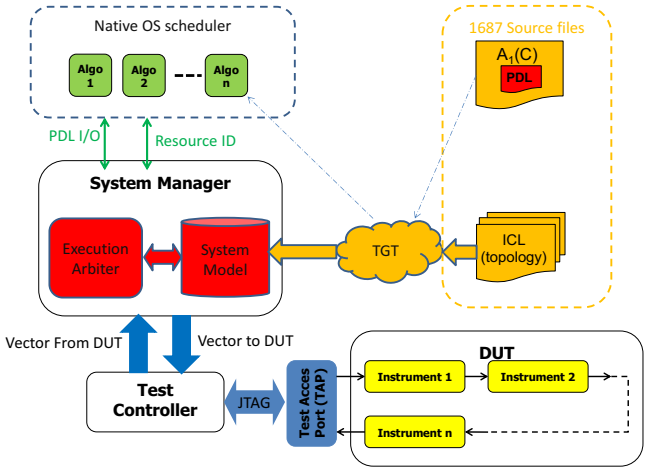

Figure 2: Proposed Functional Execution Flow

In this solution, each hardware instrument used for error and exception detection purposes, for instance (depicted in Yellow in the bottom left-hand corner) is associated with a Software Algorithm (depicted in Green in the top right-hand corner). Each algorithm A_n can access its related Instrument I_n thanks to an API based on the Procedural Description Language (PDL), defined in the IEEE 1687-2014 standard. These API requests are handled by an Execution Arbiter (EA) which polls them and translates them into system-level vectors thanks to an internal System Model (SM). This latter can be obtained from standard languages (ex: the Instrument Connectivity Language, ICL, defined by IEEE 1687-2014) through standard ADE tools, but it can also be extended to support custom structures. The System Model contains a copy of the state of the SUT and their coordination is assured by the

Execution Arbiter: when the SM and SUT states are desynchronized, the EA is able to generate a series of actions (i.e. scan operations) to restore synchronization. This behavior is exploited to perform dynamic topology configuration: PDL operations are posted as modifications to the SM, so that the EA can generate the necessarily actions to restore synchronization and therefore obtain the desired configuration.

This behavior has been completely specified and has been accompanied by an extensive analysis and generalization effort in order to define the minimal information set needed for the System Model. This knowledge has been used as the base for developing of piece of software, the MAnager for Soc Test (MAST) that implements the flow of Figure 2. The main reason for this effort is to validate the theoretical assumptions by applying them to industrial use cases, whose results can be used for scientific dissemination.

While 2016 was focused on scientific production, the fulcrum of 2017 was the Incubation of the project through Linksium with the target of creating a start-up in 2018.

On the technical side, MAST was enhanced to support as many features of the IEEE 1687 standard as possible, most noticeably by implementing a native ICL parser. A special attention was also put to packaging and user experience. We also explored applications of MAST in connected fields such for instance functional safety of mixed-signal testing [1].

On the organizational side, we explored various incubation options and launched a great survey with the main silicon and tool providers to find the best fit for MAST. This was done both remotely (email, phone) and in person, both in Grenoble and in major conferences such as the 2017 International Test Conference and 2017 European Test Symposium.

3. References

[1] M. Portolan, M. J. Barragan, R. Alhakim, S. Mir, “Mixed-signal BIST computation offloading using IEEE 1687”, 2017 22nd IEEE European Test Symposium (ETS), Year: 2017, Pages: 1 - 2, DOI: 10.1109/ETS.2017.7968222

Assertion-Based Verification for Systems on Chip

Keywords: Temporal properties for hardware and software, runtime monitoring

Members: L. Pierre, E. Brignon, M. Chabot

Contracts: SPICA (FUI)

1. Context

The design of today's systems on chip raises difficult issues, in particular regarding verification. To enhance the verification process, our on-going project (see Figure 1) targets the design and the implementation of techniques for the Assertion-Based Verification (ABV) of both *System-level* and *software requirements* of systems on chip and embedded systems [1].

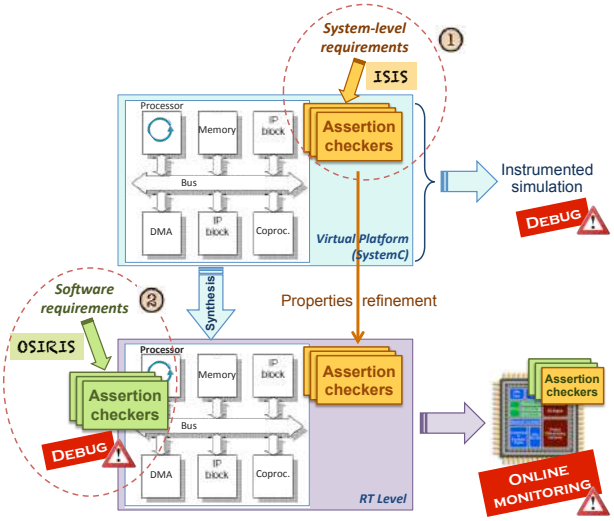

Figure 1: ABV for Systems on Chip - Current work

Two prototype tools have been developed for the automatic runtime monitoring of hardware/software requirements expressed as temporal assertions formalized using the PSL IEEE standard (Property Specification Language):

- ISIS [2] enables the automatic instrumentation of SystemC TLM virtual platforms with assertion checkers in order to check (during simulation) *system level requirements* on the interactions between hardware or hardware/software components,
- OSIRIS (still in progress) automatically generates assertion checkers from *software requirements* and instruments C programs with these verification components, together with an observation mechanism that enables their event-driven activation. It can instrument either C source codes or binary files.

Both types of assertion checkers can be used for debugging purposes.

Additionally, a refinement flow enables to convey the system-level properties along a synthesis procedure, thus allowing to check them at the Register Transfer level. At that level, they express *more accurate requirements*, refined to take into account actual protocols and timing constraints introduced during high level synthesis.

Ultimately, such concretized system-level assertion checkers, and the software requirements checkers, can be implemented with the system in the final device, to perform *online monitoring*.

2. On-going work

In this overall project, recent and on-going work is emphasized on Figure 1:

(1) Improvement of the checkers generated by ISIS. These components report satisfaction or violations of the assertions during simulation. However, a textual SystemC simulation trace can be several MB or GB large, and a large amount of checker messages may be scattered in such a simulation result. The analysis of the verification results can be extremely burdensome. To ease this analysis and to alleviate debugging activites, our verification infrastructure has been enhanced towards the *runtime identification of key verification events* (i.e., start, decision, and end of each PSL assertion evaluation) [3]. To that goal, those key events have been semantically characterized, and the checkers that are generated by ISIS can now produce additional messages pertaining to these events. Furthermore, structured verification data can also be stored in a database, and analyzed by a post-processing tool.

(2) New version of OSIRIS. This tool focuses on the *ABV of software requirements*. Two alternatives are proposed, either to instrument C source files with assertion checkers, or to dynamically instrument binary codes under execution. Our first version of the binary instrumentation solution requires suspensions of the software under observation; experimental results have shown that the resulting CPU time overhead could become prohibitive, in particular when all the checkers are kept enabled along the execution. A second version of this tool has been designed, for bare metal embedded applications. In this version, the assertion checkers are interrupt triggered. First experimental results demonstrate much better performances.

3. References

[1] L. Pierre: "Assertion-Based Verification IP's". Présentation à la Journée *Vérification en ligne du matériel au logiciel*, GDR SoC2, Dec. 2017.

[2] http://tima.univ-grenoble-alpes.fr/amfors/Isis/isis.html

[3] L. Pierre, M. Chabot: "Assertion-Based Verification for SoC Models and Identification of Key Events". Proc. Euromicro Conference on Digital System Design (DSD'2017), August 2017.

Testing Framework for Cyber-physical System Design

Keywords: Multi-disciplinary system design and requirements, test specification, test automation

Members: L. Pierre, M. Chabot, Y. Ducruy

Contracts: CIFRE PhD (with Schneider Electric)

1. Context

This work targets a specific methodology for the M&SBSE (Modeling and Simulation-Based Systems Engineering) of cyber-physical systems. The design of such systems, in which the operations of physical entities are controlled by computing cores, is particularly challenging since it requires the cooperation of many disciplines. We have defined a *requirement driven testing approach* to improve the design and verification processes. It is based on a *unified testing framework and a generic test infrastructure* [1]. As summarized on Figure 1, the advocated principle is to start from the requirements, and to create specialized test infrastructures (each of them is distinctive of a given standard or set of user requirements). Then the successive steps of the design process are as follows:

- these specific instances of the generic test infrastructure motivate a uniform system interface for the models of all disciplines that is delivered to the designers who create their models accordingly,
- from these abstract models, system designers then produce a simulatable model at a more concrete level (e.g., a Simulink model),
- in parallel, the specified behaviour of each tester is automatically translated into an executable scenario, and finally the validation engineer simply has to integrate the simulatable model in the test infrastructure and to run simulation. Note that this executable scenario can also be automatically transformed into a validation plan (e.g., for TestStand/Labview) to automate the validation of the final product with a test platform.

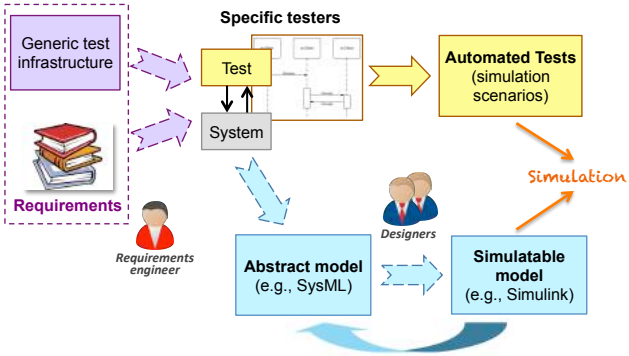

Figure 1: Requirement driven testing method

Structural aspects of the generic test infrastructure have previously been investigated with great care [1], and specified in SysML as Block Definition Diagrams (BDD) and Internal Block Diagrams (IBD). A generic Tester includes two interacting subcomponents: a TestCoordinator, and a Wrapper to encapsulate the system under test. The TestCoordinator includes an arbitrary number of Emulator components: each Emulator corresponds to one specific interaction with the system under test.

2. Test automation

More recently our work focused on the *behavioural aspects of this test infrastructure*. We have specified and implemented a method that enables to *automatically transform models of test scenarios* originating from requirements *into automated tests*, thus driving the verification process from early design phases to simulation, and ultimately to the validation of the final product [2].

We enable to express test scenarios as SysML Sequence Diagrams that specify the interactions between test emulators and the system under test. A test scenario is modeled as an ordered sequence of interactions. These interactions are either simple messages that enable to get or to set a value through the dedicated ports identified during the requirements analysis phase, or combined fragments that express repetitions of actions, optional or conditional actions, or parallel actions.

From such a SysML behavioural description of the test scenario, our tool automatically generates an executable scenario used to drive the simulation. Our method for *producing executable scenarios from sequence diagrams* consists in generating a finite state machine (FSM) able to communicate with the system through the specific ports. The communication events are modifications or readings of values in the system.

Great care must be taken in ensuring that the automatic translation process infers a total order for the actions of the tester. Our algorithm first produces a tree-like structure that represents this total order, then this tree is traversed to build the FSM. This FSM model can be concretized into various languages. Experiments have been performed for several Matlab/Simulink use cases; in that case, the FSM used to drive the simulation was expressed as a Simulink stateflow chart.

3. References

[1] M. Chabot, L. Pierre, A. Nabais-Moreno: “A Requirement Driven Testing Method for Multi-disciplinary System Design”. Proc. ACM/IEEE Int. Conference on Model Driven Engineering Languages and Systems (MODELS'2016), October 2016.

[2] L. Pierre: "Verification of Cyber-physical Systems: from Requirements to Automated Tests", presented at FETCH'2018, January 2018, https://sites.google.com/site/fetch2018a/

Synthorus-2: Automatic Compilation of Properties into Synthesizable Designs

Keywords: Correct by construction, Assertion-Based Design (ABD), PSL

Members: N. Javaheri, K. Morin-Allory, D. Borrione

In the context of verification, declarative properties about the behavior of a design (Assertions) or its environment (Assumptions) are checked using dynamic or static verification tools. Once refined down to the register transfer level (RTL), a complete set of assertions unambiguously characterizes how a module reacts to signals sent to it, logically and temporally. Assertions may be written in one of two IEEE standard languages: Property Specification Language (PSL) and System Verilog Assertions (SVA). Many tools are available to compile assertions into *monitors*, i.e. verification IP's that check the design correctness, either by simulation or emulation.

Our project considers the direct *production* of *compliant control and communication modules* from a set of assertions, which are seen as the specification of the module to be designed. We directly compile the synthesizable RTL design from its assertions. For each property, we obtain a compliant RTL component called *reactant*: its inputs and outputs are operands of the property, it reacts to the input values and produces output values so that the property holds.

This fast prototyping from assertions will not soon replace manual design, the results are not currently competitive. Still, it may come early in the design flow, and resulting reactants can be used for:

- Replacing a nonavailable module by a fast prototype of it, to check a more comprehensive design.
- Replacing the (possibly very complex) environment of a designed module by a fast prototype of just the part of the environment that interacts with it.
- Debugging specifications by generating properties to be model-checked, during the compilation process.

In contrast to most other works that are based on automata and game theory, our method is modular: it is based on the interconnection of primitive library modules for the logical and temporal operators of the property, according to its syntactic structure, a technology that we initially introduced to compile assertions into monitors (Horus). In the 2014 report, we illustrated the construction of the reactant for a property written in terms of temporal operators [1]. In 2015, we extended the method to SERE's (sequential regular expressions) [2]. To our knowledge, this is an original feature, other authors only reported processing linear temporal operators.

The years 2016 and 2017 were devoted to showing the proof of concept for producing reactants from SVA, and to formally proving the correctness of the reactant construction, using the PVS proof assistant (on the basis of the proof of monitors that had been performed ten years ago).

The main achievements of this research are:

- A new semantic definition of the logical and temporal operators of PSL and SVA, compliant with the standard formal semantics, which exhibits a dependence relation between the operator's operands. It expresses which operand may constrain the value of the other.
- A hardware interpretation for the dependence relation, foundation for our library of primitive reactants.
- A unification between the concepts of monitor and reactant, and the formal proof of correctness w.r.t. the formal semantics of PSL and SVA.
- When a same variable appears in several distinct properties, the automatic identification of which properties monitor the variable and which ones generate its value (*annotation*). If a variable is an output for several reactants, a solver is produced to generate the variable final value (*resolution*).

SynthHorus-2 is the software prototype tool that implements these principles [1,2]. It takes as input the interface declaration of the specified module and a set of properties written in the simple subset of PSL, and produces a RTL design in the synthesizable subset of VHDL (see Fig.1). The circuit is constructed as the interconnection of the reactants for all the properties, together with solvers. It is a register transfer level model that is input to a conventional industrial synthesis tool to obtain the final implementation, either on FPGA or on an ASIC.

Moreover, as an aid to specification debug, Synthorus-2 may generate a set of complementary properties to verify (by model checking) if the input set of properties are complete and consistent.

SynthHorus2 compiles several dozen properties in seconds, and produces a reasonably sized RTL circuit model. Experiments and performances are reported in [1,2].

Application prospects include the extension of our technology to the observation of real time operating systems. As a proof of concept, a preliminary study was started in cooperation with a group from Université de Nantes. We developed a RISC V based FPGA platform, and used Synthorus2 to instrument it, under the form of monitors capable of observing properties on the OS kernel.

References

[1] K. Morin-Allory, N.F. Javaheri, D. Borrione: "Efficient and Correct by Construction Assertion-Based Synthesis", IEEE Trans. on VLSI, Vol.23, N° 12, December 2015, pp. 2890-2901

[2] N.F. Javaheri, K. Morin-Allory, D. Borrione: "Synthesis of Regular Expressions Revisited: from PSL SEREs to Hardware", IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems, Vol. 36, N°5, May 2017, pp. 869-882, DOI: 10.1109/TCAD.2016.2600241

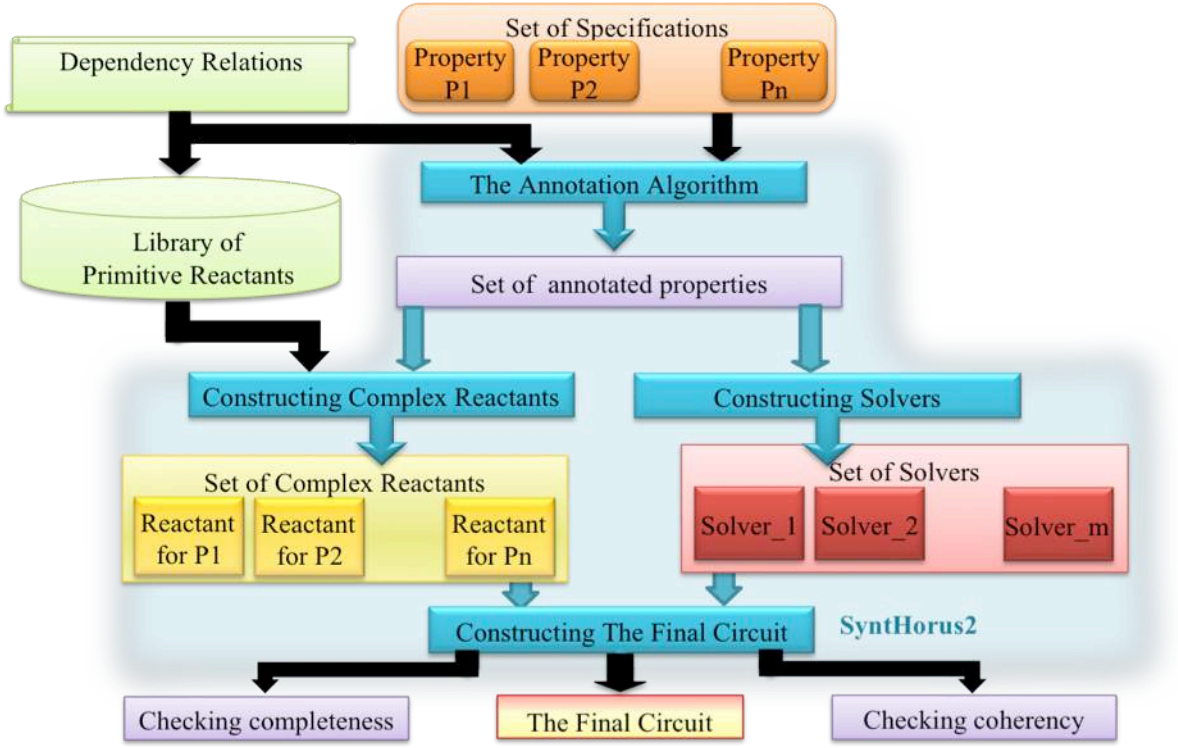

Fig.1: The SyntHorus-2 Flow

Clock Domain Crossing StaticVerification

Keywords: Multi Clock Circuits, GALS, Clock Domain Crossing, Formal Verification, Model Checking

Members: M. Kebaili, G. Plassan, K. Morin-Allory, D. Borrione

Cooperations: STMicroelectronics, Synopsys

Contracts: CIFRE contracts for M. Kebaili and G. Plassan

The design of large circuits assembles IP's and blocks coming from various design teams, each with their own clock. For instance, circuits for games or set top box applications easily include in excess of 25 blocks. To guarantee low power consumption, the clock of a block is set at the lowest frequency compatible with the execution time required the block. Thus, all the clocks have their own speed, and there is no guarantee that clock cycles are multiple of a basic master clock nor that two clocks are phased. Each clock constitutes a clock domain, and any two clock domains are considered mutually asynchronous: the design is globally asynchronous locally synchronous (GALS).

Clock domain crossing (*CDC*) is the transmission of information between two clock domains. A synchronizing module (*synchronizer*) is needed to connect a source signal, output of a flip-flop in a transmitter domain, to the input of a destination flip-flop in a receiver domain, because the sampling by the receiver clock may happen before the input signal has reached its correct stable value. It is therefore essential to guarantee the correctness of the communication protocol and the synchro-nization between the modules.

Major EDA providers propose CDC static checkers to perform two kinds of verification (Fig.1):

1. Structural verification detects the synchro-nizers in an input hardware description (at RTL or Gate level), using a library of patterns describing the synchronization structures.
2. Functional verification calls upon a Model-Checking engine to prove formal properties on each model identified during the structural ver-ification. In most CDC checkers, the properties depend on the type of synchronizer detected.

All tools recognize the control-based model of synchronizer, which is the minimal synchronization structure; typical handshake and FIFO structures may also be identified [1]. For each structure, a set of properties is associated to its recognition pattern.

In industry, designers often develop their own synchronizers for performance reasons. As a result, CAD software fails to recognize them in a flat design, or the recognition is only partial. A strong interaction with the verification engineer is needed to provide constraints, case by case, which is both costly and non reusable. We investigated this problem in parallel along two directions: the definition a meta model for synchronizers, and a new strategy to deal with the state explosion in model checking.

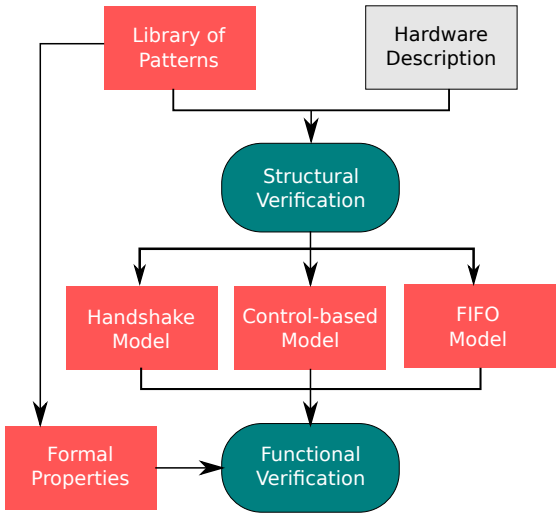

Fig. 1: State-of-the-art CDC verification flow

1. Meta-Model for Synchronizers

The meta-model aims at providing a fully automated CDC verification, by a complete recognition of all synchronization structures. It is a generalization of the handshake model. We have defined a generalized pattern, and an associated comprehensive set of properties to be formally checked: absence of metastability, absence of data loss and data consistency.

The meta-model has been applied to a multiprocessor subsystem design from STMicroelectronics, containing 129 data synchronizers. The CDC formal verification using the meta-model was completed in less than 40 minutes, while standard tools required over 45 hours [3].

2. The UsAAR Methodology in Model Checking

In even bigger designs, a conclusive answer (proven/failed) is not always obtained from the functional verification step of Fig. 1, with the tools of EDA providers. The verification of some properties reaches a timeout even after several days. The typical solution would be, for each property, to extract the CDC logic and focus the verification on just the small relevant part of the design. Yet, this manual effort is not realistic for large RTL designs.

The key idea of our method is to let the user aid the model checking process by replying yes/no to a series of questions whose answers only require local design knowledge (see Fig.2).

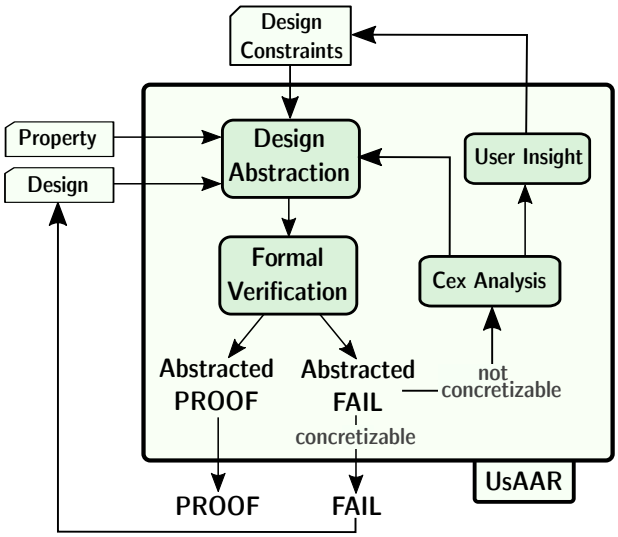

Fig.2: CDC verification with UsAAR

Technically, our underlying framework is a counter-example-guided abstraction refinement (CEGAR) algorithm: we maintain a sequence of abstractions with increasing precision until a definite result can be established. In contrast to fully automatic CEGAR approaches, the user here influences the refinement process. We therefore call our approach user-aided abstraction refinement (*UsAAR*) [2].

UsAAR was applied on a complex SoC design from STMicroelectronics, which holds over 300,000 registers and 7 million gates, with 38 clock domains and 17 primary resets. Compared to standard model checking with manual constraints, and to CEGAR-based model checking, the UsAAR method was the only one to prove all the 125 absence of metastability and data consistency properties, with a performance 20 times better than the most efficient other method [2].

3. Automatic generation of missing constraints

During the formal functional verification of RTL designs, a false failure is often observed. Most of the time, this failure is caused by an under-constrained model. The analysis of the root cause for the verification error and the creation of missing assumptions are a significant time burden. We developed an algorithm which automatically infers these missing assumptions from the counter-examples provided by the model checker upon failure. First, we force the model checker to generate multiple counter-examples. Then, we extract common root causes of the failure from the set of counter-examples, using mining techniques combined with a structural analysis of the netlist. Finally, we generate realistic temporal assumptions for the user to review. The validity of our methodology was shown on two academic and one big industrial design [3].

4. References

[1] M. Kebaili, J.C. Brignone, K. Morin-Allory: “Clock Domain Crossing Formal Verification: a Meta-Model", IEEE Int. High Level Design Validation and Test Workshop (HLDVT), pp. 136-141, Santa Cruz, USA, 7-8 Oct. 2016

[2] G. Plassan, H.J. Peter, K. Morin-Allory, S. Sarwary, D. Borrione: “Improving the Efficiency of Formal Verification: The Case of Clock-Domain Crossings”, In “VLSI-SoC: System-on-Chip in the Nanoscale Era – Design, Verification and Reliability” (revised selected contributions from VLSI-SoC'16), Ed T. Hollstein, J. Taik et al., IFIPAICT vol 508, Springer, pp 108-129, Sept. 2017

[3] G. Plassan, K. Morin-Allory, D. Borrione: “Extraction of missing formal assumptions in under-constrained designs”, Proceedings of the 15th ACM-IEEE International Conference on Formal Methods and Models for System Design (MEMOCODE’17), Vienna, Austria — Sept. 29 – Oct. 2, 2017, pp. 94-103

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

Cooperations: STMicroelectronics, LCIS, LIRMM

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, at design time and 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. Higher level descriptions may in some cases be used, with limited representation of soft error locations and reduced accuracy in terms of propagation analysis. 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 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 (with 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 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. Further work is planned on this subject and collaborative work with LIRMM laboratory has been started [3].

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 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, A. Mkhinini, V. Reynaud

Cooperations: Faculté des Sciences de Monastir (Tunisia), Univ. Sousse - ENISo (Tunisia), ENSMSE, STMicroelectronics, Thales, LIRMM

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

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 protocols and algorithms, which can be implemented in software or hardware according to the performance requirements, and to the cost constraints. Cryptography is at the heart of those systems.

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 the crypto-processor implementations, but 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 analysis. 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 symmetric (AES and lightweight implementations) and asymmetric (mainly ECC) cryptosystems. Work was also carried out on hardware implementations for Full Homomorphic Encryption (FHE) and new work is dedicated to hierarchical access authorizations to embedded test and monitoring devices.

2. Recent outcomes

A significant part of the work in 2017 was targeted towards hardware acceleration of Full Homomorphic Encryption (FHE), in collaboration with Faculté des Sciences de Monastir and ENISo (Tunisia). Some results have been published in [1] and the full work has been reported in [2]. It was demonstrated that efficient and flexible implementations can be obtained using High Level Synthesis (HLS). The tool used is AUGH that is developed at TIMA in the SLS team. Results illustrated in Figure 1 show for example that polynomial multipliers with very large size of coefficients and/or very large degrees can be efficiently generated with little effort while characteristics of hand-made designs proposed in the literature are much more limited. A complete prototype has been demonstrated for the acceleration of the scheme FV.

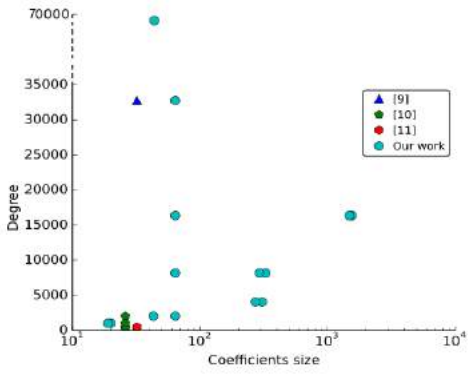

Figure 1:
Different parameters (degree, coefficients size) supported with our work compared to the state of the art [1]

A project launched in 2017 aims at experiencing a new second-order laser attack on AES crypto-processors protected by specific time redundancy (so-called DDR approach). The work is based on circuits manufactured in a previous project (ANR LIESSE) by our partner STMicroelectronics in 28 nm bulk and FDSOI technologies. It is carried out in collaboration with our colleagues from ENSMSE in Gardanne. First experimental results are under analysis and new experiments are planned in 2018.

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. In relation with this context, we are also working on a flexible generation environment for lightweight crypto-processors, including several algorithms, several types of hardware counter-measures, self-test features, … all being optimized for both ASIC and FPGA implementations.

3. References

[1] A. Mkhinini, P. Maistri, R. Leveugle, R. Tourki*, "HLS design of a hardware accelerator for homomorphic encryption", IEEE International Symposium on Design and Diagnostics of Electronic Circuits and Systems (DDECS), Dresden, Germany, April 19-21, 2017
* Univ. Monastir, EμE laboratory, Tunisia and Univ. Sousse, ENISo, Tunisia

[2] A. Mkhinini, " Implantation matérielle de chiffrements homomorphiques", http://tima.univ-grenoble-alpes.fr/tima/fr/mediatheque/PhDthesisresult_id467.html

RT-Level design for reliability/safety/availability and/or security

Keywords: dependability improvement, control flow checking, behavioral checking

Members: R. Leveugle, L. Terras

Cooperations: STMicroelectronics, Dolphin Integration

Contracts: SPICA (FUI), CIFRE STMicroelectronics

1. Context and goals

Protecting a design against natural perturbations or malicious attacks can be done at several levels. We mostly focus here on approaches that can be applied at RT-Level, therefore quite early in the design flow and easy to synthesize on several physical targets (several ASIC technologies, FPGAs …). Previous studies in the team also included operating systems or software modifications but we mainly focus on hardware protection techniques. Some protections aim at improving reliability, safety and/or availability against natural perturbations (radiations, particles, electromagnetic fields) and other disturbances caused for example by process, voltage and temperature (PVT) variations. Some others are dedicated at improving security against malicious attacks, either passive (based e.g., on power or electromagnetic measures) or active (laser-, glitch- or electromagnetic-based perturbations). For protections specifically dedicated to security, see the section entitled "Design of secured crypto-processors"; these aspects will not be developed in this section. We will focus in this part on approaches used for natural perturbations but that may also be used in a security context for fault-based attacks only.

Due to the increasing spatial multiplicity of error patterns, protecting a circuit with information redundancy is more and more difficult. This is particularly true when malicious attacks are concerned, but the problem exists also for natural perturbations. Another approach consists in using functional checks. In this context, we had proposed and demonstrated a new approach for microprocessor-based systems, called IDSM. Checks include not only the control flow, but also the validity of operations and the integrity of critical data, with several possible trade-offs between overheads and error detection. The approach allows checking not only the main application program, but also the boot phase and the calls to external functions, for which the source code is not available (e.g., pre-compiled library functions). No assumption is made on the error multiplicity.

As initially proposed, the approach is not intrusive and compatible with the norms requiring a complete separation between the nominal functions and the checking features (e.g., for automotive applications). Prototypes had been developed including (1) a specific watchdog processor (or infrastructure IP) and (2) development tools. The watchdog program was automatically generated at compile time by a modified version of the GCC compiler. Additional tools had been developed to cope with linkage constraints. A prototype had been demonstrated on the Leon3 processor. Overheads were smaller than those induced by the classical lock-step duplication. Another advantage with respect to duplication is to ensure diversity, making successful malicious attacks much more difficult to achieve.

2. Recent outcomes

The work done these last years in the SPICA project aims at transferring the principles of the approach on other processors in industrial contexts. In 2015, after a first evaluation on an ARM core, the focus was put on a ST proprietary microprocessor. The approach has been revisited, taking the new micro-architectures as well as additional industrial constraints into account. In particular, the need for taking care of indirect jumps or calls arose and was carefully analyzed [1]. Such a constraint is not covered by control-flow techniques previously published but has been identified as unavoidable in the considered industrial context. An approach to resynchronize the watchdog after an indirect branch, at the expense of limited intrusiveness in the initial system, has been proposed and evaluated [1]. A full demonstrator has been developed for this microprocessor in 2017. In parallel, work has been carried out by our second industrial partner (Dolphin Integration) and showed with similar constraints the feasibility of applying the approach on their own microprocessors.

3. References

[1] L. Terras, Y. Teglia*, M. Agoyan*, R. Leveugle, “Taking into account indirect jumps or calls in continuous Control-Flow Checking”, 11th IEEE International Design & Test Symposium (IDT), Hammamet, Tunisia, December 18-20, 2016, pp. 125-130
* STMicroelectronics, Rousset, France

Adaptive Routing for Fault Tolerance and Congestion Avoidance for 2D Torus NoCs in Many-Core Systems-on-Chip

Keywords: Networks on Chip; fault tolerant adaptive routing; reliability; 2D Mesh; 2D Torus; congestion; packet-retransmission; permanent faults; transient faults; intermittent faults; Many-Core; Systems-on-Chip

Members: M. Benabdenbi, L. Anghel

According to the International Technolo y Roadmap for Semiconductors (ITRS), as we move towards the post-CMOS era with the continuous downscaling of the feature sizes, the lowering of power supply voltages and the increasing of operating frequencies, the reliability of circuits is threatened by the increased process-voltage and frequency variations. In these environments, the designs should be able to provide either full functionality (e.g. full critical systems), or degraded one in the case of consumer applications even in the presence of high failure rates. To accomplish this, the systems should be able to adapt to manufacturing and runtime failures and continue their functioning.

Error resilience in on-chip networks is addressed in this work. Although, the presented solutions tackles the problem at the network level, it is orthogonal to other developed solutions at other levels (e.g. link-level, end to end recovery) and can work complementarily to them.

We proposed in 2015 an error resilient solution designed for 2D Mesh NoCs [1]. This solution is based on the use of local information such as state of links and routers. It consists of an adaptive fault tolerant routing algorithm named CAFTA (Congestion Aware Fault Tolerant Algorithm) enhanced with neighbor fault-aware knowledge. In order to improve transmission latencies and to effectively guide the routing decisions towards load-balanced network traffic, a new metric is introduced. In the presence of runtime errors, packet retransmission combined with novel message recovery mechanisms are utilized in order to provide fault tolerance under high failure rates. Extensive simulation experiments under various fault types (permanent, transient, intermittent) commonly encountered in advanced technology nodes demonstrated the effectiveness of the proposed fault tolerant scheme.

To evaluate the impact of the network topology on CAFTA efficiency, we extended the proposed algorithms to a 2D torus Network-on-Chip [2]. Due to its higher connectivity, torus topology offers more connectivity in terms of potential paths between any source and destination pair. The main difference between Mesh and Torus topology is that for torus NoCs vertical and horizontal wrapping links are added.

Before extending CAFTA to this topology, we must guarantee that the routing algorithm is deadlock free. Based on Dally theorem [20], a routing algorithm is deadlock free if and only if its channel dependency graph is acyclic. Variant B, the used routing algorithm, for 2D Mesh is proved deadlock free, but is it still the case when introducing the additional Torus wrapping links? We answered this question by defining a set of turn restrictions adapted to this torus topology. Extensive simulation experiments for different NoC sizes, traffic types, traffic loads and fault injections allowed us to compare the behavior of the 2D Mesh and the 2D Torus topologies.

Compared to the mesh topology, the experimental results show that the saturation threshold is improved on average of 20 % for all traffic types.

Another important result consists in the packet delivery rate for torus topology (Fig. 1). In the case of random uniform traffic, more than 99, 6 % of the packets reach their destination even in the presence of 40 % of faulty links. This ratio is higher than the case of mesh topology.

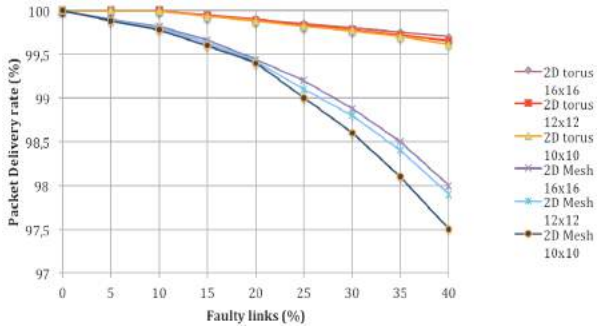

Figure: Packet delivery rate with faulty links: torus vs mesh

References

[1] M. Dimopoulos, Y. Gang, L. Anghel, M. Benabdenbi, N.E. Zergainoh, M. Nicolaidis, Fault-Tolerant "Adaptive Routing under an Unconstrained Set of Node and Link Failures for Many-Core Systems-on-Chip", Microprocessors and Microsystems journal, Ed. Elsevier, Vol. 38, No. 6, pp. 620–635,

[2] M. Benabdenbi, L. Anghel, M. Dimopoulos, Y. Gang "Adaptive Routing for Fault Tolerance and Congestion Avoidance for 2D Mesh and Torus NoCs in Many-Core Systems-on-Chip", Chapter of book "Advances in Microelectronics: Reviews" Vol. 1, december 2017, IFSA Publishing, pp 405-435.

Digital System Failure Prediction due to Variability and Aging Induced Threats: Evaluation, Methodology of Implementation and Application Results

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

Members: L. Anghel, A. Benhassain, C. Ndiaye, A. Sivadasan, R. Shah

Cooperations: STMicroelectronics Crolles

Contracts: CIFRE contracts

1. Context and goals

With CMOS technology scaling, it becomes more and more difficult to guarantee circuit functionality for all process, voltage, temperature (PVT) corners and compensate for different sources of variability. Moreover, circuit wear-out degradation lead to additional temporal variations, potentially resulting in timing and functional failures. Under normal operation conditions, a transistor can be affected by various aging effects such as Hot Carrier Injection (HCI), Bias Temperature Instability (BTI), and Time Dependent Dielectric Breakdown (TDDB). In these advanced technologies, local and global variability, aging phenomena, such as NBTI and HCI impact have become the most critical reliability issues. Hence, taking into account these phenomena during the logic cells characterization step, but also at the circuit and system design and validation levels are mandatory, especially for high reliable application such as automotive applications, or mixed critical applications. In fact, these reliability threats can severely degrade the performance and in the worst case they can provoke system failures, affecting safety goals of specific reliable systems (automotive or health care embedded applications).

Therefore, accurate simulations with physical degradation models of the aging phenomena combined with silicon actual measures are necessary to better understand and assess the reliability impact on a complex digital design. A conventional method handling such problems is to provide more safety margins (called guardbands) in the circuit design phase. Adding pessimistic timing margins to guarantee all operating points under worse case conditions is not acceptable anymore due to the huge impact on design costs, such as up to 10% increase of slack time, with an upward trend as technology moves further.

Therefore, the usage of in-situ monitors for error and pre-error detection becomes a must, as they allow decreasing the constraints imposed on the overall design. In addition to the reduction of design margin, adaptive voltage scaling (AVS) technique or Dynamic Voltage Frequency Scaling (DVFS) triggered by pre-errors in-situ timing monitors may be used to adapt dynamically the frequency and the voltage according to the operating conditions and the application needs [1]. Thus, the performance degradation is 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 2017:

- a new methodology for pre-error monitor insertion in digital block without workload and pattern availability has been proposed. Basically, the timing of data path arriving to an endpoint register is analyzed, and a weight is calculated as a figure of merit. This work accounts for stochastic dispersion, aging, global corner process, voltage and temperature variations. The important role played by the sub- critical path is illustrated with silicon measurement. [2]. This work completes the critical path selection strategy developed in the past.
- Another aspect was to finalize the process variation and aging sensitivity estimation methodology applied to a large digital circuit. The main idea was to to understand how the critical paths are aging. For doing that we used gate-level models able to capture the aging induced stress experienced by the corresponding standard cell. We performed simulation of several different digital cells with different duty cycles and showed the importance of the activity and thus the workload in the assessment of the delay increase of a gate due to aging. [3].
- The simulation flow leading to the identification of specific workloads provoking critical paths aging has also been proposed. By applying this flow the critical path ranking was observed as changing in time. This allowed us to properly quantify aging margins for a specific worst case or average workload. The accurate estimation of workload dependent aged critical path ranking allows further for a judicious use of in-situ performance monitors without an increase in area overhead. [4]

3. Perspectives

Body bias management by using In Situ Monitors feedback will be implemented and demonstrated with post silicon measurements. These data will be used to enhance delay models of standard cell designs but also to accurately compute the path delay, which will be useful to predict the lifetime of a complex system on chip or a product.

4. 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] F. Cacho; A. Benhassain; R. Shah; S. Mhira; V. Huard; L. Anghel, « Investigation of critical path selection for in-situ monitors insertion », 2017 IEEE 23rd International Symposium on On-Line Testing and Robust System Design (IOLTS), 2017, pages 247-252

[3] Sivadasan A., Benhassain A., Huard V., Cacho F., Anghel L., Architecture and Workload Dependant Digital Failure Rate, IEEE International Reliability for Physics of Semiconductors (IRPS 2017), Monterey, UNITED STATES, 2 au 6 avril 2017

[4] Sivadasan A., Huard V., Anghel L., Worload Dependent Reliability Timing Analysis Flow, DATE 2017, Lausanne, SWITZERLAND, 27 au 29 mars 2017

Aging Monitoring and Adaptation using IEEE 1687

Keywords: aging, IEEE 1687, DVFS, Adaptive Monitoring

Members: L. Anghel, M. Portolan, K. Senthamarai Kannan

Contracts: HADES (Penta)

1. Context and goals

With CMOS technology scaling, it becomes more and more difficult to guarantee circuit functionality for all process, voltage, temperature (PVT) corners and compensate for different sources of variability. Moreover, circuit wear-out degradation lead to additional temporal variations, potentially resulting in timing and functional failures. Under normal operation conditions, a transistor can be affected by various aging effects such as Hot Carrier Injection (HCI), Bias Temperature Instability (BTI), and Time Dependent Dielectric Breakdown (TDDB). In these advanced technologies, local and global variability, aging phenomena, such as NBTI and HCI impact have become the most critical reliability issues. Hence, taking into account these phenomena during the logic cells characterization step, but also at the circuit and system design and validation levels are mandatory, especially for high reliable application such as automotive applications, or mixed critical applications. In fact, these reliability threats can severely degrade the performance and in the worst case they can provoke system failures, affecting safety goals of specific reliable systems (automotive or health care embedded applications). A conventional handling method for such problems is to provide more safety margins (called guardbands) in the circuit design phase. Adding pessimistic timing margins (or equivalent voltage margins) to guarantee all operating points under worse case conditions is not acceptable anymore due to the huge impact on design costs, with an upward trend as technology moves further.

To reduce performance loss and design constraints, delays violation monitors can be inserted into the circuit to observe transistor degradation and adapt operating conditions to mitigate them, thanks to techniques such as Adaptive Voltage Scaling (AVS) technique or Dynamic Voltage Frequency Scaling (DVFS). Monitors are usually inserted 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, which is of course dependent on the workload. This means that:

- Aging is not necessarily coherent with STA results, as near-critical paths may be seldom used. Experiments demonstrated how STA performed on aged circuits can deliver a significantly different set of near-critical paths depending on the workload;
- Other than the workload, the usage of AVF/DVFS techniques directly impacts aging, making priori evaluation impossible

To tackle this issue, we decided to explore two complementary directions:

- Find a near-optimal monitor placement strategy based on predictive aging modeling;
- Develop online AVS/DVFS approaches which take into account both aging and performance issues.

In both cases, we prioritize data-driven automated learning approaches to model the specific near-unpredictable behavior of transistor aging.

2. Recent outcomes

In this context, the IEEE 1687 standard is an interesting opportunity because it allows both easy access to deeply embedded instruments and the definition of algorithmic operations. By defining a 1687-compliant infrastructure for both monitors and DVFS, the developing effort would be greatly reduced while in the same time boosting reusability. Moreover, the MAST software developed by TIMA couples traditional 1687 access with state-of-the-art functional approaches: it is therefore possible to ally run-time monitor access, on-the-fly data mining, feature extraction and decision-making to obtain a new, innovative and complete solution.

The PhD of K. Senthamarai Kannan, started in November 2017, focuses on developing the theoretical and practical foundations for this approach, with a special emphasis on the usage of machine learning.

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

Keywords: neuromorphic computing, low power, in-memory computing

1. Design for Testability and Reliability of In-Memory Computing Applications

Members: L. Anghel, E.I. Vatajelu

Cooperations: IMEP-LAHC, LIRMM, TU Delft

Today's computing systems are facing a plethora of issues, amongst which, the so called "memory wall" poses important problems in terms of speed and power consumption, limiting the Central Processing Unit's (CPU) capabilities. This is due to the growing disparity of speed between CPU and memory outside the CPU chip [1]. An important reason for this disparity is the limited communication bandwidth beyond chip boundaries. The trend of placing more and more cores on chip exacerbates the problem, since each core enjoys a relatively narrower channel to shared memory resources. 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. All these issues, together with the promising development of non-volatile technology devices, open the path for developing brand new computing paradigms such as in-memory computing. This is a non von-Neumann computing paradigm.

In-Memory Computing (IMC) concept is based on the tight integration of memory elements and computational circuitry, to minimize the time and the energy needed to move data across the processor. Actual research efforts are directed to proving the effectiveness of memristors to achieve the in-memory computing ambition through the development of new architectures with low cost, high reliability and low power consumption. In this context, **our goal is on the development of models, characterization and methods targeting the enhancement of the dependability of the non-volatile-based in-memory computing basic blocs.** We will address circuit design, test and reliability challenges related to resistive technologies, from the computing-memory cell up to the In-Memory Computing architecture.

Our current research and collaborations in relation with this topic are threefold: (i) to use enhanced compact models of memristive devices to perform an in-depth failure analysis and define pertinent fault models; (ii) to investigate meaningful reliability threats affecting memristive-based in-memory computing modules and architectures, and derive solutions for their mitigation (Design-for-Reliability), (iii) to establish design and test methodologies for these devices and architectures (2).

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, P. Prinetto, M. Taouil, S. Hamdioui: Challenges and Solutions in Emerging Memory Testing, IEEE Transactions on Emerging Topics in Computing, Ed. IEEE, Vol. PP, No. 99, DOI: 10.1109/TETC.2017.2691263, 2017

2. Test and Reliability for Switching Nano-Crossbar Arrays

Members: L. Anghel, E. I. Vatajelu

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

Project support: H2020 RISE NANOxCOMP

Beyond CMOS, new technologies are emerging to extend electronic systems with features unavailable to silicon-based devices. Emerging technologies provide new logic and interconnection structures for computation, storage and communication that may require new design paradigms, and therefore trigger the development of a new generation of design automation tools. In the last decade, several emerging technologies have been proposed and the time has come for studying new ad-hoc techniques and tools for logic synthesis, physical design and testing.

The main goal of the project and the overall 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. The proposed methodology leads to the implementation of logic, arithmetic, and memory elements by considering performance parameters such as area, delay, power dissipation, and also reliability threats. As presented in the literature, emerging devices such as spintronic, memristor but also auto-assembly nano-crossbars arrays are not only prone to fabrication defects, but also sensitive to variations, intermittent and transient faults, and the resulting defect rates can be very high (up to 15 %). To tolerate high defect rates and variations, we focus on *defect tolerance* techniques to improve the manufacturing yield combined with *fault tolerance* design for lifetime reliability and *variation mitigation* to ensure the predictability and performance. *Adaptive* and *built-in* defect, variation and fault tolerant design flows, fundamentally different from conventional

approaches, are proposed in which the objective is to ensure high manufacturing yield and runtime reliability of the circuit at extremely low costs.

Self-Test and Repair solutions for nano-crossbars have been proposed in [1]. Non-conventional defect and fault tolerant techniques are also under studies combining spatial and temporal redundancy, that includes adaptive self-mapping architectures of logic properties of the design on the valid nano-wires.

References

[1] D. Alexandrescu, M. Altun, L. Anghel, V. Ciriani, M. Tahoori, A. Bernasconi, Logic synthesis and testing techniques for switching nano-crossbar arrays, Microprocessors and Microsystems, Ed. Elsevier, Vol. 54, pp. 14-25, DOI: 10.1016/j.micpro.2017.08.004, octobre 2017

Dependable Low Power Spiking Neural Networks based on Emerging Technologies

Keywords: spiking neural networks, reliability of spintronic based circuits, MTJ spintronic synapses

Members: L. Anghel, E.I. Vatajelu

Cooperation: SPINTEC-Grenoble

Project support: EU H2020 NANOxCOMP

1. Context and goals

The power, reliability and technological issues of today's memories have led to intensive research of emergent memory technologies and emerging computing paradigms. One of the most promising emergent technology is the Spin-Transfer-Torque Magnetic Random Access Memory (STT-MRAM). It has the great advantage of favouring increasing system complexity and performance, while being CMOS compatible. Computation paradigms, such as hardware based neuromorphic computing, unfeasible a few years back due to technological limitations, can take profit from this technology. The usability of this type of devices in neural network hardware designs has been demonstrated recently. Most of these works state that emerging memory devices are ideal candidates for synaptic weight implementations. Moreover, in some cases, these devices have been shown to have good potential for artificial neuron implementation as well [1]. In these works, compound synaptic device, (CSS), consisting of multiple parallel MTJs working in the stochastic regime have been investigated together with MTJ-based stochastic spiking neuron (SSN) circuits [2].

The impact of process variation during the learning stage of a stochastic MTJ-based neuron and synapse that use bio-inspired learning rules has been investigated and presented in [3]. In this work we have presented a comprehensive reliability analysis of spintronic-based functional modules (a compound spintronic synapse and spiking neuron, proposed in [1]) for spiking neural networks by providing an in-depth study of meaningful threats. The effect of cycle-to-cycle and device-to-device variability has been evaluated, as well as the effect of non-nominal environmental conditions on the learning and inference operation modes of the spintronic functional module.

Furthermore, in paper [4] we have presented an in-depth analysis of the behaviour of a fully-connected single layer STT-MTJ-based spiking neural network designed for pattern recognition under the influence of fabrication- and environmental-induced variability. Synaptic variability has been shown to affect the efficiency of depression and potentiation during the learning mode, while neuron variability has been shown to cause deviation of the spiking time and modify the relation between pre- and post-synaptic spiking. This could lead to imprecise or slow learning process, that in turn translates in the necessity of a larger number of samples in the training set.

At the network level, accounting for these variations and their impacts leads us to the following conclusion: the process variability has indeed an effect on the learning process of the SNN under study, but this effect can be mitigated by increasing the size of the network or/and the size of the training set. Our simulations show that for the minimum resolution synaptic weight, i.e., synapse designed with 1 MTJ device, using the minimum assumed training set, the recognition error can take values between 40.1 % and 72 %, while when using larger resolution synaptic weight (synapse designed with 16 MTJ devices) using the maximum assumed training set, the recognition error can take values between 7 % and 12.4 %. Cycle-to-cycle variation coming from the stochasticity of the write operation shows an impact on the recognition error as high as 30 % when 1 MTJ synapse is used with minimum size training set. However, when the SNN was designed with 16 MTJ compound synapse and the full training set was used, the recognition error remained basically the same. Another important fact observed when performing these simulations was that increasing the size of the training set is a more efficient means to mitigate the effect of cycle-to-cycle variations, than increasing the resolution of the synaptic weight. However, increasing the size of the training set leads to increased energy consumption, while increasing the resolution of the synaptic weight leads to larger area footprint and larger power consumption. These trade-offs remain to be validated in the current and future works.

2. Perspectives: Towards Dependable Brain-Inspired Modular Architectures for Neuromorphic Computing

In future research, we'll target hardware design of the neural networks consisting in hardware design of hierarchical modular neural networks (rather than stratified networks, as it is the case for the deep neural networks) with modules re-usable over multiple applications, thanks to the ability to learn and adapt on-line. Moreover, such an approach allows to design and connect separate neural modules (much like standard IPs are being connected today), for the execution of the desired function, thus reducing the complexity of

interconnects. The dependability of such networks will be analysed, as the inherent fault-tolerance property may degrade due to strong restrictions on the size of the network. Accounting for dependability threats and using mitigation techniques when necessary, brings a new degree of freedom to the proposed neural architecture design.

3. References

[1] D. Zhang, L. Zeng, Y. Zhang, W. Zhao, J. O. Klein, “Stochastic spintronic device based synapses and spiking neurons for neuromorphic computation,” IEEE/ACM International Symposium on Nanoscale Architectures, pp. 173-178, 2016.

[2] D. Zhang, et al., “All Spin Artificial Neural Networks Based on Compound Spintronic Synapse and Neuron,” IEEE Transactions on Biomedical Circuits and Systems, vol. 10, no. 4, pp. 828-836, 2016.

[3] E. I. Vatajelu, L. Anghel, “Reliability Analysis of MTJ-based Functional Module for Neuromorphic Computing,” IEEE 23nd International Symposium on On-Line Testing and Robust System Design (IOLTS), 2017.

[4] E. I. Vatajelu, L. Anghel, Fully-Connected Single-Layer STT-MTJ-based Spiking Neural Network under Process Variability, ACM/IEEE International Symposium on Nanoscale Architectures (NANOARCH 2017), Newport, RI, UNITED STATES, 25 au 29 juillet 2017