ARCHITECTURE, IC DESIGN AND EMBEDDED SOFTWARE
Committed to Innovation, Leti Creates Differentiating Solutions for its Industrial Partners.

Leti is a research institute of CEA Tech and a recognized global leader in miniaturization technologies. Leti’s teams are focused on developing solutions that will enable future information and communication technologies, health and wellness approaches, clean and safe energy production and recovery, sustainable transport, space exploration and cybersecurity.

For 50 years, the institute has built long-term relationships with its industrial partners, tailoring innovative and differentiating solutions to their needs. Its entrepreneurship programs have sparked the creation of 64 startups. Leti and its industrial partners work together through bilateral projects, joint laboratories and collaborative research programs. Leti maintains an excellent scientific level by working with the best research teams worldwide, establishing partnerships with major research technology organizations and academic institutions. Leti is also a member of the Carnot Institutes network*.

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

CEA Tech is the technology research branch of the French Alternative Energies and Atomic Energy Commission (CEA), a key player in research, development and innovation in defense & security, nuclear energy, technological research for industry and fundamental physical and life sciences.

www.cea.fr/english

Leti at a glance

€315 million budget
800 publications per year
ISO 9001 certified since 2000

Founded in 1967
Based in France (Grenoble) with offices in the USA (Silicon Valley) and Japan (Tokyo)
350 industrial partners

1,900 researchers
2,760 patents in portfolio
91,500 sq. ft. cleanroom space, 8” & 12” wafers
64 startups created
Specialized in digital systems design, List aims at achieving excellence in technological development, helping industrial partners to create value and innovation.

List is a CEA Tech institute, develops and implements cutting-edge technological research dedicated to intelligent digital systems. Its R&D programs, all with potentially major economic and social implications, focus on advanced manufacturing (robotics, virtual & augmented reality, non destructive testing, vision), embedded systems (computing architectures, software and systems engineering, security & safety), and ambient intelligence (sensors, instrumentation & metrology, communication & sensory interfaces, data processing & multimedia).

CEA List benefits from a portfolio of patented enabling technologies and 9 R&D platforms with the latest equipment, led by high-level experts. Capitalizing on its expertise, from Embedded systems design to Cybersecurity, Software & systems engineering, Ambient intelligence, Non destructive testing, Nuclear instrumentation for energy, Virtual reality, Collaborative robotics, DOSEO Radiotherapy & imaging, the institute has collaborated with more than 100 different companies, among which half are SMEs.

List at a glance

€80 million budget
300 publications per year
ISO 9001 certified since 2003

Founded in
2001

Based in
France (Paris Saclay)
with offices in the
USA (Silicon Valley)
and Japan (Tokyo)

800 researchers
486 patents in portfolio
220 industrial partners
20 startups created
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.
ARCHITECTURE, IC DESIGN AND EMBEDDED SOFTWARE
EDITO 07

KEY FIGURES 08

SCIENTIFIC ACTIVITY 09

01 / Hardware software integration and methodologies 11
02 / Computing solutions 26
03 / Wireless solutions 37
04 / Sensors & Energy 45
05 / Secured and Reliable systems 59
06 / Emerging technologies & Paradigms 72
07 / PHD Degree awarded 84
I am now leading the Architecture, IC Design and Embedded Software Division. I would like to warmly thank Thierry Collette, the former Head of this Division for the work done during the last 7 years.

Information processing technologies and communications systems are everywhere, from mobile devices, cars, household appliances to portable medical devices. This drives the widespread development of emerging trends like the Internet of Things (IoT), Smart Cities, Autonomous Vehicles, and Smart Manufacturing — helping billions of objects around the world to connect securely and efficiently.

In particular, the rise of the IoT, cyber-physical systems (CPS) and artificial intelligence has brought the need to design novel smart devices capable of computing, sensing and communicating with extreme energy efficiency while also generating huge amounts of data in the cloud that require extensive storage and processing capacity. From devices to the infrastructure, real-time performance, fast data processing, low energy consumption, security and reliability are critical for IoT applications.

These technology challenges require a holistic approach to hardware and software. Hence, combining research between the two well-known institutes, LETI and LIST, is increasingly important. The LETI brings advanced know-how in integrated circuit design and testing, and the LIST brings expertise in architecture and embedded software development. The combination of both expertise brings unique advantages, enabling the development of smarter and more efficient solutions.

In 2018, based on our expertise and background, we launched several aggressive programs targeting 5G, cyber-security, embedded artificial intelligence, performance computing and CPS. Thus, from smart sensors and wireless connectivity to computing solutions, artificial intelligence, cyber-physical and reliable systems, in this report you will find our latest innovations that address these upcoming challenges.

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

Jean-René Lèquepeys / Thierry Collette
**Key figures**

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

198 Permanent researchers
51 Students and post-docs

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

€39M Budget
84% Funding from contracts

50 granted patents
52 papers, journals & books
175 conferences & workshops
Publications

Prize and awards
• First price at 2017 iDASH workshop on Genome Privacy & Security Competition, Sergiu Carpov et al., “Software Guard Extension (SGX) based whole genome variants search”, http://www.humangenomeprivacy.org/2017/
• Best paper award, L. Alacoque, “A 128x128, 34um pitch, 190mK NEDT, TECless Uncooled IR Bolometer image sensor with column-wise processing”, Electronic Imaging 2017.
• Student Award, E. Chabchoud et al., “High Temperature, Time Domain Sensor Interface based on Phase Shifter”, HITEN 2017.

Experts
49 CEA experts: 5 research directors, 2 international experts.
14 researchers with habilitation qualification (to independently supervise doctoral candidates).

Scientific committees
19 members of Technical Programs and Steering Committees in major conferences: ISSCC, VLSI-Circuits, ESSCIRC, DATE, ASP-DAC, ISPLED, HIPEAC, ICCAD, IJCNN, ESWeek, PATMOS, MPSoC.

Conferences and Workshops organizations

International Collaborations
Collaborations with more than 20 universities and institutes worldwide, e.g. Stanford University (USA), Massachusetts Institute of Technology (USA), Columbia University (USA), Cornell 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)…
HARDWARE SOFTWARE INTEGRATION AND METHODOLOGIES

- Sensor fusion
- Autonomous control
- Code generation and associated frameworks
- Smart environment control
- Mapping, Scheduling, Optimization
- Manycore arbitration and scheduling
EVALUATION OF OCCUPANCY GRID RESOLUTION THROUGH A NOVEL APPROACH FOR INVERSE SENSOR MODELING

RESEARCH TOPIC:
Occupancy grid, SigmaFusion®, inverse sensor model, grid resolution, sensor precision

AUTHORS:
Roxana Dia, Julien Mottin, Tiana Rakotovao, Diego Puschini, Suzanne Lesecq

ABSTRACT:
Several robotic applications imply motion in complex and dynamic environments. Occupancy Grids model the surrounding environment by a grid composed of a finite number of cells. The probability whether a cell is occupied or empty is computed and updated iteratively based on sensor measurements by considering their uncertainty through probabilistic models. Even if Occupancy Grids have been widely used in the state-of-the-art, the relation between the cell size, the sensor precision and the inverse sensor model is usually neglected. Here, a methodology to build the inverse probabilistic model for single-target sensors is proposed. Then, the impact of the variation in the sensor precision and the grid resolution on the inverse sensor model is analysed.

Context and Challenges
Modern mobile robots evolve in complex a priori unknown environments. In these scenarios, the surrounding obstacles are perceived thanks to range sensors such as LiDARs, sonars, radars, Time-of-Flight cameras, vision sensors. However, external conditions, the nature of observed obstacles or even imperfections in the sensor design introduce noise in the measurements. Common approaches translate this uncertainty into a probabilistic distribution called the Sensor Model (SM).

The SM gives the likelihood of a specific sensor value knowing a property of the environment such as the distance to the nearest object. Hence, it creates a link between the physical world and the sensor output. It gives information on the position of the obstacles rather than the occupancy of a specific region. To deduce occupancy probabilities, OGs have been introduced by Elfes in 1989. A formal methodology that produces an Inverse Sensor Model (ISM) from a SM and a specific partition of the space. This ISM relates the sensor measurements to the occupancy probability for all cells. This approach naturally propagates the sensing precision to the occupancy evaluation. Uncertainty captured by the SM is converted into an equivalent uncertainty for the ISM.

Main Results
The contributions of this study [1] are threefold.

Firstly, a new methodology for the Inverse Sensor Model construction is proposed. It introduces no additional approximation compared to the original OG formulation. This methodology can be applied to sensors that exhibit single-target behaviour (Fig. 1), i.e. whose output is caused by a unique obstacle assumed the nearest one, such as LiDARs and depth pixels of time-of-flight cameras. For these sensors, the proposed method offers a computational cost proportional to the number of cells in the grid (linear complexity). Since the new method naturally fits in the original OG framework proposed by Elfes, it also propagates the original precision of the sensor to the ISM.

Secondly, by applying our methodology to a LiDAR, we analyse the impact of the grid resolution and the sensor precision on the occupancy estimation.

Finally, a novel method (Fig. 2) is proposed to properly choose the grid resolution. This method allows achieving a predefined maximum occupancy probability in the ISM taking into account the sensor precision and its output. It is therefore useful when modelling the ISM, because it allows propagating the variations in the sensor precision and the cells size in the grid to the occupancy evaluation.

Perspectives
Since autonomous mobile robots require to use a wide type of sensors, including multi-target sensors, additional work remains to be done in this regard. We look forward to extend our study in the future and encounter the treated topics in the case of multi-target sensors.

Fig. 1: Single target sensor behaviour

Fig. 2: Link between the grid resolution, maximum occupancy probability and sensor precision.

RELATED PUBLICATIONS:
**INSPEX: AN INTEGRATED SMART SPATIAL EXPLORATION SYSTEM**

**RESEARCH TOPIC:**
Environment perception, spatial exploration, integrated system, embedded, low-power, portable, wearable, health.

**AUTHORS:**

**ABSTRACT:**
The INSPEX H2020 project (grant nb. 730953) main objective is to integrate automotive-equivalent spatial exploration and obstacle detection functionalities into a portable/wearable multi-sensor, miniaturised, low power device. The INSPEX system will be used for 3D real-time detection, location and warning of obstacles under various environmental conditions in indoor and outdoor environments with static and mobile obstacles. Potential applications range from safer human navigation in reduced visibility conditions, small robot/drone obstacle avoidance systems to navigation for the visually/mobility impaired, this latter being the primary use-case considered in the project.

**SCIENTIFIC COLLABORATIONS:** CSEM, CIT, GoSense, U. of Manchester, U. of Namur, SensL, STMicroelectronics, Tyndall-UC.

**Context and Challenges**
A large amount of work has been recently done on obstacle avoidance systems for autonomous vehicles. To detect different types of obstacles across the full range of possible lighting and weather conditions, the obstacle detection systems usually combine multiple subsystems including Lidar, radar, IR and optical. Data from these subsystems is merged with vehicle orientation and navigation subsystems. These systems are typically heavy, power hungry, and require high computational capabilities. INSPEX partners have asked themselves “what if we could miniaturise an obstacle detection system like this and reduce its power consumption so that it could be portable/wearable?” This is very challenging but it would open many new applications in e.g. assistive guidance for the visually impaired, human guidance in low visibility conditions (night, smoke, fog), obstacle detection for small or humanoid robots, and drone obstacle detection for obstacle avoidance procedures.

**Main Results**
INSPEX objective is to combine several range sensors with an IMU, environmental sensing, signal and data processing, wireless communications, energy management and user interface in a miniature, low power, small size, light weight device (Fig. 1) [1]. INSPEX first demonstrator targets the Visually Impaired and Blind (VIB) community.

The INSPEX system is split in three devices with a modular architecture that provides its versatility (Fig. 2) [2]. The mobile detection device will integrate several range sensor technologies (i.e. Lidar on chip, MEMS ultrasound, Ultra-Wide-Band (UWB) impulse Radar) brought to the project by partners and optimised in order to fulfil the system requirements. The choice for these range sensing technologies is conducted by the capability of the final mobile detection device to detect a large variety of obstacles (in shape, size, material, and colour) in different environmental conditions (temperature, humidity, luminosity, visibility) and particular situations (holes, stairs). The INSPEX system will function under various weather conditions (e.g. rain, snow) over a large temperature range (typ. -20°C to 40°C) but also in low visibility conditions (e.g. night, dust, smoke, fog). An environmental sensing module will help reconfigure the system depending on these environmental conditions.

**Fig. 2: INSPEX modular system architecture**

In parallel to sensor optimisation, the hardware and software architectures have been designed (Fig. 2) and the firmware has been defined.

**Perspectives**
An early prototype of the INSPEX system has been developed [3] to demonstrate the system functionalities. The second year of the H2020 project will be dedicated to the integration of these functionalities in a portable/wearable device compatible with the application requirements (in terms of size, power consumption, detection capabilities).

**RELATED PUBLICATIONS:**
DEVELOPMENT TOOLS FOR RULE-BASED COORDINATION PROGRAMMING IN LINC

RESEARCH TOPIC:
Coordination environment, Development tools, Distributed systems

AUTHORS:
M. Louvel, F. Pacull (Bag-Era), E. Rutten (INRIA), A. N. Sylla

ABSTRACT:
Development tools are essential to make coordination models and languages to be used by a broader range of developers. The set of design and debugging tools put in place with LINC, a coordination environment based on tuple spaces and implementing the chemical machine paradigm, allows developers not only to generate coordination rules that are validated with domain specific knowledge or by the use of formal methods, but also to monitor, analyze and update a running distributed application. To the best of the authors’ knowledge, there exist no equivalent set of development tools for tuple space programming environments.

SCIENTIFIC COLLABORATIONS: INRIA

Context and Challenges
Coordination models and languages have been around for decades, but there is a lack of development tools that can better capture the coordination models instead of relying on the understanding of software engineers. There is also a need for debugging tools and tools to monitor, introspect and update running applications without restarting them.

LINC coordination environment fulfills these needs and provides a set of tools that can be of interest to other tuple spaces based coordination environments.

Main Results
LINC coordination environment [1] uses distributed tuple spaces, a rule based language and transactional reactions implementing the chemical machine to provide applications with transactional guarantees that ensure their reliability at run-time. It is used for distributed applications, possibly large scale, that may include cyber-physical systems or the Internet of Things (IoT).

LINC coordination environment includes rule generation tools as well as analysis and debugging tools [2].

Rule generation tools automatically validate and generate correct rules based on domain specific knowledge or formal models such as Discrete Controller Synthesis (DCS). This allows developers that are experts of their domain (and not experts in coordination models or distributed systems) to focus on the application logic and not on the writing of coordination rules, and avoids design errors like conflicts between rules, circularities and violations of applicative constrains.

Analysis and debugging tools offer runtime monitoring, run time debugging and analysis of the coordination rules executions. LINC provides a web interface (called monitor) that can be easily accessed from any web browser. It is used (i) to track and modify tuples, (ii) to start, stop and update coordination rules, (iii) to look into, make a search and filter the execution traces generated by the application, and complementary, (iv) to explore the coordination rules executions (in a data flow or global view of the coordination actions) to see where the application is blocked. As a result, the monitor interface provides a central point of view to easily access all the information of the coordination entities.

![Fig. 1. Monitor interface of LINC](image1)

![Fig. 2. Exploring coordination rules executions of LINC](image2)

All these tools have proven really useful, not to say essential, for developing and monitoring dozens of demonstrators in many areas (e.g. building automation, energy management, industrial IoT), especially when these applications are based on remote components distributed over European countries and across several Local Area Networks with their own global policy for security.

Perspectives
Future works will focus on extending analysis tools to allow re-executing an application with the same scheduling as the original execution, because rules executions are not deterministic.

RELATED PUBLICATIONS:
DESIGN FRAMEWORK FOR RELIABLE AND ENVIRONMENT AWARE MANAGEMENT OF SMART ENVIRONMENT DEVICES

RESEARCH TOPIC:
Smart environments, Reliability, Transactional middleware, Transition systems

ABSTRACT:
Smart environments are equipped with numerous distributed devices that are automatically controlled to achieve a set of objectives. Controlling such an environment is not an easy task due to the heterogeneity of devices, the inconsistencies that can result from communication errors or device failures, and the conflicting decisions including those caused by environment dependencies. To solve these challenges, a design framework based on the combination of the rule-based coordination environment LINC with a transition system is proposed. It avoids conflicts and ensures behavioral and transactional reliability of smart environments.

SCIENTIFIC COLLABORATIONS: INRIA

Context and Challenges
Existing solutions for the reliability of smart environments are based on methods (e.g. model checking) that require to manually program or model the behavior of the smart environment before it is verified. Moreover, they do not handle the inconsistencies due to communication errors and hardware failures.

Our design framework consists of an abstraction layer to manage the heterogeneity of the devices, a transactional execution mechanism to avoid inconsistencies, and an environment aware controller that, based on a generic model of the environment, makes appropriate decisions and prevents from conflicts.

Main Results
The framework described in [1] relies on (i) the transactional middleware LINC, used to build distributed applications for Wireless Sensor Networks and building automation, and (ii) the reactive language Heptagon/BZR, used to build reactive systems by means of automata, equations, and a set of system properties to be enforced through Discrete Controller Synthesis (DCS).

DCS enables the declarative management of the devices while avoiding objectives violations at compilation time. It produces a step function which ensures that the controlled system will never reach an undesired state. When invoked, this function computes and returns the appropriate commands to be executed by the system. To remain consistent with the controlled system, the step function must be called whenever an event occurs. This is achieved by LINC that keeps also the internal state of the automata consistent with the controlled system, even in case of communication errors or hardware failures.

Finally, the framework allows developers to deal with self-adaptive systems [2], where multiple autonomic loops can be composed into complex managers (parallel, coordinated parallel, hierarchic) and autonomic loops adaptation to the changing objectives in realistic smart environment can be managed through the automatic reconfiguration of the controllers.

The proposed framework has been illustrated on a case study in the context of building automation, using a general model of a room equipped with many (real or emulated) sensors and actuators. It has shown its ability to reliably achieve the target objectives (related to air quality, comfort, confidentiality and energy savings) whatever the selected scenarios.

Fig. 2. Demonstration of the framework with real devices

Perspectives
In the field of building automation, we will provide a Domain Specific Language (DSL) that enables building managers to easily specify their devices and the objectives to be achieved, and generate multiple autonomic loops from these specifications.

RELATED PUBLICATIONS:
Context and Challenges
Buildings consume more than 40% of the total primary energy resources throughout the world. Moreover, inefficiencies of the deployed sensing and control strategies cause energy waste that should be avoided by a better coordination among Building Automation Systems (BMS) and appropriate control approaches. The minimization of the energy consumed by buildings is essential for their sustainability. However, this minimization may badly affect the occupants’ comfort, e.g. by reducing (resp. increasing) the temperature in the building when the outside temperature is low (resp. high).

Indoor Environmental Quality (IEQ) is related to thermal aspects, Indoor Air Quality (IAQ), acoustic and visual (lighting) levels while humidity level also affects the comfort feeling. IEQ/IAQ is an active research area, from clinical and medical viewpoints, to control, but also building construction and retrofit, communication, etc. These research works are often related to energy efficiency and HVAC control.

Main Results
This study [1][2] presents an application of Model Predictive Control (MPC) to deal with thermal and CO2 concentration control in indoor environment, see Fig. 1.

Temperature T and CO2 concentration are contradictory by nature. The CO2 concentration can be easily decreased by the injection of fresh outdoor air inside the building using for instance forced or natural ventilation. The side effect of this injection is the decrease of the indoor temperature when the outdoor one is colder, leading to more heating. Both aspects are physically coupled and can be modelled using a MIMO model.

The proposed controller is applied to an open-space office, split in three zones, located in CIT (Fig. 2). A centralized MPC approach is first designed. It requires all data to be collected in a unique control point. For large buildings, and in presence of communication issues, this can lead to bad functioning of the control law. Thus, a distributed version of the controller, with loosely coupled areas, might be of interest to overcome these situations. Such a distributed approach is therefore proposed. This Distributed MPC implements a separate local controller for each zone.

Simulation results can be seen in [1][2]. They show that the distributed MPC solution achieves control performance quite close to the centralized version with less computing effort.

Perspectives
Both approaches are currently under implementation on the real testbed in the context of the H2020 TOPAs project (GA nb. 676760 https://www.topas-eeb.eu/).

RELATED PUBLICATIONS:
OPTIMIZATION OF CIRCUIT PARTITIONING FOR MULTIPLE FPGA PLATFORMS

RESEARCH TOPIC:
SoC prototyping, multi FPGA partitioning, hypergraphs, solvers.

AUTHORS:
J-P. Appéré, F. Galea, L. Zaourar

ABSTRACT:
System on Chip (SoC) design is complexifying as technology enables higher transistor resolutions and large numbers of components. For early testing, during the development phase of a SoC, it is usually implemented on a FPGA, which is a reconfigurable on-chip logical array.

Modern fully-featured SoC designs may be too large to fit in a single FPGA. In that case, it may be possible to make use of a multi-FPGA platform, where the design has to be partitioned in a way that the overall performance decrease is minimal.

Context and Challenges
FPGA netlist partitioning is a generalization of the hypergraph partitioning problem, in which the vertices of the hypergraph are groups of resources to be placed in one or the other partition, and the hyperedges correspond to the wires (or nets) connecting the groups of resources. Resources are the available basic cell types in the FPGA (Lookup table, flip-flop, block RAM, DSP, ...). Each resulting partition will be mapped into one of the FPGAs of the target multi-FPGA platform. A good partitioning minimizes a cost function which is the number of hyperedges passing from one partition to another. The constraint is that the partitions must not contain more of each resource than available in the FPGAs.

LocalSolver can take a long time finding a feasible solution, so we proceeded in a two-step procedure, which first solves the relaxed hypergraph partitioning problem using hMetis, then uses the obtained solution as a starting point in LocalSolver.

Graph partition problems fall under the category of NP-Hard problems, so the use of heuristics in order to find a sufficiently good solution on large-scale problems is required. Our specific mathematical model is presented in Fig.2.

Hypergraph partitioning methods are very well studied in the literature, but our specific multi-resource constraint makes the use of existing generic partitioners (such as hMetis [2]) not sufficient to find feasible solutions for our specific problem. However, solutions obtained using such solvers can be used as starting points to other solution methods.

Main Results
We adapted the model on top of the LocalSolver [3] generic combinatorial optimization problems. When used alone, LocalSolver can take a long time finding a feasible solution, so we proceeded in a two-step procedure, which first solves the relaxed hypergraph partitioning problem using hMetis, then uses the obtained solution as a starting point in LocalSolver.

We experimented this procedure on a variety of test hypergraphs, coming from real-world applications. We managed to obtain results on instances up to about 50,000 vertices and the same amount of hyperedges.

The solutions we obtained using this two-step method had much better values (i.e., had less cut nets) than when using the LocalSolver-based method alone. Simpler ad-hoc initial solution methods such as a greedy algorithm or tabu search also improve the solution quality, but to a lower extent.

Perspectives
This work was a first study in the domain of netlist partitioning. It shows the difficulty of designing good netlist partitioning methods, and even industrial-class generic solvers like LocalSolver may not be suitable for obtaining good solutions, and can mostly be used to improve solutions coming from specifically-made heuristics. This opens many opportunities in various domains, like ad-hoc methods inspired from those used in hMetis but adapted to our problems, or using metaheuristics such as simulated annealing. Dealing with large-scale netlists of millions of cells also constitutes a challenge that must be dealt with.

RELATED PUBLICATIONS:
Nowadays, heterogeneous system has merged to offer high performance computing while preserving energy consumption. However, they bring new challenges in term of task scheduling and resource management. We present an efficient approximation algorithm to solve the task scheduling problem on heterogeneous platform for the particular case of sequential application represented by a linear chain of tasks. The objective is to minimize the total execution time (makespan) with respect to energy consumption bound during execution. Our goal is to provide an algorithm with a performance guarantee.

**Main Results**

We first modeled the problem by mixed integer quadratic constrained program (P) as presented in figure 2. This model take into account various constraints such as: unicity, energy consumption bound (D) during execution as well as precedence constraints and communication costs.

We propose an efficient approximation algorithm which provides a solution with small running time, and also guarantee the quality of the solution obtained compared to the optimal solution.

\[
\text{Min} (Z) = \text{start}_1 + \sum_{i=1}^{m} x_{ij} \text{execute}_{ai} \\
\sum_{j=1}^{n} x_{ij} = 1, \forall i \in [1..n] \\
\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij} \rho_{ij} \leq E \\
\text{start}_1 + x_{ij} \text{execute}_{1j} + x_{ij} x_{ij} \text{execute}_{ij} \leq \text{start}_{i+1} \\
\forall j_1 = 1..m, \forall j_2 = 1..m, j_1 \neq j_2, \forall i = 1..n - 1
\]

Figure 2 : Mixed Integer Quadratic Problem (MIQP)

The ratio obtained depends on the frequencies of two successive processing elements $PE_i$ and $PE_{i+1}$ used in preemptive scheduling. The performance ratio of our algorithm is bounded by $\frac{f_{PE}}{f_{PE}}$ where $f_j$ (resp. $f_{i+1}$) is the frequency of $PE_i$ (resp. $PE_{i+1}$).

Numerical evaluations demonstrate that the proposed algorithm achieves a close-to-optimal performance compared to exact solution obtained by Cplex for small instances. For large instances, Cplex is struggling to provide a feasible solution, whereas our approach takes less than a second to produce a solution for an instance of 10000 tasks.

**Perspectives**

As part of the future, we will focus on studying the tightness of this ratio and then the extension to more general classes of graphs to handle real application.
MAPPING AND SCHEDULING PARALLEL APPLICATIONS ON FULLY HETEROGENEOUS PLATFORMS TO OPTIMISE EXECUTION TIME AND ENERGY CONSUMPTION

RESEARCH TOPIC:
Heterogeneous platform, Parallel application, mapping, scheduling, execution time, energy consumption, optimization

ABSTRACT:
Recent applications in research and industry require massive computing systems which need a large amount of energy. Energy-efficient computing is paramount to sustain the increasing demand of computing power while keeping energy consumption reasonable. One promising research path is to build fully heterogeneous systems so as to map specific parts of applications on computing resources (processors, accelerators, etc.) that are adapted to efficiently run them. The exploitation of these platforms raises new challenges in terms of application management optimization for execution time and energy consumption.

Context and Challenges
Heterogeneous computing systems are widely used in data centers and HPC systems to increase their performance per watt ratio. Exploiting these systems brings new challenges in terms of application management optimization both in research and user community. Tackling this challenge consists in determining efficient strategies to exploit these heterogeneous platforms by finding the best mapping and scheduling strategies. For this purpose, having the right application and hardware models is paramount. The mapping and scheduling problem consists in the assignment of application tasks to a set of processing elements and then sequencing the order of tasks execution for each processing element. The sequencing must be performed in a way that precedence relationships between tasks are not violated. Orchestrating inter-node data transfers to optimize the performance per watt ratio is also an important point to consider. The aim of our work is to determine effective algorithms to exploit these heterogeneous platforms.

Main Results
Different experiments were performed on a heterogeneous micro-server platform (Figure 1) using various applications to characterize the energy consumption and performance of the system. A detailed mathematical model was formulated to map and schedule an application represented by a Directed Acyclic Graph DAG with non-independent tasks on a fully heterogeneous platform. It lead to a bi-objective Mixed Integer Quadratic Problem (MIOP) taking into account the whole set of constraints (including processing elements and communication links heterogeneity) as represented in Figure 2.

$$\begin{align*}
\text{Min} & \quad z = C_{\text{max}}, \\
\text{Min} & \quad z = \sum_{t \in T} \text{energy}_t \cdot x_t \\
\sum_{t \in T} x_t &= 1, \forall \ell, \text{Visit}_{\tau_t}, \forall \tau, \forall r \\
\sum_{t \in T} \text{exec}_t &\leq C_{\text{max}}, \forall \ell, \forall r \\
\sum_{r \in R} \text{exec}_t + \sum_{j \in J} \text{comm}_{ij} \cdot x_{t_j} &\leq \text{Comm}_{\text{max}}, \forall \ell, \forall j \\
\sum_{t \in T} \text{exec}_t - \text{visit}_t &\leq B(1 - x_{t_j}), \forall \ell, \forall j, j \\
\sum_{t \in T} \text{exec}_t - \text{visit}_t &\leq B(1 - x_{t_j}), \forall \ell, \forall j, k
\end{align*}$$

Figure 1: Christmann RECS Box© heterogeneous server (2 Intel CPUs, 1 Nvidia Tesla K80, 8 Samsung Exynos ARMv CPUs and 1 Xilinx Zynq 7045 FPGA).

To validate and efficiency solve the introduced mathematical model, different strategies were proposed. A linearization of quadratic constraints was needed to apply an exact strategy based on IBM ILOG CPLEX Optimization Studio solver. A Local Search heuristic using Innovation24 LocalSolver was adapted for larger instances, as presented in [1]. Several experiments on randomly generated instances as well as the real world Cholesky Matrix Decomposition application were performed. Optimal solutions were found for instances up to 80 tasks. For bigger instances, Local Search heuristic gave good results in terms of makespan or energy minimization. Good quality solutions within reasonable computational times for bi-objective optimization were also obtained. The objective values for both makespan and energy obtained using multi-objective mode for LocalSolver are not far from the optimal value obtained by Cplex. This is an interesting feature in practical situations where efficient algorithms are needed for a tradeoff solution. Our experiments confirmed the soundness of our mathematical formulation for mapping and scheduling parallel application on future fully heterogeneous platforms.

Perspectives
Future work will include additional tests on real-life applications and dynamic strategies to manage online mapping and scheduling of applications onto fully heterogeneous servers. Another perspective is to derive from this study new optimization algorithms together with their parallel implementations to enhance the result systems themselves and to achieve suitable solutions for future HPC systems.

AUTHORS:
L. Zaourar, J.M. Philippe, D. Briand, M. Ait Aba

RELATED PUBLICATIONS:
M2DC -- MODULAR MICROSERVERS DATACENTER WITH HETEROGENEOUS HARDWARE

RESEARCH TOPIC:
Heterogeneous architectures, microservers, hardware acceleration, acceleration framework

AUTHORS:
Ariel Oleksiak (PSNC), Thierry Goubier, Alexandre Carbon, Loïc Cudennec, Jean-Marc Philippe et al.

ABSTRACT:
The Modular Microserver DataCentre (M2DC) project investigates, develops and demonstrates a modular, highly-efficient, cost-optimized server architecture composed of heterogeneous microserver computing resources. The resulting server architecture will be able to be tailored to meet requirements from a wide range of application domains. M2DC is built on three main pillars: a flexible server architecture that can be easily customised, maintained and updated; advanced management strategies and system efficiency enhancements (SEE); well-defined interfaces to the surrounding software data centre.

SCIENTIFIC COLLABORATIONS: UBI, PSNC, POLIMI, OFFIS, XLAB

Context and Challenges
Fast growth in the data center and cloud space calls for dramatic decreases in costs and power requirement, while maintaining quality of service guarantees. New technologies with high computing power to power consumption are creating new opportunities, bring a continuum of computing resources enabling to tailor a system to the exact needs of applications and workload.

To address these emerging challenges, M2DC will investigate, develop and demonstrate as a prototype in an operational environment a modular, highly-efficient, cost-optimized server architecture composed of heterogeneous microserver computing resources, being able to be tailored to meet requirements from different application domains such as image processing, IoT, cloud computing and HPC [1].

Main Results
The project has developed a micro-server architecture (Fig.1) with a flexible interconnect enabling direct high speed serial, PCIe and Ethernet to allow the topology to adapt to application needs, and a mix of high power boards (x86, ARM64, FPGA) and high efficiency system-on-chip boards.

A specific acceleration framework was introduced, with system efficiency enhancements (SEEs). SEEs materialize various acceleration possibilities to increase compute efficiency in the micro-server, but accelerating communications, task scheduling, workload management, and bringing security and self-healing capabilities. First results have been seen for various SEEs such as encryption, intrusion detection, pattern matching and scheduling.

Demonstration at SuperComputing 2017
At the heart of the M2DC middleware sits OpenStack Ironic, which provides bare metal (micro) server software deployment and lifecycle management. OpenStack Ironic will be complemented by additional OpenStack components. It will handle the dynamics and heterogeneous nature of the M2DC microserver nodes, in particular for the hardware accelerators.

Additional results are focusing on a heterogeneous, distributed shared memory implementation [2], and optimization of task allocation on heterogeneous architectures [3] of which M2DC is a key enabler.

Perspectives
M2DC, due to its heterogeneous and reconfigurable nature, significantly lower the barriers to research on architectures, acceleration and task deployment, apart from its market potential for customized appliances.

RELATED PUBLICATIONS:
ARCHITECTURE, IC DESIGN AND EMBEDDED SOFTWARE

AN OPENMP BACKEND FOR THE SIGMA-C LANGUAGE

RESEARCH TOPIC:
Cyclo-Static Data-Flow, streaming language, compilation, runtime generation, OpenMP

AUTHORS:
S. LOUISE

ABSTRACT:
The $\Sigma$C (pronounced "Sigma-C") language is a general purpose data-flow language that was initially targeted for Kalray’s MPPA embedded many-core processor. It is designed as an extension of C, allowing the Cyclo-Static Data-Flow (CSDF) model of computation. Until now, it was only available for the first generation of the MPPA chip, and could not be used to freely study stream programming concepts or to use it to program HPC applications. In this paper, we show how we built an OpenMP backend for this language, and we used this compiler to evaluate some of the assets of stream programming and some limitations of the current implementation.

Context and Challenges
In embedded computing, parallel computing is becoming the standard, for the sake of both performance and power efficiency. Nonetheless, usual HPC programming models are not well fitted, because they are memory and power hungry. High Performance Embedded System must rely less on memory coherence and more on explicit data-movements, with an easy way to avoid both parallelism and performance pitfalls.

Dataflow concepts are not new. One of the base models is Kahn Process Networks (KPN), which provides very simple model of computation which can be represented by a directed graph, with processes as nodes and FIFO communication channels in-between. Data movements are modeled by tokens. Fully deterministic MoCs include SDF (Synchronous DataFlow) and CSDF (Cyclo-Static DataFlow) upon which the $\Sigma$C language is built.

Main Results
For the MPPA processor, we developed a front-end compiler, a graph-optimizer and a code generator targeting a specialized runtime. For OpenMP code-generation, we changed only the latter and disabled some of the Intermediate steps (e.g. the place & route step). The scheduling of the OpenMP code is generated from the so-called repetition vector which provides a base for all correct scheduling.

We use this base for OpenMP code generation and provide 2 types of parallelism: thread level parallelism, as usually provided by OpenMP and pipeline-level parallelism, which is usually trickier to obtain with OpenMP. Last, but not least, we can use OpenMP pragma into standard $\Sigma$C code to generate more parallelism.

A nice illustration is provided on the following benchmark program which is a laplacian filter on an image with an intrinsic level of parallelism below 2:

The exploitation of pipeline parallelism grants a nice speed-up of 2.5 by comparison to the sequential execution and the use of additional OpenMP directive in one of the process (image transposition) offers an additional speed-up of 4.2 compared to the sequential execution, well above the naïve speed-up which would be about 1.5 according to Amdhal’s law.

Perspectives
This work of porting the $\Sigma$C compiler to an OpenMP backend opens two perspectives: first it is a nice tool to educate future engineers with DataFlow programming concepts, but as can be seen, the generated OpenMP code despite a certain naiveté gives good performance, above what can be based on the pure Amdhal’s law. Our next step would be to use a common Intermediate Representation (IR) to natively compile both usual DataFlow and OpenMP structures.

RELATED PUBLICATIONS:

Figure 1 A simple example of a SCDF graph: Data-token input-output rates are shown in square brackets on the edges. Specifically for CSDF, there are 2 repetition vectors: one for the whole cycle of the processes and the other for the subtasks. For the simple graph of Fig. 1, the two repetition vectors are (1 2 2 4 4 4) and (3 4 4 8 8 8) respectfully.

<table>
<thead>
<tr>
<th>Number of threads</th>
<th>Speed-Up</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1.5</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>2.5</td>
</tr>
<tr>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>3.5</td>
</tr>
<tr>
<td>35</td>
<td>4</td>
</tr>
<tr>
<td>40</td>
<td>4.5</td>
</tr>
<tr>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>50</td>
<td>5.5</td>
</tr>
</tbody>
</table>

Figure 2 Speed-up of the laplacian filter as a function of number of threads.
ABSTRACT:
The contribution of this paper is articulated around a new software design approach of autonomous control systems for connected vehicle platoons. Our control system is distributed and real-time based on object-oriented component-based method of design that brakes with the industrial traditions subject to cyclic OS-free approaches. We discuss our results on an autonomous automotive platooning system. We focus on speed control and brake-command. The software is mainly implemented using the Ada and the distribution in PolyORB. The control scenarios and communication aspects are animated by wheeled robot prototypes commanded by single-board ARM Cortex computers under real-time Linux kernels.

Context and Challenges
The impact of massive usage of cars in terms of traffic jams and air pollution makes the mastery of automotive systems of a paramount importance and a hot topic in today’s politics. Few years ago, embedded electronics and software advances are transforming progressively computers to become ambient in human mobility, which renders transport systems more advanced. One of the innovations are Advanced Driver Assistance Systems (ADAS) that aim to improve efficiency, safety and comfort using advanced information and communication technologies. However, safety-related issues remain a challenge.

Main Results
As main result, we propose an automatic process to control the longitudinal speed of vehicles in Autonomous Connected Vehicles Platooning systems. Two simple scenarios are presented and studied in order to illustrate our control system: i) the tail merging of a new connected vehicle in a platoon already in circulation, and ii) the propagation of the braking alarm to followers by the leader’s detection of obstacles.

A further result concerns soundness and scalability of the Distributed Object-Oriented Component-Based Design (DOOCBD) to specify embedded software for High Integrity Safety-Critical systems. We present for that a software architecture of generic speed control in Autonomous Connected Vehicles Platooning systems using a DOOCBD approach.

A first prototype has been realized in Ada, using a PolyORB middleware. The mock-up wheeled robots are controlled by Arduino-based boards enslaved by ARM Cortex-A single-board computers on which our software distributed application is deployed and executed under real-time (Preempt RT) Linux kernels.

Perspectives
We plan to discuss the interplay between safety and security issues in the brake alarm message passing.

RELATED PUBLICATIONS:
A MULTI-START HEURISTIC FOR MULTIPLICATIVE DEPTH MINIMIZATION OF BOOLEAN CIRCUITS

RESEARCH TOPIC:
homomorphic encryption, multiplicative depth, multi-start heuristic

AUTHORS:
Sergiu Carpov, Pascal Aubry, Renaud Sirdey

ABSTRACT:
In this work[1] we propose a multi-start heuristic which aims at minimizing the multiplicative depth of boolean circuits. The multiplicative depth objective is encountered in the field of homomorphic encryption where ciphertext size depends on the number of consecutive multiplications. The heuristic is based on rewrite operators for multiplicative depth-2 paths. Even if the proposed rewrite operators are simple and easy to understand the experimental results show that they are rather powerful. The multiplicative depth of the benchmarked circuits was hugely improved. In average the obtained multiplicative depths were lower by more than 3 times than the initial ones. The proposed rewrite operators are not limited to boolean circuits and can also be used for arithmetic circuits.

Context and Challenges
An encryption scheme is said to be homomorphic when some operations on plaintext messages can be done directly in the space of ciphertexts. A noise component is added to the ciphertext during the encryption for security reasons. Each new homomorphic operation applied on the ciphertexts increases the noise component in the resulting ciphertext. After a number of homomorphic operations the noise is so large that the correctness of the decryption cannot be ensured anymore. The noise growth induced by the addition operation is smaller than the noise growth induced by the multiplication. That is why many authors consider the multiplicative depth.

Obtaining low multiplicative depth circuits is a major issue in the practical use of homomorphic encryption.

Main Results
In this work we propose a multi-start heuristic which aims at minimizing the multiplicative depth of boolean circuits. The heuristic is based on a rewrite operator for multiplicative depth-2 paths.

Even if the proposed rewrite operator is simple and easy to understand the experimental results show that it is rather powerful.

The goal of the rewrite operator is to decrease the number of AND gates that follow node X. In the left-hand side of Figure 1. AND gates follow x.

Let \(((x, y) \oplus u_1) \oplus \ldots \oplus u_r) \cdot z = r\) be the formula of the path. It can be rewritten as \(((x, y) \oplus (u_1 \oplus \ldots \oplus u_r)) \cdot z = r\) (XOR associativity) and can be transformed to \(((x, y) \oplus (u_1 \oplus \ldots \oplus u_r)) \cdot z = r\) (XOR distributivity).

By replacing nodes in term \((x, y) \cdot z\) by \((y, z), y\) (AND commutativity), we obtain a new path (right-hand side of illustration) which minimizes the multiplicative depth by one. A single AND gate follows instead of 2 as before rewriting.

When applied to the critical part of a boolean circuit, this rewrite operator potentially minimizes the multiplicative depth of a circuit. The idea behind this operator is to rewrite critical paths of multiplicative depth 2 in such a way that the overall multiplicative depth decreases.

Figure 1: 2-Depth rewrite operators

We propose a multi-start heuristic aiming to minimize the multiplicative depth of boolean circuits. It is based on multiple starts of a priority based heuristic which rewrite critical paths of multiplicative depth 2. The path to rewrite is chosen using a priority function. Several priority functions have been studied as well as a random priority function. The algorithm stops either when a termination condition (e.g. time, number of iterations) is verified or when there are no more reducible critical paths

The best solution was obtained using a non-random priority in 9 cases and using a random priority in 11 cases. Multiplicative depth of the obtained circuits is significantly smaller when compared to the multiplicative depth of input circuits. In average the multiplicative depth decreases by more than 3 times. As expected, the price to pay for a smaller multiplicative depth is an increase in the number of AND gates (approximately 1.2 times more in average).

Perspectives
The optimization method described in this paper can also be applied to arithmetic circuits. An arithmetic circuit is a generalization of boolean circuits where instead of binary field operations higher degree field/ring operations are used. An arithmetic circuit is functionally complete when defined over addition and multiplication operations. Optimization algorithm proposed in this paper together with rewrite operators can also be directly applied to arithmetic circuits and how to do so

RELATED PUBLICATIONS:
ABSTRACT:
A major challenge with multi-cores in real-time systems is contention between concurrent accesses to shared memory. Dynamic arbitration schemes allow for an optimal utilization of the system’s memory, while sacrificing time predictability. Time-Division multiplexing (TDM), on the other hand, sacrifices average-case performance in favor of predictability. In this work, we explore a dynamic arbitration scheme that is essentially based on TDM, and thus preserves many of its guarantees.

Main Results
The concurrent execution of the system’s tasks is governed by the way memory accesses are arbitrated. For simplicity, we define two classes of tasks: (1) critical tasks, and (2) non-critical tasks. Critical tasks have to complete their execution in time, i.e., respect a deadline. We do not explicitly specify the deadline though, since we assume that critical tasks meet their deadlines when executed under a standard TDM arbitration scheme. The arbiter may, however, diverge from such a scheme, e.g., by keeping track of outstanding requests and selecting one of them in order to give the owning task exclusive access to the memory. We investigated improved arbitration schemes based on TDM, that allow for a more dynamic scheduling of memory requests. We defined an arbitration scheme dubbed TDMds (Dynamic TDM with slack counter), where the arbitration decisions are driven by deadlines. For each critical tasks, we derive a deadline for each request, which simply corresponds to the end of the task’s next TDM slot after the request’s issue date. The TDMds arbiter then free to schedule memory requests dynamically, as long as the request deadlines of critical tasks are respected. This dynamic scheduling thus blurs the separation of TDM slots, i.e., any task may perform a memory request in a given TDM slot - as long as no deadlines are missed. We conceptually associate a slack counter with each critical task. The arbiter then schedules requests according to an earliest-deadline first policy. Our deadline-driven arbitration policy renders TDM slots exchangeable between cores and allows to reorder requests to improve memory utilization.

The concurrent execution of the system’s tasks is governed by the way memory accesses are arbitrated. For simplicity, we define two classes of tasks: (1) critical tasks, and (2) non-critical tasks. Critical tasks have to complete their execution in time, i.e., respect a deadline. We do not explicitly specify the deadline though, since we assume that critical tasks meet their deadlines when executed under a standard TDM arbitration scheme. The arbiter may, however, diverge from such a scheme, e.g., by keeping track of outstanding requests and selecting one of them in order to give the owning task exclusive access to the memory. We investigated improved arbitration schemes based on TDM, that allow for a more dynamic scheduling of memory requests. We defined an arbitration scheme dubbed TDMds (Dynamic TDM with slack counter), where the arbitration decisions are driven by deadlines. For each critical tasks, we derive a deadline for each request, which simply corresponds to the end of the task’s next TDM slot after the request’s issue date. The TDMds arbiter then free to schedule memory requests dynamically, as long as the request deadlines of critical tasks are respected. This dynamic scheduling thus blurs the separation of TDM slots, i.e., any task may perform a memory request in a given TDM slot - as long as no deadlines are missed. We conceptually associate a slack counter with each critical task. The arbiter then schedules requests according to an earliest-deadline first policy. Our deadline-driven arbitration policy renders TDM slots exchangeable between cores and allows to reorder requests to improve memory utilization.

Fig.1 Example of a TDMds arbitration scheme.
Fig.1 depicts a TDMds arbitration scheme on a simple taskset. We evaluated the behavior of TDMds using software simulation of a simple use case and obtained interesting results w.r.t. to the dynamic behavior of slack counters.

Perspectives
Our evaluation considers a single use case only and does not capture realistic execution scenarios in a multi-task environment. As future work, we plan to investigate more realistic execution models, and implement our arbitration policies in hardware. We also plan to allow other forms of slack to be accumulated, e.g., from execution time.

RELATED PUBLICATIONS:
• In-Memory-Computing architecture
• Variable Precision and Approximate computing
• Memory transfer optimization
• Cache coherency scalability
• Runtime optimization and power management
• SSD and FLASH memory optimization
• 3D testability
SOFTWARE PLATFORM DEDICATED FOR IN-MEMORY COMPUTING CIRCUIT EVALUATION

RESEARCH TOPIC:
In-memory Computing

AUTHORS:
M. Kooli, H. P. Charles, C. Touzet, B. Giraud, J. P. Noel

ABSTRACT:
This report presents a new software platform, co-developed by research teams composed of memory designers and software engineers, to size and evaluate a novel In-Memory Power Aware Compiling (IMPACT) system for IoT. IMPACT system is an emerging memory concept that promises to save execution time and power consumption by embedding computing abilities. The proposed platform permits to compile applications implemented with both sequential and IMPACT version, on the Low Level Virtual Machine (LLVM) and record execution traces in order to evaluate the performance and energy consumption.

Context and Challenges
In-Memory Computing (IMC) represents a new concept of data computation that has been introduced to overcome the Von Neumann bottleneck in terms of data transfer rate. This concept aims to reduce the traffic of the data between the memory and the processor. Thus, it offers significant reduction of energy consumption and execution time compared to a conventional system where the computation units and the storing elements are separated.

The IMC concept has just started to be the focus of recent research works. The objective of our research works is to focus on different technological layers of an IMC system: silicon design, system architecture, compilation and programming flow. In [2], we introduced the IMPACT system concept that is based on an SRAM architecture to perform in-memory arithmetic and logic operations (Fig. 1).

Main Results
In order to characterize a new compilation platform for IMPACT circuit, we need to investigate methods to describe the software independently from the target hardware architecture. At this stage of our research, the IMPACT hardware system is not yet implemented. Therefore, it cannot be used to evaluate the system performance when executing an application. We propose a new software emulation platform based on LLVM [2]. It permits to evaluate the system performances thanks to a high level trace generator and analyzer. This allows studying different memory parameters under the light of real world and high level applications.

The proposed software platform consists in compiling the application implemented in two modes: conventional and IMPACT mode (Fig. 2). The IMPACT implementation is based on the LLVM vector representation of the data, where the word size is dynamically configured and is not limited by the register size. The platform then generates two execution traces from each implementation that will be analyzed to compute the execution time speed factor and the energy reduction factor compared to a conventional system.

Perspectives
As perspectives, we aim to investigate more precise energy evaluation of the IMPACT system. Furthermore, we propose to experiment more applications in different fields.

RELATED PUBLICATIONS:
RESEARCH TOPIC:
Scientific computing, variable precision

AUTHORS:
Andrea Bocco, Yves Durand, Florent de Dinechin (INSA-Lyon)

ABSTRACT:
The widely used IEEE 754 standard for floating point representation of real numbers does not perfectly match the requirements of precision and accuracy of many scientific applications. We have designed a floating-point arithmetic co-processor unit which supports the new variable precision format UNUM. Variable precision means that the size of both exponent and fractional part of the operands may be dynamically adapted along with the precision requirements of the calculation. Besides, the unit implements interval arithmetic capabilities which contribute to improve the reliability of the computation.

SCIENTIFIC COLLABORATIONS: INSA-Lyon

Context and Challenges
Current computing systems for scientific applications exclusively use the IEEE 754 standard floating point representation of real numbers defined in the 80s. But a growing number of scientific computing applications, such as large linear systems, ab initio simulations and small scale physical modeling, require large and adaptive fractional part sizes – up to hundreds of digits - during the course of their processing. These applications suffer from cancellation and rounding problems, and augmenting precision may significantly increase stability, convergence and compensate the conditioning issues.

Main Results
Our first realization ([1]) implements the basic operations +, -, x, the arithmetic predicates as well as their counterparts in Interval arithmetic. It is based on the UNUM format published in 2015, which complements the usual sign, exponent and mantissa with self-descriptive fields ((1) on fig. 1).

Fig. 1: UNUM floating point format

Our UNUM arithmetic unit is tightly coupled with its host, in our case a RISC V processor ((1) on fig. 2), through a dedicated interface (2). The host instruction set has been extended to handle the variable precision operations.

Data representation in memory is point critical: augmenting precision not only augments the size of data in memory, but it may also generate misaligned accesses which degrade access times. Therefore, the UNUM arithmetic unit requires a custom “Load and Store” unit (3) for handling its specific access to memory and minimizing memory footprint.

Fig. 2: architecture of the host/ co-processor couple

Next point to take care of is the structure of the floating point unit itself. Instead of using a unique internal accumulator for intermediate calculations, we have considered grouping several extended registers into a local “scratchpad” memory (4): this considerably improves precision and speed for kernels using small and localized internal loops, provided that the compiler is able to constrain the assignment in the scratchpad memory.

Finally, the realization of the operations is an important issue by itself, because the arbitrary length of the fractional part transforms the hardware operations into complex pipe-lines.

Perspectives
Our first realization using programmable components demonstrates that the complexity of using variable format of floating point numbers is affordable with the current integration possibilities offered by modern silicon technologies.

Our next step is to develop an operation computing platform by connecting a revised version of this innovative floating-point unit to an open and mature computing platform, and to experiment benchmarks for refining and validating our design choices.

These adaptive floating point arithmetic accelerators will bring substantial benefits in computation efficiency (by speeding up convergence and reducing software complexity) and in memory occupation (by using the most appropriate format for each data set).

RELATED PUBLICATIONS:
RUNTIME ACCURACY ADAPTATION OF HARDWARE OPERATORS USING THRESHOLD VOLTAGE CONTROL

RESEARCH TOPIC:
Approximate computing, FDSOI technology

ABSTRACT:
Approximate computing trades the accuracy of numerical results for energy savings. At circuit level, this requires to scale the accuracy of selected hardware kernels at runtime according to application requirements. To enable this scaling, the kernel layout is partitioned at design time to create isolated islands. At runtime, adapting the individual threshold voltage of each island modifies the number of valid bits of the complete operator, and reduces its power accordingly. This method achieves up to 49% of power reduction compared to the best alternative method (e.g., DVAS), and avoids the need of costly voltage shifting devices.

AUTHORS:
D.Pagliari (Politecnico di Torino), Y.Durand, D.Coriat, A. Molnos, E.Beigne, E.Macii (Politecnico di Torino), M. Poncino (Politecnico di Torino)

SCIENTIFIC COLLABORATIONS: Politecnico di Torino

Context and Challenges
The principle is to selectively favor the significant functional path when available energy reduces, in an otherwise fully functional hardware computational kernel. Globally lowering the supply voltage $V_{DD}$ is not the best solution: it does not contrast enough the timings between bits in the circuitry, because CAD design tools tend to balance the traversal times for all paths into a same logic group (the “wall of slack” phenomenon). Instead, we exploit the selective tuning of $V_{TH}$ by means of back biasing on groups of supply voltage islands ([1]): this selectively speeds up of critical parts of the operator (MSB), while others (LSB) are simply set to zero by means of control signals.

This method has a low hardware overhead on the design, since back-bias domains require only small separation guardbands and no level shifters. However, the selection and grouping of the different gates according to their impact on precision is still a challenging task at design time.

Main Results
Partitioning the design into several islands sharing the same threshold voltage $V_{TH}$ is a multidimensional optimization problem, since the result has to be valid with different bitwidth/accuracy configurations. After trying diverse algorithms, we finally adopted a regular grid for dividing the initial layout into a relatively small number of rectangles.

Then, the layout is incrementally optimized and routed after the insertion of guardbands and well taps for the connection of voltage Bias rails. The final phase is a systematic exploration of the different voltage/accuracy configuration by means of timing analysis, under different operation conditions. This gives the set of threshold voltage assignments corresponding to the different bitwidth requirements under different values for the global supply voltage $V_{DD}$

We have applied our method to three different hardware designs common in DSP applications: a Booth multiplier with Wallace tree, a 30-tap FIR filter, and a FFT butterfly. We considered 16-bit fixed point implementations of the operators.

Fig. 1: initial division in isolated islands of the physical layout

Fig. 2: energy/accuracy diagram comparing DVAS method (“No Groups”) versus threshold voltage control using 2x2 groups

The diagram of Fig. 2 above savings up to 41.9% (1) of energy for a 9 bits accuracy, and a relatively small energy overhead of 0.85% for the maximum configuration (2). Similar figures were obtained for the FIR design, whereas the FFT partitioning gave a 16.5 % power gain.

Perspectives
This work demonstrates an effective method for building approximate hardware operators, with low overhead and compatible with the state-of-the-art CAD tools. Future developments include the study of alternative $V_{TH}$ domains construction methods, and an investigation of the optimal number and configuration of domains.

RELATED PUBLICATIONS:
A PROGRAMMABLE PROCESSOR FOR ENHANCING INBOUND MEMORY TRANSFERS WITHIN AN EMBEDDED MULTICORE SYSTEMS

RESEARCH TOPIC:
Embedded systems, multicore, process synchronization

AUTHORS:
Yves Durand, Christian Bernard, Romain Lemaire, Cesar Fuguet Tortolero, Emilie Garat

ABSTRACT:
We propose an improved architecture for middle-end embedded devices with emphasis on the efficiency of data transfers: we complement a classical multicore architecture including processors, local memories and DMA with a simple programmable “inbound data processor” which reorganizes incoming data when it reaches each local memory. By doing this, only useful data is transported on the network, and unpacking at destination restores their structure in the most suitable way for the application, without the need of duplication. We have prototyped this inbound data processor in a MIPS-based multicore architecture. Applied on an image compression application, we save up to 33% memory footprint and divide the processing latency by a factor of 2.

Context and Challenges
Specialized accelerators improve the global energy efficiency of a multicore system, provided that the network traffic between local embedded memories does not cancel out the energy savings. Thus, the multicore infrastructure must be enhanced in consequence. When running on multiple cores, the processing of data frequently follows complex “gather-scatter” schemes, involving arbitrary organization patterns. The generic processors are not appropriate to realize these complex transfers. Commonly, advanced DMA are used to read data in local memory using any arbitrary “scattering” pattern and forward it on the local Network-on-chip.

Main Results
Our new component ([1]) is meant to realize the symmetric task at the other side, i.e. to read incoming data from the network and store it in local memory using an arbitrary “gathering” pattern, without intermediate copy in memory and without interaction with the main processor. It is also highly optimized for efficient use of network bandwidth low latency. Also, it is easily programmable to enable any arbitrary re-ordering scheme in memory.

Although it may be programmed independently, it is primary meant for use through application specific APIs (e.g. for encryption, video, signal processing). For example, we have used the inbound transfer processor to implement a Discrete Wavelet Transform (DWT) for image compression on our MIPS-based multicore architecture equipped with both DMA and inbound processor. We only used two different kernels programs for inbound transfers: transpose rectangular block copy, and simple rectangular block copy.

The processor was only left with the actual filtering of the successive picture sub-bands during the successive iterations. The inbound processor performed the costly transposition on the fly ([1] on Fig. 2), and we save up to 33% memory footprint and divided the processing latency by a factor of 2.

Perspectives
This component is now part of our new multicore architecture portfolio based on RISC V, together with the smart DMA and the regulated NoC ([2]). However, the introduction of high bandwidth transfers with the accelerators requires a new level of integrity and versatility on the NoC infrastructure. Therefore, our next step will be to revisit the versatility and robustness of the Network-On-Chip in consequence.

Fig. 2: memory transfers for image blocks between the horizontal and vertical phases of the DWT application.

Fig. 1: transfer between local memories in an enhanced multicore architecture.

RELATED PUBLICATIONS:
**ABSTRACT:**
Doing early design space exploration for coherence protocols on manycore is a challenge. Implementing such protocols in realistic simulation models is costly both in modelling effort and execution time. We propose a trace-driven method to accurately compare cache coherence protocols while keeping cache modelling at a high level of abstraction. We have demonstrated that our approach, while being still fairly accurate, is much easier to develop and much faster to execute than state of the art low level simulators, and enables exploration of key design parameters.

**Context and Challenges**
Manycore architects need a way to select and dimension the best cache coherence protocol for their application considering performance (i.e. latency, bandwidth, power consumption) and hardware related costs. As many parameters have to be considered (protocol complexity, sharing set management, network topology, traffic shape, etc.), neither analytical models – too complex – nor cycle accurate simulation – too lengthy – are appropriate.

**Main Results**
We have proposed in [1] and [2] a method to rank cache coherence protocols by using a trace-driven approach. As described in Fig. 1, our method requires that we first extract the requests emitted by L1 caches in an existing accurate multiprocessor simulator. We typically produce the traces using gem5 as multiprocessor simulator, by inserting monitors at the output of L1 caches. These traces are stored in a binary database to be injected in the simulator in a second step.

The L2 cache includes multiple models of sharing set that are simulated simultaneously to compare them. Traffic, latency, cache block states and sharing set occupation are extracted in another database to be analyzed in a last step. Moreover, our tool is two order of magnitude faster than gem5 simulator.

Fig 2. shows an example of metric analysis on 13 benchmarks of the Parsec suite. The number of cache blocks in broadcast mode is studied in function of a threshold parameter used in a linked-list sharing set representation. This analysis allows fixing the value of this threshold and we see that there is a diminishing return past 5.

This methodology is also useful to rank multiple sharing set representations. We have modeled multiple state-of-the-art sharing set strategies (bit-vector, snoop, Ackwise and linked-list) and shown that their ranking is comparable to other publications. So we have validated that our model match the expected behavior of existing sharing set representations.

**Perspectives**
This tool will now be used to propose an innovative sharing set representation for scalability of cache coherence in a manycore context. Indeed, thanks to the speed of simulation we will be able to explore the design space of sharing sets and their parameters to rapidly converge on a performant proposal.

**RELATED PUBLICATIONS:**
IN-SITU FMAX/VMIN TRACKING FOR ENERGY EFFICIENCY AND RELIABILITY OPTIMIZATION

RESEARCH TOPIC:
Fmax Vmin tracking, low power

AUTHORS:
Ivan Miro-Panades, Edith Beigne, Olivier Billoint, Yvain Thonnart

ABSTRACT:
Achieving the lowest possible operating voltage is needed to minimize the power consumption of a circuit but also to increase its reliability w.r.t hardware errors. An in-situ technique to estimate and reduce the design margins of a circuit is presented which significantly minimizes the operating voltage and tracks it during run-time operation of a circuit without failure. A DSP core embedding this technique has been fabricated and measured. Its VMIN has been estimated within +3.5%/-2.5% at nominal clock frequency (1600MHz), thus reducing by 19% its energy per operation.

Context and Challenges
In traditional digital circuit design and to guarantee that the circuit is able to work correctly under process, voltage and temperature (PVT) variations, timing margins are added during the design. These margins constrains the circuit to a suboptimal operating point. Reducing these margins leads to a power and reliability optimization. Therefore, it is possible to reduce the voltage (VMIN tracking) for an unchanged clock frequency, or to increase the frequency (FMAX tracking) when the voltage is unchanged.

Main Results
Unlike state of the art techniques where the circuit margins are estimated only during the design of the circuit, this methodology, called TMFLT, also estimates and minimizes the actual margins in-situ. Thus the margins are computed by the circuit itself taking into account its PVT and its environment. Once the margins are estimated, it is possible to minimize them and track this minimum voltage point during the run-time of the circuit.

The TMFLT methodology is decomposed into two phases: the calibration phase where the circuit margins are estimated and compensated, and the run-time phase where the circuit is able to track the minimum voltage operation during run-time.

The calibration phase (Fig 1) has to be executed at least once in the lifetime of the circuit to obtain the calibration parameters. These parameters are used by the run-time phase to track the minimum operating voltage of the circuit during normal operating mode. The calibration phase can be re-executed in order to take into account the environment change or the ageing of the circuit. Moreover, a firmware update of the circuit may improve the calibration phase and thus improve the achievable margin reduction.

A register selection methodology has been developed and used to select the best registers to add the sensors. Bringing the usefulness of the sensor from 40% (without it) to 98%.

Fig 2. Voltage error percentage between estimated and real VMIN on 21 different dies (meas.)

A DSP was fabricated on UTBB-FDSOI technology using this technique to estimate and track the VMIN of the circuit. The estimation error was ±3 at nominal voltage 1.6GHz where its power consumption was reduced by 19%.

Perspectives
A second demonstrator embedding these methodology has been fabricated and is being measured. Moreover an artificial neural network is being used to correlate the sensor information with the real FMAX/VMIN of the circuit.

RELATED PUBLICATIONS:
REFRESH FREQUENCY REDUCTION OF DATA STORED IN SSDS BASED ON A-TIMER AND TIMESTAMPs

RESEARCH TOPIC:
Flash memory; SSD; reliability; endurance; data retention; Arrhenius law; timestamps

AUTHORS:
M. Seif, E. Farjallah, F. Badets, C. Layer, J.-M. Armani, F. Joffre, C. Anghel (ISEP), L. Dilillo (LIRMM), V. Gherman

ABSTRACT:
In flash-based solid-state drives (SSDs), periodic data refresh allows dealing with a reduced retention time and, indirectly, may be used to improve cycling endurance. A worst case data refresh frequency is not optimal in the presence of temperature variations as it may become unnecessarily pessimistic and alter the SSD response latency and energy consumption. Here, a flexible data refresh methodology is proposed based on approximations of the Arrhenius-curve employed to describe the temperature impact on the retention capability of flash memories. These approximations may be implemented with the help of a small module called A-timer. For an asymmetric temperature distribution between 30°C and 70°C, it is estimated that the refresh frequency can be reduced by more than 63× and almost 3× for respectively charge detrapping and stress-induced leakage current (SILC) failure mechanisms.

Context and Challenges
Solid-state drives (SSDs) based on flash memories offer power and performance advantages together with improved shock and vibration resistance as compared to traditional hard-disc drives (HDDs). Technology scaling and development of multi-level cells are used to fill the cost gap between SSDs and HDDs. Unfortunately, both trends affect cycling endurance, i.e., the cumulative number of program/erase (P/E) cycles that can be sustained by a flash memory cell. The number of endured P/E cycles has also an important impact on data retention time, i.e., the longest period of time during which information can be reliably stored. Consequently, solutions aimed to deal with an insufficient data retention time may also be used to improve cycling endurance. An efficient approach to deal with an insufficient retention time is to periodically check and refresh the stored data. The problem is that a worst-case refresh frequency may become excessively pessimistic with a consequent impact on response latency and energy consumption in the presence of temperature variations. Here a method is proposed to reduce the number of data refresh operations in flash memories via an integration over time of the temperature impact on their data retention capability.

Main Results
A small hardware module, called A-timer, is introduced to integrate over time the temperature impact on the data retention capability and alert against potential data retention hazards. The considered A-timer implementation is shown in Fig. 1. The output of a digital temperature sensor is used to select an increment value for a status register. Warnings are issued each time the value stored in the status register is increased above a certain threshold value. The status register only needs to be updated at a low frequency since the temperature variation rate is relatively low compared to system clock frequency. The increment value of the status register approximates the temperature impact on the guaranteed data retention time that is usually described by the Arrhenius law. Such possible approximations are illustrated by the dashed curves in Fig. 2 for the two main failure mechanisms responsible that may affect data retention in flash memory cells i.e. charge detrapping and stress-induced leakage current. The dashed curves are maintained below the corresponding Arrhenius curves in order to avoid data retention hazards.

Fig. 1: Possible A-timer implementation

Fig. 2 Arrenius curve approximations.

Despite its simple design, the A-timer is able to efficiently approximate the impact of a changing operating temperature. For asymmetric temperature distributions over a range of 40°C measured within data center SSDs, the estimated refresh frequency reductions may reach 63× and 3× for respectively charge detrapping and SILC failure mechanisms.

Perspectives
Currently, we are investigating methods that allow to reduce the uncorrectable bit error rate [1] and further cut down the number of refresh operations in flash-based SSDs.

RELATED PUBLICATIONS:
IMPROVEMENT OF THE TOLERATED RAW BIT ERROR RATE IN NAND FLASH-BASED SSDS WITH THE HELP OF EMBEDDED STATISTICS

RESEARCH TOPIC:
NAND flash; SSD; reliability; adaptability; data retention; bit error rate; embedded statistics

AUTHORS:
V. Gherman, E. Farjallah, J.-M. Armani, M. Seif, L. Dilillo

ABSTRACT:
A recent large-scale study revealed that the uncorrectable bit error rates (UBER) in data center SSDs may fall far below the JEDEC standard recommendations. Here, a technique is proposed to improve the tolerated raw bit error rate (RBER) based on the observation that (a) a small ratio of SSD units may have a much higher RBER than the rest and (b) the RBER is dominated by the retention error rate. Instead of employing stronger but costly error-correcting codes a statistical approach is used to estimate the remaining retention time of flash memory pages. It is estimated that the tolerated RBER can be increased by more than a decade over a storage period of 3 years if the stored data are verified on a monthly basis and refreshed only if necessary. The ability to adapt to the actual RBER enabled performance overhead reductions by more than 8x compared to systematic refresh schemes.

Context and Challenges
Solid-state drives (SSDs) based on NAND flash memories offer a low power and high performance storage alternative to traditional hard-disc drives (HDDs). The continuous technology scaling and emergence of flash memories with multilevel cells brought not only cost per gigabit reductions but also reliability degradations. For instance, the cycling endurance of a flash memory is decreased by an order of magnitude each time the cell storage capacity is enhanced with an additional bit. What is more, a recent large-scale study revealed that the uncorrectable bit error rate (UBER) of data center SSDs may significantly exceed the JEDEC standard recommendations. The reported UBER values are between $10^{-11}$ and $10^{-9}$ while client and enterprise class SSDs are required to provide an UBER below $10^{-15}$ and $10^{-16}$, respectively. Besides stronger error-correcting codes (ECCs) or solutions based on improved read and write algorithms, the RBER can be tempered via periodic refresh operations of the stored data. Existent refresh schemes are based on worst-case scenarios, oblivious to intra- and inter-device variations, which may lead to unnecessary overheads with respect to response latency, dissipated power and P/E cycles. Here, an approach is proposed to avoid the utilization of worst-case refresh frequencies for dealing with populations of flash memories or SSDs that may contain some error-prone units.

Main Results
The idea is to take advantage of the read operations of each flash memory page in order to estimate its remaining retention time. This can be done based on the detected number of retention errors and the retention age, i.e., the elapsed time since data was programmed. Fig. 1 illustrates examples of the remaining retention time as a function of the number of retention errors at several retention ages for a flash memory page with 8K bits vulnerable to retention errors. A valid memory page should be refreshed when the estimated remaining retention time is smaller than the timespan to the next page access operation. Such a technique is effective when a maximum time interval is imposed between consecutive read operations of any memory page. For example, the tolerated RBER can be increased by up to 28x over a storage period of 3 years if one makes sure that the stored data are read at least once in a month. The resulting data refresh frequency is not necessarily correlated to the frequency of data read operations since it depends on the actual RBER via the estimated remaining retention time. It is shown that the refresh probability is negligible at RBERs that can be managed by the available ECC alone and starts to increase only at larger RBERs. Compared to systematic refresh schemes able to ensure the same protection level, performance overhead reductions between 8x and 12x have been simulated. This is illustrated in Fig. 2 for an ECC able to correct up to 10 per page.

Fig. 1 Estimated remaining retention time as a function of the number of retention errors for several retention ages.

Fig. 2 Performance overhead reduction with respect to a systematic refresh scheme with fixed refresh frequency. Each curve stops at the maximum tolerated RBER.

Perspectives
Currently, we are looking at a linear approach to estimate remaining retention time of each flash memory page.

RELATED PUBLICATIONS:
DISTRIBUTED SHARED MEMORY OVER HETEROGENEOUS COMPUTING ARCHITECTURES

RESEARCH TOPIC:
High Performance Embedded Computing, Heterogeneous Computing, Shared Memory Programming Model

AUTHORS:
L. Cudennec

ABSTRACT:
Today's heterogeneous architectures are composed by a mix of high-performance, low-power, reconfigurable computing nodes and accelerators. They are expected to provide a tradeoff between performance and energy consumption in application domains such as data centers, embedded devices and autonomous vehicles. Among several open challenges, the programmability is one major step towards the acceptance of these architectures. In this work, we propose to tackle the problem of shared data management by deploying a distributed shared memory over the heterogeneous nodes, building an abstraction of the distributed memories.

SCIENTIFIC COLLABORATIONS: This work received support from the H2020-ICT-2015 European Project M2DC

Context and Challenges
Distributed shared memory (DSM) provides a convenient abstraction of the physically distributed memories by implementing a global logical space shared among tasks. Data localization and transfer are transparently managed by the runtime and orchestrated following the given consistency protocol. While the implementation is quite straightforward on a NUMA machine, DSM is still challenging in distributed architectures with a focus on computing performance and energy consumption.

Main Results
In this work [1] a DSM is proposed and evaluated over a heterogeneous micro-server architecture. This software DSM is implemented at the user level over the MPI message passing runtime. It allows tasks to transparently allocate and access shared data, using multiple consistency protocols within the same application. DSM are usually deployed onto homogeneous systems that allow a simple load balancing of shared data and metadata management. With heterogeneous resources, a tight configuration of the consistency protocol and the application is required to obtain performances and energy savings.

Christmann RECS|Box heterogeneous micro-server

We have conducted experimentations over a RECS|Box micro-server including Intel Core i7, Arm Cortex A15 and A9 cores as a mix of high-performance and low-power computing nodes. Image and video stream processing applications have been deployed using different configurations. Processing tasks and S-DSM servers are mapped onto heterogeneous nodes in order to highlight the intricacies between the shared memory access patterns, the consistency protocol and the hardware capabilities. We show that data coherence protocols should be designed with the possibility to adapt the metadata management load depending on the resource performances. We propose the dissociation of the data management (metadata) and the cache system to be able to deploy on different sub-systems. Another insight is the collocation of roles (and user tasks) that need to extensively communicate, which is not always possible due to hardware limitations, hence promoting a finely tuned S-DSM.

Perspectives
Not all decisions regarding the application and DSM configuration can only be of the user responsibility and we currently work on the automatic exploration of DSM sizing and placement over heterogeneous resources. This leads to a Pareto front from which the user is able to select the most relevant tradeoff between performance and energy, depending on the execution context.

RELATED PUBLICATIONS:
WIRELESS SOLUTIONS

- RF Power Amplifier
- Wireless sensor system
- Millimeter-waves high data rate
- Compressive Sensing
- Feature extraction, Cognitive Radio
A ROBUST AND VERSATILE, -40°C TO +180°C, 8 SPS TO 1KSPS, MULTI POWER SOURCE WIRELESS SENSOR SYSTEM FOR AERONAUTIC APPLICATIONS

RESEARCH TOPIC:
High temperature, robust, digitally-assisted sensor interface.

AUTHORS:

ABSTRACT:
In order to meet aggressive aeronautic environment constraints, we developed an autonomous multi-power-source wireless sensor node (WSN) interfacing with aeronautic-grade resistive bridge transducers. It operates from -40°C to +180°C, provides 8Sps to 1kSps sample rate at 10-to-12-bit resolution with a 10mV/V input signal, and can be powered either from RFID or a dedicated micro wind turbine. It consumes 794µW, including 1kΩ transducer excitation, and is implemented in 0.18µm HT SOI process. It is hermetically sealed into a metal/ceramic packaged that resists to more than 2500 temperature cycling for 2500 cumulated hours.

Context and Challenges
Developments of next-generation aircraft engines call for prototypes to be fitted with several sensors that can withstand severe operating conditions. Besides, these sensors must not require any cables in order to instrument rotating element (such as fan blades) and eventually reduce weight and cost.

To meet these needs, we developed a wide-temperature-range WSN compatible with 350Ω-3kΩ industry-standard resistive transducers [1]. It requires no battery, gets supplied from a UHF RF signal or a low-frequency AC energy harvester, and communicates wirelessly with a standard RFID reader.

Main Results
A top-level bloc diagram is shown in Fig.1. Power is derived from RF and AC inputs by, respectively, a native-NMOS 3-stage RF rectifier and a negative-voltage-converter (NVC). At 180°C, NVC shows up to 92% efficiency, while 14.5 dBm RF level is required to acquire sensor data at 1kSps with 10mV/V input signal, and can be powered either from RFID or a dedicated micro wind turbine. It consumes 794µW, including 1kΩ transducer excitation, and is implemented in 0.18µm HT SOI process. It is hermetically sealed into a metal/ceramic packaged that resists to more than 2500 temperature cycling for 2500 cumulated hours.

The WSN comes out of sleep mode upon detection of an RF level larger than -10 dBm. It secures startup by measuring the power available from both sources and selects the most powerful one. The 6-wire sensor interface includes a configurable-duty-cycle transducer driver as well as a 3rd-order sigma-delta ADC with programmable input range, sample rate, and OSR. This choice outperforms a SAR ADC approach in terms of noise, peak power consumption, and anti-aliasing properties.

Latest measurements show 62 dB SNR at 1kSps, 180°C, equivalent to 10-bit ENOB and 8.9µV RMS input-referred noise (Fig.2a).

Perspectives
A 2nd generation of this WSN will feature improved RF efficiency and sensor interface performances, as well as a specific clocking scheme that could enable much lower digital supply voltage.

Moreover, investigation work is carried out on digitally-assisted sensor interface architectures capable of reaching very high accuracy and ultra-low drifts.

RELATED PUBLICATIONS:

Figure 1: Wireless Sensor Node bloc diagram

Figure 2: (a) ENOB (b) ADC clock frequency error (c) Power consumption at 180°C

ADC clock is divided from a 10MHz ring oscillator using a digital algorithm to track the RFID standard frame delimiter. This saves a clock trim step and keeps sample rate drift below 2.2% from -40 to +180°C (Fig.2b). Overall, digital circuits represent about 50% of total power, while sensor interface account for only 6% of it (Fig.2c), thus leaving plenty of room for future improvements.
CO-SIMULATING COMPLEX ENERGY HARVESTING WSN APPLICATIONS: AN IN-TUNNEL WIND POWERED MONITORING EXAMPLE

RESEARCH TOPIC:
WSN; cross-layer design; wind energy harvesting; tunnel modelling; HarvWSNet; WSN simulation

AUTHORS:
L.-Q.-V. Tran, A. Didioui, C. Bernier, G. Vaumourin, F. Broekaert and A. Fritsch (Thalès)

ABSTRACT:
A complex wind-energy harvesting wireless sensor network (WSN) application for subway tunnels is pre-prototyped using the HarvWSNet co-simulation framework. Detailed models of every component, and physical layer (PHY), medium access control (MAC), routing and power aware protocols are described. Pre-prototyping simulations quickly identify the key parameters affecting the viability of the application (e.g., harvested energy, train-induced signal attenuation). Thanks to its native capacity to handle detailed charge flow models, HarvWSNet can address scenarios with complex and intermittent energy supplies.

Context and Challenges
While energy harvesting (EH) appears to be an extremely promising technology for extending WSN lifetime, new tools are required for accelerating time-to-market for these new systems.

In this work, a simulation based prototyping approach is proposed that overcomes the traditional limitations of event-driven WSN simulators. The HarvWSNet co-simulation framework [1] associates an open-source discrete-events WSN simulator with an industry standard simulator (Matlab) that easily supports continuous-time models for the energy harvester. The suitability of this co-simulation framework for rapidly pre-prototyping complex energy harvesting WSN scenarios is demonstrated based on an illustrative subway tunnel wind harvesting monitoring application.

Main Results
The applicative scenario is as follows: 183 wireless sensor nodes are deployed along the ceiling of a 24.3 km subway line tunnel, one approximately every 130 m. Each node is locally powered by a micro-turbine generator which harvests the wind produced by each train passage. The pre-prototyping simulation must be able to quickly respond to questions such as:

Can the wind generated by the passage of subway trains near each wind harvester furnish enough energy to power the nodes?

The block diagram of the Matlab model for the wind-flow energy harvester is given in Figure 1. A crucial element of this model concerns the supercapacitor whose large redistribution currents can be accurately modeled by a multi-time constant RC network. This accurate model, along with models for the in-tunnel RF propagation channel, node power consumption and train mobility, and backed by dedicated routing, MAC and synchronization protocols, was used to simulate the network behavior over several hours and days, and given different application constraints.

Figure 2 shows the evolution over 24 hours of the supercapacitor voltage of two nodes, node 178 (a leaf node) and node 181 (a more heavily burdened routing node) in the case where a power management algorithm at each node allows each node to independently adapt its sensor activation period. Seeing that node 181’s final voltage is smaller than its initial one, we can conclude that this particular scenario is not viable for the application. The figure also clearly illustrates that supercapacitors tend to recharge at moments of the day when the train traffic is more intense.

Perspectives
This work proves that the use of a simulation platform allowing for the detailed charge flow modelling of the energy harvester and storage module is essential to the initial prototyping exploration step. This work also demonstrates the importance of energy aware communication protocols. Indeed, the fluctuating nature of energy harvesting makes the use of context and energy aware protocols a major contributing factor in the future success of EH-WSN applications.

Fig.1 Windflow energy harvester Matlab model

Fig. 2 Evolution of supercapacitor voltage during a 24h simulation

RELATED PUBLICATIONS:
TAKING ADVANTAGE OF FD-SOI FOR RF POWER AMPLIFIER DESIGN

RESEARCH TOPIC:
Power Amplifier, 28nm UTBB FD-SOI, 5GHz, High Power technique

AUTHORS:
B. Martineau, E. Mercier, P. Vincent

ABSTRACT:
This work describes the design of a 5GHz power amplifier (PA) taking advantage of the FD-SOI technology. Fabricated in a 28nm UTBB FD-SOI process with 1.8-V thick oxide devices, the PA output exhibits up to 23dBm Psat and 18dBm with low distortion under a 3.7V power supply. The core occupies less than 1mm² while integrating transformer and baluns.

Context and Challenges
Among the research focuses on wireless connectivity, power amplifiers (PA) for the 802.11ac (5GHz – high data rate) standard are some of the most challenging to design because of their high frequency and the strict linearity specifications. Those imply a high power capability as well as high distortion performances. On the face of it, FD-SOI technology could be seen as a poor candidate to address such a challenge since it has been created for digital and low power applications. Nevertheless, thanks to the distinctive CMOS FD-SOI properties [1], a specific design technique can be adopted to match those antagonistic requirements without new process options.

Main Results
The power amplifier high output power means a high output swing voltage at the same time with a high operating voltage. In this work, a solution using stack configuration is proposed. The Figure 1 shows an area optimized, cascode NMOS-PMOS pushpull configuration. This topology allows stacking 4 MOSFET transistors in series while sharing the same n-well sub. Unlike standard bulk CMOS, this technic eliminates the need of a deep n-well for N-MOS and the large area which come out with this configuration.

Therefore, each transistor of this array sees only ¼ Vdd voltage over the drain to source and only ½ Vdd at a maximum DC voltage between each of its terminal and the n-well sub. This architecture is only possible because of the regular and flip well configuration allowed by the FD-SOI.

Perspectives
The fabricated circuit proves that FD-SOI technologies are not only suitable for low power but also for high power RF circuit if proper design taking benefits of the technology is adopted. Additionally, in a system on chip perspective, with a small area and without any specific technology option, this work demonstrates that cost effective integration is possible.

Figure 1: Core circuit cascode NMOS-PMOS push-pull configuration

Figure 2: S-parameters and AM-AM / AM-PM distortion measurements

The figure 2 shows the measured results. S-parameters confirm the 5GHz center frequency with 1GHz bandwidth. In this configuration, the power gain is 17dB, while AM-AM and AM-PM results shows that the topology performs very well to suppress distortion [2]. The circuit consumes ~200 mA Idd at rated power +18dBm and can deliver up to 23dBm Psat. These results are obtained in a compact area of 1mm² integrating matching transformer and baluns.

RELATED PUBLICATIONS:
A HIGHLY LINEAR BIDIRECTIONAL PHASE SHIFTER BASED ON VECTOR MODULATOR FOR 60GHz APPLICATIONS

RESEARCH TOPIC:
5G mobile communication, phase shifters, vector modulator, 60GHz, CMOS

AUTHORS:
Frédéric Hameau, Clément Jany, Baudouin Martineau, Aurélien Larie (NXP), E. Mercier

ABSTRACT:
This work describes the design of a 60GHz phase shifter for beam-steered systems. Based on an I/Q hybrid transformer and specific passive attenuator, it improves attenuation depth, phase accuracy and impedance matching. Fabricated in CMOS 55nm it covers the 360° phase shift with less than 1dB gain and 5° imbalance over the 55GHz to 67GHz bandwidth. This vector modulator paves the way towards mobile millimeter-waves high data rate telecommunication systems.

Context and Challenges
High data rate mobile wireless communication recently led to the widespread adoption of the 802.11ad standard in the 60GHz band. Such communication requires the use of beamforming and beamsteering in order to improve the link budget. It implies an ability to accurately control the gain and phase to support high spectral density modulation scheme. In the state-of-the-art, phase shifters are mainly composed of switching delay elements or polyphase filters for the passive ones, or vector modulator based on voltage gain control attenuators for active ones. While passive topologies allow bidirectionality at the price of poor noise figure, low bandwidth and phase coverage, active topologies exhibit unidirectional wideband operation and low NF. The proposed architecture takes advantages of vector modulator topology but add the bidirectionality by using an innovative design of passive attenuator.

Main Results
Each phase is amplitude-weighted thanks to an analog control passive attenuator which has been designed to improve the attenuation tuning range at high frequency using a neutralization crossed coupled MOS structure.

As shown on measurements on fig.2, the 3dB bandwidth covers the complete V-band with a good matching in both modes (Transmitter and receiver). An example of vector modulator complete phase and gain coverage is given at 60.48GHz using an external voltage control stepped at 10mv (equivalent to a 5bits DAC control). Limiting the gain variation to +/−0.5dB (grey circle on fig.2 polar chart) the vector modulator achieves a 4.5° phase control step with only a 1° RMS phase error which is far beyond the state-of-the-art.

Figure 1: Vector modulator architecture with the crossed coupled passive attenuator structure.

Fig.1 shows the proposed vector modulator based on a hybrid transformer followed by two baluns which provide the needed four phases (0°, 90°, 180° and 270°) to achieve the 360° phase control.

Figure 2: Measurement performances
Large signal measurements have been performed and show a compression point of 5dBm in TX mode and -16dBm in RX mode. The phase non linearity is only of 3° at compression.

Perspectives
The fabricated vector modulator occupies only 1.69mm² including Pads. It consumes 16mW for the TX and RX drivers. Based on a low cost technology (CMOS 55nm without any specific option) it paves the way for large scale antenna phase array integrations needed for the future beamforming and beamsteering 5G millimeter wave transceiver.

RELATED PUBLICATIONS:
CODE PROPERTIES ANALYSIS FOR THE IMPLEMENTATION OF A MODULATED WIDEBAND CONVERTER

RESEARCH TOPIC:
Compressive Sensing, Cognitive Radio

AUTHORS:
M. Marnat, M. Pelissier, O. Michel (GIPSA), L. Ros (GIPSA)

ABSTRACT:
This paper tackles the sampling of analog signals under the Nyquist rate with the Modulated Wideband Converter and delivers an evaluation methodology of the code sequences element, vital for implementation. First, new circulant codes with excellent correlation properties are proposed. Second, high-level properties of the codes such as coherence and distance preservation are studied. Third, the expected performances are confirmed with a simulation platform.

SCIENTIFIC COLLABORATIONS: GIPSA-LAB

Context and Challenges
Compressive Sensing proposes to go past the Nyquist rate by assuming that there is a representation space in which few coefficients suffice to describe the input signal. Exploiting this theory, the Modulated Wideband Converter (MWC) is one of the compressive sensing architectures foreseen to be able to break the usual compromise between bandwidth, noise figure and energy consumption of Analog-to-Digital Converters. A high potential is also identified for classification or estimation of spectral parameters directly from the compressed samples. In the context of Cognitive Radio, the integrality of the signal is not necessarily required, and therefore the power-hungry reconstruction step can be skipped. The pseudo-random code elements are the key for the implementation of a MWC architecture. However selecting a code, from a set of desired properties, has not yet been studied methodologically.

Main Results
The Modulated Wideband Converter, on Fig. 1, mixes the input in M branches with periodic codes. Each subband of the Nyquist band is thus folded with weights given by the Fourier Transform of the code, creating spectral diversity. The spectrum is then low-pass filtered and sampled at low rate (k₀ρ/νN). To ensure the information recovery, the codes need to optimize the measured information.

![MWC architecture](Image)

New structured codes for the MWC are proposed: circulant, based on Zadoff-Chu codes and real-valued, they satisfy excellent correlation properties. Usual binary codes (Gold and Bernoulli) are benchmarked with our proposal and we showed that our codes are very promising. First, the circularity eases the implementation. Second, the proposed codes have good coherence. This implies an almost minimal required number of measurement, hence parallel branches, and an ability to handle even crowded spectra. Our study highlights the importance of an appropriate row selection in the sensing matrix, i.e. of M among N possible codes. With the right row selector (through random selection rather than the first lines) our codes achieve lower hence better coherence than State-of-The-Art. More importantly the proposed codes have proved excellent norm and distance preservation properties. This implies a good resilience to noise, in contrast to Gold codes. Noise resilience of circulant codes based on Zadoff-Chu sequences has been verified by simulation means, see Fig. 2. Fig. 2 is an accuracy graph showing the percentage of spectral support fully recovered for both a noisy (10 dB ISNR) and a noiseless environment.

![Accuracy of the MWC with growing compression rate](Image)

The recovery rate with Gold codes (brown) collapses when noise is added (dashed compared to solid line) whereas our codes (green) show better resilience. Simulations validate that proposed codes indeed require few parallel branches and are more resilient to noise than State-of-the-Art, as expected from the analysis of mathematical properties.

Perspectives
In order to avoid the highly energy consuming signal reconstruction step, a promising solution for feature extraction is to perform, instead, estimation and classification directly on the compressed samples.

RELATED PUBLICATIONS:
NUWBS: NON-UNIFORM WAVELET BANDPASS SAMPLING FOR COMpressive RF FEATURE ACQUISITION

RESEARCH TOPIC:
Sub-Nyquist sampling, Analog to information converter, cognitive radio, compressive sensing, feature extraction

ABSTRACT:
Feature extraction from wideband radio-frequency (RF) signals, such as spectral activity, interferer energy and type, or direction-of-arrival, finds use in a growing number of applications. Compressive sensing (CS)-based analog-to-information (A2I) converters enable the design of inexpensive and energy-efficient wideband RF sensing solutions for such applications. However, most A2I architectures suffer from a variety of real-world impairments. We propose a novel A2I architecture, referred to as non-uniform wavelet bandpass sampling (NUWBS).

Context and Challenges
CS-based A2I converters that leverage spectrum sparsity are a promising solution for wideband RF feature acquisition applications. CS enables the acquisition of larger bandwidths with relaxed sampling-rate requirements, thus enabling inexpensive, faster, and potentially more energy-efficient solutions than traditional Nyquist analog-to-digital converters (ADCs). While a large number of CS based A2I converters have been proposed in the literature the generally-poor noise performance and sensitivity to real-world hardware impairments prevents their straightforward use in low-power and cost-sensitive applications.

Main Results
We propose a novel A2I converter for cognitive RF receivers, i.e., radio receivers that are assisted with an A2I converter specifically designed for RF feature extraction. The A2I converter bypasses conventional RF circuitry and extracts a small set of features directly from the incoming RF signals in the analog domain. The acquired features can then be used by the RF front-end for parameter tuning (e.g., of filters) or by the digital signal processing (DSP) (cf Fig1).

Perspectives
The proposed NUWBS methods are promising strategies for A2I converter architectures that overcome the traditional limitations of existing solutions in power and cost limited applications. The ongoing work concerns the development of practical hardware platform that demonstrates the benefits of the solution exploiting custom ASIC. Also investigation on direct specific RF features extraction for signal classification is a topic of interest.

A2I converter solutions, such as signal noise, aliasing, and stringent clocking constraints, which enables a broad range of RF feature extraction tasks (Fig2).

NUWBS has the following key advantages. First, the analog wavelet transform reduces the bandwidth of the input signal x(t), which relaxes the bandwidth of the sample-and-hold (S&H) circuit and the ADC. Second, NUWBS enables full control over a number of parameters, which enables one to tune the wavelets to the signal class to be acquired. Using this novel A2I architecture, we have shown by simulation means that NUWBS approaches the theoretical phase transition of 11-norm-based sparse signal recovery with Gaussian measurement ensembles. We have furthermore analyzed the rejection performances of NUWBS against out-of-band interferers. To demonstrate the practical feasibility of our A2I feature extractor, we have proposed a suitable wavelet generation circuit that enables the generation of tunable wavelet pulses in the GHz regime.

Fig 1: Overview of a cognitive radio receiver assisted by A2I converter for RF features extraction

Our approach, referred to as non-uniform wavelet bandpass sampling (NUWBS), combines wavelet preprocessing with non-uniform sampling in order to alleviate the main issues of existing A2I converter architectures, such as signal noise, aliasing, and stringent clocking constraints, which enables a broad range of RF feature extraction tasks (Fig2).

Fig 2 : Generic serial NUWBS architecture to acquire Gabor frame or wavelet samples.

RELATED PUBLICATIONS:
Sensors & Energy

- Smart Imagers for consumer (Event based, low noise, compression, ...)
- Dedicated Imagers (InfraRed, Xray, Thz antenna)
- GaN power devices
- Piezoelectric energy harvesters
- Time-Domain sensor interface
AN EVENT-BASED, FRAME-BASED IMAGE ACQUISITION MECHANISM FOR CMOS IMAGE SENSORS

ABSTRACT:
A novel event-based image acquisition mechanism dedicated to industrial vision system under hardware constraints is presented here. This technique relies on smoothing the data throughput of the image sensor through time, decreasing information collision and without adding complex communication systems such as address event representation (AER). This technique can also reduce the data throughput by suppressing a part of the spatial redundancies. The proposed architecture is firstly validated through a MATLAB model, then a corresponding design is presented.

Context and Challenges
Time domain image sensors are a relevant and robust solution to overcome technological challenges. Contrary to conventional image sensor where light is encoded as an amount of charge, voltage or current, time domain quantization is based on timing of pulse. To reduce the intrinsic constraints of standard asynchronous AER readout scheme, this work presents a design of an event-based acquisition mechanism with appropriate communication scheme inside the pixel array to smooth data throughput over time as well as reducing the output data flow.

Main Results
The proposed acquisition scheme illustrated in Fig. 1 is based on a basic assumption assuming that the light intensity does not change over a time frame, which is greatly superior to the reading time of information. The pixel array is divided in independent columns, each column having a set of output bus to transmit its pixels information. After the global reset, when a pixel event is generated at time $t_1$, the pixel requests access to the first bus. Two cases can occur: either the reading system is already busy receiving information from another pixel of the column, either it is free. If the reading system is busy, then the pixel cancels its request and resets its photodiode. At time $2t_1$, another event is generated by the pixel, requesting access to the second bus. If the reading system is free then pixel information, its row address, can be send. With such scheme, the time quantization error is reduced and limited to the requests occurring after the $n$th reset of the pixel (with $n$ bus).

Fig. 2 shows the results of MATLAB simulation for an acquisition, the different image shows results for an increasing number of buses available, with only one reading system. The data reduction stays constant over the image (over 60% data reduction). The more buses are implemented, the more image PSNR improves.

Perspectives
We propose a novel image acquisition mechanism based on AER vision sensor through a smart management of columns pixel outputs. This kind of architecture could be seen as a trade-off between a standard (fully synchronous) imager architecture and a fully asynchronous AER imager in terms of complexity, power consumption and standard image processing compatibility. This mechanism allows data reduction of 60% in the taken example, but can be up to 90% depending on the scene. This acquisition and communication mechanism is relevant for application such as always-on vision sensor and recent digital technologies.

RELATED PUBLICATIONS:
IMPACT OF FIXED PATTERN NOISE ON EMBEDDED IMAGE COMPRESSION TECHNIQUES

RESEARCH TOPIC:
Wavelet image compression, Compressive Sensing, advanced CMOS image sensor, Fixed Pattern Noise

AUTHORS:
W. Guicquero and L. Alacoque

ABSTRACT:
The Fixed Pattern Noise (FPN) is mainly due to technology dispersions and layout non uniformity. The Dark Signal Non-Uniformity (DSNU) represents the offset variations while Photo Response Non-Uniformity (PRNU) is the level of pixel gain variations. Once those two parameters are properly calibrated for each pixel, the image can be corrected. If this 2-points correction is not performed, some problems arise in case of embedded image compression. This work presents a novel approach combining wavelet based compression and Compressive Sensing (CS) to improve image compression performances in the case of a high FPN.

Context and Challenges
Recent advances in digital image sensor (e.g. in CMOS technology) tend to implement image compression techniques directly in the focal plane in order to take advantage of the bit rate reduction at the lowest system level. Image compression methods generally assume the sparsity or the structurality of images. For instance, mathematical transforms such as Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT) have been intensively used for that purpose during the last decades [1]. In the specific case of a certain class of sensors (e.g. cooled IR imagers) the Fixed Pattern Noise (FPN) can be very high but considered as a simple affine transformation of each pixel with various offsets and gains. In that context, the interest of transform based methods is lowered because of being not able to compensate pixel variabilities. On the other hand, Compressive Sensing (CS) [2] theoretically seems to be insensitive to it as FPN can be compensated at the reconstruction stage if properly calibrated.

Main Results
Thanks to a simple model of the FPN, we have introduced a CS reconstruction algorithm trick taking advantage of the image sensor calibration. Indeed, due to the synthesis nature of the algorithm, it is thus possible to incorporate PRNU and DSNU correction during the reconstruction.

Fig. 1: Compression schemes that are compared. Haar-wavelet based, Compressive Sensing and mixed.

In the case of a low FPN, it is commonly admitted that in terms of general performance (i.e. PSNR versus compression ratio) standard image compression strategies outperform CS at the expense of heavier on-chip embedded processing, namely related to the entropy encoder. It motivates what we called a mixed compression scheme that is aiming at combining CS based measurements with Haar transform coefficients to take advantage of both acquisition modalities. Fig. 1 illustrates the different approaches that have been investigated.

Fig. 2: Compression results for an unstructured FPN on the cameraman test image (512×512, 10bit).

Fig. 2 shows the results in terms of PSNR of decompressed images estimated from original artifact-free images vs. compression rate (number of Bit Per Pixel). The first conclusion is that recent CS implementations therefore represent an attractive and robust alternative to transform based compression for imagers exhibiting poor pixels responses uniformity. The second result is that our mixed compression strategy allows a PSNR increase of almost 4dB (on average) compared to straightforward CS while still preserving the great robustness of CS against FPN.

Perspectives
All the reconstruction algorithms suppose that a complete Non Uniform Compensation (NUC) table has been acquired for the sensor, prior to the compression. However, recent works have shown that, thanks to a proper post processing, the NUC table can be build using multiple acquisitions and without the need of a complex calibration of the system.

RELATED PUBLICATIONS:
SMART PIXEL ARCHITECTURE FOR LOW POWER CMOS IMAGE SENSOR: TIME-TO-FIRST SPIKE WITH INHIBITION MECHANISM

Context and Challenges
The internet of Things consists of an ensemble of billions of communicated devices. One of the main challenges in IOT for security, surveillance or detection is the power autonomy of the devices to reduce maintenance constraints. From those statements, two constraints of IOT can be easily highlighted: all devices have to exhibit low power consumption specifications and the amount of emitted data has to be strongly reduced. We thus propose an image acquisition scheme relying on an asynchronous read-out, to address the low power constraint since the system is solicited only when information is available for reading. The proposed pixel architecture has been simulated in the FDSOI 28nm technology.

Main Results
We propose a new low power image acquisition AER readout scheme: the Time-To-First Spike with Inhibition Mechanism (IM-TTFS). To accomplish spatial redundancy reduction during read-out, leading to data flow reduction, we group neighbor pixels in small blocks and pixels with similar luminance information under a same block will provide a single send information. The pixel matrix is split into N x N blocks. Within each block, a signal, called the inhibition signal, is generated when the first event – and only the first – is generated and it will be active for a predefined time. During this inhibition time, every pixel belonging to the block which flashes is inhibited, i.e. flashing pixels during the inhibition time do not send requests and are immediately put into idle mode to save power. After the inhibition time, every flashing pixel can send a request, the inhibited pixels stay in idle mode and do not send requests until the next global array reset, which also resets the inhibition system. The mechanism is illustrated Fig. 1 (in this example, P2 is the brightest pixel).

The table below illustrates the percentage of data which is sent with the inhibition mechanism for different images, illustrating the compression ratio and the induced PSNR.

<table>
<thead>
<tr>
<th>Image</th>
<th>Data %</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vine Sunset</td>
<td>51,43</td>
<td>41,74</td>
</tr>
<tr>
<td>Snow</td>
<td>25,51</td>
<td>44,54</td>
</tr>
<tr>
<td>Hall Monitor</td>
<td>77,73</td>
<td>43,93</td>
</tr>
<tr>
<td>Lena</td>
<td>96,65</td>
<td>55,43</td>
</tr>
<tr>
<td>Bridge</td>
<td>71,33</td>
<td>44,42</td>
</tr>
</tbody>
</table>

A pixel architecture has been designed in the FDSOI 28nm technology. The whole power consumption for on-going work is expected to be less than 100µW for a 128x128 image sensor with a pixel pitch of 3µm.

Perspectives
In an IOT application context, where power consumption challenges arise, we propose an enhanced acquisition mechanism in an advanced technology, FDSOI 28nm. IM-TTFS, inhibition mechanism time-to-first spike, allows data reduction up to 90% on homogeneous scene and at least 5% data reduction. It partially suppress spatial redundancy. For a 128 by 128 pixels image sensor, the power consumption is expected to be less than 100µW which could make this sensor an advantageous candidate for always-on vision sensor.

RELATED PUBLICATIONS:
LIQUID CRYSTAL LENS CHARACTERIZATION FOR INTEGRATED DEPTH SENSING AND ALL-IN-FOCUS IMAGING APPLICATION

RESEARCH TOPIC:
3D Imaging, Liquid Crystal lens, Depth from Focus, Lens characterization, Modular lens characterization

AUTHORS:
S. Emberger, L. Alacoque, A. Dupret, C. Lecat-Mathieu de Boissac, J-L. de Bougrenet de la Tocnaye, N. Fraval

ABSTRACT:
Depth from Focus (DfF) is a promising method for embedded 3D acquisition. It relies on the extraction of points of focus during a focus sweep. To be efficient, the scene should be invariant during the acquisition. One limitation of conventional focus lenses is the slight focal-induced zoom caused by the lens translation. Liquid Crystal Lenses (LCL) and liquid lenses (LL) do not rely on lens movement for focus which makes them good candidates for processing-inexpensive DfF techniques. We developed experimental LC lens and a dedicated methodology, in order to evaluate their compliance with a DfF application.

SCIENTIFIC COLLABORATIONS: Telecom Bretagne

Context and Challenges
One of the first source of noise in DfF 3D sensing method is the changes that takes place in the scene during the focus sweep of the acquisition process. These changes can be caused by objects in motion or optical deformations induced by the focus change. We designed an optical bench in order to measure two key parameters of these discrepancies: focus change speed and focal-induced zoom. Two affordable liquid lenses (LL) from Optotune are compared to experimental Liquid Crystal lenses (LCL) designed at Telecom Bretagne. Fig. 1 presents the optical benches, the acquired scene and an example computed depth map.

Main Results
Focus stabilization speed was measured by computing the Sum of Absolute Difference (SAD) between consecutive frames of high-speed videos acquired during various focus sweeps. The SAD value increases throughout the focus sweep as it reflects the scene fast change. This allows the precise timing of focus change speed. The following graph presents the focus speed of both LCL and LL for various focus excursions. As it can be seen, LCL exhibit much slower focus speed than LL.

In order to evaluate the focal-induced zoom for each lens, we acquired a series of images at different focus point and evaluated the cumulative zoom using Matlab imregister tool. Results are presented on Fig. 2. As it can be seen, LL exhibit much higher focal-induced zoom discrepancies than LCL lenses.

Perspectives
Experimental LCL are not subject to focal-induced zoom but are slower than COTS liquid lenses. Some new LCL will be designed to mitigate this issue through the use of lower-viscosity nematic solutions.

RELATED PUBLICATIONS:
S. Emberger et al. "Liquid Crystal lens characterization for integrated depth sensing and all in focus imaging application" Electronic Imaging 2017
S. Emberger et al. "Low complexity depth map extraction and all-in-focus rendering for close-to-the-pixel embedded platforms", ICDSC 2017
S. Emberger et al. "A near pixel Depth from Focus architecture for video rate depth estimation", Electronic Imaging 2018
LOW COMPLEXITY DEPTH MAP EXTRACTION AND ALL-IN-FOCUS RENDERING FOR CLOSE-TO-THE-PIXEL EMBEDDED PLATFORMS

RESEARCH TOPIC:
RGBZ camera, depth camera, Depth from Focus, 3D acquisition, All-in-Focus imaging

AUTHORS:

ABSTRACT:
Providing depth information for images is a recurrent topic and many methods have been proposed in the literature, but most of them at the expense of heavy processing. Among these methods, Depth-from-focus allows to estimate depth with low processing requirements, therefore representing the most convenient depth extraction method for an implementation close to the pixel. We developed a low-processing-requirements depth extraction method based on bloc contrast analysis and compared it to the literature. A confidence image is also computed in order to help the hosting system to weight measurement relevance.

Context and Challenges
In embedded computer vision, the trend is to get richer information from the scene than mere 2D images. The depth of the scene constitutes the main lack of information and many works aim at providing solutions for its extraction. However, few studies attempt to embed this processing as close to the pixel matrix as possible. This work relies on the Depth-from-Focus (DF) method because of its low processing requirements. It is designed to provide RGB-Z-C pixels, where RGB is an all in focus colored pixel, Z the object depth and C a confidence value used for data filtering or fusion.

Main Results
We developed an original Depth from Focus low complexity architecture. It is based on two distinct DfF pipelines whose results are merged thanks to a built-in confidence estimation, thus enhancing the final result quality (Fig. 1). Some examples of outputs are presented on Fig. 2.

Figure 1. Dual channels Depth from Focus estimation with embedded data fusion based on measurement confidence

Figure 2. Output of the system. From left to right: ground truth, computed depth map, confidence map and All-in-Focus images

Quantitative and visual comparisons were performed with results from the literature. The table below presents the ratio of pixels whose depth was correctly estimated with respect to the ground truth for several low complexity methods, including ours.

<table>
<thead>
<tr>
<th>Image</th>
<th>Sun</th>
<th>Panda</th>
<th>Chair</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our Method</td>
<td>0.8056</td>
<td>0.6193</td>
<td>0.9330</td>
</tr>
<tr>
<td>Max gradient modulus</td>
<td>0.6586</td>
<td>0.4644</td>
<td>0.8814</td>
</tr>
<tr>
<td>Max Laplacian</td>
<td>0.6513</td>
<td>0.4334</td>
<td>0.8916</td>
</tr>
<tr>
<td>SMD</td>
<td>0.7014</td>
<td>0.5777</td>
<td>0.9117</td>
</tr>
<tr>
<td>Yokata (first step)</td>
<td>0.3511</td>
<td>0.2520</td>
<td>0.5845</td>
</tr>
</tbody>
</table>

A great care was taken to reduce the computing complexity. By careful design of the architecture and simplification of the sharpness criterion algorithm a complexity as low as 1.4p additions per pixel (with p, the number of depth planes) and 0.06 frame of memory was reached for continuous depth maps, confidence maps and all-in-focus images generations.

Perspectives
This depth sensing architecture was implemented in software on a Jetson Tegra K1 embedded platform with a 1 full record per second throughput. A FPGA or ASIC port of this architecture would allow near video throughput of RGB-Z images.

RELATED PUBLICATIONS:
S. Emberger et al. “Liquid Crystal lens characterization for integrated depth sensing and all in focus imaging application” Electronic Imaging 2017
S. Emberger et al. “Low complexity depth map extraction and all-in-focus rendering for close-to-the-pixel embedded platforms”, ICDSC 2017
S. Emberger et al. “A near pixel Depth from Focus architecture for video rate depth estimation”, Electronic Imaging 2018
STUDY OF INTEGRATED THZ ANTENNA GEOMETRIES USING METASURFACE FOR REAL-TIME IMAGING

RESEARCH TOPIC:
Terahertz, THz, CMOS, MOSFET, antenna, bow-tie, loop, metasurface, EBG, imaging.

AUTHORS:
N. Monnier, J-A Nicolas, G. Sicard, F. Gallée (IMT), M. Ney (IMT).

ABSTRACT:
Terahertz frequency region has grown much interest especially in the imaging field. CEA-Leti works on the development of THz focal plane array compliant with low-cost sub-micron CMOS processes and therefore addresses some issues about the performances of antennas on a silicon substrate and with a reduced metal stack. In this work various antenna geometries with a rectifying MOSFET detector were studied under different electromagnetic (EM) environments and coupling characteristics.

Context and Challenges
Thanks to some of its characteristics (non-ionizing, absorbed by water, ...) THz band triggered some interest in the imaging field with different applications from medical diagnostics to industrial testing or also to security areas. To be able to generalize the use of THz imagers CEA-Leti made the choice to develop real-time THz detectors with low-cost production and at ambient temperature. Therefore, the whole THz imager has to be monolithically integrated with a standard CMOS process. Nevertheless using a standard CMOS technology for THz detection raises some issues like the losses in silicon substrate, the influence of routing metal layers, and coupling effects from pixel to pixel due to surface waves. This work tackles these challenges.

Main Results
The THz detection is realized by the on-chip antenna and an nMOSFET detector acting as THz waves rectifier; both are followed by an in-pixel high performance signal processing chain. Co-design of the antenna and this nMOS is therefore mandatory to obtain the maximum transfer of power between them, i.e. to satisfy the impedance matching.

Figure 1. (a) bow-tie antenna, (b) loop antenna.

Two different geometries of antenna were studied (Fig. 1):
- A bow-tie antenna with an inductive balun for impedance matching.
- A loop antenna, naturally inductive.

To optimize the pixel area, the antenna is implemented above the read-out chain in the 8th metal layer (M8) and therefore a ground plane (GP) in M6 is added to separate them and to avoid losses in the Si substrate or electromagnetic influence by lower metal layers. In classical design, the space between the antenna and the GP as to be around λ/4, while CMOS process imposes a shorter space between them. To reduce this space, a specific electromagnetic band gap (EBG) structure is designed in M7 for its capacity of size reduction of the classical quarter wavelength cavity and surface wave suppression.

Figure 2. Radiation patterns of bow-tie [(blue solid (a)) & (b)] and loop [(dot red (a)) & (c)] antennas with EBG surface and GP

Considering this stacking of the three metal structures (antenna, EBG surface and ground plane) included in silicon dioxide (SiO2) (εr = 4), Ansys HFSS simulations were realized [1,2]. As result, a S11 of 18.2 dB at 294 GHz is obtained with the loop antenna and -18.2 dB at 286 GHz with the bow-tie antenna. Furthermore a 5.2 dB gain is achieved for the loop antenna at 294 GHz and a 5.8dB gain for the bow-tie antenna at 286 GHz (Fig. 2).

Perspectives
These antenna / MOS elements have been implemented in a test-chip and will be tested soon. This work paves the way for the next CEA-Leti CMOS THz imager.

RELATED PUBLICATIONS:
A 128X128, 34µM PITCH, 8.9MW, 190MK NETD, TECLESS UNCOOLED IR BOLOMETER IMAGE SENSOR WITH COLUMN-WISE PROCESSING

RESEARCH TOPIC:
Thermal imaging, low power imaging, On-chip image processing, microbolometers

AUTHORS:
Laurent Alacoque, Sébastien Martin, Wilfried Rabaud, Edith Beigné, Antoine Dupret, Bertrand Dupont

ABSTRACT:
A 128x128 pixels thermal imager with on-chip processing was designed. It targets a broad range of applications such as 2-points corrected IR camera or features extraction for privacy-compliant presence detection, localization and counting. It features a new current-based ΔΣ Analogue to Digital Conversion architecture for high dynamic-range, 128 processing elements and a built-in analogue and digital pixel-level offset pre-correction to improve operability and manufacturing yields. This architecture was designed to push bolometers IR technology one step forward towards high-end applications for consumer market.

Context and Challenges
There is a growing demand for thermal imagers that operate at room temperature for a variety of domains such as home automation or automotive. However, their relatively high power consumption, system complexity, limited thermal range, and privacy concerns are obstacles to their acceptance in these industries. The goal of this project is to facilitate thermal sensor adoption by providing integrated solutions to these concerns.

Main Results
The overview of the circuit architecture is presented on Fig. 1

Figure 1. Circuit general architecture and example counting and localization application

This circuit aims to provide a sensing and computing platform that can address a variety of applications. It contains column-wise general purpose processing elements that can perform signal processing, feature extraction and data filtering. Fig. 2 presents the detailed architecture of the imager column. Pixels are acquired using a novel ΔΣ architecture that greatly improves linearity and dynamic range. At the end of the acquisition phase, the pixel value is in the processor registers, thus allowing immediate bolometer offset and gain compensation errors correction. The corrected image can be stored into the built-in 12 frames memory and further processed on-chip in order to produce video or privacy-compliant rich image features. We demonstrated a people counting and localization application (Fig. 1 right) with a 13 times lower power consumption than the lowest-power infrared imager product available.

Perspectives
This highly configurable architecture can be programmed to adapt to a variety of application, and to the adequate infrared sensors.

RELATED PUBLICATIONS:
ABSTRACT:
To achieve better and faster material discrimination in applications like security inspection, X-Ray image sensors giving a highly resolved energy spectrum per pixel are required. In this paper, a new pixel architecture for spectral imaging is presented, exhibiting a 256 bin spectrum per pixel in a single image duration, up to two orders of magnitude higher than previous works. A prototype circuit, composed of 4x8 pixels of 756µm x 800µm and hybridized to a CdTe crystal, was fabricated in a 0.13µm process. Our pixel architecture has been measured at 8 Mcounts/s/pixel while embedding on-chip charge sharing, charge induction and pile-up corrections.

Context and Challenges
X-ray solid-state imaging is used in a growing number of applications. In the particular case of security inspection application, discriminating among materials is a key asset, achieved by photon energy measurement. This technique is referred to as spectral imaging. X-ray images acquisition is usually based on hybridization of a CMOS Read Out Integrated Circuit (ROIC) with a CdTe layer. In standard photon counting, i.e. "gray-level" X-Ray image sensors, the pixels are based on the summation of the charges created by detected photons, regardless of their individual energy. In spectral imaging, the image acquisition is based on the sorting of detected photons according to their energy.

Main Results
In this work [1], we present, for the first time, a circuit providing 256 energy bin spectrum, up to 2 orders of magnitude higher than previous works. In order to obtain a cleaner spectrum and contrary to previously published works, our pixel also integrates on-chip charge sharing, charge induction and pile-up corrections relaxing off-chip post processing constraints.

Chips have been hybridized with CdTe and Fig. 3 gives our measurement results. Our maximum Output Count Rate (OCR) is measured at 8 Mcounts/s/pixel as illustrated in Fig.3 (a). It also clearly shows that our Charge Sharing Correction decreases the maximum achievable OCR by only 19%. A measured Cobalt source spectrum is illustrated on Fig.3 (b). Given our circuit bandwidth tuned to reach high OCR, a desirable low Full Width at Half Maximum (FWHM) of 12keV is reached.

Perspectives
To the best of our knowledge, this X-Ray multispectral image sensor is the first exhibiting a 256 bin spectrum per pixel in a single frame duration. It is also 2 side buttable, so a large line-scan detector system can be implemented. It will pave the way for a breakthrough in security inspection applications.

RELATED PUBLICATIONS:
GAN POWER FOR HOUSEHOLD APPLICATIONS: INNOVATIVE BRIDGELESS CONVERTERS USING BIDIRECTIONNAL GAN POWER TRANSISTORS

RESEARCH TOPIC:
AC switch; Gallium Nitride (GaN); AC/DC conversion; LED driver

AUTHORS:
D.Bergogne, O Ladhari, L Sterna, P Perichon

ABSTRACT:
The work addressed in this summary presents a new bridgeless AC/DC topology without electrolytic capacitor. The topology is described and studied by simulation. Experimental results on a LED driver demonstrator using bidirectional GaN transistors validate the proposed circuit. A volume reduction of the power electronics of a ratio of 2 can been achieved.

Context and Challenges
GaN technology, compared to silicon, benefits from remarkable properties: high breakdown field, high thermal conductivity, higher electron density, and wide band gap. These properties allow: higher voltage, higher frequency and higher operating temperature than silicon, which naturally predestines it to power conversion applications. Mixing GaN and AC switches has produced Bi-Directional Switches. To illustrate these properties, we propose a non-isolated AC-DC LED power supply that shows an improvement in lifetime and energy efficiency, therefore addressing current environmental concerns (ecology, energy).

Main Results
By using GaN for the Cuk structure, it is possible to increase the switching frequency over 1MHz, hence enabling a drastic reduction in size (>2 ratio) and weight of the transient energy storage passives: inductor L1 and capacitor C. By using Bi-Directional GaN AC switches, it is possible to remove the input diode bridge and the DC tank capacitor, thus increasing the energy efficiency. Note that Q21 and Q22 are “swapping transistors”, which operate at mains frequency, 50 or 60Hz, depending on the country energy distribution standards. D1 and D2 are freewheeling diodes used for the Cuk behavior of the circuit, switching at 1MHz and more, just as the output filter, L2 and C. Here again, the high frequency operation reduces the size and cost of the passive components such as the output filter.

Figure 1. AC input CUK converter for LEDs. Q1 is Bi-Directional. No DC bus.

A prototype was built and tested running directly from the 230VAC mains. On figure 2, one can see the control signals of the diode swapping circuit (Q21, Q22) in blue and green traces operating at 50Hz, mains frequency, the yellow-brown trace is the PWM control signal running at 1MHz, the result on the oscilloscope plot is a low pass effect.

Figure 2. 5W LED bulb with the power board implementing the circuit described in this work. Control signals and mains voltage measured on the prototype.

A novel circuit is presented and a prototype is built to power-up a LED bulb. The prototype power board fits inside a LED bulb. It is not optimized in size, as some components still use large packages and the control board and interface need to be miniaturized. However, by using high frequency and a new architecture it has permitted the reduction of the energy storage components.

Perspectives
This work was based on a discrete component approach that has shown its limits because of the high switching frequency. Future work include the integration of the GaN bridgeless converters directly on one chip using Leti’s GaN Power technology to realize the power components alongside the gate drivers and auxiliary supplies. Such a GaN Integrated Circuits (GaN_IC) will need integrated passives components at circuit board level to reduce connection parasitics and to achieve even higher efficiency in a reduced volume.

RELATED PUBLICATIONS:
ADVANCED POWER CONVERTERS TOPOLOGIES FOR SEAMLESS INTEGRATION

RESEARCH TOPIC:
Power on-chip, fully integrated DC-DC converter, harvesting interface, piezo-based power supply, resonant converter.

ABSTRACT:
We have explored multiple design aspects in power converters to enhance their power density, voltage isolation capability and efficiency at lower power delivery. We propose and experimentally prove multiple approaches by introducing alternative materials (piezoelectric device), energy transfer (resonant), feedback (back biasing) and passive optimization (serial compensation). Based on a strong collaboration with well-known research labs, we investigate some alternative solutions for seamless integration of power converters in various applications context (harvesting, wireless sensor, gate drivers, low-power computing).

Context and Challenges
Considerable efforts have been devoted to the design of high power density (W/mm³) or low-power delivery (<μW), or high high electrical isolation (kV) power supplies in the last decade. On the high power density side, the power density of fully integrated (on-die) DC-DC converters have been improved thanks to the revival of capacitive-based topology. But we are now faced to the capacitor scaling capability in CMOS, then the efficiency vs power density product achieves a rock limit. On the low-power delivery side, the growing demand of supplying ultra-low power consumption circuits e.g. less than μW wake-up radio creates a new market segment where design solutions are still to be found. In isolated converters, the magnetic core property limits their switching frequency then the passive size.

In this frame, we study hardware solutions i.e. power stage and control topologies, to implement efficient and seamless power converters in various applications such as energy harvesting, DC power supply, gate drivers.

Main Results
We have investigated the improvement of switched capacitor converters (SCC) by introducing resonant energy transfer. Including an inductor in the (dis)charge capacitor paths reduces the inherent sharing loss in SCC, thus enhances the capacitance energy potential at a fixed switching frequency. We have studied the integration capability of resonant SCC (ReSCC) using standard CMOS technology in [1,2]. The effect of interleaved structure have been studied to reduce the constraints on the on-chip air-core inductors (saturation current, magnetic flux frequency...), to reduce the inductors’ counts and bypass capacitor value. Guideline to design air-core inductor into CMOS metal stack have also been discussed in order to maximize the imaginary to resistive ratio by reducing the Foucault effect mainly due to Si substrate. Some perspectives have been drawn to improve the power density of ReSCC especially by maximizing the quality factor of the resonant tank formed by air-core inductor and flying capacitor. Lossless control ability of ReSCC has also been proven in [3].

We have also investigated electromechanical-based converters to remove the flying electrical components (inductor and capacitor). In particular, the piezoelectric cantilever offers high quality factor, high electromechanical coupling and efficiency at resonant frequency and could be an alternative to pure electrical resonant tank. We demonstrated an ultra-low input start-up converter (~10nV) for harvesting applications using Rosen piezo-transformer and Amstrong resonator topology in [4].

We have studied isolated DC-DC converters in low-power and high voltage isolated context. In [5] we analyzed the loss contribution in a sub-nW flyback converter which extracts power from microbial fuel cell. We gave some design guidelines to minimize the loss in the magnetic transformer. In [6], we described a design method to maximize the transmitted power through a cm-scale PCB air-core transformer while maintaining a targeted voltage isolation between the primary and secondary sides. The serial and parallel compensations have been compared to enhance the transformer power capability.

We have also investigated the ability of linear voltage regulator (LDO) to deliver μW level while maintaining high efficiency in [7]. We proposed duty-cycled structure to save quiescent power of the voltage and current references. Thanks to a double feedback based on transistor’s back biasing, our circuit offers stability over a wide range of output capacitor then removing the additional bypass capacitor.

The last research axis developed in 2017 has been the reversible resonant converters to supply energy-recovery logic processor (adiabatic operation). In [8], we analyzed the loss mechanism in multiple and synchronized power-clock rails to fully recover the energy of logic operation. We proposed a solution to synchronize the power clocks to be immune against inductor mismatch.

Perspectives
The new technology (3D packaging, materials, MEMS) and new application constraints (high voltage bus, nW circuits) are the new design challenges in power microelectronics community for the next decade. These proposals pave the way to design dense, high voltage ratio, isolated and efficient power converters by exploring multiple design aspects from materials to converter topology.

RELATED PUBLICATIONS:
NON-LINEAR ELECTRICAL INTERFACES FOR HIGHLY COUPLED PIEZOELECTRIC ENERGY HARVESTERS

RESEARCH TOPIC: Energy Harvesting, non-linear energy harvesting extraction strategy, frequency tuning, piezoelectric harvester, piezo

AUTHORS: G. Pillonnet, A. Morel, P. Gasnier, A. Badel (USMB)

ABSTRACT: Thanks to progresses in both mechanical design and piezoelectric materials, piezoelectric energy harvesters (PEHs) may now exhibit relatively high electromechanical coupling. We show that state-of-the-art non-linear interfaces are no longer optimal for this kind of harvester, as they overdamp the mechanical part of the PEH. We then introduce new approaches in order to reduce the damping induced by the electrical interface on this mechanical part, and even propose to use this strong backward coupling effect to tune the harvester’s resonant frequency and enlarge its energy harvesting bandwidth.

SCIENTIFIC COLLABORATIONS: Univ. Savoie Mont Blanc.

Context and Challenges
During the last decades, energy harvesting has been widely investigated as a promising alternative to batteries in order to autonomously supply sensor nodes. Piezoelectric energy harvester (PEH) is of particular interest in closed confined environments where there are few solar energies nor thermal gradients, such as motors in automotive applications. PEH are usually made of a piezoelectric ceramic deposited on a mechanical oscillator, e.g. a cantilever beam. When the oscillator is subjected to an external vibration, a strain is induced on the piezoelectric material which generates electric charges, thanks to the direct piezoelectric effect. The harvested energy density when the oscillator resonant frequency matches the vibration frequency is relatively high (100µW/cm3). However, in realistic environment, vibrations frequencies are time-variant. When the mechanical harvester is excited outside its own resonant frequency, it drastically decreases the scavenged power. Furthermore, the resonant frequency of the harvester is also time-, temperature- and amplitude-dependent and aging moves the PEH properties. In this research axis, we investigate an alternative way to enlarge the harvesting bandwidth of harvester by exploiting the electrical interface rather than changing the PEH mechanical part. Thanks to the improvement of the indirect piezoelectric effect coefficient from high-coupled piezo materials, the electrical interface has an impact on the dynamics of the mechanical resonator [1,2]. Hence, we have proposed to use the indirect effect i.e. put electrical energy to mechanical part to adapt both the quality factor of the resonator, and its resonant frequency, leading to an enhanced harvesting bandwidth while maintaining maximum power extraction.

Main Results
Non-linear strategies have been traditionally used in order to maximize the extracted energy from PEH. In particular, the well-known Synchronous Electric Charge Extraction (SECE) enhances the power extraction compared to standard energy harvesting (SEH). It consists in accumulating electrical charges in the piezoelectric material every semi-period, and harvesting them when the piezo material have stored the maximum of charges. In 2017, we proposed to extend the SECE strategy by introducing a new degree of freedom in the electrical interface thus increase the SECE ability to harvest energy outside the mechanical resonant frequency. The energy extraction events (and hence harvest the energy once every 1-to-N semi-period of vibration) or speed it up (and hence harvest the energy N times during a single vibrations semi-period). N stands for the ratio of the harvesting events rate to the standard SECE. In [1], we proved that decreasing N reduces the equivalent PEH damping leading in higher harvested power capability. In [2], we proved that increasing N also leads to a damping reduction, while adapting the equivalent PEH resonant frequency.

This analytical modelling is confirmed by experimental results shown on Figure 1. The black curve represents the standard SECE (N=1) harvested power. Both increasing and decreasing N increase harvested power over a large frequency range. High N changes the resonant frequency. Compared to the standard SECE approach, adapting N allowed us to increase the harvested power by a factor of 3x.

![Fig 1. Experimental power responses of our PEH](image)

Perspectives
Based on adaptable SECE, we have to define the best way to control dynamically the parameter N to maximize the power extraction without any ambient vibration information. We plan to integrate this max power extraction tracking (MPPT) algorithm in dedicated IC to achieve low-power consumption and high end-to-end extraction efficiency under small vibration amplitude (less than g).

RELATED PUBLICATIONS:
TIME-DOMAIN SENSOR INTERFACE DEDICATED TO HIGH-TEMPERATURE APPLICATIONS USING INJECTION LOCKED OSCILLATORS AS PHASE SHIFTERS

RESEARCH TOPIC:
Sensor readout circuit - High Temperature - Time-Domain - Injection Locked Oscillators

AUTHORS:
F. Badets, E. Chabchoub

ABSTRACT:
A sensor readout circuit dedicated to high temperature (225°C) is presented. Injection Locked Oscillators are used in order to convert the transducer output into a time shift that is measured with the help of a high resolution counter. As the architecture is mostly digital in nature, it is well-suited for high-temperature applications. A 12-bit sensor readout has been designed and fabricated onto a high-temperature 180 nm SOI technology. The measured temperature drift is lower than 2x ppm over a temperature range from -20 to 220°C have validated the principle.

SCIENTIFIC COLLABORATIONS: LIRMM, National Engineers School of Sfax

Context and Challenges
While the use of smart sensors has become pervasive in consumer applications, their use in high-temperature applications (> 220°C) is still limited. Indeed, at such temperature, silicon technologies suffer from a global increase of its leakage currents as well as a dramatic decrease of its carrier mobility and threshold voltages. Therefore, the biasing points of analog circuits dramatically vary over such a high temperature range, which compromises their robustness.

Main Results
In order to overcome the difficulty of realizing robust analog readout circuit, we choose to work on time-domain architectures. Indeed, performing the signal processing in time domain rather than in analog domain leads to mostly digital architectures which are more robust to high-temperature by nature.

transducer output signal feeds the ILO based phase shifter input. The measurement of the phase shift is made with the help of an 11-bit counter running at a much higher frequency. The same master oscillator is used as the locking frequency for the ILO phase shifter and as the reference frequency for the counter. This ensures the overall temperature robustness of the architecture as it ensures a ratio metric variation of the ILO phase shifter output and of the counter time resolution over temperature.

Fig 2. Measured characteristic function of the ILO based readout circuit

A 12-bit architecture dedicated to resistive sensor has been implemented on a High Temperature 180 nm SOI technology from XFAB [1]. The chip has been tested on board. As it can be seen in Fig. 2, the sensor characteristics presents a very low thermal drift which has been measured to 65 ppm/°C as expected.

Perspectives
This kind of architecture could find application also in Ultra Low Power sensors application. While this architecture has been developed for high temperature application purpose, it is also promising for Ultra-Low Power sensors. Indeed, applications such as always on sensors necessities ultra-low voltage interface which constrains a lot the design of classical analog architecture.

Fig 1. Injection Locked based time domain architecture

Usually, time domain conversion involves a ramp generator and an analog comparator. In the proposed architecture two Injection Locked Oscillators (ILO) are used to convert the transducer output signal into a phase shift. Indeed, once locked onto the frequency of a master oscillator, it is possible to vary the phase shift of an ILO output respect to the master oscillator frequency by tuning its free running frequency [1] [2]. Figure 1 depicts the overall architecture of the proposed readout circuit. The

RELATED PUBLICATIONS:
05 SECURED AND RELIABLE SYSTEMS

- Cryptography
- Silicon reliability
- Wire Diagnosis in aircraft engines
- Smart connectors
- Wiring topology reconstruction, fault localization
- Safety for Autonomous Cyber-Physical Systems
AN ANALYSIS OF FV PARAMETERS IMPACT TOWARDS ITS HARDWARE ACCELERATION

RESEARCH TOPIC:
Homomorphic Cryptography, FV parameters, Residue Number System, Number Theoretical Transform.

AUTHORS:
J. Cathébras, A. Carbon, R. Sirdey, N. Ventroux

ABSTRACT:
Homomorphic cryptography allows computation on encrypted data without decrypting them. Fan and Vercauteren present a promising homomorphic scheme in 2012 and this work deals with the choice of its parameters for efficient implementation in the light of mainstream acceleration approaches. The problematic is to highlight some range of parameters more suitable for hardware acceleration by studying the interdependency between application parameters (security and homomorphic evaluation capability) and implementation parameters (polynomial degree and coefficient size, error distribution). Considering RNS representation and NTT-based polynomial multiplication this work motivates choice of parameter sets with lowest degree possible.

Context and Challenges
Privacy is one of the main concerns regarding the development of cloud services in the context of applications handling sensible data. Homomorphic Encryption (HE) is a solution to exploit data on outsourced servers, making us foresee promising applications [2]. Hardware acceleration could bring currently modest applications to larger scale. This demands the definition of adequate hardware implementation strategies consistent with HE scheme parameter ranges. This work [1] analyses the FV scheme [3] and try to conciliate parameters derivation and mainstream acceleration approaches.

Main Results
This work put in perspective advantages and drawbacks of mainstream acceleration approaches. NTT-based polynomial multiplications tackle high degree influence on performance; this restricts polynomial ring choices to power of two cyclotomic fields to be efficient in hardware. RNS representation brings parallelism and avoid multi-precision arithmetic complexity. Nevertheless, some operations like divisions and rounding operations become more complex. Together these approaches require appropriate choice of RNS basis elements and proper handling of numerous precomputed twiddle factors for NTT.

Concrete choice of FV parameter relates to hardness of the Ring Learning With Errors (RLWE) problem. This influence both degree and coefficient size of the handled polynomials. We analyze parameters sets derivate from original derivation rules proposed by Fan and Vercauteren, and confront them to practical limitation of RNS and NTT coupled approach. Our hardware cost Vs performances analysis brings us to promote parameters sets that trade low polynomial degree against larger coefficients. Indeed NTT asymptotic complexity appears less exploitable at large scale than RNS parallelism. Moreover, handled elements seem smaller and less sensible to application parameters influence for such parameter sets (Fig.2). To validate the feasibility of such suggestion we show that it is always possible to find large enough appropriate RNS basis for considered range of applications.

Perspectives
This analysis invites us to consider acceleration approach scalable with wide range of parameters (polynomial degree and coefficient size). Well-implemented RNS approach seems the key of an efficient and scalable FV hardware acceleration.

RELATED PUBLICATIONS:
SAFETY AND DEGRADED MODE IN CIVILIAN APPLICATIONS OF UNMANNED AERIAL SYSTEMS

RESEARCH TOPIC:
Safety for Autonomus Cyber-Physical Systems

AUTHORS:
Emine Laarouchi, Daniela Cancila, Hakima Chaouchi (Institut Mines Telecom)

ABSTRACT:
The large increase in the number of drones entering the airspace is leading to real concerns about safety and security issues; and small incidents are more and more frequent. The causes are twofold. On one hand the unmanned inherent nature of drones involves a less maintained and consequently less reliable properties than manned aircrafts. On the other hand, amateurs with no particular experiences/backgrounds pilot civilian drones. In this paper, we analyze the existing state of the art and real use cases to identify the key factors involved in safety issues for UASs in civilian applications.

Context and Challenges
The use of small UASs (Unmanned Aircraft Systems) or drones is tremendously increasing and is expected to increase even more in the upcoming years. In 2016, within the United States alone, the total civil drone fleet was estimated at nearly two million aircraft. Despite this trend, however, safety related issues remain a challenge problem. Let us introduce two real examples.
(1) Seattle 2015. A drone non-intentionally crashed during a parade and knocked a woman unconscious. Hopefully, the lady did not have serious injuries but the drone owner was sentenced for 30 days in jail and a 500 $ fine. (2) UK 2012: despite the police interest in using Unmanned Aircraft Systems to monitor the Olympic Games for surveillance purpose, the UK Civil Aviation Authority unlikely did not allow the use of drones for lack of safety and for fear that some accidents may occur.

All these concerns leads to one conclusion: the importance of ensuring safety for the Unmanned Aircraft Systems in civilian application.

Main Results
The main result is the identification of the major flight safety factors for UASs (Figure 1) which tailor work [3] to safety. These key-factors are strongly related to ensure the drone survival and the failure of a single link is sufficient for an accident to occur.

The human factor regroups all the human operators included for the UAS flight. Some missions are complex and require the coordination between multiple human operators to ensure the accomplishment of the given mission in safe conditions. The drone pilot is the main responsible for the flight but some other contributors such as the payload operator or the control station technician may be essential for given applications. The autonomy level of the drone varies depending on the drone type and the mission complexity. The significant difference between UAS and manned ones is the skill of the human operator. In this regard, the pilots of manned aircrafts need to have specific licenses and satisfy certain requirements in order to be authorized to fly with the aircrafts. On the other hand, the civilian drones are operated by anyone without any specific background. This difference is important especially in a matter of safety because the unexperienced pilots can lead to serious accidents.

The environment covers all the external elements that can eventually impact the flight. This includes the meteorological conditions, the communication between the drone and the remote control and also the physical obstacles. There are two different types of environments: static and dynamic. An environment is considered static when all the external elements are determined before the flight and their behavior is predictable and controlled, no further elements can be added during the flight. This type of environment is mainly used for tests and simulations to highlight predefined aspects of the drone. The dynamic environments are more realistic, because it includes unexpected events during the flight. Simulations that includes dynamic environments are more complex and more accurate because they are close to the real world environment.

Figure 1 : the major flight safety factors for UAS

The drone is the unmanned aerial system. Drones can be equipped with information gathering tools such as cameras or sensors, which allows to extract data from its environment and have more or less autonomy, that impacts safety.

In this work, we highlight the safety-related properties with respect to the application domain and discuss some of these.

Perspectives
We plan to introduce a framework based on contracts to guarantee robust ACPS design.

RELATED PUBLICATIONS:
TOWARDS ON-LINE ESTIMATION OF BTI/HCI-INDUCED FREQUENCY DEGRADATION

RESEARCH TOPIC:
BTI, HCI, Circuit-level model, aging.

AUTHORS:
M. Altieri, S. Lesecq, E. Beigne, O. Héron

ABSTRACT:
This work proposes a new bottom-up approach for on-line estimation of circuit performance loss due to BTI/HCI effects. Built on the top of device-level models, it takes into account all factors that influence global circuit aging, namely, process, topology, workload, voltage and temperature variations. The proposed model is fed by voltage and temperature monitors that on-line track dynamic variations. This allows an accurate assessment of the evolution of the circuit critical path delays during its operation. Its accuracy is evaluated on two circuits implemented in 28nm FD-SOI technology.

Context and Challenges
Continuous miniaturization of transistors has exacerbated Front-End-Of-Line (FEOL) aging effects such as Bias Temperature Instability (BTI) and Hot Carrier Injection (HCI). Both phenomena increase the transistors threshold voltage ($V_T$), resulting in a larger propagation delay in digital circuits. Thus, safe margins must be added to the circuit design in order to avoid timing faults. This means that either a lower frequency than the maximum allowed one or a higher voltage than the minimum necessary one has to be applied. All these margins lead to a considerable loss of energy efficiency. In addition, aging is a factor to be considered when applying strategies for power reduction, such as Dynamic Voltage and Frequency Scaling (DVFS). Besides the energy efficiency, one should also consider the long-term consequences when choosing which V-F level must be applied to the circuit, i.e. the impact on the degradation rate. Moreover, actual processors contain tens or even hundreds of cores. Task migration could therefore make use of the information about the state of each core to favour the fresher ones over the more degraded ones. Therefore, there is a need for a mechanism to on-line assess the circuit reliability by estimating the degradation of its critical paths under the actual stress conditions.

Main Results
This work [1,2] proposes a new methodology for tackling this problem by creating accurate but nonetheless simplified circuit-level aging models from existing device-level models and using in-situ monitors to follow dynamic variations.

The proposed solution consists of two parts, namely, an Off-line Modelling step and an On-line Estimation step.

The first step (Fig. 1) consists in generating a propagation delay model $\text{Delay}(V, T)$. Aging simulations are then conducted to find the parameter shift $\Delta \rho$. A formula $\Delta \rho(V, T, t)$ is constructed at the end. All parameters from both $\text{Delay}$ and $\Delta \rho$ formula are stored to be used during circuit operation. Then, the circuit degradation is estimated on-line (Fig. 2) by feeding the previously created model with records of the voltage $V$, the temperature $T$ and workload variations. Note that the process calibration is done only once at the beginning of the circuit lifetime.

The proposed methodology has been successfully evaluated on two different circuit architectures, namely, a DSP and a RISC processor, both circuits implemented in 28nm FDSOI technology. The reader is invited to refer to [1,2] for a complete analysis of the results.

Fig. 1: methodology for obtaining a circuit-level aging model from device-level models (first step)

Fig. 2: On-line estimation of circuit degradation from previously created models (second step)

Perspectives
Future work directions will consider the implementation costs of the proposed on-line estimation in a real system. This includes how the different formula will be computed and how much resources are needed for it. In addition, the effects of body bias changes in FDSOI UTBB technology is under study and will be included in both models.

RELATED PUBLICATIONS:
COMPRESSED SIGNAL ACQUISITION IN WIRE DIAGNOSTIC

RESEARCH TOPIC:
Reflectometry, compressive sensing, Shannon-Nyquist frequency, sparsity, Fractional Fourier Transform (FrFT).

AUTHORS:
Tzila Ajamian, Antoine Dupret and Saïd Moussaoui (ECN)

ABSTRACT:
Reflectometry is a diagnostic technique that allows detection of defects in cables. Analog-to-Digital Converters (ADC) are one of the most required components in the architectures of reflectometry systems. However, the respect of the sampling theorem and the need of performing real-time analysis restrict the highest frequency of the injected signals, thus preclude the detection of some kind of defects. The contribution of this paper is to introduce an alternative ADC based on Compressive Sensing (CS) of the reflected signal in order to overcome the limits of Nyquist. Therefore, a complex linear chirp signal is injected, a dictionary of sparsity based on Fractional Fourier Transform (FrFT) is used and a Random Demodulator is employed as an analog encoder. The main result is that the new ADC scheme allows to reconstruct the reflectogram with a sampling frequency of $18/\text{Nyquist limit}$.

Context and Challenges
Cables are subject to aggressive environmental conditions, which might cause several defects such as short or open circuits. To inspect their status, a reflectometry system injects a signal in the damaged wire and then analyzes the reflected waveforms to locate and characterize the damage. To achieve an accurate localization, high frequency signals should be injected in the wires and the reflected signal according to the Nyquist-Shannon theorem should be sampled at a frequency of several hundred MHz. The objective of this paper is to introduce a compressed acquisition of the reflected signal in order to avoid the limits of Nyquist. The proposed compressive sampling scheme is based on an analog encoder also known as Analog to Information Converter (AIC) that requires the definition of an appropriate dictionary of sparsity. The main idea is to reconstruct perfectly $x \in \mathcal{S}_M$ from an observation $y = \Phi x$, where $y \in \mathbb{R}^M$, with $M$ much lower than $N$ and $\Phi \in \mathbb{R}^{M \times N}$ is a measurement matrix, modeling the sub-sampling and verifying the incoherency property. Actually, the signal $x$ should be sparse, otherwise it will be expressed in a basis $\Psi$, such that $x = \Psi s$ with $s$ containing few non-zero coefficients. In order to adapt the CS to reflectometry, here the purpose is to define both matrices $\Psi$ and $\Phi$.

Main Results
In our study, the injected signal is a linear complex chirp allowing spanning a wide range of frequency band. Moreover, a FrFT with an order $\alpha = 1.69$ of this signal will approve a sparse representation. Since, a hard defect causes a total reflection no matter what the frequency of the signal is, the rate of variation of the $x_\alpha(t)$ frequency is identical to that of $x(t)$. Therefore, the same fractional order $\alpha$ calculated for the injected signal $x(t)$ can be used for the reflected one $x_\alpha$. In this case, $x_\alpha = \Psi s_\alpha$ and $\Psi = F_\alpha$. To implement the idea of CS in the analog domain, the AIC is proposed. The Random Demodulator (RD) architecture is adapted of the reflectometry signals which defines the measurement matrix $\Phi$. This methodology is typically decomposed into three steps: modulation, filtering and uniform sampling. In the Figure 1, we represent the proposed new architecture for reflectometry combined with the RD (as seen in red lines).

Figure 1: The new architecture of reflectometry

Figure 2: Reflectogram corresponding to $f_{\text{Nyq}}/8$ on the reflectogram, Figure 2, we can reconstruct and measure the distance of the defect located at 100m from the point of injection, which is fundamental for reflectometry.

Perspectives
The perspective of this work is to apply the compressed acquisition for different kinds of signal as well as to validate our simulations by an experimental part.

RELATED PUBLICATIONS:
OMTDR-BASED EMBEDDED CABLE DIAGNOSIS FOR MULTIPLE FIRE ZONES DETECTION AND LOCATION IN AIRCRAFT ENGINES

RESEARCH TOPIC:
Aircraft Engines; Fire zone; Detection; Location; Cable diagnosis; Reflectometry; OMTDR.

AUTHORS:

ABSTRACT:
Nowadays, the safety of passengers on board is the first preoccupation for aircraft manufacturers. For this reason, next-generation helicopters are now being designed with a wide range of advanced technologies to reinforce safety. In this context, a new fire detection system based on reflectometry is proposed for monitoring the engines of helicopters. The main idea is to mesh engines by a fireproof cable that serves not only as a physical support of test signals, but also as a fire detection sensor. To do so, OMTDR technology is integrated into an electronic device analyzing in permanence the impedance variation of the sensor caused by a temperature increase.

Context and Challenges
Fire constitutes a severe hazard, which may eventually lead to the crash of the aircraft. Aircraft manufacturer hence design engine compartments to mitigate the risks of fire. These zones must to be equipped by fixed fire detection systems to continuously monitor the conditions leading to fire outbreak. A novel reflectometry-based fire detection system is introduced for monitoring the engines of helicopters. The proposed approach is based on the measurement of the reflection caused by the characteristic impedance variation due to temperature changes. The measurement is achieved by an electronic board that implements Orthogonal Multi-tone Time Domain Reflectometry (OMTDR) [1,2]. Here, the cable serves not only as a physical support of reflectometry signal, but also as a fire detection sensor. The conductor resistance varies according to $R(t) = R_0 + [1 + \alpha(T - T_0)]$ where $R(T)$ is the resistance at elevated temperature $T$, $R_0$ is the obtained resistance at temperature $T_0$ and $\alpha$ is the temperature coefficient. The signature of a heating fault depends strongly on temperature $T$. The heating fault is characterized by a peak of very low amplitude on the reflectogram for $T=100^\circ C$. A direct measurement is complex to achieve since the peak is near to the parasitic peaks. A solution, a differential reflectogram is computed between a reference (at 20°C) and the current reflectogram. The proposed approach is sensitive enough to detect the heating fault at 100°C. Moreover, it is not disturbed by parasitic peaks thanks to the use of reference.

Main Results
The feasibility of the proposed solution is proved. The developed demonstrator includes an unshielded twisted pair EN 2346-005 with 15-meters length that is used in the onboard electrical systems of aircraft as shown in Figure 1. The fire outbreak, at each fire zone, is simulated by a thermo-regulated resistance and concerns 5 cm-length of the cable. Four potential fire zones are considered. The performance of our system is tested in the presence of a single or multiple fire zones simultaneously. Here, the detection of the start of a fire is performed within 15 seconds. Firstly, the cable is heated using a single resistance #1, #2, #3 or #4. Then, the cable is heated using the fourth resistances simultaneously. Despite the signal attenuation at each heating point, our fire detection system permits to detect and locate multiple fire zones without loss of accuracy (Figure 2).

Perspectives
For future works, the demonstrator will be tested in engines compartments with the presence of temperature, vibration, etc.

RELATED PUBLICATIONS:
This paper introduces the first embedded Smart Connector (SmartCo) able to detect and locate incipient defect such as chafing, bending radius, etc., and this even in complex wiring networks. Based on Orthogonal Multi-tone Time Domain Reflectometry (OMTDR) technology, the SmartCo uses the test signal to generate the reflectometry signal and to transmit useful information to a master connector. The proposed approach permits to aggregate data from a set of slave connectors. Indeed, since incipient defect is hard to distinguish from noise, the master performs an advanced post-processing to eliminate false alarms.

Electrical incipient defects represent 45% of faults found in aging aircraft (i.e. chafing (35%), broken insulation (5%), etc.). In practice, incipient defect is difficult to detect based on reflectometry since it is characterized by a very low amplitude on the measurements. Our main challenge is to diagnose such defects and this, even in complex wiring systems. Although interesting methods have been proposed to overcome this issue such as Self-Adaptive Correlation Method (SACM) [1], Signature Magnification by Selective Windowing (SMSW) [2], etc., they are prone to false alarms. An advanced post-processing is used. It combines several algorithms that stress the incipient defect [3]. In complex wiring networks, incipient defect diagnosis becomes much more challenging due to signal loss. Indeed, resistance of the cable, splices and connectors are the main causes for the attenuation of the signal. In such networks, ambiguity problems are added due to multiple paths phenomenon, which introduce complexity to the measurements analysis. To cope with these problems, a distributed diagnosis strategy is proposed where several connectors make reflectometry measurements alternately. Since the diagnosis function is distributed and integrated at different points in the network, the communication is enabled between those smart connectors for information aggregating to a central one. To do so, the OMTDR is suitable since it uses a signal carrying information [4]. The SmartCo fusion permits to maximize the network coverage, avoid fault location ambiguity, decrease blind zones and check SmartCo reliability.

For feasibility proof, an OMTDR-based demonstrator has been developed where wire diagnosis and data communication are embedded into a board card named “Ariane” as shown in Figure 1. The SmartCo fusion is used to transmit the reflectogram obtained by each slave to master SmartCo for post-processing. Experimental results obtained on a 30 m long STP EN 2714-013 MLB. A 5 mm long, 2 mm wide- chafing defect is present at 20 m from the injection port. From the sole measured reflectogram, the incipient defect detected is difficult since the reflected power is less than 0.1%. The advanced post-processing makes possible the detection and localization of the incipient defect with a detection probability of 80%.

For future works, the electronic board and this latter will be miniaturized for its integration into an aeronautical connector.
BASELINING: A CRITICAL APPROACH USED FOR SOFT FAULT DETECTION IN WIRE NETWORKS

RESEARCH TOPIC:
Wire networks, soft fault detection and location, clutter, difference system.

AUTHORS:
M. Kafal, J. Benoit.

ABSTRACT:
During the last few decades, vast efforts have been invested in projects related to the detection and location of faults in wire networks. Although, reflectometry based methods have been the standard choice for fault-detection techniques, their results can be of more difficult interpretation when soft faults and complex branched networks are addressed. Baselining, a natural fitting concept formed an important basis as a post-processing approach and succeeded in bypassing the aforementioned obstacles. In the following, we will verify that a soft fault’s signature can be easily masked by normal impedance variation in the cable’s environment.

Context and Challenges
Although, reflectometry-based methods have proven to be effective when dealing with hard faults e.g., short or open circuits, the detection and location of soft faults, e.g., chafes, cracks, etc. which are mainly characterized by their weak reflectivity, seemed to be difficult. Unfortunately, weak echoes produced by soft faults might be easily masked by either pseudo echoes generated by higher reflectivity impedance discontinuities (junctions) or any perturbation or noise. This situation is expected to worsen complexity with an increased network’s complexity. Baselining, an inverse problem based postprocessing technique, showed feasibility in enhancing the fault’s signature. It is based on comparing the response of the faulty network with either the pre-measured or simulated response of its (known) healthy condition. However, such a technique requires very controlled laboratory environments with no movement or vibration.

Main Results
The concept of baselining have been adopted from radar applications where to enhance the visibility of a target in the presence of clutter, a difference system is applied between that containing the target and clutter and a reference one containing just the clutter as demonstrated in Fig. 1. The same applies for soft faults present in complex networks where clutter can be resembled by junctions and other impedance discontinuities.

Fig. 1: An example of free space propagation: (a) with the absence of a target, and (b) with its presence.

Although, the baselining approach served in enhancing the detectability of soft faults, it is very difficult to obtain in a realistic environment. Indeed, archiving baseline networks in specific applications as that in aircraft systems is possible, but it is limited to a very small number of networks, as there exist enormous wiring configurations whose archiving process is a costly task. In addition, moving a wire, which is inevitable in high-vibration environments such as that in planes, can make impedance changes that are as large as or even larger than that of a soft fault. It was proven in [1] that simply moving a wire around in ways that could occur during regular use, created more impedance changes than frays (soft faults). In other words, regardless of how accurate and sensitive the systems are made, the environmental impedance changes can be expected to be larger than the soft fault impedance change, making it impossible to locate such frays using reflectometry methods.

Fig. 2: Reflectograms of the reference, faulty, and difference systems corresponding to the single-junction single-fault network while applying a change in the length of the first branch.

Perspectives
The most significant observation of [1] is that soft faults might have signatures, which are much weaker than impedance changes that can affect a cable during normal operation. As a result, implementing such techniques necessitates very controlled measurement environments, which is not always feasible in real life applications especially for embedded network configurations.

RELATED PUBLICATIONS:
SOFT FAULT DIAGNOSIS IN WIRE NETWORKS USING TIME REVERSAL CONCEPT AND SUBSPACE METHODS

RESEARCH TOPIC:
Wire networks, soft fault detection and location, time reversal concept (TR), multiple signal classification method (MUSIC).

AUTHORS:
M. Kafal, J. Benoit, A. Cozza (GeePs, CentraleSupelec), and L. Pichon (GeePs, CentraleSupelec)

ABSTRACT:
Time-reversal (TR) techniques recently adopted from acoustics to guided wave propagation along transmission lines have proven to be effective in locating soft faults in different wire network configurations. In the following, the TR multiple signal classification (TR-MUSIC) technique is applied to cable testing. TR-MUSIC is shown to provide spatial resolutions in the mm range, while using continuous-wave test signals when the well-known time-domain reflectometry methods would require large bandwidths in the order of gigahertz for a similar performance.

Context and Challenges
Guaranteeing a trustworthy usage of cables necessitates investing in techniques qualified for precisely detecting the presence of faults in wire networks. Although, we have witnessed during the last few decades a wide range of methods, reflectometry remained the flagship. Despite the promising results obtained when dealing with hard faults, difficulty in detection was encountered once soft faults are addressed. Regrettably, the situation worsened once branched complex networks were addressed. On the other hand, TR concept, the process of emitting a time-reversed version of a source’s captured response so as to retro-focus around it, was adapted to wire networks and showed efficiency in dealing with the aforementioned problems. Remarkably, MUSIC, a subspace method, succeeded in efficiently providing high spatial resolution estimates of single as well as multiple faults’ locations when applied in conjunction with TR processing which became better known as TR-MUSIC. Significantly, this can be accomplished using continuous wave excitations at even low frequencies.

Main Results
To ensure the proper application of the TR-MUSIC method, the network under test's (NUT) scattering matrix must be available. It can be obtained by measurement accomplished with a vector network analyzer. This matrix is then used in the TR-MUSIC procedure as detailed in [1,2], where computing what is referred to as pseudo-spectrum requires an a priori model of the voltage propagation at each position along the NUT. The position of the fault is thus determined by a singularity obtained at its location.

Experimental tests were conducted on different complexity NUT layouts affected by either a single or a double soft fault configuration. Fig. 1 presents a double junction NUT containing two soft faults at two different locations.

Fig. 2: Pseudo-spectrum of the TR-MUSIC applied to the NUT of Fig. 1.

The TR-MUSIC method is applied on the NUT of Fig. 1 whereas the pseudo-spectrum of Fig. 2 attests, two highly resolved singularities are obtained on the positions of the faults while suing a single frequency continuous wave at 25 MHz. Significantly, TR-MUSIC has proven its ability to provide a sub-millimeter resolution at the fault’s position while using a single frequency excitation. Eventually, this demonstrates the technique’s applicability in band limited systems without the need of complex testing setups.

Perspectives
Since TR-MUSIC do not depend on the chosen test frequency, it appears as more readily adaptable to the case of live testing, where the NUT would be under use over a given set of frequencies and thus test signals could ideally be allocated outside these frequencies. The same rationale clearly applies to Electromagnetic Compatibility constraints. Future work will also need to deal with the robustness of the TR-MUSIC and its sensitivity to the presence of noise.

Fig. 1: Double junction NUT affected by two soft faults.

Fig. 2: Pseudo-spectrum of the TR-MUSIC applied to the NUT of Fig. 1.

RELATED PUBLICATIONS:
A STATISTICAL STUDY OF DORT METHOD FOR LOCATING SOFT FAULTS IN COMPLEX WIRE NETWORKS

RESEARCH TOPIC:
Complex wire networks, DORT method, fault location, noise, statistical study

AUTHORS:
M. Kafal, J. Benoit, A. Cozza (GeePs, CentraleSupelec), and L. Pichon (GeePs, CentraleSupelec)

ABSTRACT:
Decomposition of the time reversal operator (DORT), a recently applied time reversal method to transmission lines, has shown to be effective in detecting and locating soft faults in complex wire networks. In the following, we will conduct a statistical study investigating the influence of several parameters, namely the number of testing ports in addition to noise, on the performance of DORT technique after having proposed a suitable fault location criterion. Notably, this would allow a closer inspection of the method’s practicability for future implementation in real-life networks.

SCIENTIFIC COLLABORATIONS: GeePs laboratory, CentraleSupelec

Context and Challenges
DORT method, a computational time reversal (TR) method, allows synthesizing testing signals from the analysis of the network’s TR operator (TRO) which are bound to focus on the fault's position once injected into the ports of a network under test (NUT). It is noteworthy that the process of locating a fault by DORT is accomplished by fetching the energy peak associated to the constructive interference of the injected testing signals on the position of the fault, along a predefined path of the NUT. Although, this allowed a clear identification when dealing with simple network structures and with a sufficient number of testing ports, an increasing complexity in the interpretation is noticed when the NUT’s complexity increases and when the number of testing ports decreases. Accordingly, we will propose a contrast enhancing method for a more accurate location of soft faults in complex NUTs. This will be followed by a statistical investigation studying the influence of the number of testing ports on the DORT’s performance.

Main Results
A fault location criterion is derived and well established in [1]. It is based on defining each of the delimiting intervals in the energy diagram of an NUT followed by calculating the plateau related to each one of them along different paths of the network.

After that we proceed to the calculation of the contrast of the local peaks as the ratio of the amplitude of the peak to that of the corresponding plateau previously calculated. Consequently, the highest contrast peak hints at the location of the fault. For illustration purposes, if we consider a double junction network affected by a soft fault on one of its branches, applying the aforementioned criterion produces the energy diagram presented in Fig. 1. An important fact is pointed out: we are able to directly pinpoint the location of the soft fault using only two testing ports. It is situated at 2 m on the third or fourth path of the network.

TABLE I
LOCATION PROBABILITIES PMLE OF THE FAULTS IN THE NUT OF FIG. 1 IN TERMS OF THE NUMBER OF TESTING PORTS.

<table>
<thead>
<tr>
<th>Fault</th>
<th>Number of Testing Ports</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>F1</td>
<td>0</td>
</tr>
<tr>
<td>F2</td>
<td>0</td>
</tr>
<tr>
<td>F3</td>
<td>0</td>
</tr>
</tbody>
</table>

The previous example showed that the number of testing ports influences the performance of DORT method. In fact, a change in their number might allow or not to locate the faults in an NUT. Consequently, a statistical study has been conducted to analyze this parameter [1], where a database containing a large number of NUTs has been considered. In particular, table I shows that one testing port leads to a null location probability, whilst adding more testing ports increases this probability so as it converges to 1 once all the NUT’s testing ports are used. Thus, based on the DORT properties and the preliminary analysis obtained, one can expect to have a higher location probability when testing ports exist from both sides of the fault and a low location probability in the opposite cases.

Perspectives
An element influencing the DORT performance method has been investigated through a statistical study. In this event, it would also be greatly interesting to conduct a more extended study; starting with a wider database, including more network configurations, as well as an extension of the influencing parameters to include the bandwidth, losses, noise, model perturbations, etc.
"DETECTING VERY SHORT INTERMITTENT FAULTS IN ELECTRICAL WIRES"

RESEARCH TOPIC:
Intermittent faults, wires, reflectometry, chaos

AUTHORS:
F. Auzanneau, C. Layer

ABSTRACT:
Intermittent faults are temporary malfunctions appearing at random or periodic moments with normal behavior in between. They are measurable symptoms of the degradation of a system. Their frequency and severity increase over time, eventually leading to permanent faults. In order to optimally plan preventive maintenance, it is essential to catch intermittent faults soon enough to repair them before they turn permanent. This paper shows that the recent Chaos Time Domain Reflectometry method can detect intermittent short circuits of a duration of a few microseconds or less.

Context and Challenges
Live wire diagnosis is required to detect failures for the health monitoring of embedded cables. This is most important when Intermittent Faults (IFs) occur, because they are impossible to replicate in laboratory. If a maintenance operator must repair a system that was declared faulty and cannot find any anomaly, a phenomenon called "No Fault Found" (NFF), the defective unit is sent back to service unchanged, which may cause a potential threat to safety. NFF is one of the most time consuming and expensive maintenance operations, reportedly responsible for 90% of all maintenance costs in avionics.

Contrary to other reflectometry methods, Chaos Time Domain Reflectometry (CTDR) shows the capacity of detecting very short IFs, down to less than 1µs, with a good noise robustness.

Main Results
An experimental device is realized to create duration controlled short-circuits in a 30 meters long cable (Fig. 1). A mechanical relay cannot be used because the commutation time is too slow when the metallic part moves, and the contact bounces a few times before becoming permanent.

An FPGA controls the system, i.e., the signal generator, the generation and duration of the trigger signal that makes the transistors commute, the recording of the signal before, during and after the short-circuit and the reflectogram's computation.

Experimental measurements show the possibility to detect and locate intermittent short-circuits of 1 microsecond. A theoretical model predicts that even shorter faults can be detected (down to 200 ns), and that their minimal duration is related to the sampling frequency of the Digital to Analog Converter (Fig. 2).

The inherent noise robustness of the CTDR method is an important asset for IF detection. This method enables the very early detection of the graceful degradation of a connection system.

Perspectives
Recent works have shown that it is possible to not only detect and locate an IF but also to characterize it, by estimating its time of appearance and its duration. This was made possible by an improvement of the CTDR method, which continuously injects a signal, computes the short-time reflectogram and analyses the peaks created by the part of the signal trapped between the IF and the end of the line. An accuracy of 2 µs for the time of appearance and duration of a 15 µs IF was obtained.

RELATED PUBLICATIONS:
CONSTRUCTING THE TOPOLOGY OF COMPLEX WIRE NETWORKS USING REFLECTOMETRY RESPONSES AND OPTIMIZATION BASED ALGORITHMS

RESEARCH TOPIC:
Wiring networks, topology reconstruction, time domain reflectometry, optimization-based algorithms

AUTHORS:
M. Kafal, J. Benoit, A. Dupret, C. Layer, G. Beck

ABSTRACT:
Despite the fact that wireless systems are overwhelming most of nowadays applications, wiring networks are still forming a pivotal anchor in many controls and security units. Thus, ensuring the reliable operation of such networks necessitates investing in techniques dedicated for their protection. A vast majority of such methods rely on the presence of a reference model of the network. Accordingly, we propose a nondestructive testing approach based on the tenets of reflectometry methods and optimization based algorithms to retrieve the topology and load impedances of unknown embedded complex wiring networks.

Context and Challenges
The proper functioning and good performance electrical systems is of critical importance. Reflectometry methods deployed relatively simple test setups for the analysis of a network under test (NUT) and provided valuable information about any impedance discontinuity (presence, position, and nature). In general, the accuracy of these methods is affected by several factors including: pre-fault system conditions, testing restrictions (disconnecting the NUT, predefined loads at the extremities, etc.) and most importantly the network’s topology. Consequently, analyzing networks of unknown topology can therefore be particularly very difficult. Therefore, the aim is to propose an efficient automated method to retrieve all parameters of an unknown or partially unknown wiring network, namely its topology branch lengths and extremities’ load impedances.

Main Results
The proposed method considers the tested NUT as a black-box with a definite number of extremities. The approach consists of two main parts.

- Generate all possible hypothesized network topologies \( NT_{hyp} \).
- Direct Numerical Model (DNM) control \( V_{ref} \) for each \( NT_{hyp} \).
- Generate initial population
- Particle Swarm Optimization (PSO) techniques
- Sort \( P_{best} \) and eliminate \( NT_{hyp} \) with highest \( P_{best} \).
- \( P_{best} = 0 \)
- No
- Best Network
- Generate Algorithms

In order to validate the proposed approach, we opted to test it against numerical and experimental data. Different complexity NUTs have been considered with different types of cables. The obtained results have proven the effectiveness and feasibility of the approach in precisely reconstructing unknown complex networks. Significantly, this can be clearly noticed from the close agreement of the reflectometry responses acquired for a tested double junction NUT as demonstrated in Fig. 2. The surprising ability of the proposed approach in ensuring a 1 cm length precision along with a 5 ohms impedance accuracy in a relatively short processing time (tens of seconds) is good news, as no similar performance has been presented so far [1,2].

Perspectives
Future work will need to deal with integrating the network connections as an additional variable set in the optimization algorithm and embrace the location and characterization of soft and hard faults.

RELATED PUBLICATIONS:
EMERGING TECHNOLOGIES & PARADIGMS

- Spiking Neurons
- Unsupervised learning
- Memory circuits in advanced technologies
- Crosspoint non volatile memories
- Content addressable memories
- Adiabatic Logic using MEMS
- Monolithic 3D technology, partitioning and thermal
ARCHITECTURE EXPLORATION OF A FIXED POINT COMPUTATION UNIT USING PRECISE TIMING SPIKING NEURONS

RESEARCH TOPIC:
Spiking Neural Network, Precise Timing, Event Driven processing, DSP, Architecture

AUTHORS:
Thomas Mesquida, Alexandre Valentian, David Bol (ICTEAM) and Edith Beigne

ABSTRACT:
This proposed device is intended to fulfill the role of a Digital Signal Processor in the spiking domain, becoming an essential tool to Spiking Neuromorphic Sensors such as Dynamic Vision Sensors. Our approach is based on the use of Spiking Neural Networks with preset topology and weights in order to implement Fixed Point basic arithmetic and signal processing functions using the timing of spikes to convey information. The proposed architecture supporting this application is composed of synchronous clusters of neurons served by an Asynchronous Network on Chip in order to take advantage of the networks particularities. We demonstrate here how requirements generated by the developed topologies lead our way through the architecture exploration.

SCIENTIFIC COLLABORATIONS: ICTEAM, Université catholique de Louvain

Context and Challenges
Event driven sensors are keys for breakthrough in IoT devices. Legacy sensors are designed to record frames of their environment at a frequency corresponding to the data. Their event driven counterparts are designed to send asynchronous events using Address Event Representation (AER) only when there is data to be sent. Each pixel in a Dynamic Vision Sensor (DVS) will send asynchronous event only when it records an input intensity change or will send it with temporal interval proportional to the intensity, depending on the DVS design. Such sensors tend to have lower idle power and produce lower processing workload as the data rate is directly lower linked to information.

However, processing solutions answering needs from those sensors are not mature yet. Spiking Neural Networks (SNN) inspired from brain offer promising possibilities [1] as they can directly exploit data contained in AER events. In order to match applicative needs, complete SNNs can be trained or dedicated SNN building blocks can be assembled as operators are assembled to build a function [2]. We explore in this paper architectures possibilities for supporting SNNs derived from [2].

Main Results
Using the frequency of the events can lead to flaws when implementing operators [3]. Fig.1 shows main strategies when for retrieving data from spikes. Precise Timing (PT) SNN operators use a few fixed number of spikes per operation when frequency coded operators figures are linked to required accuracy.

Architecture have been thought to answer those special needs. Using Xnet [2] and a SystemC model, we developed the GALS architecture in Fig.2, which computes AER events. Synchronous clusters process stored spikes in a 4-stage pipeline implemented to minimize hardware cycles used per active neuron. Those pipeline stages end with the mutualized computation unit responsible for the update of neurons’ membrane potential.

Perspectives
A fully synchronous cluster will be taped out in FDSOI 28nm as proof of concept. Future work will also focus on local DVFS usage and designing SNNs for specific vision and audio applications.

RESEARCH TOPIC:
Spiking Neural Network, Precise Timing, Event Driven processing, DSP, Architecture

AUTHORS:
Thomas Mesquida, Alexandre Valentian, David Bol (ICTEAM) and Edith Beigne

ABSTRACT:
This proposed device is intended to fulfill the role of a Digital Signal Processor in the spiking domain, becoming an essential tool to Spiking Neuromorphic Sensors such as Dynamic Vision Sensors. Our approach is based on the use of Spiking Neural Networks with preset topology and weights in order to implement Fixed Point basic arithmetic and signal processing functions using the timing of spikes to convey information. The proposed architecture supporting this application is composed of synchronous clusters of neurons served by an Asynchronous Network on Chip in order to take advantage of the networks particularities. We demonstrate here how requirements generated by the developed topologies lead our way through the architecture exploration.

SCIENTIFIC COLLABORATIONS: ICTEAM, Université catholique de Louvain

Context and Challenges
Event driven sensors are keys for breakthrough in IoT devices. Legacy sensors are designed to record frames of their environment at a frequency corresponding to the data. Their event driven counterparts are designed to send asynchronous events using Address Event Representation (AER) only when there is data to be sent. Each pixel in a Dynamic Vision Sensor (DVS) will send asynchronous event only when it records an input intensity change or will send it with temporal interval proportional to the intensity, depending on the DVS design. Such sensors tend to have lower idle power and produce lower processing workload as the data rate is directly lower linked to information.

However, processing solutions answering needs from those sensors are not mature yet. Spiking Neural Networks (SNN) inspired from brain offer promising possibilities [1] as they can directly exploit data contained in AER events. In order to match applicative needs, complete SNNs can be trained or dedicated SNN building blocks can be assembled as operators are assembled to build a function [2]. We explore in this paper architectures possibilities for supporting SNNs derived from [2].

Main Results
Using the frequency of the events can lead to flaws when implementing operators [3]. Fig.1 shows main strategies when for retrieving data from spikes. Precise Timing (PT) SNN operators use a few fixed number of spikes per operation when frequency coded operators figures are linked to required accuracy.

Architecture have been thought to answer those special needs. Using Xnet [2] and a SystemC model, we developed the GALS architecture in Fig.2, which computes AER events. Synchronous clusters process stored spikes in a 4-stage pipeline implemented to minimize hardware cycles used per active neuron. Those pipeline stages end with the mutualized computation unit responsible for the update of neurons’ membrane potential.

Perspectives
A fully synchronous cluster will be taped out in FDSOI 28nm as proof of concept. Future work will also focus on local DVFS usage and designing SNNs for specific vision and audio applications.

RESEARCH TOPIC:
Spiking Neural Network, Precise Timing, Event Driven processing, DSP, Architecture

AUTHORS:
Thomas Mesquida, Alexandre Valentian, David Bol (ICTEAM) and Edith Beigne

ABSTRACT:
This proposed device is intended to fulfill the role of a Digital Signal Processor in the spiking domain, becoming an essential tool to Spiking Neuromorphic Sensors such as Dynamic Vision Sensors. Our approach is based on the use of Spiking Neural Networks with preset topology and weights in order to implement Fixed Point basic arithmetic and signal processing functions using the timing of spikes to convey information. The proposed architecture supporting this application is composed of synchronous clusters of neurons served by an Asynchronous Network on Chip in order to take advantage of the networks particularities. We demonstrate here how requirements generated by the developed topologies lead our way through the architecture exploration.

SCIENTIFIC COLLABORATIONS: ICTEAM, Université catholique de Louvain

Context and Challenges
Event driven sensors are keys for breakthrough in IoT devices. Legacy sensors are designed to record frames of their environment at a frequency corresponding to the data. Their event driven counterparts are designed to send asynchronous events using Address Event Representation (AER) only when there is data to be sent. Each pixel in a Dynamic Vision Sensor (DVS) will send asynchronous event only when it records an input intensity change or will send it with temporal interval proportional to the intensity, depending on the DVS design. Such sensors tend to have lower idle power and produce lower processing workload as the data rate is directly lower linked to information.

However, processing solutions answering needs from those sensors are not mature yet. Spiking Neural Networks (SNN) inspired from brain offer promising possibilities [1] as they can directly exploit data contained in AER events. In order to match applicative needs, complete SNNs can be trained or dedicated SNN building blocks can be assembled as operators are assembled to build a function [2]. We explore in this paper architectures possibilities for supporting SNNs derived from [2].

Main Results
Using the frequency of the events can lead to flaws when implementing operators [3]. Fig.1 shows main strategies when for retrieving data from spikes. Precise Timing (PT) SNN operators use a few fixed number of spikes per operation when frequency coded operators figures are linked to required accuracy.

Architecture have been thought to answer those special needs. Using Xnet [2] and a SystemC model, we developed the GALS architecture in Fig.2, which computes AER events. Synchronous clusters process stored spikes in a 4-stage pipeline implemented to minimize hardware cycles used per active neuron. Those pipeline stages end with the mutualized computation unit responsible for the update of neurons’ membrane potential.

Perspectives
A fully synchronous cluster will be taped out in FDSOI 28nm as proof of concept. Future work will also focus on local DVFS usage and designing SNNs for specific vision and audio applications.

RELATED PUBLICATIONS:
COMPRESSED SIGNAL ACQUISITION IN WIRE DIAGNOSTIC

RESEARCH TOPIC:
Reflectometry, compressive sensing, Shannon-Nyquist frequency, sparsity, Fractional Fourier Transform (FrFT).

AUTHORS:
Tzila Ajamian, Antoine Dupret and Saïd Moussaoui (ECN)

ABSTRACT:
Reflectometry is a diagnostic technique that allows detection of defects in cables. Analog-to-Digital Converters (ADC) are one of the most required components in the architectures of reflectometry systems. However, the respect of the sampling theorem and the need of performing real-time analysis restrict the highest frequency of the injected signals, thus preclude the detection of some kind of defects. The contribution of this paper is to introduce an alternative ADC based on Compressive Sensing (CS) of the reflected signal in order to overcome the limits of Nyquist. Therefore, a complex linear chirp signal is injected, a dictionary of sparsity based on Fractional Fourier Transform (FrFT) is used and a Random Demodulator is employed as an analog encoder. The main result is that the new ADC scheme allows to reconstruct the reflectogram with a sampling frequency of $18/N$ Nyquist limit.

Context and Challenges
Cables are subject to aggressive environmental conditions, which might cause several defects such as short or open circuits. To inspect their status, a reflectometry system injects a signal in the damaged wire and then analyzes the reflected waveforms to locate and characterize the damage. To achieve an accurate localization, high frequency signals should be injected in the wires and the reflected signal according to the Nyquist-Shannon theorem should be sampled at a frequency of several hundred MHz. The objective of this paper is to introduce a compressed acquisition of the reflected signal in order to avoid the limits of Nyquist. The proposed compressive sampling scheme is based on an analog encoder also known as Analog to Information Converter (AIC) that requires the definition of an appropriate dictionary of sparsity. The main idea is to reconstruct perfectly $x \in \mathbb{R}^M$ from an observation $y = \Phi x$, where $y \in \mathbb{R}^M$, with $M$ much lower than $N$ and $\Phi \in \mathbb{R}^{M \times N}$ is a measurement matrix, modeling the sub-sampling and verifying the incoherency property. Actually, the signal $x$ should be sparse, otherwise it will be expressed in a basis $\Psi$, such that $x = \Psi s$ with $s$ containing few non-zero coefficients. In order to adapt the CS to reflectometry, here the purpose is to define both matrices $\Psi$ and $\Phi$.

Main Results
In our study, the injected signal is a linear complex chirp allowing spanning a wide range of frequency band. Moreover, a FrFT with an order $\alpha = 1.69$ of this signal will approve a sparse representation. Since, a hard defect causes a total reflection no matter what the frequency of the signal is, the rate of variation of the $x_r(\sigma)$ frequency is identical to that of $x_\sigma(\sigma)$. Therefore, the same fractional order $\alpha$ calculated for the injected signal $x_\sigma(\sigma)$ can be used for the reflected one $x_r$. In this case, $x_r = \Psi s_r$ and $\Psi = \Phi \alpha$. To implement the idea of CS in the analog domain, the AIC is proposed. The Random Demodulator (RD) architecture is adapted of the reflectometry signals which defines the measurement matrix $\Phi$. This methodology is typically decomposed into three steps: modulation, filtering and uniform sampling. In the Figure 1, we represent the proposed new architecture for reflectometry combined with the RD (as seen in red lines).

Figure 2: Reflectogram corresponding to $f_{\text{sys}}/8$

Despite the presence of the secondary lobes for $f_{\text{sys}}/8$ on the reflectogram, Figure 2, we can reconstruct and measure the distance of the defect located at 100m from the point of injection, which is fundamental for reflectometry.

Perspectives
The perspective of this work is to apply the compressed acquisition for different kinds of signal as well as to validate our simulations by an experimental part.
PROPOSED SOLUTIONS FOR MEMORY CIRCUITS IN ADVANCED TECHNOLOGIES

RESEARCH TOPIC:
Volatile and non-volatile memories, SRAM, CAM, crosspoint, advanced technologies, 3D CoolCube™, TFET and RRAM

ABSTRACT:
Emerging CMOS and non-volatile memory (NVM) technologies pave the way to new circuit design solutions that can thus enhance energy efficiency, performance and silicon area. In this study, we investigated memory circuits in emerging technologies: SRAM in monolithic 3D CoolCube™ technology, CAM/SRAM in 28nm FD-SOI technology, SRAM and Flip-Flop in TFET technology, assist technique for crosspoint memory using resistive RAM (RRAM) technology.

SCIENTIFIC COLLABORATIONS: ISSEP(1), IM2NP(2)

Context and Challenges
The increasing demand for highly energy efficient systems in IoT markets brings new circuit design challenges that can be solved with advanced solutions adapted for new MOS and/or memory technologies. In this report, only memory circuits will be discussed by exposing the proposed approaches and solutions for SRAM, CAM, Flip-Flop and crossbar in FD-SOI, 3D CoolCube™, TFET and RRAM technologies [1].

Main Results
High density SRAM bitcell using monolithic 3D CoolCube™ technology has been analyzed. This 4T Driver-Less (DL) shows 30% area reduction with regards to 0.078µm planar 6T bitcell. The proposed configuration (Fig. 1-a) enables a smart dynamic back bias (BB) that can adjust the transistor Vt of the bitcell in a non-symmetrical manner. Therefore each bitcell will have unbalanced access transistors according to the stored data in the internal nodes which corresponds to a data-dependent back biasing (DDBB). This BB will be done without area penalty thanks to the fact that the used 3D technology enables to implement the NMOS access transistors on the top tier and the PMOS pull-up transistors on the bottom one. This leads to no need for area dedicated to bias the back plan (BP) or Well as the signals used to polarize the BP of the NMOS located on the top tier are coming directly from the bottom tier.

The DDBB increases the leakage current of access transistor connected to internal node storing a ‘0’, leading to more stable retention and read operations that allow a functionality from -10°C to 125°C, from 0.7V to 1V (Fig. 1-b) [1-2].

Regarding the study on reconfigurable CAM/SRAM, 1.56GHz operation frequency with 0.13fJ/bit per search has been demonstrated on silicon in 28nm FD-SOI [3].

Besides, TFET technology has allowed proposing novel SRAM and Flip-Flop architectures that present drastic reduction of leakage power while maintaining similar performances [1].

Concerning crosspoint memory, metal line resistance combined with high programing current conducts to a significant voltage drop (Fig. 2) depending on the accessed bitcell location. The resulting degradation of the NVM resistance variability can be mitigated by implementing a compensation circuit that can modify the injected programing voltage. The programing voltage will be thus shifted either positively or negatively, according to accessed bitcell location by monitoring its address [1] [4]. The energy will be therefore reduced and the endurance of the entire array will be enhanced, while mitigating the programing failure rate.

Fig. 2 Voltage drop in programing operation in crosspoint memory

Perspectives
New technologies available in Leti such as sequential 3D, NVM, TFET, etc. offer the opportunity to revisit the conventional or disused circuit designs, leading to new design paradigms and promising solutions adapted to the highly-constrained market.

Fig. 1 (a) 4T Driver-Less (DL) SRAM bitcell exploiting the data-dependent back biasing (DDBB) in 3D CoolCube™ technology (b) functional space with and without DDBB

RELATED PUBLICATIONS:
INNOVATIVE MODELING FOR DESIGN SPACE EXPLORATION FOR CROSSPOINT MEMORY CIRCUIT

RESEARCH TOPIC:
Non-volatile memories: RRAM and Vertical-RRAM, crosspoint architecture

AUTHORS:
A. Levisse, B. Giraud, J.P. Noel, M. Moreau (IM2NP), J.M. Portal (IM2NP)

ABSTRACT:
This work focus on exploring the possibilities and limits of crosspoint memory based on RRAM (Resistive RAM) or PCM (Phase Change Memory) technologies. This design space exploration is made possible thanks to innovative modeling of crosspoint circuit features such as voltage drop, area overhead and timing estimation. The study takes into account technology scaling and multiple bit-write in the same bitcell array in conventional (i.e. planar) crosspoint and in Vertical-RRAM technologies.

SCIENTIFIC COLLABORATIONS: IM2NP

Context and Challenges
High quantity of data needed for IoT market require non-volatile memory (NVM) for both connected devices and servers. New technologies, such as Resistive RAM (RRAM) appear as promising candidates to replace Flash memories whose process complexity leads to scaling issues and high fabrication cost.

Main Results
A key parameter for properly designing a Resistive RAM (RRAM) chip is the RC delay of bitcell array. In [1] and more precisely in [3], an analytical methodology is developed to quickly estimate this parameter as a function of array size for crosspoint memory. For that purpose, the lateral capacitance value (the one between two metal line on the same level) and the vertical capacitance (the one between 2 consecutive metal level) are estimated.

Figure 1 illustrates the relationship between the metal half-pitch and the array size (in number of lines). The increase of the metal resistivity with the technology scaling is taken into account. The half-pitch variation has a limited impact on the timing on the contrary to the number of lines. It is worth to note that a reduction by 10x of the number of lines reduces by 100x the charge/discharge time.

Increase of bit number to program at the same time in the same array. The sum of selector leakage current can be limited by reducing the array size. It also can be reduced by using a more efficient selector but at the expense of a better material stack (costly in term of development time). The needed programing current can also be reduced by improving the material stack but more efficiently by increasing the programing time.

Fig. 2 Relationship between maximum selector leakage current and maximum RRAM programing current for 1 to 8 bits written per array for metal half pitch F=25nm.

In [1-2], we developed models to assess the limitations of the crosspoint memory architecture either in conventional crosspoint made of Back End of Line (BEoL) RRAM or in Vertical-RRAM. In [2], there are clear relationships between the number of stacked planes and their thicknesses. Indeed, these two parameters have a direct impact on the access resistance and thus on the voltage drop for a fixed programing current. The number of bits per array is maximized for a number of stacked planes ranging from 50 to 150 when the plane thickness is ranging from 10nm to 100nm.

Fig. 1 Charging and discharge times of a crosspoint array versus array size and metal half pitch F using the timing estimation model described in [1] and detailed in [3].

Fig. 2 shows that to limit the area of the peripheral circuit (programing drivers, ...) in order to keep the same area of the bitcell array (100% area overhead), the programing and leakage currents must be limited. This limit must be reduced with the increase of bit number to program at the same time in the same array.

Perspectives
These works will lead to full memory macro design, manufacturing and test. The used technology will be either RRAM or PCM, stacked with Ovonic Threshold Switching selector.

This array made entirely in the BEoL can be stacked on top of an SRAM or CAM memory. These volatile memories will be used most of the time by the system cores whereas the crosspoint memory will store permanently data and/or read instructions. A more advanced study could consist in embedding computation features inside these volatile memories.

RELATED PUBLICATIONS:
In this work we propose a high-speed 6T Re-CAM/SRAM (Reconfigurable CAM/SRAM) with new energy efficient sensing technique. The implementation is compatible with compact 6T-SRAM foundry bitcells requiring only minor modifications on the bitcell level in order to pass the ground connection vertically along the column and split the wordlines into two separate connections. Test-macro of 8kb is implemented in 28nm FDSOI CMOS and reaches up to 1.56GHz at 0.9V with 0.13fJ/bit energy consumption per search, giving an improvement of between 4.2x and 6.24x and between 4.6x and 14.3x as compared to the similar designs in the state of the art, respectively.

**ABSTRACT:**

In this work we propose a high-speed 6T Re-CAM/SRAM (Reconfigurable CAM/SRAM) with new energy efficient sensing technique. The implementation is compatible with compact 6T-SRAM foundry bitcells requiring only minor modifications on the bitcell level in order to pass the ground connection vertically along the column and split the wordlines into two separate connections. Test-macro of 8kb is implemented in 28nm FDSOI CMOS and reaches up to 1.56GHz at 0.9V with 0.13fJ/bit energy consumption per search, giving an improvement of between 4.2x and 6.24x and between 4.6x and 14.3x as compared to the similar designs in the state of the art, respectively.

**SCIENTIFIC COLLABORATIONS:** ISEP

**Context and Challenges**

In modern SoC, the embedded memory can easily dominate the overall chip area and power consumption. Not only many applications require a large amount of SRAM but some will also need the CAM to be available on-chip. In order to accommodate larger variety of applications on the same chip while reducing the total memory footprint, we propose a reconfigurable CAM/SRAM macro to be implemented as a part of embedded memory [1]. At the same time, we address another important issue relative to large amount of embedded memory which is the memory dominating the overall power consumption. To this end, the proposed memory uses only a 6T bitcell and a novel read/match approach allowing a major reduction of read/search energy as compared to the state of the art.

**Main Results**

Fig. 1 depicts the organization of the SRAM/CAM array. SRAM words are horizontal and CAM words vertical. CAM search is executed by passing the data pattern by activating one of the wordlines per row, setting RD_EN to GND and keeping the bitlines precharged to VDD. If all cells store ‘1’ on the side where the wordline was activated the VVSS for that column stays at GND.

![Fig.1 Reconfigurable CAM/SRAM array organization](image)

Should at least one of the cells store ‘0’ on the side of the activated wordline, the VVSS for that column will increase as the current will pass through the cell from the bitline to VSS, thus indicating a miss. SRAM read is performed in the same way, corresponding effectively to a CAM match or single bit miss.

This approach is very energy efficient as only VVSS will shift by a limited voltage value while all other nodes remain precharged. Measurement revealed energy/search/bit as low as 0.13fJ showing an improvement of over 4x versus state of the art designs. As indicated in Fig.2, the speed of read/search is very fast, requiring only a 190ps wordline pulse to correctly detect a single bit miss (or read an SRAM cell). The memory remains functional down to 0.4V, with application of write assist techniques below 0.6V. As in the test macro the controller was not implemented, taking an approximation on the decoding delay, we estimated the operation frequency of 1.56GHz at 0.9V giving over 4x improvement as compared to other CAM macros.

![Fig.2 Minimum wordline pulse for single bit miss detection (silicon measurement data) [1]](image)

**Perspectives**

Further work on this subject should consist in implementing a full reconfigurable CAM/SRAM macro in a full system and benchmarking the gain in performance and area on the system level.

**RELATED PUBLICATIONS:**

TUNNEL FET BASED REFRESH-FREE DRAM

RESEARCH TOPIC: DRAM, Tunnel FET, TFET, Negative Differential Resistance, eDRAM

AUTHORS: N. Gupta (ISEP/CEA), A. Makosiej, A. Vladimirescu (ISEP), A. Amara (ISEP), C. Anghel (ISEP)

ABSTRACT:
A refresh free and scalable ultimate DRAM (uDRAM) bitcell and architecture is proposed for embedded applications [1]. The uDRAM 1T1C bitcell is designed using access Tunnel FETs taking advantage of their negative differential property (NDR) feature. By exploiting the NDR and capacitor leakage, a refresh free DRAM bitcell behavior is achieved. uDRAM allows scaling of storage capacitor by 87% and 80% in comparison to DDR and eDRAMs, respectively. Bitcell area of 0.0275μm² is obtained in 28nm FDSOI-CMOS and is scalable further with technology shrink. Owing to the removal of refresh, we estimate the gain in throughput of 3.8% to 18% in comparison to CMOS DRAMs.

SCIENTIFIC COLLABORATIONS: ISEP

Context and Challenges
In order to address the problem of area and power consumption of SRAMs in modern SoCs, the use of embedded DRAMs (eDRAMs) as an alternative has been explored in recent years. While the eDRAMs provide a higher-than-SRAM density with relatively easy fabrication with CMOS technology for digital logic, they suffer from lower performance relative to intrinsic DRAM operation principle and the requirement of refresh operation. Another important challenge is related to the capacitor scaling. This is related to the increasing device leakage with technology shrink and the need to maintain the refresh frequency reasonably low. Conversely, in eDRAMs capacitor size is reduced at the cost of retention time in order to optimize the cost of process cost and silicon footprint.

Main Results
In order to address those challenges we proposed a TFET refresh free bitcell. The bitcell, same as standard DRAM cell consists of one transistor and one capacitor. The bitcell takes advantage of the unique TFET properties, notably the fact that it is unidirectional and under reverse VDS bias it exhibits a NDR property with a current profile as depicted in Fig. 1. Refresh-free retention is obtained by ensuring the sufficient ratio between the NDR current (I_{NDR}), capacitor leakage (I_{CAPL}) and the TFET leakage (I_{OFF}), as depicted in Fig. 2.

Fig. 2 depicts the mechanisms for data retention for storing ‘0’ (a) and ‘1’ (b). In both cases the wordline is maintained at high voltage (here 1V) in order to ensure large enough value of the NDR current (Fig. 1, region I). In case (a), the capacitor leakage tries to charge the node Q. Once the voltage of node Q increases the access transistor enters NDR range, thus discharging it to 0V through the bitline (BL) that is maintained at 0V. In case (b) the access transistor is reverse biased in region II (Fig. 1) where the TFET current is at minimum. Therefore, as long as the TFET I_{OFF} is much lower than the capacitor leakage I_{CAPL}, the high logic state in the bitcell is maintained.

The read operation is performed in the same fashion as in the standard DRAM cell, i.e. by charge sharing between the bitline and the bitcell capacitor and thus it is also destructive and requires a data write-back. The size of the bitcell capacitance is now only enforced by ensuring sufficient voltage drop during read and not by long data retention and thus a much smaller capacitor can be used of up to 87% as compared to standard DRAMs.

Write operation is done by applying a pulse on the G node and modifying the values of the corresponding bitlines and collapsing all wordlines except the accessed line and is detailed in [1]. As a consequence the energy consumption of this design during write is expected to be larger than that of the standard DRAM and remain in total on a similar level for both read and write operations. Owing to the removal of refresh we estimate a gain in throughput while maintaining a very high memory density.

Fig.2 Refresh free data retention conditions for storing ‘0’ (a) and ‘1’ (b)

Perspectives
Further work on the subject will be a silicon demonstrator, as soon as the technology reaches sufficient maturity.

RELATED PUBLICATIONS:
CONTACTLESS ENERGY-RECOVERY LOGIC COMPUTATION USING ELECTROMECHANICAL DEVICES

RESEARCH TOPIC: Adiabatic logic, energy-recovery computation, MEMS

ABSTRACT: A new logical operation called capacitive adiabatic logic (CAL) has been proposed to break the inherent tradeoff between leakage and conduction losses in existing digital circuits. We report a particular implementation of this paradigm based on electromechanical device. Any digital processing unit can be built in MEMS-based CAL including i) cascade logical operations, ii) fan-out, iii) any Boolean and sequential operations, iv) information storage. Thanks to adiabatic operation, our proposal can almost fully recover the energy to code the information, thus these research results pave the way of energy-dissipation breakthrough.

SCIENTIFIC COLLABORATIONS: Delft Tech. Univ., ESIEE, LIRMM

Context and Challenges
Considerable efforts have been devoted to the design of low-power digital electronics. However, after decades of improvements and maturation, CMOS technology could face a power efficiency ceiling. This is due to the trade-off between leakage and conduction losses inherent to the limited subthreshold slope of any semiconductor junction e.g. FET. Consequently, the lowest dissipation per operation (~10^-15 W) remains nowadays few decades higher than the theoretical Landauer’s limit (3.6 fJ at 300 K). CMOS scaling, better electrostatic gate control (FinFET, FDSOI…) or alternative FET structure (SET) do not drastically change this previous statement. It is therefore clear that there is room to introduce new devices eventually based on new logic operation concept to dramatically reduce the power dissipation of digital operations.

On logic operation side, energy-recovery computation (also called adiabatic) is a good candidate for reducing the dynamic losses due to charge-based information coding. But adiabatic operation reduces operating frequency, thus exacerbating the inherent leakage loss of semiconductor devices (FET). Consequently, we need to propose new device or break the previously mentioned tradeoff to work in adiabatic way.

In [1-4], we propose in a new logic operation called contactless adiabatic logic (CAL) to break the inherent trade-off between existing leakage and conduction losses by removing the need of electrical contact: the devices are always in OFF region.

Main Results
In CAL approach, the elementary device is a four-terminal variable capacitance element (VC) instead of resistive one e.g. FET. VC has two voltage terminals to control the capacitance across two other terminals. The main interesting property of this new logic family is that no electrical contact occurs during logic operations, meaning the family is inherently immune to leakage losses [1]. It is a hardware independent proposal: the device can be implemented in MEMS, by modifying FETs or any element which permits to vary a capacitance value from two independent electrodes.

To experimentally prove the concept of CAL, VCs have been implemented using a well-known MEMS structure: the comb drive actuator [2-4]. To have 4-terminal element, we have designed two mechanically coupled (electrically isolated) comb drives. Compared to other MEMS based computation (such as µ-relay), no mechanical contact is needed, thus CAL improves the mechanical reliability and scalability potential (no adhesion force).

By driving adiabatically a VCs pipeline (gradual information coding and erasing @10’s Hz), we succeed to recover 99% of mechanical energy stored in the double comb-drives during Boolean operations. The uncompressed loss is only due to the leakage of electrical field between the comb. As we are now limited by the process lithography (2µm) and by the spring sizing ratio, the loss ceiling is about 1µJ per operation which is still far from nanoscale FETs (10’s aJ). A basic gate area e.g. inverter is mm² scale and its operation frequency is limited by the resonant frequency of comb drive’s mobile part (1’s kHz).

We also show how we can organize CV elements in order to create all combinational logic gates (AND, XOR…), pipeline structure or more complex digital operations (N-bits adder, Fredkin gate…). Additionally, we have proposed comb-drives structure to perform combinational operation (ANR, XOR…) using single device. Furthermore, volatile and non-volatile memories are available to build a processor structure.

We have also built on a specific computer aided design (CAD) flow to speed up the electromechanical simulations of multiples CV based gates. Fitted on finite-element simulations (MEMS+), some Verilog-A models included in a SPICE environment can be simulated to predict the operation of multiple MEMS gates combination such as N-bits adder.

As any adiabatic logic operation, reversible and power-efficiency power-clock generation has also been studied [5].

Perspectives
This proposal paves the way to an alternative operation, called CAL, by breaking the inherent tradeoff between leakage and dynamic losses. Based on comb drive MEMS, it is possible to implement CAL-based processor. The next step is to process thinner comb drive (100nm lithography node) to achieve the same energy dissipation than actual FET-based logic.

RELATED PUBLICATIONS:
THERMAL PERFORMANCE OF COOLCUBE™ MONOLITHIC AND TSV-BASED 3D INTEGRATION PROCESSES

RESEARCH TOPIC: 3D technologies - Coolcube™ - Thermal impact

AUTHORS: Cristiano Santos, Pascal Vivet, Sébastien Thurillès, Olivier Billoint, Jean-Philippe Colonna, Peresoval Coudrain, Lee Wang

ABSTRACT: CoolCube™ is a monolithic 3D technology which has the potential to solve the interconnection density limitation of the existing TSV-based 3D integration processes. Since the active devices are fabricated on extremely thin die substrates, heat dissipation has been pointed as a potential showstopper issue for this emerging technology. This work provides a comparative study of the thermal performance of the CoolCube and TSV-based 3D integration processes. Results show that CoolCube exhibits thermal performance similar to or even better than the TSV-based technologies thanks to its very tight die-to-die thermal coupling.

SCIENTIFIC COLLABORATIONS: Mentor Graphics

Context and Challenges
Power density and associated thermal dissipation is one the main challenges of 3D technologies. Due to the increased number of die layers, a too large power density may arise, that would imply a large thermal flux, which might be difficult to evacuate using regular packaging technologies. As a result, thermal runaway could arise that would potentially destroy the 3D circuits.

The thermal dissipation of regular TSV-based 3D technologies has already been widely studied. A systematic thermal exploration [1] 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.

However, little work has been carried out on thermal study of more advanced 3D technologies. In this work [4], we perform a systematic comparative study of the thermal performance of the 3 different 3D technologies: the advanced CoolCube process (CoolCube), the advanced fine pitch hybrid bonding Integration scheme (CuCu-DB), and finally the more regular TSV-based 3D integration scheme (Cu-pillar). The thermal comparison between the 3 different technologies is performed for multiple-die stacks and considering a full range of technology parameters and different application scenarios.

Main Results
In terms of thermal modelling, this work uses a new thermal modelling simulation tool, from Mentor Graphics, which allows to cover the whole thermal simulation steps: from early thermal exploration - as done here - to final thermal sign-off, including 3D-stack dies, with its package, board and boundary conditions (heatsink, fan), with detailed static or dynamic power map. The thermal methodology is presented in more details in [2].

As a result, the Copper Pillar technology show poor results for thermal hot spots, due to underfill layers which have poor thermal conductivity. On the other hand, Copper Hybrid technologies and CoolCube technology present a very good thermal coupling between the different circuit layers, where the impact of interconnect density is rather limited. More details and results can be found in [4].

Perspectives
The very tight CoolCube die-to-die thermal coupling allows easy heat dissipation of internal hot spots without any noticeable temperature difference between the stacked layers.

This proposed thermal exploration comparison study needs to be pursued by real architecture explorations and validated with silicon measurements.

Figure 1: 3D Stacking cross-section and associated technological parameters for 3 different 3D technologies for thermal comparison

Peak Temperature (°C)

Figure 2: Peak temperature in the case of hot spot dissipation considering very few die-to-die connections or max connection density between tiers, and associated Thermal maps of the middle and topmost tiers for a) Cu-pillars, b) Cu-DB, and c) CoolCube.

RELATED PUBLICATIONS:
**ABSTRACT:**

CoolCube™ is a monolithic 3D (M3D) technology offering a vertical density of integration 20 times higher than face to face copper hybrid bonding (F2F Cu-Cu), thanks to ultra-thin Monolithic Inter-tier Vias (MIVs). In this work, we propose a new partitioning tool exploiting this characteristic for 2-tier Cell-on-Cell ICs before placement. It is based on a fast and iterative algorithm that explores the space of solutions and minimizes the estimated cost of wires with balanced area between tiers without limiting the number of MIVs. A mathematical formulation of the 3D partitioning problem and a comprehensive framework, based on simulated annealing (SA) algorithm coupled with a dedicated cost function, are detailed and compared with Min-Cut (MC) partitions commonly used. It appears that our solution can decrease the estimated total cost of wires by 41% and 45% for the LDPC and FFT/AES units. It also reduces the total cost of wires by 30% to 44% compared to the MC algorithm for the same units and with no significant increase in runtime.

**Main Results**

As seen in Fig. 2, the number of cuts (MIVs) is closely linked to the hyperedge weight reduction (compared to the 2D hyperedge weight taken as a reference). Indeed, the MC algorithm has a small number of cuts and it almost does not reduce the hyperedge weight, around 1% or even 11% in reduction, while our SA algorithm can reduce the hyperedge weight up to 45%. The MC partitioning have a total hyperedge weight close to the 2D design hyperedge weight because, as its name implies, it minimizes the number of cuts. Therefore, the MC partitioning returns 2 graphs that are almost independent, thus having two 2D circuits. Since the sum of the hyperedge weight of two 2D circuits is the same as the 2D original design, the MC hyperedge weight is close to the 2D hyperedge weight. As expected, the MC algorithm minimizes the number of cuts while our SA algorithm tends to maximize them in order to reduce the total hyperedge weight.

**Perspectives**

The presented work is a first step toward a full Monolithic 3D placer that consists in future works to address in order to get signoff Physical Implementation for Monolithic 3D - CoolCube™
ABSTRACT:
This paper focuses on test and characterization of Cu-Cu 3D interconnects and proposes smart test vehicles to measure several informations after bonding: perfect alignment, misalignment (direction, value) and contact resistance. These test structures are implemented as stand-alone structures for process development and also in an application circuit to assess performance of 3D-IC thanks to the DFT infrastructure. This work includes simulation and logic synthesis results of both implementations.

Context and Challenges
3D integration technology consists in interconnecting the integrated circuits in three dimensions. 3D stacking reduces the interconnect length and increases the performance in terms of RC delay, power consumption and form factor. Cu-Cu direct bonding is very promising for fine pitch and high density interconnects, with pitch scaling in the 1-2µm range.

However, despite the rapid development of the new stacking technologies bonding is very challenging, and some bonding defects between chips will still remain especially when targeting the smallest pitches.

As a consequence, the following defects can be observed: shift in one or two directions, rotation, which are leading to electrical micro-voids or even contact failure.

Main Results
The mis-alignment test structure is proposed: it is composed of a top Cu-pad (Red) and four Bottom Cu-pads (blue). In case of misalignment, top & bottom pads are in contact. Using various pitch and offset values, it is then possible using current probing to extract the mis-alignment value. It can be used as a stand-alone structure for process development and qualification, but it can also be integrated in an application circuit to assess performance of the final 3D-IC thanks to a complete 3D Design-for-Test (DFT) infrastructure [1] using a dedicated BIST engine [2].

Perspectives
The test structure has been designed and silicon measurement will be carried out to validate it.

On the longer term, it is envisaged to study other test structures for high density TSV, and as a final objective to pursue system level Design-for-Test architecture, with fault tolerance and repair mechanisms.

RELATED PUBLICATIONS:
• Simon EMBERGER
• Mauricio ALTIERI-SCARPATO
• Alexandre LEVISSE
• Camille DUPOIRON
• Emma CHABCHOUB
• Thierno BARRY
• Julie DUMAS
• Soundous CHAIRAT
• Nicolas MONNIER
• Armande CAPITAINE
• Anthony GOAVEC
• Vincent LORRAIN
• Hela GUESMI
• Julien COLLET
• Adja Ndeye SYLLA
• Guillaume GOURLAT
Acquiring the depth of a scene in addition to its image is a desirable feature for many applications which depend on the near environment. The state of the art in the field of depth extraction offers many methods, but very few are well adapted to small embedded systems. Some of them are too cumbersome because of their large optical system. Others might require a delicate calibration or processing methods which are difficult to implement in an embedded system. In this PhD thesis, we focus on methods with low hardware complexity in order to propose algorithms and optical solutions that extract the depth of the scene, provide a relevance evaluation of this measurement and produce all-in-focus images. We show that Depth from Focus (DfF) algorithms are the most adapted to embedded electronics constraints. This method consists in acquiring a cube of multi-focus images of the same scene for different focusing distances. The images are analyzed in order to annotate each zone of the scene with an index relative to its estimated depth. This index is then used to build an all in focus image.

The continuous scaling of transistor dimensions has increased the sensitivity of digital circuits to PVT variations and, more recently, to aging effects such as BTI and HCI. Large voltage guard bands, corresponding to worst-case operation, are thus necessary and leads to a considerable energy loss. Current solutions to increase energy efficiency are mainly based on Adaptive Voltage and Frequency Scaling (AVFS). However, as a reactive solution, it cannot anticipate the variation before it occurs; It has to be improved for handling long-term reliability issues. This thesis proposes a new methodology to generate simplified but nevertheless accurate models to estimate the circuit maximum operating frequency Fmax. A model is created for the modelling of the propagation delay of the critical path(s) as a function of PVT variations. Both BTI/HCI effects are modelled as a shift in the parameters of the first model. Built on the top of device-level models, it takes into account all factors that impact global aging, namely, circuit topology, workload, voltage and temperature variations. The proposed method is evaluated on two architectures in 28nm FD-SOI. Finally, an aging-aware recalibration method has been developed for a particular V T monitor. This allows the evaluation of different strategies regarding performance, energy and reliability, under aging variations, of a multicore AVFS circuit.
ALEXANDRE LEVISSE
3D HIGH DENSITY MEMORY BASED ON EMERGING RESISTIVE TECHNOLOGIES: CIRCUIT AND ARCHITECTURE DESIGN
Université Aix Marseille (France)

While conventional non-volatiles memories, such as floating gate Flash memories, are becoming more and more difficult and costly to integrate and suffer of reduced performances and reliability, emerging resistive switching memories (RRAM), such as OxRAM, CBRAM, MRAM or PCM, are seen in the scientific community as a good way for tomorrow’s high-density memories. However, standard RRAM architectures (such as 1 Transistor-1 RRAM) are not competitive with flash technology in terms of density. Thereby, this thesis proposes to explore the opportunities opened by transistor-less RRAM architectures: Crosspoint and Vertical RRAM (VRRAM) architectures. First, the positioning of Crosspoint and VRRAM architectures in the memory hierarchy is studied. New constraints such as the sneak-path currents, the voltage drop through the metal lines or the periphery area overhead are identified and modeled. In a second time, circuit solutions answering to previously mentioned effects are proposed. Finally, this thesis proposes to explore new opportunities opened by the use of innovative transistors to improve the density or the performances of RRAM-based memory architectures.

CAMILLE DUPOIRON
NEWS PARADIGMS FOR JOINED IMAGE ACQUISITION AND PROCESSING IN ADVANCED NANOMETRIC PROCESS
Université Grenoble Alpes (France)

The goal of this thesis is to study new image acquisition paradigm in integrated vision circuits to enhance their robustness and scalability using nanometric technologies (such as the 28nm FDSOI) in order to satisfy the imaging constraints imposed by applications such as Internet of Things. In this case, a heterogeneous system-on-chip (SoC) designed in advanced technology would meet the energy consumption constraints. Using standard imagers is not compatible with this requirement because of their excessive power consumption and their architectures non-compatible with 28nm FDSOI technologies. Two event-based architectures were studied during this thesis in order to fit with the constraints of an implementation in 28nm FDSOI: a "Time-to-first-Spike" (TTFS) architecture with an inhibition system and an architecture called "multi-bus "using the dense interconnections possibilities offered by the technology. These two architectures aim to reduce the data throughput as well as energy consumption. Finally, a part of this thesis has been dedicated to the exploration of FDSOI 28nm capabilities in terms of pixel implementation. Notably, by studying pixels using a photodiode...
EMMA CHABCHOUB
NEW SENSOR READOUT ARCHITECTURES IN SOI TECHNOLOGY DEDICATED TO HIGH TEMPERATURE APPLICATIONS (> 225°C)

Université de Montpellier (France)

The design of analog integrated circuits for high temperature applications is not easy because of the exponential increase of the leakage currents and because of the large drift of the transistor parameters which implies a large drift of the biasing points and a sharp degradation of analog functions at high temperature. On the other hand there is a demand for smart sensors dedicated to application working in harsh environment. The purpose of the thesis was to study a time-domain architecture which minimizes the number of analog functions. The proposed sensor interface converts the differential sensor output voltage into a phase-shift by means of a pair of Injection Locked Oscillators (ILOs). The phase-shift is then digitized with the help of a counter. As the ILO frequency tracks the frequency of the on chip reference oscillator, the overall sensor readout is insensitive to temperature variations. The sensor interface has been fabricated using a 0.18µm PD-SOI technology from XFab. The fabricated sensor interface measurements showed low temperature dependence as expected. Worst case measured thermal variation is about 65ppm/°C over the +/- 40 mV input full scale and over a wide operation temperature range extended from -10°C to 220°C.

THIERNO BARRY
SECURING SOFTWARE AGAINST FAULT ATTACKS AT COMPILE TIME

Université de Lyon (France)

Embedded systems are increasingly present in our daily lives (e.g. credit cards, smartphones and biometric passports). Given the sensitivity of the data they handle, the safety of these systems has become a major concern for industry and state organizations. It is established that a fault injection in an embedded system can compromise the security of the data it contains, for example obtaining a secret key or bypassing an authentication mechanism. The goal of this thesis is the automatic generation of software protections against fault injection attacks on embedded systems. The source and binary approaches consist in inserting the protections respectively in the source and binary code of the application; this thesis explores the use of a compilation approach
In the context of High Performance Computing, cache coherence protocol scalability problem for parallel architecture is also a problem for on chip architecture, following the emergence of many-cores architectures. There are two protocol classes: snooping and directory-based. Protocols based on snooping, which send coherence information to all caches, generate a lot of messages whose few are useful. On the other hand, directory-based protocols send messages only to caches which need them. To scale, a coherence protocol must produce a reasonable number of messages and limit hardware resources used by the coherence and in particular for the sharing set.

To evaluate and compare protocols and their sharing set, we first propose a method based on trace injection in a high-level cache model. This method enables a very fast architectural exploration of cache coherence protocols. We then propose a new dynamic sharing set for cache coherence protocols, which is fully scalable. With 64 cores, 93% of cache blocks are shared by up to 8 cores. Several algorithms for coherent rectangle placement are proposed and evaluated, and compared with the state of the art.

Wireless sensor network (WSN) have experienced an incredible success these past years, especially due to the Internet of Thing (IoT) paradigm, which opened the door to much more interesting applications. The wireless sensor network nodes (WSNN) are used in nearly all smart houses applications, as a network of wearables or as entertainment devices. This keen interest in WSN is not without consequences, as many of these applications require from the node to be autonomous and thus energy efficient. One of the most promising solutions is the integration of adaptive blocks in the node, which can adapt their performances and thus their energy expenditure according to the application, environment or the energy budget. The aim of this work is to efficiently transfer the control data and the sense&react data throughout the node to and from the corresponding adaptive blocks. An asynchronous communication network is proposed offering a fast and independent wake and sleep mode, mainly the transfer of configuration data throughout a network, in an event-driven fashion, hence the use of the QDI asynchronous logic. This network is digital only and two versions were designed, a serial and a hybrid one, and the serial version was implemented in FDSOI 28nm.
This thesis is focused on the design of several THz image sensors, sensitive to the sub-THz band (100GHz – 1THz). Those sensors are fully integrated monolithically with a standard CMOS technology. The aim is to obtain a low cost, robust and reliable sensor, capable to work at room temperature with real time acquisition constraint. Each sensor is composed of an integrated antenna thanks to the top available metallic layers, and with a ground plane coupled with an electronic band-gap (EBG) in order to isolate the readout circuit and the substrate of the CMOS circuit from the antenna. The antenna is co-designed with its rectifying THz NMOS transistor in order to have the best adaptation. A total of 17 test cases, 16 with 3x3 pixels array and a 9x9 pixels array have been designed. Those cases implement several types of antennas with different magnetic surfaces (ground planes and EBG). Several rectifying NMOS have also been integrated and address a fully high performance signal processing chain (amplification, enhancement of SNR).

NICOLAS MONNIER
DESIGN OF HIGH PERFORMANCES CMOS MONOLITHIC THZ IMAGE SENSORS
Université de l’IMT Atlantique (France)

Regarding the ever-increasing number of remote sensors, harvesting the energy from the surrounding environment is an advantageous solution to meet the energy needs and reduce the use of conventional batteries. However, in various area such as seafloors, the lack of sunlight, vibrations and temperature gradients hinders the use of conventional energy harvesting systems. The microbial fuel cell (MFC) is a system harvesting the biochemical energy from a large range of organic substrates (compost, sewage, seafloor sediment...) thanks to the metabolism of electro-active bacteria present in the surroundings. To adapt and store the MFC energy to power a sensor, a harvesting electrical interface is required, i.e. a power management unit (PMU) including a DC/DC converter extracting and shaping the energy prior to storage. During this thesis, we designed lab-made BMFCs with environmental conditions to the natural ones. Then we studied their electrical behavior by characterizing them with a potentiostat. The polarization curves highlighted a maximum power point and allowed as to determine a static electrical model. Using impedance spectroscopy we also identified a dynamic model of the BMFC revealing a double RC behavior.

ARMANDE CAPITAINE
ENERGY HARVESTING FROM MICROBIAL FUEL CELLS
Université de Lyon (France)
This thesis focuses on the power emission constraints defined by regulations and standards for every kind of ultra-wide band impulse radio transmitters. In fact, these power emission constraints have to be respected all along the device life. Also, an integrated sensor able to extract the essential information for an on-chip estimation of the power spectral density has been realized. Then, an embedded algorithm is added to the system and detects if a power limit is broken. If necessary, it acts on the transmitter to solve the problem. In the first chapter, a large variety of power constraints shapes and several architectures of impulse generators have been observed and studied. Therefore, the aim of this thesis is to realize a calibration system which would be universal to all impulse radio transmitters. After its extraction at the output of the transmitter, information have to be down converted in order to reduce the constraints on conversion stage but without using a local oscillator and a mixer. A model for the impulse signal based on the instantaneous envelop and on the instantaneous frequency has been proposed in the second chapter. A new calibration method based on these two signals is also presented. The last chapter concentrates on detailing the extraction of the instantaneous envelop and the instantaneous frequency. The design of the electronic devices essential to this extraction is presented and a chip has been realised and the viability of the solution shown.

In this thesis, we study the dedicated computational approaches of deep neural networks and in particular the convolutional neural network (CNN). We highlight that the efficiency of convolutional neural networks is an interesting choice for many applications. We study the different implementation options for this type of networks and characterize their computational complexity. We show that the computational complexity of this type of structure can be incompatible with the constraints of embedded systems. To address this issue, we explore various models of neurons and architectures in order to minimize the resources required for embedded applications. In a first step, we explore the benefits we can expect by changing the model of the neurons. We show that the spiking models can reduce the computational complexity while offering interesting dynamical properties. However, this model requires a complete rethinking of the hardware architecture. We propose a novel architecture based on the spiking model to implement convolutional neural networks. We have set up a software and hardware simulation toolchain in order to explore the different paradigms of computation and hardware implementation and evaluate their suitability to embedded environments. The toolchain allows us to validate computational aspects and evaluate the relevance of our architectural choices. We validate our theoretical approach by simulating the proposed architecture with a 28 nm FDSOI technology node. We show that our proposed approach is both computationally efficient and technologically interesting in terms of scaling and dynamical precision.
With the increasing complexity of embedded applications, their a-posteriori validation becomes, at best, a major factor in the development costs and, at worst, simply impossible. It is therefore necessary to define a method that allows the development of correct-by-construction systems while simplifying the specification process. High-level component-based design frameworks that allow design and verification of hard real-time systems are very good candidates for structuring the specification process as well as verifying the high-level model. The goal of this thesis is to couple a high-level component-based design approach based on the BIP (Behaviour-Interaction-Priority) framework with a safety-oriented real-time execution platform implementing the TT approach (the PharOS Real-Time Operating System). To this end, we propose an automatic transformation process from BIP models into applications for the target platform (i.e. PharOS). The process consists in a two-step semantics-preserving transformation. The first step transforms a BIP model coupled to a user-defined task mapping into a restricted one, which lends itself well to an implementation based on TT communication primitives. The second step transforms the resulting model into the TT implementation provided by the PharOS RTOS. We provide a tool-flow that automates most of the steps of the proposed approach and illustrate its use on an industrial case study for a flight Simulator application and a medium voltage protection relay application. In both applications, we compare functionalities of both original, intermediate and final model in order to confirm the correctness of the transformation. For the first application, we study the impact of the task mapping on the generated implementation. And for the second application, we study the impact of the transformation on some performance aspects compared to a manually written version.

This thesis work deals with the exploration of graph-processing algorithms on distributed architectures, notably using GraphLab, a state of the art graph-processing framework. Two use-cases are considered. For each, a parallel implementation is proposed and deployed on several distributed architectures of varying scales. This study highlights operating ranges, which can eventually be leveraged to appropriately select a relevant operating point with respect to the datasets processed and used cluster nodes. Further study enables a performance comparison of commodity cluster architectures and higher-end compute servers using the two use-cases previously developed. This study highlights the particular relevance of using clustered commodity workstations, which are considerably cheaper and simpler with respect to node architecture, over higher-end systems in this applicative context. Then, this thesis work explores how performance studies are helpful in cluster design for graph-processing. In particular, studying throughput performances of a graph-processing system gives fruitful insights for further node architecture improvements. Moreover, this work shows that a more in-depth performance analysis can lead to guidelines for the appropriate sizing of a cluster for a given workload, paving the way toward resource allocation for graph-processing. Finally, hardware improvements for next generations of graph-processing servers are proposed and evaluated. A flash-based victim-swap mechanism is proposed for the mitigation of unwanted overloaded operations. Then, the relevance of ARM-based microservers for graph-processing is investigated with a port of GraphLab on a NVIDIA TX2-based architecture.
In the context of pervasive computing and internet of things, systems are heterogeneous, distributed and adaptive (e.g., transport management systems, building automation). The design and the deployment of these systems are made difficult by their heterogeneous and distributed nature but also by the risk of conflicting adaptation decisions and inconsistencies at runtime. Inconsistencies are caused by hardware failures or communication errors. This thesis proposes a middleware support, called SICODAF, for the design and the deployment of reliable adaptive systems. SICODAF combines a behavioral reliability (absence of conflicting decisions) by means of transition systems and an execution reliability (absence of inconsistencies) through a transactional middleware. SICODAF is based on autonomic computing. It allows to design and deploy an adaptive system in the form of an autonomic loop which consists of an abstraction layer, a transactional execution mechanism and a controller. SICODAF supports three types of controllers (based on rules, on continuous or discrete control theory). SICODAF also allows for loop reconfiguration, to deal with changing objectives in the considered system, and the integration of a hardware failure detection system. Finally, SICODAF allows for the design of multiple loops for systems that consist of a high number of entities or that require controllers of different types.

The extreme sensitivity of nano electro mechanical system (NEMS) to atomic scale physical variations has led to the breakthrough development of NEMS-based mass spectrometry systems capable of measuring a single molecule. Parallel sensing using thousands of devices will help to circumvent the small effective sensing area while opening new perspectives for applications which require spatial mapping. While the development of NEMS-CMOS co-integration technology is of paramount importance to achieve high density sensor arrays (>1000 devices), the readout circuitry capable of tracking NEMS resonator frequency shifts is still the limiting factor for the very large scale integration of individually addressed sensors. Moreover, in order to resolve the mass and position of an adsorbed analyte, single particle mass sensing applications require to track simultaneously and in real time at least two modes of the resonators. To respond to the size, power consumption and resolution constraints linked to NEMS array measurement, this work focuses on the development of a new readout architecture based upon a dual mode heterodyne oscillator.
Greetings

Editorial Committee
- Pascal Vivet
- Christian Gamrat
- Edith Beigné
- Marc Belleville
- Jean-René Lèquepeys

Edition Management
- Caroline Vacher
- Hélène Vatouyas
- Caroline Arnaud

Technical Committee
- Suzanne Lesecq
- Valentin Gherman
- Belgacem Ben Hedia
- Gilles Sicard
- Dominique Morche
- Gaël Pillonnet
- Thierry Goubier
- Karim Ben Chehida
ARCHITECTURE, IC DESIGN AND EMBEDDED SOFTWARE

Contacts

Christian GAMRAT
International expert, Scientific director – Paris-Saclay
christian.gamrat@cea.fr

Jean-René LEQUEPEYS
Head of Architecture, IC design & Embedded Software division
jean-rene.lequepeys@cea.fr

Thomas SIGNAMARCHEIX
Sr business development manager
thomas.signamarcheix@cea.fr

Pascal VIVET
Research director, Scientific director - Grenoble
pascal.vivet@cea.fr