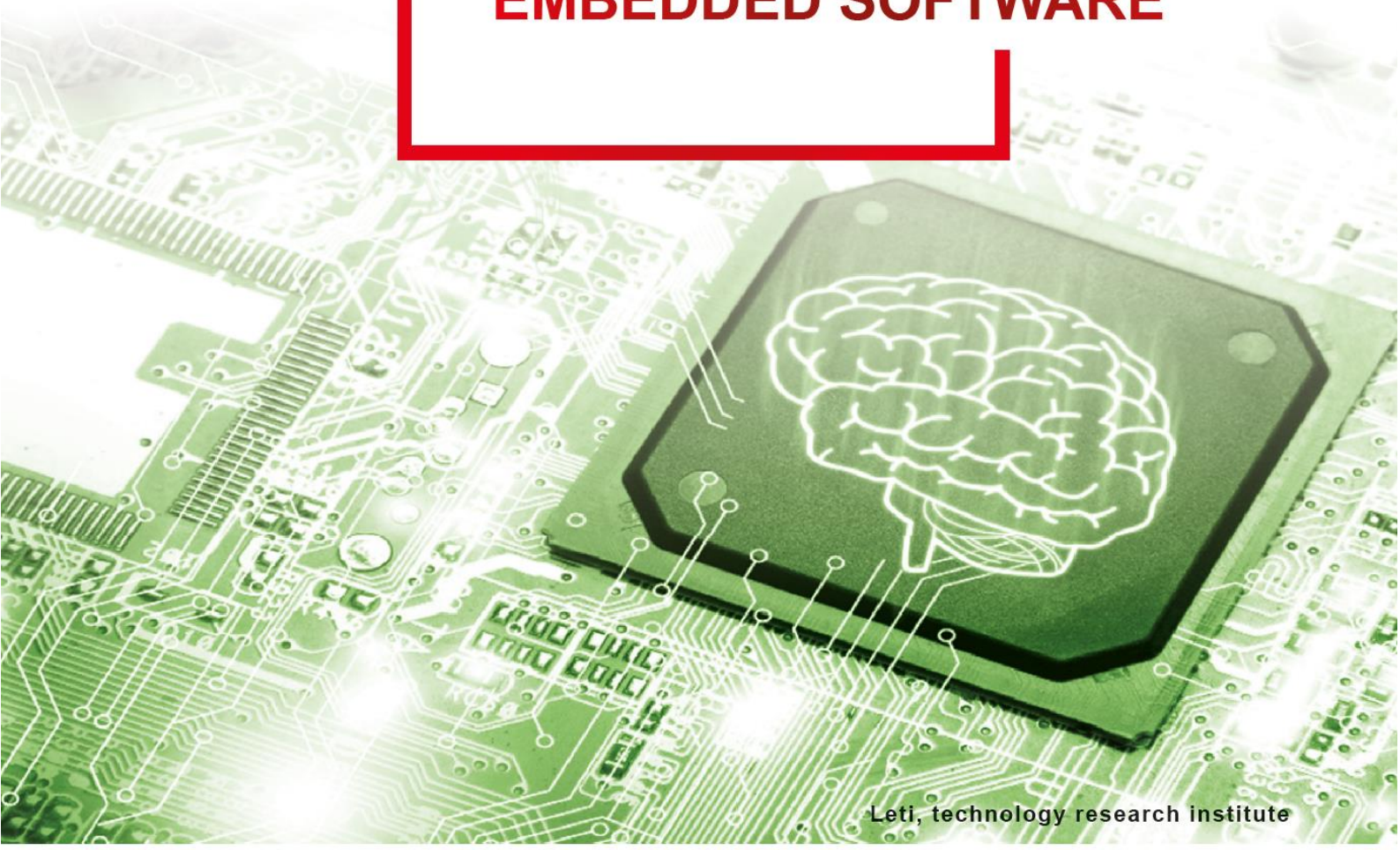
leti
cea tech

ANNUAL
RESEARCH REPORT
2015

# ARCHITECTURE AND IC DESIGN, EMBEDDED SOFTWARE



Leti, technology research institute

# TECHNOLOGY RESEARCH INSTITUTE

## LETI AT A GLANCE

Founded in **1967**

Based in **France** (Grenoble)
with offices in **USA** (Silicon Valley)
and **Japan** (Tokyo)

**350**
industrial partners

**1,900**
researchers

Committed to innovation, Leti's teams create **differentiating solutions in miniaturization and energy-efficient technologies** for its industrial partners.

Leti is a technology research institute at CEA Tech and a recognized global leader focused on miniaturization technologies enabling energy-efficient and secure IoT. Leti delivers solid expertise throughout the entire IoT chain, from sensors to data processing and computing solutions. Leti pioneered FDSOI low power platform for IoT, M&NEMS technology for low cost multisensors solutions, CoolCube™ integration for highly connected and cost effective devices.

Leti's mission is to pioneer new technologies, enabling innovative solutions to ensure Leti's industrial partners competitiveness while creating a better future. It tackles most current global issues such as the future of industry, clean and safe energies, health and wellness, sustainable transport, information and communication technologies, space exploration and safety & security.

For 50 years, the institute has built long-term relationships with its partners: global industrial companies, SMEs and startups. It tailors innovative and differentiating solutions that strengthen their competitiveness and contribute to creating new jobs. Leti and its partners work together through bilateral projects, joint laboratories and collaborative research programs. Leti actively contributes to the creation of startups through its startup program.

Leti has signed partnerships with major research technology organizations and academic institutions. It is a member of the Carnot Institutes network*.

**Carnot Institutes network: French network of 34 institutes serving innovation in industry.*

**2,670**
patents in portfolio

**60**
startups created

**€315**
million budget

**700**
publications each year

**ISO 9001**
certified since 2000

# ARCHITECTURE AND IC DESIGN, EMBEDDED SOFTWARE

CEA Tech institutes Leti and List share design, architecture & embedded software research activity in a dedicated division.

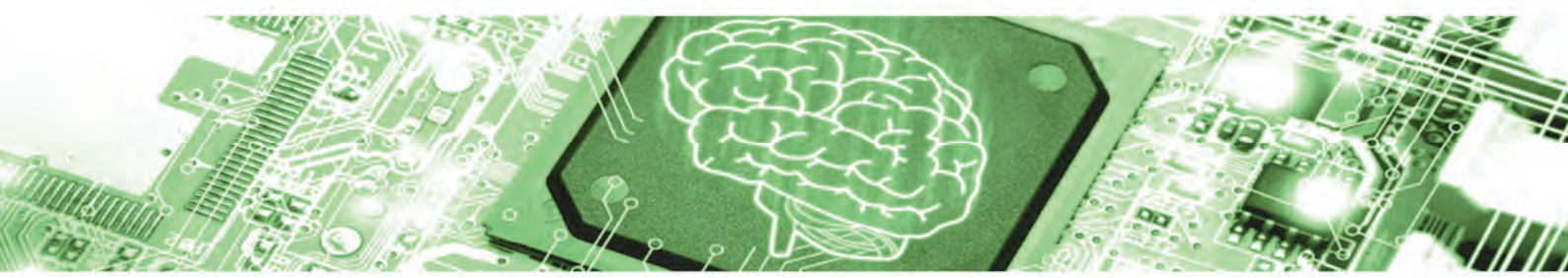
List is a key player in information and communication technologies. Its research activities are focused on digital systems that will have a major impact on society and the economy: embedded systems, ambient intelligence and information processing. List is based in CEA's Paris-Saclay campus.

In the dedicated division, more than 245 people focus on radio frequency, digital and SoC, imagers and sensors, integrated circuits, design environments and embedded software. Our research activities target the major challenges of tomorrow's systems. These include energy efficiency; complexity, especially in advanced technology nodes; reliability, including real-time constraints, security, and confidentiality; and the design of mixed-signal and heterogeneous systems (analog/digital, radio frequency, multi-physics, hardware and software).

We are preparing future systems in which computation, communication and real-world interactions will be tightly coupled. The Internet of Things and autonomous vehicles are primary examples of such applications.

Leti and List researchers collaborate on projects for both CEA Tech's internal needs and outside customers, ranging from startups and SMEs to large international companies.

# Contents



# Edito

**Thierry COLLETTE**

*Head of the Architecture & IC Design, Embedded Software Division*

The emergence of the Internet of Things is a real opportunity for designing new low-power smart devices with capabilities for communication and computing. These smart devices generate a huge quantity of data that must be stored and processed in real time inside data servers embedding high performance computing (HPC).

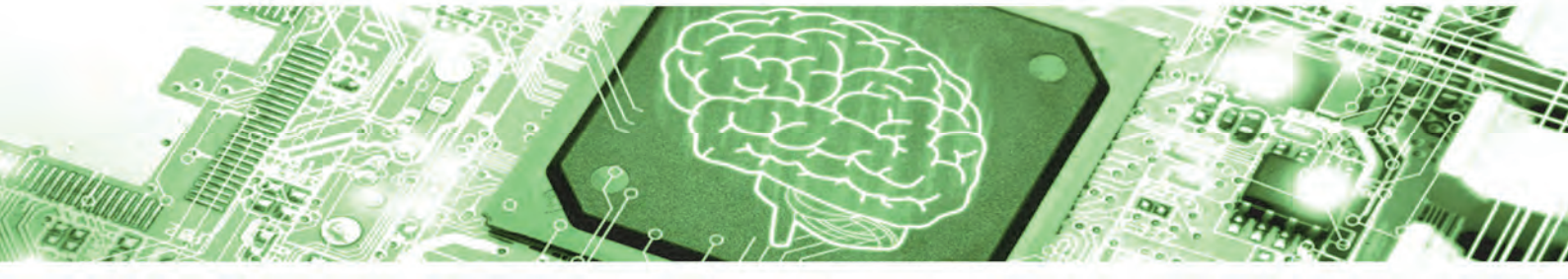
For these new HPC servers, we are working on innovations to increase their performance in real-time, fast-data computing, low energy consumption, and security and dependability. Research topics like secure hypervisor, neuromorphic computation, silicon photonics data communication and architectures using non-volatile resistive RAM or 2.5-D integration are at the center of our strategy.

You will find in this report presentations of our innovations in the domains of computing, wireless, smart sensors and imagers, reliable systems and design methodologies.

Many of these innovations were silicon- and/or system-proved in 2015 and are integrated in demonstrators. Come and visit us in our facilities on the Minatec or Paris-Saclay (France) campuses. You will have the opportunity to see that we are paving the way to future applications, such as artificial intelligence or autonomous systems.

We hope you enjoy reading this overview of our latest research.

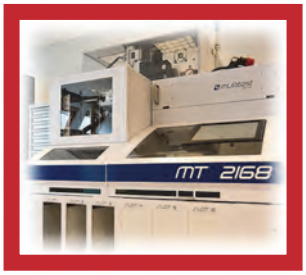
Thierry Collette



# Key figures

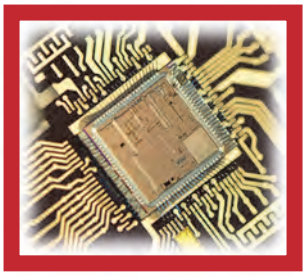
**2 locations:**
**Minatec** campus (Grenoble, France)
**Nano-INNOV** Paris-Saclay campus (Palaiseau, France)

**191** Permanent researchers
**51** PhDs and post-docs

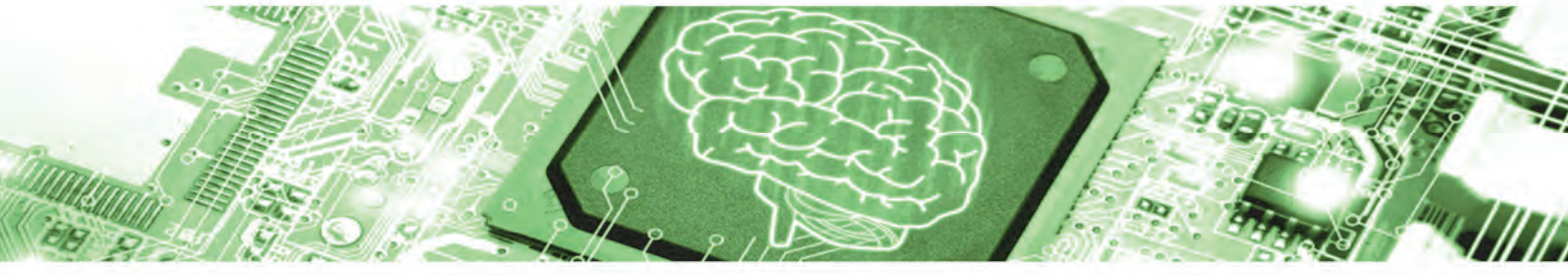



**Full suite** of IC CAD tools,
hardware emulators,
& industrial test equipment

**€41M** budget
**87%** funding from contracts




**48** granted patents
**37** papers, journals & books
**210** conferences & workshops



# Scientific Activity

## Publications

247 publications in 2015, including journals and major conferences like ISSCC, VLSI Circuits, ESSCIRC, IMS, ISCAS, DAC, DATE, SPIE, ISLPED, HIPEAC, IJCAI, CDC, ECC, and ESWeek.

## Prize and Awards

- Best Paper Award, IEEE S3S 2015 – N. Jovanovic et al.
- Best Paper Award, Async 2015 – E. Zianbetov et al.

## Experts

- 46 CEA experts: 5 research directors, 2 international experts.
- 11 researchers with habilitation qualification (to independently supervise doctoral candidates).

## Scientific Committees

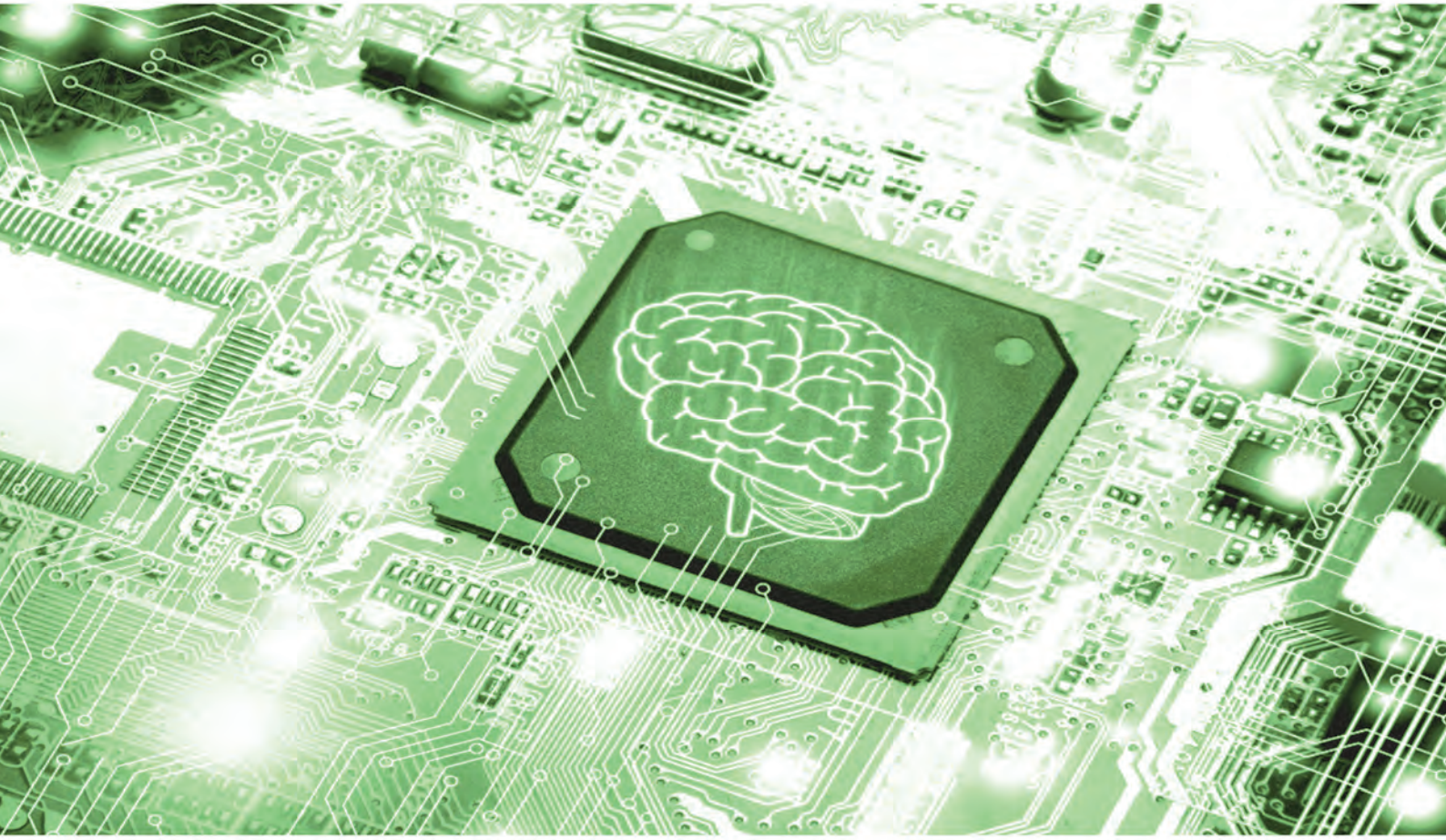
- Editorial boards: IEEE Journal of Solid State Circuits, Journal of Low Power Electronics, ISTE-Editions.
- 19 members of Technical Programs and Steering Committees in major conferences: ISSCC, VLSI-Circuits, ESSCIRC, DATE, ASP-DAC, ISPLED, ISVLSI, HiPEAC, ICCAD, IJCNN.

## Conferences and Workshops Organization

- 13th IEEE International New Circuits and Systems (NEWCAS) Conference, Grenoble (France) 2015.
- 8th Workshop on Design for 3D Silicon Integration (D43D), Grenoble (France) 2015.
- 9th International Workshop on Verification and Evaluation of Computer and Communication Systems (VECoS), Bucharest (Romania) 2015.

## International Collaborations

Collaborations with more than 20 universities and institutes worldwide, e.g. University of California, Berkeley (USA), Columbia University (USA), Cornell University (USA), Carnegie Mellon University (USA), EPFL (Switzerland), ETHZ (Switzerland), CSEM (Switzerland), UCL (Belgium), UNIBO (Italy), Polito Torino (Italy), KIT (Germany), Chalmers University (Sweden), Tongji (China), Keio University (Japan), NII (Japan)...



# 01

# METHODOLOGIES & HARDWARE - SOFTWARE INTEGRATION

- 3D technologies
- Thermal effects
- Performance prediction
- Communication modeling
- Code generation

# From 2D to Coolcube™ 3D: Cell on Cell Design using Commercial 2D tools

*Research topic: 3D Place&Route, Coolcube™, Sequential Integration*

**Authors: O. Billoint, S. Thuries, H. Sarhan, M. Brocard, G. Berhault**

**Abstract: Design of conventional 2D integrated circuits is becoming more and more challenging as we strive to keep on following Moore's law. Cost, thermal behavior, transistor characteristics, variability and back end properties are creating an increasingly complex equation to solve for designers. A possible solution could be to stay at the same node and use Coolcube™. Its main features are a 3D sequential process of MOS layers and tier to tier interconnect size allowing fine grain 3D partitioning of designs. Expected benefits are wire length reduction, power savings and increased operating frequency.**

## Context and Challenges

As 2D Place&Route commercial tools are handling one and only one cell layer, it is impossible to place cells from different tiers in parallel. However, in case we achieve to place cells in a 3D way, it is possible without increasing computing time (which could be due to placement errors detected by the tool) to route the cells using a user-defined back end. This concept will be the main driving argument of the methodology we are presenting below

## Main Results

The methodology (Fig. 1 (a)) allows an emulated-3D two tiers physical implementation of any design using 2D commercial tools. Place and Route is achieved through similar steps as required by 2D designs: pre clock tree synthesis (including placement), clock tree synthesis and routing; to which we added a folding step in order to emulate the 3D placement. Routing of both tiers in parallel using inter-tier metal layers is made possible by modifying input files of the tools. Benchmark results on two tiers 3D Monolithic integration have been done on several IPs (microcontroller, reconfigurable FFT and LDPC) using as reference ST 28nm FDSOI technology and show the correlation between cell density, routing congestion, wire length, operating frequency and power consumption.

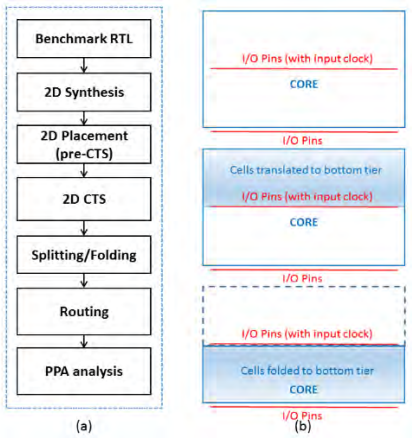



*Figure 1: (a) Emulated-3D flow / (b) Folding technique*

In order to emulate 3D placement, we are doing a 2D Place and Route of the 2D synthesized netlist by putting the top I/O ports (including control signals like clock) at the middle of the chip in the y direction as shown on figure 1 (b). Once 2D pre-Clock Tree Synthesis (pre-CTS) and Clock Tree Synthesis (CTS) are done, we are changing cells above the middle I/O ports from top tier to bottom tier. Bottom tier cells are then folded over top tier cells. By construction a 2D clock tree is transformed into a 3D clock tree without any errors. Once clock tree cells have been assigned to a tier, routing phase is then relatively straightforward as the tool has full knowledge of the back end and access to all layers for bottom to top tier Benchmark results are shown in table 1 below.

*Table 1: 3D Physical implementation PPA results*

| | Density | WL (mm) | | | Power (mW) | | | Operating Frequency |
|---|---|---|---|---|---|---|---|---|
| Open MSP | 76% | top | bottom | total | Clock | Data | total | |
| 2D | | 78,8 | 0 | 78,8 | 3,5 | 5,4 | 8,9 | 1,09 GHz |
| 3D 4BM | | 33,4 | 42,9 | 76,3 | 3,5 | 5,9 | 9,4 | 1,14 GHz |
| FFT | 77% | | | | | | | |
| 2D | | 265,7 | 0 | 265,7 | 15,1 | 43,4 | 58,5 | 1,75 GHz |
| 3D 4BM | | 129,2 | 124,4 | 253,6 | 14,4 | 47,2 | 61,6 | 1,83 GHz |
| LDPC | 46% | | | | | | | |
| 2D | | 1613,9 | 0 | 1613,9 | 23,2 | 67,3 | 90,5 | 0,83 GHz |
| 3D 4BM | | 1080,4 | 456,8 | 1537,2 | 21 | 63,3 | 84,3 | 0,81 GHz |

## Perspectives

We have demonstrated an emulated-3D physical implementation flow using a folding technique that goes through the same Place and Route steps as required by a 2D design with the addition of one extra step that is splitting/folding. Cells are first placed in a 2D design footprint before clock tree synthesis, then 50% area reduction is done using the splitting/folding technique, finally inter-tier routing is done in one run and connects correctly all pins from both tiers while meeting timing closure. This "one tool only" methodology is applicable to any design including those with memories and allows to easily estimate the impact of different options on 3D physical implementation. In order to fully evaluate the benefit of Monolithic 3D technology, optimization of cell placement between tiers is mandatory as it is the only possibility to reach the optimal cell to tier configuration.

*Related Publications:*
[1] F. Clermidy, O. Billoint, H. Sarhan, S. Thuries; "Technology scaling: The CoolCube™ paradigm", SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), 2015.
[2] P. Batude et al.; "3DVLSI with CoolCube process: An alternative path to scaling", VLSI Technology (VLSI Technology), 2015.
[3] O. Billoint et al.; "A comprehensive study of Monolithic 3D cell on cell design using commercial 2D tool", Design, Automation & Test in Europe Conference & Exhibition (DATE), 2015
[4] O. Billoint et al.; "From 2D to Monolithic 3D: Design Possibilities, Expectations and Challenges", International Symposium on Physical Design (ISPD), 2015

# 3D Thermal Effects:
# 3D Thermal Modeling, 3D Technology Analysis, Packaging using Heat Spreader

*Research topic: 3D CAD tools, TSV, Heat Spreader, System Level modeling and optimization*

**Authors: C. Santos, P. Vivet, J.P.Colonna, P.Coudrain (STMicro), C.Weiss (UKL), N.Peltier (DOCEA/INTEL)**

**Abstract: Heat dissipation is one of the main challenges in 3D technology. This is due to two main effects: increased power density in the same volume, and the use of extremely thin layers reducing the lateral thermal conductivity. In this work, we propose to address 3D thermal issues in various directions: thermal modeling methodology for fast, accurate and early thermal estimation of 3D thermal effects, systematic thermal exploration of the main effects (power density, die thinning, impact of TSVs), optimization by using advanced packaging materials such as heat spreaders, and lastly system level optimization like DRAM retention time dependency with 3D thermal effect.**

## Context and Challenges

Heat dissipation is frequently pointed as one of the main challenges in 3D integration technology. The heat generated in one die must travel through adjacent dies before reaching heat sinks. In TSV-based 3D ICs, aggressively thinned silicon dies present reduced lateral heat spreading capacity while poorly conductive adhesive materials used to bond dies together contribute to increase the thermal resistance of the vertical stack. The extremely thin layers reduce the lateral thermal conductivity which results in heavy tier-to-tier thermal coupling and thus similar temperature profiles for all tiers in the 3D stack. The 3D integration also brought new challenges to system-level thermal analysis. Thinned dies and fine-grain structures like µ-bumps and TSVs have considerable impact on the thermal performance. However, accounting for such small feature sizes increases the model complexity and the cost of simulations.

## Main Results

A new thermal modeling approach has been developed and combines material homogenization and model reduction techniques within the DOCEA Thermal Profiler™ tool. This approach handles all fine-grain structures required for 3D integration while producing a compact thermal model to be used for fast and accurate thermal simulations. Comparisons to temperature measurements on a Memory-on-Logic 3D circuit instrumented with heaters and thermal sensors show an average error lower than 4% for steady-state and very similar time responses [1].

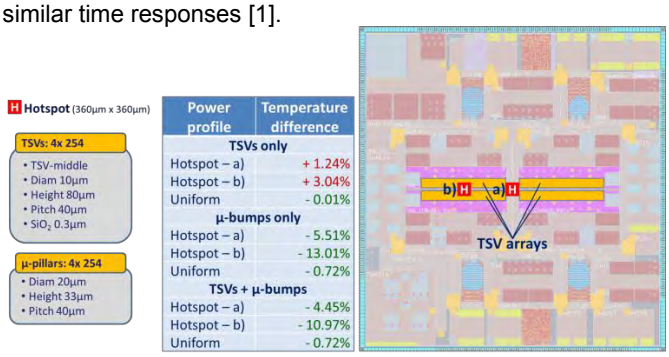



*Figure 1: Thermal impact of TSVs in a Memory-on-Logic circuit [2].*

A systematic thermal exploration [2] has been carried out to investigate the main aspects differentiating TSV-based 3D ICs: power density, die thinning, die-to-die thermal coupling and the impact of TSVs. The study reveals that non-thinned dies in a 3D stack may act as heat spreaders while TSVs may even provoke exacerbated hotpots due to the SiO2 used for TSV isolation. When TSVs are properly modeled with their SiO2 layer, TSVs induce actually a decreased horizontal heat transfer coefficient for a light increase of the vertical heat path. TSVs lead to larger thermal hotspots in case of lateral blockage, contrarily to the common believes. Measurements on a dedicated test-chip and systematic simulations (figure 1) have been carried out leading to the same conclusions [2].

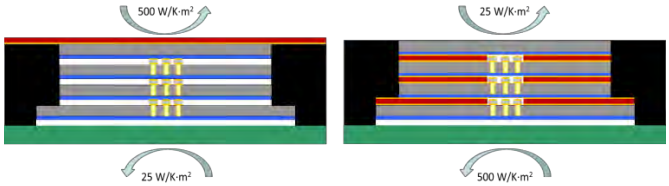



*Figure 2: 3D Stack Thermal optimization using Graphite Heat Spreader [3][4].*

Graphite-based materials present very high thermal conductivity and can be integrated into 3D stacks to compensate the poor heat spreading capacity of thinned silicon dies. Silicon measurements and simulations of a variety of circuit configurations (figure 2) indicate that the use of graphite-based heat spreaders is an effective approach to mitigate the strong hotspot dissipation in 3D ICs. Results show a reduction of up to 40% in the peak temperature for 3D ICs [3][4].

## Perspectives

Better knowledge of thermal impact in 3D circuit is mandatory to estimate and optimize all 3D design aspects: floorplan, packaging, but also dynamic system execution using advanced thermal mitigation control schemes. For instance, in case of the 3D stacking of a WideIO compatible DRAM memory, the bottom die may dissipate large power budget, inducing degration of DRAM retention time due to 3D thermal dissipation. A systematic study of DRAM thermal sensitivity has been proposed, including a system level model, for fault estimation and retention time system optimization [5].

*Related Publications:*
[1] C. Santos, P. Vivet, G. Matter, N. Peltier, S. Kaiser, R. Reis, "Thermal Modeling Methodology for Efficient System-Level Thermal Analysis", Custom Integrated Circuits Conference, 2014.
[2] P. Coudrain et al. "Experimental Insights into Thermal Dissipation in TSV-Based 3D Integrated Circuits", IEEE Design & Test, Issue 99, Dec 2015.
[3] C. Santos, R. Prieto, P. Vivet, J.P. Colonna, P. Coudrain, R. Reis, "Graphite-based Heat Spreaders for Hotspot Mitigation in 3D ICs", 3DIC'2015.
[4] R. Prieto, J.P. Colonna, P. Coudrain, C. Santos, P. Vivet, S. Cheramy, D. Campos, A. Farcy, Y. Avenas, "Thermal measurements on flip-chipped system-on-chip packages with heat spreader integration", SEMITHERM'2015, San Jose, USA, March 2015.
[5] C. Weis, M. Jung, C. Santos, P. Vivet, O. Naji, A. Hanssonz, N. When, "Thermal Aspects and High-Level Explorations of 3D stacked DRAMs", ISVLSI'2015, Montpellier, France, July 2015.

# A Simulation Framework for Rapid Prototyping and Evaluation of Thermal Mitigation Techniques in Many-Core Architectures

*Research topic: EDA, Virtual Prototyping, Power and Thermal simulation*

**Authors: T. Sassolas, C. Sandionigi, A. Guerre, J. Mottin, P. Vivet, H. Boussetta (Intel), N. Peltier (Intel)**

**Abstract: Modern SoCs are characterized by increasing power density and consequently increasing temperature that directly impacts performances, reliability and device packaging cost. Thermal issues need to be predicted and mitigated as early as possible in the design flow, when the optimization opportunities are the highest. We present an efficient framework for the design of dynamic thermal mitigation schemes based on a high-level SystemC virtual prototype tightly coupled with efficient power and thermal simulation tools. We demonstrate the benefit of our approach through silicon comparison with the SThorm 64-core architecture.**

## Context and Challenges

Thanks to technology scaling, SoC designers have been able to pack more and more transistors into the same chip for the great benefit of the end user. Unfortunately, this scaling in size is not equally matched by a scaling in power consumption, resulting in an increase in junction temperature. Such increase has negative impacts on the system reliability. It can also impact system performances as the power budget must be distributed among the processing resources.

Dynamically balancing the power consumption of processing resources is now a necessity. For many-core architectures, this is often managed by online scheduling techniques. To validate that a system can provide enough processing power while respecting its power and thermal budget, such scheduling techniques shall be developed as early as possible in the flow.

Efficient thermal simulation frameworks that take into account the complete behavior of the system are required to enable the development of dynamic thermal management. To be able to develop applications for architectures that are still under design, software designers usually rely on Virtual Prototypes (VPs). Accurate VPs are efficient to simulate an architecture's behavior for small periods of times in the time scale of an application's length. However, temperature phenomena timescale ranges from an order of magnitude of few milliseconds up to hours. Simulation time of accurate VPs would drastically soar if they were to be used for thermal evaluation. As a result, accurate VPs fail to encompass thermal issues.

## Main Results

To tackle this performance issue we propose a high-level simulation framework [1] whose speed allows to develop efficient thermal mitigation early in the flow. The framework is composed of a functional SystemC simulator [2]; a power and thermal simulation tool Aceplorer; and a thermal model generation tool ThermalProfiler. Based on an x86 implementation of the HARS runtime [3], the functional simulator trades accuracy for speed as it abstracts the architecture's components and relies on host processor's cycle counter to estimate execution duration and provide activity data to Aceplorer. Aceplorer simulation speed is also greatly accelerated by the generation of compact thermal models from physical description of the design. Automated simplification of the description (Fig.1) allows to keep transient thermal simulation accuracy with high simulation speed.

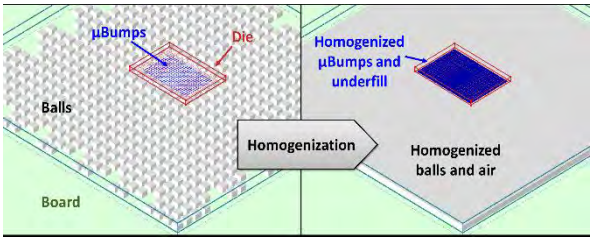



*Figure 1 : homogenization of component physical thermal description*

With this framework we modelled a 64-core SThorm architecture and designed efficient thermal mitigation for an image processing application case. Comparison with the silicon demonstrates the validity of the modelling approach. The thermal accuracy was further compared to show a maximal error of +-5°C at all timescales (Fig.2), which highlights the system-level accuracy of the thermal model.

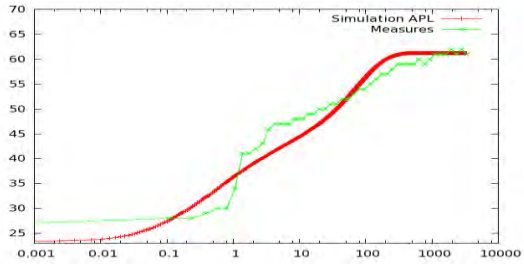



*Figure 2 : Thermal step response of the SThorm 64-core architecture*

## Perspectives

Future work will focus on the design of advanced thermal mitigation techniques for many-core systems using this framework. We also plan to extend the framework deeper in the flow by benefiting from emulation technique to refine the design behavior when RTL models are available.

*Related Publications:*

[1] T. Sassolas, C. Sandionigi, A. Guerre, J. Mottin, P. Vivet, H. Boussetta, N.Peltier, "A Simulation Framework for Rapid Prototyping and Evaluation of Thermal Mitigation Techniques in Many-Core Architectures", in 'International Symposium on Low Power Electronics and Design (ISLPED), 2015.

[2] T. Sassolas, C. Sandionigi, A. Guerre, A. Aminot, P. Vivet, H. Boussetta, L. Ferro, N. Peltier, "Early Design Stage Thermal Evaluation and Mitigation: the Locomotiv Architectural Case", Design Automation and Test in Europe (DATE), 2014

[3] Y. Lhuillier et al., "HARS: A hardware-assist,ed runtime software for embedded many-core architectures," ACM Transactions on Embedded Computing Systems

# Enabling Efficient Validation of Temperature-Dependent System Behavior Through Co-Emulation

*Research topic: EDA, Emulation, MPSoC, Power and Thermal simulation*

**Authors:** T. Sassolas, C. Andriamisaina, S. Bacles-Min, P. Vivet, S. Kaiser[1], N. Peltier[1], H. Boussetta[1] ([1]Intel)

**Abstract: Modern SoCs are characterized by increasing power density and consequently increasing temperature that directly impacts performances, reliability and device packaging cost. Thermal issues need to be predicted and mitigated as early as possible in the design flow, when the optimization opportunities are the highest. Emulation is the first step in the design flow that combines hardware accuracy, software execution and speed for accurate system bring-up. In this work, we propose to associate thermal estimation with emulation as a key addition to the IC design flow to reach thermal evaluation and mitigation objectives.**

## Context and Challenges

As temperature issues become a rising concern in modern integrated circuit, the definition of efficient thermal-aware design flow is compulsory. To cope with this situation, Intel has developed a power and thermal simulation environment, composed of two tools Aceplorer and ThermalProfiler, adapted in terms of simulation speed to the earliest design stages. However, accurate power and thermal simulation require detailed SoC activity profiling. Past works [2] have been conducted to efficiently couple Intel's tools with functional SystemC simulators at various levels of description but they either sacrificed speed or accuracy. As emulation is the first step in the design flow that combines hardware accuracy, software execution and speed, for accurate system bring-up, we propose to couple emulators with Intel's flow creating a co-emulation framework. The power and thermal evaluation of a 64-core MPSoC architecture emulated on a ZeBu server 2 was studied as a test case.

## Main Results

Our proposition (Fig.1) is to efficiently retrieve activity data from the emulated architecture (aka Design Under Test (DUT)) to supply the power model and thermal model simulated in Aceplorer. The compact thermal model is automatically generated by Thermal Profiler from a physical description of the design. The temperature of the design is also injected back in the DUT to allow thermal sensor emulation and therefore thermal mitigation design and mitigation.

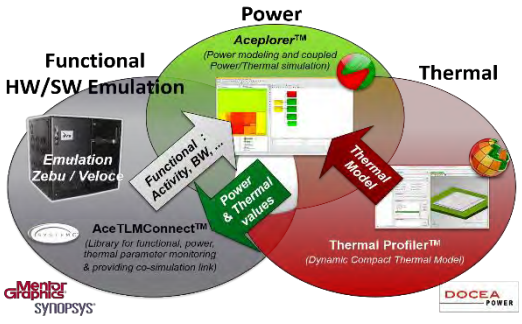



*Figure 1 : Co-emulation flow*

Though very similar to thermal co-simulation frameworks, the difficulty lies in setting up an efficient communication interface. Indeed, emulation can offer fast functional evaluation only if synchronization and data exchange are kept limited. In this work, we propose to rely on both a compaction of activity data and the usage of transactional co-emulation to reach high co-emulation speed. We specifically designed a processor activity monitor dedicated to counting execution of instructions in *instruction classes*. Such classes were defined according to the power consumption of each individual instruction.

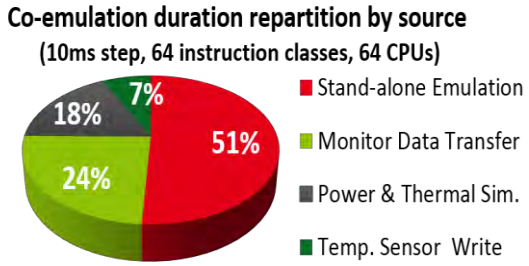



*Figure 2 : Thermal co-emulation performance*

This setup allowed to limit the monitor resource occupancy (<7% of the DUT) while reducing data exchange. Overall, the thermal emulation only doubles the standard emulation duration (Fig.2) making it a key addition to the classical IP design flow. Being able to emulate thermal sensors, we demonstrated the benefit of the solution by designing a reactive power and thermal management on this MPSoC architecture.

## Perspectives

Future work will focus on the generalization of the approach to other IP types (memories, caches, interconnects) as well as other emulation environment such as Mentor Graphics' Veloce or fast prototyping FPGA environments. This work could also be extended to study long term phenomena such as ageing during the emulation phase.

*Related Publications:*
[1] T. Sassolas, C. Andriamisaina, S. Bacles-Min, P. Vivet, S. Kaiser[1], N. Peltier[1], H. Boussetta[1] ([1]Intel), "Enabling Efficient Validation of Temperature-Dependent System Behavior Through Co-Emulation", Design Automation Conference (DAC) user track, 2015
[2] T. Sassolas, C. Sandionigi, A. Guerre, J. Mottin, P. Vivet, H. Boussetta[1], N.Peltier[1], "A Simulation Framework for Rapid Prototyping and Evaluation of Thermal Mitigation Techniques in Many-Core Architectures", in 'International Symposium on Low Power Electronics and Design (ISLPED), 2015.

# A performance prediction for automatic placement of heterogeneous workloads on many-cores

*Research topic: Compilation, heterogeneous systems, multi/many-cores*

**Authors: N. Benoit, S. Louise**

**Abstract: The multicore revolution is changing the compilation flow for the new many-core targets of the embedded world. Another new challenge is also coming as the future is foresighted to be more heterogeneous for the sake of power consumption while still allowing high levels of performance. This can only come in an economical way if and only if the compilation flow can automatically map computation kernels on the most relevant cores of a heterogeneous platform. In this set of papers, we introduce one element of such a compilation flow: a performance predictor that allows to choose the best configuration of heterogeneous mapping for a parallelized workload [1,2].**

## Context and Challenges

Adding multiple independent execution cores is one of the proposed solutions to the power wall. It is simple and promising but requires a paradigm shift in the field of programming. Instead of writing a single monolithic task, programmers are now required to distribute collaborating tasks to multiple cores.

Another trend also suggests that heterogeneous systems will conquer a large share of the architectural landscape. To take advantage of such designs, a compilation flow must be able to detect the affinity of code portions with the capabilities of the available cores. A successful parallelizing compiler on a heterogeneous target would require several steps: one to distinguish what parts of a given code should fit best on a given core, and another to find the best configuration of job distribution. The former can be reached by working on the Intermediate Representation (IR) and pattern matching as we did, e.g. in our Gomet extension of GCC [3] with the Kimble IR. The latter requires exploring the place and allocation space. To that end, we designed an oracle of the performance obtained for a given choice of allocation for kernels on the available cores.

## Main Results

The chosen approach was to define an efficient model of execution for heterogeneous multi or many-cores that takes into account the distribution of the work between generic cores and specialized accelerators, as seen in figure 1

.

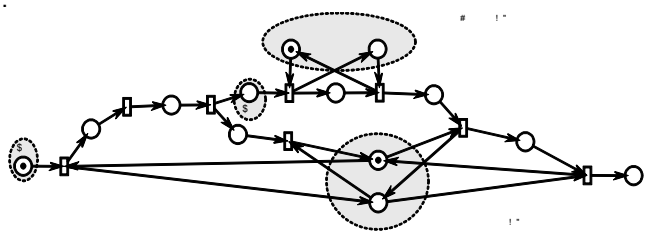


*Figure 1 : A Petri Net of the model of execution for SPMD heterogeneous workload, with General Purpose Cores (GPP) and specialized one (SP).*

This approach is easily generalized when several kind of accelerators or specialized cores can be used. The principle is to use the model of execution and an experimental value of the execution time of each computation kernel on the cores, to generate the prediction of performance, and then choose the best configuration both for the distribution of processors and for the number of jobs the program is sliced into

.

We tested this approach with several benchmark programs on both homogeneous and heterogeneous systems. The comparison between the predicted performance and the actual performance of the associated configuration can be shown, as can be seen on figure 2, where we evaluated the result on a PC using both a 4-core Xeon processor and one NVidia GPGPU. The comparison is done with regards to the best performance of 1 Xeon core and 1 GPGPU, for several job distributions

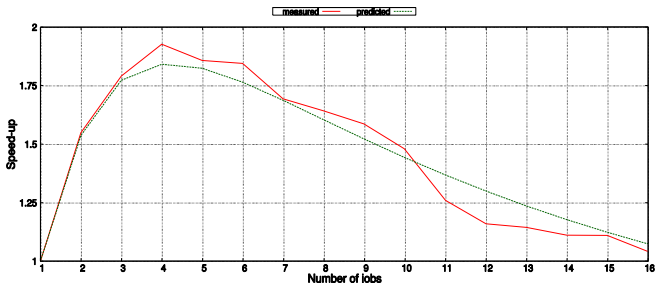



*Figure 2 : Measured and predicted speed-up on a GPGPU-enabled architecture with a varying number of jobs relative to an execution with one GPP (x86 core) and one SPP (one Nvidia GPGPU)*

Here the best configuration is achieved when the DCT offloading on the GPGPU is divided into 8 jobs

.

## Perspectives

Future work aim at achieving automatic compilation of programs and compare the result with hand optimized programs on the same heterogeneous target.

*Related Publications:*
[1] « Toward Performance prediction for Heterogeneous workloads on Manycores », N. Benoit, S. Louise. International Conference on Computational Science (ICCS - 2015).
[2] " A performance prediction for automatic placement of heterogeneous workloads on many-cores", N. Benoit, S. Louise. IEEE 9th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC-15)
[3] ""Kimble: a hierarchical intermediate representation for multi-grain parallelism", N. Benoit, S. Louise, Proceedings of the Workshop on Intermediate Representations, CGO 2011."

# Modeling of communications of a real-time system by SDFG

*Research topic: Synchronous data flow graph, real-time systems, latency*

**Authors: J. Khatib, E-C. Klikpo (IRT), A. Munier-Kordon (LIP6), K. Trabelsi-Colibet**

**Abstract: The dataflow models allow to synthetize the communications of an application composed of different communicating processes. They are used to deploy applications over physical architectures. In this paper, we propose a general intuitive mode of communications between real-time tasks with different periods. These tasks are executed based on the model of Liu and Layland. We show that the communications may be expressed directly as "Synchronous DataFlow Graph" (SDFG). In this case, we equally show that the latency between two communicating tasks can be simply bounded according to their respective periods.**

## Context and Challenges

Embedded Systems are complex systems which join different processing with different characteristics. For example, the advanced driving assistance systems (ADAS), which combine tasks of intensive computation (image processing) with real–time tasks (emergency brakes). One of the major difficulties encountered in the design of such systems is managing the communications between the different tasks of the application in order to ensure that the temporal constraints are satisfied.
Dataflow models are a formalized class which allow to represent in a simple and compact way the communications of regular applications (for example of video encoding type). A model of this class is usually represented in the form of a communicating tasks network whose execution flow is guided by the data's dependencies. On the other hand, synchronous languages are used to model real-time systems. The description of the exchanges is then directly extracted from the application.

## Main Results

Our first contribution in [1] is to define a communication scheme for multi-periodic tasks. For this purpose, we consider a set of tasks based on the model of Liu and Layland. In this model each task $t_i$ is characterized by an activation period $T_i$, an execution time $C_i$, a deadline $D_i$, and eventually a release date $r_i$.

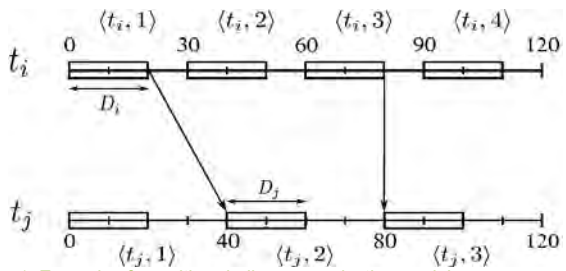



*Figure 1: Example of a multi-periodic communication model*

In order to describe the scheme of communication, we simplify the execution model of the tasks by equalizing the deadline of each task with its execution time. This type of communication happens between an emitting and a receiving task.

Figure 1 illustrates the communication scheme between the periodic tasks $t_i$ and $t_j$, such that $t_i$ is the emitting task and $t_j$ is the receiving one. We notice that the second execution of $t_j$ cannot begin before the first execution of $t_i$. Furthermore, we note the absence of this constraint between $< t_i, 1 >$ and $< t_j, 1 >$ on one hand and between $< t_i, 2 >$ and $< t_j, 2 >$ on the other hand. In both cases, the corresponding execution of $t_j$ begins before the end of the execution of $t_i$. The second contribution of our study is to prove that a set of communication constraints between two periodic tasks can be modeled by a buffer $a = (t_i, t_j)$ of a SDFG: Its production (*resp.* consumption) rate is equal to $T_i$ (*resp.* $T_j$), while the initial marking is computed by a closed formula (see figure 2).

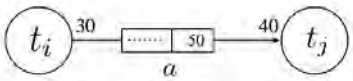



*Figure 2: The buffer $a = (t_i, t_j)$ associated to the precedence constraints induced by the communication from $t_i$ to $t_j$.*

In general, latency is the time gap between the moment when simulation appears and the associated reaction begins or ends. Evaluate the latency $\mathcal{L}$ between two tasks $t_i$ and $t_j$, is equivalent to compute the duration between the end of execution of $t_i$ and the start of execution of $t_j$. In this context, we show that the maximum latency between the different executions of these tasks is calculated according to each task parameters. Then, we bounded its value as follow:

$$T_i - max\{T_i - T_j, 0\} - T_a^* \leq \mathcal{L}_{\max}(t_i, t_j) < T_i - max\{T_i - T_j, 0\}$$

Where, $T_a^* = \gcd(T_i, T_j)$

## Perspectives

In this article we have demonstrated that the communication in a real time system can be modeled with an SDFG. Furthermore, we express the maximum latency between two communicating tasks and bound its value according to their period. One of the main perspectives of this work is to evaluate the upper bound of the latency between the entire system input and output in order to ensure a fixed response time.

*Related Publications:*
[1] J.Khatib, E-C.Klikpo, A.Munier-Kordon, K.Trabelsi-Colibet, "Modélisation des communications d'un système temps-réel par un SDFG".*Summer School for Real-time (ETR 15),* Rennes 2015.

# Data and Hardware Dependent Binary Code Generation

*Research topic: Compiler, Architecture, Data Dependent optimization, Code Generation, Dynamic Compilation*

**Authors: HP Charles, D. Couroussé, S. Lesecq, JF Méhaut (UGA), F Endo, L. Vincent & N. Halli (TIMA)**

**Abstract: Binary code generation is generally considered as « old technology » but new hardware developments and data dependent applications performances (multimedia, telecommunications, linear algebra, etc) provide new challenges and opportunities to compiler optimizations. This work adresses several aspects ranging from new hardware prefetcher to high level HotSpot Java JIT compiler. Our results show that major improvements can be reach and that classical static tools are no more "good enough" to reach the peak performances provided by new hardwares.**

## Context and Challenges

The "Speed race" for hardware architecture needed by new applications has hit many fundamental limits such as: sub-nanometric scaling, memory wall, parallelism handling, etc. To overcome these limitations hardware architects has developed new disruptive technologies: 3D stacking, highly parallel System on Chips, heterogeneous multiprocessors, specialized Arithmetic Units, etc. The side effects of these innovations break the classical programming model and make the hardware architecture more sensitive to data-sets in terms of performances and power.
In this new context, software researchers have to innovate a lot and revisit some old problems. Challenges are on dynamicity, hardware / software interactions, compilation even with strict memory or power constraints.

## Main Results

One of the first step was to show that it's possible to use dynamic code generation on embedded systems. Dynamic compilers (Java JIT or LLVM) are usually used on big HPC systems without any memory constraints. But small embedded systems can benefits of dynamic code adaptation as well.

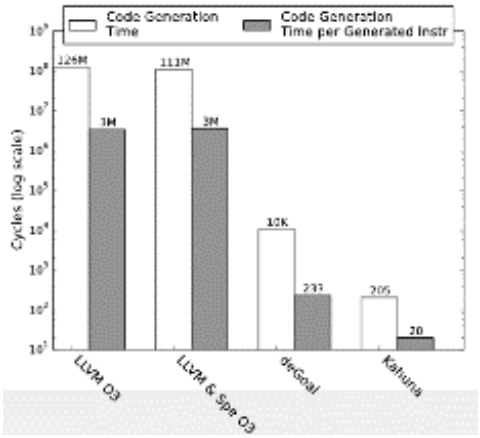



*Figure 1: Comparison of the code generation speed from 4 different tools : LLVM, LLVM with a manual code specialization, deGoal our "in house" tool for code generation and Kahuna (an experimental tool based on LLVM).*

[1] shows that, using specifics tools, dynamic compilation can be used on embedded systems with hard memory constraints and with a code generation fast enough to amortize the code generation for each function call.
The main results are shown on Figure 1. In [3] we showed that improvements can be achieved on power consumption and applied on more benchmarks applications.

At the opposite side, it's important to understand how high level software infrastructure consider the achievable performances. For example Java Hotspot compiler contain a very complex infrastructure. It was as surprise to see in [2] that a source transformation can have a huge impact on performance. This article show that there is many research to do in this domain.

At hardware level, we proved that using control/command theory, it is possible to implement data prefetching that improve performance from 25% to 99% on multimedia applications that are highly data dependent (image filtering and transformation, linear algebra).

*Table 1 Average cache miss reduction using the proposed prefetcher for various applications*

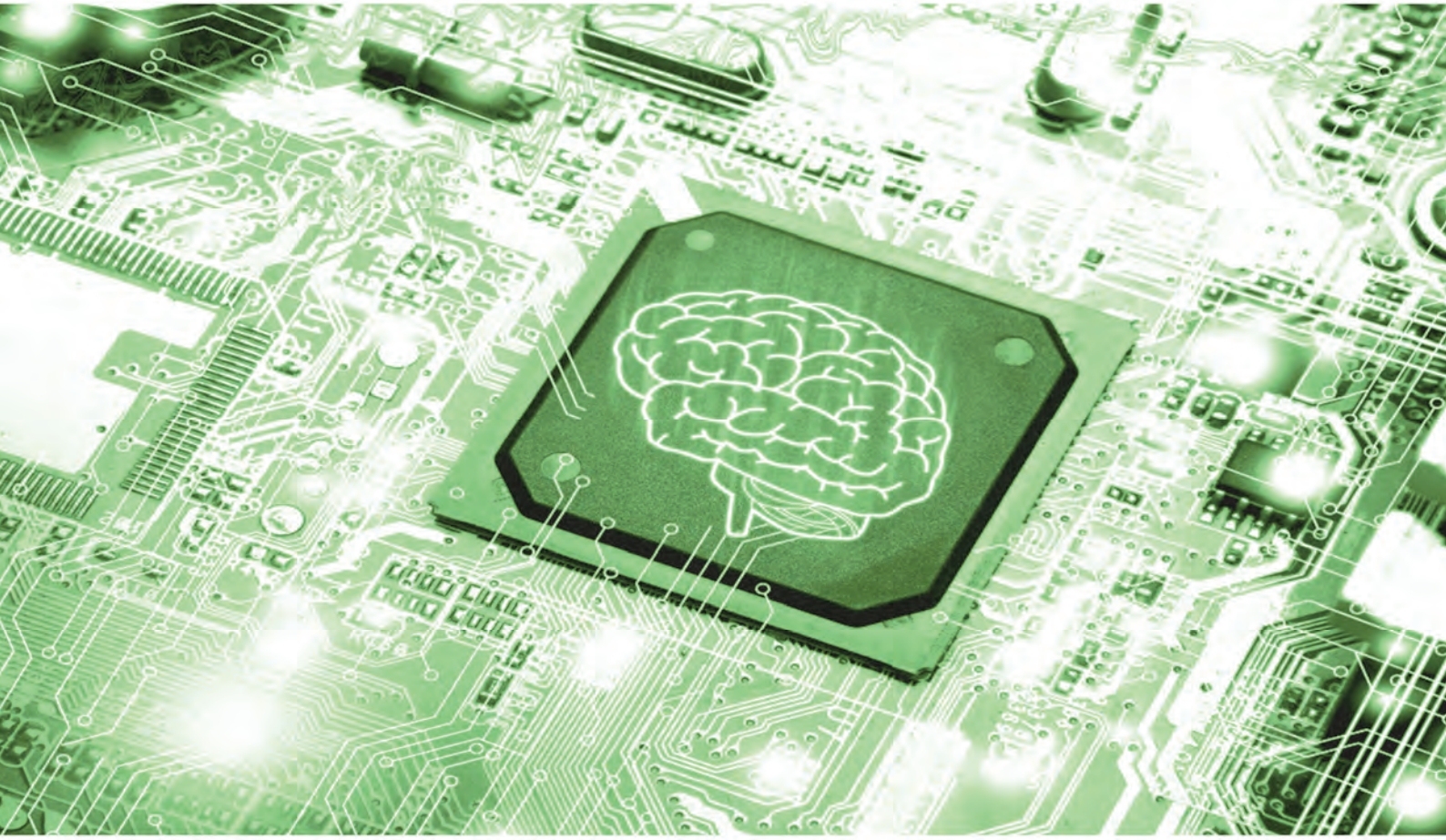
| Benchmark | No prefetch | SPT agg. 1 | SPT adapt |
|---|---|---|---|
| Rotate | 76.6% | 69.5% | 68.1% |
| Cjpeg | 98.5% | 92.6% | 92.6% |
| Matmul | 99.0% | 63.9% | 63.9% |
| Fisheye | 47.8% | 35.0% | 25.5% |

Having a vision on high and low level aspects from application to hardware behaviour help to understand how an application can benefit from dynamic optimization. The article [4] demonstrates that embedded software can beneficit from a dynamic approach both in term of performance and power consumption. The experimentations showed that even an embedded application can benefit of a run-time code generation optimization similar to a high level JIT compiler but adapted to the embedded domain.

## Perspectives

We have shown that very good results can be achieved on performances and power consumptions by a vertical approach for applications on embedded systems. This approaches needs trans-disciplinarily.

*Related Publications:*
[1] Charles, H. & Lomuller, V. (2015), 'Is dynamic compilation possible for embedded system?" Proceedings of the 18th International Workshop on Software and Compilers for Embedded Systems
[2] Nassim A. Halli, H.P. Charles, J.F. Mehau : « Performance comparison between Java and JNI for optimal implementation of computational micro-kernels » Proceedings of the 5th International Workshop on Adaptive Self-tuning Computing Systems 2015
[3] Charles, H.; Lomuller, V.; Endo, F. & Rekik, W. (2015), 'Low Overhead Runtime Code Specializations: A Case Study of the Impact on Speed, Energy and Memory"18th International Workshop on Compilers for Parallel Computing (CPC 2015)
[4] Endo, F.; Courousse, D. & Charles, H. (2015), 'Towards a dynamic code generator for run-time self-tuning kernels in embedded applications" Proceedings of The 4th International Workshop DCE-2015 (Dynamic Compilation Everywhere)'.



# 02

# COMPUTING SOLUTIONS : MULTI/MANY-CORES, ARCHITECTURES & SOFTWARE

- Benefits of advanced technologies
- Energy management
- Real time computing

# 3D Many-Core Architecture: Active Interposer system, 3D Network-on-Chip, and 3D asynchronous link

*Research topic: 3D technology, TSV, 3D Many Core Architecture, 3D Communication Links*

**Authors: P. Vivet, E. Guthmuller, I. Miro-Panadès, C. Bernard, Y. Thonnart, J. Pontes**

**Abstract: With the era of massive multi-core architecture targeting cloud computing for high end performances or advanced consumer electronics with tighter power consumption constraints, 3D integration technology will allow to design large scale many-cores. Thanks to 3D technology, it will be possible to maintain power consumption, increase chip-to-chip bandwidth, and preserve overall system cost by smart system partitioning. This work proposes a new partitioning for 3D Many-Core, using active interposer, and the design of advanced 3D Network-on-Chip (NoC) using robust asynchronous logic and new protocol converters to design efficient 3D communication plugs.**

## Context and Challenges

For designing large scale Many-Cores, one main limitation is system partitioning: how to integrate more cores into a single package and provide efficient core-to-core communication bandwidth. 3D technology is a next step, allowing integration of more cores, with reduced interconnect distance and thus reduced communication power consumption. Current 3D technology and architectures are using passive interposers, with power hungry synchronous SerDes communication links. In this work, we propose to use active interposer partitioning and energy efficient asynchronous communication links.

## Main Results

By splitting the 3D Many-Core system into multiple chiplets stacked onto an active interposer (fig. 1), one can achieve: 1) Yield optimization: chiplets are composed of clusters of cores, manufactured in advanced technology, tested and assembled on a large size interposer, made in a mature technology; 2) Power & Thermal dissipation mitigation: compared to homogeneous vertical 3D stacking, power density is reduced, power delivery networks are simplified, and thermal dissipation is comparable to standard 2D dies; 3) Smart Interposer features: by using active logic within the interposer, it is possible to provide advanced Network-on-Chip interconnects and embedded power management, (compared to passive interposers being limited to wire only).

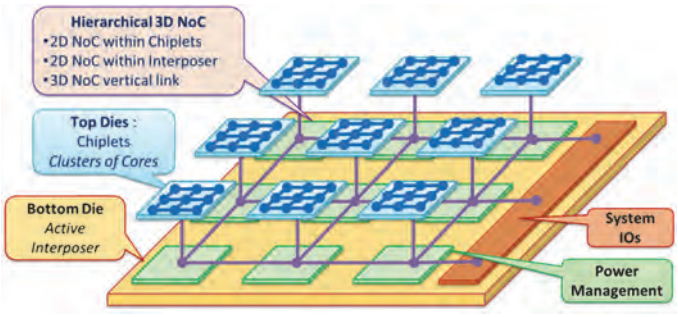



*Figure 1: 3D Many-Core composed of chiplets stacked onto an active Interposer, providing system IOs and power management, while chiplets are interconnected by 3D Network-On-Chip and efficient 3D vertical links [1,2].*

In such active interposer based many-core [1], the system level interconnect is hierarchical: 2D NoC within the chiplets in order to connect the clusters of cores, 2D NoC within the active interposer in order to connect the chiplets, and 3D vertical links between the chiplets and the interposer [2]. The 3D communication Plug is composed of: the NoC router, the NoC logical link, some additional DFT logic for testability, and the 3D physical interface itself, including TSVs and µ-bumps, and lastly µ-buffer cells. The 3D Plug is implemented with robust asynchronous logic to avoid timing hypothesis at the 3D interface.

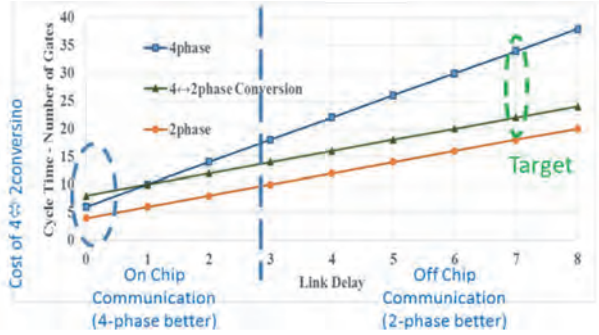



*Figure 2: Asynchronous 4-phase and 2-phase protocols trade-offs [3]*

In order to optimize the 3D NoC link, and get rid of the 3D link delay, we use 2-phase asynchronous communication (fig. 2). The principle consists in using the regular 4-phase asynchronous protocol for on-chip communication, and use 2-phase protocol for the 3D off-chip communication. Efficient 4-phase/2-phase protocol converters are proposed, using a new 1T-of-n transition encoding [3]. A throughput of 500 MFlit/s can be achieved on 32bits asynchronous parallel link using std-cell implementation.

## Perspectives

A large 3D ManyCore architecture composed of cache coherent tiles of cores [4] is currently partitioned in 3D using an active interposer, with embedded power management, and using optimized 3D NoC links, either asynchronous links or source-synchronous links. The 3D system integrates 6 chiplets in FDSOI 28nm on an active interposer in 65nm, offering a total of 96 cores.

*Related Publications:*
[1] P. Vivet, C. Bernard, F. Clermidy, D. Dutoit, E. Guthmuller, I. Miro-Panadès, G. Pillonnet, Y. Thonnart, A. Garnier, D. Lattard, A. Jouve, F. Bana, T. Mourier, S. Chéramy, "3D Advanced Integration Technology for Heterogeneous Systems", 3DIC'2015, Sendai, Japon, Sept 2015.
[2] P. Vivet, C. Bernard, E. Guthmuller, I. Miro-Panades, Y. Thonnart, F. Clermidy, "Interconnect Challenges for 3D Multi-cores: from 3D Network-on-Chip to Cache Interconnect", ISVLSI'2015, Montpellier, France, July 2015.
[3] Julian Hilgemberg Pontes, Pascal Vivet, Yvain Thonnart, "Two-phase Protocol Converters for 3D Asynchronous 1-of-n Data Links", ASP-DAC'15, Tokyo, Japan, January 2015.
[4] E. Guthmuller, I. Miro-Panades, and A. Greiner, "Architectural exploration of a fine-grained 3D cache for high performance in a manycore context,' in VLSI-SoC 2013, Istanbul, Turkey, Oct 2013.

# Distributed Synchronization of All-Digital PLLs Network for Clock Generation

*Research topic: PLL network, clock synchronization.*

Authors: C. Shan (LIP6), E. Zianbetov (LIP6), F. Anceau (LIP6), O. Billoint, D. Galayko (LIP6)

**Abstract: This work presents a Cartesian network of CMOS oscillators distributed on a chip and synchronized by a network of all-digital PLLs in phase and in frequency. The originality of the work is in the use of a solution essentially based on digital circuits that offers many opportunities for implementation of different synchronization algorithms. Our solution is based on a PI control applied to the phase error measured between neighbors, achieving global synchronization through a local control. This work presents a prototype demonstrating the feasibility and reliability of the proposed solution for synchronization.**

## Context and Challenges

This study addresses the problem of global clock generation inside complex and large SoC, in order to allow a fully-synchronous communication on the chip. Although asynchronous communication techniques in the SOCs have recently gained the ground, fully synchronous operation of the digital system is desirable in many cases, especially where the reliability of the system is the first priority. The main problem for implementation of fully synchronous SOCs is a generation of a global clock. Our study is motivated by deficiencies of conventional clock generation techniques.

## Main Results

The structure of the clocking network is presented in Fig.1. The local clocks are generated by digitally controlled oscillators (DCOs). Digital phase-frequency detectors (PFD) measure the timing error between each couple of neighboring DCOs. The network is coupled with the external reference clock through a PFD placed in upper left corner of the network. The digital error signals from PFDs are processed by the digital proportional-integral (PI) loop filters.

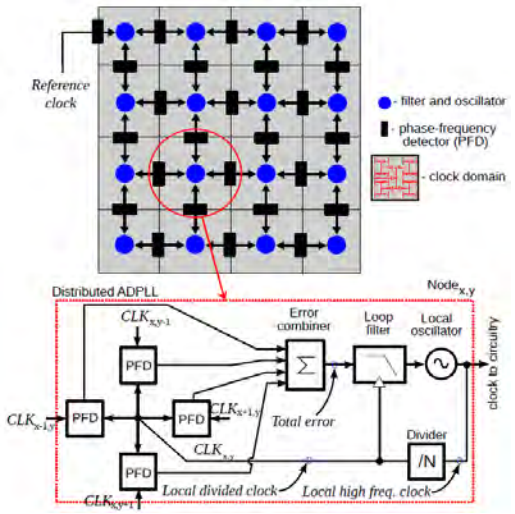



*Figure 1: Architecture of the ADPLL network and of a single node*

Fig.2. traces the maximum absolute value of the synchronization error in function of the distance of the node from the reference, for two network topologies. The first one is a unidirectional topology, where the information about the phase error is transmitted from the left upper corner toward lower right corner. The second one is a bidirectional topology, where each node is connected with its neighbors along all Cartesian directions. We can observe that in unidirectional network the phase error is accumulated as the reference phase information travels further, just like in a conventional clock tree. While in bidirectional network, which is the configuration at which the network works at steady state, phase errors between all the nodes in the network and the reference are well constrained within ±3 times PFD quantification steps. This shows the scalability of the synchronization solution we propose.

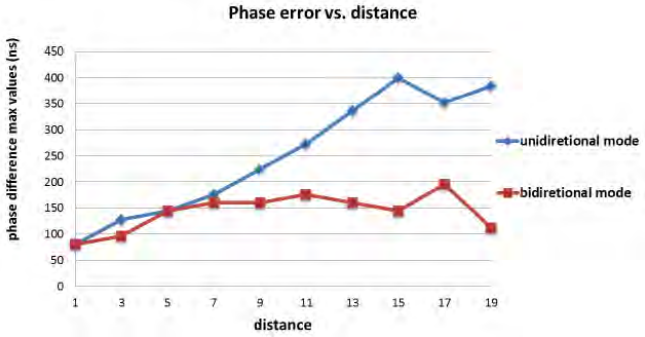



*Figure 2: Max. value of phase error in function of distance to the ref. clock obtained on the FPGA prototype of a 10x10 ADPLL network*

## Perspectives

Although the experimental study proved the operation and functionality of the solution, several theoretical issues remains uncleared. In particular, deeper study is required in order to link the residual synchronization error with the network parameters, and to study the topology of the basin of attraction related with the multiple synchronized modes. These questions are subjects of the ongoing work.

*Related Publications:*
[1] C. Shan, E. Zianbetov, F. Anceau, O. Billoint; "A distributed synchronization of all-digital PLLs network for clock generation in synchronous SOCs", New Circuits and Systems Conference, France, 2015
[2] A. Korniienko, G. Scorletti, E. Colinet, E. Blanco, J. Juillard, D. Galayko; "Control Law Synthesis for Distributed Multi-Agent Systems: Applications to Active Clock Distribution Network", American Control Conference, San Francisco, CA, 2011
[3] E. Zianbetov, D. Galayko, F. Anceau, M. Javidan, C. Shan, O. Billoint, A. Korniienko, E. Colinet, G. Scorletti, J-M. Akrea; "Distributed clock generator for synchronous SoC using ADPLL network", Custom Integrated Circuits Conference, San Jose, CA, 2013

# Body Bias usage in UTBB FDSOI designs: a parametric exploration approach

*Research topic: FDSOI, Dynamic Voltage Scaling, Dynamic Body Bias*

**Authors: D. Puschini, J. Rodas, E. Beigne, S. Lesecq**

**Abstract: This work presents a parametric exploration approach that analyzes the benefits of using Body Bias in 28nm UTBB FDSOI circuits. The exploration is based on electrical simulations of a ring-oscillator structure. These experiences show that a Body Bias strategy is not always required but, they underline the large power reduction that can be achieved when mandatory. Results are summarized in order to help designers to analyze how to choose the best dynamic power management strategy for a given set of operating conditions in terms of temperature, circuit activity and process choice. This exploration contributes to the identification of conditions that make DBB more efficient than DVS.**

## Context and Challenges

Fully-Depleted Silicon-on-Insulator (FDSOI) technology promises highly efficient designs with Ultra-Wide Voltage Range (UWVR) thanks to extended Body Bias properties. From power management perspective, this new opportunity is considered as a new degree of freedom in addition to the classical Dynamic Voltage Scaling (DVS), increasing the complexity of the power optimization problem at design time. So far, no formal or empiric tool allows to early evaluate the real need for a Dynamic Body Bias (DBB) mechanism on future designs.

## Main Results

A parametric exploration approach has been used to analyze the benefit of using Body Bias in 28nm FDSOI. These experiences showed that a Body Bias strategy is not always required. However, they underlined the large power reduction that can be achieved when it is mandatory.

During these experiments, it has been observed that depending on the power balance between dynamic and static consumptions, the management strategy to be adopted at run time strongly differs. Figure 1 shows examples with different power balance behaviors.

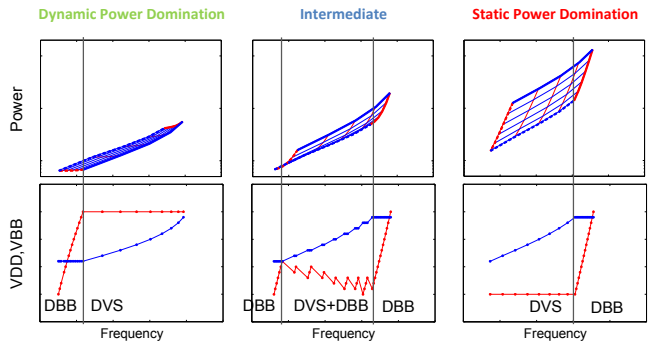



*Figure 1: Power balance between Pstat and Pdyn, and optimal Vdd, Vbb*

As can be seen, the circuit power consumption is either dominated by dynamic power (Pdyn domination), by static power (Pstat domination) or is in an intermediate state. By analyzing the power balance as a function of the operating conditions (switching activity range and temperature range) and design choices (well type, cells,…), it is possible to identify the power management strategy that will optimize the circuit operation (see Figure 2) in terms of power consumption.

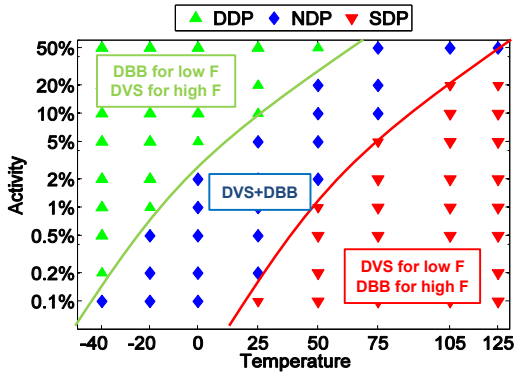



*Figure 2: Management strategy in function of power domination for various operating conditions*

This analysis improves design methodologies for circuits in FDSOI. To optimize power consumption taking advantage of Body Bias opportunities in FDSOI, this analysis helps the designer to determine the best strategy to be adopted for run time power management.

## Perspectives

The circuit studied, composed of a unique type of gates is not always representative of a complex circuit. Thus, a new sample circuit should be designed integrating different gate poly biasing (PBs) as in a whole complex circuit. The study presented here can then be conducted on this new sample. The resulting power analysis will therefore take into account a more realistic PB repartition. Once identified the dominant power, the corresponding management strategies as described in [1] will be applied.

*Related Publications:*

[1] D. Puschini, J. Rodas, E. Beigne, M. Altieri, S. Lesecq, "Body Bias usage in UTBB FDSOI designs: A parametric exploration approach", in Elsevier Solid-State Electronics - Journal, Volume 117, March 2016, Pages 138-145, ISSN 0038-1101, http://dx.doi.org/10.1016/j.sse.2015.11.019.

# Single-Well Design in FDSOI Technology

*Research topic: Energy-efficient digital circuit, high performance, low voltage, process compensation*

**Authors: A. Valentian, Y. Thonnart, B. Pelloux-Prayer (ST), P. Flatresse (ST)**

**Abstract: The Single Well option, offered by the FDSOI technology, is demonstrated to enable the design of energy-efficient digital circuits, operating on an ultra-wide voltage range. Silicon measurements of a DSP test chip fabricated in the 28nm node show performance ranging from 2.2GHz at 1.3V down to 65MHz at 440mV. Compared to a full LVT implementation, the leakage power is reduced by up to 6x, for a performance penalty of less than 10%. Additionally, balancing the n-MOS/p-MOS trees over this large voltage range only requires a single back-bias voltage, thus simplifying the power management scheme.**

## Context and Challenges

Driven by the current wave of Internet of Things devices, the need for high energy efficiency can be partially fulfilled by extreme Dynamic Voltage and Frequency Scaling (DVFS): digital circuits are made to operate on an Ultra Wide Voltage Range (UWVR), *i.e.* from the nominal supply voltage value, when high processing power is needed, down to the minimum operating voltage in low standby power mode.

The Ultra-Thin Body and Box (UTBB) FDSOI technology is very much tailored to implement such energy-efficient UWVR circuits. The low electrical parameters variability and reduced short-channel effects enable this technology to offer both a lower minimum operating voltage and better performance at low voltage than same node Bulk technology. However, UWVR circuits in FDSOI are usually made using Low Threshold Voltage (LVT) transistors for keeping good performance at low voltage, at the expense of leakage power. Additionally, mixing those LVT transistors with Regular Threshold Voltage (RVT) ones, at a fine-grain level, is not straightforward.

Using Single Well n-MOS and p-MOS transistor alleviates those issues, while providing an additional VT flavor with reduced leakage currents (Single Well means that both transistors are built on the same Well type).

## Main Results

A set of UWVR "Single N-Well" standard cells libraries has been built, starting from the same set of UWVR libraries available in the 28nm LVT option. Modifications have been performed on the cells schematics and layouts by scripting: modification of the p-MOS transistor type, the well geometries and the back-bias ties. For building a test chip, the Digital Signal Processor (DSP) of [1] has been considered. The proposed DSP is a 32 bits datapath Very Long Instruction Word (VLIW) structure, organized around a Multiplier-Accumulator.

The test chip has been fabricated in STMicroelectronics 28nm node FDSOI technology [2]. Measurements results are shown in Fig. 1: the Blue curve represents the default Single N-Well back bias value, which is equal to 0V, while the Red curve gives results obtained from slightly boosting the p-MOS transistors. As can be seen, there is no perceived impact at high voltage: the maximum clock frequency is equal to 2.2GHz in both cases, at VDD=1.3V. But as the power supply is reduced, boosting the p-MOS transistor helps reduce the imbalance that builds up between the N- and P-trees of logical gates. Thus, the performance is improved (by +55% at VDD=0.5V). Additionally, the minimum operating voltage (VDD-min) is reduced: it can go down to 412mV with a -0.5V bias voltage, starting from 471mV with the default bias. The fact that a single voltage is needed for optimizing UWVR operation (and for process compensation) eases the power management structure.

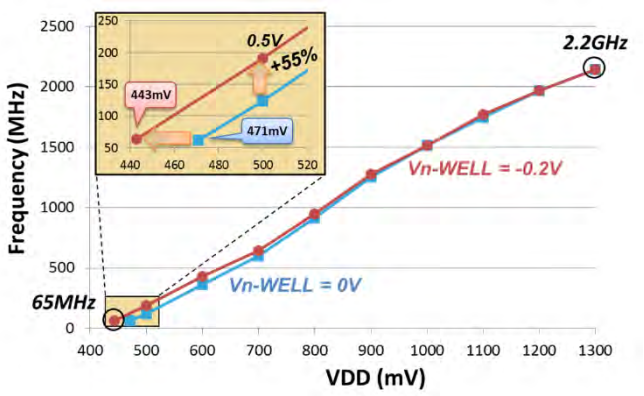



*Figure 1: Maximum operating frequency as function of the supply voltage value, for two different Single Well biases*

## Perspectives

The Single Well option, available in the 28nm FDSOI technology, enables the design of energy-efficient UWVR digital circuits, with performance close to a full LVT implementation and a 4-6x leakage power gain. Compared to this more classical implementation, the Single Well option offers the possibility to balance n-MOS/p-MOS trees with a single lever, i.e. a unique back bias tie.

This therefore opens up the perspective of designing low-power, high energy-efficiency, circuits for sensor nodes and biomedical applications.

*Related Publications:*

[1] E. Beigne, A. Valentian, I. Miro-Panades, R. Wilson, P. Flatresse, F. Abouzeid, T. Benoist, C. Bernard, S. Bernard, O. Billoint, S. Clerc, B. Giraud, A. Grover, J. Le Coz, J.P. Noel, O. Thomas, Y. Thonnart,, "A 460 MHz at 397 mV, 2.6 GHz at 1.3 V, 32 bits VLIW DSP Embedding FMAX Tracking," IEEE Journal of Solid-State Circuits, vol. 50, no. 1, pp. 125-136, January 2015

[2] A. Valentian, Y. Thonnart, B. Pelloux-Prayer and P. Flatresse, "Single-Well Design in FDSOI Technology – Towards energy-efficient ultra-wide voltage range digital circuits," IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), Rohnert Park, CA, USA, October 2015

# Energy estimation and management for multi-processor systems

*Research topic: Feedback control, system-on-a-chip, dynamic operating point scaling*

**Authors: A. Molnos, J. Mottin, D. Puschini, S. Lesecq**

**Abstract: Energy consumption remains a crucial issue in embedded multi-processor systems, despite many advances of the state-of-the-art. The challenge is to build low-overhead energy managers able to adapt to application's dynamics. We propose a proportional-integral controller that sets the operating points of processors in a system-on-chip. We address data-parallel applications with throughput constraints. Results on a test-chip indicate better performance when compared to state-of-the-art. Furthermore we introduce a power estimation framework that enables early power management design; comparisons against real hardware measurements suggest a high accuracy.**

## Context and Challenges

State-of-the-art embedded platforms comprise multiple processor cores on a chip and support operating point switching, e.g., dynamic frequency and/or voltage scaling (DVFS) at the level of processor cores, groups of cores or entire chips. Reducing energy consumption in such compute systems is still a challenge, despite many recent advances on the subject. Taking scaling decisions is difficult because variable workload demands have to be taken into account, such that application performance is not negatively impacted. Having a design-time power model to investigate these decisions is crucial.

## Main Results

The first contribution is a Proportional-Integral (PI) based controller [1] that addresses data-parallel applications with throughput constraints, as illustrated in Fig.1.

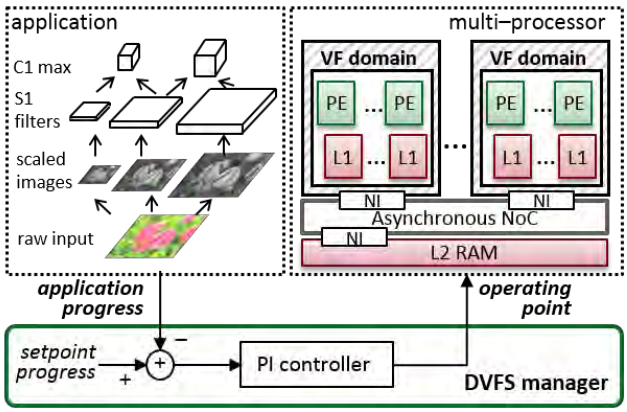



*Figure 1: Operating point setting for an application on a SoC platform*

These applications execute on an architecture comprising several voltage-frequency (VF) domains, which can be controlled independently. We perform an in-depth experimental analysis of the sensitivity of the PI controller to different configurations of an object recognition application. The outcome of this analysis are the best controller coefficients depending on the QoS level and buffer size of the application.

The controller is compared against an existing non-linear controller with two variants. Experiments on a test chip indicate that for our use-case, the PI controller attains on average 25% less number of operating point switches, leads to slightly better energy savings than the two non-linear ones. In addition, the PI controller meets the throughput constraint in cases where the other two approaches fail. All investigated controllers have low overhead.

The second contribution is a power estimation framework suitable for the evaluation and benchmarking of power management strategies [2]. It is based on off-line trace analysis and hardware state machine models. Our framework has been compared against real hardware measurements showing less than 8% of relative error in average.

*Figure 2: Power measurement on real hardware*

## Perspectives

The efficiency of a resource management policy severely depends on the ability of the manager to react to changing workload or platform variability. Feedback control methods require building an accurate model of the dynamics of applications, which may be a challenging job. To avoid this, we move towards reinforcement learning, a method able to find good adaptation policies, by trial-and-error and on-line, non-supervise training.
Future work will investigate the applicability of this learning method on standard operating systems, e.g, embedded Linux. Furthermore, other relevant criteria for state-of-the-art hardware platforms, i.e., temperature and reliability, will be addressed along with energy.

*Related Publications:*
[1] A. Molnos, W. Lombardi, D. Puschini, J. Mottin, S. Lesecq, A. Tonda: "Energy management via PI control for data parallel applications with throughput constraints", International Workshop on Power And Timing Modeling Optimization and Simulation (PATMOS) 2015.
[2] Y Ben Atitallah, J. Mottin, N. Hili, T. Ducroux, G. Godet-bar: "A Power Consumption Estimation Approach for Embedded Software Design using Trace Analysis", Euromicro Conference series on Software Engineering and Advanced Applications (SEAA 2015),
[3] M. Brieda, A. Molnos, J. Mottin, "Computation and data migration in an embedded many-core SoC", Workshop HiPPES4CogApp: High-Performance Predictable Embedded Systems for Cognitive Applications, co-located with 10th International Conference on High-Performance and Embedded Architectures and Compilers (HIPEAC) 2015.

# Task mapping and communication routing model for minimizing power consumption in multi-cores

*Research topic: task mapping, communication routing, power minimization*

**Authors: S. Carpov**

**Abstract: In this paper we introduce a novel MILP formulation for the problem of mapping tasks and routing communications on multi-core systems with power minimization objective. The cores have several power consumption modes. Dynamic and static power consumptions are modeled independently and the dynamic power consumption depends on core load rate. Three types of communication routing are examined: single-path, multi-path and fractional multi-path. Initially a mathematical model is introduced and afterwards a linearized mixed-integer program formulation is proposed. We conclude the paper by presenting computational results on task graph instances obtained from StreamIt applications.**

## Context and Challenges

In this work we propose a novel MILP (Mixed-Integer Linear Program) model which maps application tasks and routes inter-task communications, all at once, onto multi-core architectures. The objective is to minimize the total power consumption. We consider multi-core architectures where several computation cores are connected via a communication network (e.g. Network-on-Chip or NoC), as for example the MPPA clustered many-core processor. The applications are modeled as task graphs. Task graph nodes are mapped to architecture cores and communications between tasks are routed through the network.

Related to this work is [1] where the authors present several task mapping and communication routing methods for dataflow applications on clustered many-cores. A GRASP heuristic for the joint problem of placement and routing of dataflow applications is given in [2]. Contrary to our work the authors optimize the communication network bandwidth.

## Main Results

A mathematic program for mapping application tasks and routing communications to a parallel architecture is proposed. Linearized constraints are proposed for the non-linear constraints of the mathematical model, a MILP is obtained. In what follows we describe the used architecture, application model and minimization objectives.

A multi-core architecture is composed of a set of computational cores and a set of communication l inks between pairs of cores. The cores can operate in several operating modes. The operating modes can be due to either frequency/voltage scaling or a combination of both. Dynamic power consumption is supposed to be linearly proportional to the computational load of a core. Communication links between cores are directed. The cores are not fully connected to each other. Each communication link has a maximal data throughput.

An application to be mapped is composed of a set of tasks and a set of communications between pairs of tasks. The applications are executed continuously (i.e. data-flow like applications). The communications between tasks can be performed in three ways (routing strategies): (1) single-path routing – whole communication takes a single route, (2) fractional multi-path routing – fractional parts of communication take several routes, (3) multi-path routing – integer parts of communication take multiple routes.

Performance of the proposed MILP model was tested on a benchmark of StreamIt applications. An instance is composed of two parts: an application to map and route and an architecture. Task graphs are obtained from StreamIt applications. We choose to adapt (dimension) several multi-core architectures to applications in order to test MILP performance in different conditions. The linearized constraints increase the performance of MILP formulation in terms of relaxation compute time and MILP relaxed value. In Fig. 1 are presented the ratio between relaxation execution times of the MILP and the mathematical model formulations.

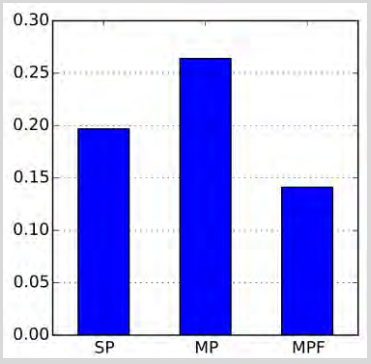



*Figure 1: Ratio between linearized and non-linearized averaged relaxation times. Routing types are depicted on the horizontal axis.*

## Perspectives

Perspectives for this work include the proposal of heuristic methods for solving the task mapping and communication routing problem in order to support larger instances. Another track will be to study other minimization objectives.

*Related Publications:*
[1] Galea, F. and Sirdey, R., "A Parallel Simulated Annealing Approach for the Mapping of Large Process Networks", in Parallel and Distributed Processing Symposium Workshops PhD Forum (IPDPSW), 2012 IEEE 26th International (, 2012), pp. 1787--1792.
[2] Stan, O. and Sirdey, R. and Carlier, J. and Nace, D., "A GRASP for Placement and Routing of Dataflow Process Networks on Many-Core Architectures", in 3PGCIC (, 2013), pp. 219--226.

# Self-Timed Periodic Scheduling of Data-Dependent Tasks in Embedded Streaming Applications

*Research topic: Self-Timed, Periodic Scheduling, Embedded applications*

**Authors: Xuan Khanh Do, Amira Dkhil, Stéphane Louise**

**Abstract: In this report, we consider the problem of multiprocessor scheduling for safety-critical streaming applications modeled as acyclic data-flow graphs. We introduce a new scheduling policy noted Self-Timed Periodic (STP), which is an execution model combining self-timed scheduling with periodic scheduling. The proposed framework shows that STS improves the performance metrics of the programs, while the periodic model captures the timing aspects. We evaluate the performance of our scheduling policy for a set of 10 real-life streaming applications and in most of the cases, our approach gives a significant improvement in latency.**

## Context and Challenges

There is an increasing interest in developing applications on multiprocessor platforms due to their broad availability and the looming horizon of many-core chip, such as the MPPA-256 chip from Kalray (256 cores) or the Epiphany from Adapteva (64 cores). Given the scale of these new massively parallel systems, programming languages based on the data-flow model of computation have strong assets in the race for productivity and scalability. Nonetheless, as streaming applications must ensure data-dependency constraints, scheduling has serious impact on performance. Hence, multiprocessor scheduling for data-flow languages has been an active area and therefore many scheduling and resource management solutions was suggested.

## Main Results

We introduce four classes of STP schedules based on two different granularities and two types of deadline: implicit and constrained for applications modeled as Cyclo-Static Dataflow (CSDF) graphs. Two first schedules, denoted $\mathrm{STP}^{\mathrm{I}}_{\mathrm{q_i}}$ and $\mathrm{STP}^{\mathrm{C}}_{\mathrm{q_i}}$, are based on the repetition vector $\mathrm{q_i}$ without including the sub-tasks of actors. Two remaining schedules, denoted $\mathrm{STP}^{\mathrm{I}}_{\mathrm{r_i}}$ and $\mathrm{STP}^{\mathrm{C}}_{\mathrm{r_i}}$, have a finer granularity by including the sub-tasks of actors. It is based on the repetition vector $\mathrm{r_i}$.

The effect of Self-timed Periodic (STP) scheduling can be modeled by replacing the period of the actor in each level by its worst-case execution time under periodic scheduling. The worst-case execution time is the total time of computation and communication parts of each actor. For a graph G, a period Φ, which represents the period, measured in time-units, of the levels in G is given by the solution to:

$$\Phi \geq \max_{\mathrm{j}=1\rightarrow\alpha} (\widehat{\mathrm{W}}_{\mathrm{j}} + \widehat{\varphi}_{\mathrm{j}})$$

where $\alpha$ is the number of levels, $\widehat{\mathrm{W}}_{\mathrm{j}}$ is the maximum workload and $\widehat{\varphi}_{\mathrm{j}}$ is the worst-case communication time of all levels in the Timed Graph G.

Let $\mathrm{a_1}$ denote the level-1 actor. $\mathrm{a_1}$ will complete one iteration when it fires $\mathrm{q_1}$ times. Assume that $\mathrm{a_1}$ starts executing at time t = 0.Then, by time $t = \Phi \geq \mathrm{q_1}\omega_1$, $\mathrm{a_1}$ is guaranteed to finish one iteration in a self-timed mode (start the next sub-task immediately after the end of the precedent). $\mathrm{a_1}$ will also generate enough data such that every actor $\mathrm{a_k} \in \mathrm{V_2}$ can execute $\mathrm{a_k}$ times (i.e. one iteration). By repeating this over all the $\alpha$ levels, a schedule $S_\alpha$ (shown in Figure 1) is constructed:

| time | $[0,\phi)$ | $[\phi,2\phi)$ | $[2\phi,3\phi)$ | ... | $[(\alpha-1)\phi,\alpha\phi)$ |
|---|---|---|---|---|---|
| level | $V_1(1)$ | $V_2(1)$ | $V_3(1)$ | ... | $V_\alpha(1)$ |
| | | $V_1(2)$ | $V_2(2)$ | ... | $V_{\alpha-1}(2)$ |
| | | | $V_1(3)$ | ... | $V_{\alpha-2}(3)$ |
| | | | | ... | $V_{\alpha-3}(4)$ |
| | | | | ... | ... |
| | | | | | $V_1(\alpha)$ |

*Figure 1: Schedule $S_\alpha$*

We evaluate the proposed STP representation using a set of 10 real-life applications. Figure 2 the ratios of latency obtained under $SPS$, $\mathrm{STP}^{\mathrm{I}}_{\mathrm{q_i}}$, $\mathrm{STP}^{\mathrm{I}}_{\mathrm{r_i}}$, $\mathrm{STP}^{\mathrm{C}}_{\mathrm{q_i}}$, $\mathrm{STP}^{\mathrm{C}}_{\mathrm{r_i}}$ schedules to the $STS$ latency. In most of the cases, our approach gives a significant improvement in latency compared to the STS.

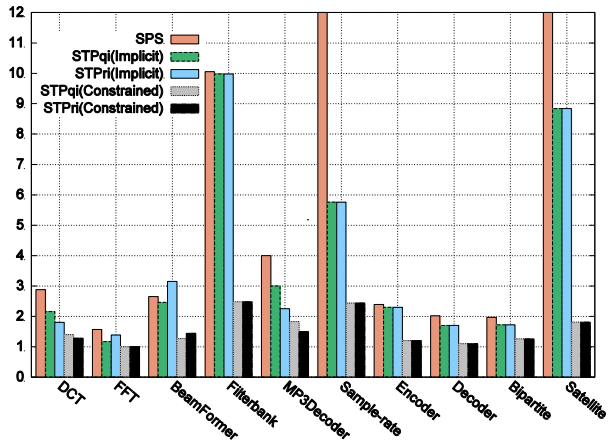



*Figure 2: Ratios of the latency under $SPS$, $STP^I_{q_i}$, $STP^I_{r_i}$, $STP^C_{q_i}$, $STP^C_{r_i}$ schedules to the $STS$ latency. It must be noted that the Sample-Rate and Satellite programs have a ratio for $SPS$ much larger than 12, but the graph is zoomed to display accurately the results for most of the programs.*

## Perspectives

As a future work, we want to improve our scheduling policy for STP model using the constrained deadline which requires different schedulability analysis.

*Related Publications:*
[1] X. Do, A. Dkhil, S. Louise, "Self-Timed Periodic Scheduling of Data-Dependent Tasks in Embedded Streaming Applications" in 15th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP), 2015.
[2] A. Dkhil, X. Do, S. Louise, and C. Rochange, "A Hybrid Algorithm based on Self-Timed and Periodic Scheduling for Embedded Streaming Applications" in International Conference on Parallel, Distributed, and Network-Based Processing (PDP), 2015.

# An Empirical Evaluation of a Programming Model for Context-Dependent Real-time Streaming Applications

*Research topic: Parallelism, Real-time, Model of Computation, Many-core*

**Authors: XK. Do, S. Louise, T. Goubier, P. Dubrulle, P. Dore, A. Cohen (INRIA)**

**Abstract: We present and evaluate a Programming Model for real-time streaming applications on high performance embedded multi- and many-core systems, which encompasses both real-time requirements, determinism of execution and context dependency. It is an extension of the well-known Cyclo-Static Dataflow (CSDF), for its desirable properties, with two new important data-flow filters: Select-duplicate, and Transaction which retain the main properties of CSDF graphs and also provide useful features to implement real-time computational embedded applications. We evaluate the performance of our programming model thanks to several real-life case-studies.**

## Context and Challenges

Programmers currently lack a clear path to navigate the landscape of many-core programming models, and programming tools are unfortunately lagging behind the fast paced architectural evolutions. Dataflow programming address some of these issues, meeting the requirements in terms of scalable task parallelism, functional determinism, and temporal and spatial data reuse in these systems. Unfortunately, these streaming languages are often too static to meet the needs of emerging embedded applications, such as context- and data-dependent dynamic adaptation. For this reason, we must find a way to combine the parallelism management of dataflow principles and the reliability of real-time systems, and a common base to define such a Programming Model. In this paper, we are looking at the evaluation of a previously defined Model of Computation (MoC) which extends the CSDF-based usual MoC with 2 important data-filters: Select-duplicate and Transaction.

## Main Results

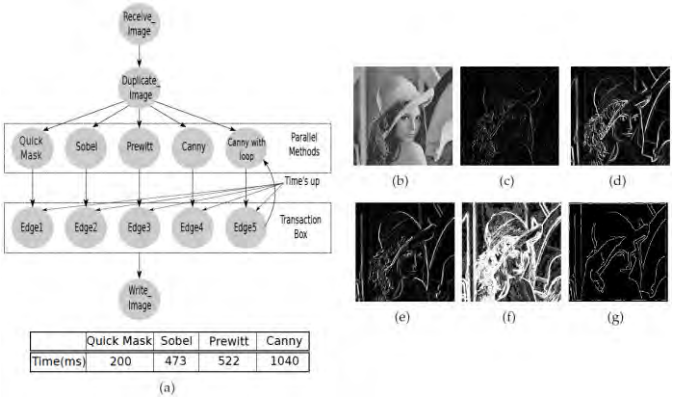
Modern data-flow languages, based on CSDF (e.g. ΣC and StreamIt) support special filters for data management only (system agents in ΣC). Such filters do not modify the values they receive, but only route them or organize them. This particular feature is very useful for optimization of process networks. Usual data redistribution filters include splitters (or split filters), joiners (or join filters) and duplicate filters, which are usually combined as *split/join* constructs or *duplicate/join* constructs. Nonetheless, these data-filters cannot resolve the case when no guarantee is provided on the availability of a given dataset as input, or no guarantee can be given on the availability of a computing resource for the process to be fired.

To deal with these new requirements, extensions of CSDF graphs by combining the 2 new data-filters: *Select-duplicate* and *Transaction* which do not affect the desirable properties of this model are introduced in [2]:

**Select-duplicate filters:** as an ordinary duplicate filter, the select-duplicate filter has one entry and *n* outputs (*n* is a maximum number which can be specified to enable automatic sizing). At a given time any combination of the *n* outputs can be enabled. Nonetheless, if there is a timing constraint on this filter (either directly, or indirectly), then at least one output must be enabled at any given time, and the enabled path will bear the timing constraints.

**Transaction filters:** the transaction filter implements an important action not available in usual data-flow MoC, which can be useful in a wide set of use-cases: Speculation, Redundancy with vote, Best-of at a given deadline, Select an active data-path among several, and so on. Let us suppose we have several algorithms and several characteristics for data-paths. Only one data-path is required to finish before the deadline. The others can be much loosely defined: algorithms can be highly refining, with uncertain execution times, and execution delays. When the deadline comes, the transaction box choose the best result available at that time.

We study applications from image processing, as can be seen in Figure 1, as they convey both computation and time-constraint requirements.




| | Quick Mask | Sobel | Prewitt | Canny |
|---|---|---|---|---|
| Time(ms) | 200 | 473 | 522 | 1040 |

*Figure 1: (a) CSDF graph of the Edge Detection application. Execution time of different Edge Detection methods (measured on Intel core i3@2.53GHz). Canny with loop is an incremental evaluation. At the end of the deadline, the best result will be chosen, according to the order: (c) Quick Mask < (d) Sobel < (e) Prewitt < (f) Kirsch < (g) Canny < Canny with loop.*

## Perspectives

As a future work, we intend to resolve the problem of scheduling and translating this model more into a formal language to have a more coherent view of many-cores.

*Related Publications:*

[1] X. Do, S. Louise, T. Goubier, P. Dubrulle, P. Dore, A. Cohen "An Empirical Evaluation of a Programming Model for Context-Dependent Real-time Streaming Applications" in International Conference on Computational Science (ICCS), 2015.

[2] S. Louise, P. Dubrulle, T. Goubier "A Model of Computation for real-time applications on embedded Manycores" in International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC), 2014.

# Wormhole networks properties and their use for optimizing worst case delay analysis of many-cores

*Research topic: Worst-case delay, Network-on-Chip, Wormhole switching*

**Authors: L. Abdallah, M. Jan, J. Ermont (IRIT), C. Fraboul (IRIT)**

**Abstract: Wormhole switching is widely used within Network-on-Chip (NoC), due to its improved throughput. Several NoCs targeting hard real-time systems have been designed and the Worst-Case Traversal Time (WCTT) of flows analyzed. However, none is currently available in commercially existing NoCs that instead rely on wormhole switching and Round-Robin arbitration. In this paper, we demonstrate three properties of such NoC-based wormhole networks to identify worst-case scenarios and reduce the pessimism when modeling flows in contentions. We then describe and evaluate an algorithm to compute WCTT of flows that uses these properties.**

## Context and Challenges

In hard real-time systems, the worst-case end-to-end delay of all the packets generated by a flow must be lower than a predetermined deadline. Such real-time packet schedulability analysis has been done for various types of networks by taking into account the type of contentions that can occur. In fact, a flow can be classified as either being in direct or indirect contention with the analyzed flow $f_a$. A direct flow shares at least one link with $f_a$, while an indirect flow shares at least one link only with a direct flow of $f_a$. The challenge thus lies in the ability to define methods that have a limited complexity in order to compute WCTT values that are not too pessimistic. Thus, in this work, we focus on defining network properties on the behavior of packet transmission over many-core architectures which rely on wormhole switching and round-robin arbitration ([1], [2]). In these networks, a packet is divided into flits of fixed size, which are transmitted one by one by routers. The header flit contains the routing information that defines the path the next flits will follow in a pipeline way. In our previous work, we have shown that modeling this pipeline transmission of flits of NoCs can greatly reduce the pessimism of the WCTT of flows [1].

## Main Results

Our first contribution in this paper is therefore to demonstrate three properties of wormhole-switched NoC-based many-cores when analyzing contentions between flows. The first property reduces the number of possible scenarios by studying the arrivals of the flows. Thus, we consider that the worst-case scenario is obtained when the headers of contented flows arrive synchronously on each router. The following ones concern the pipeline behavior and its utilization in the computation. First, we can reduce the pessimism by computing the maximal delay an unblocked flow can block the next one. Then, we can eliminate some scenarios by identifying an indirect flow as non-influent on the analysis of $f_a$.

Our second contribution is then to implement these properties in an algorithm and evaluate it. Our algorithm is divided into two parts. The goal of the first part is to determine all the blocking flows of $f_a$. The second part computes the delay of $f_a$, by adding each time the needed delay that the flows affecting $f_a$ can progress to unblock $f_a$.

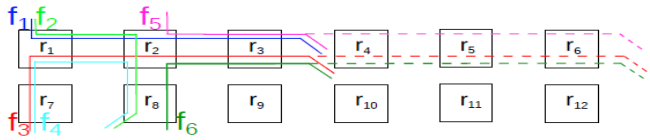



*Figure 1: Mapping of flows*

Fig.1 illustrates our case study. In order to quantify the improvements due to our method compared to a method of the state-of-the-art such as Recursive Calculus (RC), we compute the gain which corresponds to $(d_{RC} - d_A)/d_{RC}$ where $d_{RC}$ returns the worst-case latency of $f_1$ using RC method and $d_A$ the value returned by our algorithm. To identify this gain, we modify the destination of flows presented in dotted lines. Fig.2 shows the gain obtained in function of the number of hops. Our property shows that a flow blocked by an unblocked direct flow, does not wait for the direct flow to reach its destination as in RC method. This explains the shape of the curve for a fixed size of packets. For a fixed number of hops, the values of delay in both methods increase which lead to a decreasing value of the gain.

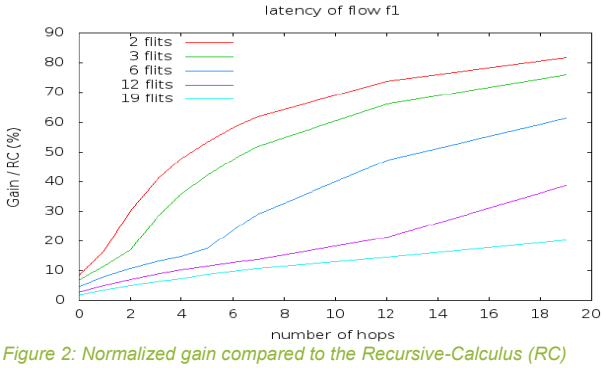



*Figure 2: Normalized gain compared to the Recursive-Calculus (RC)*

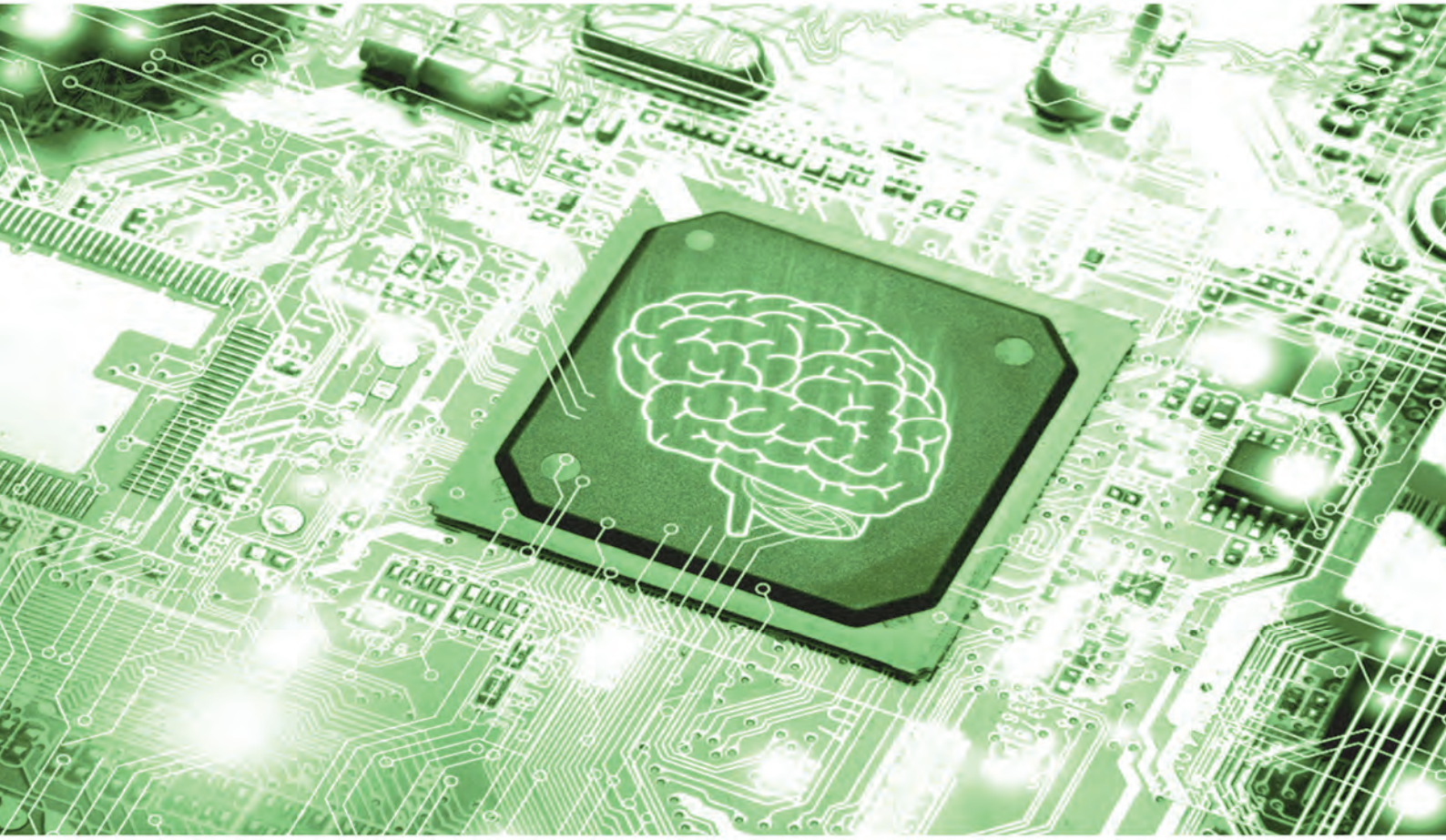
## Perspectives

The results on different synthetics benchmarks show that our properties provide an optimization of the worst-case delay of $f_a$. Further works will be done in order to study the worst-case delay in a model where the buffer has a capacity more than one flit and flows with different length of packets.

*Related Publications:*

[1] L. Abdallah, M. Jan, J. Ermont, C. Fraboul. "Optimizing worst-case delay analysis on wormhole networks by modeling the pipeline behavior". In *Proceedings of the 13th International Workshop on Real-Time Networks,* 2014.

[2] L. Abdallah, M. Jan, J. Ermont, C. Fraboul. "Propriétés des réseaux wormhole pour optimiser l'analyse de délai pire cas dans les many-coeurs". In *Summer School for Real-time (ETR),* 2015.



# 03

# COMMUNICATION: WIRELESS COMMUNICATIONS & CYBER-PHYSICAL SYSTEMS

- Low Power RF circuits for IoT
- RF Power Amplifiers
- Millimeterwave integrated circuits
- Optical communications
- Test methologies
- Middleware & Sensors network optimisation
- Cyber Physical Systems

# 2.45 GHz 0.8mW voltage-controlled ring oscillator (VCRO) in 28 nm fully depleted silicon-on-insulator (FDSOI) technology

*Research topic: FDSOI, UTBB, PLL, VCO*

**Authors: G.Jaquemod (UNS-EPOC), A. Fonseca, E. de Foucauld and Y. Leduc (UNS-EPOC)**

**Abstract: The FDSOI capability to bias the back-gate allows to implement calibration techniques without adding transistors in critical blocks. This technique is illustrated on a very low power voltage-controlled oscillator (VCO) based on a ring oscillator (RO). Despite the fact that such VCO topology exhibits a larger phase noise, this design addresses aggressively the size and power consumption reduction. The reasons to use the FDSOI technology to reach the specifications of this PLL are presented. The VCRO exhibits a 0.8mW power consumption, with a phase noise about -94 dBc/Hz@1 MHz.**

## Context and Challenges

To perform 1 MHz step channels of Bluetooth (BT), a fractional phase divider (FPD) PLL architecture is chosen because it produces no fractional division error, and is energy-saving since it allows fractional-N PLL without sigma-delta, time-to digital converter, or other averaging technique. The drawback is the control of spurious level.

## Main Results

In this work, we developed a 2.45 GHz voltage-controlled ring oscillator (VCRO) for BT applications. The proposed architecture takes advantage of the FDSOI back-gate biasing opportunity to reduce this phase noise spurious. This back-gate biasing is used to calibrate the mismatch between transistors of the VCRO, and it decreases jitter and thus phase noise. This ring oscillator topology has been chosen for its simplicity of realization (single ended inverters and their buffers) and to measure and verify all specifications (consumption, frequency gains and phase noise) for short rings. Figure 1 shows the schematic of the ring oscillator and its buffers, inputs and outputs. The back-gate biasing is used on VBGP, VBGN nodes for oscillation and DCC node for mismatch compensation. The duty cycle correction (DCC) will calibrate PLL output (Fosc) duty cycle through the back-gate voltage of an inverter, to at most +/ – 2.5% error (+/ – 10 ps). The DCC avoids the calibration loop clamping the VCRO back-gate voltages to Vdd/Gnd rails.

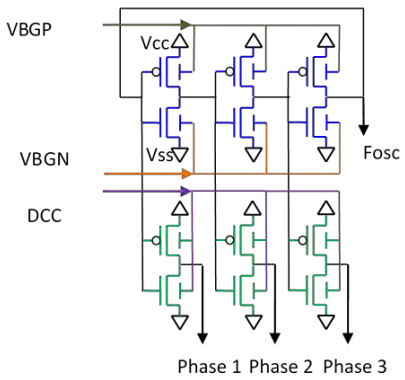



*Figure 1: Schematic of the ring oscillator*

It also corrects other output phases with the same voltage. The major part of duty cycle error being the PMOS/NMOS threshold voltage mismatch, it is quite similar on all phase outputs.

The final version of the VCRO consists of 15 delay cells (inverters). Figure 2 shows the layout of this ring oscillator: delay cells are at the bottom and buffers at the top of the picture. The area is about 600 µm2.

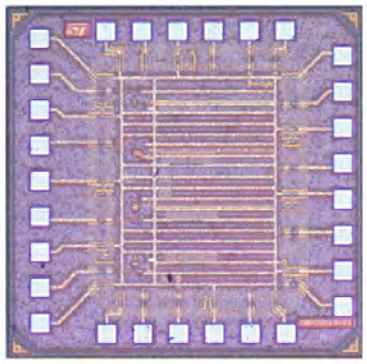


*Figure 2: MicroPhotograph of the VCRO*

We have shown how to overcome some limitations due to process variation on aggressive nodes. FDSOI technology reduces by 3 the standard deviation on the threshold voltage compared with CMOS bulk technology. Secondly, the capability to bias the back gate allows us to implement calibration techniques without adding transistors in critical blocks, thus reducing parasitic capacitance and power consumption. A first VCRO test-chip has been fabricated in 28 nm FDSOI and tested. It exhibits a 0.8mW power consumption, with a measured phase noise about -94 dBc/Hz@1 MHz.

## Perspectives

We have illustrated this technique on a very low power PLL designed in 28 nm FDSOI technology. This allowed us to verify the behavior and specifications of the main component of the PLL (VCR). Next step will be the design and manufacturing of the whole PLL.

*Related Publications:*

[1] G.Jacquemod,A. Fonseca, E. de Foucauld, Y.Leducand and Lorenzini, P. "2.45 GHz 0.8 mW voltage-controlled ring oscillator (VCRO) in 28 nm fully depleted silicon-on-insulator (FDSOI) technology." Frontiers of Materials Science 9/2 (2015): 156-162.

[2] A. Fonseca, E. De Foucauld, P. Lorenzini, and G. Jacquemod, “Process variation compensation for PLL on FDSOI 28nm,” in Proceedings of VARI 2014, Palma, Spain, 2014.

[3] A. Fonseca, E. De Foucauld, P. Lorenzini, and G. Jacquemod, “Low Power 28nm FDSOI 2.45 GHz PLL,” J. Low Power Electron. JOLPE, vol. 10, no. 1, pp. 149–162, Mar. 2014.

# Continuous Time Analog to Digital Converter for Ultra Low Power Radios in 28nm FD-SOI

*Research topic: CT-ADC/DSP, Ultra-Low-Power, Wake-Up Radio*

**Authors: A.Ratiu, D.Morche, S. Patil and Y.Tsividis (Columbia University)**

**Abstract: Continuous time digital signal processing represents a class of systems in which the information is encoded in a set of discrete levels in the amplitude domain as well as in the timing between the arrivals of these levels. These systems present a set of interesting properties (activity-dependent power dissipation, require no clocks, good programmability) which we exploit for the design of an ultra-low power radio back-end. The circuit which has been achieved this year demonstrates that it is possible to cumulate both excellent power efficiency (similar to the best classical ADC) and activity-dependent power dissipation.**

## Context and Challenges

CT analog-to-digital conversion can be viewed as defining a set of discrete digital levels and triggering a digital output whenever a level is crossed by the input signal. The information is thus encoded in the digital level as well as in the time of the level crossings. The output signal can be processed by a continuous time DSP made up of asynchronous delay cells and asynchronous adders which preserve the timing between events. Such systems have been proven to be energy efficient in performing interferer rejection for ultra-wide band signals [1]. However, up to now, the efficiency of the achieved Analog to Digital Converter were not competitive with respect to classical ADC. This design is focused on a power-efficient implementation suitable for ultra-low power radios.

## Main Results

We have proposed a novel continuous-time ADC architecture that allows a programmable, highly compact, and power-efficient circuit implementation, while preserving the benefits of continuous-time ADC/DSP systems. It is a modified version of delta modulator as shown in Figure 1.

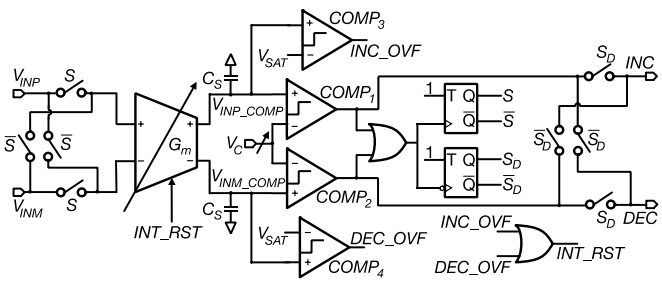



*Figure 1: Proposed CT ADC architecture*

In classical delta modulators, the loop delay is limited by the N-bit feedback DAC that generates the $2^N$ levels. However, in a uniform-resolution async ADC, the amplitude separation between two successive samples is always 1 LSB. The proposed clockless ADC exploits this by replacing the N-bit DAC with chopping switches that implement a 1-bit DAC. Through these, the fully differential input (S=1) or its negative version (S=0) is fed to a Gm-C integrator. Once the threshold, VC, is crossed by one of the integrator outputs, a comparator (COMP1-2) generates a narrow pulse. Therefore, the output pulse rate, and hence power dissipation, is proportional to the slope of the integral of the input, and thus to the input signal value. This makes it partially behave like a voltage-to-frequency converter (VFC). But, unlike VFCs, our ADC is event driven and produces no pulses when the input is zero. Besides, flipping causes no charge loss and is thus more power efficient than a VFC's integrate-and-reset.

Following the proposed architecture, a circuit has been design using the 28nm FD-SOI technology. It takes benefit of the body bias to reach high speed when needed and to reduce power consumption when possible. The ADC core occupies only 45x72µm2 (0.0032mm$^2$). No calibration is required and no complex post-processing is used for reconstruction. The noise and distortion power was integrated over the 10MHz–50MHz BW, resulting in 32dB–42dB SNDR, which is sufficient for ultra-low-power applications like wake-up receivers. The total power dissipation is 24µW, resulting in a FoM of 3–10fJ/conv-step over the BW which compares favorably with state of the art (as shown in Figure 2).

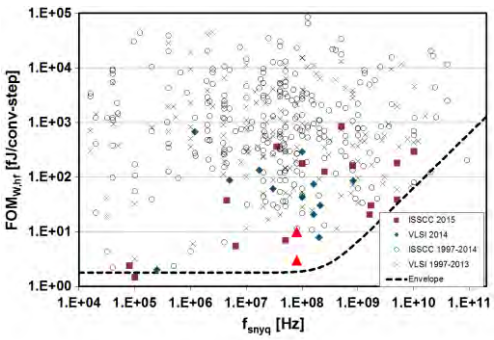



*Figure 2: Performance comparison with state of the art (B.Murmann Survey)*

## Perspectives

We have demonstrated the efficiency of the CT ADC approach for medium resolution ADC. The next step will be the implementation of the following stage whose purpose is to process the signal efficiently. A tunable band-pass filter would be particularly adapted to the application to reject out of band blockers.

*Related Publications:*

[1] M. Kurchuk, C. Weltin-Wu, D. Morche, Y. Tsividis, "Event-Driven GHz-Range Continuous-Time Digital Signal Processor with Activity-Dependent Power Dissipation", IEEE Journal of Solid State Circuits (JSSC), 2012

[2] A. Ratiu, D. Morche. A. Arias, B. Allard, X. Lin-Shi, J. Verdier, "Efficient Simulation of Continuous Time Digital Signal Processing RF Systems", Sampling Theory and Applications (SAMPTA), 2013.

[3] A. Ratiu, D. Morche. A. Arias, B. Allard, X. Lin-Shi, J. Verdier, "Continuous Time Analog to Digital Conversion in Interferer Resistant Wake Up Radios", IEEE PhD Research in Microelectronics (PRIME), 2014.

[4] Patil, S. ; Ratiu, A. ; Morche, D. ; Tsividis, Y. "A 3–10fJ/conv-step 0.0032mm2 error-shaping alias-free asynchronous ADC", VLSI-Circuits 2015

# A Low Power 4 and 8GHz IR-UWB Receiver Front-End for Long Range, IOT Devices and Precise Ranging.

*Research topic: IR-UWB, Low Power, Ranging, IoT*

**Authors: L. Ouvry, G. Masson, F. Hameau**

**Abstract: A Dual-Band CMOS receiver front-end for IR-UWB communications and ranging was designed with the aim to fully benefit from the scarce nature of the received impulsive signal. Instead of relying on a continuous local oscillator, it is based on a 2.75ns template generation and a lower frequency 1GHz frequency synthesis. The duty cycling with less than 1ns settling time enables up to 82% current savings for the front-end elements at 15.6MHz pulse repetition frequency with marginal performance degradation. The receiver can operate in two band groups centered respectively around 4 and 8GHz and sub-divided in three different channels each in compliance with IEEE802.15.4a/6 band plans.**

## Context and Challenges

Today, Bluetooth Smart, or similar widely spread integrated solutions, are used in many IoT devices to bring low power connectivity and coarse ranging useful to find tagged belongings. With IR-UWB technology and coherent integration over long trains of pulses, one gets a much more accurate ranging measurement at higher range. However IR-UWB suffers from higher power consumption. To make it disruptive for the mass market and the IoT devices, we propose to drastically reduce the consumption with a windowed-based approach, without making use of a continuous high frequency LO generation.

## Main Results

Our work is based on an existing IR-UWB coherent transceiver architecture, particularly well-suited for long range and precise ranging. This former architecture is modified with the aim to fully benefit from the scarce nature of the received impulsive signal. Thus, each front-end element is modified to be compatible with a duty cycled processing, and the continuous 4 or 8GHz local oscillator is suppressed. The down-conversion is ensured by a 2.75ns template generation, and a duty cycling controller is introduced, synchronized on a 1GHz clock reference.

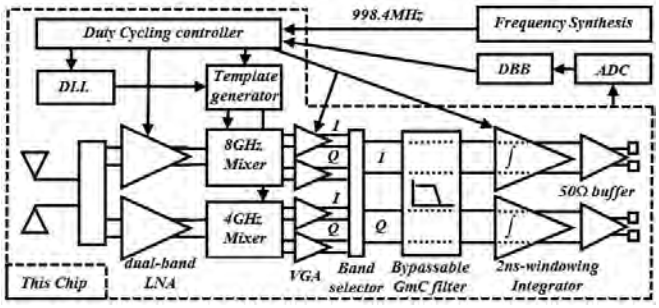



*Figure 1: Block diagram of the dual-band duty cycled receiver architecture.*

Figure 1 shows the architecture. The signal from the differential antenna is first amplified by a dual-band 4GHz-8GHz resonant LNA. Two distinct passive mixers fed by a channel-dependent 2.75ns-width pulse template from the template generator, serving as LO, ensure the quadrature down conversion. The quadrature base-band signals are then amplified by a VGA.

Out-of-band interferers are mostly filtered in the analog baseband domain thanks to 2ns-windowing integrators. If the out-of-band filtering and blocking constraints cannot be fulfilled solely by the integrators, a programmable 4$^{th}$ order Gm-C filter can be switched on, which however cannot be duty-cycled.

The pulse template generator is based on a voltage controlled delay line (VCDL) composed of cells duplicated from a duty-cycled delay loop line (DLL) which produces and stores the exact LO frequency from the 1GHz clock reference. A start template rising edge signal propagates along the VCDL, generating the rising and falling edges which are used in an impulse generator to produce the effective short template.

Thanks to a very low settling time lower than 1ns, the current saving with this approach is up to 82% for the front-end elements (LNA, LO, Mixer, VGA and Integrator-Filter) in duty cycling mode at 15.6MHz pulse repetition frequency (Tab. 1).

*Table 1: Comparison of simulated and measured current consumption in full mode and duty cycled mode.*

| | Simulated (@1.2V) (mA) | | Measured (@1.2V) (mA) | |
|---|---|---|---|---|
| **Supply** | **Normal** | **Duty Cycled** | **Normal** | **Duty Cycled** |
| **LNA** | **5.95** | **0.6** | **5.72** | **0.6** |
| **Mixer** | **0.61** | **0.61** | **0.68** | **0.68** |
| **VGA** | **5.1** | **0.6** | **4.97** | **0.64** |
| **Integrator** | **0.72** | **0.72** | **0.81** | **0.81** |
| **DLL** | **4.2** | **0.2** | **4.34** | **0.2** |
| **Sub Total** | **16.58** | **2.73** | **16.52** | **2.93** |
| Biasing | 1.3 | 1.3 | 1.7 | 1.7 |
| DC Controller | 1 | 1 | 1.28 | 1.28 |
| Sub Total | 18.88 | 5.03 | 19.5 | 5.91 |
| *Gm-C filter* | *9.5* | *9.5* | *8.68* | *8.68* |
| *Total* | *28.38* | *14.53* | *28.18* | *14.59* |

## Perspectives

Future works aim at designing a complete low power IR-UWB transceiver compliant with IEEE802.15.4a/6 band plans. The work will consist in completing the present architecture with a digital baseband, an existing low power transmitter, an existing 1GHz frequency synthesis, and new low power ADCs, both interleaved, which would enable multi-paths processing during a pulse repetition period.

*Related Publications:*
[1] L. Ouvry, G. Masson, F. Hameau, B.G Gaillard, B. Caillat, "A CMOS Duty-Cycled Coherent RF Front-End IC for IR-UWB Systems", IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB), 2015.
[2] G. Masson, L. Ouvry, F. Hameau, B. Caillat, "A 4 and 8GHz CMOS130nm Duty-Cycled Front-End for Ultra Low Power IR-UWB Receivers", IEEE International Conference on Electronics, Circuits, and Systems (ICECS), 2015.

# A SOI CMOS Reconfigurable Matching Network for Multimode Multiband PA

*Research topic: RF SOI, Power Amplifier, Reconfigurable RF PA*

**Authors: A. Giry, P. Ferris**

**Abstract: A RF SOI reconfigurable output matching network (OMN) for multimode multiband power amplifier applications has been designed, implemented and tested. The proposed reconfigurable OMN covers up to ten E-UTRA frequency bands ranging from 700 MHz to 920 MHz and presents low impedance levels. It has been designed with four distinct RF output paths, one path being dedicated to 2G PA applications and three paths being dedicated to 3G/4G PA applications. The realized prototype occupies a silicon area of 1.5 mm2 in a 130 nm SOI CMOS PD process.**

## Context and Challenges

In the last two decades, the continuous growth in wireless applications has led to the development of multiple cellular standards (2G/3G/4G) with different requirements in terms of output power and linearity. In the meantime, a large number of cellular frequency bands was defined to support global roaming of mobile handsets. The need for wireless devices with smaller form factor and reduced cost is driving research towards flexible multimode multiband power amplifiers (MMPA) with a high integration level to efficiently afford multimode multi-band requirements. Today's MMPA modules are mostly based on multi-chip module approach that uses GaAs technology and multiple PA cores.

This research activity targets the design and implementation of reconigurable MMPAs [1] covering several modes and frequency bands using a reduced number of PA cores in RF SOI technology. To get optimum MMPA performances, the output matching network (OMN) has to present an optimal load impedance to the power stage with reduced insertion loss (<1dB). This optimal load impedance depends on the required output power and linearity and is generally low (<3Ohm) when considering low voltage high power (>1W) PA design.

## Main Results

In this work [2], a reconfigurable OMN for 2G/3G/4G MMPA applications has been designed and implemented in a RF SOI process. The proposed circuit has four RF output ports, covers up to ten E-UTRA frequency bands ranging from 700 MHz to 920 MHz and can provide different low impedance levels (<3Ohm) to a high efficiency SOI LDMOS MMPA. Schematic of the proposed reconfigurable output matching network (OMN) is depicted in Fig. 1. It provides optimum matching to a reconfigurable PA [3] using L-type matching networks and four distinct RF paths. One path (P4) is dedicated to 2G PA applications operating in bands 5 and 8, the three other paths (P1,P2,P3) being dedicated to 3G/4G PA applications operating in LTE bands ranging from 700MHz to 920MHz. All capacitors and RF switches are integrated into a RF SOI 130nm process with high resistivity substrate. Series inductors L2 and L3 are implemented on chip whereas input inductor L0 and inductor L1 are realized using bondwires to get higher Q (>50).

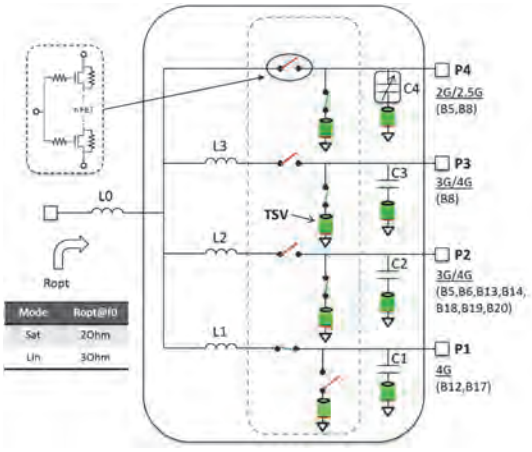



*Figure 1: Schematic of the RF SOI reconfigurable OMN*

Fig.2 shows a photograph of the reconfigurable OMN. The proposed circuit occupies an area of 1.5 mm2 including pads, negative voltage generator and control interface (SPI).

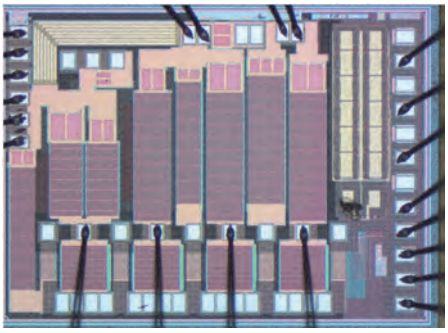


*Figure 2: Micrograph of the die*

## Perspectives

The obtained results compare favorably to the state-of-the-art, with a higher integration level. Future design with reduced loss and silicon area are expected by using advanced RF SOI process with thick copper back-end.

*Related Publications:*
[1] A.Giry, "SOI CMOS PA for Mobile Applications," in IEEE MTT-S International Microwave Symposium, Workshop WSA, Tampa, Jun. 2014
[2] G. Tant, A. Giry, P. Ferris, G. Pares, J.-D. Arnould, J.-M. Fournier, C. Raynaud, P. Vincent, "A SOI CMOS reconfigurable output matching network for multimode multiband power amplifiers," in Microwave Symposium (IMS), 2015 IEEE MTT-S International , vol., no., pp.1-4, 17-22 May 2015
[3] P. Ferris, G. Tant, D. Parat, A. Giry, J-D. Arnould, J-M. Fournier, « Un Amplificateur de Puissance Multimode Multibande reconfigurable en technologie CMOS SOI », XIXèmes Journées Nationales Microondes (JNM), 3-5 Juin 2015

# A 270-to-300 GHz sub-harmonic injection locked oscillator for frequency synthesis in sub-mmW systems

*Research topic: Sub-harmonic injection-locked oscillator, VCO, THz, mmW, BiCMOS*

**Authors: A. Siligaris, Y. Andee, C. Jany, V. Puyal, J. Moron Guerra, J.L. Gonzalez Jimenez, P. Vincent**

**Abstract: This work describes a sub-mmW oscillator with more than 10% tuning range oscillation frequency. In particular, the circuit includes a high sensitivity sub-harmonic injection technique that forces the 270-to-300-GHz output signal to multiply by 6 a 45-to-50-GHz injection signal. Thus, the output spectral and phase noise properties depend only on the injected lower frequency signal. The circuit is fabricated in a BiCMOS 55 nm technology and it is measured using probes. It achieves a 30 GHz tuning range and locks all over it. The measured phase noise is -105 dBc/Hz at 1 MHz offset at 297 GHz and the power consumption is 52 mW.**

## Context and Challenges

Sub-mmW frequency band is gaining an increased interest for RF system development. Indeed, within the J-band (220-325 GHz) a large spectrum is available where various applications are developed like short range radar, THz imaging, chip-to-chip high throughput communications or backhaul communications. Such systems need a stable and low phase-noise local oscillator that must provide a reference frequency for up and down conversion. In this work, we describe an oscillator that can be locked from a sub-harmonic external frequency reference. The oscillator provides an output signal tuned from 270 GHz to 300 GHz and is locked by a 45-to-50-GHz sub harmonic injection signal. Thus, the PLL operation can be shifted to lower frequency where phase noise performance is better and design constrains are smaller. The circuit is fabricated in a BiCMOS 55nm technology.

## Main Results

The oscillator core is designed to operate in the 90-to-100-GHz band. The third harmonic is exploited in order to obtain an output signal in the 270-to-300-GHz band. For that, the triple push architecture has been applied. Indeed, the triple-push architecture is a good compromise for energy efficiency and 3rd harmonic extraction. It uses three common emitter stages that are connected in a closed loop.

Oscillation frequency spans from 272 GHz to 303 GHz and the output power is up to -9 dBm (figure 2.a). The oscillator draws a current ranging from 33 mA to 52 mA from a 1.1 V supply. The sub-harmonic injection locked oscillator sensitivity curves are shown in figure 2.b. In locking conditions, the output frequency and phase noise properties are copied from the injected signal and do not depend any more on the intrinsic oscillator characteristics. Figure 3.a shows the measured phase noise at the output when locked at 297 GHz ($F_{inj}$=49.5 GHz). We observe that the output phase noise reproduces the source phase noise augmented by *20log(6)* (multiplication factor). This result demonstrates a good locking of the oscillator. Figure 3.b illustrates the measured output spectrum at 297 GHz where we observe a very clean signal thanks to locking.

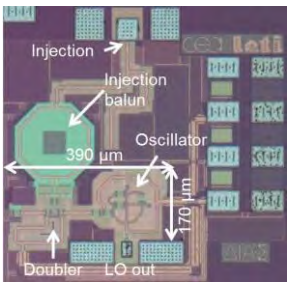



*Figure 1: Micro-photograph of the oscillator fabricated in a BiCMOS 55nm technology.*

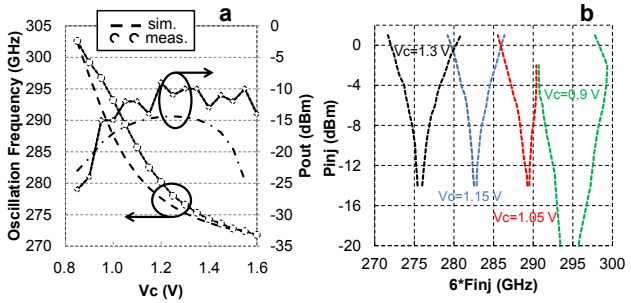



*Figure 2: a. Measured free running oscillation frequency versus the control voltage and corresponding output power. b. Measured sensitivity curves of the sub-harmonic injection locked oscillator for 4 different control voltages.*

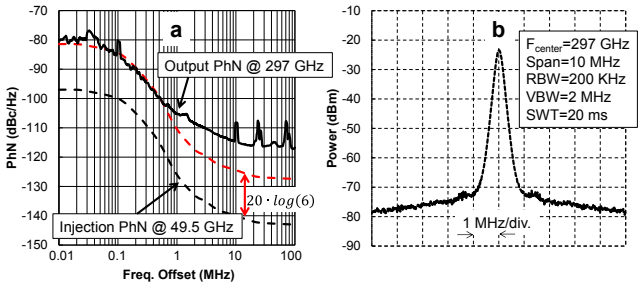



*Figure 3: a Measured output phase noise at 297 GHz under locking conditions on an injection signal at 49.5 GHz. b. Output spectrum locked at 297 GHz.*

## Perspectives

This work sets the first bases for THz systems development for high data rate communications and THz imaging.

*Related Publications:*

[1] A. Siligaris, Y. Andee, C. Jany, V. Puyal, V.; J.M. Guerra, J.L. Gonzalez Jimenez, P. Vincent, "A 270-to-300 GHz Sub-Harmonic Injection Locked Oscillator for Frequency Synthesis in Sub-mmW Systems," *IEEE Microwave and Wireless Components Letters*, vol.25, no.4, pp.259-261, April 2015.

[2] J.M. Guerra, A. Siligaris, J.-F. Lampin, F. Danneville, P. Vincent, "A 285 GHz sub-harmonic injection locked oscillator in 65nm CMOS technology," *IEEE International Microwave Theory and Techniques Symposium, MTT-S*, pp. 1-3, June 2013.

[3] J.M. Guerra, A. Siligaris, J.-F. Lampin, F. Danneville, P. Vincent, "A 283 GHz low power heterodyne receiver with on-chip local oscillator in 65 nm CMOS process," *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pp. 301-304, 2-4 June 2013. .

# Millimeter wave access point for future 5G Heterogeneous Network

*Research topic: 5G heterogeneous network, access point, millimeter wave*

Authors: C. Dehos, L. Dussopt, O. El Bouayadi, J. Luna, & Y. Lamy

**Abstract: Small cell millimeter wave access point is proposed for offloading part of the mobile data traffic in future 5G heterogeneous network. A top down analysis is adopted to specify the mmW devices from uses cases and required KPI. A versatile access point composed of multiple antenna array modules is proposed to address multiple users in the cell with beam steering. Antenna array module gathers on the same organic interposer package a 60GHz transceiver, switches, active phase shifters, and an antenna array fabricated in different technologies.**

## Context and Challenges

The exponential increase of mobile data traffic, driven by smartphone and tablets, requires disrupting approaches in the definition of the future 5G network. The trend is to reduce the network cell size and offload a great part of this traffic to small cell access points, optically or wirelessly linked together and backhauled to the core network. In this scope the huge frequency bands available at millimeter wave should be good candidates for opportunistic high data rate data transfer.

The latest breakthrough in CMOS and BiCMOS technologies are paving the way for the development of mmW devices at low cost for 5G small cells. We propose a heterogeneous network infrastructure based on the superimposing of millimeter wave access point and backhaul to the former cellular infrastructure. Our works focus mainly on the tricky mmW access point architecture and design.

## Main Results

From use cases and system studies, access point device has been specified with the aim of delivering from 150Mbps to 4Gbps connectivity to every user terminals, located into a 50m radius cell. Such a high range is obtained at 60GHz thanks to highly directive antenna arrays with beam steering capabilities.

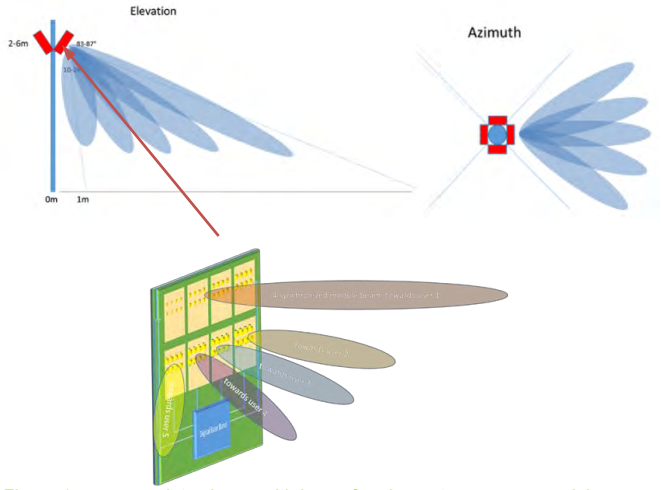



*Figure 1: access point scheme with beam forming antenna array module*

Indeed the proposed access point is composed of sectors, composed of antenna array modules. Each module may address a different user in the cell, or many modules may cooperate to address the same user with improved range or data rate.

The versatile antenna array module is composed of a CMOS 65nm transceiver operating in one of the four 60GHz channels, a Transmit/Receive switch, and active phase shifters for antenna beam forming. The phase shifter front-ends, designed in BICMOS55nm technology, achieves strong signal amplification in both Transmit/Receive directions. The antenna array is integrated into the multilayer organic interposer module.

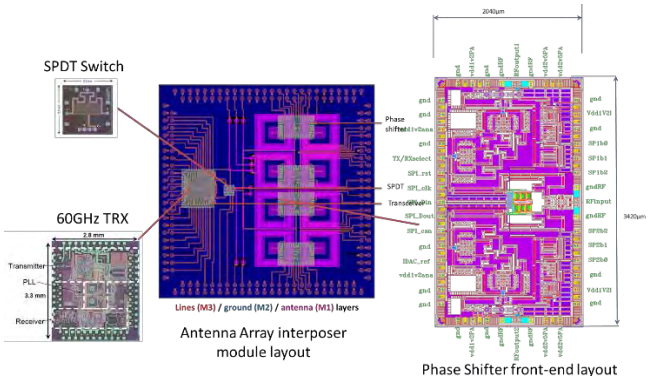



*Figure 2: antenna array module layout with active components*

Different phase shifter topologies has been designed and evaluated. Vector modulators and polyphase filters are promising solutions to reduce the signal losses while keeping gain and phase precision throughout the 9GHz bandwidth.

## Perspectives

Proof of concept access point prototypes should be realized in 2016. Further design works deal with new phase shifters topologies for better performance and power consumption.

*Related Publications:*

[1] Millimeter-Wave Access and Backhauling: The Solution to the Exponential Data Traffic Increase in 5G Mobile Communications Systems? Cedric Dehos, Jose Luis González, Antonio De Domenico, Dimitri Kténas, and Laurent Dussopt IEEE Communications Magazine • September 2014.

[2] Dussopt, L.; El Bouayadi, O.; Luna, J.A.Z.; Dehos, C. & Lamy, Y.. "Millimeter-wave antennas for radio access and backhaul in 5G heterogeneous mobile networks." in Proceedings of the 9th European Conference on Antennas and Propagation, EuCAP 2015, 13 May 2015 through 17 May 2015: 1-4.

# Clock receiver and data transceivers for Optical Network on Chip

*Research topic: Silicon Photonics, Optical-NoCs*

**Authors: R. Polster, G. Waltener, J.L. Gonzalez**

**Abstract: Optical networks on-chip (ONoC) have specific requirements regarding devices foot-print and power consumption for silicon-photonics components. This is particularly true for photonics IPs that aimed to be integrated in microprocessors and memories ICs connected to this network. CEA-LETI is developing small foot-print, very low power consumption drivers and receivers dedicated to silicon photonics transceivers. Moreover specific clock receivers have been developed which enable low power optical clock distribution across the network for small latency and improved synchronization.**

## Context and Challenges

Data Centers (DC) and High Performance Computers (HPC) are one of the key elements of today's information based society. The computational power and data bandwidth requirements are increasing at an exponential rate. Newer architectures for these DC and HPC that allow increasing the computer power and data bandwidth without exploding the power consumption are currently under development. Optical cables are nowadays replacing electrical cable for the long, medium and even meter-range interconnections between the computer racks and blades use to build such massively parallel computers. The next step is to bring optical interconnections to an integration scale at the level of the package, which would eventually allow connecting the microprocessors and memories inside the package. This will allow breaking the current bottleneck found at the interface between microprocessor and memory for increased memory bandwidth density, but requires having power efficiencies high enough to compete with short electronic interconnections currently used at this scale.

## Main Results

Electronic receivers [1] and drivers [2], able to operate at data rates up to 10 Gbps with energy efficiencies well below 1pJ/bit, have been demonstrated. These circuits are developed on a CMOS 65nm node. They have been designed to be connected to a Ge photodiode and a ring modulator, respectively, fabricated using Leti's silicon photonics technology [3]. Fig. 1 and 2 show measurement results for the TIA and modulator driver.

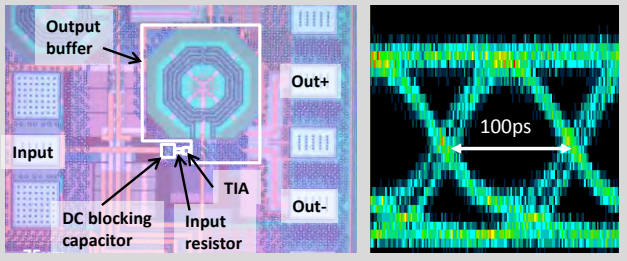



*Figure 1: 10Gbps Transimpedance amplifier.*

The photonic IC is currently under fabrication. 3D technology will be used to connect the electronic IC to the silicon IC. This allows reducing dramatically the interconnection capacitance, with significant benefits to the power efficiency of the circuits. For examples, the TIA consumes just 0.26 mW providing a gain of 70 dBΩ thanks to a total input capacitance (including the photodiode) of only 7 fF. The driver is able to drive 70 fF of load with a 2.4 V swing consuming just 0.5 pJ/bit.

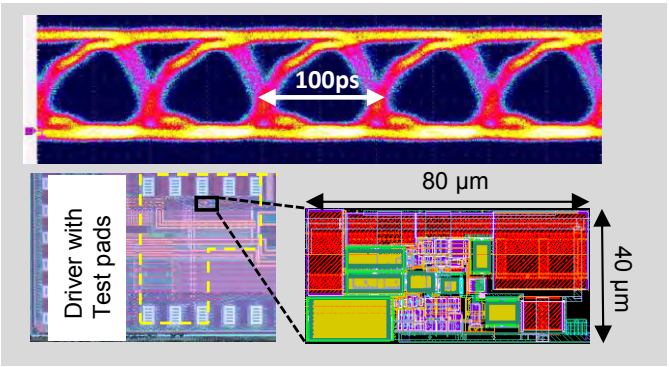



*Figure 2: 10Gbps optical modulator driver.*

In addition to data transceiver, two specific receivers for optical clock have been developed [4, 5]. They exploit the possibility to use a short pulsed laser for generating the clock signal, which reduces its power consumption. The optical clock is received by the electronic ICs composing the ONoC and transformed into a 50% duty cycle electrical clock usable by the receiving and transmitting electronics.

## Perspectives

Optical test on a full link realized with flip-chip electronic ICs mounted on a silicon photonics interposer will be carried out during 2016. Furthermore, the TIA and modulator driver are currently migrated to a 28nmFDSOI technology node, forecasting integration in state-of-the arte microprocessors and memories developed for this node.

*Related Publications:*

[1] R. Polster, J.L. González, E. Cassan, L. Vivien, (2015), "A TIA for optical networks-on-chip in 65nm CMOS," OIC 2015, pp. 109-110.
[2] G. Waltener, J.L. González (2015), "Driver de modulateur optique à haute tension et 10 Gbps," 19èmes Journ. Nat. Microondes, pp. 188-190, June 2015.
[3] B. Charbonnier, Sylvie Menezo, P. O'Brien, Aurélien Lebreton, J. M. Fedeli, and B. Ben Bakir, "Silicon Photonics for Next Generation FDM/FDMA PON," J. Opt. Commun. Netw. 4, A29-A37 (2012)
[4] R. Polster, J.L. González, E. Cassan, "A Novel Optical Integrate and Dump Receiver for Clocking Signals," Proc. of 13th IEEE International New Circuits and Systems Conference (NEWCAS), June 2015.
[5] R. Polster, J.L. González, I. Miro-Panades, E. Cassan, "An optical clock receiver based on an injection locked ring oscillator featuring auto-calibration", Proc. of IEEE Midwest Symposium on Circuits and Systems (MIDWEST), August 2015.

# De-embedding the noise figure of differential amplifiers using the correlation of output noise waves

*Research topic: noise measurement, differential amplifier, hybrid coupler, network analyzer*

**Authors: Y.Andee, C.Arnaud, P.Seurre (Rohde&Schwarz), F.Danneville (IEMN)**

**Abstract: This work presents a measurement technique for de-embedding the noise figure of differential amplifiers. It is based on the measurement of the correlation of the noise waves at the output ports of the differential amplifier. It makes use of an hybrid coupler and takes into account the phase and amplitude imbalances of the latter. Measurement results of a radio-frequency low-noise amplifier demonstrate the validity of this general technique.**

## Context and Challenges

The classical technique for noise figure measurements of differential amplifiers consists in embedding the latter between two hybrid couplers or baluns. The noise figure of the cascaded two-port system is measured using the hot/cold method with a calibrated noise source and a noise figure analyzer. The noise figure of the differential DUT is then de-embedded from the cascaded system using the Abidi technique. This is convenient as all measurements are performed in a single-ended way using conventional two-port equipment. It has however some limitations as the phase and amplitude imbalances of the couplers are not considered during the de-embedding procedure. These imbalances contribute to measurement errors in the de-embedding of the differential noise figure. The aim of this work is to develop an accurate technique that takes into account these imbalances.

## Main Results

It has been demonstrated in [1] that the noise figure of a differential circuit is function of the correlation of the noise waves at the two output ports of the circuit. This correlation cannot be measured directly with available equipment.
The principle proposed in this work to measure this correlation consists in connecting only one hybrid coupler at the output ports of the differential DUT as we can observe on Fig.1.

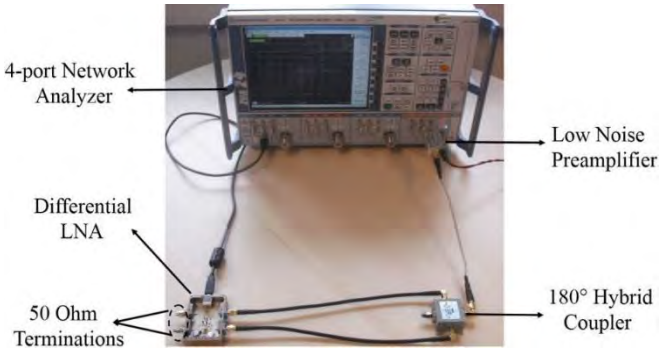



*Figure 1: Photograph of the measurement setup of the cascaded system*

The noise power measured at the output port of the cascaded system is function of the noise powers at the output ports of the differential DUT, the S-parameters of the coupler and the correlation of output noises [2]. The noise powers and S-parameters are measured easily with a Rohde & Schwarz 4-port Network Analyzer. The correlation is a complex term, i.e. it consists of a real part and an imaginary part. As there are therefore two unknown terms, two configurations are required. The second configuration is obtained by switching the connections between the output ports of the DUT and the input ports of the coupler.
The output noise powers of the 2 configurations and the S-parameters of the coupler are used to calculate the real and imaginary parts of the correlation of output noises.

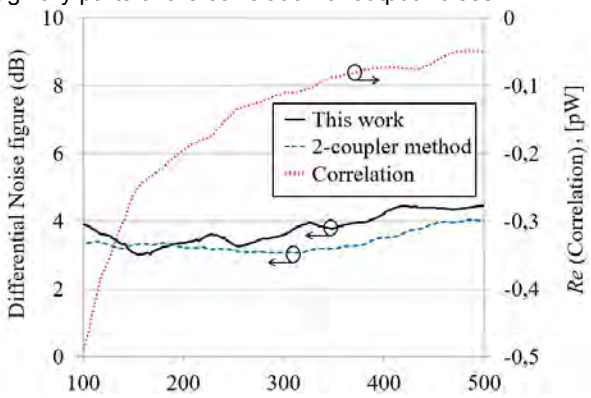



*Figure 2: Comparison of the 2 methods*

Fig.2. shows the comparison between the differential noise figures obtained with our new technique and with the classical 2-coupler approach. There is a quite good agreement between the two noise figures. Some deviations are due to measurement uncertainties and to the coupler's imbalances that are not taken into account in the 2-coupler method.

## Perspectives

This work proposes a de-embedding technique for the noise figure of differential amplifiers. It is based on the measurement of the correlation of the noise waves at the output ports of the amplifier. Future research will be oriented towards improving the theoretical study, by considering differential amplifiers with arbitrary (non-100 Ω) input and output impedances.

*Related Publications:*
[1] Y. Andee et al. "Determination of noise figure of differential circuits using correlation of output noise waves," Electronic Letters, vol. 50, no. 9, Apr. 2014.
[2] Y. Andee et al. "On-wafer differential noise figure measurement without couplers on a vector network analyzer," 84th ARFTG Conference, May 2015.

# Test Precision Optimization for Cost Reduction

*Research topic: RF & Analog Circuits Test, Indirect Test, Test Cost Reduction*

**Authors: M. Verdy, S. Lesecq, E. De Foucauld, D. Morche, J.P. Mallet, C. Mayor (Presto-Engineering)**

**Abstract: . Reducing test time or using simplified circuit characterization are well known approaches to reduce the cost of the test. However, they usually translate into a reduction of the precision of the test which can be tolerated up to a certain threshold. This work presents a novel method to evaluate the impact of precision reduction on yield accuracy, using only measured values and easy-to-obtain uncertainty models. Thanks to this approach, it is possible to obtain from a single set of measurements of the circuit the value of the minimum precision which is required for the considered measurement.**

## Context and Challenges

Reducing test costs of analog and RF circuits is a complex challenge, for which intuitive solution is to reduce test time. However, such reduction usually leads to a degradation of measurement accuracy not easy to handle when no model is available to understand the impact of the reduction. Another solution is to resort to alternate simplest test [1] whose cost is much lower. However, a model should be constructed to link the specified performances to the measured one. The precision of the model will dictate the efficiency of the indirect test. An improvement of the precision can only be obtained at the cost of bigger data base and more complex models. In both cases, the optimization of the precision is thus mandatory to reduce the test cost. Up to now, mainly empirical approaches have been used because the influence of the noise is strongly influenced by the distribution of the circuit performances.

## Main Results

We have developed a new method [2] to evaluate the impact of test time reduction on yield accuracy, using only measured values and easy-to-obtain uncertainty models. The results proposed by this method provide a balance between test time reduction and yield accuracy. To demonstrate the efficiency of the proposed approach, it has been applied to the evaluation of a SNR measurement of an Analog to Digital Converter.
The approach proposed is based on a de-convolution in the Fourier domain. Unfortunately this approach, if applied directly, would lead us to solve an ill-conditioned deconvolution problem.

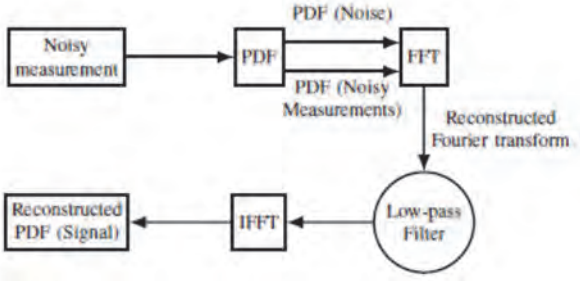



*Figure 1: Proposed method flow*

To circumvent this issue, we have chosen to use a low-pass filter which rejects the frequencies where the magnitude of the measurement noise is no longer the dominant one. The cutoff frequency of that filter is computed from the knowledge of the noise Probability Density Function. The method flow, including the low-pass filter, is represented in Fig.1. The measurement noise distribution is acquired by measuring several times a noisy measurement from a single device. This noise is considered independent from the signal.
The proposed approach has been applied to the evaluation of the Signal to Noise Ratio (SNR) measurement of a 15MHz signal, which is the output of an ADC [3]. The SNR measurement is highly linked to the number of observed samples since it is based on a Fourier Transform.
In Figure 2, the yield losses obtained by the proposed approach (called reconstructed) is compared with the real ones. The saw-tooth behavior of the “Real” curve is due to the limit of precision of the measured yield for the considered number of circuits. It shows the good obtained precision. Then, as a function of the tolerated yield losses, the optimal number of measured point can be directly obtained.

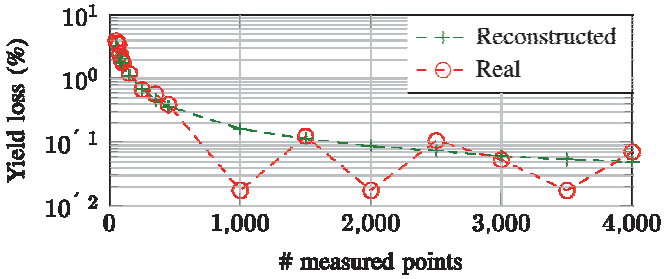



*Figure 2 : Yield losses as a function of the number of samples*

## Perspectives

The results could be improved by taking into account the correlation between the measurement noise and the circuit performances. However, this was not needed in the considered ADC. Additional methods will be established in order to be able to rapidly characterize, for a given circuit, the number of samples which are needed to fully characterize this circuit behavior according to the targeted yield losses. Such method will exploit the approach proposed in this paper.

*Related Publications:*
[1] Verdy, M.; Ratiu, A.; Morche, D.; De Foucauld, E.; Lesecq, S.; Mallet, J.-P.; Mayor, C .”Cost-driven statistical analysis for selection of alternative measurements of analog circuits” Electronics, 21st IEEE International Conference on Circuits and Systems (ICECS), 2014
Pages: 104 - 107
[2] Verdy, M.; Morche, D.; De Foucauld, E.; Lesecq, S.; Mallet, J.-P.; Mayor, C. “Balancing test cost reduction vs. measurements accuracy at test time” 13th IEEE International New Circuits and Systems Conference (NEWCAS), 2015 Pages: 1 – 4
[3] Patil, S. ; Ratiu, A. ; Morche, D. ; Tsividis, Y. “A 3–10fJ/conv-step 0.0032mm2 error-shaping alias-free asynchronous ADC”, VLSI-Circuits 2015

# sensiNact: Service-oriented Horizontal IoT Platform for Smart Cities

*Research topic: smart city, internet of things, open platforms, fog computing*

Authors: L. Gurgen, C. Munilla, R. Gruillhe, E. Gandrille, J. Botelho do Nascimento

**Abstract: Internet of things has the potential to achieve more efficient management of the urban resources (infrastructure, economic, natural, etc.) to be shared by the increasing population being concentrated in urban areas. The unprecedented number of connected things raises new technical challenges in terms of interoperability, scalable and online data processing, actionable knowledge extraction, dependable application development and deployment, etc. We present sensiNact, an interoperable IoT platform addressing those challenges with a service oriented approach. The platform has been validated in real-life testbeds and field trials within several collaborative projects.**

## Context and Challenges

The world is facing a number of critical challenges such as global warming, economic crisis, security threats, inequality, natural disasters and ageing society. Urban areas are particularly affected, given that the world population is increasingly concentrated in those areas. IoT is a key driver for world sustainability helping to increase the efficiency in using shared urban infrastructure, economic and natural resources. However, the unprecedented number of connected things and the need to process the associated big data naturally raises new technical challenges in terms of interoperability, data and connectivity heterogeneity, scalable and online data processing, actionable knowledge extraction, dependable application development and deployment, etc.

## Main Results

We have developed an IoT platform, namely sensiNact, to address the above mentioned challenges. It provides interoperability among various heterogeneous IoT protocols and platforms (e.g. Zigbee, CoAP, enOcean, MQTT, XMPP, OMA LWM2M, FIWARE, etc.) with a service oriented approach [1].

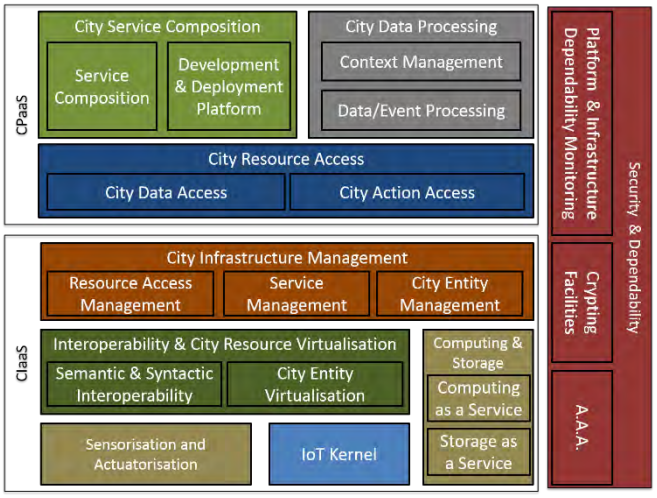



*Figure 1: Reference architecture of IoT + Cloud integration for smart cities.*

Based on a City Infrastructure as a Service (ClaaS) model, all data from the city (sensor devices, legacy devices, mobile applications, social networks, etc.) are virtualized and made available via the City Platform as a Service (CPaaS), which provides secure access to and complex event processing on the data [2]. It also provides tools such as sensiNact Studio for rapid and dependable prototyping and deployment of applications [3]. The tool also provides means to monitor and manage the deployed applications.

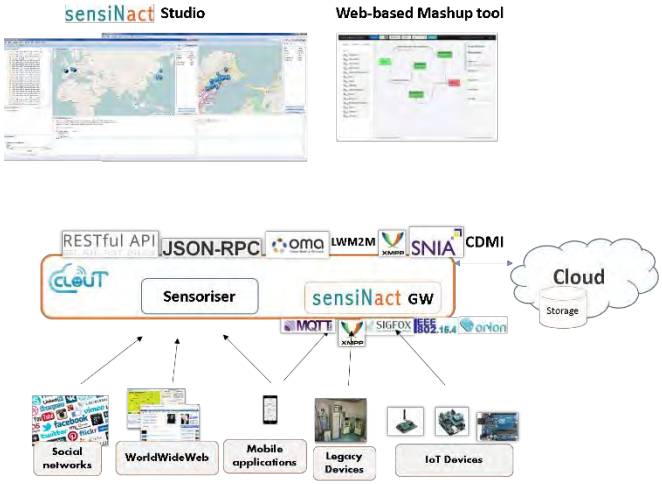



*Figure 2: ClouT smart city platform and associated tools*

The platform has been validated within several collaborative projects such as ClouT [4] and FESTIVAL [5].

## Perspectives

One of the future work is to integrate the platform within a P2P framework (such as PIAX, www.piax.org) in order to allow distributed communication among the developed services and applications, instead of the current hierarchical model. We are also currently investigating means of validating the developed applications with a models@run-time approach to empower the platform with self-management capabilities.

*Related Publications:*
[1] L. Gurgen, O. Gunalp, Y. Benazzouz, M. Gallissot, "Self-aware cyber-physical systems and applications in smart buildings and cities", International Conference on design, Automation & Test in Europe (DATE), 2013.
[2] T. Yonezawa, J.A. Galache, L. Gurgen, I. Matranga, H. Maeomichi, T. Shibuya. A Citizen-centric Approach towards Global-scale Smart City Platform.. International Conference on Recent Advances in Internet of Things (RIOT) 2015.
[3] J. Cano, E. Rutten, G. Delaval, Y. Benazzouz, L. Gurgen. ECA rules for IoT environment: a case study in safe design. Workshop on Quality Assurance for Self-adaptive, Self-organising Systems (QA4SASO); In conjunction with the 8th IEEE Int. Conference on Self-Adaptive and Self-Organizing Systems. Sep 2014
[4] ClouT: Cloud of Things for empowering citizen clout in smart cities. http://clout-project.eu
[5] FESTIVAL: FEderated interoperable SmarT ICT services deVelopment And testing pLatform. www.festival-project.eu

# Power management of a set of sensors at application level

*Research topic: Wireless Sensor Network, Energy Management, Model Predictive Control, Hybrid Dynamical System*

**Authors: Mokrenko, O.; Vergara Gallego, M.; Lesecq, S.; Albea, C. (LAAS); Zaccarian, L. (LAAS)**

**Abstract: Energy is a key resource in Wireless Sensor Networks (WSNs), especially when sensor nodes are powered by batteries. Here, we investigate how to save energy for the whole WSN, at the application level, thanks to control strategies, in real time and in a dynamic way. The first strategy investigated is based on Model Predictive Control (MPC), aiming to reduce the energy consumption of the set of sensor nodes while ensuring a given service, for the sensor network. The second strategy is based on a Hybrid Dynamical System (HDS) approach. This choice is motivated by the hybrid inherent nature of the WSN system when energy management is considered.**

## Context and Challenges

Lifespan is an important metric in assessing the performance of a Wireless Sensor Network. Indeed, in a constrained environment, any limited resource has to be taken into account. The lifespan of the network is related to the energy consumption of the sensor nodes. Thus, lifespan increase implies a better power management of the nodes at the sensor node level but also at the application level, the application being built on top of the sensor network (e.g. monitoring, control application).

## Main Results

In this work, two strategies are proposed to minimize the energy consumption of the set of nodes at application level. These strategies are based respectively on Model Predictive Control (MPC) [1] and on a Hybrid Dynamical System (HDS) approach [3]. The strategies are first evaluated and compared in simulation on a realistic test-case. Then, they are implemented on a real test-bench [1, 2].

Simulation results obtained with both control strategies are shown on Fig.1. As can be seen, the mission lifespan is increased up to +8.6% using the MPS strategy when compared to the HDS strategy. The basic control corresponds to the situation when just enough nodes are switched on to fulfil the mission at any time.

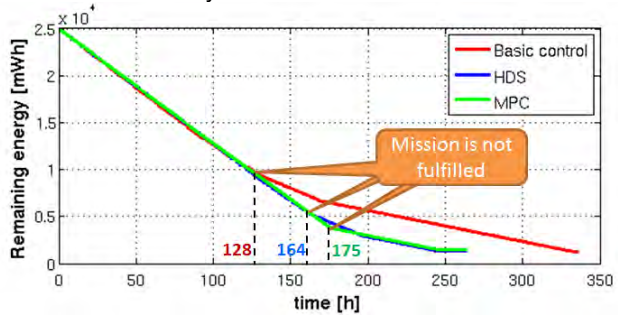



*Figure 1: Total remaining energy of the set of nodes*

A first analysis of the complexity of the control strategies has been conducted. The results obtained in simulation are summarized in Table 1. Here we report the simulation time for one control sampling time and for different implementations of the control strategies, in the Matlab environment. MPC-MIQP means that the MPC problem has to minimize a quadratic cost with binary and real variables while MPC-MILP corresponds to an optimization problem expressed with a linear cost. As can be seen, the HDS strategy offers a nice solution, without any optimization problem to be solved, its implementation requiring very simple equations computation. However, when the number of nodes increases HDS is no more interesting because of its combinatory explosion.

*Table 1: Computational time- for one step of control law evaluation for MPC and HDS strategies*

| Strategy | 6 nodes [sec] | 18 nodes [sec] | 54 nodes [sec] |
|---|---|---|---|
| MPC-MIQP | 0,36 | 5,13 | 10,81 |
| MPC-MILP | 0,051 | 0,052 | 0,053 |
| HDS | 0,004 | 0,043 | 0,375 |

A dynamic mission has been implemented, in which the number of sensor nodes to be switched on evolves over times. This shows that the proposed strategies can cope with adaptive context, this dynamic mission fitting with the changes in the test-case environment needs.

Both energy management strategies have been successfully implemented [1][2] on a real test case using the LINC middleware in order to collect temperature and humidity measurements in a large office. It has been proved that, in real-life conditions, the WSN lifespan can be extended (the mission being fulfilled) when compared to the basic control strategy. During the experiments, the radio link quality is not monitored (nor controlled). It is known that this radio link quality strongly influences the WSN lifespan results and, consequently, the comparison of both control strategies must be conducted carefully, due to the different radio perturbation profiles that are in practice not known. Despite this aspects, both control strategies show very promising results.

## Perspectives

Extra real-life experiments must be conducted to evaluate the efficiency of the proposed strategies, increase the number of nodes, and properly evaluate their computational cost.

*Related Publications:*

[1] Mokrenko, O.; VergaraGallego, M.; Lombardi, W.; Lesecq, S.; Puschini, D.; Albea, C. "Design and Implementation of a Predictive Control Strategy for PowerManagement of a Wireless Sensor Network." European Control Conference ECC 2015, Linz, Autriche.

[2] Mokrenko, O.; Vergara-Gallego, M.-I.; Lombardi, W.; Lesecq, S., Albea, C. "WSN power management with battery capacity estimation." 13th IEEE International NEW Circuits and Systems Conference NEWCAS 2015, Grenoble, France.

[3] Mokrenko, O.; Albea, C.; Zaccarian, L.; Lesecq, L. "Feedback scheduling of sensor network activity using a hybrid dynamical systems approach.". Conference on Decision and Control CDC 2015, Osaka, Japan.

# Cyber-Physical System and Contract-Based Design:A Three Dimensional View

*Research topic: cyber-physical systems, contract-based design, 3D simulation, mixed-criticality, railway*

**Authors: D. Cancila, H. Zaatiti, Roberto Passerone (University of Trento)**

**Abstract: The main goal of this work is to confront us with realistic mixed-critical smart CPS systems, using the railway domain and autonomous trains as a case study. We use contract-based design to properly deal with the integration and composition of heterogeneous components, where safety aspects require special attention. The main scientific and technical results concern the implementation of contract-based design in a 3D tool. Finally, we discuss the teaching methodology underlying the internship and the competences required to address the design of a (critical) CPS by the new generation of students**

## Context and Challenges

One important and characteristic feature of CPS is the property of autonomy, i.e., the system's ability \of being sufficiently independent in controlling its own structural and behavioral properties" . Autonomy involves intelligent CPS able to dynamically change their behavior. This property has an impact in satisfying safety-related requirements.
These, in turn, play a crucial and leading role in critical CPS, where the design and the implementation must comply to strict safety norms.
This work addresses a feasibility study in a railway system and on preexisting scientific results. To improve the design of critical CPS, we adopt Contract-Based Design (CBD). Several European projects highlight the added-value of contracts to address non-functional properties during the system design and, more recently, to address modular certification and safety-related issues. Complex systems require collaborative engineering between teams Contracts are a means to structure such communications.

## Main Results

We address the Communication Based Train Control system (CBTC) which is a distinctive example of CPS.
The high level architecture of the CBTC system comprises the Automatic Train Protection (ATP) and the Automatic Train Operation (ATO) sub-systems. Both systems have on-board and wayside equipment, which encompass software and physical systems. We model this architecture using information which exploits the IEEE 1474 standard. We specify two models: the first is concerned with the static block principle of train movement while the second is based on the moving block principle, a more modern version of the CBTC functionality. The latter uses high-resolution train location determination and provides better rail capacity.

In order to better visualize the behavior of the system and their respective impact on contracts, we construct a 3D model of a train station of an automatic metro, where the platform and the train doors have to align and open synchronously. We focus on two CBTC functionality: the passenger exchange function of the ATO and the train speed supervision of the ATP. We then define an animated scenario describing the train coming into the station, the automatic train doors and platform doors opening, passengers being exchanged, the same doors closing and finally the train departing from the station. This scenario, simple to describe, involves complex safety procedures that are not visible to the eye. Mixed criticality resides in, but is not limited to, the communication between ATP and ATO which holds different safety integrity level.

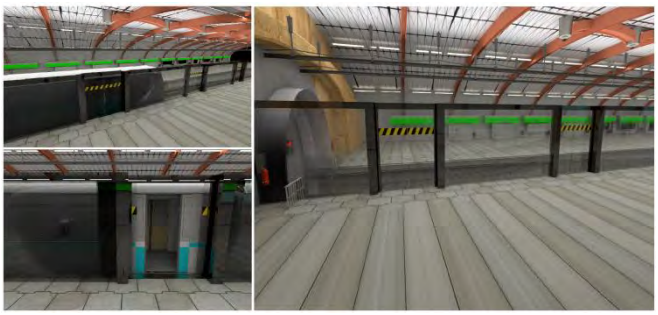


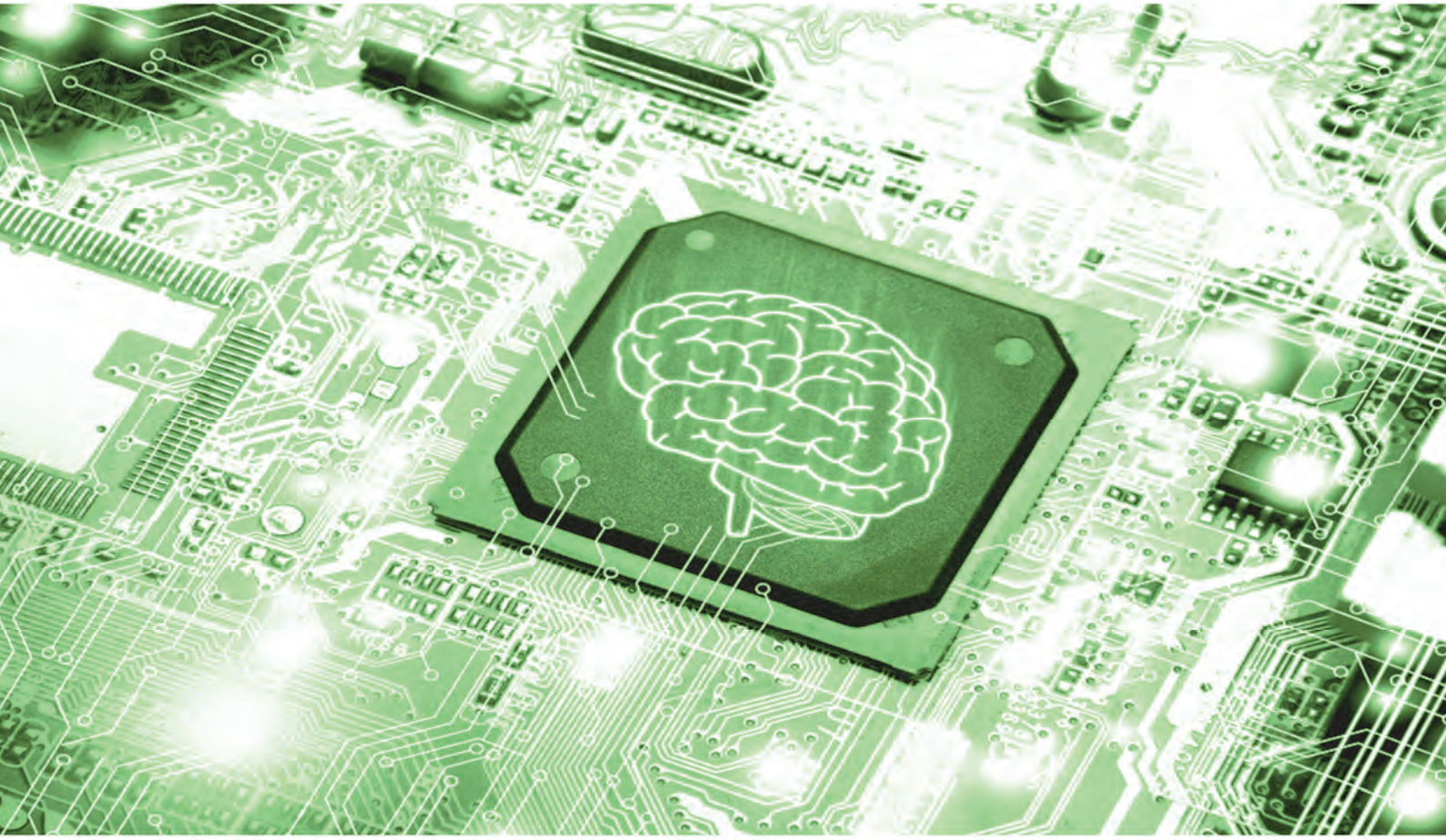
*Figure 1 : images extracted from the animation of an automatic metro*

We implement an interface for temporal constraints verification in an animated 3D model. Using a similar GUI, we construct contracts over the model using the pre-defined temporal constraint(s). Finally, we automatically verify the system meets the contracts.

## Perspectives

This work highlights the importance of using 3D scenarios to deal with CPS, and introduces contract-based design in a 3D scenario – thus improving the design of critical CPS. The industrial and scientific international feedbacks are positive encouraging.

*Related Publications:*
[1] D. Cancila, E. Soubiran, R. Passerone. Feasibility study in the use of contract-based approaches to deal with safety-related properties in CPS. Ada User Journal, 35(4):272-277, December 2014.
[2] A. Sangiovanni-Vincentelli, W. Damm, and R. Passerone. Taming Dr. Frankenstein: Contract-based design for cyber-physical systems.European Journal of Control, 18(3):217{238, 2012.



# 04

# SENSORS: ENERGY, SENSORS & DIAGNOSIS

- NEMS interfaces
- Imagers
- Multi-sensor fusion & navigation
- Accurate actuation
- Power conversion
- Diagnosis

# Heterodyne oscillator architectures for mass sensing applications

*Research topic: Nano Electro Mechanical System (NEMS), Resonator, Oscillator, Co-integration, Mass Sensing*

**Authors: G. Gourlat, M. Sansa, P. Villard, G. Sicard**

**Abstract: The full integration of NEMS sensors with their readout electronic is the key toward the design of a dense array capable of detecting small analytes adsorbed at the sensor surface. This co-integration will allow to overcome today's limitations in terms of speed and capture efficiency for mass sensing applications with NEMS resonators. This work aims to design a readout architecture capable of reading thousands of sensors without sacrificing integration density or matrix reading rate.**

## Context and Challenges

Nano Electro Mechanical Systems (NEMS) constitute a promising solution for mass sensing application, which requires very high capture efficiency of the analytes, only achievable by the increase of the sensing area brought by the co-integration of sensors arrays with readout circuitry. Two architectures have been reported in the literature. The homodyne self-oscillating (SOL) scheme is the most compact, but is very sensitive to parasitic coupling and is not able to handle multimode operations that are necessary to deduce the mass of an individual particle landing on an unknown position of the NEMS. On the other hand, the Phase Lock Loop (PLL) handles multimode resonators with robustness *vis-à-vis* the parasitic capacitances but it implies power consuming and bulky circuitry, which does not scale favorably for large arrays of sensors. In this context, we present two improved oscillator architectures that overcome the actual SOL limitations.

## Main Results

This year's work demonstrates for the first time the multimode operation of an heterodyne NEMS self-oscillator scheme (Fig. 1) [1] built around a symmetric doubly clamped resonator with no frequency stability degradation of the resonator (Fig. 2).

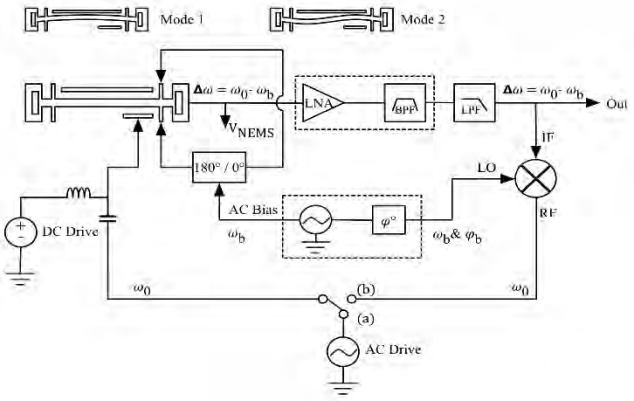



*Figure 1: Diagram of the oscillator setup used on the first and second mode of the silicon resonator. Switch position (a) and (b) corresponds respectively to the open loop and closed loop scheme.*

The second architecture [2] developed replaces the mixer from the first scheme by an adder and uses the electrostatic force quadratic component to mix the two signals. When implemented on-chip, these readout schemes will allow the reduction of the size and power consumption of readout CMOS circuitry.

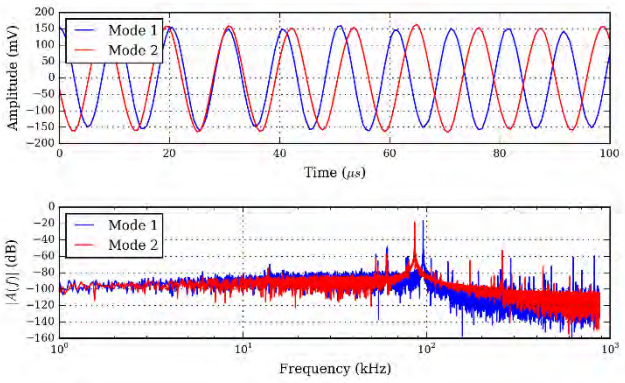



*Figure 2: Oscilloscope acquisition (a) and FFT (b) of the low intermediate frequency signal after amplification and filtering in closed loop for mode 1 in blue and mode 2 in red.*

Those topologies are very suitable for multi-mode operation and CMOS integration. The heterodyne set-up allows a control over the feedback signal frequency independently of the frequency of the resonance mode making this setup very resilient towards NEMS process variations.
Reduced silicon footprint and power consumption, compatibles with large sensor arrays required for mass sensing applications, are achievable.

## Perspectives

We believe that the 3D co-integration of these heterodyne oscillator's architectures in CMOS process with the NEMS resonators will lead to compact sensors that fit the speed and resolution requirements of mass sensing applications. This co-integration work is already underway and we expect the first 3D co-integrated sample to be available in 2016 for testing.

*Related Publications:*
[1] Gourlat, G., M. Sansa, G. Jourdan, P. Villard, G. Sicard, and S. Hentz. "Dual-Mode NEMS Self-Oscillator for Mass Sensing." In Frequency Control Symposium the European Frequency and Time Forum (FCS), 2015 Joint Conference of the IEEE International, 222–25, 2015.
[2] Sansa, M., G. Gourlat, G. Jourdan, P. Villard, G. Sicard, and S. Hentz. "Compact Heterodyne NEMS Oscillator for Sensing Applications." In Solid State Device Research Conference (ESSDERC), 2015 45th European, 146–48, 2015...

# Incremental sigma–delta ADC for compressive sensing based image sensors

*Research topic: Image sensors, compressive sensing, analog to information conversion, sigma-delta*

**Authors: W. Guicquero, A. Verdant, A. Dupret**

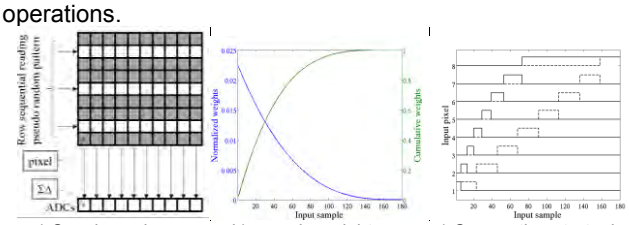
**Abstract: ΣΔ ADCs show promising perspectives in the field of compressive sensing (CS). A major limitation related to first-order ΣΔ is its high oversampling ratio (OSR) that is required for a certain bit resolution. An extension to high-order incremental ΣΔ for the implementation of a dedicated image sensor is presented. By extending the concept of ΣΔ CS ADC, optimal working points can be reached between the OSR and the compression ratio. Simulations of a particular fourth order ADC combined with the column-based Bernoulli CS scheme show a severe relaxation on the ADC master clock.**

## Context and Challenges

The compressive sensing (CS) has been intensively studied from a theoretical point of view. On the practical implementation side, the design of CS-dedicated image sensors has been investigated because of leading to low-power embedded systems, with relaxed constraints. For the moment, one of the most relevant and promising CS-dedicated architecture showing accurate experimental results relies on ΣΔ ADCs used in combination with a pseudo-random multiplexer at the end of column circuitry of the imager. It allows the use of a CS acquisition without modifying already optimized 4T pixels thanks to the intrinsic nature of the incremental 1st-order ΣΔ conversion performing the summation/averaging operation. The major limitation of this is related to the fact that such a ΣΔ ADC requires a high oversampling ratio (OSR) to ensure a proper conversion (i.e. with a sufficient number of quantization bits).

## Main Results

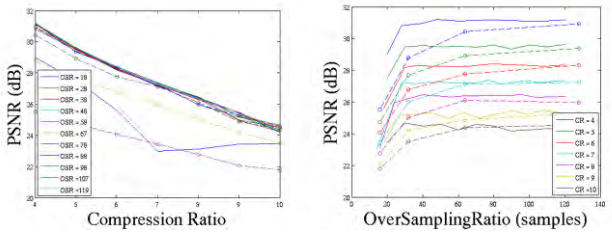
For CS purposes, a large variety of sensing schemes can be found in the literature. In the specific case of a CMOS image sensor, due to intrinsic design limitations, “parallel CS” is generally preferred. This “parallel” term means that the same set of measurement vectors is applied independently to each sub support of the signal. Regarding 2D signals such as images, two sensing schemes are relevant in that sense, corresponding either to block-based CS or line-based CS. The row-based CS scheme is presented in Fig1 a). This sensing scheme facilitates both the acquisition and the reconstruction implementations thanks to the parallelization of the unitary operations.




*a) Sensing scheme* *b) sample weights* *c) Summation strategies*

*Figure 1: Model of the compressive sensing scheme performed by the imager.*

The first order ΣΔ ADCs are highly limited due to their high OSR. In the case of the 4th-order ΣΔ ADC we consider, the quantized output after the digital filtering no more corresponds to an averaging (DC value) of the input but to a weighted sum of the successive inputs. Those weight values are reported in Fig. 1b). Two strategies regarding the number of successive samples for each pixel involved in a single ΣΔ CS measurement (ie. one conversion) have been considered. The strategy based on equaling the weights (see plain curve Fig1 c)) associated to each pixel input at the ΣΔ inputs has shown the best efficiency in terms of CS.




*a) PSNR in function of the CR* *b) PSNR in function of the OSR*

*Figure 2: PSNR results for an approximately equal weighting of the pixel inputs during each Analog to Digital Conversion.*

Fig 2 presents the results in terms of PSNR of the reconstructed image depending on the compression ratio (CR) and the OSR. The results for the 4th order ΣΔ (plain curves) clearly outperform the ones obtained for a basic 1st-order (dashed curves).

## Perspectives

We demonstrate the efficiency of using a high-order ΣΔ instead of a first order for specific CS applications. A sensing scheme using this ADC for CMOS image sensor shows great results compared to state-of-the-art techniques in terms of reconstruction quality. The proposed sensing scheme allows to read each pixel once to alleviate multiple reading issues (the same pixel read multiple times during a single frame acquisition, leading to multiple integration times). In particular, compared with a first-order ΣΔ, the proposed high order allows an OSR of 32 instead of 256 for a CR of 8. Future work will consist in optimizing reconstruction method to take advantage of all the binary information that can be contained in the bit data flow at the output of the ΣΔ modulation, before the digital filtering operation.

*Related Publications:*

[1] W. Guicquero, A. Verdant, A. Dupret, "A high order incremental Sigma Delta for compressive sensing and its application to image sensors ", IEEE Electronic Letters, 2015..

# Burst CMOS Image Sensor with on-chip A/D Conversion

*Research topic: High speed image sensor, burst video, digital burst, 3D integration*

**Authors: R. Bonnard, J. Segura Puchades, F. Guellec, W. Uhring (ICube, Strasbourg)**

**Abstract: This PhD work aims to study the inflows of the 3D integration technology to ultra-high speed CMOS imaging. The acquisition speed range considered here is between one million to one billion images per second. However above ten thousand images per second, classical image sensor architectures are limited by the data bandwidth of the output buffers. To reach higher acquisition frequencies, a burst architecture is used where a set of about one hundred images are acquired and stored on-chip (Fig. 1). Thanks to 3D technology, this work proposes to evaluate performances of a digital storage burst images sensors. A test chip has been fabricated and tested and a full 3D demonstrator is under fabrication**

## Context and Challenges

3D integration technologies are considered as a complementary solution to the technological improvements of the devices. In our case, integrated circuits are stacked on the top of each other (3D-SIC). The interconnection density between the circuits is high enough to enable interconnections at the pixel level. The 3D integration offers some significant advantages: Mainly, it allows deporting the readout electronic below the pixel and increases the fill factor of the pixel while offering a wide area to the signal processing circuit. For burst imaging, this technology provides more room to the images memory while staying close to the pixel. It also allows implementing analog to digital converter (ADC) on-chip.

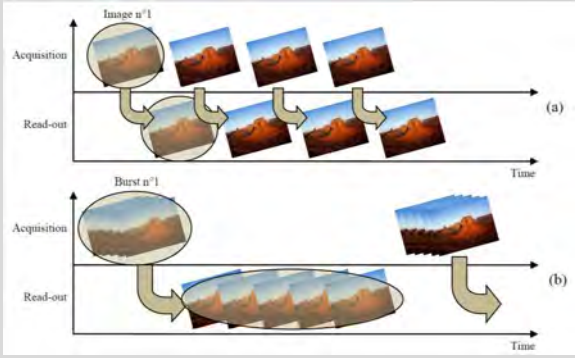



*Figure 1: Timing diagram for (a) a classical acquisition of 5 images and (b) a burst acquisition of a burst of 5 images*

## Main Results

First, we have proposed a model to assess the performances of burst sensors with analog or digital storage. The first one stores images into analog memory and then performs the digital conversion during the memory reading. This architecture enables very high frame rates up to billion images per second. However, the analog storage has some drawbacks as sampling noise and limited retention time. The analog storage architecture is able to store basically hundreds of images. The second architecture converts images into digital data before their storage. The frame rate is then limited by the ADC to tens millions images per second. Using digital memories allows storing thousands of images at each acquisition. This architecture provides an improvement of the memory depth of a factor ten compared to the analog storage architecture.

During the performance study, it has appeared that the power consumption of this 3D sensor is very high. To assess the risk of overheating, a thermal model of the sensor has been made. It has confirmed the feasibility of this structure for the acquisition of a single burst of images and has defined some operating limits to the multi burst acquisition mode [3].

A test chip circuit has been designed which enables global shutter acquisition (Fig. 2). It includes several pixels and readout architectures to optimize the sensitivity and the power consumption: A pixel with a direct injection circuit and a pixel without current source have been implemented [2]. Electrical and optical tests have been carried out. The speed both of these pixels has been evaluated thanks to electronic aperture measurements. A frame rate of 1.6 and 5 million images per second have been achieved.

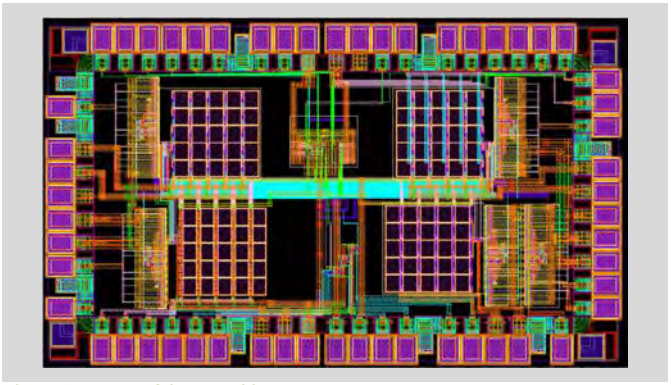


*Figure 2: Layout of the test-chip*

## Perspectives

This year, a 3D digital storage burst image sensor prototype has been designed. This image sensor is made of a stack of two pixel arrays. The top tier pixel contains the photodiode, the global shutter circuit and the comparator of a single slope ADC. On the bottom tier pixel, the 8 bit counter of the converter and the digital memory are implemented. This circuit should be tested end of 2016.

*Related Publications:*
[1] R. Bonnard, F. Guellec, J. Segura, A. Dupret, W. Uhring ; “New 3D-integrated burst image sensor architectures with in-situ A/D conversion”, DASIP conference, pages 215-222, ECSI (Eds.), IEEE, Cagliari, Italy, October, 2013.
[2] R. Bonnard, J. Segura Puchades, F. Guellec, W. Uhring ; “Signal conditioning circuits for 3D-integrated burst image sensors with on-chip A/D conversion”, Electronic Imaging Conference, San Francisco, USA, February 2015.
[3] R. Bonnard, M. Garci, J. Kammerer, W. Uhring ; “Electrothermal Analysis of 3D Integrated Ultra-fast Image Sensor With Digital Frame Storage”, Therminic Conference, Paris, France, October, 2015.

# A 0.4 e-rms Temporal Readout Noise 7.5 µm Pitch and a 66% Fill Factor Pixel for Low Light CMOS Image Sensors

*Research topic: Image sensor, Low-noise, CMOS*

**Authors: Assim Boukhayma, Arnaud Peizerat and Christian Enz (EPFL)**

**Abstract: This paper explores a new way to reduce the readout noise for CMOS image sensors by using a typical 4T pixel embedding a PMOS source follower with reduced oxide thickness and gate dimensions. This approach is confirmed by a test chip designed in a 180 nm CIS CMOS process, and embedding small arrays of the proposed new pixels together with state-of-the-art 4T pixels for comparison. The new pixels feature a pitch of 7.5 µm and a fill factor of 66%. A 0.4 e-rms input-referred noise and a 185 µV/e- conversion gain are obtained. Compared to state-of-the-art pixels, also present onto the test chip, the RMS noise is divided by more than 2 and the conversion gain is multiplied by 2.2.**

## Context and Challenges

It has been shown that buried channel transistors feature lower 1/f noise compared to surface channel transistors. This is due to the fact that the 1/f noise is a result of the process of trapping and de-trapping at the silicon silicon-oxide interface. It is also known that the gate oxide thickness reduction comes with a higher electrical field density and hence a better control of the gate over the channel. Thus, gate oxide reduction is also expected to reduce 1/f noise. In order to combine the two noise reduction mechanisms, this work presents a new low noise CIS pixel based on a PMOS source follower transistor with a reduced gate oxide thickness.

## Main Results

In order to evaluate the impact of the proposed noise reduction technique, the classical NMOS source follower reference pixel embedding a source follower dedicated for low light CIS and already optimized for low noise by the foundry, has been integrated in the same chip together with the new proposed pixel. The 1×5 mm2 test chip is designed in a 180 nm CIS process. The pixels are made of 4 transistors and a standard pinned photo diode. The reference pixel features a pitch of 6.5 µm. The chip includes a total of 24 pixels of new and reference pixels.

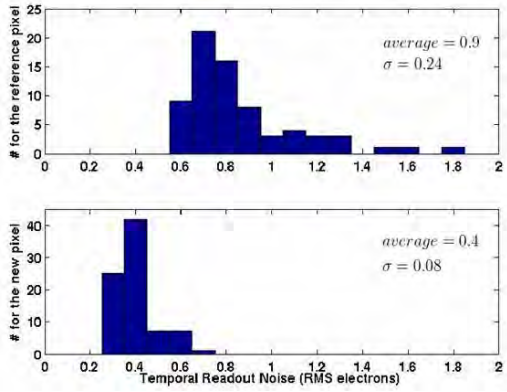



*Figure 1 : Histograms of input referred noise of readout chains based on the reference pixel and the new pixels for a column gain of 64*

Each pixel is surrounded by 8 dummy pixels for proper characterization and each pixel is connected to its own column amplifier offering an adjustable gain between 8 and 64 in order to verify the impact of column gain on noise. The column amplifiers also limit the bandwidth of the readout chain to 265 kHz.

*Table 1 : comparative table*

| Technique | Noise [e-rms] | Conversion gain [µV/e-] | Pixel pitch [µm] | Fill factor [%] | Reference |
|---|---|---|---|---|---|
| PMOS common source pixel amplifier | 0.86 | 300 | 11 | 50 | [2] |
| Buried channel NMOS source follower, column gain and CMS | 0.7 | 45 | 10 | 33 | [1] |
| **Optimized PMOS source follower and column amplification** | **0.4** | **185** | **7.5** | **66** | **This work** |

As expected, the reference pixel used for comparison and already optimized by the foundry for low noise shows a state-of-the-art performance. But the new proposed pixel features a very promising result of 50% noise reduction and 2.2 times higher conversion gain compared to the reference pixel integrated on the same chip for the same 1.5 µA pixel bias current.

The test results presented in this paper are summarized in Table I and compared to other sub-electron rms pixels integrated in similar 180 nm CIS processes: Chen et al., ISSCC 2012 [1], Lotto et al., ISSCC 2011 [2]. The new proposed pixel offers significantly lower noise with a smaller pixel pitch and higher fill factor.

## Perspectives

To confirm this new state of the art noise figure with a full imager.

*Related Publications:*

[1] Boukhayma, A.; Peizerat, A.; Enz, C., "Temporal Readout Noise Analysis and Reduction Techniques for Low-Light CMOS Image Sensors," in Electron Devices, IEEE Transactions on , vol.63, no.1, pp.72-78, Jan. 2016

# A 3T or 4T pixel compatible DR extension technique suitable for 3D-IC imagers : a 800x512 and 5µm pixel pitch 2D demonstrator

*Research topic: image sensor, pixel, High dynamic Range, 3D-IC technology,*

**Authors: A. Peizerat, F. Guezzi, M. Benetti, A. Dupret, Y. Blanchard (ESIEE)**

**Abstract: In this paper is presented a High Dynamic Range (HDR) extension technique that applies to an imager without modifying any of its other specifications (as speed, noise floor or pixel scheme). The technique relies on the division of the focal plane into blocks that are able to choose the integration time from a set of eleven exposures, reaching +60db extension compared to a standard CMOS imager. The exposure time of each block can also be controlled by an external frame buffer, paving the way to advanced bracketing techniques. This system has been explored with a proof of concept sensor processed in a 0.18µm CMOS.**

## Context and Challenges

In order to preserve details in both very dark and very bright parts of images, numerous imaging applications, e.g. automotive or video-surveillance, require the acquisition of High Dynamic Range (HDR) images, i.e. typically above 100dB. “Bracketing”, i.e. the successive acquisition of several Low Dynamic Range (LDR) images with different exposure times, is one way to circumvent this issue when using a conventional sensor, which has a typical DR of 60dB. In this work, our new architecture aims to add the HDR feature to an imager without any modifications in its specifications (speed, noise floor, pixel scheme, etc.). Using a classic 0.18µm process, it has been implemented in an 800x512 proof-of-concept imager with a 5µm pixel pitch and features a 120dB DR. Mapping this architecture on a 3D-IC technology, in a next step, will bring fully optimized capabilities

## Main Results

The fabricated 2D-IC architecture, as well as the targeted 3D-IC one, are illustrated Figure 1. The dark blue blocks, further referred to as “primary” image sensor blocks, are common to both architectures and could be part of any image sensor. The light blue blocks are added to provide the HDR feature to this primary image sensor. Without changing the pixel architecture (3T or 4T), the technique consists in dividing the pixel array into pixel blocks (32x32 pixel blocks for the 2D-IC architecture, 8x8 pixel blocks for the 3D-IC architecture), each having its own self adjusted integration time.

An 800x512 image sensor with 5µm pixel pitch has been fabricated using a 0.18µm 1P4M process. The image sensor can be operated in the conventional LDR mode or the HDR mode that extends the DR by +60dB. Three images of the same scene (whose DR exceeds 120dB) are given in Figure 2, the first one in the conventional LDR mode, the second one in the HDR mode and the third one after a basic FPN blind reconstruction algorithm. On this second image are superimposed the exponents of the pixel blocks. As expected these exponents span from 0 to 10 in order to avoid saturation inside a pixel block. On the third image, block artifacts appear but will be removed after a more sophisticated image processing or improving the acquisition system.

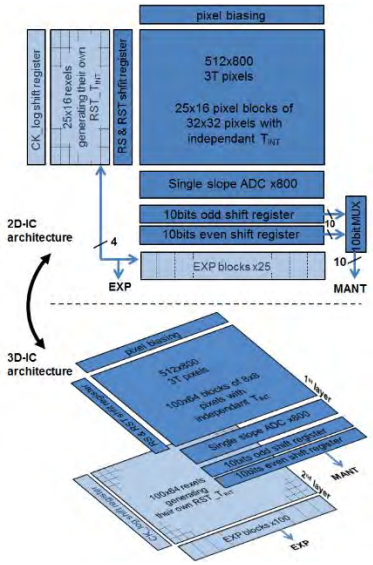



*Figure 1 : Overall block diagram of the chip for 2 different architectures*

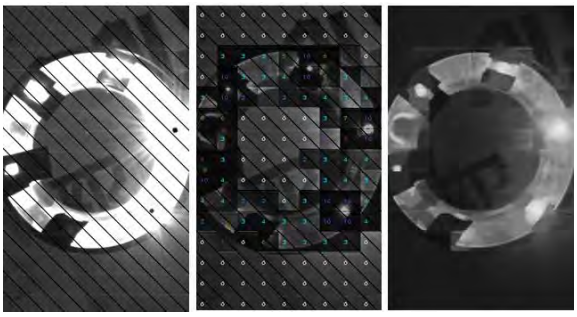


*Figure 2 : Three cropped-images of the same scene (an LED lamp).*

## Perspectives

Once implemented using a 0.18µm 3D-IC technology and a BSI imager, this architecture will easily accept pixel blocks of 8x8 pixels without the inconvenience of dead pixels inside the matrix. A more advanced technology node for the 2nd layer of a 3D-IC imager would reduce even more the block size.

*Related Publications:*

[1] Messaoud, F.G.; Peizerat, A.; Dupret, A.; Blanchard, Y., "On-chip compression for HDR image sensors," Design and Architectures for Signal and Image Processing (DASIP), 2010 Conference on , vol., no., pp.176,182, 26-28 Oct. 2010

[2] Peizerat, A.; Guezzi, F.; Benetti, M.; Dupret, A.; Jalby, R.; de Sa, L.B.; Guicquero, W.; Blanchard, Y., "A 3T or 4T pixel compatible DR extension technique suitable for 3D-IC imagers: A 800×512 and 5µm pixel pitch 2D demonstrator," in Circuits and Systems (ISCAS), 2015 IEEE International Symposium on , vol., no., pp.1094-1097, 24-27 May 2015

# A 278 GHz heterodyne receiver with on-chip antenna for THz imaging in 65 nm CMOS process

*Research topic: mmW, CMOS, THz, imaging, heterodyne receiver*

**Authors: A. Siligaris, Y. Andee, E. Mercier, J. Moron Guerra, J.-F. Lampin (IEMN), G. Ducournau , Y. Quere**

**Abstract: A low power 278-GHz CMOS zero-IF heterodyne receiver is presented. The circuit includes a passive mixer, a baseband amplifier, a 278-GHz triple push sub-harmonic injection locked oscillator and an integrated antenna. The receiver measured maximum conversion gain is -12 dB and the DC power consumption is 47 mW. The on-chip antenna size is 390x280 µm². The heterodyne receiver is used as a THz detector for imaging. It is shown that thanks to the heterodyne structure and the oscillator locking the THz image quality and contrast increases significantly. The imaging system achieves a noise equivalent power (NEP) of 0.2 fW/Hz.**

## Context and Challenges

There is a growing interest in THz and Sub-THz frequency bands (200 GHz to 3 THz). The main reasons for this are the potential industrial applications of those bands: astronomy, short range radar, wireless high data rate communication and THz imaging. In this work we explore the advantages of the heterodyne structure in CMOS. Indeed, heterodyne detection offers the possibility of phase information and a significantly higher sensitivity for the amplitude detection. For direct detection, the typical noise equivalent power (NEP) is in the order of magnitude of a few hundreds of $pW/\sqrt{Hz}$, while the heterodyne receivers offer equivalent noise power lower than 1 *fW/Hz*. In this work, a fully CMOS 65nm integrated low power Sub-THz (278 GHz) heterodyne detector with on-chip local oscillator and on-chip antenna is presented. The particularity is that the oscillator is able to lock on an external lower frequency signal and copy the phase and frequency characteristics (sub-harmonic injection locked oscillator). It is shown how the local oscillator frequency stability and the phase noise impact on the image quality.

## Main Results

A zero-IF architecture is employed for the heterodyne receiver. The 277 GHz local oscillator uses triple-push architecture (third harmonic generation). The triple-push oscillator is modified in order to add a 46.33 GHz injection port. The receiver mixer is a passive structure (cold-FET resistive mixer) which allows mixing beyond CMOS cut-off frequencies ($f_t$, $f_{max}$). Finally, the down converted signal is amplified by a three-stage baseband amplifier and analyzed in the 10 MHz to 1 GHz base-band.

Fig. 2 shows various 278 GHz raster scan images of a wasp (*megascolia maculata*). In Fig. 2.a. the image was made with a free running oscillator. Fig. 2.b. was performed with the oscillator being locked on a 46.375-GHz injection signal. We observe that the locked oscillator image is cleaner (as expected) than the free running oscillator image. This demonstrates that a locked local oscillator increases the quality of THz imaging using heterodyne receiver. Indeed, the image dynamic range (ratio between white and black) is increased by 10 dB in Fig. 2.b. respect to 2.a. Finally, Fig. 2.c. shows the resulting photograph when the IDR is only 20 dB, whereas Fig. 2.d. shows the photograph when the image dynamic range is 50 dB (locked oscillator). We observe clearly in Fig. 2.c that many details in the wasp tissues are completely black. This reveals that high sensitivity is needed in order to increase the image contrast.

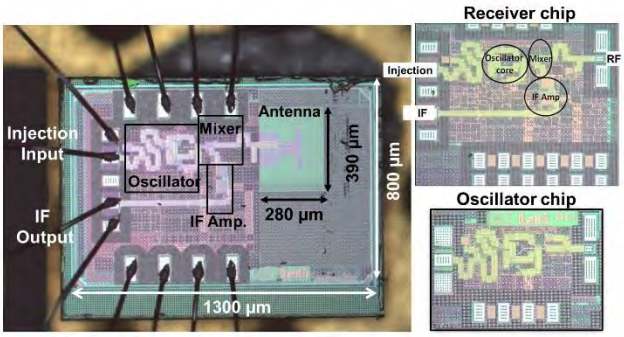



*Figure 1: Microphotographs of the heterodyne 278 GHz pixel and stand-alone chips of the receiver and 278 GHz sub-harmonic injection locked oscillator.*

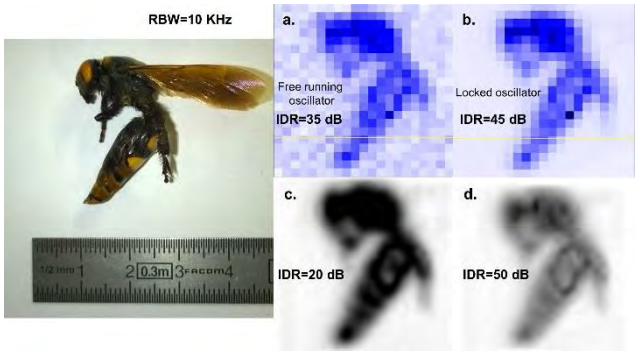



*Figure 2: 278-GHz images of a wasp. a. and b.: images without and with locked oscillator. c. and d.: impact of the sensitivity on the contrast of the image. RBW= 10 KHz. IF frequency = 250 MHz.*

## Perspectives

This work paves the way for larger application fields such as high data rate wireless communication systems in the THz band.

*Related Publications:*
[1] A. Siligaris, Y. Andee, E. Mercier, J.M. Guerra, J.-F. Lampin, G. Ducournau, Y. Quere, "A 278 GHz heterodyne receiver with on-chip antenna for THz imaging in 65 nm CMOS process," *IEEE European Solid-State Circuits Conference (ESSCIRC)*, pp.307-310, Sept. 2015.
[2] J.M. Guerra, A. Siligaris, J.-F. Lampin, F. Danneville, P. Vincent, "A 285 GHz sub-harmonic injection locked oscillator in 65nm CMOS technology," *IEEE International Microwave Theory and Techniques Symposium, MTT-S*, pp. 1-3, June 2013.
[3] J.M. Guerra, A. Siligaris, J.-F. Lampin, F. Danneville, P. Vincent, "A 283 GHz low power heterodyne receiver with on-chip local oscillator in 65 nm CMOS process," *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pp. 301-304, 2-4 June 2013.

# Embedded Multi-Sensor Fusion for Automotive Perception

*Research topic: Multi-Sensor Fusion, System Integration, Autonomous Vehicles*

**Authors: J. Mottin, D. Puschini, T. Rakotovao**

**Abstract: Safe autonomous vehicles will become reality when reliable, precise, and integrated environment perception systems will emerge. To understand their environment, autonomous vehicles rely on multiple heterogeneous sensors providing information on the distance of closest objects. All the sensory information is fused and integrated in an environment model called Occupancy Grid. Present work proposes new methods for real-time Occupancy Grid computation on industrial standard automotive circuits. Experiments were conducted on instrumented vehicles in real life traffic conditions, demonstrating real-time performance for Multi-Sensor Fusion on embedded platforms.**

## Context and Challenges

Advanced driving functionalities for autonomous navigation such as obstacle detection & avoidance or target tracking rely on deep understanding of the surrounding environment. This is traditionally achieved by mapping measurements from multiple heterogeneous range sensors in a discrete model of the space called Occupancy Grid (OG). The performance of the perception system is crucial, and is dictated by the output rates of the range sensors. The major challenge is then to integrate OG computation on industrial automotive targets, while guaranteeing real-time performance.

## Main Results

The first contribution is an architecture exploration, to identify and benchmark electronic hardware capable of executing real-time OG computation in a vehicle.

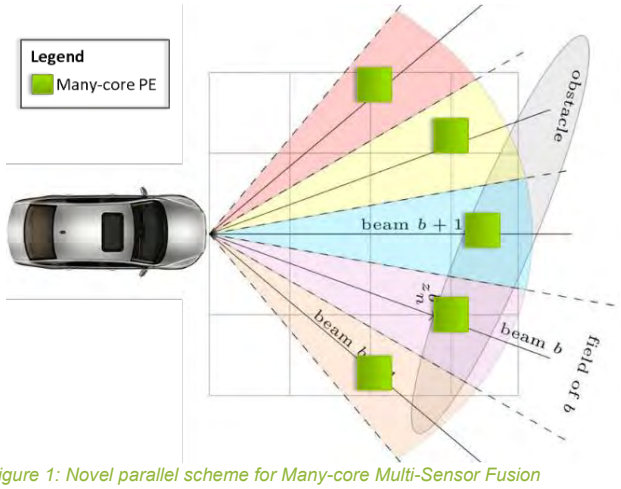



*Figure 1: Novel parallel scheme for Many-core Multi-Sensor Fusion*

A comparison among Graphical Processing Units (GPU), multi-core and many-core platforms indicates that many-core platforms could match both the OG performance requirements and the automotive constraints in terms of cost, size and power consumption [1].

The second contribution is a novel parallel computing scheme for Multi-Sensor fusion in OG. As opposed to GPU implementation where regular data parallel scheme yields best performance, our proposed “Beam-by-beam” (Fig. 1) scheme targets many-core platforms where Processing Elements (PEs) have more control flow capabilities. Compared to state-of-art methods, our method achieves 9x speed-up, yielding real-time performance in an automotive case-study [2].

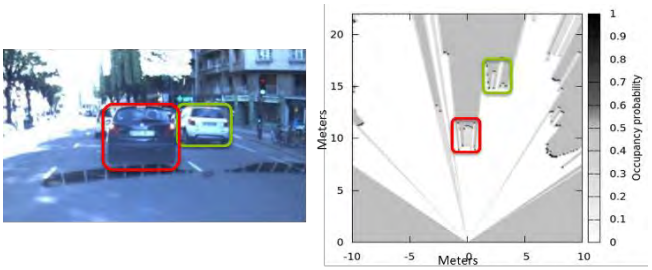



*Figure 2: Real embedded automotive integration of Multi-Sensor Fusion*

The third contribution is novel theoretical formulation of the Multi-Sensor Fusion. This formulation drastically reduces hardware requirements for real-time OG computation on industrial targets [3]. Preliminary implementation shows that it would now be feasible to run environment perception task on standard automotive circuits, such as Freescale iMX6 Quad Auto chips, and even on micro-controller platforms. Experiments were conducted in a real instrumented vehicle where OG computation is executed on a ARM Cortex-M3 micro-controller platform at a 5 Hz output rate (Fig 2.), a promising step towards real-time micro-controller perception implementation

## Perspectives

The advances described above allow now to build an industrial perception system, suitable for advanced automotive applications. Computing power required for real-time OG computation could be delivered by last generation of micro-controller platforms. Current work could also be applied to other application domains were high efficiency is critical such as wearable devices or drones.

*Related Publications:*
[1] T. Rakotovao, D. Puschini, J. Mottin, L. Rummelhard, A. Negre, C. Laugier, "Intelligent Vehicle Perception: Toward the Integration on Embedded Many-core", 6th Workshop on Parallel Programming and Run-Time Management Techniques for Many-core Architectures (PARMA-DITAM), 2015.
[2] T. Rakotovao, J. Mottin, D. Puschini, C. Laugier, "Real-Time Power-Efficient integration of Multi-Sensor Occupancy Grid on Many-Core", IEEE International Workshop on Advanced Robotics and its Social Impacts (ARSO), 2015.
[3] T. Rakotovao, J. Mottin, D. Puschini, "Perception d’obstacles : fusion efficace de capteurs pour la construction des grilles d’occupation", Groupe de Travail et Journées Automatique et Automobile (GTAA-JAA), 2015

# Exploratory Digraph Navigation

*Research topic: Autonomous planning and exploration under uncertainty, Robotics, Networks, Graph Theory*

**Authors: F. Mayran de Chamisso, L. Soulier, M. Aupetit (Qatar Computing Research Institute)**

**Abstract: Exploratory Digraph Navigation is a new paradigm whose main point is to take into account yet unknown paths and places while planning a path to some destination on an incomplete map of an environment. Indeed, yet uncharted places and paths may contain a shortcut to the destination. One of the key benefits of our approach is being able to balance the risks and benefits of exploration.**

## Context and Challenges

There are more and more autonomous and semi-autonomous vehicles surrounding us, from automatic lawn mowers and vacuum cleaners to rescue robots in disaster areas or rovers on other planets. However, these either use a map of the environment to plan their path, or explore the environment first in order to produce said map. In the same context, an animal would navigate directly with an incomplete map, exploring only areas it is interested in, especially paths that lead from its lair to its favorite food sources. An Exploratory Digraph Navigation algorithm allows precisely this. Exploratory Digraph Navigation also finds applications in video game AI and network theory.

## Main Results

We implemented the Exploratory Digraph Navigation paradigm inside the well-known path planning algorithm A*, for a new algorithm called EDNA* [1]. The paradigm can easily be adapted to other planning algorithms, especially A* variants such as theta*. A* uses an heuristic (under)estimating the remaining length to destination to guide its search for a path to destination.

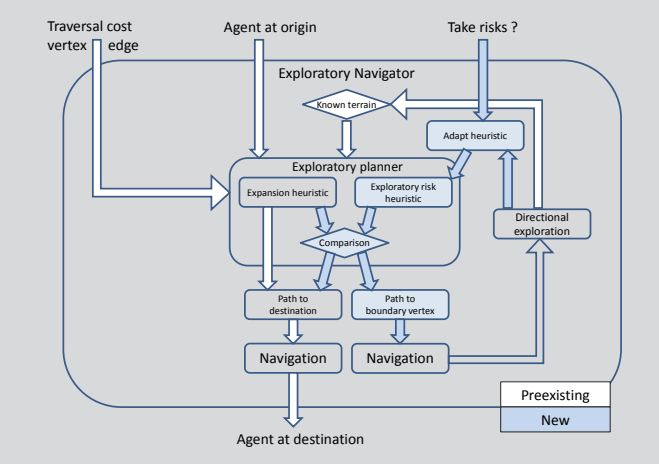



*Figure 1 : Conceptual view of EDNA**

EDNA* introduces a second heuristic, (over)estimating the length of a shortcut traversing yet uncharted space. The interaction of both heuristics (Fig. 1) expresses a balance between risk and reward.

Compared to a non-exploratory algorithm such as A*, EDNA* can lead to paths shorter (Fig. 2) by more than 50% in average and up to 24 times shorter in extreme cases (depending on the layout of obstacles and unknown places). EDNA* can also be tuned to privilege exploration of uncharted space for future traversals.

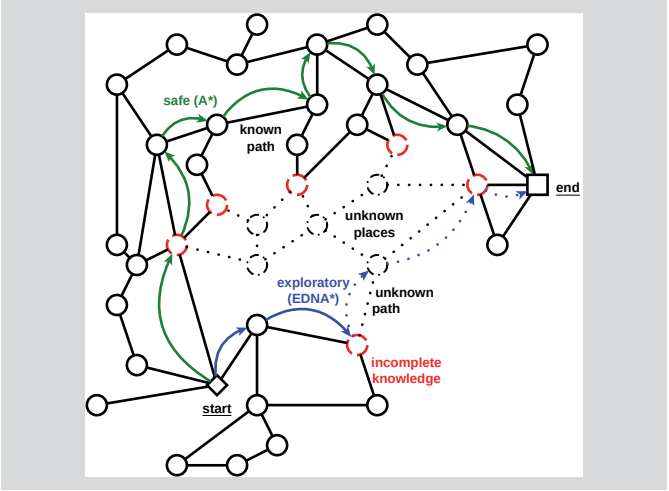



*Figure 2 : EDNA* may find shorter paths using exploration*

## Perspectives

We are currently investigating the use of Exploratory Digraph Navigation for routing purposes as well as the balance between map completeness and average path lengths, with the aim to compress the map of an environment with minimal impact on the motion of an agent using it to navigate.

*Related Publications:*
[1] F. Mayran de Chamisso, L. Soulier, M. Aupetit, “Exploratory Digraph Navigation using A*”, International Joint Conference on Artificial Intelligence (IJCAI), 2015

# Autofocus performance realization using automatic control approach

*Research topic: Imaging systems, Autofocus, Lens control*

**Authors: Zarudniev M., Alacoque L., Tonda A., Bolis S., Pouydebasque A., Jacquet F.**

**Abstract: In this work we propose a new approach for autofocus systems design that uses results from linear control theory to determine the optimal feedback gain in the sense of quadratic sharpness hypothesis. The method proposed allows to improve the existing autofocus systems that are based on the sharpness autofocus using the optimal feedback gain. It is shown that the method is generic and can be integrated into an image signal processing unit of a modern camera.**

## Context and Challenges

In modern autofocus cameras, main research directions for the sharpness based algorithms are the study of operators that transform the flow of pixels into a sharpness value, and the study of control schemes for the sharpness peak search. The sharpness operators are often searched with properties such as high signal to noise (SNR) level, noise filtering, linearity and low implementation complexity. This allows to obtain robust metrics for the maximum sharpness search algorithms. These methods use an empirical approach where the feedback gain is chosen using simulations, or adaptation of chosen parameters to some experiment results.

Recently, it was suggested to choose the feedback gain that realizes the sharpness peak search using optical and image spatial frequency decomposition models that allow to connect analytically the autofocus performance and the choice of the feedback gain. Despite the model-based approach that allows to change the feedback gain in the loop, no optimality for the gain choice is given. Moreover, this technique involves some matrix computations, to compute the feedback parameters at each iteration, leading to high power consumption.

## Main Results

We developed a new feedback gain control that uses the optimal gain value in the loop under quadratic sharpness hypothesis. This method minimizes the number of control steps, efficiently filters the image sensor noise and allows to be accurate in the peak search. The new sharpness operator minimizes the computation complexity. The autofocus design method has a strong analytic foundation for the lens step choice; it is based on the sharpness evaluation and its evolution is supposed quadratic near the maximum. From this quadratic hypothesis, one can define an analytical solution for the feedback gain $K_i = |u^*(k)-u^*(k-1)|$ that minimizes the sharpness gradient with efficient calculations of the optimal control value $u^*(k)$ (Figure 1). The simplicity of the control structure and of the sharpness operator leads to a successful implementation of the new autofocus control approach, either in software or hardware. This new approach of autofocus control transfers the noise filtering properties from the sharpness operator to the feedback control gain. Since the filtering properties are represented by single input and single output feedback control law, it is more advantageous compared to developing a new sharpness operator that uses matrix computations. This might decrease the computational time for the sharpness value evaluation as well as reduce the autofocus response time (Figure 2).

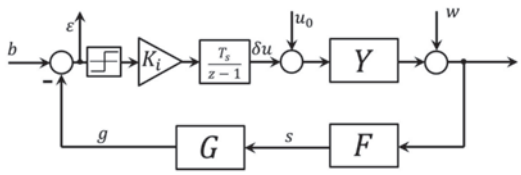



*Figure 1: Autofocus system with integral control.*

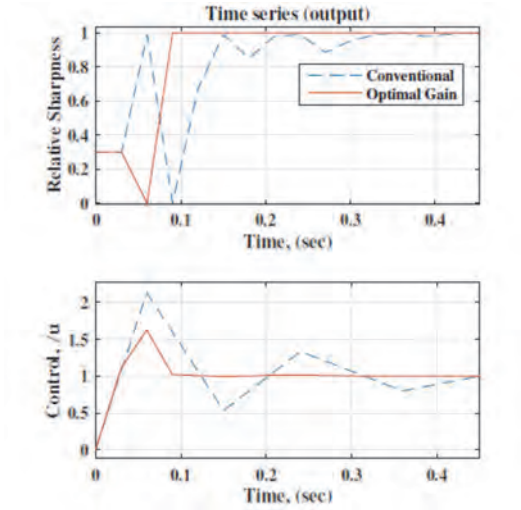



*Figure 2: Simulation of the autofocus system with conventional gradient search control and the optimal gain control*

## Perspectives

Substantial improvements of the autofocus performance can be achieved if more complex hypotheses on the sharpness variation are taken. However, the optimal gain can be calculated using a moving average that takes in account a number of preceding measurements to compute the optimal gain. This approach can be useful in order to exclude the use of optimization algorithms, being more precise when the quadratic hypothesis is not true.

*Related Publications:*
[1] Zarudniev, M., Alacoque, L., Tonda, A., Bolis, S., Pouydebasque, A., Jacquet, F., "Autofocus performance realization using automatic control approach", Conference Proceedings - 13th IEEE International NEW Circuits and Systems Conference, NEWCAS 2015, art. no. 7182058
.

# Application Specific circuits for GaN HEMT based power conversion systems

*Research topic: Gallium Nitride (GaN), Gate Drivers, Power Supplies, Electric Power Conversion*

**Authors: D. Bergogne, R. Grézaud, F. Ayel, Y. Wanderoild, C. Gillot, R. Escoffier, W. Vandendaele**

**Abstract: Energy efficient solutions have brought a specific incentive on Wide Band Gap (WBG) Power Devices. Our work implements specific circuits to control Normally-On and Normally-Off Power Devices. In this work, DC and AC current GaN Power Switches have been successfully implemented using specifically designed Gate Drivers ASICs. Future work extends to GaN integrated Power Circuits.**

## Context and Challenges

In recent years SiC and now GaN power devices have come out on the market and engineers are starting to implement them in promising high frequency, high efficiency power converters, especially for high bus voltages requiring breakdown voltages in the 600V to 1200V range. However, commercial integrated circuits are either single gate drivers without insulation or limited voltage bridge drivers, up to 100V for high temperature above 175°C. In addition, depletion Mode transistors, (Normally-On) need specific control circuits and High operating temperatures are required more and more often. These elements are the motivation to study and develop Gate Drivers with their insulated power supply for GaN transistors and devices.

## Main Results

The main result this year has been the demonstration of a GaN AC switch produced by Leti [1], connected to a 110V AC voltage source. The Power Switch uses the Leti GaN HEMT Technology in a patented arrangement to provide a Bidirectional Device with a Single Reference electrode (SRBD). This device enables new solutions with reduced impact on material supply and energy consumption. Figure 1 shows the demonstration of the operation at a frequency of 1MHz under 250V with a switched current of 2A.

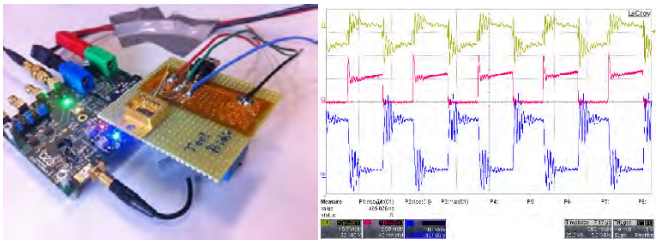


*Figure 1: Left. Experiential setup showing a GaN device from Leti under test with a specific gate driver board. .Right. Experimental verification of the behaviour of the SRBD in a High Side Switch for a DC chopper switching cell. Parameters are: 250V supply, 2A inductive load and 1MHz operating frequency. Yellow: Gate Voltage 10V/div, Magenta: Drain Current 2A/div, Blue: Drain-Source Voltage 100V/div Time base is 500ns/div*

Developing applications for GaN HEMT devices [2], led our laboratory to propose an innovative gate driver sub-function enabling to further reduce the component count and the environmental impact.

*Figure 2: 200°C motor controller with WBG power stage and ASIC gate driver.*

Another step forward was the demonstration of a High Temperature solution embedding a Gate Driver Application Specific Integrated Circuit with Wide Bad Gap (WBG) Power Transistors. The solution was successfully tested at 200°C cooling ambient temperature, see figure 2.

## Perspectives

Future work includes two complementary aspects: GaN Integrated Power circuits, and High Temperature Versatile Application Specific Controller, SOI integrated circuits for GaN power systems. These themes will require to co-design the integrated circuit with the interposers and the external passives additional functions.

Recent actions address GaN Power Integration, where GaN devices become Power Integrated circuits. Alongside, the implementation of physical models for GaN HEMT in the design flow will lead to a new technological offer with prototype Power Solutions.

Applications include for instance down hole mining, the more electrical aircraft, high temperature automotive electro-actuators…

*Related Publications:*

[1] D. Bergogne, O. Ladhari, L. Sterna, C. Gillot, R. Escoffier, W. Vandendaele, "The single reference Bi-Directional GaN HEMT AC switch" in Power Electronics and Applications (EPE'15 ECCE-Europe), 17th European Conference on ECCE, 2015.

[2] R. Grezaud, F. Ayel, N. Rouger, J.-C. Crebier, "Monolithically Integrated Voltage Level Shifter for Wide Bandgap Devices-Based Converters," IEEE conference on PRIME, 2014.

# Power Supply Integration on Chip

*Research topic: Fully integrated power supply, Granular DC-DC converter*

**Authors: G. Pillonnet, A. Quelen, P. Vivet**

**Abstract: The full integration of DC-DC converters on chip offers great promise for dramatic reduction in power consumption and number of board-level components, in complex systems on chip. Within this frame, this work compares in the same context the merits and potentialities of some converter topologies (capacitive-, inductive- and resonant-based switching converters), some widely-used CMOS nodes (28, 65 and 130nm), and the assembly of passive and active multilayers using 3D technology.**

## Context and Challenges

Fully integrated dc-dc converters are an attractive solution to supply high performance computing units on the same chip. The advantage is to provide clean, high speed and individual power supply modulation for the multicore processors without bulky external components or additional circuits. Nonetheless, the recently published on-die power converters still achieve performances far from the industrial targets in terms of power density, voltage regulation, efficiency versus conversion ratio or direct battery connection compatibility.

## Main Results

[1] provides a technology independent method and a practical use case to compare three particular, but well-spread, DC-DC converter topologies for chip-scale power management. This method aims to help system-level designers to select the passive nature of the converter, predict the achievable efficiency in hard-switching conditions, validate a CMOS technology choice, or compare some technologies. To derive the proposed method, a standard 65nm CMOS has been chosen to fully integrate the active and passive elements of the converters. Above 1W/mm$^2$ power density, the resonant-based converter appears more appropriate, mainly due to a best passive energy utilization.

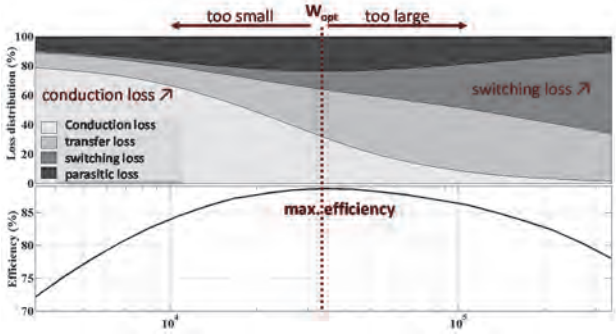



*Fig. 1. Optimal point to maximize the power efficiency*

[2] confirms that 3D technology is a relevant candidate to provide an efficient integrated power supply in a multi-core processor context. As converters do not directly benefit from scaling, the additional 3D active or passive allows higher voltage transistors or higher capacitance density which better suits the converter requirements. This solution saves the expensive die-area of digital cores, improves the silicon yield, increases the acceptable input power supply and therefore reduces the pin number for the external package.

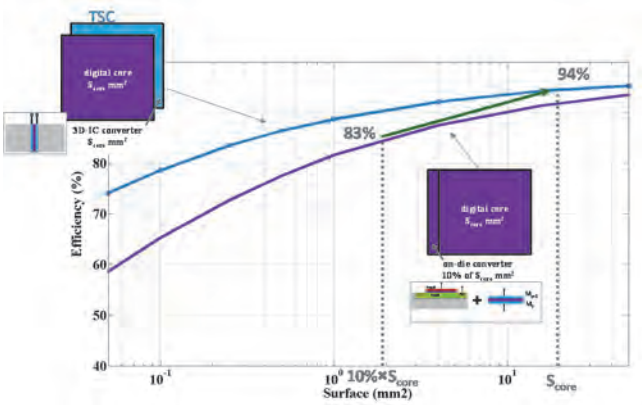



*Fig. 2. Efficiency trends vs. 3D technology choice*

[3] proposes a novel switched capacitor converter structure called MISO-CSC to achieve a more constant efficiency over a large conversion ratio. For two inputs, the MISO converter generates 18 ratios instead of 3 in SISO mode. Efficiency analysis led us to select only seven efficient ratios. The MISO converter was designed and compared to SISO topology in CMOS 65nm. In the on-die power supply multi-core processor, the MISO topology could be used to efficiently refine the DVFS with no extra cost if two power rails are available on the PCB board.

## Perspectives

Concerning the converter topology, the promising resonant-based converter seems to be relevant mainly due to the best passive energy utilization, but further research has to study its lossless controllability and its active and passive component stress management related to CMOS technology. Concerning the future trends, we believe 3D with dedicated layers for power is the promising choice rather than on-die converters. 3D design also allows multiple layers to integrate N converters in parallel to again improve the power density.

*Related Publications:*
[1] Y. Pascal, G. Pillonnet, "Efficiency Comparison of Inductor-, Capacitor- and Resonant-based Converters Fully Integrated in CMOS Technology," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2015
[2] G. Pillonnet, N. Jeanniot, P. Vivet, "3D ICs: An Opportunity for Fully Integrated Power Supply", IEEE conference on 3DIC, 2015
[3] G. Pillonnet, A. Andrieu, E. Alon, "Dual-Input Switched Capacitor Converter Suitable for Wide Voltage gain Range", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2015

# Integrating live diagnosis of complex topology wires inside the network

*Research topic: Embedded cable diagnosis, OMTDR, connector*

Authors: W. Ben Hassen, F. Auzanneau, L. Incarbone, S. Evain, F. Peres (ENIT), A. Zanchetta (Nicomatic)

**Abstract: The innovative OMTDR method (Orthogonal Multi-tone Time Domain Reflectometry) was designed for the distributed diagnosis of live complex topology electrical networks. Adding communication to reflectometry, and using a specific protocol, it enables data fusion from multiple sensors and provides a complete perspective of the network. The location of several defects can be determined with no ambiguity. OMTDR has been integrated into 2 standard aerospace connectors and tested on the harness of a mockup of a fighter aircraft. Several defects scenarios were tested, showing the added value of sensors communication : 3% to 4% location accuracy was obtained.**

## Context and Challenges

In many application domains, wires faults can have dramatic consequences. Among many methods, reflectometry has proven to be the most efficient one. Similarly to Radar, reflectometry injects a probe signal at one end of the network under test and analyses the reflected signal to locate the defects. Live wire diagnosis is often required to ensure permanent monitoring of the health of embedded cables, and complex topology networks needs the use of several diagnosis systems in parallel, each one providing a different perspective of the network, thus eliminating location ambiguities.

The new OMTDR method enables a precise sharing of the frequency spectrum between the native users of the network and several reflectometry sensors, for a harmless distributed diagnosis. Adding communication between sensors inside the probe signal, it enables to implement a complete diagnosis strategy for the accurate location of all the defects in the network.

## Main Results

OMTDR applies the principles of OFDM (Orthogonal Frequency Division Multiplexing) to wired network diagnosis [1], dividing the bandwidth into multiple sub-bands using orthogonal and overlapped subcarriers, thus maximizing the spectral efficiency and the spectrum control.

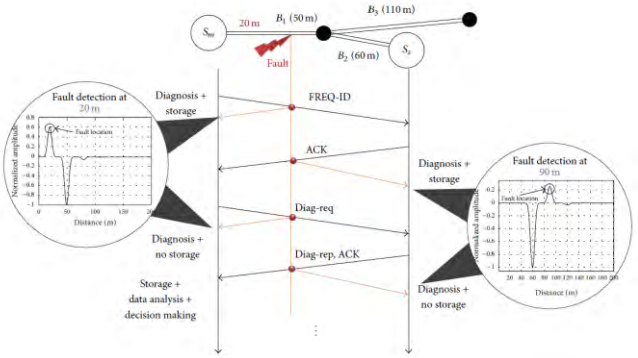



*Figure 1: Communication protocol for online diagnosis*

Sensor communication enables to deploy a specific data fusion strategy, based on a master / slaves protocol (Fig. 1), for the unambiguous location of several defects in a complex topology network (such as a communication bus). The cooperation of several sensors enables to detect a defect event when it is out of reach of the master sensor.

OMTDR has been tested on the harness of a mockup of a fighter aircraft [2]. OMTDR was integrated into the harness, embedding both real time distributed wire diagnosis and sensor communication in 2 standard aerospace connectors (Fig. 2).

*Figure 2: Miniaturized OMTDR systems were realized and integrated into 2 standard aerospace connectors (DMM and 38999)*

Several defects scenarios were tested, showing the added value of sensors communication. Even for a simple network, OMTDR obtains a better location accuracy (4 to 7 cm) by combining different sensor data.

## Perspectives

The innovative OMTDR diagnosis method has proven the possibility of integrating the diagnosis into the network without interfering with the operating signal and building a complete perspective of the network's status. A Graphical User Interface (GUI) running on an Android device shows the diagnosis results, the communication and analysis status. The Smart Connector's embedded diagnosis capability also permits location and detection of an intermittent fault and soft faults (such as shafing, cable flexion, etc.)

*Related Publications:*

[1] Ben Hassen, W.; Auzanneau, F.; Incarbone L.; Peres, F. & Tchangani A, 'Distributed Sensor Fusion for Wire Fault Location Using Sensor Clustering Strategy', International Journal of Distributed Sensor Networks, 2015. (http://dx.doi.org/10.1155/2015/538643)

[2] Incarbone, L.; Auzanneau, F.; Bonhomme, Y.; Ben Hassen, W.; Dupret, A.; Evain, S.; Morel, F.; Gabet, R.; Solange L. & Zanchetta, A, 'OMTDR Based Integrated Cable Health Monitoring System SmartCo: An embedded reflectometry system to ensure harness auto-test', IEEE Conference on Industrial Electronics and Applications (ICIEA 2015)

# Soft defects localization by signature magnification with selective windowing

*Research topic: Soft defect, detection, time domain reflectometry, cable diagnosis, selective amplification*

**Authors: S.Sallem (WIN-MS), N. Ravot**

**Abstract: The SMSW (Signature Magnification by Selective Windowing) method, based on a temporal processing treat the reflectogram so as to make soft defect signatures detectable. The algorithm first performs a localization of critical points of the reflectogram (zero-crossing or mean-crossing). Then, it selects points having enough energy in both sides (above threshold). This threshold is to be determined statistically or set depending on the application. Once the defect area is windowed, we proceed to the magnification procedure which amplify the defect signature while reducing the noise level on the reflectogram. The proposed method shows very convincing results both in simulation and experiments and for different types of cable and transmission line.**

## Context and Challenges

Cables are present in every systems and their role is fundamental, from energy supply to information transmission. However their maintenance is often neglected and systems for quick and accurate diagnosis are required. In terms of wire diagnosis and SHM (Structural Health Monitoring), methods based on reflectometry are among the most used. They show great performance in detecting hard defects, such as short circuits and open circuits. These significant defects may have serious consequences economically and materially, it is necessary to be able to detect them early, before they arise (called soft defects).

These soft defects do not prevent the signal propagation and are difficult to detect. Unfortunately, no current method is powerful enough to face this kind of defect.

Reflectometry is a non-destructive method, based on the radar principle. In fact, it consists in injecting a signal into the wiring network or structure and analyzing the signal reflected by the characteristic impedance of discontinuities.

Actually, reflectometry methods exhibit good performance for the detection of hard defects. In contrast, soft defects evoke much difficulty. Indeed, such defects result in a very small local variation of characteristic impedance (few ohms). It is so needed to be able to detect reflected signals with very low amplitude. Various techniques have been proposed in the literature to overcome the soft defect detection problem. [1] - [2]. But these methods have two main limitations: complexity and false alarms. Recently, a new method was proposed, called Self - Signature Magnification by Selective Windowing (SMSW). The advantage of this method is its low complexity (low cost) and allows us to avoid false alarms [3].

## Main Results

The SMSW algorithm allows to reduce the amount of noise and amplify the soft defect signature. SMSW targets only the defect zone and doesn't treat all the reflectogram.

The blue curve in Figure 1 shows the reflectogram of a cable with a soft defect located approximately at 2.50 meters. Here, the electrical pattern (first positive pulse followed by a second negative pulse), which characterizes a soft defect, has a very low amplitude and is difficult to detect.

The pink curve of Figure 1 shows that the SMSW method has enabled to amplify the signature of the soft defect without degrading the localization of its position. Similar results were performed experimentally (Figure 2).

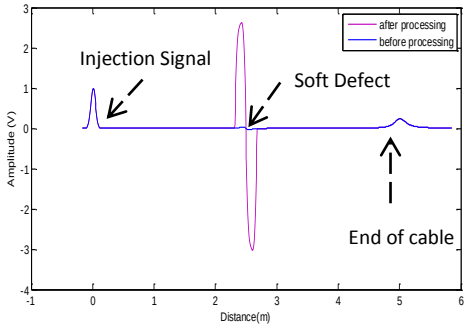



*Figure 1 : Simulation results with the shielded twisted pair*

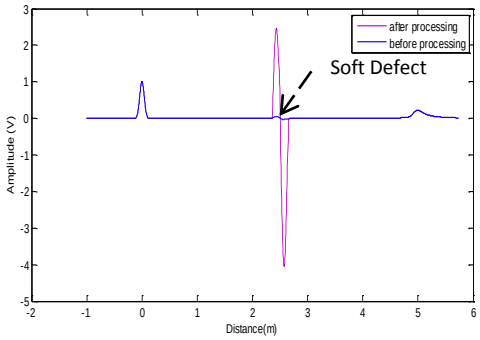
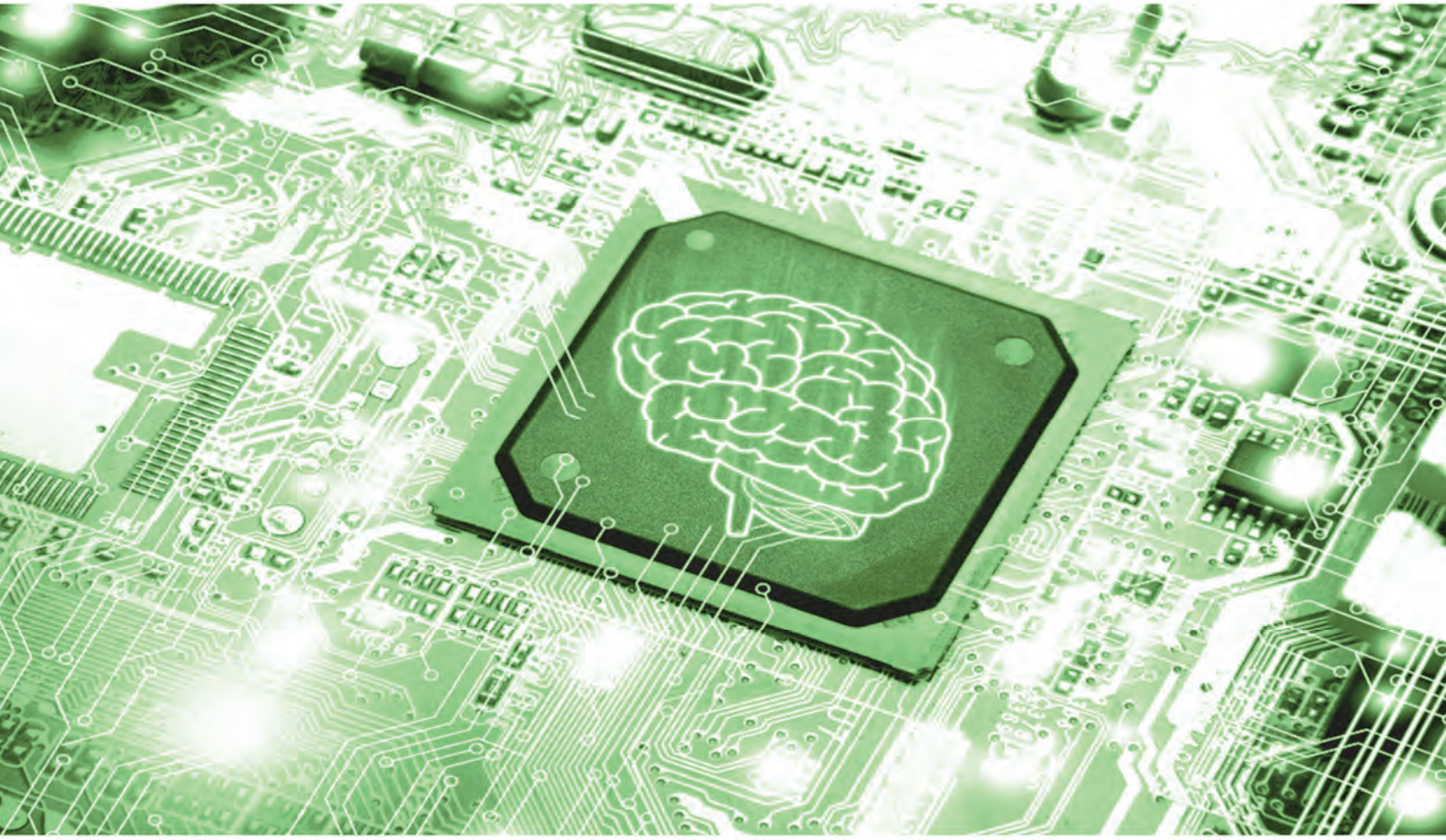



*Figure 2 : Experimental results with the shielded twisted pair*

## Perspectives

This research work introduces a new method of signal processing. The Soft defects localization by signature magnification with selective windowing method is promising with the gain brought in terms of soft defects detecting and its robustness to noise. Moreover, it may be easily implemented on a target processor such as DSP or FPGA with a minimum silicon area».

*Related Publications:*
[1] M. Franchet, N. Ravot and O. Picon, "On a useful tool to localize jacks in wiring network ", PIERS 2012 Kuala Lumpur. 2012.
[2] S. Sallem and N. Ravot, "Self-adaptive Correlation Method for Soft Defect Detection in cable by Reflectometry," IEEE SENSORS 2014 Conference, 2014
[3] S.Sallem ans N. Ravot, " Soft defects localization by signature magnification with selective windowing", IEEE SENSORS 2015 Conference, 2015



# 05

# RELIABLE SYSTEMS

- Polymorphic code for security
- Hypervision for security
- Real Time management
- Ageing & harsh environment

# Code Generation for Software Components Secured against Physical Attacks

*Research topic: CyberSecurity, Physical Attacks, Compilation, Runtime Code Generation*

**Authors: D. Couroussé, T. Barry, B. Robisson**

**Abstract: Physical attacks represent the strongest security threat against embedded systems. A secured system usually integrates protection mechanisms at several levels: attack detection mechanisms in the hardware, dedicated cryptographic IPs, and hardened software components. In this work, we provide code generation tools to automatically apply software protection schemes against both side channel attacks and fault attacks (1) using traditional static compilation techniques, and (2) in a new paradigm shift: code polymorphism, which involves runtime code generation.**

## Context and Challenges

Physical attacks are a mean to break the security protections of a computing system. They are mostly used against embedded systems, where the device is in the hands of the attacker, but are also effective against other computing systems such as computers and servers. The research literature contains several hundreds of known attacks, and a secured component needs to be protected against all these. Usually, the protection schemes are ad hoc; and are added iteratively to the design. The industry still misses tools in order to automatically apply these protection schemes.

## Main Results

We developed a secured compiler based on the compilation framework LLVM [1]. The compiler is able to automatically apply a protection scheme against fault attacks, on a portion of the source code selected by the developer. The scheme consists in duplicating every instruction in the protected section of the program, so that an attack with an instruction skip fault model is not possible. This scheme has been formally verified for the ARM thumb instruction set [3].
Figure 1 provides a performance comparison of the same protection scheme applied in this work and by [3], where the protection scheme was applied semi-manually after the compilation step. It illustrates the fact that in our case the cost of the protection is mitigated by the performance optimisations applied by the compiler on the protected code.

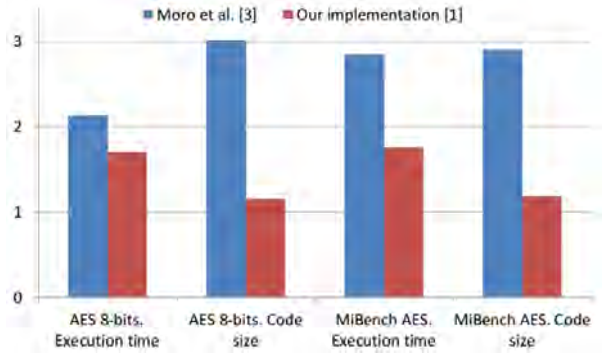



*Figure 1: Security overheads in terms of execution time and code size for two implementations of AES.*

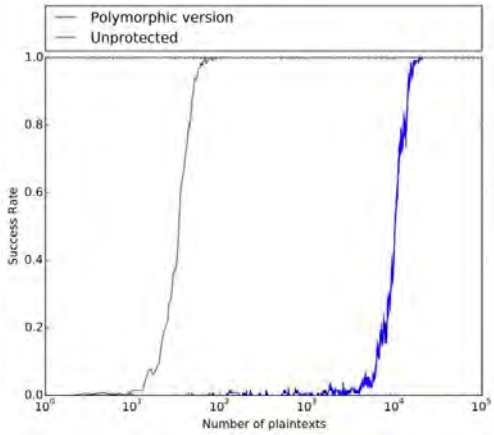



*Figure 2: success rate of a CPA attack on AES, comparing the resistance of an unprotected version and of a version protected with polymorphism.*

We also develop a new security mechanism for embedded systems, code polymorphism [2], which consists in regularly changing the behaviour of a secured component at runtime while maintaining unchanged its functional properties, in order to decrease the effectiveness of physical attacks. It is achieved by means of runtime code generation, and can tackle security issues at several levels: it increases the difficulty of reverse engineering, and it provides dynamically changing temporal and spatial properties that also increase the difficulty of side channel and fault attacks. Our implementation can target embedded devices with a few kilo-bytes of memory only. Furthermore, this solution is generic and compatible with other state of the art protection schemes
Figure 2 illustrates the security improvement of an implementation of AES protected with code polymorphism as compared to an unprotected version, in terms of resistance against a side channel attack using Correlation Power Analysis (CPA). The quantification of the security level is indicated by the success rate of the attack. Figure 2 illustrates that, in the case of this attack, the security margin is increased by a factor of 300.

## Perspectives

We will add other protection schemes supported by our tools, and we will work on the verification of the secured code.

*Related Publications:*
[1] T. Barry, D. Couroussé, and B. Robisson, “Compiler-based Countermeasure against Fault Attacks.” Workshop on Cryptographic Hardware and Embedded Systems (CHES), Saint-Malo, september-2015.
[2] D. Couroussé, T. Barry, B. Robisson, P. Jaillon, J. Lanet, and O. Potin, “Runtime Code Polymorphism as a Protection against Physical Attacks.” Workshop on Cryptographic Hardware; Embedded Systems (CHES), Saint-Malo, september-2015.
[3] Moro, Nicolas, et al. "Formal verification of a software countermeasure against instruction skip attacks." Journal of Cryptographic Engineering 4.3 (2014): 145-156.

# Blind hypervision to protect with a high level of assurance the privacy of Virtual Machines against software hypervisor vulnerabilities

*Research topic: Virtualization, Hypervision, Virtual Machine privacy, Security, Confidentiality, Integrity*

**Authors: M. Ait Hmid, M. Aichouch, P. Dubrulle**

**Abstract: The proposed "blind hypervision" concept intends to warrant the data privacy of Virtual Machines even if the Hypervisor is not trustworthy. The claimed security properties (confidentiality and integrity of the Virtual Machines) come with a high level of assurance as the software part of the Trusted Computing Base is of small size and then easily formally verifiable. Two proof of concept have been prototyped: a software implementation on an existing embedded SOC relying on the ARM TrustZone hardware security feature and a hardware evolution of the TSAR many-core architecture.**

## Context and Challenges

Type-1 hypervision is traditionally used to securely share resources between several application domains, some of which may be trustworthy and some potentially vulnerable. In such a configuration, the hypervisor has full access over hardware resources, especially memory content. The hypervisor traditionally acts as Trusted Computing Base, and must be fully trusted. However, full-featured hypervisors are too large to be entirely verified, thus become the target of privilege escalation attacks, as illustrated by e.g. the VENOM vulnerability revealed May 2015.

The objective of the proposed concept is to warrant confidentiality and integrity of the VMs, with a high level of assurance (formal verification feasible), from software attacks including those from the untrusted hypervisor. The goal is also to implement this concept with a low performance impact, with a full-featured and rich market hypervisor, in an already available hardware.

## Main Results

The blind hypervisor concept described in [1]) consists in isolating two required security functions from the remaining functions of the untrusted hypervisor (Figure 1).

The "Secure Memory management Unit" function is responsible of managing memory partitions for each VM and the hypervisor. Each VM and the Hypervisor run in isolation in a single memory partition. The Secure MMU securely handles the memory partition switches by managing execution contexts and performing the required clean up.

The "Trusted Loader" function is in charge of deciphering (resp. ciphering) and copying the VM from (resp. to) a hard disk (or network) to (resp. from) its memory partition allocated by the Secure MMU.

The untrusted Hypervisor blindly manages the VMs through the Secure MMU and the Trusted Loader function, i.e. without ever being able to access the VMs' code and data.

A first proof of concept demonstration has been implemented on a SOC using the ARM security extensions (TrustZone hardware feature), the Freescale iMX6 ARM v7 Cortex A9 processor. A Secure Kernel (less than 3kLoC) implementing the two security functions (Secure MMU and Trusted Loader) relying on the TrustZone feature runs in the Secure World, protected from attacks from the VMs and a lightweight Hypervisor running on the Normal World (Figure 2).

A second implementation (see [2] and [3]) is an evolution of the TSAR many-core architecture. Each VM runs independently (i.e. hypervisor is disengaged) in a set of dedicated clusters statically allocated by the untrusted hypervisor running on a dedicated cluster. Two new hardware components implement the Secure MMU and Trusted Loader functions. The Hardware Address Translation (HAT) confines memory access from a VM in its dedicated clusters. The cryptographic processor H-Crypt is used to (de)cipher and load VM images into dedicated clusters.

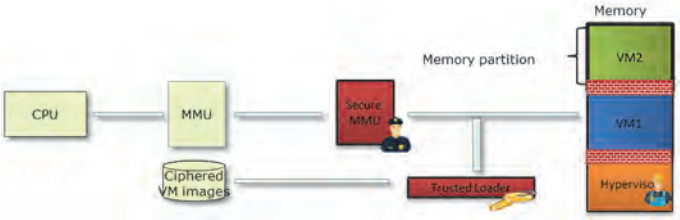



*Figure 1: Untrusted hypervisor blindly manages VMs running on isolated memory partitions through two trusted functions: Secure MMU and Trusted Loader.*


*Figure 2: Implementation with ARM TrustZone feature: The trusted Secure Kernel (formally verifiable) is responsible of the security properties of the VMs.*

## Perspectives

Next studies include porting an existing market hypervisor on the TrustZone implementation, formally verifying the Secure Kernel.

*Related Publications:*

[1] Dubrulle, P.; Ohayon, E. & Dore, P. (2015), 'Blind Hypervision To Protect Virtual Machine Privacy Against Hypervisor Escape Vulnerabilities"13th IEEE International Conference on Industrial Informatics (INDIN'15)', IEEE Computer Society, Royaume-uni.

[2] Aichouch, M. & Ait Hmid, M. (2015), Towards an Implementation of a Blind Hypervisor, in 'SEC2 INRIA Workshop'.

[3] Aichouch, M.; Ait Hmid, M. & Guy, G. (2015), Applications security in manycore platform: from operating system to hypervisor - Implementation of a Blind Hypervisor on a Manycore Architecture, in 'Workshop International sur les Architectures pour la Cryptographie dans les Systèmes Embarqués (CryptArchi - 2015)'.

# Managing the Latency of Data-Dependent Tasks in Embedded Streaming Applications

*Research topic: Parallelism, Real-time, Latency, Many-core*

**Authors: XK. Do, S. Louise, A. Cohen (INRIA)**

**Abstract: Because stream languages enable periodic scheduling, Cyclo-Static Dataflow (CSDF) models of computation and their variants are well fitted to modern real-time applications. Nevertheless, they can still violate timing constraints and safety requirements of critical real-time embedded systems (e.g. avionics). In this report, we investigate the applicability of the theory of hard-real-time scheduling for CSDF graphs while considering variable interprocessor communication (IPC) times, and real-time constraints imposed by hardware devices or control engineers. We evaluate the performance of our scheduling policy by using a set of 12 real-life streaming applications.**

## Context and Challenges

There is an increasing interest in developing applications on multiprocessor platforms due to their broad availability and the looming horizon of many-core chip, such as the MPPA-256 chip from Kalray (256 cores), which provide new opportunities, but introduces also several challenges on systems designers. The first challenge "How to express parallelism found in applications efficiently?" could be resolved by using data-flow models of computation because of its correctness of the parallel design. However, for the second challenge "How to provide guaranteed services against unavoidable interferences which can affect real-time performance?", no such de-facto solution exists. For this reason, in this report, we investigate the applicability of the hard-real-time scheduling theory for periodic tasks to streaming applications modeled as acyclic CSDF graphs, while considering variable IPC overhead and real-time constraints imposed by hardware devices or control engineers.

## Main Results

We introduce an extended hard-real-time scheduling (RTS) algorithm for applications modeled as CSDF graph. A Static Periodic Schedule of a Cyclo-Static Dataflow graph $G = \langle A, E \rangle$, (where $A$ is a set of actors, $E$ is a set of communication channels) is a schedule such that:

$$s(i,k) = s(i,0) + \alpha \times k \quad \forall a_i \in A$$

where $s(i,k)$ represents the time at which the $k$-*th* iteration of actor $a_i$ is fired and is an equal iteration period for every complete repetition of all the actors. Moreover, it is possible to schedule a graph $G$ actors as static periodic tasks using periods given by the following equation:

$$\alpha = q_1\lambda_1 = q_2\lambda_2 = \cdots = q_n\lambda_n$$

where $q_i \in \vec{q}$ (The basic repetition vector of $G$) and $\lambda_i \in \vec{\lambda}$ (The minimum period vector of G), given by:

$$\lambda_i^{min} = \frac{Q}{q_i}\left\lceil\frac{\eta}{Q}\right\rceil \quad for\ a_i \in A$$

where $\eta = max_{a_i \in A}(\omega_i q_i)$ (with $\omega_i$ the execution time of $a_i$) and $Q = lcm(q_1, q_2, \cdots, q_n)$ ($lcm$ is the least common multiple operator).

However, in a real-time applications, temporal constraints are usually imposed by the control engineers or by electronic devices. For instance, an audio output sink should not experience any hiccups due to the aperiodic behavior caused by either the initial transition phase of the STS or by the variation of execution times from iteration to iteration. In this case, a throughput constraint could be imposed for the sink node (i.e. terminal actor) by the programmer. This constraint could be converted into a periodic constraint for the sink node. Moreover, the control engineer in the domain of avionics and automotive sector usually impose real-time constraints on actors for safety requirements or hardware features. In this case, we take care of these real-time constraints by defining:

$$\eta = max_{a_i \in A}(\omega_i^* q_i, T_i^* q_i)$$

where $T_i^*$ is the period imposed by the control engineer and $\omega_i^* = \omega_i + \varphi_i + \Delta_{clock}$ where $\varphi_i$ is the communication time from $a_i$ to its successors and $\Delta_{clock}$ is the sum of the maximum clock offset between 2 consecutive distributed nodes. As a result, the minimum period of each actor is given by:

$$T_i^{min} = \frac{\eta}{q_i} \quad for\ a_i \in A$$

We evaluate our proposed scheduling policy by performing an experiment on a set of 12 real-life streaming applications. As can be seen in Table I, we clearly see that our approach delivers the same maximum throughput as STS for 11 out of 12 applications.

Table I
RESULTS OF THROUGHPUT COMPARISON

| Application | $\text{Throughput}_{\text{STS}}$ | $\text{Throughput}_{\text{RTS}}$ | Ratio |
|---|---|---|---|
| DCT | $2.22 \times 10^{-3}$ | 4/1800 | 1.0 |
| FFT | $3.33 \times 10^{-3}$ | 3/900 | 1.0 |
| Beamformer | $5.13 \times 10^{-4}$ | 4/7800 | 1.0 |
| Filterbank | $2.64 \times 10^{-5}$ | 3/113430 | 1.0 |
| Sample-rate | $1.67 \times 10^{-1}$ | 1/6 | 1.0 |
| MP3 | $2.22 \times 10^{-1}$ | 8/36 | 1.0 |
| Motion Detection | $1.74726 \times 10^{-8}$ | 1/57232627 | 1.0 |
| Encoder | $4.73 \times 10^{-6}$ | 1/382000 | 1.8 |
| Decoder | $1.0 \times 10^{-4}$ | 1/10000 | 1.0 |
| Bipartite | $6.35 \times 10^{-2}$ | 16/252 | 1.0 |
| Satellite | $9.46 \times 10^{-4}$ | 1/1056 | 1.0 |
| Modem | $6.25 \times 10^{-2}$ | 1/16 | 1.0 |

## Perspectives

We view this work as an important first step to provide a failure-handling strategy for distributed real-time streaming applications.

*Related Publications:*

[1] X. Do, S. Louise, A. Cohen "Managing the Latency of Data-Dependent Tasks in Embedded Streaming Applications" in International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC), 2015.

# Toward a Model of Computation for Time-Constrained Applications on Manycores

*Research topic: Model of computation, compilation, Many-core systems, data-flow*

**Authors: S. Louise**

**Abstract: As the number of cores on computing systems increases, we can see a lack of a universal programming model for many-core systems and the challenge it convey. Ideally, a program should be written only once, and making it run on a given target should be the role of the compiler. But addressing the problem of programming these systems requires a good Model of Computation (MoC) as a base for both programming and compilation tools. We propose a first base for such a MoC. It would take the CycloStatic DataFlow (CSDF) MoC for its good properties, and extend it to overcome its limitations while retaining the good properties.**

## Context and Challenges

Nowadays, the limits on power usage and dissipation for single-chips are encouraging a trend toward more parallelism in both HPC and embedded systems. This is why off-the-shelf processors went from single-core in the 1990s to the generalization of the multi-core at the beginning of the 2010s. And now come the many-cores which are new platforms where the use of OpenMP is becoming an issue as large scale shared memory can only be used at a high cost in term of latencies and power consumption.

Manycores differ from multicores not only by the number of cores but also the communication means between the cores. We change from a single bus or an evolution of a single-bus to a Network on Chip (NoC) because approaches based on single buses are no longer sustainable.

The future also looms toward new applications that will be more computational, more dynamic and more real-time oriented, especially in embedded systems. Examples of such applications can be autonomous vehicles, augmented and virtual reality or Software Defined Radio.

## Main Results

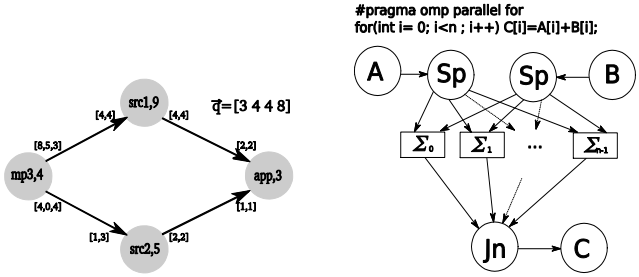
Ideally speaking a programmer would like to write universal code: write once, run anywhere. This goal was arguably reached for single process applications with a C-compiler (or other usual programming language), because C exposes a good generic abstraction of the model of computation of a general purpose processor. This is what is currently lacking for many-core systems. But before heading for compilers, we need to find a sound generic base for a Model of Compilation (MoC), as MoCs provide an abstraction for both programming languages and Intermediate Representations (IR) in a compilation toolchain.

We can identify several problems: to efficiently exploit this amazing processing power, there is a need of an exponential growth of the expressed parallelism of target applications. But usual parallel programming paradigms do not scale well.

Then we still have a problem to access to enough data to feed the cores. This is the memory barrier. And finally, we have the problem of power consumption.

Usual approaches from HPC, especially OpenMP do not work for real-time embedded manycores. Shared data consistency is an expensive features in terms of power budget and timing uncertainties. Lately, there has been a movement toward using message passing (e.g. MPI) to avoid that. In the embedded world, this idea gave birth to a renewed interest in dataflow concepts and its derivatives.

The bases rely on Kahn Process Networks (KPN) and their derivations, such as Cyclo-Static Data Flows –CSDF. Applications are defined as directed graphs whose nodes are processes and edges are communication channels. Channels are the only communication means between processes and reading is blocking if an insufficient number of tokens is present on any input channel of a process. KPN and CSDF are locally deter-
ministic for their execution and the possibility to run a CSDF in bounded memory and without deadlocks is a decidable problem. This means well-formed CSDF applications can be statically determined and they are globally deterministic in that case and easy to schedule, as seen in the figure below.




Such model is well fit to express high level of parallelism while retaining the deterministic execution like on a sequential program. But we need to improve on these models to insert more dynamic execution and more real-time compliance.

## Perspectives

From the base of CSDF, we can try and define new models of computation (MoC) that would express large scale parallelism, context based dynamism, real-time compliance and a good determinism to ensure ease of programming and debugging [2].

*Related Publications:*
[1] "Toward a Model of Computation for Time-Constrained Applications on Manycores", S. Louise. 10th conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2015).
[2] "A model of computation for real-time applications on embedded manycores", Louise, S., Dubrulle, P., and Goubier, T. IEEE 8th International Symposium on Embedded Multicore/Manycore SoCs (MCSoC 2014), pages 333–340..

# Towards Time-triggered Component-based System Models

*Research topic: Time-Triggered; correct-by-construction; component-based design; model transformation*

**Authors: H.Guesmi, B.Ben Hedia, S.Bliudze (EPFL, RiSD), S.Bensalem(Verimag), J.Combaz(Verimag)**

**Abstract: In this paper, we propose a methodology for producing correct-by-construction Time-Triggered (TT) physical model by starting from a high-level model of the application software in Behaviour, Interaction, Priority (BIP). BIP is a component-based framework with formal semantics that rely on multi-party interactions for synchronizing components. Commonly in TT implementations, processes interact with each other through a communication medium. Our methodology transforms, depending on a user-defined task mapping, high-level BIP models where communication between components is strongly synchronized into TT physical model that integrates a communication medium.**

## Context and Challenges

Analysis and design of hard real-time systems often starts with developing a high-level model of the system. Building models allows designers to abstract away implementation details and validate the model regarding a set of intended requirements through different techniques such as formal verification, simulation, and testing. However, deriving a correct TT implementation from a high-level model is always challenging, since adding TT implementation details involves many subtleties that can potentially introduce errors to the resulting system. Thus it is highly advantageous if designers can somehow derive a model with implementation details in a systematic and correct way from high-level models. We call such a model physical model. It can be automatically translated to the programming language specific to the target TT platform.

## Main Results

In [1], we propose a generic framework for transforming a high-level BIP model into a TT physical model that respects the TT communication pattern. The main subtlety in this work is how to move from strongly synchronized communication to a communication through a medium. In order to address this subtlety, we first defined the pattern of the obtained model. And then we defined transformation rules leading to such models.

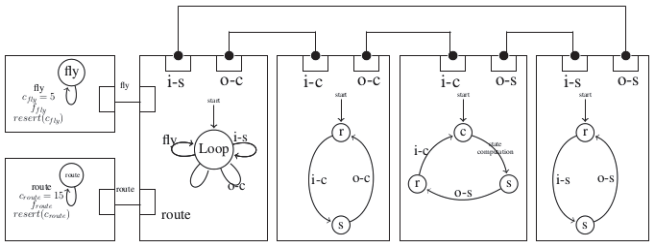



*Figure 1 : Initial model of Flight Simulator application*

The final model should be structured in three layers: task components layer, layer of communication components (TTCC) which replace initial intertask connectors, and the layer responsible for conflicts resolution (CRP component).
The transformation process leading to such a model consists of: (1) breaking atomicity of actions in ATC components by replacing strong synchronizations with asynchronous send/receive interactions, (2) inserting TTCC components that coordinate execution of inter-task interactions according to a user-defined task mapping, (3) extending the model with an algorithm for handling conflicts between TTCC and (4) adding local priority rules in task components for handling conflicts between inter-task and intra-task interactions.
In Fig.1. We display a BIP model of a flight simulator application. We applied our automatic transformation to this model and we obtained the model displayed in Fig.2. .

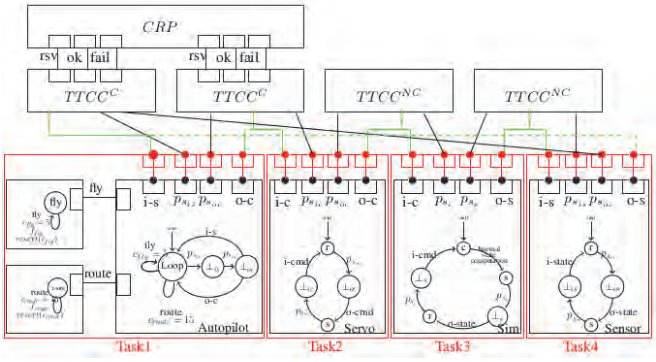



*Figure 2 : Obtained model after transformation of model of Fig.1.*

In order to prove that the transformation is correct (i.e., semantics preserving), we used the weak bisimulation technique since it induces trace equivalence.
This transformation is an important result since it guarantees that all properties verified on the initial model, are respected in the obtained one with no need to a posteriori verification.

## Perspectives

For future work, we plan to automate the transformation process in BIP modeling environment. Furthermore, we are working on the tool allowing code generation for the target TT platform,that depends on the offered communication services of target operating system. For each specific target operating system, it translates an adapted version of the model.

*Related Publications:*
[1] Guesmi, H.; Ben Hedia, B.; Bliudze, S.; Bensalem, S. & Combaz, J. (2015), 'Towards Time-Triggered component-based system models''The Tenth International Conference on Software Engineering Advances (ICSEA - 2015),'.

# GenetIC toolbox for ageing-aware processor design and exploration

*Research topic: Embedded systems reliability*

Authors: O. Heron, C. Sandionigi, E. Piriou, V.Huard (STMicroelectronics), F.Cacho (STMicroelectronics)

**Abstract: Transistor ageing is recognized as a key bottleneck for performance and reliability of embedded systems; it introduces delay on the paths of the circuit and may induce the failure of the system due to timing variations. As safety-critical control functions begin to be integrated in multicore processors, this issue must be addressed as early as possible in the IC design flow. ISO26262 automotive standard now recommends methods and tools able to guaranty the failure rate of processor architectures at design time. This work presents GenetIC toolbox, an environment to aid design teams to take the right design decisions under ageing constraint.**

## Context and Challenges

Based on ISO26262 automotive standard, the integration of safety-critical control functions in multicore architectures must guarantee reliability and availability levels. Technology is facing the challenges of nano-scaling, which causes faster transistor ageing. The effects of ageing on the behavior of processors depend on environment, technology, design and applications. Hence, design teams deal with crucial challenges: How to get an accurate failure rate at design time? Which manufacturing technology to choose? Which protection strategy against ageing to select?

The GenetIC toolbox is an integrated environment that enables exploration and design of processor architectures under ageing conditions. It implements a methodology that accurately estimates the global failure rate of a processor. The estimations aid to define protection strategies. The main characteristics are the following:

- The estimations are done at RTL without the need of a physical implementation, hence saving time.
- The environment relies on State-of-the-Art CAD tools, ready to be integrated in product design flow
- The estimation accuracy is already validated with manufacturer reference values.

## Main Results

An overview of the steps to achieve a robust processor design with the GenetIC toolbox is illustrated in Figure 1.

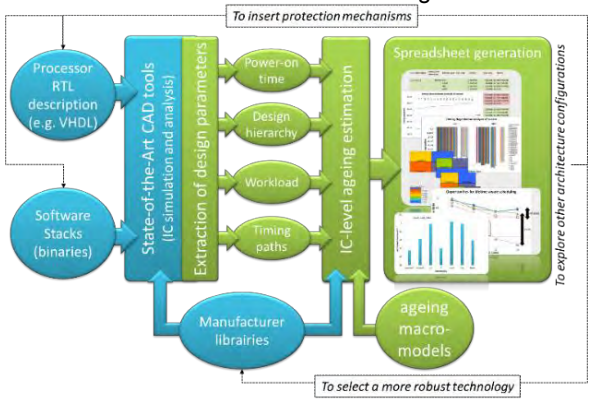



*Figure 1: Overview of GenetIC toolbox*

First, the relevant parameters for ageing analysis are extracted from RTL design description, software binary and manufacturer libraries with the aid of simulation and analysis tools. Then, the global failure rate is estimated by using a mathematical ageing macro-model, which aggregates the extracted parameters and post-stress characteristics from the manufacturer libraries. Finally, the various spreadsheets aid the design teams to decide whether to validate the selected inputs or to explore alternative design or technology solutions. This work considers NBTI and HCI mechanisms.

The first contribution of the work was the definition of a solution to propagate State-of-the-Art models from transistor to higher abstraction levels. In [1] a method to estimate the aging effects at RTL is presented. Figure 2 shows good accuracy with respect to gate level; the absolute error is bounded to 18% and the median error is about 1%.

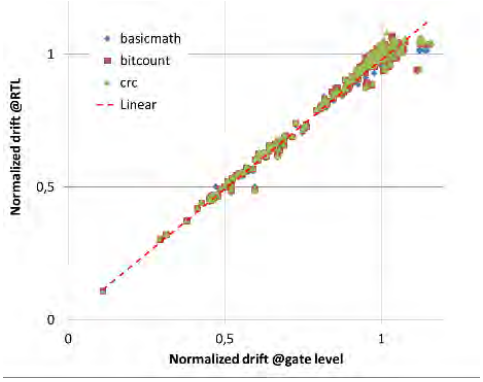



*Figure 2: Correlation of estimations between RTL and gate level*

The second contribution of this work was to further speed up the analysis execution and to make results more concise to be ready to address multicore architectures. A solution to only extract a subset of timing paths representative of the behavior of the processor under ageing conditions is proposed in [2].

As a side outcome of this work, memories were also addressed [3].

## Perspectives

The current work aims at raising the working abstraction level, evaluating aging at TLM, and at preparing on-chip monitoring and on-line management of ageing.

*Related Publications:*

[1] O. Heron, C. Sandionigi, E. Piriou, S. Mbarek, V. Huard, "Workload-dependent BTI analysis in a processor core at high level", IEEE International Reliability Physics Symposium (IRPS), 2015

[2] C. Sandionigi, O. Heron, "Identifying aging-aware representative paths in processors", IEEE International On-Line Testing Symposium (IOLTS), 2015

[3] F. Cacho, E. Piriou, O. Heron, V. Huard, “Simulation Framework for Optimizing SRAM Power Consumption under Reliability Constraint”, MEDIAN Workshop held in conjunction with DATE’15, 2015.

# Field Programmable Gate Arrays and Memory Devices Radiation Tolerance

*Research topic: Radiation hardening, Radiation effects*

**Authors: J. M. Armani, J. L. Leray, R. Gaillard (consultant) and V. Iluta (ERMES)**

**Abstract: Several FPGAs and memory devices of various types have been irradiated in order to evaluate their total ionizing dose tolerance. The FPGAs showed significant degradation after 200 Gy and were all broken after 450 Gy. The SRAM failure level was above 2.4 kGy, while two Flash memories failed just after 1 kGy.**

## Context and Challenges

In the nuclear industry, FPGAs are gaining increased attention worldwide for application in nuclear power plant I&C systems since old analog and microcontroller-based systems are becoming obsolete and need to be replaced.
Radiation hardened FPGAs are available for the space community but they are much more expensive than COTS devices and their radiation tolerance may be too limited. Indeed, the robustness of electronic systems has to be orders of magnitude higher for a use in nuclear environment.
In this work [1] we have tested the Total Ionizing dose response of several FPGAs and memory devices in a view to design a hardened FPGA-based system for nuclear applications.

## Main Results

During the experiment we tested two FPGAs: the Lattice SRAM-based LFE3-35EA and the Microsemi Flash-based A3PE1500. We have also irradiated five memory devices : the 23LC1024, a 1 Mbit SRAM from Microchip, the MR25H40, a 4 Mbit MRAM sold by Everspin, the EPCS16SI8 and the M25PX16, two 16 Mbit Flash from Altera and Micron respectively, and the N25Q128, a 128 Mbit Flash manufactured by Micron.

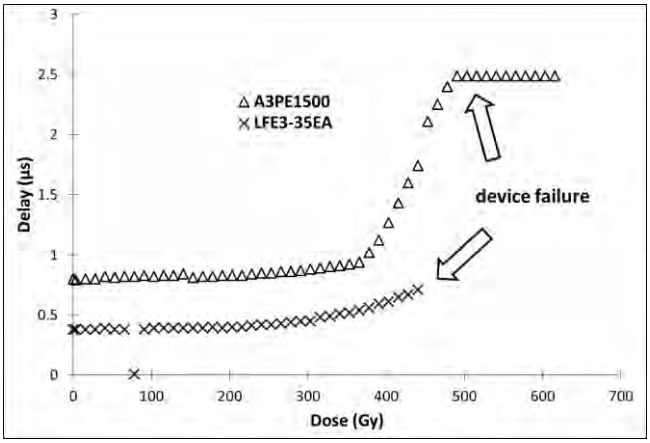



*Figure 1: Evolution of the flip-flop chain propagation delay for the Microsemi A3PE1500 and Lattice LFE3-35EA FPGAs during irradiation.*

Both FPGAs included two chains of logic elements in order to measure degradation of the propagation delays: one of 800 AND-gates and another of 800 D-type flip-flops configured as a ripple clock divider.
Fig. 1 shows the evolution of the clock propagation delay in the flip-flop chain for Microsemi and Lattice FPGAs. Both devices start degrading significantly after 200 Gy and exhibit total failure at 450 Gy.
The memories have been tested in a static mode to evaluate their data retention and writing capabilities. Eight pre-defined patterns (00, FF, AA, 55, CC, 33, 66 and 99) were stored manually in the devices before an irradiation step.

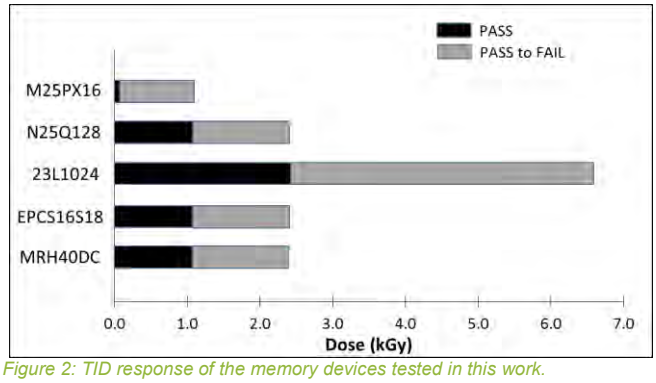


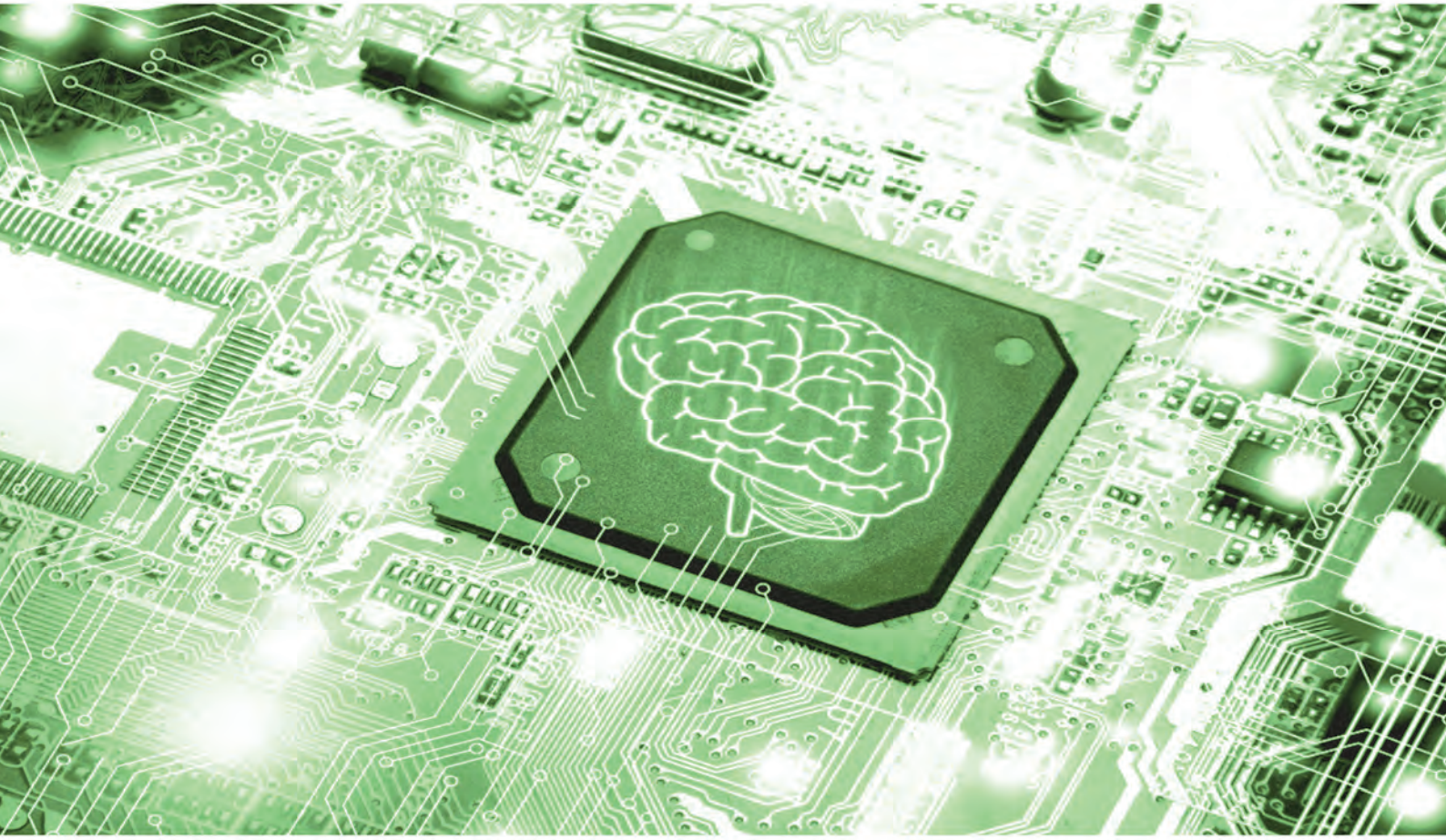

*Figure 2: TID response of the memory devices tested in this work.*

Fig. 2 shows the failure dose level for the different memory devices. The three Flash memories failed before 2 kGy as the MRAM component. The SRAM was still functional at 2400 Gy but failed before 6600 Gy. However, it is the most resistant of the tested devices.

## Perspectives

The objective of this work was to assess the TID response of several FPGAs and memory devices from different vendors.
The total dose tolerance of the FPGAs is around 200 Gy. The SRAM memory has a failure dose level lower than 6.6 kGy. These levels are not sufficient for our objectives. So a new irradiation test with other components should be planned.

*Related Publications:*
[1] J. M. Armani, J. L. Leray, R. Gaillard and V. Iluta, "TID Response of Various Field Programmable Gate Arrays and Memory Devices, " IEEE International Conference on Radiation and its Effects on Components and Systems (RADECS), Moscow, Sept. 2015.



# 06

# EMERGING TECHNOLOGIES & PARADIGMS

- Monolithic 3D
- NEMS-based logic
- TFET Memory
- ReRAM circuits
- Neuromorphic architectures

# Partitioning and Intermediate BEOL process influence on Power and Performance for Monolithic 3D IC

*Research topic: 3D IC, CoolCube, 3D Monolithic,*

**Authors: H. Sarhan, S. Thuries, O. Billoint, F. Clermidy**

**Abstract: Monolithic 3D (M3D) integration technology offers fine grain gate level stacking capability compared to 3D Through Silicon Vias (3D-TSV) which is well adapted to coarse-grain applications. Though, design partitioning, i.e. which cell on which tier, affects the 3D design performance. In this work we demonstrate that un-balancing the tier to tier area ratio of the M3D design brings better performance than classical balanced 3D design approach and study impact in Performance, Power and Area (PPA) using W/SiO2 compared to Cu/low-k IBEOL.**

## Context and Challenges

3D-IC have been demonstrated with significant performance and power gains compared to 2D-IC for applications like Memory Cubes, Memory-on-Logic and Multi-Processor System on Chip with Through Silicon Via (TSV) from 50µm down to 5µm pitch while 3D VNAND Flash memory has taken huge benefits by increasing the density of integration going into 3D. Currently, 3D technologies scale down, providing up to $10^8$ 3D Vias per mm² in case of Monolithic 3D (M3D) and CoolCube™ process, thus offering high density gate-level integration[1]. M3D differs from state of the art 3D technologies by its sequential process avoiding the bonding phase after fabrication of each die or wafer i.e. the second transistor layer is directly fabricated on top of the first one within the same process while dealing with 3D vias of about 50nm diameter. This aggressive pitch can open new perspectives of Power Performance and Area (PPA) gains for Internet of Things, Neural Networks or Processing in Memory applications. Consequently, fine grain partitioning at architecture level and technological parameters impact on PPA are new challenges that are addressed by CEA-Leti.

## Main Results

Un-balancing the tier to tier area ratio of the M3D design brings better performance than classical balanced 3D design approach. Our technique that consists in sorting the wires from the longest to the smallest and cutting long wires first in order to reduce as much as possible the resulting Total Wire Length (TWL) may lead to unbalance the area ratio. Simulations are done using ATRENTA SpyGlass Physical 2D and 3D as physical implementation prototyping tool to early evaluate PPA gains from one partitioning trial to another and highlight up to 24% performance improvement compared to 2D and 15% better performance than the state-of the-art technique, without extra power penalty. Table 1 shows all results [2].

Impact in Performance, Power and Area (PPA) of using W/SiO2 compared to Cu/low-k in the Intermediate BackEnd of Line (IBEOL) are presented in Figure 1; simulations are done using ATRENTA SpyGlass Physical 2D and 3D as well. Using W/SiO2 shows limited impact on performance with maximal 1.93% degradation compared to Cu/low-k IBEOL[3].

*Table 1: Results comparison between 2D and 3D cases (hMetis and PAP partitioning) for openMSP, Reconf-FFT, LDPC blocks.*

| | | No. Standard Cells | Area (µm²) | | TWL (µm) | No. M3D-VIA | Max. Perf. (GHz) | Perf. Gain (%) | Power @2D Max-Perf. (mW) | Power Gain (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Top | Bottom | | | | | | |
| openMSP | 2D | 6122 | 11390 | | 91278 | NA | 1.21 | NA | 11.54 | NA |
| | hMetis [14] | 5456 | 6000 | 6000 | 89642 | 1110 | 1.24 | 2.5% | 11.53 | 0.1% |
| | PAP (31-69) | 5466 | 7875 | 3530 | 89630 | 3775 | 1.42 | 17.5% | 11.74 | -1.7% |
| Reconf-FFT | 2D | 29673 | 27600 | | 420547 | NA | 1.33 | NA | 52.50 | NA |
| | hMetis [14] | 20093 | 13853 | 13853 | 316490 | 1158 | 1.45 | 9.0% | 30.87 | 41% |
| | PAP (40-60) | 20604 | 16656 | 10902 | 321269 | 13894 | 1.64 | 24.0% | 28.75 | 45% |
| LDPC | 2D | 68179 | 88800 | | 3277363 | NA | 1.58 | NA | 103.5 | NA |
| | hMetis [14] | 56896 | 52496 | 52496 | 1431432 | 12511 | 1.74 | 9.9% | 98.3 | 5.2% |
| | PAP (25-75) | 56896 | 78660 | 25380 | 965889 | 26917 | 1.76 | 11.5% | 100.7 | 3.0% |

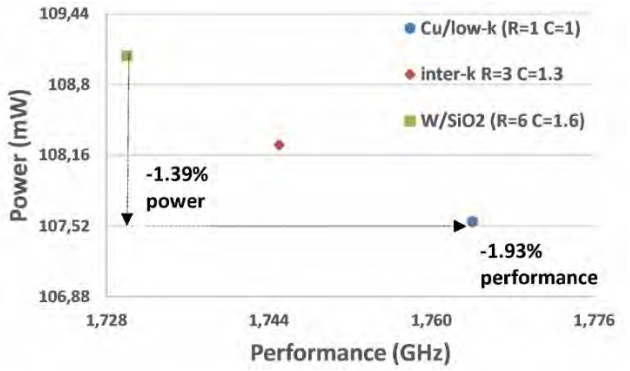



*Figure 1: IBEOL effect on Performance and Power results for LDPC block*

## Perspectives

Memories are already taking large benefits of 3D implementation at coarse grain level with HBM or HMC memory cube and up to fine grain with 3D V-NAND or 3D Crossbar architectures. With growing cost of interconnect, logic need to be pushed to 3D while addressing higher density of integration to decrease long interconnects impact on Power, Performance and Area. Monolithic 3D based CoolCube™ technology offers a promising alternative to 2D scaling for Logic-on-Logic application. Fine grain partitioning also raises new design and methodology challenges to achieve high quality of results. Impact on performance of IBEOL technological parameters has to be studied.

*Related Publications:*
[1] P. BATUDE, et al. "3DVLSI with CoolCube process: An alternative path to scaling." VLSI technology symposium 2015.
[2] H. SARHAN, et al. “An Unbalanced Area Ratio Study for High Performance Monolithic 3D Integrated Circuits”. ISVLSI 2015
[3] H. SARHAN, et.al. “Intermediate BEOL process influence on Power and Performance For 3DVLSI”. 3DIC 2015.

# Adiabatic Logic for Quasi-Zero Power Dissipation

*Research topic: Adiabatic/Reversible/Sub-threshold logic, Energy recovery, Electromechanical relays*

**Authors: H. Fanet, G. Pillonnet, G. Billiot, S. Houri (Univ. Delft), A. Valentian, M. Belleville**

**Abstract: A detailed analysis and comparison of electromechanical systems and CMOS technologies for low power adiabatic logic implementation is presented. Fundamental limits of CMOS-based adiabatic logic are identified. The interest of combining N/MEMS technology and adiabatic logic is described, and the key electromechanical switch parameters that govern the dissipation-performance relationship are identified. NEMS-based adiabatic gates architectures are also described and the contribution of the power-clock is estimated in order to compare CMOS and N/MEMS-based adiabatic architectures at the system level.**

## Context and Challenges

In order to reduce active power dissipation, standard solutions for current CMOS technologies consist in developing components working at low voltage and/or to design low frequency architectures. However, such simplified approaches are not able to solve the trade-off between energy dissipation and performance if ultra-low power constraints are required. Therefore various solutions have been suggested to drastically improve the performance of low-power circuits. On a *device level*, the use of advanced device technologies such as FDSOI improves the electrostatic control over the conducting channel. A promising emerging solution is the use of electromechanical relays which eliminates the leakage current. On a *circuit level* architectures and power management solutions can be adapted in order to reduce power consumption such as sub-threshold circuits, power gating, asynchronous architecture, charge recovery and adiabatic charging.

## Main Results

[1] focuses particularly on the charge recovery and adiabatic charging solutions that will be explored in combination with both CMOS- and NEMS-based circuits in order to determine their envelope of operation in terms of performance and power consumption. We have analyzed CMOS and NEMS adiabatic logic taking into account device performance as well as the necessary power-clock generator. The analysis shows that while adiabatic charging is in principle an attractive means to low power logic, its advantages are negligible in the case of CMOS technology. In sharp contrast, NEMS-based adiabatic logic is exceptionally suitable to tap into the energy savings allowed by the adiabatic principle.

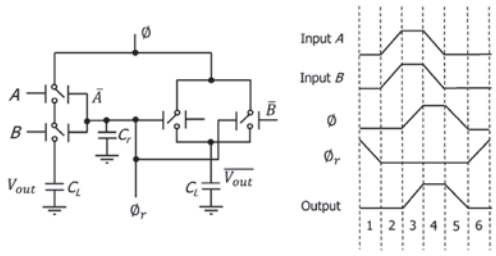



*Fig. 1: The Bistable Switch Adiabatic Gate (BISAG) architecture, showing an example of an AND function and the associated waveforms.*

Furthermore, three NEMS-based adiabatic logic architectures are proposed. The key NEMS switch parameters that govern the dissipation-performance relationship are identified as the switch commutation frequency, its actuation voltage, and the switch series resistance, itself dominated by the contact resistance between the switch electrodes. Therefore low contact resistance NEMS switch technology is crucial to implement such ultra-low power solutions. If NEM relays are able to combine reliable low contact resistance and sub-1 volt operating voltage, then the energy saving afforded by such a NEM-adiabatic logic combination is at least an order of magnitude better than what is possible with state of the art CMOS technology.

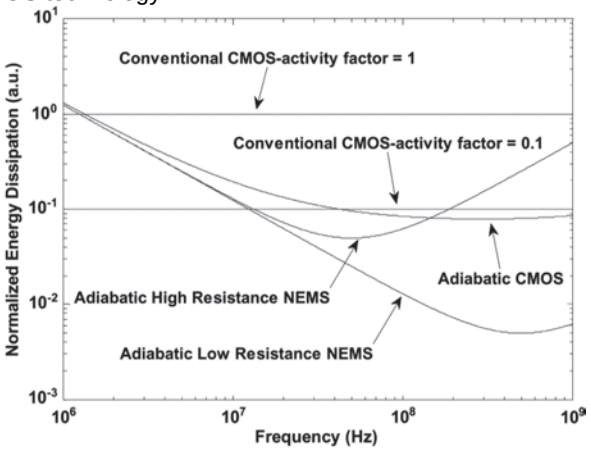



*Fig. 2: Energy dissipation versus operating frequency for adiabatic CMOS and NEMS switch circuits. Conventional CMOS dissipation for mean activity factors of and are also shown for comparison.*

The contribution of the power-clock or energy recovery generator has been estimated in order to compare CMOS and NEMS-based adiabatic architectures at the system level.

## Perspectives

NEM-adiabatic logic could reduce the actual power consumption of digital core by few decades. However, improving the low reliability of the existing relays by technology and circuit improvements is a key requirement to fully exploit this potential.

*Related Publications:*
[1] S. Houri, G. Billiot, M. Belleville, A. Valentian, H. Fanet, "Limits of CMOS technology and interest of NEMS relays for adiabatic logic applications" IEEE Transactions on Circuits and Systems, 2015

# CMOS/TFET Hybrid SRAMs for ultra-low leakage Wireless Sensor Node Applications

*Research topic: Hybrid Tunnel FET/CMOS Memory Design*

**Authors:** N. Gupta(ISEP), A. Makosiej, O. Thomas, A. Amara(ISEP), A. Vladimirescu(ISEP), C. Anghel (ISEP)

**Abstract: In this work applicability of Tunnel FET (TFET) for SRAMs is analyzed for ultra-low leakage embedded applications like Internet of Things (IoT). To this end a TFET/CMOS hybrid memory architecture is proposed using a novel 8T TFET SRAM cell operating at VDD=1V or lower. Owing to the properties of TFET the bitcell leakage is below 5fA at VDD=1V. The proposed application is a sensor node architecture in a hybrid CMOS/TFET process with a large memory consuming as little as 2fW/cell or 48pW for a 4kB array. The Read and Write Static Noise Margins (SNM) of the proposed bitcell are evaluated at 120mV and 200mV, with the operation speed of 3.8GHz and 800MHz in read and write, respectively.**

## Context and Challenges

The first and foremost requirement for wireless sensor nodes is to operate with energy harvested from the environment as replacing batteries even at long time intervals is impractical. Therefore a system optimization study is necessary to find the right balance between processing power, storage and communication required to ensure that it fits into the energy budget obtainable from a rechargeable source like a solar cell. A processor design for sensing applications in CMOS has demonstrated the necessity to address the minimization of the standby power. While the obvious supply voltage reduction and innovative standby mode techniques were applied they could not significantly overcome the leakage of the memory in standby leading the authors of contending with only 2k data RAM and 640b instruction RAM. The relatively high leakage of CMOS represents a serious problem that limits the complexity of sensor node. To address this challenge one can look at alternative devices that can coexist with CMOS. The Tunnel Field Effect Transistor (TFET) was proposed as a new possible solution to reduce the power dissipation.

## Main Results

A solution for future generations of sensor nodes is proposed that can save one to two orders of magnitude in power compared to existing CMOS implementations, as shown in Figure 1. The solution consists in using a different device, the TFET, for building the memory. TFETs have up to five orders of magnitude lower, very weakly dependent on supply voltage, leakage than the equivalent CMOS process.

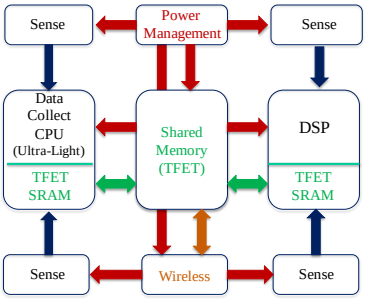



*Figure 1 Wireless Sensor Node Architecture with TFET Memory*

TFET SRAMs can operate up to 1 GHz if needed. Sensor nodes using TFET memory can afford larger embedded memories due both to the low leakage and the possibility to use the most recent process and minimize the occupied area. A new integrated TFET/CMOS single/dual port SRAM based scratchpad memory has been introduced (see Figure 2). The memory array is built with TFETs for ultra-low leakage and the periphery with FDSOI MOSFETs for speed. The proposed design uses new imbalanced sense amplifier based single ended read scheme and negative bitline write assist. Ultra-low leakage current of memory cells has been achieved without cell area increase in comparison to an 8T CMOS-DualPort SRAM cell and 29% area increase in comparison to a 6T SinglePort-SRAM. Gain up to 79.2% in write speed at sub-1V supply has been achieved.

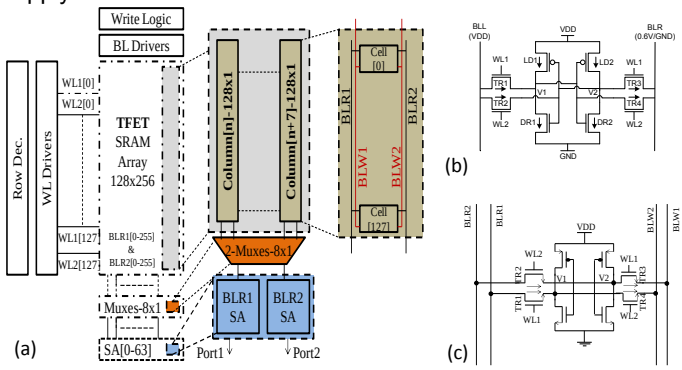



*Figure 2: a) Dual-Port Hybrid TFET/CMOS Memory Architecture, (b) TFET single port SRAM cell, and c) Dual-Port TFET SRAM cell*

## Perspectives

Sensor nodes using TFET SRAMs can afford larger on chip memory due both to the low leakage and the possibility to use the most recent process and minimize the occupied area. In contrast, sensor nodes reported to consume minimum power have to rely on older CMOS processes in order to avoid increased leakage and variability; this latter approach requires too much space also leading to limiting the size of the embedded memory. Last but not least, TFET devices can be implemented on the same FDSOI chip with CMOS transistors.

*Related Publications:*
[1] Gupta, N., Makosiej, A., Thomas, O., Amara, A., Vladimirescu, A., & Anghel, C. Ultra-low leakage sub-32nm TFET/CMOS hybrid 32kb pseudo DualPort scratchpad with GHz speed for embedded applications. ISCAS, 2015
[2] Gupta, N., Makosiej, A., Thomas, O., Amara, A., Vladimirescu, A., & Anghel, C. TFET/CMOS hybrid pseudo dual-port SRAM for scratchpad applications. In Ultimate Integration on Silicon (EUROSOI-ULIS), 2015
[3] Vladimirescu, A., Anghel, C., Amara, A., Gupta, N., & Makosiej, A. (2015, June). Sub-picowatt retention mode TFET memory for CMOS sensor processing nodes. In Advances in Sensors and Interfaces (IWASI), 2015.

# SneakPath Compensation Circuit for Programming and Read Operations in RRAM-based CrossPoint Architectures

*Research topic: CrossPoint, NonVolatile Memory, RRAM, compensation circuit, Multiple Level Cell*

**Authors: A. Levisse, B. Giraud, J.P. Noël, M. Moreau (IM2NP), J.M. Portal (IM2NP).**

**Abstract: Passive crossbar memories based on resistive switching bit-cells are today seen as the most promising candidates for flash memories replacement. However, inherent sneak currents through unselected devices lead to low operating margins and over-consumption during read and programing operations.Peripheral circuits have to be designed in order to mitigate the sneak currents that impact memory operations. In this work, we propose a dynamic sneak current compensation circuit for Set and read operations, enabling multi-level cell programming.**

## Context and Challenges

While conventional non-volatile memories (NVM) turns out to be more and more costly and complex to process, emerging resistive NVM (RRAM) are becoming more and more attractive for high density memories. Oxide-based RRAM (OxRAM) are one of the most promising RRAM technologies due to non-exotic materials, high scalability and fast switching. Beyond the standard 1Transistor-1RRAM architecture, transistor-less crosspoint architecture using 1Selector-1RRAM (1S1R) device allow ultra-high density ($4F^2$) and 3D stackability. However, inherent sneak currents (Sneak Paths, SP) through unselected resistances lead to performances reduction. Thus, beyond the technology improvements of the selector, innovative peripheral circuitry has to be developed to mitigate the SP impact on programing and reading operations. Mitigating SP impact allows a fine control of the programing operations and a precise sense of the read current.

## Main Results

SP impact on programming operations is studied and a novel SP compensation circuit is proposed. The presented circuit provides immunity to SP, which can be used for Multi-Level Cell (MLC) programming, and also for precise read operation. The SP is dynamically monitored and compensated in order to provide a one-step programming operation. The presented circuit is also studied for use as a sense amplifier on MLC programming operations.

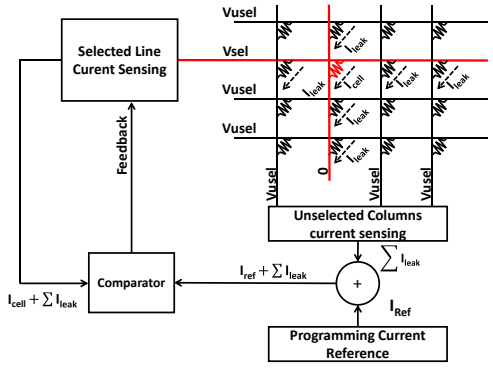



*Figure 1: Block diagram of the proposed programing circuit providing SP immunity with fine programming current tuning.*

Figure 1 presents the block diagram of the proposed SP compensation. The SP is monitored on the non-selected columns, added to the reference programming current and then compared to the selected line current. When this current pass over the selected line current, the programming current in the selected RRAM is limited to the aimed current. This leads to SP immune programming operations.

This SP compensation approach can also be used for read operations in order to maintain a good read margin in MLC RRAM in spite of the SP. Figure 2 presents the schematic diagram of the MLC read circuit with SP compensation.

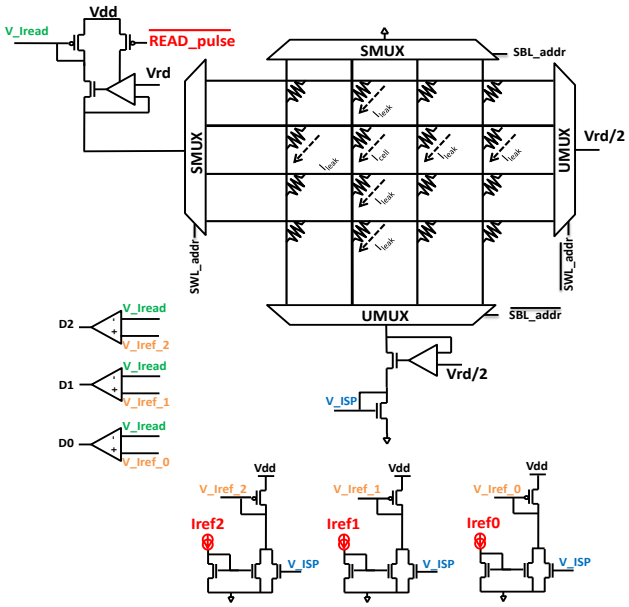



*Figure 2: Schematic diagram of the proposed circuit for read MLC (3 bits) in 1 cycle using ½ bias polarization scheme.*

## Perspectives

Read and Write peripheral circuits with SP compensation technique are presented. The proposed structures opens the way for MLC operations in crosspoint array which is an efficient way to increase the memory density.

*Related Publications:*
[1] A. Levisse, B. Giraud, J-P. Noel, M. Moreau, J-M. Portal, “SneakPath Compensation Circuit for Programming and Read Operations in RRAM-based CrossPoint Architectures”, NVMTS 2015
[2] A. Levisse, B. Giraud, J-P. Noel, M. Moreau, J-M. Portal, “Design Considerations during Programing Steps for bipolar OxRAM-Based Crossbar Architecture”, Workshop E-NVM 2015

# Twin neurons for efficient real-world data storage in networks of neural cliques

*Research topic: Neural network, Associative memory*

Authors: B. Boguslawski, V. Gripon (Telecom Bretagne), F. Seguin (Telecom Bretagne), F. Heitzmann

**Abstract: Associative memories are data structures that allow retrieval of previously stored messages given part of their content. Neuro-inspired sparse memories achieve a great power efficiency with a small latency. However, their performance may be highly degraded when stored messages show non-uniformity. By changing the structure of the model, it is possible to compensate the non-uniformity in stored messages. The concept of twin neurons was introduced, which approach almost optimal performance, while the implementation cost remains low.**

## Context and Challenges

Neuro inspired architectures are very promising, not only for new applications but also for energy efficiency. Neural Cliques Networks show great performances for retrieving partially erased messages, in terms of power efficiency and latency. This allows to use them as associative memory for quick-loop optimization in embedded systems, where low latency and low energy matters.

However, real-world applications with real-world data – i.e. with a non-uniform non-random distribution – has a strong impact on performances. We explored structural changes in the networks, which showed a great resilience to non-uniformity in stored messages, with a reasonable increase in the number of connections and neurons.

## Main Results

When a message is stored in a network of neural cliques, it is split into some segments. Each segment is mapped onto a cluster of neurons in the network. Then, each segment value is stored in one dedicated neuron. During the retrieval process, when only part of a message is known, the corresponding neurons are activated. Activation signals flow through connection to other neurons, which propagates to other neurons in return. This process converges in tens of nanoseconds to one and only one set of neurons, a fully connected graph, also named a *clique*. This clique represents the search message.

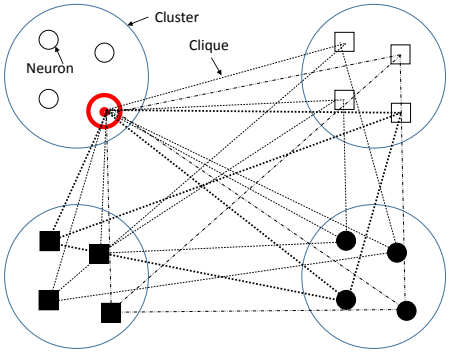



*Figure 1: One « hot neuron » in upper-left cluster.*

Problem arises when, for a given segment, a value occurs in many messages. When the corresponding neuron is activated, many neurons become potentially active, with a poor capability to discriminate between them. The iterative process may converge to a wrong message, or even several messages at the same time. Fig 1 depicts such a situation, with a “hot neuron” in the upper-left cluster.

Observations in Biology show both numbers of connections and neurons in the brain increase when they received a repeated stimulus over time.

The idea of twin neurons is to allocate a new neuron to store a segment value each time previously allocated neurons becomes overloaded, i.e. when a given limit is reached. This process splits the recurrent pieces of information among several neurons, the more often it occurs the wider it is spread among the global storage capacity. Network on Fig 2 stored the same messages as Fig &, with the hot neuron split into two twin neurons.

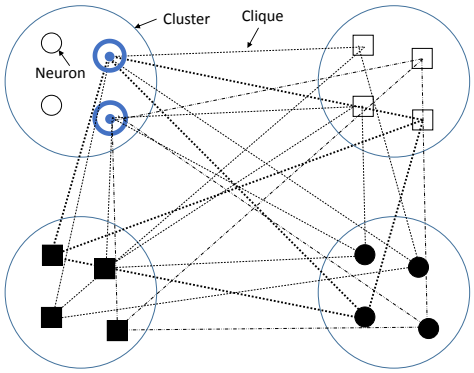



*Figure 2: Same network as Fig1 with two twin neurons. Information is better spread other the whole network.*

## Perspectives

These results were assessed in realistic use-cases of power management in a Multi-Processor System-On-Chip (MPSOC) in [1]. Performances in terms of energy, silicon area, and latency were also compared with other classical associative memory implementations in [2]. Next step is then to implement and validate this technique in a low-power circuit.

*Related Publications:*

[1] B. Boguslawski, V.Gripon, F. Seguin, and F. Heitzmann, “Twin Neurons for Efficient Real-World Data Distribution in Networks of Neural Cliques Applications in Power Management in Electronic Circuits”, IEEE Transactions on Neural Networks and Learning Systems, Volume 27, Issue 2, pp. 375-387, 2015

[2] B. Boguslawski,, F. Heitzmann, B. Larras, F. Seguin, “Energy Efficient Associative Memory Based on Neural Cliques”, IEEE Transactions on Circuits and Systems II: Express Briefs, Volume 63, Issue 4, pp. 376-380, 2015

# Memristive based device arrays combined with Spike based coding can enable efficient implementations of embedded neuromorphic circuits

*Research topic: Memristive devices, Neuromorphic design, Emerging technologies*

**Authors: C. Gamrat, O. Bichler, D. Roclin**

**Abstract: This work is focused on implementing neuromorphic circuits suitable for embedded applications. Such an objective puts the emphasis on two majors concerns: integration and energy efficiency. In our quest for ultimate integration, we report our investigations on the feasibility of large crossbars of synapse-like devices using memristive technologies. In an effort to tackle the energy problem, we introduce spike based coding for embedded deep neuromorphic architecture as a solution that paves the way for future embedded neuromorphic circuits**

## Context and Challenges

The operation of a neuromoprhic architecture involves two closely intertwined phases: 1) a learning phase during which network parameters (weights array and biases) are computed according to a given learning rule, and 2) a recalling phase during which outputs of neurons are computed according to parameter values set during the learning phase. Depending on architecture propositions, those two phases can be clearly separated or operated simultaneously.

In this work we focus on the recalling phase with weight arrays computed offline using the standard backpropagation algorithm..

## Main Results

In order to have comparison points, we have implemented a spike based network on the standard MNIST database benchmark. We have transposed the network topology into a four layers convolutional network (CONV) composed of two convolutional layers and two fully-connected layers as shown on figure 1.

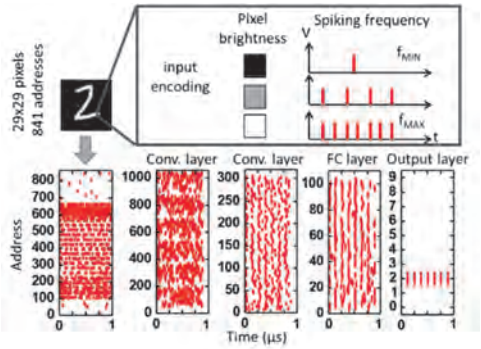



*Figure 1 : Spikes propagation for one pattern through the four layers of the CONV network: 2 Convolutional layers, 2 fully connected layers.*

Early results that we have obtained shows that our spike based version exhibit very similar recognition performances when compared to its classical counterpart with a test error rate of 0.5% close to the best published result on the MNIST database with a test error rate of 0.23% [Cireşan, 2012].

On table 1 we report the average spiking activity for each layers of the CONV networks. We observe a very low average activity in the network with a minimum of 0.86 events per connections on layer CONV 2 and an average of 1.26 for the entire network. This result indicates a very low stress on each synaptic device. Less events also meaning less energy, this suggests that spike coding could pave the way to very energy efficient implementations of Deep Neural Networks.

| Layer | Synapses (shared) | Connections | Events/ digit | Events/ connection |
|---|---|---|---|---|
| Conv 1 | 256 | 30,976 | 43,219 | 1.40 |
| Conv 2 | 2,250 | 36,000 | 31,594 | 0.88 |
| FC 1 | 57,600 | 57,600 | 89,841 | 1.56 |
| FC 2 (output) | 1,500 | 1,500 | 1,799 | 1.20 |
| Total | 61,606 | 126,076 | 166,453 | 1.26 (avg) |

*Table 1 : Details of the CONV network spiking activity.*

Finally we have studied the effect of the weight precision on the test error rate. We observe that above 12 discrete weight levels (4 bits) we reach the same test error rate of 0.6% regardless of the topology (FC or CONV) and the activation function.

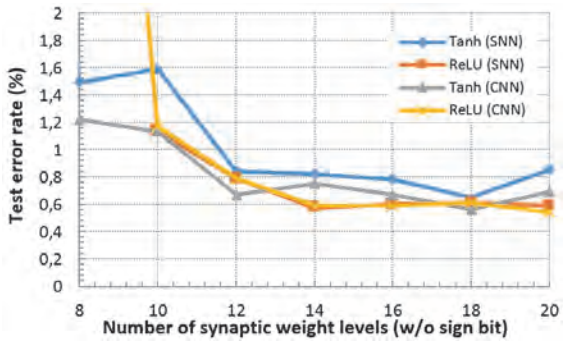



*Figure 2 : Test error rate as a function of weight resolution*

## Perspectives

Although this is a report of a work in progress, our results show that spike encoded implementations of DNN are possible with performances similar to classically encoded versions. We are currently investigating for ways of implementing a spiking learning rule. We believe that the combination of spike based coding and memristive technologies is a promising way for the energy efficient implementation of embedded DNNs.

*Related Publications:*
[1] C. Gamrat, O. Bichler, D. Roclin "Memristive based device arrays combined with Spike based coding can enable efficient implementations of embedded neuromorphic circuits", IEEE International Electron Devices Meeting (IEDM), 2015..

# Bioinspired Programming of Memory Devices for Implementing an Inference Engine

*Research topic: Neuromorphic design, Emerging technologies*

**Authors: D. Querlioz (CNRS), O. Bichler, A.F. Vincent (CNRS), C. Gamrat**

**Abstract: In this work, we use a neuroscience-inspired model of learning, spike-timing-dependent plasticity, to develop a bioinspired approach for programming memory devices, which naturally gives rise to an inference engine. The method is then adapted to different memory devices. By means of system-level simulations, we investigate several applications: image recognition, pattern detection within video and auditory data. These results highlight a possible new bioinspired paradigm for programming emerging memory devices, allowing the natural learning of a complex inference engine.**

## Context and Challenges

Cognitive tasks are essential for the modern applications of electronics, and rely on the capability to perform inference. The Von Neumann bottleneck is an important issue for such tasks, and emerging memory devices offer an opportunity to overcome this issue by fusing computing and memory, in nonvolatile instant ON/OFF systems. A vision for accomplishing this is to use brain-inspired architectures, which excel at inference and do not differentiate between computing and memory.

## Main Results

In this work, we have explored a bioinspired methodology for programming nanodevices, which naturally implements an inference engine. The system puts memory at the very core of computing, with the physics of the memory devices playing an active role in the process of learning and inference. The system is capable of learning using an unsupervised paradigm. More precisely, it can identify the latent variable underlying the input to which it is exposed, and at the same time infer its classification among the latent variable values.

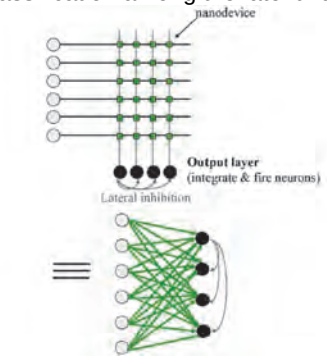



*Figure 1: Simplified crossbar architecture of the inference engine and its equivalent neural network.*

The system is tolerant to device variations and read disturb issues. We have shown that it may be adapted to various emergent memory device technologies. The most natural implementation uses cumulative memristive or Phase-Change Memory devices, in an analog manner. Another implementation can use stochastic devices such as Spin-Transfer Torque Magnetic Junctions. In many situations, stochastic and analog devices are equivalent. However, for most problems, this requires adding redundancy in the stochastic version.

We have explored several applications for the inference engine: image classification (figure 2), detection of patterns in video and auditory data. Although the results are impressive, they are currently far from production and need to be associated with other circuits capable of exploiting the inference results. The methodology introduced in this work provides a new paradigm for nanoelectronics. Bioinspired programming of memory devices could identify latent variables and perform inference, which is the biggest challenge for processing big data.

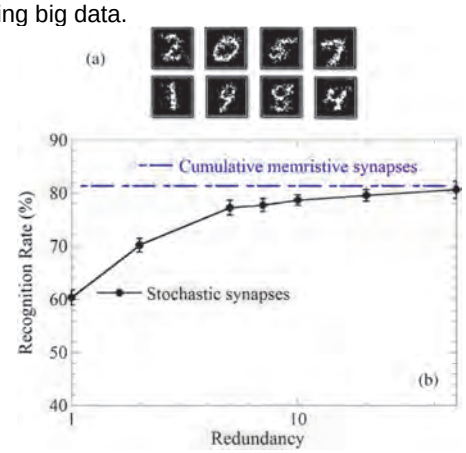
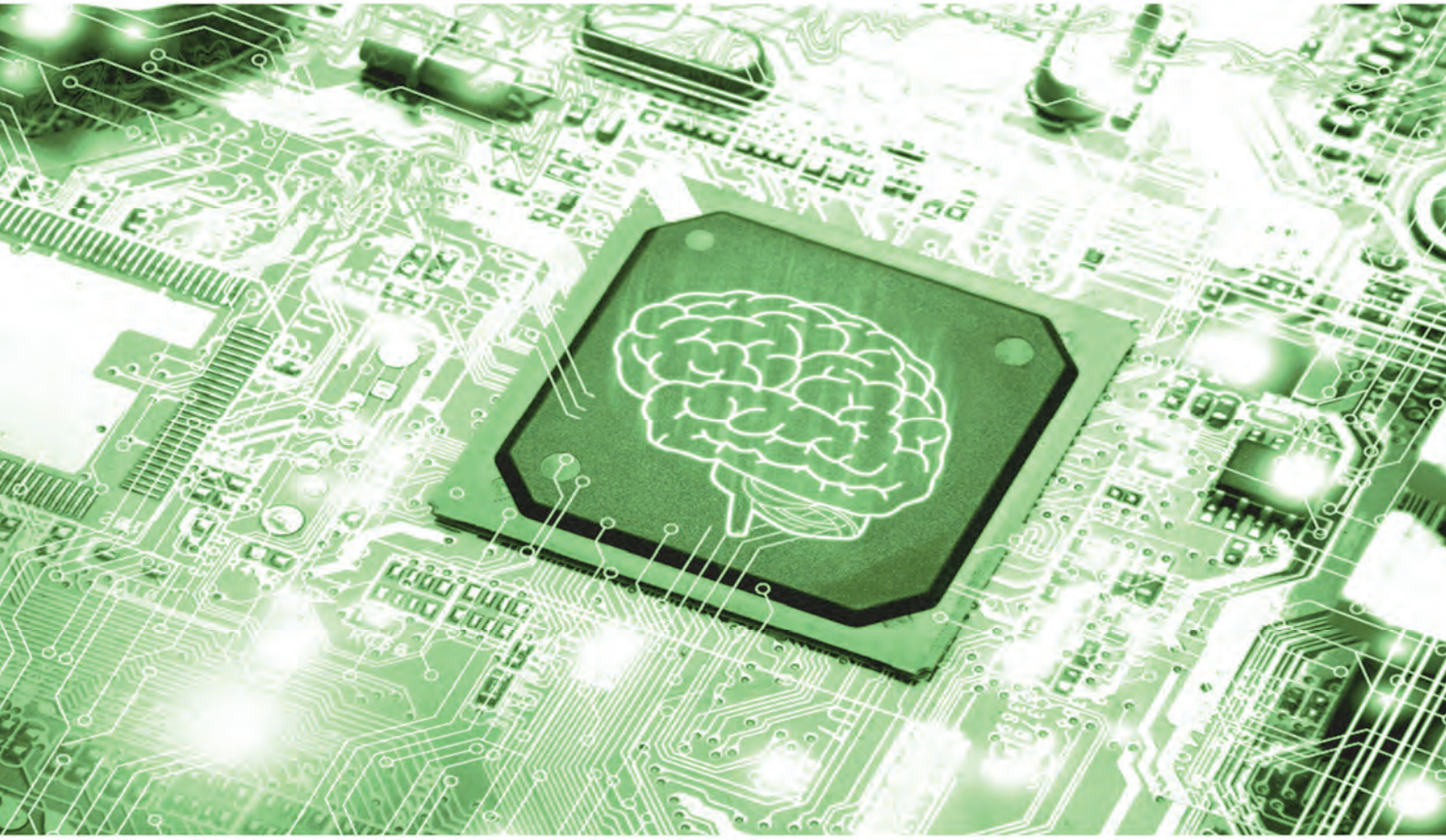



*Figure 2 : (a) Representation of the final conductance of stochastic synapses at the end of handwritten digit learning (MNIST). (b) Recognition rate as a function of device redundancy (number of devices connecting each input to each output)*

## Perspectives

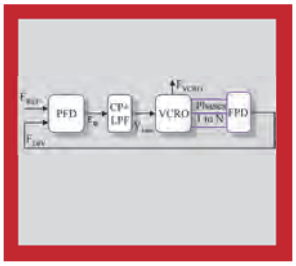
We have already connected the inference engine with Bayesian inference and the algorithm of expectation maximization. By advancing this comparison further, it should be possible to develop significantly more complex and comprehensive inference engines, while retaining the vision of this work: using the bioinspired idea of fusing computing and memory, we intend to put inference at the core of future nanoelectronics.

*Related Publications:*
[1] D. Querlioz, O. Bichler, A.F. Vincent, C. Gamrat, “Bioinspired Programming of Memory Devices for Implementing an Inference Engine”, Proceedings of the IEEE, August 2015, Vol. 103, No 8, pp 1398-1416, DOI:10.1109/JPROC.2015.2437616



# 07

## PHD DEGREES AWARDED IN 2015

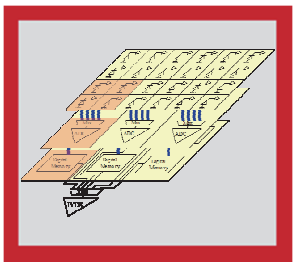



**Alexandre FONSECA**

*Université Nice Sophia Antipolis (France)*

## ARCHITECTURE AND DESIGN OF FREQUENCY SYNTHESIS CIRCUITS IN FDSOI 28NM.

The aim of this Ph.D. is, on one hand to develop a very low power consumption frequency synthesis and on the other hand to demonstrate the performances of the FDSOI technology for analog and radiofrequency applications. We chose as application example the BlueTooth Low Energy, BLE, standard.
System modelling and simulation allowed us to analyze the frequency synthesis pre-selected architecture based on phase synchronization. Especially, noises and spurious generation were studied. This allowed us to define the specifications of the VCRO (Voltage Controlled Ring Oscillator) that has been developed next in 28nm FDSOI technology. Several topologies were developed and integrated on a test-chip. The first measurement results were encouraging, but have to be completed. Indeed, the phase noise measurements are very complicated due to the pushing (about 5 GHz/V). The moderate power consumption remains lower than 0.8 mW and the die area is around 600 µm2.
In addition, we proposed a topology of the Phase Locked Loop using the previous VCRO and phases resynchronization. Each block was designed and optimized at the circuit level. Finally, a self-calibration scheme, using back-gate auto-biasing of complementary logic cells, is proposed

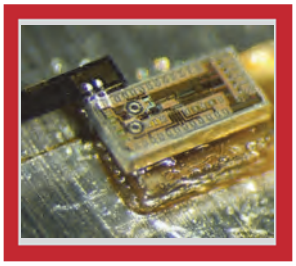


**Remi BONNARD**

*Université de Strasbourg (France)*

## BURST CMOS IMAGE SENSOR WITH ON-CHIP ANALOG TO DIGITAL CONVERSION.

This thesis starts by a review of the burst imaging and the 3D integration technologies. Two architectures of 3D burst image sensors are then proposed. The first architecture stores the images in analog form while the other stores the images in digital form. Models have been provided to assess the performances of those architectures in a 3D integration scheme. It appears that the digital storage architecture maximizes the number of images per burst. However, this analysis also highlights a high power consumption of the digital storage architecture. Then, a thermal study is presented which enables identifying the operating limits of the sensor. The rest of this work is focused on the design of pixels for the digital storage architecture. Different pixels are proposed to improve sensitivity and reduce power consumption. A test chip is presented as well as its electrical and optical performances. Finally, this work presents a prototype of burst image sensor integrated with a 3D technology.

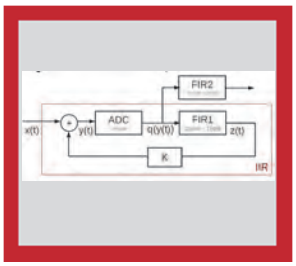



## Robert POLSTER

*Université de Paris Sud (France)*


### ARCHITECTURE OF SILICON PHOTONIC LINKS.

Future high performance computer (HPC) systems will face two major challenges: interconnection bandwidth density and power consumption. Silicon photonic technology has been proposed recently as a cost-effective solution to tackle these issues. Several works have demonstrated the possibility of integrating all elements needed for the realization of short optical links on a silicon substrate. The first contribution of this thesis is the optimization of a silicon photonic link for highest energy efficiency in terms of energy per bit. The optimization provides energy consumption models for the laser, the de- and serialization stage, the ring resonator as modulator and supporting circuitry, the receiver front-end and the decision stage. The optimal bit rate is found in the range between 8 Gbps and 22 Gbps, depending on the used CMOS technology. To apply the results of the optimization an 8 Gbps optical link receiver was designed and fabricated. It uses time interleaving to reduce the needed clock speed and elevate the need of a dedicated deserialization stage. The front-end was adapted for a wide dynamic input range, including a fast mechanism to find the optimal threshold voltage to distinguish ones from zeros. Furthermore, optical clock channels were explored. Using silicon photonics a clock can be distributed to several processors with very low skew. The thesis contributes to this by presenting two novel optical clock receivers.



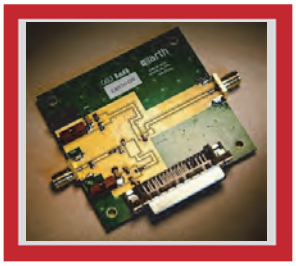


## Alin RATIU

*INSA de Lyon (France)*


### CONTINUOUS TIME SIGNAL PROCESSING FOR WAKE-UP RADIOS.

Wake-Up Receivers (WU-RX) have been recently proposed as candidates to reduce the communication power budget of wireless networks. Existing WU-RXs achieve very high sensitivities for power consumptions below 50µW but severely degrade their performance in the presence of out-of-band blockers. We attempt to tackle this problem by implementing an ultra low power, tunable, intermediate frequency filtering stage. We used Continuous Time Digital Signal Processing (CT-DSP) which offers an excellent scalability of the power consumption with respect to the characteristics of the input signal. A CT-DSP chain can be divided into a CT-ADC and the CT-DSP itself. The CT-ADC is based on a novel, delta modulator-based architecture which achieves a very low power consumption. Moreover, the CT nature of the ADC means that it does not do any sampling in time, hence no anti-aliasing filter is required. The proposed ADC requires only 24µW to quantize signals in the [10MHz 50MHz] bandwidth for an SNR between 32dB and 42dB.
The architecture of the CT-DSP which is divided into a CT-IIR and a CT-FIR. The CT-IIR is implemented by placing a standard CT-FIR in a feedback loop around the CT-ADC. It allows to cancel out some frequencies from the CT-ADC input (corresponding to out-of-band interferers) while boosting the power of the useful signal. The CT-DSP consumes around 100µW while achieving more than 40dB of out-of-band rejection; for a bandpass implementation, a 2MHz passband can be shifted over the entire ADC bandwidth.

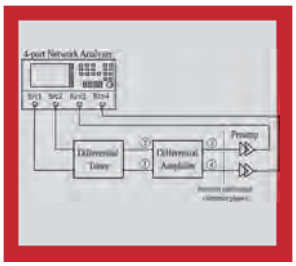


## Gauthier TANT

*Université Grenoble Alpes (France)*

### HIGH EFFICIENCY RECONFIGURABLE RF POWER AMPLIFIERS IN CMOS SOI TECHNOLOGY FOR MULTI STANDARD APPLICATIONS.

This work focuses on the study and integration of a reconfigurable multi-mode multi-band power amplifier (MMPA) supporting 2G/3G/4G at several frequency bands in SOI CMOS 130nm technology. The original solution presented in this thesis is a significant step toward the integration of MMPA. A study on PA efficiency under 3G / 4G modulated signals is also presented by comparing load and supply modulation PA architectures. A design methodology based on the study of different operating classes is proposed, which allows pre-sizing of power cells and optimal load impedance determination for high efficiency reconfigurable PA design. This led to the implementation of PA demonstrators integrated in SOI CMOS 130nm technology. One is a two stage reconfigurable MMPA operating from 700MHz to 900MHz and supporting saturated and linear modes. Tunable matching networks based on switched capacitor arrays allow performance optimization according to the mode and band. The measured prototype delivers up to 35dBm of output power in saturated mode with more than 58% efficiency. In linear mode, the measured output power exceeds 30dBm with efficiency higher than 47%. The second demonstrator is a passive load modulation PA architecture. It includes a SOI LDMOS power cell and a tunable matching network that allows presenting an optimal load trajectory to the PA in order to maximize its back-off efficiency.

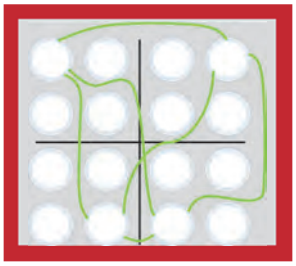


## Yogadissen ANDEE

*Université de Lille (France)*

### ADVANCED TECHNIQUES FOR NOISE FIGURE AND NOISE PARAMETERS MEASUREMENTS OF DIFFERENTIAL AMPLIFIERS.

Differential circuits have major advantages over single-ended circuits regarding immunity to common-mode noise, double voltage swing and reduction of even-order distortion. This creates a need for the development of techniques for differential noise figure measurement. His work shows that the noise figure of a 4-port circuit depends on the correlation of the noise waves at the output ports. Thus, we propose an original technique for measuring this correlation. It makes use of a hybrid coupler connected to the differential DUT. The correlation is determined by using two configurations of connection between the DUT and the coupler. A rigorous and general technique for the noise figure measurement of differential amplifiers is developed from this approach. Next, we propose an original approach where no extra coupler is required. A study of the structure of a differential amplifier is performed where an expression of the correlation is calculated in terms of output noise powers and of the 4-port S-parameters. A fast and functional measurement technique using this method is developed on a 4-port network analyzer. This coupler-free approach is extended to the measurement of the noise parameters of differential amplifiers. The noise parameters are determined from differential source-pull measurements using a differential impedance tuner.

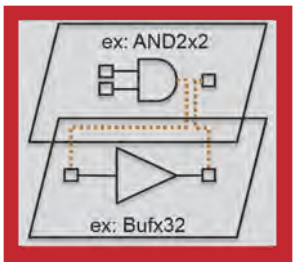


## Bartosz BOGUSLAWSKI

*Université Européenne de Bretagne (France)*

### DYNAMIC POWER MANAGEMENT OF MPSOC USING NETWORKS OF NEURAL CLIQUES.

The challenge of combining high-performance and energy efficiency is a strong demand for today's electronics applications. This objective can be reached with Multiprocessor System-on-Chip (MPSoC) platforms that provide high-level of adaptability, performance, reliability and energy efficiency. Nevertheless, they have to be accompanied with a power management decision unit that continuously adapts their operation.
This leads us to neural networks that can provide comparable properties to those of the cerebral cortex. We use an associative memory that relies on neural cliques to store information and sparse activity to retrieve it. We analyze these networks in power management applications using real-world data. We show that in order to be compatible with real-world applications, the initially proposed model of networks of neural cliques has to be improved. We propose several methods to do so within which the most efficient relies on a proposed concept of twin neurons. Sparse activity in these networks opens opportunities for low-power implementations. We show that neural cliques are more energy-efficient than power management solutions relying on game theory or Content-Addressable Memory (CAM). Furthermore, we propose a novel 3D interconnect approach for high-performance neural cliques' implementation. We show important gains in terms of total interconnect length and power consumption.

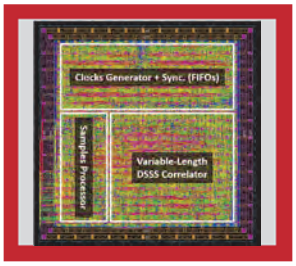



## Hossam SAHRAN

*Université Grenoble Alpes (France)*

### DESIGN METHODOLOGY AND TECHNOLOGY ASSESSMENT FOR HIGH-DENSITY 3D TECHNOLOGIES.

Scaling limitations of advanced technology nodes are increasing and interconnect parasitics are becoming more dominant. This has led to an increasing interest in 3D technologies to overcome such limitations. 3D technologies vary according to the fabrication process creating a wide spectrum including Through-Silicon-VIA (TSV), Copper-to-Copper (Cu-Cu) and Monolithic 3D (M3D). The high vertical interconnect density of Monolithic 3D creates new design paradigms.
In this context this thesis brings three main contributions: a new 3D standard cell approach named '3D Cell-on-Buffer' (3DCoB). 3DCoB cells are created by splitting 2D cells into functioning gates and driving buffers stacked one over the other. Simulation results show gain in timing performances compared to 2D. By applying an additional multi-VDD low-power approach, iso-performance power gain has been achieved.
Cell-on-cell approach has been explored. A partitioning methodology is needed to distribute cells between different tiers. A physical-aware partitioning methodology has been introduced which improves power-performance-area results comparing to the state-of-the-art partitioning techniques.
Finally a full high-density 3D technology assessment study has been done to explore the trade-offs between different 3D technologies, block complexities and partitioning methodologies

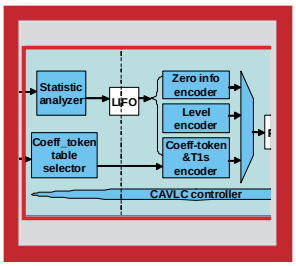



## Vincent LENOIR

*Université Grenoble Alpes (France)*

### LOW-POWER ADAPTIVE ARCHITECTURES FOR WIRELESS COMMUNICATIONS.

In the context of the Internet of Things (IoT), the energy efficiency is mandatory. However, traditional design paradigm implies a highly constrained sizing to be compatible with the worst operating conditions, even if it is not effective most of the time. It's typically the case with wireless communications since the radio channel is a medium characterized by a strong variability due to propagations effects and interferences. Thus, this study focused on the design of a communication chain whose link budget can be dynamically tuned depending on the actual signal attenuation, in order to reduce the system power consumption. This thesis has contributed to the design of a self-adaptive IEEE 802.15.4 receiver, by proposing both a reconfigurable digital baseband architecture and an automatic control method of the operating mode. More precisely, the work relied on two approaches, the compressive sampling and the partial sampling, to reduce the data's size to process, decreasing the internal activity of arithmetic operators. This solution, offers a low hardware overhead compared to conventional architecture of only 23,4k gates. Thanks to the circuit that has been developed, the power consumption for a packet demodulation is estimated to 278 µW when the baseband is fully activated. It can however be gradually scaled down to 119 µW, corresponding to a sensitivity reduction of 10 dB. Thus, the proposed digital baseband and its control loop save 30 % of energy in average in a typical use case.

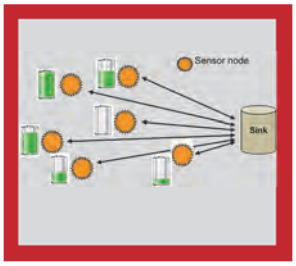



## Ngoc-Mai NGUYEN

*Université Grenoble Alpes (France), Vietnam National University.*

### POWER OPTIMIZATION STRATEGIES WITHIN A H.264 ENCODING SYSTEM-ON-CHIP.

Video codec chips are used in various applications ranging from video conferencing, security and monitoring systems, to entertainment applications. To meet the performance and power consumptions constraints encountered for mobile applications, video codecs are favorably implemented in hardware rather than in software. Nowadays, one of the most efficient standard is the H.264 Advanced Video Coding (H.264/AVC). This work has been conducted in the framework of the VENGME H.264/AVC hardware encoder design. The Entropy Coder and bytestream Network Abstraction Layer data packer (EC-NAL) module has been designed during this PhD thesis, taking into account and combining several state-of-the-art solutions to minimize the power consumption. From simulation results, it has been seen that the EC-NAL module presents better power figures than the previous solutions. Then, the encoder architecture has been analyzed and power estimations at RTL level have been performed to extract the platform power figures. From these power figures, it has been decided to implement power control on the EC-NAL module. In this latter a Dynamic Frequency Scaling (DFS) approach based on the control of the FIFO occupancy level has been proposed. The control law has been implemented in hardware and the closed-loop system stability has been studied. Simulation results show the effectiveness of the proposed DFS strategy that should be extended to the whole H.264 encoder platform

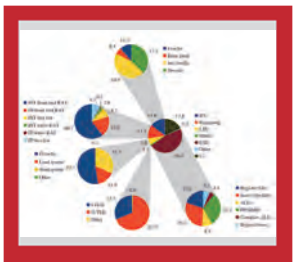



## Olesia MOKRENKO

*Université Toulouse III - Paul Sabatier (France)*

### ENERGY MANAGEMENT OF A WIRELESS SENSOR NETWORK AT APPLICATION LEVEL.

Energy is a key resource in Wireless Sensor Networks (WSNs), especially when sensor nodes are powered by batteries. This thesis investigates how to save energy for the whole WSN, at the application layer, thanks to control strategies, in real time and in a dynamic way. The first energy management strategy investigated is based on Model Predictive Control (MPC). The choice of MPC is motivated by the global objectives that are to reduce the energy consumption of the set of sensor nodes while ensuring a given service, named mission, for the sensor network. Moreover, a set of constraints on the binary control variables and on the sensor modes must be fulfilled. The second energy management strategy at the application layer is based on a Hybrid Dynamical System (HDS) approach. This choice is motivated by the hybrid inherent nature of the WSN system when energy management is considered. The hybrid nature basically comes from the combination of continuous physical processes, namely, the charge and discharge of the node batteries; while the discrete part is related to the change in the functioning modes and to the Unreachable condition of the nodes. The proposed strategies are evaluated and compared in simulation on a realistic test-case. They have also been implemented on a real test-bench and the results obtained have been discussed



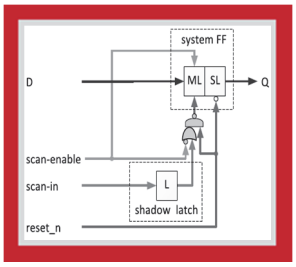
## Fernando ENDO

*Université Grenoble Alpes (France)*

### ONLINE AUTO-TUNING FOR PERFORMANCE AND ENERGY THROUGH MICRO-ARCHITECTURE DEPENDENT CODE GENERATION.

Energy consumption is limiting the performance growth experienced in the last decades. New architectural and micro-architectural designs improve the energy efficiency of hardware, thanks to hardware specialization and core heterogeneity. Hence, software needs to cope with the lack of performance portability.

This thesis proposes a run-time auto-tuning framework for embedded systems. The proposed framework can both adapt code to a micro-architecture unknown prior compilation and explore auto-tuning possibilities that are data-dependent. The methodology and proof of concept were developed around two kernel-based applications, extracted from the PARSEC 3.0 suite and its hand vectorized version PARVEC. In the favorable application, average speedups of 1.26 and 1.38 were obtained in real and simulated cores, respectively, going up to 1.79 and 2.53 (all run-time overheads included). This work also demonstrated through simulations that run-time auto-tuning of SIMD instructions to in-order cores can outperform the reference vectorized code run in similar out-of-order cores, with an average speed-up of 1.03 and an energy efficiency improvement of 39 %. The unfavorable application was chosen to show that the proposed approach has negligible overheads when better kernel versions cannot be found.

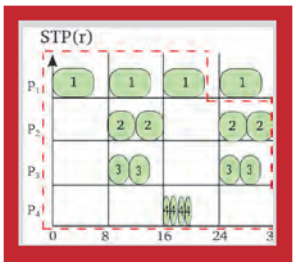



**Sébastien SARRAZIN**

*Université Paris 6 (France)*

## ON-LINE MONITORING AND TEST OF PROCESSORS FOR EMBEDDED APPLICATIONS.

Slack-time reduction is a way to improve the performance of synchronous sequential circuits.
In the presence of circuit wear-out, supply voltage fluctuations and temperature variations, aggressive slack-time reduction can be achieved based on adaptive voltage and frequency scaling with feedback from in-situ slack-time monitoring.
The first contribution of this work consist of a new shadow-scan solution which facilitates the implementation of faster scan Flip-Flops (FFs), enables in-situ slack-time monitoring and can be transparently handled by commercial tools for automated scan stitching and automated test pattern generation. A natural approach is to place in-situ slack-time monitors close to all sequential elements with incoming timing-critical paths or susceptible to become timing-critical due to wear-out or manufacturing variability. In latency constrained circuits with large ratios of timing-critical paths, this methodology may result in large area overheads and minor power improvements.
The second contribution of this work is an evaluation methodology of the monitoring quality delivered by a set of FFs. This methodology estimates monitor activation probabilities based on which two evaluation metrics are provided.
Finally, based on these metrics, it is shown that the monitoring quality can be significantly improved if the size of the detection window of each in-situ slack-time monitor is correlated to the slack-time of the monitored timing-critical paths.

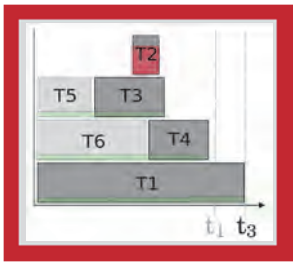



**Amira DKHIL**

*Université de Toulouse (France)*

## MEAN AND WORST CASE EXECUTION TIME EVALUATIONS FOR MPSOC TARGETS.

The increasing complexity and concurrency in digital multi-processor systems requires a design flow that covers different abstract layers.
In this PhD thesis, we take dataflow as the basic model of computation, as it fits the streaming application domain. Cyclo-Static Dataflow (CSDF) is particularly interesting because this variant is one of the most expressive dataflow models while still being analyzable at design time. Describing the system at higher levels of abstraction is not sufficient, e.g. dataflow have no direct means to optimize communication channels generally based on shared buffers.
This thesis proposes a design flow that meets these challenges, while enabling features such as temporal isolation and taking into account other challenges such as predictability and ease of validation. To this end, we propose a new scheduling policy noted Self Timed Periodic (STP), which is an execution model combining Self-Timed Scheduling (STS) with periodic scheduling. We aim at modeling four types of STP scheduling based on two different granularities naturally offered by Cyclo-Static Dataflow model: 1) Coarse-Grained description which considers instances of actors, and 2) Fine-Grained description which considers smaller components (i.e., sub-tasks of actors) the actors are composed of. In our experiments, the results are very promising, e.g., 86.4 % of improvement for constrained STP compared to the SPS latency. The solution proposed in this thesis, is a new analytical communication performance model, for the Kalray MPPA-256 architecture, that models computation, communication and arbitration.




## Alexandre AMINOT

*Université de Grenoble (France)*

### DYNAMIC AND FLEXIBLE TASK PLACEMENT ON FUNCTIONALLY ASYMMETRIC MULTI-CORE PROCESSOR

This thesis focus on the emerging functionally asymmetric multi-core processor (FAMP) architectures. These architectures are characterized by non-uniform implementation of hardware extensions in the cores (ex. Floating Point Unit (FPU)). The thesis investigates the nature of the use of extensions in state-of-the-art's applications and compares various existing methods. To optimize the tasks mapping and increase efficiency, we propose a dynamic solution at scheduler level, called relaxed scheduler. We compare multiple extension management solutions at different levels of temporal granularity of action, to understand the limitations of these solutions and thus define at which level we must act. We propose a solution for estimating online performance degradation to run a task on a core without a given extension. Placing a task on a core with no extension can have important consequences for energy and performance. But to date, there is no way to estimate this potential degradation. We propose a relaxed scheduler, based on our degradation estimation model, which maps the tasks on a set of heterogeneous cores effectively. Our experiments with simulators have shown that the scheduler can achieve a significantly higher mapping flexibility with a performance degradation of less than 2 %. Compared to a symmetrical multi-core, our solution enables an average energy gain at core level of 11 %. These results are very encouraging and contribute to the development of a comprehensive FAMP platform.


## Aziz DZIRI

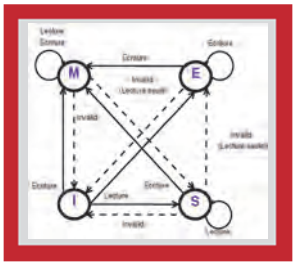
*Université Clermont-Ferrand (France)*

### VISUAL MULTI-OBJECT TRACKING IN A NETWORK OF EMBEDDED SMART CAMERAS.

Multi-object tracking constitutes a major step in several computer vision applications. In this thesis, we designed a multi-object tracking system that achieves real time processing on a low cost and a low power embedded smart camera.

The tracking pipeline is composed of a detection module based on a background subtraction method and on a tracker using the probabilistic Gaussian Mixture Probability Hypothesis Density (GMPHD) filter. The background subtraction, we developed, is a combination of the segmentation resulted from the Zipfian Sigma-Delta method with the gradient of the input image. The results show that besides the low complexity of the pipeline, the tracking quality of our method is close to the state of the art methods. A frame rate of 15 – 30 fps was achieved on the smart camera depending on the image resolution.

In the second part of the thesis, we designed a distributed approach for multi-object tracking in a network of non-overlapping cameras. The approach was developed based on the fact that each camera in the network runs a GMPHD filter as a tracker. Our approach is based on a probabilistic formulation that models the correspondences between objects as an appearance probability and space-time probability. Several simulations were performed to validate the approach. The obtained results are promising for the use of this approach in a real network of smart cameras.

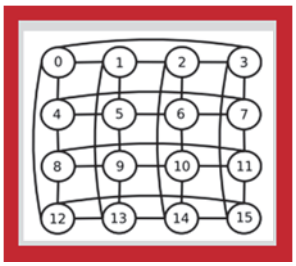



## Safae DAHMANI

*Université de Bretagne Ouest (France)*

### DATA CONSISTENCY MODELS AND PROTOCOLS, DECISION AND OPTIMIZATION AT COMPILE TIME FOR MASSIVELY PARALLEL ARCHITECTURES.

In the context of manycore architectures, providing a shared memory system has to rely on efficient hardware and software mechanisms and data consistency protocols. Numerous works explored consistency mechanisms designed for highly parallel architectures. They lead to the conclusion that there won't exist one protocol that fits to all applications and hardware contexts. In this work we propose a multi-protocol compilation toolchain, in which shared data of the application can be managed by different protocols. Protocols are chosen and configured at compile time, following the application behavior and the targeted architecture specifications.
We first use a traffic analytical model which gives some Network on Chip communication statistics but no timing information. Therefore, we propose two cycle- based evaluation models that provide more accurate performance metrics while taking into account contention effect due to the consistency protocols communications. We also propose a cooperative cache consistency protocol improving the cache miss rate by sliding data to less stressed neighbors. An extension of this protocol is proposed in order to dynamically define the sliding radius assigned to each data migration. This extension is based on the mass-spring physical model. Experimental validation of different contributions uses the sliding based protocols versus a four-state directory-based protocol.




## Karl-Eduard BERGER

*Université Versailles St- Quentin (France)*

### MAPPING OF LARGE TASK NETWORK ON MANYCORE ARCHITECTURE.

This PhD thesis is devoted to the study of the mapping problem related to massively parallel embedded architectures. This problem arises from industrial needs like energy savings, performance demands for synchronous dataflow applications. In this thesis, tasks are organized in communication networks, modeled as graphs. In order to address this problem, two heuristics are proposed. The first heuristic, denoted as Task-wise Placement, places tasks one after another using a notion of task affinities. The second algorithm, named Subgraph wise Placement, gathers tasks in small groups then place the different groups on processors using a notion of affinities between groups and processors. These algorithms are tested on tasks graphs with grid or logic gates network topologies. The results obtained are then compared to an algorithm from the literature.
A progressive construction heuristic named Regret Based Approach, based on game theory, is proposed. This heuristic maps tasks one after another. The process of task selection is based on a notion of regret, present in game theory. The task with the highest value of regret for not placing it, is pointed out and is placed in priority. In order to check the strength of the algorithm, many types of task graphs (grids, logic gates networks, series parallel, etc.) with various size are generated. Tasks and edges weights are randomly chosen using a bimodal law parameterized in order to have similar values than industrial applications. Finally the results obtained are compared to the Task Wise placement, especially adapted for non-unitary values.

# Notes

# Greetings


## EDITORIAL COMMITTEE

Marc Belleville

Christian Gamrat

Hélène Vatouyas

Jean-Baptiste David

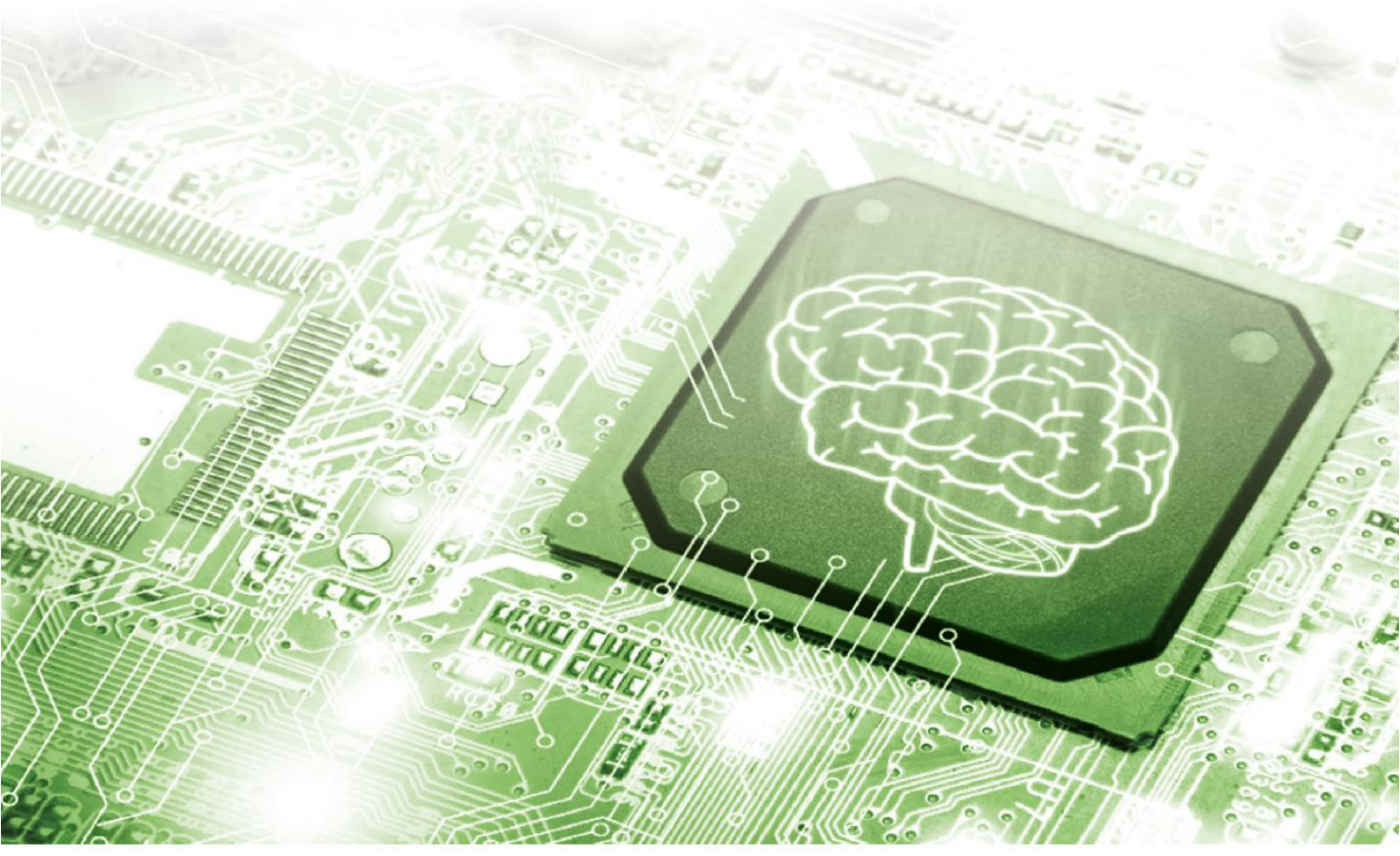
## READER

Thierry Collette

## GRAPHIC ART

Eve Issartel (Design by Eve)

Hélène Vatouyas



# ARCHITECTURE AND IC DESIGN, EMBEDDED SOFTWARE

## Contacts

**Thierry Collette**
Head of Architecture, IC Design
& Embedded Software division
thierry.collette@cea.fr

**Marc Belleville**
Research director, Scientific director-Grenoble
marc.belleville@cea.fr

**Christian Gamrat**
International Expert, Scientific director-Saclay
christian.gamrat@cea.fr

**Marc Duranton**
International Expert, European Projects Coordinator
marc.duranton@cea.fr

**Denis Dutoit**
Sr Strategic Marketing Manager,
Advanced Computing Technology & Solutions
denis.dutoit@cea.fr

**Michel Durr**
Sr Business Development Manager, Analog & Wireless IC
michel.durr@cea.fr

**Ali Erdengiz**
Sr Business Development Manager, Silicon Impulse
ali.erdengiz@cea.fr

**Olivier Heron**
Sr Business Development Manager, Dependable systems
olivier.heron@cea.fr

**Marie-Sophie Masselot**
Sr Business Development Manager, IoT
marie-sophie.masselot@cea.fr

**Bénédicte Robin**
Sr Business Development Manager, EDA & Digital IC
benedicte.robin@cea.fr

**Frédéric Surleau**
Sr Business Development Manager,
Embedded Systems & Computing
frederic.surleau@cea.fr

**Michael Tchagaspanian**
Sr Business Development Manager,
Smart Imaging & Sensors Solutions
michael.tchagaspanian@cea.fr

**Sandrine Varenne**
Sr Business Development Manager,
Embedded Systems & Software
sandrine.varenne@cea.fr

**Leti, technology research institute**
Commissariat à l'énergie atomique et aux énergies alternatives
Minatec Campus | 17 rue des Martyrs | 38054 Grenoble Cedex 9 | France
www.leti.fr