OPTIMA project revolutionizes HPC for industrial applications: interview

The EuroHPC JU-funded project OPTIMA is reshaping HPC in Europe with field-programmable gate array (FPGA) technologies.

OPTIMA, a EuroHPC JU (Joint Undertaking)-funded project launched in 2021, has developed an FPGA-based chip which allows the optimization and transfer of industrial applications on HPC (high performance computing) systems. FPGAs (field programmable gate arrays) are programmable computer chips that boost the performance of industrial applications, minimize energy consumption, and will help propel Europe towards global leadership in supercomputing.

The OPTIMA project is an initiative driven by a consortium of small and medium-sized enterprises (SMEs) and is the result of the teamwork of ten partners, six of which are SMEs from six different countries: Greece, Germany, Spain, Italy, Switzerland and the Netherlands. The project had a budget of €4,100,000 from Horizon Europe, the EU's funding programme for research and innovation, and completed its work in November 2023.

The EuroHPC JU interviewed Iakovos Mavroidis, researcher at the Technical University of Crete (TUC) and OPTIMA's project coordinator. He explains the key features of the project and explains what has been accomplished so far.

Can you please describe the OPTIMA project in your own words?

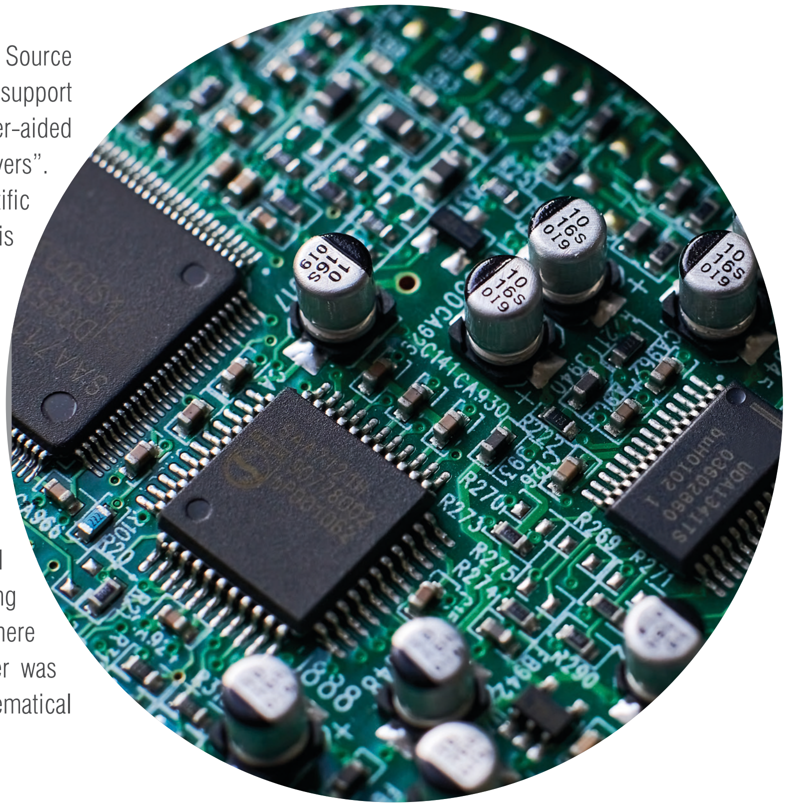
Sure! The primary objectives of the OPTIMA project were to optimize and test industrial applications and open-source libraries on HPC systems using FPGA chips.

These special chip technologies known as field-programmable gate arrays (FPGAs) function as high-performance engines and accelerators for specific applications needed to run a supercomputer. FPGA chip technologies are known for their lower power consumption compared to other chips like server-class CPUs (computer processing units) and GPUs (graphical processing units). FPGAs reduce power consumption in various applications for supercomputing and could foster new, more economic and environmentally friendly, approaches to HPC.

To accomplish this task, the OPTIMA project used high-performance servers based on JUMAX and Alveo technology equipped with FPGA chips to test different types of simulation in the fields of robotics, geosciences and computational fluid dynamics (CFD).

What was achieved?

A major achievement is the development of the OPTIMA Open Source (OOPS) library, which includes 31 hardware components to support fundamental linear algebraic operations and CAE (computer-aided engineering) problem-solving methods, also known as “solvers”. The OOPS library not only enhances raw performance for scientific algorithms, but also promotes high energy efficiency. This is achieved through the optimization of basic mathematical operations, known as BLAS kernels, which are fundamental components in computational mathematics. These result in potential energy savings of up to 50 times per BLAS kernel.



In addition, OPTIMA demonstrated remarkable results in tests on hardware prototypes: it doubled the speed of the preconditioned conjugate gradient (PCG) algorithm, which is a widely used method for solving complex mathematical equations. Tests with Robotic simulation software (involving convolutional Neural Networks) for autonomous driving where 3.4 times faster on average, CFD Lattice Boltzmann solver was 3.4 times faster and seven times faster than a basic mathematical operation on standard HPC software.

Can you give some examples of how OPTIMA supports European HPC users and how it promotes greener and more sustainable supercomputing?

Certainly: to support European HPC users, OPTIMA has made the OOPS library openly available. This allows developers to port legacy applications and code to FPGA-supported HPC systems without restrictions. We encourage users to download the template project of the OOPS library and try any of the available kernels/components. The library is accessible to anyone on GitHub [1, 2, 3].

As far as sustainable supercomputing is concerned, I am proud to say that the OOPS library enables significant performance improvements and reduced power consumption, which aligns perfectly with EuroHPC’s goals for energy-efficient supercomputing!

What were the main challenges you encountered during the project's development, if any?

We encountered several challenges while developing use cases on OPTIMA FPGA-based infrastructures. The most notable of these were the following:

- FPGA-based MPSoCs (multiprocessor systems-on-chip) offer flexibility, but managing resources efficiently poses a significant challenge. There are several hardware constraints related to the number of DDR (double data rate) memory controllers and distribution and FPGA chip technologies. We had to carefully balance functionality between these two technologies to ensure optimal performance without exceeding hardware constraints.
- While there have been significant advancements in FPGA-chip-related tools, there are still difficulties in developing efficient FPGA-based applications. Providing a simple interface for the programmer is still a challenge, and it remains easier to manually implement this design on an FPGA chip.
- In today’s FPGA-based infrastructures, communication between an FPGA-accelerated application and the host processor requires high-latency memory transfer via PCIe (peripheral component interconnect express), a high-speed serial computer expansion bus standard.
- As FPGA chip designs become increasingly complex, it becomes more challenging to ensure that all signals meet timing requirements.

Overall, we were able to overcome most of these challenges through careful planning, rigorous testing, and collaboration among team members with FPGA design, hardware, and software development experience.

How is the development of such a platform supporting the ambition of the EuroHPC JU to make Europe a world-leader in supercomputing?

By advancing HPC capabilities through FPGA optimization tools and chip technologies for industrial applications, OPTIMA contributes to Europe's competitiveness in the global HPC landscape. In this way, the OPTIMA project aligns with the ambition of the EuroHPC JU to position Europe as a global leader in HPC. OPTIMA is also a good example of an innovative SME-driven EuroHPC project.

What's next for OPTIMA?

The participating SMEs are already using the experience they gained from porting applications to OPTIMA's heterogeneous platforms. They are developing new applications for advanced cloud systems, AI accelerators, GPUs, and other chip designs.

In addition, the OOPS library will be continuously updated with additional support for FPGA chips and tools, new kernel implementations, and further improvements to existing ones. OPTIMA expects the OOPS library to become a valuable tool for software developers who want to map their applications onto FPGA-supported HPC platforms.

Furthermore, the advanced FPGA-based infrastructure developed by OPTIMA is currently being used both to reproduce complex hardware systems under design (SuDs) and to execute high-end AI-powered applications.

For more information,
please visit OPTIMA Project Website
optima-hpc.eu and eurohpc-ju.europa.eu/optima-project-revolutionises-hpc-industrial-applications-interview-2024-04-24_en

References

[1] github.com/cslab-ntua/OPTIMA-ICCS

[2] github.com/cyberbotics/optima

[3] github.com/GabrieleRolleri/optima_lbm

EnginSoft is proud to have been recognized as Key Innovator by the European Commission for its work in the OPTIMA project and for its innovative results achieved in a Custom HPC solution for CFD. We plan to provide custom CFD solvers for embedded controllers for edge computing to enable real-time simulation and analysis of complex fluid dynamics problems. In the energy sector, edge CFD can be used to optimize the performance of wind turbines and solar panels. This innovation is still under development and requires a further feasibility study and an internal prototyping phase before being introduced to the market for commercial exploitation, but improvements are already being transferred to and shared with customers in the automotive and aerospace markets. In addition, the sectors of smart, green and integrated transport, and of safe, clean and efficient energy could benefit from our achievements in the OPTIMA project.

We concentrated on the final CPU+FPGA hybrid version of the Lattice-Boltzmann (LBM) code, working in message passing interface (MPI) on several digital front ends (DFEs). To implement the new version, we started from the PALABOS library and made various modifications to the D2Q9 lattice. Beginning with pure CPU code, our team analysed the instrumented code and developed three different sub-versions:

- a CPU buffered code (exactly resembling the FPGA code);
- an FPGA code with vector buffering (one domain line per call); and
- a full domain buffer running on FPGA.

For the LBM-CFD method, our implementation of a D2Q9 lattice within multi-FPGA and MPI tasks increased performance 1.5—3 times (depending on the number of MPI tasks) compared to the Jumax supercomputing infrastructure, according to the final evaluation. We evaluated the performance of applications developed on a hybrid CPU+FPGA system (Jumax), focusing specifically on their optimization from a CPU-only system to the hybrid system.

The results showed that the hybrid CPU+FPGA application for medium-size models performed best, both in terms of speed and energy savings compared to the CPU-only version. Energy KPIs were evaluated based on the mean consumption of the boards/processors due to the lack of other information or monitoring systems.

We have confirmed the various exploitation paths already identified for FPGA-based HPC and are willing to invest time and effort in:

- (i) building on the knowledge and expertise gained from OPTIMA and other successful projects in order to expand our capabilities in extreme scale computing, seamless hardware-software integration, and the ability to tackle complex network challenges and to use these strengths to develop innovative, tailored solutions that meet the specific needs of our clients;
- (ii) evaluating and recommending energy-efficient solutions to the heterogeneous private cloud systems market, especially in simulation engineering, where FPGA accelerators are becoming increasingly integrated, and particularly as the use of FPGAs becomes more prevalent in fluid dynamic applications due to their ability to perform high-speed calculations and data processing;
- (iii) assessing and proposing embedded solutions for customized devices for the automotive and aerospace market, in order to perform reduced order modelling techniques.

By applying the principles and techniques learnt from our past successes, we can confidently take on new and exciting projects that require state-of-the-art problem-solving strategies.

For more information:
Marisa Zanotti - EnginSoft
m.zanotti@enginsoft.com