2018-00469 – Automatic parallelization of sparse matrix operations - Post-Doctorant Inria Grenoble Research center

Contract type: Public service fixed-term contract
Level of qualifications required: PhD or equivalent
Fonction: Post-Doctoral Research Visit

About the research centre or Inria department

Grenoble Rhône-Alpes Research Center groups together a few less than 800 people in 35 research teams and 9 research support departments.

Staff is localized on 5 campuses in Grenoble and Lyon, in close collaboration with labs, research and higher education institutions in Grenoble and Lyon, but also with the economic players in these areas.

Present in the fields of software, high-performance computing, Internet of things, image and data, but also simulation in oceanography and biology, it participates at the best level of international scientific achievements and collaborations in both Europe and the rest of the world.

Context

The CASH (Compilation and Analysis, Software and Hardware) group works on compilation techniques for high-performance computing. We are currently a team at the LIP laboratory (Lyon), and a subgroup of the ROMA team at Inria.

The overall objective of the CASH team is to take advantage of the characteristics of the specific hardware (generic hardware, hardware accelerators or FPGA) to compile energy efficient software and hardware. The long-term objective is to provide solutions for the end-user developers to use at their best the huge opportunities of these emerging platforms. The research directions of the team are:

* Dataflow models for HPC applications: We target representations that are expressive enough to express all kinds of parallelism and allow further optimizations.

* Compiler algorithms and tools for irregular applications: The extensions of these intermediate representations to enable complex control flow and complex data structures, and the design of associated analysis for optimized code generation for multicore processors and accelerators.

* Compiler Algorithms, Simulation and Tools for Reconfigurable Circuits: The application of the two preceding activities on High Level Synthesis, with additional resource constraints.

* Simulation of Systems on a Chip: A parallel and scalable simulation of Systems-on-Chips, which, combined with the preceding activity, will result in a complete workflow for circuit design.

Assignment

Post-doc proposal:

The polyhedral model [1] allows to represent a set of operations on arrays together with their dependencies. This representation can be used to schedule operations (for example, change the execution order of a set of nested loops), or to extract parallelism from a sequential program. It has successfully been applied to many matrix-based compute-kernels for aggressive compiler optimization and synthesis of efficient hardware circuits.

A strong limitation, however, is that the model is limited to regular applications: it supports for loops and array accesses, but loop bounds and array indices must be expressible as affine functions of variables and parameters of the compute kernel.

The newly created CASH team works on novel methods to compile and analyze programs and targets in particular hardware generation. One of the research axis is to extend the polyhedral model to irregular applications, e.g. while loops and data-structures other than arrays.

Main activities

2018-00469 – Automatic parallelization of sparse matrix operations - Post-Doctorant Inria Grenoble Research center

Contract type: Public service fixed-term contract
Level of qualifications required: PhD or equivalent
Fonction: Post-Doctoral Research Visit

About the research centre or Inria department

Grenoble Rhône-Alpes Research Center groups together a few less than 800 people in 35 research teams and 9 research support departments.

Staff is localized on 5 campuses in Grenoble and Lyon, in close collaboration with labs, research and higher education institutions in Grenoble and Lyon, but also with the economic players in these areas.

Present in the fields of software, high-performance computing, Internet of things, image and data, but also simulation in oceanography and biology, it participates at the best level of international scientific achievements and collaborations in both Europe and the rest of the world.

Context

The CASH (Compilation and Analysis, Software and Hardware) group works on compilation techniques for high-performance computing. We are currently a team at the LIP laboratory (Lyon), and a subgroup of the ROMA team at Inria.

The overall objective of the CASH team is to take advantage of the characteristics of the specific hardware (generic hardware, hardware accelerators or FPGA) to compile energy efficient software and hardware. The long-term objective is to provide solutions for the end-user developers to use at their best the huge opportunities of these emerging platforms. The research directions of the team are:

* Dataflow models for HPC applications: We target representations that are expressive enough to express all kinds of parallelism and allow further optimizations.

* Compiler algorithms and tools for irregular applications: The extensions of these intermediate representations to enable complex control flow and complex data structures, and the design of associated analysis for optimized code generation for multicore processors and accelerators.

* Compiler Algorithms, Simulation and Tools for Reconfigurable Circuits: The application of the two preceding activities on High Level Synthesis, with additional resource constraints.

* Simulation of Systems on a Chip: A parallel and scalable simulation of Systems-on-Chips, which, combined with the preceding activity, will result in a complete workflow for circuit design.

Assignment

Post-doc proposal:

The polyhedral model [1] allows to represent a set of operations on arrays together with their dependencies. This representation can be used to schedule operations (for example, change the execution order of a set of nested loops), or to extract parallelism from a sequential program. It has successfully been applied to many matrix-based compute-kernels for aggressive compiler optimization and synthesis of efficient hardware circuits.

A strong limitation, however, is that the model is limited to regular applications: it supports for loops and array accesses, but loop bounds and array indices must be expressible as affine functions of variables and parameters of the compute kernel.

The newly created CASH team works on novel methods to compile and analyze programs and targets in particular hardware generation. One of the research axis is to extend the polyhedral model to irregular applications, e.g. while loops and data-structures other than arrays.
Objectives of the post-doc:

The goal of the post-doc is to study the applicability of polyhedral methods to irregular applications. As a starting point, we will focus on sparse matrices (large matrices containing mostly 0 values, using data-structures that represent only non-zero values).

How does the polyhedral model extend to such applications? In particular how do the notions of iteration domain and affine schedule translate? Which static/dynamic analysis are required? How do the dynamic analysis translate to source code generation?

The goal is to generate process networks corresponding to the input program. This work will be the first building block towards the compilation of hardware for FPGA accelerators. Even though hardware generation itself is out of the scope of the post-doc, the source code generated is expected to fit the input of a C-to-FPGA compiler and to be tested on a FPGA.

Indicative outline:

Here is a more precise list of tasks that could be carried out.

* If needed: study of existing polyhedral techniques for automatic parallelization [1] and related work on parallelization techniques for sparse matrices (for example [2]).

* Study how the shape of sparse matrices [3] can drive the parallelization of simple regular kernels such as matrix sum and matrix multiplication. In these examples, the dependencies between computations are very simple hence polyhedral scheduling and tiling is easier to figure out with sparse matrices.

* Study more complex regular kernels from the Polybench/C suite [4] and propose a parallelization algorithm.

References:

http://dblp.uni-trier.de/rec/bib/reference/parallel/FeautrierL11


http://web.cs.ucla.edu/~pouchet/software/polybench/

Skills

The candidate should have good background in compilation (knowledge of polyhedral methods would obviously be appreciated), and should be familiar with parallel algorithms.

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

Remuneration

Gross salary: 2650 Euros per month

Defence Security:

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy:

As part of its diversity policy, all Inria positions are accessible to people with disabilities.

Warning: you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.