4th national RISC-V student contest 2023-2024

Sponsored by Thales, the GDR SOC² and the CNFM

Accelerating AI applications on a RISC-V processor

You are students and like new challenges.
You are interested in electronics, computer architectures and embedded AI. You want to participate in the adventure of a renowned OpenHW processor core.
Then join this contest and win up to € 5,000!

Introduction

Thales is a world leader for mission critical information systems for the security, defense, space, aerospace and ground transportation domains. It employs about 80,000 people in 68 countries. Following the success of the three previous national RISC-V student contests from 2020 to 2023, Thales Research & Technology France, together with the GDR SOC² and the CNFM, are proud to announce the 4th edition of the contest (2023-2024).

This year's contest will focus on **architectural modifications** of the CV32A6 RISC-V soft-core to **accelerate** an AI application, namely the **MNIST** digit recognition. This application has been generated from the N2D2 tool from CEA List.

RISC-V is a recent open ISA that is gaining every day more attraction. From this ISA, ETH Zürich has designed a mid-range open-source application core named **ARIANE**. It has the capacity to execute rich operating systems and integrates an MMU function and several privilege levels. In 2019, the **OpenHW Group** was created with the ambition to design industrial-grade RISC-V processors. It has integrated the ARIANE core as its new 64-bit application core under the name CV64A6. Thales engineers have created a more compact 32-bit version from the original design, named **CV32A6**. CV32A6 and CV64A6 share the same source code and are together referred to as CVA6.

This year, the students are challenged to accelerate the MNIST algorithm with architectural evolutions of the CV32A6. Winning teams will propose a solution that needs the lowest number of cycles to recognize a given set of digits with MNIST algorithm.

Targeted participants

You are a team of students:

- 1 to 4 **Master 2 (M2) students** (or equivalent: final year in engineering schools)
- All registered in a **French engineering school or university**
- Coached by one or several **supervisors** (teachers, assistant professors, professors)
- A PhD student can additionally provide some coaching.

Although the contest does not primarily target M1 students, they are also accepted.

You have the following skills or will get some or all of them:

- Digital electronics
- HDL languages
- Digital simulation
- Computing architecture
- Embedded programming
- AI acceleration
- FPGA design and tools

The broad timeline of the contest should make it a good fit for your last-year academic project. You can start a few months after the official kick-off date or submit results weeks or months before the deadline, e.g. before you start an internship.

Description of the contest

Inputs

Thales will prepare a kit[1] composed of:

- A testbench to simulate the CV32A6 in its RTL form;
- The parameters of the CV32A6 core to consider;
- Scripts for synthesis of the baseline core;
- A BSP;
- A reference design to run the CV32A6 core on a FPGA development board;
- The MNIST application running in bare metal on the processor core (generated by the N2D2 tool from CEA List). .

The kit will be available on a Thales GitHub repository. The repository address will be communicated through the Discord platform (see next section).

You will get support from your university/school and coaching from your supervisor. The organizers (Thales, GDR SOC², CNFM) will remain in contact with the teams and the supervisors.

Communication and support

Supervisors will provide level 1 support to the team. Level 2 support will be addressed to the contest organizers and may only be exercised if the issue cannot be solved at level 1.

A Discord server will be created by the organizers to:

- Communicate announcements and practical information;
- Allow level 2 support;
- Host forums and discussions open to all participants.

[1] The kit will look like the one used in 2021-2022, with two main evolutions: CV32A6 will be updated to a newer version and power estimations based on gate level simulations will not be used.

The Discord server will log messages so that teams, which will start later, can recover past information. No mailing list will be maintained; all messages will go through Discord after the team is registered. The invitations to the Discord server will be sent when the registration is accepted.

The French language will be used on Discord and during the interim event (English can be used by foreign students). Results can be reported in French or English.

Prerequisites

Most parts of the CVA6 processor are written in **SystemVerilog**, a language widely adopted by the industry. You will have to use this language to modify the core.

The list of CAD tools and the FPGA board are unchanged w.r.t. the previous editions, so that schools/universities having already participated may have the pre-requisite for this contest.

The **FPGA development board** selected for this context is the Digilent Zybo Z7-20[2].

Because of the current tensions on supply chains, you should **anticipate as soon as possible the provisioning of the mandatory FPGA development environment** to avoid unnecessary delays. The list of necessary material is:

Reference	URL	Remark
Zybo Z7-20	https://store.digilentinc.com/zybo-z7-zynq-7000-arm-fpga-soc-development-board/	Zybo Z7-10 is too small for CVA6.
Pmod USBUART	https://store.digilentinc.com/pmod-usbuart-usb-to-uart-interface/	Used for the console output
JTAG-HS2 Programming Cable	https://store.digilentinc.com/jtag-hs2-programming-cable/	
Connectors	https://store.digilentinc.com/pmod-cable-kit-2x6-pin-and-2x6-pin-to-dual-6-pin-pmod-splitter-cable/	At least a 6-pin connector Pmod is necessary; other references may offer it.

Supervisors must **ensure that student teams will have this FPGA environment**. They can order the complete set from CNFM (contact: fpga@cnfm.fr).

The Siemens EDA (formerly Mentor Graphics) **Questa** digital simulator will be needed to test your solutions and will be supported by Thales[3]. Questa licenses can be obtained through CNFM. As the Questa licence tokens are hosted on a CNFM server, you will likely need support from your IT (e.g. configure a firewall, communicate IP addresses with CNFM…).

AMD Xilinx Vivado will be used for the synthesis and place & route to port CV32A6 to the FPGA board. Participants can use evaluation licenses or the AMD University Program.

The kit prepared by Thales will run on Linux Ubuntu 20.04. You can use other host OSes, but you will not receive IT support by the organizers.

CAD tools are available in schools/universities through CNFM. For any information, supervisors can get in touch with cao@cnfm.fr.

From previous years' experience, you should **anticipate the provisioning of CAD licences and FPGA board, and the setup of the IT infrastructure** to avoid delays.

Work to perform and constraints

Your goal is to modify the CV32A6 architecture to accelerate the recognition of a set of digits with the MNIST algorithm.

Your creativity should prevail as long as the following constraints are fulfilled:

- Your solution recognizes the set of digits with the same recognition probabilities on the FPGA board.
- The size of your hardware design fits in the FPGA.

[2] No other FPGA boards will be accepted as the jury will replay results on its board.
[3] Teams can use other simulators, e.g. Verilator, to develop their solution, but won't receive support on these tools.

- Your solution must mainly rely on the CV32A6 processor or in a tightly-coupled coprocessor attached to the CV-X-IF interface (not on a “coarse grain” accelerator on the AXI bus).
- You cannot increase the total memory size of data caches.
- You cannot increase the total memory size of instruction caches.
- You cannot decrease the CV32A6 operational frequency by more than 20% (as measured by Vivado timing analysis)
- You cannot alter the source code of the MNIST CNN application except to exploit your architectural evolutions, e.g. to express parallelism, to interface with a coprocessor, to address custom instructions… In such case, your modifications shall be clearly explained so that the jury can understand your code.
- Hardware-accelerated operations are logically equivalent to the original software: same algorithm, same data precision and types.
- To address RISC-V extensions or coprocessors, you can use built-in assembly language or adapt the compiler.
- Optimisations of the environment (compiler options, synthesis options, SoC peripherals…) will only be accepted if they relate to your architectural changes.
- You cannot replace the CV32A6 by another core.
- You can use existing open-source code, as long as you respect its licence terms and the compatibility with the Apache/Solderpad licence[4].

The organizers cannot anticipate all teams’ ideas. If you have a doubt about your solution eligibility, get in touch with the organizers. Discussions will never be shared with other teams.

Outputs

You shall provide:

- A 6-page report presented as a scientific paper, in French or English, with illustrations[5]
- A 10-minute recorded video, that will be presented during the defense session (and during the prize ceremony for the winning teams), in French or English
- A link to your new source code and results uploaded on GitHub (under Apache or Solderpad licenses)
- The reports from the tools that justify the results presented in the report.

During the course of the contest, the organizers will provide additional instructions or clarifications. A standardized way to report results, according to the defined criteria, will also be provided.

Contributions to the OpenHW Group

Thales may get back in touch with the students to promote some of their contributions to the OpenHW Group. The OpenHW Group criteria for accepting contributions may differ from those used in this contest.

Prizes and jury’s criteria

To be eligible for the final selection, your solution must simulate correctly, work on the FPGA board, provide the correct results (recognized digits with the same recognition probabilities) and fulfill the constraints listed above.

The jury will then **rank the results based on the number of cycles needed to recognize the set of digits** on the FPGA board.

If there are ties, i.e. teams needing a close number of cycles (<1%), the results will be separated according to the FPGA resources (number of LUTs + number of flip-flops) used by the solution.

The jury will replay the finalists’ results in its environment to avoid any differences due to local configurations.

The jury will consider specific situations (if not clearly defined in the rules) while maintaining the fairness between teams.

[4] This means for instance that you shall not include GPL or LGPL code.

[5] An IEEE template is recommended: https://www.ieee.org/conferences/publishing/templates.html.

Thales will award **€ 5,000** to the winning team and **€ 3,000** to the second best team (for teams between 2 and 4 participants). Prizes[6] will be equally shared among the team members. If an awarded team only has 1 participant, the individual will get € 2,500 for the first prize and € 1,500 for the second prize.

The winning teams will present their video at an event organized by Thales and/or the GDR SOC[2]. In addition to the prize, they will get a diploma.

Registration

Teams may register anytime until **1st February 2024** and run the contest at their own pace. They will recover past information from the Discord server.

Teams from a same university/school will register separately and can start at different dates.

Teams may deliver their results at any time before the end of the contest on **May 13th, 2024**. For instance, they can deliver their results before internships start.

To register, the supervisor will send an email to Jérôme Quévremont, Sébastien Pillement and Pascal Benoit (addresses below) with the following details:

- Name of the university/school
- Name of the team[7]
- Supervisor(s): Name, position (e.g. assistant professor), email address
- For each student: Name, option/major/"filière", Master 1 / Master 2, email address

The organizers will then check whether the registration is valid and send Discord time-limited invitations to the students and the supervisors.

A few additional rules:

- Individual candidates and teams without a supervisor are not accepted.
- If you are more than 4 students, register as several teams.
- There may be several teams from a given university/school.
- A team may include students from different options/majors/"filières".
- A student may not be in more than one team.
- A supervisor (and the PhD student if there is one) may supervise several teams and ensure no co-operation between these teams on their technical solutions. Joint teaching/learning (e.g. SystemVerilog, ISA, architecture courses) is of course accepted and encouraged.
- The supervisor(s) may be replaced during the project.
- Once the team is registered, the list of students shall not change (except under exceptional circumstances, with the organizers' agreement).
- The organizers will consider specific situations (if not clearly defined in the rules) while maintaining the fairness with other teams.

Abbreviations

AI	Artificial Intelligence
AXI	Advanced eXtensible Interface
ASIC	Application-Specific Integrated Circuit
BSP	Board Support Package
CAD	Computer-Aided Design
CNFM	Coordination Nationale pour la Formation en Micro-électronique et en nanotechnologies
CNN	Convolutional Neural Network

[6] The prizes will be provided as prepaid Mastercard cards shortly after the prize ceremony. Award-winning team members will have to provide personal details needed by the payment organization.

[7] It is important if there are several teams from a same university/school. If only one team registers, it can have the name of its university/school. The team name will be printed on the winners' diploma.

EDA	Electronic Design Automation
FPGA	Field-Programmable Gate Array
GDR	Groupe de Recherche
HDL	Hardware Description Language
IC	Integrated Circuit
IT	Information Technology
ISA	Instruction Set Architecture
LUT	Look-Up Table
M1/M2	Master 1 / Master 2
MMU	Memory Management Unit
MNIST	Modified National Institute of Standards and Technology database
RISC	Reduction Instruction Set Computer
RTL	Register Transfer Level
SoC	System on Chip
TBA	To be announced

Planning

	Thales	GDR SOC²	CNFM	Teams and universities/schools
Back to school 2023	Launch of the contest			You can register teams as soon as you read this document. You can anticipate the contest by gaining knowledge on the CV32A6, SystemVerilog, the MNIST algorithm, neural networks… **You should anticipate the provision of tools and the FPGA board (see above).** No support from the organizers is offered until the kit is delivered.
First half of November 2023	Delivery of the kit as a GitHub repo.			The teams can start later than this date.
November to May	Level 2 support (fix bugs in the kits…)			Run the project. Supervisors provide level 1 support to the student team.
January 10th, 2024 14:00	Organization of an online event: teams' introduction, guidelines to prepare submission, Q&A…			Participation to the event
February 1st, 2024				Deadline to register teams
May 13th, 2024 at 23:59				**Deadline to submit results in a GitHub repository (reports, source code…)**. The teams can submit results sooner. The defense videos will have to be provided about one week after.
End of May 2024 (TBA)	Organize the online defense session as online meeting(s)			Attend the online defense session: recorded video (10') + live Q&A (5')
June 10th-12th, 2024	Final event at Colloque GDR SOC² in Toulouse: prize announcement, presentation by the winning teams (recorded videos).			You are welcome to receive the diploma and congratulations.

Contacts

Thales Research and Technology	Jérôme Quévremont	jerome.quevremont@thalesgroup.com
GDR SOC²	Sébastien Pillement	sebastien.pillement@univ-nantes.fr
CNFM	Pascal Benoit	pascal.benoit@lirmm.fr