Activity Report 2017

Project-Team CAIRN

Energy Efficient Computing ArchItectures

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

Table of contents

Project-Team CAIRN

Creation of the Project-Team: 2009 January 01

> CAIRN *is located on two campuses: Rennes (Beaulieu) and Lannion (*ENSSAT*).*

Keywords:

Computer Science and Digital Science:

A1.1. - Architectures
A1.1.1. - Multicore, Manycore
A1.1.2. - Hardware accelerators (GPGPU, FPGA, etc.)
A1.1.8. - Security of architectures
A1.1.9. - Fault tolerant systems
A1.1.10. - Reconfigurable architectures
A1.1.12. - Non-conventional architectures
A1.2.5. - Internet of things
A1.2.6. - Sensor networks
A2.2. - Compilation
A2.2.1. - Static analysis
A2.2.4. - Parallel architectures
A2.2.5. - GPGPU, FPGA, etc.
A2.2.6. - Adaptive compilation
A4.4. - Security of equipment and software
A8.10. - Computer arithmetic

Other Research Topics and Application Domains:

B4.5. - Energy consumption
B4.5.1. - Green computing
B4.5.2. - Embedded sensors consumption
B6.2.2. - Radio technology
B6.2.4. - Optic technology
B6.6. - Embedded systems
B8.1. - Smart building/home
B8.1.1. - Energy for smart buildings
B8.1.2. - Sensor networks for smart buildings

1. Personnel

Research Scientists

Olivier Sentieys [Team Leader, Senior Researcher (DR) Inria, HDR]
François Charot [Inria, Researcher, Rennes]
Tomofumi Yuki [Inria, Researcher, Rennes]

Faculty Members

Emmanuel Casseau [Professor, Univ. Rennes, ENSSAT, Lannion, HDR]
Daniel Chillet [Professor, Univ. Rennes, ENSSAT, Lannion, HDR]
Steven Derrien [Professor, Univ. Rennes, ISTIC, Rennes, HDR]

Cédric Killian [Associate Professor, Univ. Rennes, IUT, Lannion]
Angeliki Kritikakou [Associate Professor, Univ. Rennes, ISTIC, Rennes]
Patrice Quinton [Ecole Normale Supérieure de Rennes, Emeritus, Rennes]
Christophe Wolinski [Professor, Univ. Rennes, Director of ESIR, Rennes, HDR]

Post-Doctoral Fellows

Mansureh Shahraki Moghaddam [Inria, Rennes, from Dec 2017]
Lei Mo [Inria, Rennes, from Apr 2017]
Atef Dorai [Univ. Rennes, ATER, ENSSAT, Lannion, until Aug 2017]
Imen Fassi [Univ. Rennes, Rennes, until Sep 2017]
Ashraf El-Antably [Inria, Lannion, until May 2017]
Imran Wali [Univ. Rennes, Lannion, until May 2017]

PhD Students

Gabriel Gallin [CNRS, granted by CominLabs, from Oct. 2014]
Jiating Luo [Univ. Rennes, granted by China Gov., from Nov. 2014]
Van Dung Pham [Inria, granted by CominLabs, Lannion, from Dec. 2014]
Aymen Gammoudi [Univ. Rennes, Lannion, from Sep. 2015]
Rafail Psiakis [Univ. Rennes, MENRT grant, from Oct. 2015]
Simon Rokicki [Univ. Rennes, granted by ENS Rennes, from Oct. 2015]
Audrey Lucas [CNRS, granted by DGA-PEC, Lannion, from Jan. 2016]
Genevieve Ndour [Univ. Rennes, granted by CEA Leti, Grenoble, from May 2016]
Joel Ortiz Sosa [Inria, Lannion, from Oct. 2016]
Nicolas Roux [Inria, granted by Brittany Region/LTC, Lannion, from Oct. 2016]
Mael Gueguen [Univ. Rennes, MENRT grant, Rennes, from Nov. 2016]
Minh Thanh Cong [Univ de Rennes, granted by USTH, Rennes, from May 2017]
Thibaut Marty [Univ de Rennes, granted by H2020 ARGO and Brittany Region, Rennes, from Sep. 2017]
Petr Dobias [Univ de Rennes, MENRT grant, Lannion, from Oct. 2017]
Van Phu Ha [Inria, granted by ANR Artefact, Rennes, from Nov. 2017]
Gaël Deest [Univ. Rennes I, MENRT grant, Rennes, until Jan. 2017]
Rengarajan Ragavan [Univ. Rennes I, Lannion, until Jan. 2017]
Xuan Chien Le [Inria, granted by Brittany Region/LTC, Lannion, until Mar. 2017]
Baptiste Roux [Inria, granted by DGA and Inria, Rennes, until Sep. 2017]
Benjamin Barrois [Univ. Rennes I, MENRT grant, Lannion, until Dec. 2017]
Kleanthis Papachatzopoulos [Inria, Rennes, until Mar. 2017]
Tara Petric [Inria, Rennes, until May 2017]

Technical staff

Arnaud Carer [Research Engineer (half time), Univ. Rennes, Lannion]
Pierre Guilloux [Univ. Rennes, Lannion]
Pierre Halle [Inria, from Nov. 2017]
Ali Hassan El Moussawi [Inria, Rennes]
Mickael Dardaillon [Univ. Rennes, Rennes, from Dec. 2017]
Thomas Lefeuvre [Univ. Rennes, Rennes, until Dec. 2017]
Christophe Huriaux [Univ. Rennes, Rennes, until Dec. 2017]

Administrative Assistants

Nadia Derouault [Assistant, Inria, Rennes]
Emilie Carquin [Assistant, Univ. Rennes, ENSSAT, Lannion]

2. Overall Objectives

2.1. Overall Objectives

Abstract — The CAIRN project-team researches new architectures, algorithms and design methods for flexible, secure, fault-tolerant, and energy-efficient domain-specific system-on-chip (SOC). As performance and energy-efficiency requirements of SOCs, especially in the context of multi-core architectures, are continuously increasing, it becomes difficult for computing architectures to rely only on programmable processors solutions. To address this issue, we promote/advocate the use of reconfigurable hardware, i.e., hardware structures whose organization may change before or even during execution. Such reconfigurable chips offer high performance at a low energy cost, while preserving a high level of flexibility. The group studies these systems from three angles: (i) The invention and design of new reconfigurable architectures with an emphasis on flexible arithmetic operator design, dynamic reconfiguration management and low-power consumption. (ii) The development of their corresponding design flows (compilation and synthesis tools) to enable their automatic design from high-level specifications. (iii) The interaction between algorithms and architectures especially for our main application domains (wireless communications, wireless sensor networks and digital security).

Keywords — **Architectures:** Embedded Systems, System-on-Chip, Reconfigurable Architectures, Hardware Accelerators, Low-Power, Computer Arithmetic, Secure Hardware, Fault Tolerance. **Compilation and synthesis:** High-Level Synthesis, CAD Methods, Numerical Accuracy Analysis, Fixed-Point Arithmetic, Polyhedral Model, Constraint Programming, Source-to-Source Transformations, Domain-Specific Optimizing Compilers, Automatic Parallelization. **Applications:** Wireless (Body) Sensor Networks, High-Rate Optical Communications, Wireless Communications, Applied Cryptography.

The scientific goal of the CAIRN group is to research new hardware architectures for domain-specific SOCs, along with their associated design and compilation flows. We particularly focus on on-chip integration of specialized and reconfigurable accelerators. Reconfigurable architectures, whose hardware structure may be adjusted before or even during execution, originate from the possibilities opened up by Field Programmable Gate Arrays (FPGA) [57] and then by Coarse-Grain Reconfigurable Arrays (CGRA) [60], [72] [1]. Recent evolutions in technology and modern hardware systems confirm that reconfigurable systems are increasingly used in recent and future applications (see e.g. Intel/Altera or Xilinx/Zynq solutions). This architectural model has received a lot of attention in academia over the last two decades [63], and is now considered for industrial use in many application domains. One first reason is that the rapidly changing standards or applications require frequent device modifications. In many cases, software updates are not sufficient to keep devices on the market, while hardware redesigns remain too expensive. Second, the need to adapt the system to changing environments (e.g., wireless channel, harvested energy) is another incentive to use runtime dynamic reconfiguration. Moreover, with technologies at 28 nm and below, manufacturing problems strongly impact electrical parameters of transistors, and transient errors caused by particles or radiations also often appear during execution: error detection and correction mechanisms or autonomic self-control can benefit from reconfiguration capabilities.

As chip density increased, power or energy efficiency has become "the Grail" of all chip architects. With the end of Dennard scaling [67], multicore architectures are hitting the *utilisation wall* and the percentage of transistors in a chip that can switch at full frequency drops at a fast pace [61]. However, this unused portion of a chip also opens up new opportunities for computer architecture innovations. Building specialized processors or hardware accelerators can come with orders-of-magnitude gains in energy efficiency. Since from the beginning of CAIRN in 2009, we advocate the interest of heterogeneous multicores, in which general-purpose processors (GPPs) are integrated with specialized accelerators, especially when built on reconfigurable hardware, which provides the best trade-off between power, performance, cost and flexibility. During the period, it therefore turns out that the time has come for these heterogeneous manycore architectures.

> Standard multicore architectures enable flexible software on fixed hardware, whereas reconfigurable architectures make possible **flexible software on flexible hardware**.

However, designing reconfigurable systems poses several challenges: the definition of the architecture structure itself, along with its dynamic reconfiguration capabilities, and its corresponding compilation or synthesis tools. The scientific goal of CAIRN is therefore to leverage the background and past experience of its members to tackle these challenges. We propose to approach energy efficient reconfigurable architectures from three angles: (i) the invention and the design of new reconfigurable architectures or hardware accelerators, (ii) the development of their corresponding compilers and design methods, and (iii) the exploration of the interaction between applications and architectures.

3. Research Program

3.1. Panorama

The development of complex applications is traditionally split in three stages: a theoretical study of the algorithms, an analysis of the target architecture and the implementation. When facing new emerging applications such as high-performance, low-power and low-cost mobile communication systems or smart sensor-based systems, it is mandatory to strengthen the design flow by a joint study of both algorithmic and architectural issues.
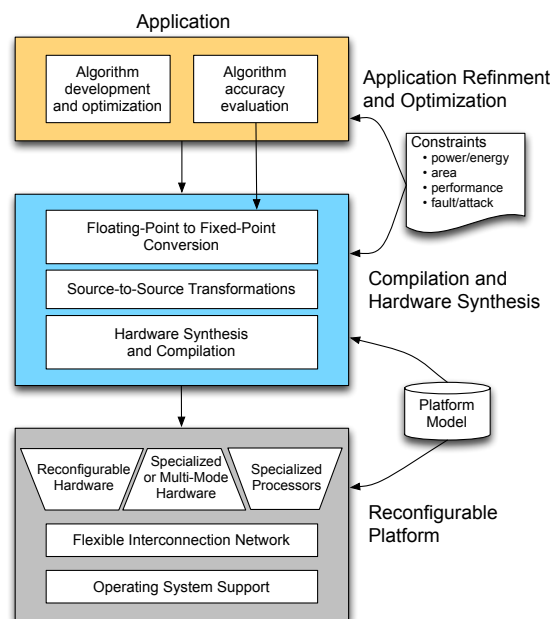
Figure 1. CAIRN*'s general design flow and related research themes*

Figure 1 shows the global design flow we propose to develop. This flow is organized in levels which refer to our three research themes: application optimization (new algorithms, fixed-point arithmetic, advanced representations of numbers), architecture optimization (reconfigurable and specialized hardware, application-specific processors, arithmetic operators and functions), and stepwise refinement and code generation (code transformations, hardware synthesis, compilation).

In the rest of this part, we briefly describe the challenges concerning **new reconfigurable platforms** in Section 3.2 and the issues on **compiler and synthesis tools** related to these platforms in Section 3.3.

3.2. Reconfigurable Architecture Design

Nowadays, FPGAs are not only suited for application specific algorithms, but also considered as fully-featured computing platforms, thanks to their ability to accelerate massively parallelizable algorithms much faster than their processor counterparts [75]. They also support to be dynamically reconfigured. At runtime, partially reconfigurable regions of the logic fabric can be reconfigured to implement a different task, which allows for a better resource usage and adaptation to the environment. Dynamically reconfigurable hardware can also cope with hardware errors by relocating some of its functionalities to another, sane, part of the logic fabric. It could also provide support for a multi-tasked computation flow where hardware tasks are loaded on-demand at runtime. Nevertheless, current design flows of FPGA vendors are still limited by the use of one partial bitstream for each reconfigurable region and for each design. These regions are defined at design time and it is not possible to use only one bitstream for multiple reconfigurable regions nor multiple chips. The multiplicity of such bitstreams leads to a significant increase in memory. Recent research has been conducted in the domain of task relocation on a reconfigurable fabric. All of the related work was conducted on architectures from commercial vendors (e.g., Xilinx, Altera) which share the same limitations: the inner details of the bitstream are not publicly known, which limits applicability of the techniques. To circumvent this issue, most dynamic reconfiguration techniques are either generating multiple bitstreams for each location [59] or implementing an online filter to relocate the tasks [69]. Both of these techniques still suffer from memory footprint and from the online complexity of task relocation.

Increasing the level and grain of reconfiguration is a solution to counterbalance the FPGA penalties. Coarse-grained reconfigurable architectures (CGRA) provide operator-level configurable functional blocks and word-level datapaths [76], [64], [74]. Compared to FPGA, they benefit from a massive reduction in configuration memory and configuration delay, as well as for routing and placement complexity. This in turns results in an improvement in the computation volume over energy cost ratio, although with a loss of flexibility compared to bit-level operations. Such constraints have been taken into account in the design of DART[7], Adres [72] or polymorphous computing fabrics[9]. These works have led to commercial products such as the PACT/XPP [58] or Montium from Recore systems, without however a real commercial success yet. Emerging platforms like Xilinx/Zynq or Intel/Altera are about to change the game.

In the context of emerging heterogenous multicore architecture, CAIRN advocates for associating general-purpose processors (GPP), flexible network-on-chip and coarse-grain or fine-grain dynamically reconfigurable accelerators. We leverage our skills on microarchitecture, reconfigurable computing, arithmetic, and low-power design, to discover and design such architectures with a focus on: -reduced energy per operation, -improved application performance through acceleration, - hardware flexibility and self-adaptive behavior, -tolerance to faults, computing errors, and process variation, - protections against side channel attacks, - limited silicon area overhead.

3.3. Compilation and Synthesis for Reconfigurable Platforms

In spite of their advantages, reconfigurable architectures, and more generally hardware accelerators, lack efficient and standardized compilation and design tools. As of today, this still makes the technology impractical for large-scale industrial use. Generating and optimizing the mapping from high-level specifications to reconfigurable hardware platforms are therefore key research issues, which have received considerable interest over the last years [62], [77], [73], [71], [70]. In the meantime, the complexity (and heterogeneity) of these platforms has also been increasing quite significantly, with complex heterogeneous multi-cores architectures becoming a *de facto* standard. As a consequence, the focus of designers is now geared toward optimizing overall system-level performance and efficiency [68]. Here again, existing tools are not well suited, as they fail at providing an unified programming view of the programmable and/or reconfigurable components implemented on the platform.

In this context, we have been pursuing our efforts to propose tools whose design principles are based on a tight coupling between the compiler and the target hardware architectures. We build on the expertise of the team members in High Level Synthesis (HLS) [4], ASIP optimizing compilers [10] and automatic parallelization for massively parallel specialized circuits [2]. We first study how to increase the efficiency of standard programmable processors by extending their instruction set to speed-up compute intensive kernels. Our focus is on efficient and exact algorithms for the identification, selection and scheduling of such instructions [5]. We address compilation challenges by borrowing techniques from high-level synthesis, optimizing compilers and automatic parallelization, especially when dealing with nested loop kernels. In addition, and independently of the scientific challenges mentioned above, proposing such flows also poses significant software engineering issues. As a consequence, we also study how leading edge software engineering techniques (Model Driven Engineering) can help the Computer Aided Design (CAD) and optimizing compiler communities prototyping new research ideas [3].

Efficient implementation of multimedia and signal processing applications (in software for DSP cores or as special-purpose hardware) often requires, for reasons related to cost, power consumption or silicon area constraints, the use of fixed-point arithmetic, whereas the algorithms are usually specified in floating-point arithmetic. Unfortunately, fixed-point conversion is very challenging and time-consuming, typically demanding up to 50% of the total design or implementation time. Thus, tools are required to automate this conversion. For hardware or software implementation, the aim is to optimize the fixed-point specification. The implementation cost is minimized under a numerical accuracy or an application performance constraint. For DSP-software implementation, methodologies have been proposed [6] to achieve fixed-point conversion. For hardware implementation, the best results are obtained when the word-length optimization process is coupled with the high-level synthesis [65]. Evaluating the effects of finite precision is one of the major and often the most time consuming step while performing fixed-point refinement. Indeed, in the word-length optimization process, the numerical accuracy is evaluated as soon as a new word-length is tested, thus, several times per iteration of the optimization process. Classical approaches are based on fixed-point simulations [66]. Leading to long evaluation times, they can hardly be used to explore the design space. Therefore, our aim is to propose closed-form expressions of errors due to fixed-point approximations that are used by a fast analytical framework for accuracy evaluation [8].

3.4. Software Frameworks Developed by the Team

With the ever raising complexity of embedded applications and platforms, the need for efficient and customizable compilation flows is stronger than ever. This need of flexibility is even stronger when it comes to research compiler infrastructures that are necessary to gather quantitative evidence of the performance/energy or cost benefits obtained through the use of reconfigurable platforms. From a compiler point of view, the challenges exposed by these complex reconfigurable platforms are quite significant, since they require the compiler to extract and to expose an important amount of coarse and/or fine grain parallelism, to take complex resource constraints into consideration while providing efficient memory hierarchy and power management.

Because they are geared toward industrial use, production compiler infrastructures do not offer the level of flexibility and productivity that is required for compiler and CAD tool prototyping. To address this issue, we designed an extensible source-to-source compiler infrastructure that takes advantage of leading edge model-driven object-oriented software engineering principles and technologies.

Figure 2 shows the global framework that is being developed in the group. Our compiler flow mixes several types of intermediate representations. The baseline representation is a simple tree-based model enriched with control flow information. This model is mainly used to support our source-to-source flow, and serves as the backbone for the infrastructure. We use the extensibility of the framework to provide more advanced representations along with their corresponding optimizations and code generation plug-ins. For example, for our pattern selection and accuracy estimation tools, we use a data dependence graph model in all basic blocks instead of the tree model. Similarly, to enable polyhedral based program transformations and analysis, we introduced a specific representation for affine control loops that we use to derive a Polyhedral Reduced Dependence Graph (PRDG). Our current flow assumes that the application is specified as a hierarchy of
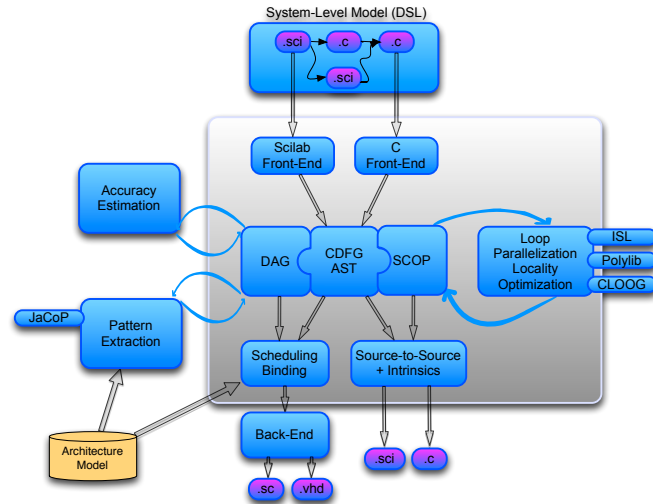
Figure 2. CAIRN*'s general software development framework.*

communicating tasks, where each task is expressed using C or Matlab/Scilab, and where the system-level representation and the target platform model are often defined using Domain Specific Languages (DSL).

Gecos (Generic Compiler Suite) is the main backbone of CAIRN's flow. It is an open source Eclipse-based flexible compiler infrastructure developed for fast prototyping of complex compiler passes. Gecos is a 100% Java based implementation and is based on modern software engineering practices such as Eclipse plugin or model-driven software engineering with EMF (Eclipse Modeling Framework). As of today, our flow offers the following features:

- An automatic floating-point to fixed-point conversion flow (for ASIC/FPGA and embedded processors). **ID.Fix** is an infrastructure for the automatic transformation of software code aiming at the conversion of floating-point data types into a fixed-point representation.

- A polyhedral-based loop transformation and parallelization engine (mostly targeted at HLS).

- A custom instruction extraction flow (for ASIP and dynamically reconfigurable architectures). **Durase** is developed for the compilation and the synthesis targeting reconfigurable platforms and the automatic synthesis of application specific processor extensions. It uses advanced technologies, such as graph matching together with constraint programming methods.

- Several back-ends to enable the generation of VHDL for specialized or reconfigurable IPs, and SystemC for simulation purposes (e.g., fixed-point simulations).

Gecos, ID.Fix or Durase have been demonstrated during "University Booths" in various conference such as IEEE/ACM DAC or DATE.

4. Application Domains

4.1. Panorama

keywords: Wireless (Body) Sensor Networks, High-Rate Optical Communications, Wireless Communications, Applied Cryptography.

Our research is based on realistic applications, in order to both discover the main needs created by these applications and to invent realistic and interesting solutions.

Wireless Communication is our privileged application domain. Our research includes the prototyping of (subsets of) such applications on reconfigurable and programmable platforms. For this application domain, the high computational complexity of the 5G Wireless Communication Systems calls for the design of high-performance and energy-efficient architectures. In **Wireless Sensor Networks** (WSN), where each wireless node is expected to operate without battery replacement for significant periods of time, energy consumption is the most important constraint. Sensor networks are a very dynamic domain of research due, on the one hand, to the opportunity to develop innovative applications that are linked to a specific environment, and on the other hand to the challenge of designing totally autonomous communicating objects.

Other important fields are also considered: hardware cryptographic and security modules, high-rate optical communications, machine learning, and multimedia processing.

5. Highlights of the Year

5.1. Highlights of the Year

Members of CAIRN published six papers accepted at IEEE/ACM Design Automation and Test in Europe for 2017, one of the major events in design automation.
[30] was among the few papers nominated for best paper at IEEE FPL.

6. New Software and Platforms

6.1. Gecos

Generic Compiler Suite
KEYWORDS: Source-to-source compiler - Model-driven software engineering - Retargetable compilation
SCIENTIFIC DESCRIPTION: The Gecos (Generic Compiler Suite) project is a source-to-source compiler infrastructure developed in the Cairn group since 2004. It was designed to enable fast prototyping of program analysis and transformation for hardware synthesis and retargetable compilation domains.

Gecos is Java based and takes advantage of modern model driven software engineering practices. It uses the Eclipse Modeling Framework (EMF) as an underlying infrastructure and takes benefits of its features to make it easily extensible. Gecos is open-source and is hosted on the Inria gforge.

The Gecos infrastructure is still under very active development, and serves as a backbone infrastructure to projects of the group. Part of the framework is jointly developed with Colorado State University and between 2012 and 2015 it was used in the context of the FP7 ALMA European project. The Gecos infrastructure is currently used by the EMMTRIX start-up, a spin-off from the ALMA project which aims at commercializing the results of the project, and in the context of the H2020 ARGO European project.

FUNCTIONAL DESCRIPTION: GeCoS provides a programme transformation toolbox facilitating parallelisation of applications for heterogeneous multiprocessor embedded platforms. In addition to targeting programmable processors, GeCoS can regenerate optimised code for High Level Synthesis tools.

- Participants: Tomofumi Yuki, Thomas Lefeuvre, Imèn Fassi, Mickael Dardaillon, Ali Hassan El Moussawi and Steven Derrien
- Partner: Université de Rennes 1
- Contact: Steven Derrien
- URL: http://gecos.gforge.inria.fr

6.2. ID-Fix

Infrastructure for the Design of Fixed-point systems

KEYWORDS: Energy efficiency - Dynamic range evaluation - Accuracy optimization - Fixed-point arithmetic - Analytic Evaluation - Embedded systems - Code optimisation

SCIENTIFIC DESCRIPTION: The different techniques proposed by the team for fixed-point conversion are implemented on the ID.Fix infrastructure. The application is described with a C code using floating-point data types and different pragmas, used to specify parameters (dynamic, input/output word-length, delay operations) for the fixed-point conversion. This tool determines and optimizes the fixed-point specification and then, generates a C code using fixed-point data types (ac_fixed) from Mentor Graphics. The infrastructure is made-up of two main modules corresponding to the fixed-point conversion (ID.Fix-Conv) and the accuracy evaluation (ID.Fix-Eval)

FUNCTIONAL DESCRIPTION: ID.Fix focuses on computational precision accuracy and can provide an optimised specification using fixed point arithmetic from a C source code with floating point data types. Fixed point arithmetic is very widely used in embedded systems as it provides better performance and is much more energy efficient. ID.Fix used an analytic programme model which means it can explore more solutions and thereby produce much more efficient code.

- Participant: Olivier Sentieys
- Partner: Université de Rennes 1
- Contact: Olivier Sentieys
- URL: http://idfix.gforge.inria.fr

6.3. Platforms

6.3.1. Zyggie

KEYWORDS: Health - Biomechanics - Wireless body sensor networks - Low power - Gesture recognition - Hardware platform - Software platform - Localization

SCIENTIFIC DESCRIPTION: Zyggie is a hardware and software wireless body sensor network platform. Each sensor node, attached to different parts of the human body, contains inertial sensors (IMU) (accelerometer, gyrometer, compass and barometer), an embedded processor and a low-power radio module to communicate data to a coordinator node connected to a computer, tablet or smartphone. One of the system's key innovations is that it collects data from sensors as well as on distances estimated from the power of the radio signal received to make the 3D location of the nodes more precise and thus prevent IMU sensor drift and power consumption overhead. Zyggie can be used to determine posture or gestures and mainly has applications in sport, healthcare and the multimedia industry.

FUNCTIONAL DESCRIPTION: The Zyggie sensor platform was developed to create an autonomous Wireless Body Sensor Network (WBSN) with the capabilities of monitoring body movements. The Zyggie platform is part of the BoWI project funded by CominLabs. Zyggie is composed of a processor, a radio transceiver and different sensors including an Inertial Measurement Unit (IMU) with 3-axis accelerometer, gyrometer, and magnetometer. Zyggie is used for evaluating data fusion algorithms, low power computing algorithms, wireless protocols, and body channel characterization in the BoWI project.

The Zyggie V2 prototype includes the following features: a 32-bit microcontroller to manage a custom MAC layer and processe quaternions based on IMU measures, and an UWB radio from DecaWave to measure distances between nodes with Time of Flight (ToF).

- Participants: Arnaud Carer and Olivier Sentieys
- Partners: Lab-STICC - Université de Rennes 1
- Contact: Olivier Sentieys
- URL: http://www.bowi.cominlabs.ueb.eu/fr/zyggie-wbsn-platform

Figure 3. CAIRN*'s Ziggie platform for WBSN*

7. New Results

7.1. Reconfigurable Architecture Design

7.1.1. *Voltage Over-Scaling for Error-Resilient Applications*

Participants: Rengarajan Ragavan, Benjamin Barrois, Cédric Killian, Olivier Sentieys.

Voltage scaling has been used as a prominent technique to improve energy efficiency in digital systems, scaling down supply voltage effects in quadratic reduction in energy consumption of the system. Reducing supply voltage induces timing errors in the system that are corrected through additional error detection and correction circuits. In [43], we proposed voltage over-scaling based approximate operators for applications that can tolerate errors. We characterized the basic arithmetic operators using different operating triads (combination of supply voltage, body-biasing scheme and clock frequency) to generate models for approximate operators. Error-resilient applications can be mapped with the generated approximate operator models to achieve optimum trade-off between energy efficiency and error margin. Based on the dynamic speculation technique, best possible operating triad is chosen at runtime based on the user definable error tolerance margin of the application. In our experiments in 28nm FDSOI, we achieved maximum energy efficiency of 89% for basic operators like 8-bit and 16-bit adders at the cost of 20% Bit Error Rate (ratio of faulty bits over total bits) by operating them in near-threshold regime.

7.1.2. *Stochastic Computation Elements with Correlated Input Streams*

Participants: Rengarajan Ragavan, Rahul Kumar Budhwani, Olivier Sentieys.

In recent years, shrinking size in integrated circuits has imposed a big challenge in maintaining the reliability in conventional computing. Stochastic Computing (SC) has been seen as a reliable, low-cost, and low-power alternative to overcome such issues. SC computes data in the form of bit streams of 1s and 0s. Therefore, SC outperforms conventional computing in terms of tolerance to soft error and uncertainty at the cost of increased computational time. Stochastic Computing with uncorrelated input streams requires streams to be highly independent for better accuracy. This results in more hardware consumption for conversion of binary numbers to stochastic streams. Correlation can be used to design Stochastic Computation Elements (SCE) with correlated input streams. These designs have higher accuracy and less hardware consumption. In [38], we proposed new SC designs to implement image processing algorithms with correlated input streams. Experimental results of proposed SC with correlated input streams show on average 37% improvement in accuracy with reduction of 50-90% in area and 20-85% in delay over existing stochastic designs.

7.1.3. *Fault Tolerant Architectures*

Participants: Olivier Sentieys, Angeliki Kritikakou, Rafail Psiakis.

Error occurrence in embedded systems has significantly increased, whereas critical applications require reliable processors that combine performance with low cost and energy consumption. Very Long Instruction Word (VLIW) processors have inherent resource redundancy which is not constantly used due to application's fluctuating Instruction Level Parallelism (ILP). Approaches can benefit these additional resources to provide fault tolerance.

The reliability through idle slots utilization can be explored either at compile-time, increasing code size and storage requirements, or at run-time only inside the current instruction bundle, adding unnecessary time slots and degrading performance. To address this issue, we proposed a technique in [41] to explore the idle slots inside and across original and replicated instruction bundles reclaiming more efficiently the idle slots and creating a compact schedule. To achieve this, a dependency analysis is applied at run-time. The execution of both original and replicated instructions is allowed at any adequate function unit, providing higher flexibility on instruction scheduling. The proposed technique achieves up to 26% reduction in performance degradation over existing approaches.

When permanent and soft errors coexist, spare units have to be used or the executed program has to be modified through self-repair or by using several stored versions. However, these solutions introduce high area overhead for the additional resources, time overhead for the execution of the repair algorithm and storage overhead of the multi-versioning. To address these limitations, a hardware mechanism is proposed in [42] which at run-time replicates the instructions and schedules them at the idle slots considering the resource constraints. If a resource becomes faulty, the proposed approach efficiently rebinds both the original and replicated instructions during execution. In this way, the area overhead is reduced, as no spare resources are used, whereas time and storage overhead are not required. Results show up to 49% performance gain over existing techniques.

7.1.4. *Hardware Accelerated Simulation of Heterogeneous Platforms*

Participants: Minh Thanh Cong, François Charot, Steven Derrien.

When considering designing heterogeneous multi-core platforms, the number of possible design combinations leads to a huge design space, with subtle trade-offs and design interactions. To reason about what design is best for a given target application requires detailed simulation of many different possible solutions. Simulation frameworks exist (such as gem5) and are commonly used to carry out these simulations. Unfortunately, these are purely software-based approaches and they do not allow a real exploration of the design space. Moreover, they do not really support highly heterogeneous multi-core architectures. These limitations motivate the study of the use of hardware to accelerate the simulation, and in particular of FPGA components. In this context, we are currently investigating the possibility of building hardware accelerated simulators using the HAsim simulation infrastructure, jointly developed by MIT and Intel. HAsim is an FPGA-accelerated simulator that is able to simulate a multicore with a high-detailed pipeline, cache hierarchy and detailed on-chip network on a single FPGA. A model of the RISC-V instruction set architecture suited to the HAsim infrastructure has been developed, its deployment on the Xeon+FPGA Intel platform is in progress. This work is done with the perspective of studying hardware accelerated simulation of heterogeneous multicore architectures mixing RISC-V cores and hardware accelerators.

7.1.5. *Optical Interconnections for 3D Multiprocessor Architectures*

Participants: Jiating Luo, Ashraf El-Antably, Van Dung Pham, Cédric Killian, Daniel Chillet, Olivier Sentieys.

To address the issue of interconnection bottleneck in multiprocessor on a single chip, we study how an Optical Network-on-Chip (ONoC) can leverage 3D technology by stacking a specific photonics die. The objectives of this study target: i) the definition of a generic architecture including both electrical and optical components, ii) the interface between electrical and optical domains, iii) the definition of strategies (communication protocol) to manage this communication medium, and iv) new techniques to manage and reduce the power consumption of optical communications. The first point is required to ensure that electrical and optical components can be used together to define a global architecture. Indeed, optical components are generally larger than electrical components, so a trade-off must be found between the size of optical and electrical parts. For the second

point, we study how the interface can be designed to take applications needs into account. From the different possible interface designs, we extract a high-level performance model of optical communications from losses induced by all optical components to efficiently manage Laser parameters. Then, the third point concerns the definition of high-level mechanisms which can handle the allocation of the communication medium for each data transfer between tasks. This part consists in defining the protocol of wavelength allocation. Indeed, the optical wavelengths are a shared resource between all the electrical computing clusters and are allocated at run time according to application needs and quality of service. The last point concerns the definition of techniques allowing to reduce the power consumption of on-chip optical communications. The power of each Laser can be dynamically tuned in the optical/electrical interface at run time for a given targeted bit-error-rate. Due to the relatively high power consumption of such integrated Laser, we study how to define adequate policies able to adapt the laser power to the signal losses.

In [37] we designed an Optical-Network-Interface (ONI) to connect a cluster of several processors to the optical communication medium. This interface, constrained by the 10 Gb/s data-rate of the Lasers, integrates Error Correcting Codes (ECC) and a communication manager. This manager can select, at run-time, the communication mode to use depending on timing or power constraints. Indeed, as the use of ECC is based on redundant bits, it increases the transmission time, but saves power for a given Bit Error Rate (BER). Moreover, our ONI allows for data to be sent using several wavelengths in parallel, hence increasing transmission bandwidth. From the design of this interface, estimation in terms of power consumption and execution time have been obtained, as well as the energy per bit of each communication.

The optical medium can support multiple transactions at the same time on different wavelengths by using Wavelength Division Multiplexing (WDM). Moreover, multiple wavelengths can be gathered as high-bandwidth channel to reduce transmission time. However, multiple signals sharing simultaneously a waveguide lead to inter-channel crosstalk noise. This problem impacts the Signal to Noise Ratio (SNR) of the optical signal, which increases the Bit Error Rate (BER) at the receiver side. In [39], we formulated the crosstalk noise and execution time models and then proposed a Wavelength Allocation (WA) method in a ring-based WDM ONoC to reach performance and energy trade-offs based on the application constraints. We showed that for a 16-core ONoC architecture using 12 wavelengths, more than 10^5 allocation solutions exist and only 51 are on a Pareto front giving a tradeoff between execution time and energy per bit (derived from the BER). These optimized solutions reduce the execution time by 37% or the energy from 7.6fJ/bit to 4.4fJ/bit.

We also proposed to explore the selection of laser power for each communication. This approach reduces the global power consumption by ensuring the targeted Bit Error Rate for each communication. To support laser power selection, we have also studied, designed and evaluated at transistor level different configurable laser drivers using a 28NM FDSOI technology.

7.1.6. *Adaptive Dynamic Compilation for Low-Power Embedded Systems*
Participants: Steven Derrien, Simon Rokicki.

Single ISA-Heterogeneous multi-cores such as the ARM big.LITTLE have proven to be an attractive solution to explore different energy/performance trade-offs. Such architectures combine Out of Order cores with smaller in-order ones to offer different power/energy profiles. They however do not really exploit the characteristics of workloads (compute-intensive vs. control dominated).

In this work, we propose to enrich these architectures VLIW cores, which are very efficient at compute-intensive kernels. To preserve the single ISA programming model, we resort to Dynamic Binary Translation as used in Transmeta Crusoe and NVidia Denver processors. Our proposed DBT framework targets the RISC-V ISA, for which both OoO and in-order implementations exist.

Since DBT operates at runtime, its execution time is directly perceptible by the user, hence severely constrained. As a matter of fact, this overhead has often been reported to have a huge impact on actual performance, and is considered as being the main weakness of DBT based solutions. This is particularly true when targeting a VLIW processor: the quality of the generated code depends on efficient scheduling; unfortunately

scheduling is known to be the most time-consuming component of a JIT compiler or DBT. Improving the responsiveness of such DBT systems is therefore a key research challenge. This is however made very difficult by the lack of open research tools or platform to experiment with such platforms.

To address these issues, we have developed an open hardware/software platform supporting DBT. The platform was designed using HLS tools and validated on a FPGA board. The DBT uses RISC-V as host ISA, and can be retargeted to different VLIW configurations. Our platform uses custom hardware accelerators to improve the reactivity of our optimizing DBT flow. Our results [44] show that, compared to a software implementation, our approach offers speed-up by $8\times$ while consuming $18\times$ less energy.

Our current research work investigates how DBT techniques can be used to support runtime configurable VLIW cores. Such cores enable fine grain exploration of energy/performance trade-off by dynamically adjusting their number of execution slots, their register file size, etc.). More precisely, we build on our DBT framework to enable dynamic code specialization. Our first experimental results suggest that this approach leads to best-case performance and energy efficiency when compared against static VLIW configurations [54].

7.1.7. *Design Space Exploration for Iterative Stencil computations on FPGA accelerators*
Participants: Steven Derrien, Gaël Deest, Tomofumi Yuki.

Iterative stencil computations arise in many application domains, ranging from medical imaging to numerical simulation. Since they are computationally demanding, a large body of work addressed the problem of parallelizing and optimizing stencils for multi-cores, GPUs, and FPGAs. Earlier attempts targeting FPGAs showed that the performance of such accelerators is the result of a complex interplay between the FPGA's raw computing power, the amount of on-chip memory it has, and the performance of the external memory system. They also illustrate how each application may have different requirements. For example, in the context of embedded vision, the designer's goal is often to find the design with minimum cost that matches real-time performance constraints (e.g., 4K@60fps). In an exascale context, the designer's goal is to maximize performance (measured in ops-per-second) for a given FPGA board, while maintaining power dissipation to a minimum. Based on these observations, we explore a family of design options that can accommodate a large set of requirements and constraints, by exposing trade-offs between computing power, bandwidth requirements, and FPGA resource usage. We have developed a code generator that produces HLS-optimized C/C++ descriptions of accelerator instances targeting emerging System on Chip platforms, (e.g., Xilinx Zynq or Intel SoC). Our family of designs builds upon the well-known tiling transformation, which we use to balance on-chip memory cost and off-chip bandwidth. To ease the exploration of this design space, we propose performance models to hone in on the most interesting design points, and show how they accurately lead to optimal designs. Our results demonstrate that the optimal choice depends on problem sizes and performance goals [30].

7.1.8. *Energy-driven Accelerator Exploration for Heterogeneous Multiprocessor Architectures*
Participants: Baptiste Roux, Olivier Sentieys.

Programming heterogeneous multiprocessor architectures combining multiple processor cores and hardware accelerators is a real challenge. Computer-aided design and development tools try to reduce the large design space by simplifying hardware software mapping mechanisms. However, energy consumption is not well supported in most of design space exploration methodologies due to the difficulty to fast and accurately estimate energy consumption. To this aim,we proposed and validated an exploration method for partitioning applications on software cores and hardware accelerators under energy-efficiency constraints. The methodology is based on energy and performance measurement of a tiny subset of the design space and an analytical formulation of the performance and energy of an application kernel mapped on a heterogeneous architecture. This closed-form expression is captured and solved using Mixed Integer Linear Programming, which allows for very fast exploration resulting in the optimal solution. The approach is validated on two applications kernels using Zynq-based architecture showing more than 12% acceleration speed-up and energy saving compared to standard approaches. Results also show that the most energy-efficient solution is application- and platform-dependent and moreover hardly predictable, which highlights the need for fast exploration.

7.2. Compilation and Synthesis for Reconfigurable Platform

7.2.1. *Superword-Level Parallelism-Aware Word Length Optimization*

Participants: Steven Derrien, Ali Hassan El Moussawi.

Many embedded processors do not support floating-point arithmetic in order to comply with strict cost and power consumption constraints. But, they generally provide support for SIMD as a mean to improve performance for little cost overhead. Achieving good performance when targeting such processors requires the use of fixed-point arithmetic and efficient exploitation of SIMD data-path. To reduce time-to-market, automatic SIMDization – such as superword level parallelism (SLP) extraction – and floating-point to fixed-point conversion methodologies have been proposed. In [33], we showed that applying these transformations independently is not efficient. We proposed an SLP-aware word length optimization algorithm to jointly perform floating-point to fixed-point conversion and SLP extraction. We implemented the proposed approach in a source-to-source compiler framework and evaluated it on several embedded processors. Experimental results illustrated the validity of our approach with performance improvement by up to 40% for a limited loss in accuracy.

7.2.2. *Automatic Parallelization Techniques for Time-Critical Systems*

Participants: Steven Derrien, Imen Fassi, Thomas Lefeuvre.

Real-time systems are ubiquitous, and many of them play an important role in our daily life. In hard real-time systems, computing the correct results is not the only requirement. In addition, the results must be produced within pre-determined timing constraints, typically deadlines. To obtain strong guarantees on the system temporal behavior, designers must compute upper bounds of the Worst-Case Execution Times (WCET) of the tasks composing the system. WCET analysis is confronted with two challenges: (i) extracting knowledge of the execution flow of an application from its machine code, and (ii) modeling the temporal behavior of the target platform. Multi-core platforms make the latter issue even more challenging, as interference caused by concurrent accesses to shared resources have also to be modeled. Accurate WCET analysis is facilitated by *predictable* hardware architectures. For example, platforms using ScratchPad Memories (SPMs) instead of caches are considered as more predictable. However SPM management is left to the programmer-managed, making them very difficult to use, especially when combined with complex loop transformations needed to enable task level parallelization. Many researches have studied how to combine automatic SPM management with loop parallelization at the compiler level.It has been shown that impressive average-case performance improvements could be obtained on compute intensive kernels, but their ability to reduce WCET estimates remains to be demonstrated, as the transformed code does not lends itself well to WCET analysis.

In the context of the ARGO project, and in collaboration with members of the PACAP team, we have studied how parallelizing compilers techniques should be revisited in order to help WCET analysis tools. More precisely, we have demonstrated the ability of polyhedral optimization techniques to reduce WCET estimates in the case of sequential codes, with a focus on locality improvement and array contraction. We have shown on representative real-time image processing use cases that they could bring significant improvements of WCET estimates (up to 40%) provided that the WCET analysis process is guided with automatically generated flow annotations [31].

7.2.3. *Operator-Level Approximate Computing*

Participants: Benjamin Barrois, Olivier Sentieys.

Many applications are error-resilient, allowing for the introduction of approximations in the calculations, as long as a certain accuracy target is met. Traditionally, fixed-point arithmetic is used to relax accuracy, by optimizing the bit-width. This arithmetic leads to important benefits in terms of delay, power and area. Lately, several hardware approximate operators were invented, seeking the same performance benefits. However, a fair comparison between the usage of this new class of operators and classical fixed-point arithmetic with careful truncation or rounding, has never been performed. In [27], we first compare approximate and fixed-point arithmetic operators in terms of power, area and delay, as well as in terms of induced error, using many

state-of-the-art metrics and by emphasizing the issue of data sizing. To perform this analysis, we developed a design exploration framework, *ApxPerf*, which guarantees that all operators are compared using the same operating conditions. Moreover, operators are compared in several classical real-life applications leveraging relevant metrics. In [27], we show that considering a large set of parameters, existing approximate adders and multipliers tend to be dominated by truncated or rounded fixed-point ones. For a given accuracy level and when considering the whole computation data-path, fixed-point operators are several orders of magnitude more accurate while spending less energy to execute the application. A conclusion of this study is that the entropy of careful sizing is always lower than approximate operators, since it require significantly less bits to be processed in the data-path and stored. Approximated data therefore always contain on average a greater amount of costly erroneous, useless information.

In [26] we performed a comparison between custom fixed-point (FxP) and floating-point (FlP) arithmetic, applied to bidimensional K-means clustering algorithm. First, FxP and FlP arithmetic operators are compared in terms of area, delay and energy, for different bitwidth, using the *ApxPerf2.0* framework. Finally, both are compared in the context of K-means clustering. The direct comparison shows the large difference between 8-to-16-bit FxP and FlP operators, FlP adders consuming 5-12$\times$ more energy than FxP adders, and multipliers 2-10$\times$ more. However, when applied to K-means clustering algorithm, the gap between FxP and FlP tightens. Indeed, the accuracy improvements brought by FlP make the computation more accurate and lead to an accuracy equivalent to FxP with less iterations of the algorithm, proportionally reducing the global energy spent. The 8-bit version of the algorithm becomes more profitable using FlP, which is 80% more accurate with only 1.6$\times$ more energy.

7.2.4. *Dynamic Fault-Tolerant Mapping and Scheduling on Multi-core systems*

Participants: Emmanuel Casseau, Petr Dobias.

Demand on multi-processor systems for high performance and low energy consumption still increases in order to satisfy our requirements to perform more and more complex computations. Moreover, the transistor size gets smaller and their operating voltage is lower, which goes hand in glove with higher susceptibility to system failure. In order to ensure system functionality, it is necessary to conceive fault-tolerant systems. One way to tackle this issue is to makes use of both the redundancy and reconfigurable computing, especially when multi-processor platforms are targeted. Actually, multi-processor platforms can be less vulnerable when one processor is faulty because other processors can take over its scheduled tasks.

In this context, we investigate how to dynamically map and schedule tasks onto homogeneous faulty processors. We developed a run-time algorithm based on the primary/backup approach which is commonly used for its minimal resources utilization and high reliability. Its principal rule is that, when a task arrives, the system creates two identical copies: the primary copy and the backup copy. Several policies have been studied and their performances have been analyzed. We are currently refining the algorithm to reduce its complexity without decreasing performance. This work is done in collaboration with Oliver Sinnen, PARC Lab., the University of Auckland.

7.2.5. *Energy Constrained and Real-Time Scheduling and Mapping on Multicores*

Participants: Olivier Sentieys, Angeliki Kritikakou, Lei Mo.

Multicore architectures are now widely used in energy-constrained real-time systems, such as energy-harvesting wireless sensor networks. To take advantage of these multicores, there is a strong need to balance system energy, performance and Quality-of-Service (QoS). The Imprecise Computation (IC) model splits a task into mandatory and optional parts allowing to tradeoff QoS. We focus on the problem of mapping, i.e. allocating and scheduling, IC-tasks to a set of processors to maximize system QoS under real-time and energy constraints, which we formulate as a Mixed Integer Linear Programming (MILP) problem. However, state-of-the-art solving techniques either demand high complexity or can only achieve feasible (suboptimal) solutions. We develop an effective decomposition-based approach in [40] to achieve an optimal solution while reducing computational complexity. It decomposes the original problem into two smaller easier-to-solve problems: a master problem for IC-tasks allocation and a slave problem for IC-tasks scheduling. We also

provide comprehensive optimality analysis for the proposed method. Through the simulations, we validate and demonstrate the performance of the proposed method, resulting in an average 55% QoS improvement with regards to published techniques.

7.2.6. *Real-Time Scheduling of Reconfigurable Battery-Powered Multi-Core Platforms*

Participants: Daniel Chillet, Aymen Gammoudi.

Reconfigurable real-time embedded systems are constantly increasingly used in applications like autonomous robots or sensor networks. Since they are powered by batteries, these systems have to be energy-aware, to adapt to their environment and to satisfy real-time constraints. For energy harvesting systems, regular recharges of battery can be estimated, and by including this parameter in the operating system, it is then possible to develop strategy able to ensure the best execution of the application until the next recharge. In this context, operating system services must control the execution of tasks to meet the application constraints. Our objective concerns the proposition of a new real-time scheduling strategy that considers execution constraints such as the deadline of tasks and the energy for heterogeneous architectures. For such systems, we first addressed homogeneous architectures and extended our work for heterogeneous systems for which each task has different execution parameters. For these two architectures models, we formulated the problem as an ILP optimisation problem that can be solved by classical solvers. Assuming that the energy consumed by the communication is dependent on the distance between processors, we proposed a mapping strategy to minimise the total cost of communication between processors by placing the dependent tasks as close as possible to each other. The proposed strategy guarantees that, when a task is mapped into the system and accepted, it is then correctly executed prior to the task deadline. Finally, as on-line scheduling is targeted for this work, we proposed heuristics to solve these problems in efficient way. These heuristics are based on the previous packing strategy developed for the mono-processor architecture case.

7.2.7. *Run-Time Management on Multicore Platforms*

Participant: Angeliki Kritikakou.

In real-time mixed-critical systems, Worst-Case Execution Time analysis (WCET) is required to guarantee that timing constraints are respected —at least for high criticality tasks. However, the WCET is pessimistic compared to the real execution time, especially for multicore platforms. As WCET computation considers the worst-case scenario, it means that whenever a high criticality task accesses a shared resource in multi-core platforms, it is considered that all cores use the same resource concurrently. This pessimism in WCET computation leads to a dramatic under utilization of the platform resources, or even failing to meet the timing constraints. In order to increase resource utilization while guaranteeing real-time guarantees for high criticality tasks, previous works proposed a run-time control system to monitor and decide when the interferences from low criticality tasks cannot be further tolerated. However, in the initial approaches, the points where the controller is executed were statically predefined. We propose a dynamic run-time control in [19] which adapts its observations to on-line temporal properties, increasing further the dynamism of the approach, and mitigating the unnecessary overhead implied by existing static approaches. Our dynamic adaptive approach allows to control the ongoing execution of tasks based on run-time information, and increases further the gains in terms of resource utilization compared with static approaches.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. *Labex CominLabs - 3DCORE (2014-2018)*

Participants: Olivier Sentieys, Daniel Chillet, Cédric Killian, Jiating Luo, Van Dung Pham, Ashraf El-Antably.

3DCORE (3D Many-Core Architectures based on Optical Network on Chip) is a project investigating new solutions based on silicon photonics to enhance by 2 to 3 magnitude orders energy efficiency and data rate of on-chip interconnect in the context of a many-core architecture. Moreover, 3DCore will take advantage of 3D technologies to design a specific optical layer suitable for a flexible and energy efficient high-speed optical network on chip (ONoC). 3DCORE involves CAIRN, FOTON (Rennes, Lannion) and Institut des Nanotechnologies de Lyon. For more details see http://www.3d-opt-many-cores.cominlabs.ueb.eu.

8.1.2. *Labex CominLabs - RELIASIC (2014-2018)*

Participants: Emmanuel Casseau, Imran Wali.

RELIASIC (Reliable Asic) will address the issue of fault-tolerant computation with a bottom-up approach, starting from an existing application as a use case (a GPS receiver) and adding some redundant mechanisms to allow the GPS receiver to be tolerant to transient errors due to low voltage supply. RELIASIC involves CAIRN, Lab-STICC (Lorient) and IETR (Rennes, Nantes). For more details see http://www.reliasic.cominlabs.ueb.eu In this project, CAIRN is in charge of the analysis and design of arithmetic operators for fault tolerance. We focus on the hardware implementations of conventional arithmetic operators such as adders, multipliers. We also propose a lightweight design and assessment framework for arithmetic operators with reduced-precision redundancy.

8.1.3. *Labex CominLabs & Lebesgue - H-A-H (2014-2017)*

Participants: Arnaud Tisserand, Gabriel Gallin, Audrey Lucas.

H-A-H for *Hardware and Arithmetic for Hyperelliptic Curves Cryptography* is a project on advanced arithmetic representation and algorithms for hyper-elliptic curve cryptography. It will provide novel implementations of HECC based cryptographic algorithms on custom hardware platforms. H-A-H involves CAIRN (Lannion) and IRMAR (Rennes). For more details see http://h-a-h.inria.fr/.

8.1.4. *Labex CominLabs - BBC (2016-2020)*

Participants: Olivier Sentieys, Cédric Killian, Joel Ortiz Sosa.

The aim of the BBC (on-chip wireless Broadcast-Based parallel Computing) project is to evaluate the use of wireless links between cores inside chips and to define new paradigms. Using wireless communications enables broadcast capabilities for Wireless Networks on Chip (WiNoC) and new management techniques for memory hierarchy and parallelism. The key objectives concern improvement of power consumption, estimation of achievable data rates, flexibility and reconfigurability, size reduction and memory hierarchy management. For more details see http://www.bbc.cominlabs.ueb.eu In this project, CAIRN will address new low-power MAC (media access control) technique based on CDMA access as well as broadcast-based fast cooperation protocol designed for resource sharing (bandwidth, distributed memory, cache coherency) and parallel programming.

8.1.5. *Labex CominLabs - SHERPAM (2014-2018)*

Participant: Patrice Quinton.

Heart failure and peripheral artery disease patients require early detection of health problems in order to prevent major risk of morbidity and mortality. Evidence shows that people recover from illness or cope with a chronic condition better if they are in a familiar environment (i.e., at home) and if they are physically active (i.e., practice sports). The goal of the Sherpam project is to design, implement, and validate experimentally a monitoring system allowing biophysical data of mobile subjects to be gathered and exploited in a continuous flow. Transmission technologies available to mobile users have been improved a lot during the last two decades, and such technologies offer interesting prospects for monitoring the health of people anytime and anywhere. The originality of the Sherpam project is to rely simultaneously and in an agile way on several kinds of wireless networks in order to ensure the transmission of biometric data, while coping with network disruptions. Sherpam also develops new signal processing algorithms for activity quantification and recognition which represent now a major social and public health issue (monitoring of elderly patient, personalized quantification activity, etc.). Sherpam involves research teams from several scientific domains and from several laboratories of Brittany (IRISA/CASA, LTSI, M2S, CIC-IT 1414-CHU Rennes and LAUREPS). For more details see http://www.sherpam.cominlabs.ueb.eu

8.1.6. DGA RAPID - FLODAM (2017–2021)

Participants: Olivier Sentieys, Angeliki Kritikakou.

FLODAM is an industrial research project for methodologies and tools dedicated to the hardening of embedded multi-core processor architectures. The goal is to: 1) evaluate the impact of the natural or artificial environments on the resistance of the system components to faults based on models that reflect the reality of the system environment , 2) the exploration of architecture solutions to make the multi-core architectures fault tolerant to transient or permanent faults and 3) test and evaluate the proposed fault tolerant architecture solutions and compare the results under different scenarios provided by the fault models.

8.2. European Initiatives

8.2.1. H2020 ARGO

Participants: Steven Derrien, Olivier Sentieys, Imen Fassi, Ali Hassan El Moussawi.

Program: H2020-ICT-04-2015

Project acronym: ARGO

Project title: WCET-Aware Parallelization of Model-Based Applications for Heterogeneous Parallel Systems

Duration: Feb. 2016 - Feb. 2019

Coordinator: KIT

Other partners: KIT (DE), UR1/Inria/CAIRN (FR), Recore Systems (NL), TEI-WG (GR), Scilab Ent. (FR), Absint (DE), DLR (DE), Fraunhofer (DE)

Increasing performance and reducing cost, while maintaining safety levels and programmability are the key demands for embedded and cyber-physical systems, e.g. aerospace, automation, and automotive. For many applications, the necessary performance with low energy consumption can only be provided by customized computing platforms based on heterogeneous many-core architectures. However, their parallel programming with time-critical embedded applications suffers from a complex toolchain and programming process. ARGO will address this challenge with a holistic approach for programming heterogeneous multi- and many-core architectures using automatic parallelization of model-based real-time applications. ARGO will enhance WCET-aware automatic parallelization by a cross-layer programming approach combining automatic tool-based and user-guided parallelization to reduce the need for expertise in programming parallel heterogeneous architectures. The ARGO approach will be assessed and demonstrated by prototyping comprehensive time-critical applications from both aerospace and industrial automation domains on customized heterogeneous many-core platforms.

8.2.2. ANR International ARTEFaCT

Participants: Olivier Sentieys, Benjamin Barrois, Tara Petric, Tomofumi Yuki.

Program: ANR International France-Switzerland

Project acronym: ARTEFaCT

Project title: AppRoximaTivE Flexible Circuits and Computing for IoT

Duration: Feb. 2016 - Dec. 2019

Coordinator: CEA

Other partners: CEA-LETI (FR), CAIRN (FR), EPFL (SW)

The ARTEFaCT project aims to build on the preliminary results on inexact and exact near-threshold and sub-threshold circuit design to achieve major energy consumption reductions by enabling adaptive accuracy control of applications. ARTEFaCT proposes to address, in a consistent fashion, the entire design stack, from physical hardware design, up to software application analysis, compiler optimizations, and dynamic energy management. We do believe that combining sub-near-threshold with inexact circuits on the hardware side and, in addition, extending this with intelligent and adaptive power management on the software side will produce outstanding results in terms of energy reduction, i.e., at least one order of magnitude, in IoT applications. The project will contribute along three research directions: (1) approximate, ultra low-power circuit design, (2) modeling and analysis of variable levels of computation precision in applications, and (3) accuracy-energy trade- offs in software.

8.3. International Initiatives

8.3.1. Inria Associate Teams

8.3.1.1. IoTA

Title: Ultra-Low Power Computing Platform for IoT leveraging Controlled Approximation

International Partner (Institution - Laboratory - Researcher):

Ecole Polytechnique Fédérale de Lausanne (Switzerland) - Christian Enz

Start year: 2017

See also: https://team.inria.fr/cairn/IOTA

Energy issues are central to the evolution of the Internet of Things (IoT), and more generally to the ICT industry. Current low-power design techniques cannot support the estimated growth in number of IoT objects and at the same time keep the energy consumption within sustainable bounds, both on the IoT node side and on cloud/edge-cloud side. This project aims to build on the preliminary results on inexact and exact sub/near-threshold circuit design to achieve major energy consumption reductions by enabling adaptive accuracy control of applications. Advanced ultra low-power hardware design methods utilize very low supply voltage, such as in near-threshold and sub-threshold designs. These emerging technologies are very promising avenues to decrease active and stand-by-power in electronic devices. To move another step forward, recently, approximate computing has become a major field of research in the past few years. IoTA proposes to address, in a consistent fashion, the entire design stack, from hardware design, up to software application analysis, compiler optimizations, and dynamic energy management. We do believe that combining sub-near-threshold with inexact circuits on the hardware side and, in addition, extending this with intelligent and adaptive power management on the software side will produce outstanding results in terms of energy reduction, i.e., at least one order of magnitude, in IoT. The main scientific challenge is twofold: (1) to add adaptive accuracy to hardware blocks built in near/sub threshold technology and (2) to provide the tools and methods to program and make efficient use of these hardware blocks for applications in the IoT domain. This entails developing approximate computing units, on one side, and methods and tools, on the other side, to rigorously explore trade-offs between accuracy and energy consumption in IoT systems. The expertise of the members of the two teams is complementary and covers all required technical knowledge necessary to reach our objectives, i.e., ultra low power hardware design (EPFL), approximate operators and functions (Inria, EPFL), formal analysis of precision in algorithms (Inria), and static and dynamic energy management (Inria, EPFL). Finally, the proof of concept will consist of results on (1) an adaptive, inexact or exact, ultra-low power microprocessor in 28 nm process and (2) a real prototype implemented in an FPGA platform combining processors and hardware accelerators. Several software use-cases relevant for the IoT domain will be considered, e.g., embedded vision, IoT sensors data fusion, to practically demonstrate the benefits of our approach.

8.3.2. Inria International Partners

8.3.2.1. LRS

Title: Loop unRolling Stones: compiling in the polyhedral model

International Partner (Institution - Laboratory - Researcher):

Colorado State University (United States) - Department of Computer Science - Prof. Sanjay Rajopadhye

8.3.2.2. HARAMCOP

Title: Hardware accelerators modeling using constraint-based programming

International Partner (Institution - Laboratory - Researcher):

Lund University (Sweden) - Department of Computer Science - Prof. Krzysztof Kuchcinski

8.3.2.3. SPINACH

Title: Secure and low-Power sensor Networks Circuits for Healthcare embedded applications

International Partner (Institution - Laboratory - Researcher):

University College Cork (Ireland) - Department of Electrical and Electronic Engineering - Prof. Liam Marnane and Prof. Emanuel Popovici

Arithmetic operators for cryptography, side channel attacks for security evaluation, energy-harvesting sensor networks, and sensor networks for health monitoring.

8.3.2.4. DARE

Title: Design space exploration Approaches for Reliable Embedded systems

International Partner (Institution - Laboratory - Researcher):

IMEC (Belgium) - Francky Catthoor

Methodologies to design low cost and efficient techniques for safety-critical embedded systems, Design Space Exploration (DSE), run-time dynamic control mechanisms.

8.3.2.5. Informal International Partners

LSSI laboratory, Québec University in Trois-Rivières (Canada), Design of architectures for digital filters and mobile communications.

Department of Electrical and Computer Engineering, University of Patras (Greece), Wireless Sensor Networks, Worst-Case Execution Time, Priority Scheduling.

Karlsruhe Institute of Technology - KIT (Germany), Loop parallelization and compilation techniques for embedded multicores.

Ruhr - University of Bochum - RUB (Germany), Reconfigurable architectures.

University of Science and Technology of Hanoi (Vietnam), Participation of several CAIRN's members in the Master ICT / Embedded Systems.

8.4. International Research Visitors

8.4.1. *Visits of International Scientists*

Mattia Cacciotti, Ecole Polytechnique Fédérale de Lausanne (Switzerland), from May 2017 until June 2017.

Emna Hammami, University of Tunis, from April 2017 until June 2017.

Prof. Stanislaw Piestrak, Univ de Lorraine, June 2017.

8.4.2. *Visits to International Teams*

P. Quinton was invited in Passau University (Passau, Germany) by Prof. Chris Lengauer during one week in June 2017, and gave an invited seminar on the synthesis of parallel architectures.

P. Quinton was invited by Prof. Daniel Massicotte of Université de Trois-Rivières (Québec) in October 2017 to cooperate on the design of FPGA hardware accelerators for electric simulation. His stay was supported by a grant of the RESMIQ (regroupement stratégique en microsystèmes du Québec). He gave an invited seminar on the synthesis of data-flow parallel systems.

8.4.3. Sabbatical programme

Casseau Emmanuel

Date: Aug 2016 - Jul 2017

Institution: University of Auckland (New Zealand), Parallel and Reconfigurable Research Lab. of the Electrical and Computer Engineering department.

The goal of the project was to propose dynamic mapping and scheduling algorithms dedicated to unreliable heterogeneous platforms, enabling self-adaptive and resource-aware computing.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Chair of Conference Program Committees

- O. Sentieys was Track Chair at IEEE NEWCAS.

9.1.2. Member of the Conference Program Committees

- D. Chillet was member of the technical program committee of HiPEAC RAPIDO, HiPEAC WRC, MCSoC, DCIS, ComPAS, DASIP, LP-EMS, ARC.
- S. Derrien was a member of technical program committee of IEEE FPL and ARC conferences and of WRC and Impact workshops.
- O. Sentieys was a member of technical program committee of IEEE/ACM DATE, IEEE FPL, ACM ENSSys, ACM SBCCI, IEEE ReConFig, FPGA4GPC.

9.1.3. Member of the Editorial Boards of Journals

- D. Chillet is member of the Editor Board of Journal of Real-Time Image Processing (JRTIP).
- O. Sentieys is member of the editorial board of Journal of Low Power Electronics and International Journal of Distributed Sensor Networks.

9.1.4. Invited Talks

- O. Sentieys gave an invited talk at FETCH (École d'hiver Francophone sur les Technologies de Conception des Systèmes embarqués Hétérogènes), Mont Tremblant, Canada, January 2017 on "Need more Energy Efficiency? Agree to Compute Inexactly".
- O. Sentieys gave an invited talk at GDR SoC^2, Paris, France, November 2017 on "Controlling Inexact Computations at Compile Time and Runtime".
- O. Sentieys gave an invited talk at IoT2Sustain Workshop, London, UK, July 2017 on "Challenges in Energy Efficiency of Computing Architectures: from Sensors to Clouds".
- O. Sentieys gave an invited course at ARCHI Spring School, Nancy, France, March 2017 on "Design of VLSI Integrated Circuits – A (very) deep dive into processors".

9.1.5. Leadership within the Scientific Community

- D. Chillet is member of the Board of Directors of Gretsi Association.
- D. Chillet is co-animator of the topics "Connected Objects" and "Near Sensor Computing" of GDR SoC^2.

- F. Charot and O. Sentieys are members of the steering committee of a CNRS Spring School for graduate students on embedded systems architectures and associated design tools (ARCHI).
- C. Killian was Co-Organizer of the Thematic Day on "Emerging Interconnect Technologies in Many Core Architectures" of GDR SoC^2, November 27, 2017.
- O. Sentieys is a member of the steering committee of a CNRS spring school for graduate students on low-power design (ECOFAC).
- O. Sentieys is a member of the steering committee of GDR SoC^2.

9.1.6. Scientific Expertise

- E. Casseau served as an expert for the Natural Sciences and Engineering Research Council of Canada (NSERC), program Discovery Grant 2017.
- O. Sentieys served as a jury member in the EDAA Outstanding Dissertations Award (ODA).

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

E. Casseau: signal processing, 16h, ENSSAT (L3)

E. Casseau: low power design, 6h, ENSSAT (M1)

E. Casseau: real time design methodology, 24h, ENSSAT (M1)

E. Casseau: computer architecture, 24h, ENSSAT (M1)

E. Casseau: SoC and high-level synthesis, 24h, Master by Research (SISEA) and ENSSAT (M2)

S. Derrien: component and system synthesis, 20h, Master by Research (ISTIC) (M2)

S. Derrien: computer architecture, 12h, ENS Rennes (L3)

S. Derrien: computer architecture, 24h, ISTIC (L3)

S. Derrien: introduction to operating systems, 8h, ISTIC (M1)

S. Derrien: embedded architectures, 48h, ISTIC (M1)

S. Derrien: high-level synthesis, 6h, ISTIC (M1)

S. Derrien: software engineering project, 40h, ISTIC (M1)

F. Charot: processor architecture, 25h, Univ. of Science and Tech. of Hanoi (M1)

D. Chillet: embedded processor architecture, 20h, ENSSAT (M1)

D. Chillet: multimedia processor architectures, 24h, ENSSAT (M2)

D. Chillet: low-power digital CMOS circuits, 6h, Telecom Bretagne (M2)

C. Killian: digital electronics, 62h, IUT Lannion (L1)

C. Killian: signal processing, 36h, IUT Lannion (L2)

C. Killian: automated measurements, 56h, IUT Lannion (L2)

C. Killian: measurement chain, 58h, IUT Lannion (L2)

C. Killian: embedded systems programming, 12h, IUT Lannion (L2)

C. Killian: automatic control, 18h, IUT Lannion (L2)

A. Kritikakou: computer architecture 1, 32h, ISTIC (L3)

A. Kritikakou: computer architecture 2, 44h, ISTIC (L3)

A. Kritikakou: C and unix programming languages, 102h, ISTIC (L3)

A. Kritikakou: operating systems, 96h, ISTIC (L3)

A. Kritikakou: multitasking operating systems, 20h, ISTIC (M1)

O. Sentieys: digital signal processing, 40h, ENSSAT (M1)

O. Sentieys: VLSI integrated circuit design, 40h, Enssat (M1)

C. Wolinski: computer architectures, 92h, Esir (L3)

C. Wolinski: design of embedded systems, 48h, Esir (M1)

C. Wolinski: signal, image, architecture, 26h, Esir (M1)

C. Wolinski: programmable architectures, 10h, Esir (M1)

C. Wolinski: component and system synthesis, 10h, Master by Research (istic) (M2)

9.2.2. Teaching Responsibilities

- C. Wolinski is the Director of Esir.
- S. Derrien was the responsible of the first year (M1) of the Master of Computer Science at ISTIC until Aug. 2017.
- O. Sentieys is responsible of the "Embedded Systems" major of the SISEA Master by Research.
- D. Chillet is the responsible of the ICT Master of University of Science and Technology of Hanoi.
- C. Killian is the responsible of the second year of the Physical Measurement DUT at IUT of Lannion.

Enssat stands for *"École Nationale Supérieure des Sciences Appliquées et de Technologie"* and is an *"École d'Ingénieurs"* of the University of Rennes 1, located in Lannion.
istic is the Electrical Engineering and Computer Science Department of the University of Rennes 1.
Esir stands for *"École supérieure d'ingénieur de Rennes"* and is an *"École d'Ingénieurs"* of the University of Rennes 1, located in Rennes.

9.2.3. Supervision

PhD: Benjamin Barrois, Methods to Evaluate Accuracy-Energy Trade-Off in Operator-Level Approximate Computing, Dec. 2017, O. Sentieys.

PhD: Gaël Deest, Implementation Trade-Offs for FPGA Accelerators, Dec. 2017, S. Derrien.

PhD: Xuan Chien Le, Improving performance of non-intrusive load monitoring with low-cost sensor networks, Apr. 2017, O. Sentieys, B. Vrigneau.

PhD: Rengarajan Ragavan, Error handling and energy estimation for error resilient near-threshold computing, Sep. 2017, O. Sentieys, C. Killian.

PhD: Baptiste Roux, Methodology and Tools for Energy-aware Task Mapping on Heterogeneous Multiprocessor Architectures, Nov. 2017, O. Sentieys, M. Gautier.

PhD in progress: Minh Thanh Cong, Hardware Accelerated Simulation of Heterogeneous Multicore Platforms, May 2017, F. Charot, S. Derrien.

PhD in progress: Petr Dobias, Towards efficient application execution on resilient multi-core architectures, Oct. 2017, E. Casseau.

PhD in progress: Gabriel Gallin, Hardware Arithmetic Units and Crypto-Processor for Hyperelliptic Curves Cryptography, Oct. 2014, A. Tisserand.

PhD in progress: Aymen Gammoudi, New Visual Adaptive Real-Time OS for Embedded Multi-Core Architecture, Oct. 2015, D. Chillet, M.Khalgui.

PhD in progress: Mael Gueguen, Improving the performance and energy efficiency of complex heterogeneous manycore architectures with on-chip data mining, Nov. 2016, O. Sentieys, A. Termier.

PhD in progress: Van-Phu Ha, Application-Level Tuning of Accuracy, Nov. 2017, T. Yuki, O. Sentieys.

PhD in progress: Audrey Lucas, Software support resistant to passive and active attacks for asymmetric cryptography on (very) small computation cores, Jan. 2016, A. Tisserand.

PhD in progress: Jiating, Luo, Communication protocol exploration in the context of 3D integration of multiprocessors interconnected by Optical Network-on-Chip with energy constraints, Nov. 2014, D. Chillet, C. Killian, S. Le-Beux.

PhD in progress: Thibaut Marty, Compiler support for speculative custom hardware accelerators, Sep. 2017, T. Yuki, O. Sentieys.

PhD in progress: Genevieve Ndour, Approximate Computing with High Energy Efficiency for Internet of Things Applications, Apr. 2016, A. Tisserand, A. Molnos (CEA LETI).

PhD in progress: Joel Ortiz Sosa, Study and design of a digital baseband transceiver for wireless network-on-chip architectures, Nov. 2016, O. Sentieys, C. Roland (Lab-STICC).

PhD in progress: Van Dung Pham, Design space exploration in the context of 3D integration of multiprocessors interconnected by Optical Network-on-Chip, Dec 2014, O. Sentieys, D. Chillet, C. Killian, S. Le-Beux.

PhD in progress: Rafail Psiakis, A Self-Healing Reconfigurable Accelerator Structure for Fault-Tolerant Multi-Cores, Oct. 2015, A. Kritikakou, O. Sentieys.

PhD in progress: Simon Rokicki, Hybrid Hardware/Software Dynamic Compilation for Adaptive Embedded Systems, Oct. 2015, S. Derrien.

PhD in progress: Nicolas Roux, Sensor-aided Non-Intrusive Appliance Load Monitoring: Detecting Activity of Devices through Low-Cost Wireless Sensors, Oct. 2016, O. Sentieys, B. Vrigneau.

PhD in progress: Mai-Thanh Tran, Hardware Synthesis of Flexible and Reconfigurable Radio from High-Level Language Dedicated to Physical Layer of Wireless Systems, Oct. 2013, E. Casseau, M. Gautier.

10. Bibliography

Major publications by the team in recent years

[1] R. DAVID, S. PILLEMENT, O. SENTIEYS. *Energy-Efficient Reconfigurable Processsors*, in "Low Power Electronics Design", C. PIGUET (editor), Computer Engineering, Vol 1, CRC Press, August 2004, chap. 20

[2] S. DERRIEN, S. RAJOPADHYE, P. QUINTON, T. RISSET. *12*, in "High-Level Synthesis From Algorithm to Digital Circuit", P. COUSSY, A. MORAWIEC (editors), Springer Netherlands, 2008, pp. 215-230, http://dx.doi.org/10.1007/978-1-4020-8588-8

[3] J.-M. JÉZÉQUEL, B. COMBEMALE, S. DERRIEN, C. GUY, S. RAJOPADHYE. *Bridging the Chasm Between MDE and the World of Compilation*, in "Journal of Software and Systems Modeling (SoSyM)", October 2012, vol. 11, n° 4, pp. 581-597 [*DOI :* 10.1007/S10270-012-0266-8], https://hal.inria.fr/hal-00717219

[4] B. LE GAL, E. CASSEAU, S. HUET. *Dynamic Memory Access Management for High-Performance DSP Applications Using High-Level Synthesis*, in "IEEE Transactions on VLSI Systems", 2008, vol. 16, n° 11, pp. 1454-1464

[5] K. MARTIN, C. WOLINSKI, K. KUCHCINSKI, A. FLOCH, F. CHAROT. *Constraint Programming Approach to Reconfigurable Processor Extension Generation and Application Compilation*, in "ACM transactions on Reconfigurable Technology and Systems (TRETS)", June 2012, vol. 5, n° 2, pp. 1-38, http://doi.acm.org/10.1145/2209285.2209289

[6] D. MENARD, D. CHILLET, F. CHAROT, O. SENTIEYS. *Automatic Floating-point to Fixed-point Conversion for DSP Code Generation*, in "Proc. ACM/IEEE CASES", October 2002

[7] S. PILLEMENT, O. SENTIEYS, R. DAVID. *DART: A Functional-Level Reconfigurable Architecture for High Energy Efficiency*, in "EURASIP Journal on Embedded Systems (JES)", 2008, pp. 1-13

[8] R. ROCHER, D. MENARD, O. SENTIEYS, P. SCALART. *Analytical Approach for Numerical Accuracy Estimation of Fixed-Point Systems Based on Smooth Operations*, in "IEEE Transactions on Circuits and Systems. Part I, Regular Papers", October 2012, vol. 59, n° 10, pp. 2326 - 2339 [*DOI : 10.1109/TCSI.2012.2188938*], http://hal.inria.fr/hal-00741741

[9] C. WOLINSKI, M. GOKHALE, K. MCCABE. *A polymorphous computing fabric*, in "IEEE Micro", 2002, vol. 22, n° 5, pp. 56–68

[10] C. WOLINSKI, K. KUCHCINSKI, E. RAFFIN. *Automatic Design of Application-Specific Reconfigurable Processor Extensions with UPaK Synthesis Kernel*, in "ACM Trans. on Design Automation of Elect. Syst.", 2009, vol. 15, n° 1, pp. 1–36, http://doi.acm.org/10.1145/1640457.1640458

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] B. BARROIS. *Methods to Evaluate Accuracy-Energy Trade-Off in Operator-Level Approximate Computing*, Université de Rennes 1, December 2017, https://hal.inria.fr/tel-01665015

[12] G. DEEST. *Implementation Trade-Offs for FGPA accelerators* , Université de Rennes 1 [UR1], December 2017, https://tel.archives-ouvertes.fr/tel-01665020

[13] X.-C. LE. *Improving performance of non-intrusive load monitoring with low-cost sensor networks*, Université Rennes 1, April 2017, https://tel.archives-ouvertes.fr/tel-01622355

[14] R. RAGAVAN. *Error handling and energy estimation for error resilient near-threshold computing*, Université Rennes 1, September 2017, https://tel.archives-ouvertes.fr/tel-01654476

[15] R. RAGAVAN. *Error Handling and Energy Estimation Framework For Error Resilient Near-Threshold Computing*, Rennes 1, September 2017, https://hal.inria.fr/tel-01636803

[16] B. ROUX. *Methodology and Tools for Energy-aware Task Mapping on Heterogeneous Multiprocessor Architectures*, Université de Rennes 1, November 2017, https://hal.inria.fr/tel-01672814

Articles in International Peer-Reviewed Journals

[17] A. DORAI, V. FRESSE, C. COMBES, E.-B. BOURENNANE, A. MTIBAA. *A collision management structure for NoC deployment on multi-FPGA*, in "Microprocessors and Microsystems: Embedded Hardware Design (MICPRO)", March 2017, vol. 49, pp. 28 - 43 [*DOI : 10.1016/J.MICPRO.2017.01.006*], https://hal-univ-bourgogne.archives-ouvertes.fr/hal-01484378

[18] M. FYRBIAK, S. ROKICKI, N. BISSANTZ, R. TESSIER, C. PAAR. *Hybrid Obfuscation to Protect against Disclosure Attacks on Embedded Microprocessors*, in "IEEE Transactions on Computers", 2017, https://hal.inria.fr/hal-01426565

[19] A. KRITIKAKOU, T. MARTY, M. ROY. *DYNASCORE: DYNAmic Software COntroller to increase REsource utilization in mixed-critical systems*, in "ACM Transactions on Design Automation of Electronic Systems (TODAES)", September 2017, vol. 23, n° 2, art ID n°13 [*DOI : 10.1145/3110222*], https://hal.archives-ouvertes.fr/hal-01559696

[20] H. LI, S. LE BEUX, M. J. SEPULVEDA FLOREZ, I. O'CONNOR. *Energy-Efficiency Comparison of Multi-Layer Deposited Nanophotonic Crossbar Interconnects*, in "ACM Journal on Emerging Technologies in Computing Systems", July 2017, vol. XX, https://hal.inria.fr/hal-01508192

[21] T. H. NGUYEN, M. GAY, F. GOMEZ AGIS, S. LOBO, O. SENTIEYS, J.-C. SIMON, C. PEUCHERET, L. BRAMERIE. *Impact of ADC parameters on linear optical sampling systems*, in "Optics Communications", November 2017, vol. 402, pp. 362-367 [*DOI : 10.1016/J.OPTCOM.2017.06.013*], https://hal.archives-ouvertes.fr/hal-01576164

[22] T. H. NGUYEN, P. SCALART, M. GAY, L. BRAMERIE, O. SENTIEYS, J.-C. SIMON, C. PEUCHERET, M. JOINDOT. *Blind transmitter IQ imbalance compensation in M-QAM optical coherent systems*, in "Journal of optical communications and networking", September 2017, vol. 9, n° 9, pp. D42-D50 [*DOI : 10.1364/JOCN.9.000D42*], https://hal.archives-ouvertes.fr/hal-01573632

[23] B. ROUXEL, S. DERRIEN, I. PUAUT. *Tightening Contention Delays While Scheduling Parallel Applications on Multi-core Architectures*, in "ACM Transactions on Embedded Computing Systems (TECS)", October 2017, vol. 16, n° 5s, pp. 1 - 20 [*DOI : 10.1145/3126496*], https://hal.archives-ouvertes.fr/hal-01655383

[24] C. XIAO, S. WANG, W. LIU, E. CASSEAU. *Parallel Custom Instruction Identification for Extensible Processors*, in "Journal of Systems Architecture", May 2017, vol. 76, pp. 149-159 [*DOI : 10.1016/J.SYSARC.2016.11.011*], https://hal.inria.fr/hal-01587020

International Conferences with Proceedings

[25] L. AUDREY, A. TISSERAND. *ECC Protections against both Observation and Pertubation Attacks*, in "CryptArchi 2017: 15th International Workshops on Cryptographic architectures embedded in logic devices", Smolenice, Slovakia, June 2017, https://hal.archives-ouvertes.fr/hal-01545752

[26] B. BARROIS, O. SENTIEYS. *Customizing Fixed-Point and Floating-Point Arithmetic - A Case Study in K-Means Clustering*, in "SiPS 2017 - IEEE International Workshop on Signal Processing Systems", Lorient, France, October 2017, https://hal.inria.fr/hal-01633723

[27] B. BARROIS, O. SENTIEYS, D. MENARD. *The Hidden Cost of Functional Approximation Against Careful Data Sizing – A Case Study*, in "Design, Automation & Test in Europe Conference & Exhibition (DATE 2017)", Lausanne, France, 2017, https://hal.inria.fr/hal-01423147

[28] T. BOLLENGIER, L. LAGADEC, M. NAJEM, J.-C. LE LANN, P. GUILLOUX. *Soft timing closure for soft programmable logic cores: The ARGen approach* , in "ARC 2017 - 13th International Symposium on Applied Reconfigurable Computing", Delft, Netherlands, Delft University of Technology , April 2017, https://hal.archives-ouvertes.fr/hal-01475251

[29] S. CHERUBIN, G. AGOSTA, I. LASRI, E. ROHOU, O. SENTIEYS. *Implications of Reduced-Precision Computations in HPC: Performance, Energy and Error*, in "International Conference on Parallel Computing (ParCo)", Bologna, Italy, September 2017, https://hal.inria.fr/hal-01633790

[30] G. DEEST, T. YUKI, S. RAJOPADHYE, S. DERRIEN. *One size does not fit all: Implementation trade-offs for iterative stencil computations on FPGAs*, in "FPL - 27th International Conference on Field Programmable Logic and Applications", Gand, Belgium, IEEE, September 2017 [*DOI : 10.23919/FPL.2017.8056781*], https://hal.inria.fr/hal-01655590

[31] S. DERRIEN, I. PUAUT, P. ALEFRAGIS, M. BEDNARA, H. BUCHER, C. DAVID, Y. DEBRAY, U. DURAK, I. FASSI, C. FERDINAND, D. HARDY, A. KRITIKAKOU, G. RAUWERDA, S. REDER, M. SICKS, T. STRIPF, K. SUNESEN, T. TER BRAAK, N. VOROS, J. †. BECKER. *WCET-aware parallelization of model-based applications for multi-cores: The ARGO approach*, in "Design Automation and Test in Europe (DATE), 2017", Lausanne, Switzerland, March 2017, pp. 286 - 289 [*DOI :* 10.23919/DATE.2017.7927000], http://hal.upmc.fr/hal-01590418

[32] A. DORAI, O. SENTIEYS, H. DUBOIS. *Evaluation of NoC on Multi-FPGA Interconnection Using GTX Transceiver*, in "24th IEEE International Conference on Electronics, Circuits and Systems (ICECS)", Batumi, Georgia, December 2017, https://hal.inria.fr/hal-01633785

[33] A. H. EL MOUSSAWI, S. DERRIEN. *Superword Level Parallelism aware Word Length Optimization*, in "Design, Automation & Test in Europe Conference & Exhibition (DATE 2017)", Lausanne, Switzerland, D. ATIENZA, G. D. NATALE (editors), IEEE, March 2017, https://hal.inria.fr/hal-01425550

[34] G. GALLIN, T. OZLUM CELIK, A. TISSERAND. *Architecture level Optimizations for Kummer based HECC on FPGAs*, in "IndoCrypt 2017 - 18th International Conference on Cryptology in India", Chennai, India, International Conference in Cryptology in India : Progress in Cryptology – INDOCRYPT 2017, Springer, December 2017, vol. 10698, pp. 44-64 [*DOI :* 10.1007/978-3-319-71667-1_3], https://hal.archives-ouvertes.fr/hal-01614063

[35] G. GALLIN, A. TISSERAND. *Hardware Architectures for HECC*, in "CryptArchi 2017: 15th International Workshops on Cryptographic architectures embedded in logic devices", Smolenice, Slovakia, June 2017, https://hal.archives-ouvertes.fr/hal-01545625

[36] G. GALLIN, A. TISSERAND. *Hyper-Threaded Multiplier for HECC*, in "Asilomar Conference on Signals, Systems, and Computers", Pacific Grove, CA, United States, IEEE, October 2017, https://hal.archives-ouvertes.fr/hal-01620046

[37] C. KILLIAN, D. CHILLET, S. LE BEUX, O. SENTIEYS, V. D. PHAM, I. O'CONNOR. *Energy and Performance Trade-off in Nanophotonic Interconnects using Coding Techniques*, in "DAC 2017 - IEEE/ACM Design Automation Conference DAC", Austin, United States, June 2017, 6 p. , https://hal.inria.fr/hal-01495468

[38] R. KUMAR BUDHWANI, R. RAGAVAN, O. SENTIEYS. *Taking Advantage of Correlation in Stochastic Computing*, in "ISCAS 2017 - IEEE International Symposium on Circuits and Systems", Baltimore, United States, May 2017, https://hal.inria.fr/hal-01633725

[39] J. LUO, A. ELANTABLY, D. D. PHAM, C. KILLIAN, D. CHILLET, S. LE BEUX, O. SENTIEYS, I. O'CONNOR. *Performance and Energy Aware Wavelength Allocation on Ring-Based WDM 3D Optical NoC*, in "Design, Automation & Test in Europe Conference & Exhibition (DATE 2017)", Lausanne, Switzerland, March 2017, https://hal.inria.fr/hal-01416958

[40] L. MO, A. KRITIKAKOU, O. SENTIEYS. *Decomposed Task Mapping to Maximize QoS in Energy-Constrained Real-Time Multicores*, in "35th IEEE International Conference on Computer Design (ICCD)", Boston, United States, IEEE, November 2017, 6 p. , https://hal.inria.fr/hal-01633782

[41] R. PSIAKIS, A. KRITIKAKOU, O. SENTIEYS. *NEDA: NOP Exploitation with Dependency Awareness for Reliable VLIW Processors*, in "ISVLSI 2017 - IEEE Computer Society Annual Symposium on VLSI",

Bochum, Germany, May 2017, pp. 391-396 [*DOI :* 10.1109/ISVLSI.2017.75], https://hal.inria.fr/hal-01633770

[42] R. PSIAKIS, A. KRITIKAKOU, O. SENTIEYS. *Run-Time Instruction Replication for Permanent and Soft Error Mitigation in VLIW Processors*, in "NEWCAS 2017 - 15th IEEE International New Circuits and Systems Conference", Strasbourg, France, June 2017, pp. 321-324 [*DOI :* 10.1109/NEWCAS.2017.8010170], https://hal.inria.fr/hal-01633778

[43] R. RAGAVAN, B. BARROIS, C. KILLIAN, O. SENTIEYS. *Pushing the Limits of Voltage Over-Scaling for Error-Resilient Applications*, in "Design, Automation & Test in Europe Conference & Exhibition (DATE 2017)", Lausanne, Switzerland, March 2017, https://hal.archives-ouvertes.fr/hal-01417665

[44] S. ROKICKI, E. ROHOU, S. DERRIEN. *Hardware-Accelerated Dynamic Binary Translation*, in "IEEE/ACM Design, Automation & Test in Europe Conference & Exhibition (DATE)", Lausanne, Switzerland, March 2017, https://hal.inria.fr/hal-01423639

[45] B. ROUXEL, S. DERRIEN, I. PUAUT. *Tightening contention delays while scheduling parallel applications on multi-core architectures*, in "International Conference on Embedded Software (EMSOFT), 2017", Seoul, South Korea, International Conference on Embedded Software, October 2017, 20 p. [*DOI :* 10.1145/3126496], http://hal.upmc.fr/hal-01590508

[46] Y. UGUEN, F. DE DINECHIN, S. DERRIEN. *Bridging High-Level Synthesis and Application-Specific Arithmetic: The Case Study of Floating-Point Summations*, in "27th International Conference on Field-Programmable Logic and Applications (FPL)", Gent, Belgium, IEEE, September 2017, 8 p. , https://hal.inria.fr/hal-01373954

[47] I. WALI, E. CASSEAU, A. TISSERAND. *An Efficient Framework for Design and Assessment of Arithmetic Operators with Reduced-Precision Redundancy*, in "Conference on Design and Architectures for Signal and Image Processing (DASIP)", Dresden, Germany, September 2017, https://hal.inria.fr/hal-01586983

National Conferences with Proceedings

[48] D. P. VAN, D. CHILLET, C. KILLIAN, O. SENTIEYS, S. LE BEUX, I. O'CONNOR. *Electrical to Optical Interface for ONoC*, in "GRETSI 2017 - XXVIème colloque", Juan les Pins, France, September 2017, pp. 1-4, https://hal.inria.fr/hal-01655417

Conferences without Proceedings

[49] G. GALLIN, A. TISSERAND. *Hardware Architectures Exploration for Hyper-Elliptic Curve Cryptography*, in "Crypto'Puces 2017- 6ème rencontre Crypto'Puces, du composant au système communicant embarqué", Porquerolles, France, May 2017, 31 p. , https://hal.archives-ouvertes.fr/hal-01547034

Research Reports

[50] H. YVIQUEL, A. SANCHEZ, R. MICKAËL, E. CASSEAU. *Multicore Runtime for Dynamic Dataflow Video Decoders*, IETR/INSA Rennes ; IRISA, Inria Rennes, April 2017, https://hal.archives-ouvertes.fr/hal-01503378

Other Publications

[51] D. CHILLET, D. P. VAN, C. KILLIAN, O. SENTIEYS, S. LE BEUX, I. O'CONNOR. *Integration of an Optical NoC into multicore architecture*, June 2017, pp. 1-2, 2017 - XIIème Colloque National du GDR SoC-SiP, Poster, https://hal.inria.fr/hal-01655420

[52] P. DOBIAS, E. CASSEAU, O. SINNEN. *Poster: Fault-Tolerant Multi-Processor Scheduling with Backup Copy Technique*, September 2017, Conference on Design and Architectures for Signal and Image Processing (DASIP), Poster, https://hal.inria.fr/hal-01610745

[53] G. GALLIN, A. TISSERAND. *Finite Field Multiplier Architectures for Hyper-Elliptic Curve Cryptography*, June 2017, Colloque National du GDR SOC2, Poster, https://hal.archives-ouvertes.fr/hal-01539852

[54] S. ROKICKI, E. ROHOU, S. DERRIEN. *Supporting Runtime Reconfigurable VLIWs Cores Through Dynamic Binary Translation*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01653110

[55] Y. UGUEN, F. DE DINECHIN, S. DERRIEN. *A high-level synthesis approach optimizing accumulations in floating-point programs using custom formats and operators*, January 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01498357

[56] Y. UGUEN, F. DE DINECHIN, S. DERRIEN. *High-Level Synthesis Using Application-Specific Arithmetic: A Case Study*, April 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01502644

References in notes

[57] S. HAUCK, A. DEHON (editors). *Reconfigurable Computing: The Theory and Practice of FPGA-Based Computation*, Morgan Kaufmann, 2008

[58] V. BAUMGARTE, G. EHLERS, F. MAY, A. NÜCKEL, M. VORBACH, M. WEINHARDT. *PACT XPP — A Self-Reconfigurable Data Processing Architecture*, in "The Journal of Supercomputing", 2003, vol. 26, n⁰ 2, pp. 167–184

[59] C. BECKHOFF, D. KOCH, J. TORRESEN. *Portable module relocation and bitstream compression for Xilinx FPGAs*, in "24th Int. Conf. on Field Programmable Logic and Applications (FPL)", 2014, pp. 1–8

[60] C. BOBDA. *Introduction to Reconfigurable Comp.: Architectures Algorithms and Applications*, Springer, 2007

[61] S. BORKAR, A. A. CHIEN. *The Future of Microprocessors*, in "Commun. ACM", May 2011, vol. 54, n⁰ 5, pp. 67–77, http://doi.acm.org/10.1145/1941487.1941507

[62] J. M. P. CARDOSO, P. C. DINIZ, M. WEINHARDT. *Compiling for reconfigurable computing: A survey*, in "ACM Comput. Surv.", June 2010, vol. 42, 13:1 p. , http://doi.acm.org/10.1145/1749603.1749604

[63] K. COMPTON, S. HAUCK. *Reconfigurable computing: a survey of systems and software*, in "ACM Comput. Surv.", 2002, vol. 34, n⁰ 2, pp. 171–210, http://doi.acm.org/10.1145/508352.508353

[64] J. CONG, H. HUANG, C. MA, B. XIAO, P. ZHOU. *A Fully Pipelined and Dynamically Composable Architecture of CGRA*, in "IEEE Int. Symp. on Field-Program. Custom Comput. Machines (FCCM)", 2014, pp. 9–16, http://dx.doi.org/10.1109/FCCM.2014.12

[65] G. Constantinides, P. Cheung, W. Luk. *Wordlength optimization for linear digital signal processing*, in "IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems", October 2003, vol. 22, n⁰ 10, pp. 1432- 1442

[66] M. Coors, H. Keding, O. Luthje, H. Meyr. *Fast Bit-True Simulation*, in "Proc. ACM/IEEE Design Automation Conference (DAC)", Las Vegas, june 2001, pp. 708-713

[67] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, A. R. LeBlanc. *Design of ion-implanted MOSFET's with very small physical dimensions*, in "IEEE Journal of Solid-State Circuits", 1974, vol. 9, n⁰ 5, pp. 256–268

[68] A. Hormati, M. Kudlur, S. Mahlke, D. Bacon, R. Rabbah. *Optimus: efficient realization of streaming applications on FPGAs*, in "Proc. ACM/IEEE CASES", 2008, pp. 41–50

[69] H. Kalte, M. Porrmann. *REPLICA2Pro: Task Relocation by Bitstream Manipulation in Virtex-II/Pro FPGAs*, in "3rd Conference on Computing Frontiers (CF)", 2006, pp. 403–412

[70] J.-E. Lee, K. Choi, N. D. Dutt. *Compilation Approach for Coarse-Grained Reconfigurable Architectures*, in "IEEE Design and Test of Computers", 2003, vol. 20, n⁰ 1, pp. 26-33, http://doi.ieeecomputersociety.org/10.1109/MDT.2003.1173050

[71] H. Lee, D. Nguyen, J.-E. Lee. *Optimizing Stream Program Performance on CGRA-based Systems*, in "52nd IEEE/ACM Design Automation Conference", 2015, pp. 110:1–110:6, http://doi.acm.org/10.1145/2744769.2744884

[72] B. Mei, S. Vernalde, D. Verkest, H. De Man, R. Lauwereins. *ADRES: An architecture with tightly coupled VLIW processor and coarse-grained reconfigurable matrix*, in "Proc. FPL", Springer, 2003, pp. 61–70

[73] N. R. Miniskar, S. Kohli, H. Park, D. Yoo. *Retargetable Automatic Generation of Compound Instructions for CGRA Based Reconfigurable Processor Applications*, in "Proc. ACM/IEEE CASES", 2014, pp. 4:1–4:9, http://doi.acm.org/10.1145/2656106.2656125

[74] Y. Park, H. Park, S. Mahlke. *CGRA express: accelerating execution using dynamic operation fusion*, in "Proc. Int. Conf. on Compilers, Architecture, and Synthesis for Embedded Systems", New York, NY, USA, CASES'09, ACM, 2009, pp. 271–280, http://doi.acm.org/10.1145/1629395.1629433

[75] A. Putnam, A. Caulfield, E. Chung, D. Chiou, K. Constantinides, J. Demme, H. Esmaeilzadeh, J. Fowers, G. Gopal, J. Gray, M. Haselman, S. Hauck, S. Heil, A. Hormati, J.-Y. Kim, S. Lanka, J. Larus, E. Peterson, S. Pope, A. Smith, J. Thong, P. Xiao, D. Burger. *A reconfigurable fabric for accelerating large-scale datacenter services*, in "ACM/IEEE 41st International Symposium on Computer Architecture (ISCA)", June 2014, pp. 13-24, http://dx.doi.org/10.1109/ISCA.2014.6853195

[76] G. Theodoridis, D. Soudris, S. Vassiliadis. *2*, in "A survey of coarse-grain reconfigurable architectures and CAD tools", Springer Verlag, 2007

[77] G. Venkataramani, W. Najjar, F. Kurdahi, N. Bagherzadeh, W. Bohm, J. Hammes. *Automatic compilation to a coarse-grained reconfigurable system-on-chip*, in "ACM Trans. on Emb. Comp. Syst.", 2003, vol. 2, n⁰ 4, pp. 560–589, http://doi.acm.org/10.1145/950162.950167