

2025 Activity Report

RESEARCH CENTRE: Inria Centre at Université Grenoble Alpes

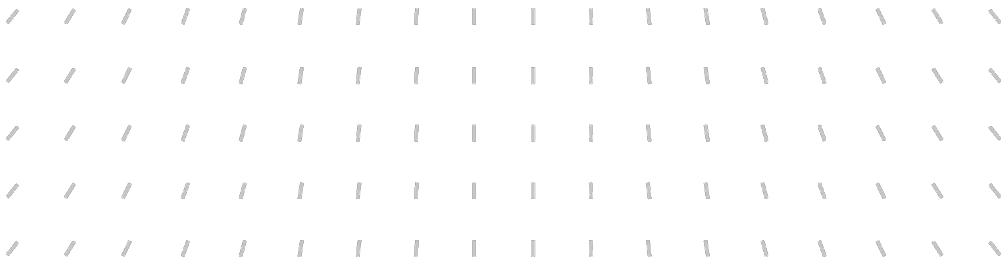
IN PARTNERSHIP WITH: CNRS, Institut polytechnique de Grenoble, Université de Grenoble Alpes

Project-Team

MADMAX

Moore, Amdahl, Dennard to their MAXimum

In collaboration with Techniques de l'Informatique et de la Microélectronique pour l'Architecture des systèmes intégrés



Project-Team MADMAX

Creation of the Project-Team: 2025 August 01

Each year, Inria research teams publish an Activity Report presenting their work and results over the reporting period. These reports follow a common structure, with some optional sections depending on the specific team. They typically begin by outlining the overall objectives and research programme, including the main research themes, goals, and methodological approaches. They also describe the application domains targeted by the team, highlighting the scientific or societal contexts in which their work is situated. The reports then present the highlights of the year, covering major scientific achievements, software developments, or teaching contributions. When relevant, they include sections on software, platforms, and open data, detailing the tools developed and how they are shared. A substantial part is dedicated to new results, where scientific contributions are described in detail, often with subsections specifying participants and associated keywords. Finally, the Activity Report addresses funding, contracts, partnerships, and collaborations at various levels, from industrial agreements to international cooperations. It also covers dissemination and teaching activities, such as participation in scientific events, outreach, and supervision. The document concludes with a presentation of scientific production, including major publications and those produced during the year.

Keywords

Computer sciences and digital sciences

- A1.1. – Architectures
 - A1.1.1. – Multicore, Manycore
 - A1.1.2. – Hardware accelerators (GPGPU, FPGA, etc.)
 - A1.1.10. – Reconfigurable architectures
- A1.6. – Green Computing
- A2.3. – Embedded and cyber-physical systems
- A4.5. – Formal method for verification, reliability, certification
- A9.7. – AI algorithmics

Other research topics and application domains

- B6.6. – Embedded systems

Contents

Project-Team MADMAX	1
1 Team members, visitors, external collaborators	4
2 Overall objectives	4
3 Research program	4
4 Application domains	5
5 Social and environmental responsibility	5
5.1 Footprint of research activities	5
5.2 Impact of research results	5
6 Highlights of the year	5
6.1 Awards	5
7 Latest software developments, platforms, open data	6
7.1 Latest software developments	6
7.1.1 cv-hpdcache	6
7.1.2 gemu-riscv128	6
7.1.3 rv128-toolchain	7
8 New results	7
9 Bilateral contracts and grants with industry	11
9.1 Bilateral contracts with industry	11
9.2 Bilateral Grants with Industry	12
10 Partnerships and cooperations	12
10.1 International initiatives	12
10.2 European initiatives	13
10.2.1 Horizon Europe	13
10.3 National initiatives	13
11 Dissemination	14
11.1 Promoting scientific activities	14
11.1.1 Scientific events: organisation	14
11.1.2 Journal	14
11.1.3 Leadership within the scientific community	15
11.1.4 Research administration	15
11.2 Teaching - Supervision - Juries - Educational and pedagogical outreach	15
11.2.1 Heavy teaching duties	15
11.2.2 Supervision	16
11.2.3 Juries	16
12 Scientific production	16
12.1 Major publications	16
12.2 Publications of the year	16

1 Team members, visitors, external collaborators

Research Scientists

- Cesar Fuguet Tortolero [INRIA, Researcher, from Aug 2025]
- Arthur Pérais [CNRS, Researcher, from Aug 2025]

Faculty Members

- Frédéric Pétrot [Team leader, Grenoble INP - UGA, Professor, from Aug 2025, HDR]
- Liliana Andrade [Grenoble INP - UGA, Associate Professor, from Aug 2025, TIMA T423 / Polytech B217]
- Julie Dumas [Grenoble INP - UGA, Associate Professor, from Aug 2025]
- Olivier Muller [Grenoble INP - UGA, Associate Professor, from Aug 2025]
- Laurence Pierre [UGA, Professor, from Aug 2025, HDR]
- Frédéric Rousseau [Grenoble INP - UGA, Professor, from Aug 2025, HDR]

PhD Students

- Andrei Ilin [UGA, from Dec 2025]
- Kilian Mc Govern [UGA, from Aug 2025]
- Abdallah Meebed [UGA, from Aug 2025]
- Olivier Romane [UGA, from Aug 2025]
- Johan Soderstrom [INRIA, from Nov 2025]

Administrative Assistant

- Myriam Etienne [INRIA]

2 Overall objectives

Our motto is "*Making the most of transistors in times of dwindling semiconductor technology*".

There are many challenges to tackle in the context of Moore's law slowing down and Dennard scaling law not applying anymore for the design and implementation of computing machines. Among those, our objectives are (1) to contribute to more power efficient mid to high-performance processor architectures, needed for sequential performance, (2) to propose dedicated hardware support to Artificial Intelligence workloads, to support execution of relatively large neural networks at an acceptable cost, for edge applications, and (3) define new design and verification methods for software centric hardware systems, and develop tools to put them into practice.

3 Research program

Micro-architecture: Discovery of yet more execution parallelism is challenging, but also rewarding. It can be at instruction level by leveraging speculation where possible, and at thread level by letting users provide synchronization or data-sharing hints, either within or among processors. For the latter case, the efficiency of the memory accesses remains key, and we therefore also work on the details of the cache coherency protocols: the high-level state machines are now well established, but a safe path to implementation is challenging,

AI acceleration: Hardware acceleration of AI workloads is inevitable. For example, 100x in power-efficiency compared to CPU is necessary to sustain current AI deployment according to the GAFAM. Although developing hardware at the RTL is hard, we believe that it is a necessary evil to reach best-in-class performance required by these ubiquitous devices. We work on extreme quantization for weights and activations, on the tiling and scheduling of memory accesses for hardware level folded networks, and weight compression,

Computer-aided-design: Although it does not provide any formal guarantees, simulation is the technology used to validate, in particular, software centric systems. Speed and accuracy are key metrics of a simulation infrastructure, but as we know, architects cannot have their cake and eat it too, and fast simulation is inaccurate, while accurate simulation is slow. We work with QEMU, a fast simulator we contributed to, for modeling cache coherence protocols, 128-bit architectures, etc. Semi-formal verification (runtime verification, not exhaustive but fully automatable) or formal verification methods (less automatic but exhaustive) can provide additional guarantees about functional or non-functional properties. We in particular look into SAIL, a framework for formal ISA description, to validate dynamic binary translation.

4 Application domains

The team develops core technologies that is meant to be usable in many different contexts, and does not focus on one or several specific applications.

5 Social and environmental responsibility

MADMAX work is at the boundary between hardware and software, and semiconductor manufacturing, used for the fabrication of the devices we design, consumes significant energy and water and emits toxic chemicals and greenhouse gases. But without this very sophisticated industry and its capability to shrink the size of the devices, there will be no software and no software evolution as we know it.

5.1 Footprint of research activities

In the absence of monitoring on our premises, we can hardly give a quantitative information. We nevertheless host two large scale servers (96 and 128 cores respectively), and a server with a Blackwell GPU, consuming around 700 W each.

5.2 Impact of research results

The team focuses on better power efficiency for computations, be it for general purpose computing or hardware acceleration of AI. In that sense, it has both a positive and negative environmental impact. Indeed, by lowering the amount of power needed for some computing task, it makes this task feasible at a larger scale. This vicious circle is call Jevon's paradox, and is typical of our modern societies. LED lightning is quite interesting in that regard (see <https://hal.science/hal-05396402/>).

6 Highlights of the year

6.1 Awards

- Best paper award for our paper "HPC Workload Analysis Using Distributed Cross-ISA Binary Instrumentation", presented at the 40th Conference on Design of Circuits and Integrated Systems, 2025 [8]
- Best paper award for the paper to which we contributed "FetchFlare: An Open-Source Strided Data Prefetcher for High-Performance Cache Hierarchies", presented at the 28th Euromicro Conference on Digital System Design (DSD), 2025 [6]

- Best paper award for the paper to which we contributed "Ramping Up Open-Source RISC-V Cores: Assessing the Energy Efficiency of Superscalar, Out-of-Order Execution", presented at the 22nd ACM International Conference on Computing Frontiers, 2025 [5]

7 Latest software developments, platforms, open data

7.1 Latest software developments

7.1.1 cv-hpdcache

Name: OpenHW Core-V High-Performance L1 Dcache (CV-HPDcache)

Keywords: Cache, Memory hierarchy

Scientific Description: L1 instruction and data cache supporting several out-of-order transactions and hit under miss.

Functional Description: This cache hold the data and instructions close to the processor to minimize the latency of memory accesses. Its microarchitecture supports multiple simultaneous accesses, with requests returned in random order to allow pending instructions to be released as quickly as possible. It uses separate memory banks to activate only the necessary parts, with a view to saving energy.

Release Contributions: This releases include some bugfixes and optimizations.

Added

Support responses for CMO operations when need_rsp in the request is set Support not-power-of-two number of entries in the Flush controller

Fixed

Fix implementation of the data merge logic in the write buffer to improve the area Fix assertions syntax Fix CMO flushes shall unset the dirty bit in the cache directory Fix handling of bus errors on write misses Fix prefetch requests shall update PLRU bits Fix initialization of the CMO handler flush request valid register

URL: <https://github.com/openhwgroup/cv-hpdcache>

Publications: hal-05452261v1, cea-04110679v1

Contact: Cesar Fuguet Tortolero

7.1.2 qemu-riscv128

Name: Quick emulation of the rv128 instruction set architecture

Keywords: Full system simulation, Emulation

Functional Description: This program is a QEMU fork that supports elf128. Emulation of 128-bit riscv instructions has been available in QEMU (<https://github.com/qemu/qemu>) since january 2022. This is the first (and currently only) simulator of a 128-bit processor.

Release Contributions: Support for reading executables using the elf128 format.

URL: <https://github.com/fpetrot/qemu-riscv128/>

Contact: Frederic Petrot

7.1.3 rv128-toolchain

Name: Cross-compilation toolchain for 128-bit riscv

Keyword: Compilers

Functional Description: This tool-chain allows to cross-compile, assemble, link and debug C programs to target the rv128 riscv instruction set architecture. This cross-compilation environment is based on gcc, the binutils, and gdb. It also contains the definition of what is the elf128 data format.

Release Contributions: Bug corrections

URL: <https://github.com/fpetrot/riscv-gnu-toolchain.git>

Contact: Frederic Petrot

8 New results

CVA6S+: A Superscalar RISC-V Core with High-Throughput Memory Architecture

Participants: Riccardo Tedeschi (University of Bologna), Gianmarco Ottavi (University of Bologna), Côme Allart (Thales and Mines Saint-Etienne, CEA, Leti, Centre CMP), Nils Wistoff (ETH Zurich), Zexin Fu (ETH Zurich), Filippo Grillotti (STMicroelectronics, Agrate Brianza), Fabio de Ambroggi (STMicroelectronics, Agrate Brianza), Elio Guidetti (STMicroelectronics, Agrate Brianza), Jean-Baptiste Rigaud (Mines Saint-Etienne, CEA, Leti, Centre CMP), Olivier Potin (Mines Saint-Etienne, CEA, Leti, Centre CMP), Jean-Roch Coulon (Thales), César Fuguet, Luca Benini (ETH Zurich and University of Bologna), Davide Rossi (University of Bologna).

Open-source RISC-V cores are increasingly adopted in high-end embedded domains such as automotive, where maximizing instructions per cycle (IPC) is becoming critical. Building on the industry-supported open-source CVA6 core and its superscalar variant, CVA6S, we introduce CVA6S+, an enhanced version incorporating improved branch prediction, register renaming and enhanced operand forwarding. These optimizations enable CVA6S+ to achieve a 43.5% performance improvement over the scalar configuration and 10.9% over CVA6S, with an area overhead of just 9.30% over the scalar core (CVA6). Furthermore, we integrate CVA6S+ with the OpenHW Core-V High-Performance L1 Dcache (HPDCache) and report a 74.1% bandwidth improvement over the legacy CVA6 cache subsystem.

Published as "CVA6S+: A superscalar RISC-V core with high-throughput memory architecture." arXiv preprint arXiv:2505.03762 (2025).

Depth-first: A deterministic and scalable NoC routing protocol for 3.5D packaged architectures [13]

Participants: Davy Million (CEA List), César Fuguet, Adrian Evans (CEA List), Rim El Cheikh (Université Clermont-Auvergne), Alireza Monemi (Barcelona SuperComputing Center), Jonathan Balkind (University of Santa Barbara), Frédéric Pétrot.

New high-volume commercial products combine 2.5D silicon-interposer based assemblies with 3D monolithic stacks of chiplets. This combination is called 3.5D packaging and makes it possible to assemble dense compute solutions. Components communicate via a Network-On-Chip, but current solutions do not support 3.5D Network-On-Chip topologies. To this end, this work proposes Depth-First, the first Deterministic, Virtual Channel based, Network-On-Chip routing protocol supporting 3.5D network topologies. The protocol

prevents deadlocks using additional Virtual Channels only in the upper chiplets, while imposing no VC constraints on the base interposer. Depth-First also features an efficient node naming scheme, enabling highly compact routing tables. Since vertical links must be assigned to routers, we present a Mixed-Integer Linear Programming formulation that greatly speeds up execution time compared to a reference implementation from prior work, which was based on an exhaustive search. We formally prove that the protocol is deadlock-free, study its performance using an open-source cycle-accurate simulator, and compare it with other protocols (on a comparable topology). A partial implementation of Depth-First in an open-source router results in a small 4.9% area impact (7nm process) compared to an implementation without our routing algorithm.

Published as "Depth-first: A deterministic and scalable NoC routing protocol for 3.5 D packaged architectures." IEEE Journal on Emerging and Selected Topics in Circuits and Systems (2025) [4].

Hardware-software co-design for supporting shared distributed virtual memory

Participants: Eduardo Tomasi Ribeiro (CEA List), César Fuguet, Christian Fabre (CEA List), Frédéric Pétrot.

With network technologies now offering latencies approaching that of memory, sharing distributed memory becomes an increasingly feasible approach. However, memory virtualization remains largely unexplored on distributed supercomputers, and developers rely on complex programming models to achieve high performance. We propose a hardware-software co-design approach to enable a single virtual global address space. Our method introduces the concept of virtual nodes, which are identified by the most significant bits of the virtual address, thereby defining address space ranges that correspond to different nodes within the distributed system. To support this approach, we provide the necessary hardware support and validate its efficacy using two open-source simulators: gem5 and SST. Experimental results demonstrate that programming this system closely resembles the shared memory programming model, and we present its scalability using a benchmark. These results demonstrate the viability of implementing a shared virtual address space in distributed systems, simplifying the development of high-performance computing applications.

Published as "Hardware-software co-design for supporting shared distributed virtual memory." Proceedings of the 22nd ACM International Conference on Computing Frontiers, 2025 [9].

Variable and extended precision (VRP) accelerator implemented in a 22 nm SoC

Participants: Eric Guthmuller (CEA), César Fuguet, Andrea Bocco (CEA), Jérôme Fereyre (CEA), Adrian Evans (CEA), Yves Durand (CEA), Jérôme Fereyre (CEA).

Linear solvers and eigensolvers are the heart of HPC scientific applications. Among them, iterative projection methods are preferred to direct algorithms for large problems because of their lower memory usage, but they are prone to roundoff errors. Using an enhanced working precision inside the linear computing kernels mitigates this issue and accelerates convergence. However, only software libraries support variable and extended precision Floating Point (FP) computations beyond 80 bits. We introduce the VaRiable and extended Precision Accelerator (VRP), a RISC-V accelerator implemented on a System-on-Chip (SoC) using GF22FDX technology. The VRP supports FP computations with a range of significand bits from 2 to 512. This accelerator delivers an average 19.25x application speedup compared to the well-known MPFR software library running on a 2400+ MHz Intel Xeon processor. Additionally, extended precision facilitates the convergence of linear solvers for problems that would otherwise fail to converge and reduces energy-to-solution.

Published as "Variable and extended precision (VRP) accelerator implemented in a 22 nm SoC." Electronics Letters 61.1 (2025): e70255 [12].

On Benefits of Modeling the HPDcache in LNT

Participants: Zachary Assoumani, César Fuguet, Radu Mateescu, Wendelin Serwe.

Stepping from natural language towards modern formal languages such as LNT is beneficial for specifying hardware architectures. We illustrate this on the HPDcache, the informal specification of which contains numerous fragments in pseudo-code. Due to the syntactical similarities between the latter and LNT, modeling the HPDcache's informal specification in LNT was greatly facilitated. The CADP tools supporting LNT enabled us to spot an error in the informal specification of the HPDcache, which might have led to a violation of the memory consistency rules of the RISC-V.

Published as "On Benefits of Modeling the HPDcache in LNT." RISC-V 2025-RISC-V Summit Europe. 2025 [11].

HPC Workload Analysis Using Distributed Cross-ISA Binary Instrumentation

Participants: Eduardo Tomasi Ribeiro (CEA), César Fuguet, Christian Fabre (CEA), Frédéric Pétrot.

Developing distributed High Performance Computing (HPC) applications is challenging, with complex interactions between application, runtime environment, processing cores, and network to obtain the highest performance that a given distributed computing system can provide. HPC systems are evolving at a fast pace, so applications must often be ported. Generally, developers natively run their applications on current machines and extrapolate the performances on future ones. However, modern and future HPC machines contain multiple nodes, each with multiple general-purpose processor cores, possibly with an Instruction Set Architecture (ISA) different from the previous generations, as well as new domain-specific accelerators, so simple extrapolations may not be accurate. Instead, we propose an automated approach to execute and non-intrusively characterize distributed HPC applications on a QEMU-based, cross-ISA, distributed simulation platform. As part of this automated approach, we propose a QEMU plugin to extract metrics at runtime during the execution of distributed applications. The approach is demonstrated on a RISC-V-based distributed multinode architecture. It achieves an average speedup of almost 3.5× on a single host machine with 16 virtual nodes in comparison with a single node. Using QEMU plugins for collecting Message Passing Interface (MPI) runtime metrics slows the simulation by 1.62× in average, but overall, our approach remains much faster than other simulation platforms.

Published as "HPC Workload Analysis Using Distributed Cross-ISA Binary Instrumentation." 2025 40th Conference on Design of Circuits and Integrated Systems (DCIS). IEEE, 2025 [8]. Best Paper Award.

On the Hardware Implementation of Lala's 64-bit SEcDED Codes

Participants: Frédéric Pétrot, César Fuguet.

Protecting memories, and particularly caches, is necessary for devices running in harsh environments. The standard approach, and for good reason, is to use single-error correction, double-error detection (SEcDED) codes. Among the solutions, the codes introduced by Lala have not received a lot of attention, because they cost one more bit in memory. However, for the 64-bit word granularity, they feature the lowest number of '1' in their parity check matrices, which translates in less logical operations for encoding and correcting. This work covers the design space of Lala's codes from a practical point of view, through synthesis on a mature silicon technology. We in particular show that the circuit timing, area, and power characteristics highly vary with the actual matrix, which allows to devise what could be the most appropriate matrix for a given set of system level constraints.

Published as "On the Hardware Implementation of Lala's 64-bit SEcDED Codes." 2025 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT). IEEE, 2025 [7].

Ramping Up Open-Source RISC-V Cores: Assessing the Energy Efficiency of Superscalar Out-of-Order Execution

Participants: Zexin Fu (ETH Zurich), Riccardo Tedeschi (University of Bologna), Gianmarco Ottavi (University of Bologna), Nils Wistoff (ETH Zurich), César Fuguet, Davide Rossi (University of Bologna), Luca Benini (iETH Zurich and University of Bologna).

Open-source RISC-V cores are increasingly demanded in domains like automotive and space, where achieving high instructions per cycle (IPC) through superscalar and out-of-order (OoO) execution is crucial. However, high-performance open-source RISC-V cores face adoption challenges: some (e.g. BOOM, Xiangshan) are developed in Chisel with limited support from industrial electronic design automation (EDA) tools. Others, like the XuanTie C910 core, use proprietary interfaces and protocols, including non-standard AXI protocol extensions, interrupts, and debug support. In this work, we present a modified version of the OoO C910 core to achieve full RISC-V standard compliance in its debug, interrupt, and memory interfaces. We also introduce CVA6S+, an enhanced version of the dual-issue, industry-supported open-source CVA6 core. CVA6S+ achieves 34.4% performance improvement compared to the scalar configuration. We conduct a detailed performance, area, power, and energy analysis on the superscalar out-of-order C910, superscalar in-order CVA6S+ and vanilla, single-issue in-order CVA6, all implemented in GF22FDX technology and integrated into Cheshire, an open-source modular SoC platform. We examine the performance and efficiency of different microarchitectures using the same ISA, SoC, and implementation with identical technology, tools, and methodologies. The area and performance rankings of CVA6, CVA6S+, and C910 follow expected trends: compared to the scalar CVA6, CVA6S+ shows an area increase of 6% and an IPC improvement of 34.4%, while C910 exhibits a 75% increase in area and a 119.5% improvement in IPC. However, efficiency analysis reveals that CVA6S+ leads in area efficiency (GOPS/mm²), while the C910 is highly competitive in energy efficiency (GOPS/W). This challenges the common belief that high performance in superscalar and out-of-order cores inherently comes at a significant cost in terms of area and energy efficiency.

Published as "Ramping up open-source RISC-V cores: Assessing the energy efficiency of superscalar, out-of-order execution." Proceedings of the 22nd ACM International Conference on Computing Frontiers, 2025 [5].

FetchFlare: An Open-Source Strided Data Prefetcher for High-Performance Cache Hierarchies

Participants: Golnaz Korkian (Barcelona Supercomputing Center), Neiel Leyva (Barcelona Supercomputing Center and Universitat Politècnica de Catalunya), Arnau Bigas (Barcelona Supercomputing Center), Noelia Oliete-Escuín (Barcelona Supercomputing Center and Universitat Politècnica de Catalunya), Abbas Haghi (Barcelona Supercomputing Center), Alireza Monemi (Barcelona Supercomputing Center), César Fuguet, Lluc Alvarez (Barcelona Supercomputing Center).

In recent years, the rise of open-source hardware has transformed the landscape of technology development. In particular, RISC-V has offered hardware designers the possibility of designing processors in a much cheaper way by leveraging a rich ecosystem of open-source designs that can be easily reused, extended, and customized. Although the RISC-V ecosystem is rapidly growing and open-source processors are becoming increasingly sophisticated, some advanced architectural techniques typically employed in commercial high-performance processors are still not prevalent in RISC-V open-source architectures. Among them, hardware prefetchers have been ubiquitous in highend processors for many years, but they are not as commonly found in open-source RISC-V processors. To bridge this gap, this work presents FetchFlare, a stride prefetcher for highperformance cache hierarchies. FetchFlare is able to capture the memory access patterns of applications, predict future memory accesses, and issue prefetch requests for them. We provide an open-source RTL implementation of FetchFlare and integrate it into a complete open-source setup formed by the OpenPiton framework, the Sargantana core, and the High-Performance Data Cache (HPDCache). Compared to a baseline system

without prefetching, FetchFlare achieves an average speedup of 63%, avoids cache misses in the L1D and the L2 caches, and presents an average accuracy, coverage, and timeliness of 86%, 39%, and 99%, respectively.

Published as "FetchFlare: An Open-Source Strided Data Prefetcher for High-Performance Cache Hierarchies." 2025 28th Euromicro Conference on Digital System Design (DSD). IEEE, 2025 [6].

Address/Data Instruction Steering in Clustered General Purpose Processors

Participants: Chandana S. Deshpande, Arthur Perais, Frédéric Pétrot.

Although they differentiate between integer and floating-point datum, modern Instruction Set Architectures and their implementations do not differentiate integer datum used to address memory from integer datum used in purely arithmetic and logical computations. This is a perfectly reasonable choice as addresses are, in fact, integral quantities. However, in many cases, there is already a fundamental difference between addresses and integer data: Their width. As computer systems moved from 16 to 32, then to 64-bit pointers, with a potential future where 128-bit might be used for specific systems, the data width required to compute a given output with a given algorithm has remained the same, e.g., an ASCII character is still represented on a byte. This work aims to leverage this dichotomy to revisit hardware clustering, a well known microarchitectural technique used to mitigate the cost of scaling processor backend structures by dividing the backend into several mostly independent execution clusters. We show that by treating instructions as manipulating addresses or data and steering them to a "data" or an "address" cluster accordingly, reasonable cluster load balancing can be achieved without the need for complex steering policies that can lead to performance on par with the baseline with limited hardware overhead. Moreover, we highlight two possible optimizations stemming from this distribution. First, the registers of the "address" cluster can easily be compressed thanks to address spatial and temporal locality. Second, if a processor requires a large address space but only processes narrow data (e.g., 32-bit data with 64-bit pointers or 64-bit data with 128-bit pointers), the "data" cluster datapath can be kept narrower than the "address" cluster datapath.

Published as "Address/data instruction steering in clustered general purpose processors". ACM Transactions on Architecture and Code Optimization, 22(3), 1-24, 2025 [3].

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

HPDCache Hardening

Participants: César Fugueta, Frédéric Pétrot.

Contract with Thales, 130 k€ (Floralis), 1/11/2025-29/2/2026

One of the major concerns in aerospace applications, such as the ones targeted by the SCAF project, is that processor-based devices are subject to radiation, which may affect their functioning. A common consequence of radiation is bit-flips in the embedded (on-chip) memories: cache or scratchpad memories. Indeed, memories occupy more than 50% of modern processor-based chips to reduce the memory traffic to external DRAMs, which suffer from low bandwidth and high latency. This important footprint makes memories more vulnerable to radiation-induced soft errors. The goal of this project, titled "Hardening of the HPDCache for RISC-V processors", is to introduce fault-tolerant mechanisms in the HPDCache integrated to the CVA6 processor. The CVA6 is an open-source RISC-V processor, whose development is driven by Thales under the umbrella of the OpenHW Group, whilst the HPDCache is an open-source high-performance L1 Data Cache for RISC-V processors, initially developed by CEA and now part of the OpenHW Group IP portfolio. Specifically, this project has three objectives: (1) we aim to introduce error correcting codes (ECCs) in embedded SRAMs (Static Random Access Memories) of the HPDCache. We plan to use SECDED (Single Error Correction, Double Error Detection) codes; (2) we aim to introduce a periodic memory scrubbing

mechanism to limit the probability of multi-bit errors in the HPDcache SRAMs; (3) we aim to introduce necessary modifications in the interface from the CVA6 towards next level of cache or the main memory to support error detection and correction in memory transactions.

9.2 Bilateral Grants with Industry

“Back to the Future” – Predicting Hard to Predict Branches

Participants: Arthur Perais, Yiannakis Sazeides (University of Cyprus), Ioannis Constantinou (University of Cyprus).

Intel Grant, 10 k€ (UGA) and 150k€ (University of Cyprus) 1/9/2025-31/7/2026

Work on branch prediction for high performance processor, with a focus on hard to predict branches, in collaboration with the University of Cyprus.

10 Partnerships and cooperations

10.1 International initiatives

Cooperation with University of Santa Barbara

Participants: Davy Million (CEA), César Fuguet, Frédéric Pétrot, Adrian Evans (CEA), Jonathan Balkind (University of Santa-Barbara).

1/9/2022 - 31/12/2026

Cooperation between UCSB, CEA, UGA and Inria around contributions within the OpenPiton framework developed at UCSB. Work on Network on Chip and heterogeneous computing for chiplet based SoC, including routing algorithms, CPU/GPU tile, ... Also actual implementation on large FPGA fabrics hosted at CEA. Goal is to deliver an opensource hardware/software platform. PhD grant for Davy Million by CEA.

Cooperation with Barcelona SuperComputing Center

Participants: Andrei Ilin, César Fuguet, Frédéric Pétrot, Lluc Alvarez (Barcelona SuperComputing Center).

1/9/2025 - 31/8/2028

Cosupervision of Andrei Ilin PhD with BSC. PhD grant through UGA Idex call. The work is focused on chiplet based cache coherency, including optimization for CPU/GPU workloads. It takes place at the

Cooperation with University of Cyprus

Participants: Arthur Perais, Yiannakis Sazeides (University of Cyprus), Ioannis Constantinou (University of Cyprus).

1/3/2025 - 31/8/2025

Master student (Ioannis Constantinou) cosupervision on branch prediction [10]. Intel grant.

Cooperation with ETH Zurich and University of Bologna

Participants: César Fuguet, Riccardo Tedeschi (University of Bologna), Davide Rossi (University of Bologna), Luca Benini (ETH Zurich and University of Bologna).

Since 2024

Informal research cooperation around evolution of the CVA6 memory interface and memory hierarchy.

Cooperation with University of Murcia

Participants: Arthur Perais, Alberto Ros (Universidad de Murcia), Alexandra Jimborean (Universidad de Murcia), Sawan Singh (Universidad de Murcia), Ravikiran Ravindranath Reddy (Universidad de Murcia).

Informal research cooperation around high performance processor microarchitecture since 2021 within the framework of Alberto Ros ERC grant. Research is about speculation for branches, memory accesses, addresses, values, ...

10.2 European initiatives

10.2.1 Horizon Europe

EdgeAI

Participants: Ana Pinzari, Frédéric Pétrot.

Huge European project (35 M€), with 48 partners, funding of 200 k€ for Grenoble INP - UGA. Our contributions are on power efficient neural networks for grapewine diseases detection, using a mix of hardware and software solution. French partners are STMicroelectronics, Pommery, and Université Reims Champagne Ardennes.

10.3 National initiatives

PEPR IA Holigrail Project

Participants: Abdallah Meebed, Van Quan Pham, Olivier Romane, Adrien Prost-Boucle, Olivier Muller, Frédéric Pétrot.

Grant: 900 k€, Grenoble INP - UGA

We contribute to the Holigrail project with work on design and implementation of highly quantized neural networks, including pruning and compression issues.

PEPR Cloud Archi-CESAM Project

Participants: Arthur Perais, Louka Yerly.

Grant: 150 k€, Grenoble INP - UGA

We contribute to the Archi-CESAM project by characterizing the differences between user and kernel code at the microarchitectural level.

Défi Inria Cocorisco

Participants: Julie Dumas, Arthur Perais, Johan Söderström, Nevena Vasilevska.

Grant: 300 k€, Inria

We contribute both on HW/SW interaction to improve the performance of distributed garbage collection systems and to improve the performance of multithreaded programs through software-hinted management of the hardware coherency mechanism. Arthur Perais is co-PI with O. Sentiyeis (TARAN team at IRISA)

11 Dissemination

Participants: Liliana Andrade, Julie Dumas, César Fuguet, Olivier Muller, Arthur Perais, Frédéric Pétrot, Laurence Pierre, Frédéric Rousseau.

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

Chair of conference program committees

- Frédéric Pétrot, program chair of the 40th Conference on Design of Circuits and Integrated Systems

Member of the conference program committees

- Arthur Perais, member of ISCA, MICRO and HPCA program committee
- César Fuguet, member of CF, ICCD, ISCA program committee
- Laurence Pierre, member of DATE technical program committee
- Frédéric Pétrot, member of DSD technical program committee
- Liliana Andrade, member of DSD, MovVe4SPS workshop (CPS-IoT week) and COMPAS program committee

Member of steering committees

- Arthur Perais, steering committee chair for the ARCHI winter school
- Liliana Andrade, steering committee for the FETCH winter school

11.1.2 Journal

Member of the editorial boards

- Frédéric Pétrot, co-guest-editor for a special issue of Elsevier Microprocessors and Microsystems.

Reviewer - reviewing activities

- Frédéric Pétrot, reviewer for IEEE Transactions on Computer Aided Design of Circuits and Systems.
- César Fuguet, reviewer for IEEE Transactions on Computer Aided Design of Circuits and Systems and Elsevier Microprocessors and Microsystems.
- Liliana Andrade, reviewer for IEEE Transactions on Computer Aided Design of Circuits and Systems.
- Arthur Perais, reviewer for ACM Transactions on Architecture and Code Optimization and IEEE Computer Architecture Letters
- Frédéric Rousseau, reviewer for Elsevier Microprocessors and Microsystems.

11.1.3 Leadership within the scientific community

- César Fuguet, co-chair of the Interconnection Task Group within the OpenHW Foundation.
- Arthur Perais, co-coordinator of the "High-Performance Embedded Computing" track of GDR SOC²
- Frédéric Pétrot, member of the steering committee of GDR SOC²

11.1.4 Research administration

- Laurence Pierre is member of the UGA IM2AG Reseach Commission.
- Laurence Pierre is member of the IM2AG Concil on education and research (UFR)
- Frédéric Rousseau is head of the EEATS Doctoral School HdR council (ED220)
- Frédéric Pétrot is member of the MSTII Doctoral School HdR council (ED217)
- Liliana Andrade is member of the Polytech Grenoble School council
- Julie Dumas is member of the Ensimag School council and restricted council
- Frédéric Pétrot is member of the Ensimag School council and restricted council
- Frédéric Rousseau is member of Polytech restricted council
- Liliana Andrade is member of scientific council of TIMA laboratory
- Frédéric Rousseau is member of scientific council of TIMA laboratory
- Olivier Muller is member of TIMA laboratory council and scientific council

11.2 Teaching - Supervision - Juries - Educational and pedagogical outreach

11.2.1 Heavy teaching duties

- Frédéric Rousseau, head of admission at Polytech Grenoble
- Liliana Andrade, head of 5th year of the apprenticeship programme at Polytech Grenoble
- Liliana Andrade, co-head of admission in the apprenticeship programme at Polytech Grenoble
- Julie Dumas, co-head of "Ingénierie des systèmes d'information" programme at Ensimag
- Liliana Andrade, steering committee for the CFA FormaSup Auvergne-Rhône-Alped and Polytech Grenoble - E2i Engineering Degree
- Frédéric Rousseau, steering committee for the CFA FormaSup Auvergne-Rhône-Alped and Polytech Grenoble - E2i Engineering Degree

- Olivier Muller, responsible for student-entrepreneurship within Ensimag

Note that 6 of the 8 members of the MADMAX team are teachers, and they are responsible for modules with sometimes many students, which represents a lot of work too.

11.2.2 Supervision

Student Name	Year	State	School	Supervisors
Ravenel Pierre	4A	(s)	MSTII	Pétrot Frédéric - Perais Arthur
Deshpande Chandana	4A	(s)	MSTII	Pétrot Frédéric - Perais Arthur
Isaac-Chassande Valentin	4A	(c)	EEATS	Rousseau Frédéric - Durand Yves
Tomasi Ribeiro Eduardo	3A	(s)	MSTII	Pétrot Frédéric - Fuguet César Fabre Christian
Million Davy	3A	(c)	MSTII	Pétrot Frédéric - Balkind Jonathan Evans Adrian - Fuguet César
Romane Olivier	3A	(c)	MSTII	Pétrot Frédéric Muller Olivier Prost-Boucle Adrien
Yerly Louka	2A	(c)	MSTII	Pétrot Frédéric - Perais Arthur
Pham Van Quan	2A	(c)	MSTII	Pétrot Frédéric - Prost-Boucle Adrien Muller Olivier
McGovern Killian	2A	(c)	EEATS	Rousseau Frédéric - Charles Henri-Pierre
Vasilevska Nevena	2A	(c)	IP Paris	Thomas Gaël - Dumas Julie Derumigny Nicolas
Dubois Jules	1A	(c)	EEATS	Rousseau Frédéric - Evans Adrian Guthmuller Eric
Söderström Johan	1A	(c)	MSTII	Pétrot Frédéric - Dumas Julie - Perais Arthur
Meebed Abdallah	1A	(c)	MSTII	Pétrot Frédéric - Muller Olivier
Ilin Andrei	1A	(c)	MSTII	Pétrot Frédéric - Fuguet César
Sartori Dorian	1A	(c)	EEATS	Rousseau Frédéric - Fuguet César

(s) means PhD defended during the period, (c) means currently pursuing the PhD

11.2.3 Juries

- Frédéric Pétrot, reviewer of Aurélie Saulquin PhD thesis, Cristal, Lille, France, November 2025.

12 Scientific production

12.1 Major publications

- [1] C. S. Deshpande, A. Perais and F. Pétrot. ‘Address/Data Instruction Steering in Clustered General Purpose Processors’. In: *ACM Transactions on Architecture and Code Optimization* 22.3 (19th Sept. 2025), pp. 1–24. DOI: [10.1145/3744908](https://doi.org/10.1145/3744908). URL: <https://hal.science/hal-05215160>.
- [2] D. Million, C. Fuguet, A. Evans, R. El Cheikh, A. Monemi, J. Balkind and F. Pétrot. ‘Depth-first: A deterministic and scalable NoC routing protocol for 3.5D packaged architectures’. In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* (17th July 2025). DOI: [10.1109/JETCAS.2025.3590106](https://doi.org/10.1109/JETCAS.2025.3590106). URL: <https://hal.science/hal-05190367>.

12.2 Publications of the year

International journals

- [3] C. S. Deshpande, A. Perais and F. Pétrot. ‘Address/Data Instruction Steering in Clustered General Purpose Processors’. In: *ACM Transactions on Architecture and Code Optimization* 22.3 (19th Sept. 2025), pp. 1–24. DOI: [10.1145/3744908](https://doi.org/10.1145/3744908). URL: <https://hal.science/hal-05215160> (cit. on p. 11).

- [4] D. Million, C. Fuguet, A. Evans, R. El Cheikh, A. Monemi, J. Balkind and F. Pétrot. ‘Depth-first: A deterministic and scalable NoC routing protocol for 3.5D packaged architectures’. In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* (17th July 2025). doi: [10.1109/JETCAS.2025.3590106](https://doi.org/10.1109/JETCAS.2025.3590106). URL: <https://hal.science/hal-05190367> (cit. on p. 8).

International peer-reviewed conferences

- [5] Z. Fu, R. Tedeschi, G. Ottavi, N. Wistoff, C. Fuguet, D. Rossi and L. Benini. ‘Ramping Up Open-Source RISC-V Cores: Assessing the Energy Efficiency of Superscalar, Out-of-Order Execution’. In: *CF ’25: Proceedings of the 22nd ACM International Conference on Computing Frontiers*. CF 2025 - 22nd ACM International Conference on Computing Frontiers. Cagliari, Sardinia, Italy: ACM, July 2025, pp. 12–20. doi: [10.1145/3719276.3725186](https://doi.org/10.1145/3719276.3725186). URL: <https://hal.science/hal-05213064> (cit. on pp. 6, 10).
- [6] G. Korkian, N. Leyva, A. Bigas, N. Oliete-Escuín, A. Haghi, A. Monemi, C. Fuguet and L. Alvarez. ‘FetchFlare: An Open-Source Strided Data Prefetcher for High-Performance Cache Hierarchies’. In: *2025 28th Euromicro Conference on Digital System Design (DSD)*. 2025 28th Euromicro Conference on Digital System Design (DSD). Salerno, Italy: IEEE, 9th Dec. 2025, pp. 276–284. doi: [10.1109/DSD67783.2025.00047](https://doi.org/10.1109/DSD67783.2025.00047). URL: <https://hal.science/hal-05452261> (cit. on pp. 5, 11).
- [7] F. Pétrot and C. Fuguet. ‘On the Hardware Implementation of Lala’s 64-bit SEcDED Codes’. In: *2025 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*. 2025 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT). Barcelona, Spain: IEEE, 25th Nov. 2025, pp. 1–4. doi: [10.1109/DFT66274.2025.11257516](https://doi.org/10.1109/DFT66274.2025.11257516). URL: <https://hal.science/hal-05452289> (cit. on p. 9).
- [8] E. Tomasi, C. Fuguet, C. Fabre and F. Pétrot. ‘HPC Workload Analysis Using Distributed Cross-ISA Binary Instrumentation’. In: *40th Conference on Design of Circuits and Integrated Systems*. 2025 40th Conference on Design of Circuits and Integrated Systems (DCIS). Santander, Spain: IEEE, 11th Dec. 2025, pp. 144–149. doi: [10.1109/DCIS67520.2025.11281913](https://doi.org/10.1109/DCIS67520.2025.11281913). URL: <https://hal.science/hal-05452524> (cit. on pp. 5, 9).
- [9] E. Tomasi Ribeiro, C. Fuguet, C. Fabre and F. Pétrot. ‘Hardware-software co-design for supporting shared distributed virtual memory’. In: *CF ’25 Companion: Proceedings of the 22nd ACM International Conference on Computing Frontiers: Workshops and Special Sessions*. CF 2025 - 22nd ACM International Conference on Computing Frontiers: Workshops and Special Sessions. Cagliari, Italy: ACM (Association for Computing Machinery), 2025, pp. 58–61. doi: [10.1145/3706594.3726976](https://doi.org/10.1145/3706594.3726976). URL: <https://hal.science/hal-05163673> (cit. on p. 8).

Reports & preprints

- [10] I. Constantinou, A. Perais and Y. Sazeides. *The Non-Predictability of Mispredicted Branches using Timing Information*. University of Cyprus, 21st Jan. 2026. URL: <https://hal.science/hal-05482302> (cit. on p. 12).

Other scientific publications

- [11] Z. Assoumani, C. Fuguet, R. Mateescu and W. Serwe. ‘On Benefits of Modeling the HPDCache in LNT’. In: *RISC-V 2025 -RISC-V Summit Europe*. Paris, France, 2025, pp. 1–1. URL: <https://inria.hal.science/hal-05106202> (cit. on p. 9).
- [12] E. Guthmuller, C. Fuguet, A. Bocco, J. Fereyre, A. Evans, Y. Durand and J. Fereyre. ‘Variable and extended precision (VRP) accelerator implemented in a 22 nm SoC’. In: *Electronics Letters* 61 (2025), ell2.70255. doi: [10.1049/ell2.70255](https://doi.org/10.1049/ell2.70255). URL: <https://cea.hal.science/cea-05123539> (cit. on p. 8).

- [13] R. Tedeschi, G. Ottavi, C. Allart, N. Wistoff, Z. Fu, F. Grillotti, F. de Ambroggi, E. Guidetti, J.-B. Rigaud, O. Potin, J. R. Coulon, C. Fuguet, L. Benini and D. Rossi. ‘CVA6S+: A Superscalar RISC-V Core with High-Throughput Memory Architecture’. In: RISC-V Summit Europe 2025. Paris, France: arXiv, 2025. DOI: [10.48550/arXiv.2505.03762](https://doi.org/10.48550/arXiv.2505.03762). URL: <https://hal.science/hal-05213082> (cit. on p. 7).