IN PARTNERSHIP WITH:
CNRS

Institut polytechnique de Grenoble

Université Pierre Mendes-France (Grenoble)

Université Joseph Fourier (Grenoble)

Activity Report 2013

Project-Team MOAIS

PrograMming and scheduling design fOr Applications in Interactive Simulation

IN COLLABORATION WITH: Laboratoire d'Informatique de Grenoble (LIG)

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Distributed and High Performance Computing

Table of contents

Project-Team MOAIS

Keywords: Scheduling, Interactive Computing, Parallel And Distributed Algorithms, High Performance Computing, Fault Tolerance, Parallel Programming Model

Creation of the Project-Team: 2006 January 01.

1. Members

Research Scientists
- Thierry Gautier [Inria, Researcher]
- Bruno Raffin [Inria, Researcher, HdR]

Faculty Members
- Jean-Louis Roch [Team leader, Grenoble-INP, Associate Professor]
- François Broquedis [Grenoble-INP, Associate Professor]
- Vincent Danjean [Univ. Grenoble I, Associate Professor]
- Pierre-François Dutot [Univ. Grenoble II, Associate Professor]
- Guillaume Huard [Univ. Grenoble I, Associate Professor]
- Grégory Mounié [Grenoble-INP, Associate Professor]
- Clément Pernet [Univ. Grenoble I, Associate Professor, until Aug 2013]
- Denis Trystram [Grenoble-INP, Professor, HdR]
- Frédéric Wagner [Grenoble-INP, Associate Professor]

Engineers
- Pierrick Brunet [Inria, IJD, from Oct 2013]
- Ludovic Jacquin [[Inria, from Sep 2013 until Dec 2013]
- Jérémy Jaussaud [Inria]
- Philippe Virouleau [Inria, IJD]
- Christian Séguy [CNRS/LIG Engineer, 40%]
- Pierre Neyron [CNRS/LIG, Research engineer, 40%]
- Xavier Martin [Inria, Apprentice ENSIMAG]

PhD Students
- David Beniamine [Univ. Grenoble I]
- Stéfano Drimon Kurz Mor [joint Univ. Grenoble and UFRGS, Brazil, until Sep 2014]
- Damien Dosimont [Inria, granted by Rhône-Alpes Region]
- Matthieu Dreher [Inria, granted by ANR EXAVIZ project]
- Marie Durand [Inria, until Nov 2013]
- Mathias Ettinger [Inria, granted by EDF R&D]
- Joao Ferreira Lima [joint Univ. Grenoble and UFRGS, Brazil, until Sep 2014]
- Abhinav Srivastav [Univ Grenoble, LabEx Persyval with VERIMAG, from Jan 2013]
- David Glesser [Lig, granted by Bull CIFRE, from Jan 2013]
- Raphaël Bleuse [CNRS, from Feb 2013]
- Amrit Kumar [Univ. Grenoble, LabEx Persyval with VERIMAG and Privatics, from May 2013]
- Fernando Machado Mendonca [COFECUB, Brazil, from Aug 2013]
- Florence Monna [CNRS]
- Vinicius Pinheiro [COFECUB, Brazil]
- Marwa Sridi [CEA, from Apr 2013]
- Ziad Sultan [Univ. Grenoble I]

Post-Doctoral Fellows
- Alexandre Ancel [Inria, granted by CEA, until Apr 2013]

Joachim Lepping [Inria, until Jan 2013]

Visiting Scientists

Jacek Blazewicz [Polish Academy of Sciences, Poznań, funding by INP, until Feb 2013]
Afredo Goldman [USP visitor, July 2013]
Daniel Cordeiro [USP visitor, Dec 2013]
Matthias Diener [UFRGS visitor, Dec 2013]
Mario Cesar Lopez Loces [Inria, Dec 2013]
Adel Safi [Teaching assistant Institut Supérieur d'Informatique et de Gestion, Kairouan, Tunisia]

Administrative Assistants

Annie Simon [Inria, 50%]
Christine Guimet [CNRS, 30% from Apr 2013]

Others

Élie Attigui [Grenoble-INP, student internship, from Feb 2013 until Jun 2013]
Nathanaël Cheriere [ENS Cachan, student internship, from May 2013 until Jul 2013]
Sebastien Curt [Inria, student internship, from Feb 2013 until Aug 2013]
Guillerme Duvillie [Inria, student internship, from Mar 2013 until Jun 2013]
Birahima Fall [Grenoble-INP, student internship, from Mar 2013 until Jul 2013]
Hugo Guiroux [Inria, student internship, from Jun 2013 until Sep 2013]
Neha Nayak [Grenoble-INP, student internship, from Mar 2013 until Jun 2013]
Iullia Petrova [Univ. Grenoble I, student internship, from Feb 2013 until Jun 2013]
Valentin Reis [Grenoble-INP, student internship, until Mar 2013]
Gaëtan Steininger [Inria, student internship, from Feb 2013 until Jun 2013]
Armand Vassal [Inria, student internship, from Feb 2013 until Jun 2013]
Alexis Martin [Inria, student internship, from Feb 2013 until Jul 2013]

2. Overall Objectives

2.1. Introduction

The goal of the MOAIS team-project is to develop the scientific and technological foundations for parallel programming that enable to achieve provable performances on distributed parallel architectures, from multi-processor systems on chips to computational grids and global computing platforms. Beyond the optimization of the application itself, the effective use of a larger number of resources is expected to enhance the performance. This encompasses large scale scientific interactive simulations (such as immersive virtual reality) that involve various resources: input (sensors, cameras, ...), computing units (processors, memory), output (videoprojectors, images wall) that play a prominent role in the development of high performance parallel computing.

To reach this goal, MOAIS gathers experts in : algorithm design, scheduling, parallel programming (both low level and high level API), interactive applications. The research directions of the MOAIS team are focused on scheduling problems with a multi-criteria performance objective: precision, reactivity, resources consumption, reliability, ... The originality of the MOAIS approach is to use the application's adaptability to control its scheduling:

- the application describes synchronization conditions;
- the scheduler computes a schedule that verifies those conditions on the available resources;
- each resource behaves independently and performs the decision of the scheduler.

To enable the scheduler to drive the execution, the application is modeled by a macro data flow graph, a popular bridging model for parallel programming (BSP, Nesl, Earth, Jade, Cilk, Athapascan, Smarts, Satin, ...) and scheduling. A node represents the state transition of a given component; edges represent synchronizations between components. However, the application is malleable and this macro data flow is dynamic and recursive: depending on the available resources and/or the required precision, it may be unrolled to increase precision (e.g. zooming on parts of simulation) or enrolled to increase reactivity (e.g. respecting latency constraints). The decision of unrolling/enrolling is taken by the scheduler; the execution of this decision is performed by the application.

The MOAIS project-team is structured around four axis:

- **Scheduling**: To formalize and study the related scheduling problems, the critical points are: the modeling of an adaptive application; the formalization and the optimization of the multi-objective problems; the design of scalable scheduling algorithms. We are interested in classical combinatorial optimization methods (approximation algorithms, theoretical bounds and complexity analysis), and also in non-standard methods such as Game Theory.
- **Adaptive parallel and distributed algorithms**: To design and analyze algorithms that may adapt their execution under the control of the scheduling, the critical point is that algorithms are either parallel or distributed; then, adaptation should be performed locally while ensuring the coherency of results.
- **Programming interfaces and tools for coordination and execution**: To specify and implement interfaces that express coupling of components with various synchronization constraints, the critical point is to enable an efficient control of the coupling while ensuring coherency. We develop the **Kaapi** runtime software that manages the scheduling of multithreaded computations with billions of threads on a virtual architecture with an arbitrary number of resources; Kaapi supports node additions and resilience. Kaapi manages the *fine grain* scheduling of the computation part of the application. To enable parallel application execution and analysis. We develop runtime tools that support large scale and fault tolerant processes deployment (**TakTuk**), visualization of parallel executions on heterogeneous platforms (**Triva**).
- **Interactivity**: To improve interactivity, the critical point is scalability. The number of resources (including input and output devices) should be adapted without modification of the application. We develop the **FlowVR** middleware that enables to configure an application on a cluster with a fixed set of input and output resources. FlowVR manages the *coarse grain* scheduling of the whole application and the latency to produce outputs from the inputs.

Often, computing platforms have a dynamic behavior. The dataflow model of computation directly enables to take into account addition of resources. To deal with resilience, we develop softwares that provide **fault-tolerance** to dataflow computations. We distinguish non-malicious faults from malicious intrusions. Our approach is based on a checkpoint of the dataflow with bounded and amortized overhead.

2.2. Highlights of the Year

- Best Paper - HeteroPar 2013
- Best Long Paper - Second Prize at Web3D 2013

BEST PAPERS AWARDS :

[23] **Scheduling Independent Tasks on Multi-Cores with GPU Accelerators in Proceedings of the 11th HeteroPar workshop (Algorithms, Models and Tools for Parallel Computing and Heterogeneous Platforms)**. S. KEDAD-SIDHOUM, F. MONNA, G. MOUNIÉ, D. TRYSTRAM.

[17] **VCoRE: a web resource oriented architecture for efficient data exchange in 18th International Conference on 3D Web**. T. FRANKE, V. SETTGAS, J. BEHR, B. RAFFIN.

3. Research Program

3.1. Scheduling

Participants: Pierre-François Dutot, Guillaume Huard, Grégory Mounié, Jean-Louis Roch, Denis Trystram, Frédéric Wagner.

The goal of this theme is to determine adequate multi-criteria objectives which are efficient (precision, reactivity, speed) and to study scheduling algorithms to reach these objectives.

In the context of parallel and distributed processing, the term *scheduling* is used with many acceptations. In general, scheduling means assigning tasks of a program (or processes) to the various components of a system (processors, communication links).

Researchers within MOAIS have been working on this subject for many years. They are known for their multiple contributions for determining the target dates and processors the tasks of a parallel program should be executed; especially regarding execution models (taking into account inter-task communications or any other system features) and the design of efficient algorithms (for which there exists a performance guarantee relative to the optimal scheduling).

Parallel tasks model and extensions. We have contributed to the definition and promotion of modern task models: parallel moldable tasks and divisible load. For both models, we have developed new techniques to derive efficient scheduling algorithms (with a good performance guaranty). We proposed recently some extensions taking into account machine unavailabilities (reservations).

Multi-objective Optimization. A natural question while designing practical scheduling algorithms is "which criterion should be optimized ?". Most existing works have been developed for minimizing the *makespan* (time of the latest tasks to be executed). This objective corresponds to a system administrator view who wants to be able to complete all the waiting jobs as soon as possible. The user, from his-her point of view, would be more interested in minimizing the average of the completion times (called *minsum*) of the whole set of submitted jobs. There exist several other objectives which may be pertinent for specific use. We worked on the problem of designing scheduling algorithms that optimize simultaneously several objectives with a theoretical guarantee on each objective. The main issue is that most of the policies are good for one criterion but bad for another one.
We have proposed an algorithm that is guaranteed for both *makespan* and *minsum*. This algorithm has been implemented for managing the resources of a cluster of the regional grid CIMENT. More recently, we extended such analysis to other objectives (makespan and reliability). We concentrate now on finding good algorithms able to schedule a set of jobs with a large variety of objectives simultaneously. For hard problems, we propose approximation of Pareto curves (best compromizes).

Incertainties. Most of the new execution supports are characterized by a higher complexity in predicting the parameters (high versatility in desktop grids, machine crash, communication congestion, cache effects, etc.). We studied some time ago the impact of incertainties on the scheduling algorithms. There are several ways for dealing with this problem: First, it is possible to design robust algorithms that can optimized a problem over a set of scenarii, another solution is to design flexible algorithms. Finally, we promote semi on-line approaches that start from an optimized off-line solution computed on an initial data set and updated during the execution on the "perturbed" data (stability analysis).

Game Theory. Game Theory is a framework that can be used for obtaining good solution of both previous problems (multi-objective optimization and incertain data). On the first hand, it can be used as a complement of multi-objective analysis. On the other hand, it can take into account the incertainties. We are curently working at formalizing the concept of cooperation.

Scheduling for optimizing parallel time and memory space. It is well known that parallel time and memory space are two antagonists criteria. However, for many scientific computations, the use of parallel architectures is motivated by increasing both the computation power and the memory space. Also, scheduling for optimizing both parallel time and memory space targets an important multicriteria objective. Based on the analysis of the dataflow related to the execution, we have proposed a scheduling algorithm with provable performance.

Coarse-grain scheduling of fine grain multithreaded computations on heterogeneous platforms. Designing multi-objective scheduling algorithms is a transversal problem. Work-stealing scheduling is well studied for fine grain multithreaded computations with a small critical time: the speed-up is asymptotically optimal. However, since the number of tasks to manage is huge, the control of the scheduling is expensive. We proposed a generalized lock-free cactus stack execution mechanism, to extend previous results, mainly from Cilk, based on the *work-first principle* for strict multi-threaded computations on SMPs to general multithreaded computations with dataflow dependencies. The main result is that optimizing the sequential local executions of tasks enables to amortize the overhead of scheduling. This distributed work-stealing scheduling algorithm has been implemented in **Kaapi**.

3.2. Adaptive Parallel and Distributed Algorithms Design

Participants: François Broquedis, Pierre-François Dutot, Thierry Gautier, Guillaume Huard, Bruno Raffin, Jean-Louis Roch, Denis Trystram, Frédéric Wagner.

This theme deals with the analysis and the design of algorithmic schemes that control (statically or dynamically) the grain of interactive applications.

The classical approach consists in setting in advance the number of processors for an application, the execution being limited to the use of these processors. This approach is restricted to a constant number of identical resources and for regular computations. To deal with irregularity (data and/or computations on the one hand; heterogeneous and/or dynamical resources on the other hand), an alternate approach consists in adapting the potential parallelism degree to the one suited to the resources. Two cases are distinguished:

- in the classical bottom-up approach, the application provides fine grain tasks; then those tasks are clustered to obtain a minimal parallel degree.
- the top-down approach (Cilk, Cilk+, TBB, Hood, Athapascan) is based on a work-stealing scheduling driven by idle resources. A local sequential depth-first execution of tasks is favored when recursive parallelism is available.

Ideally, a good parallel execution can be viewed as a flow of computations flowing through resources with no control overhead. To minimize control overhead, the application has to be adapted: a parallel algorithm on p resources is not efficient on $q < p$ resources. On one processor, the scheduler should execute a sequential algorithm instead of emulating a parallel one. Then, the scheduler should adapt to resource availability by changing its underlying algorithm. This first way of adapting granularity is implemented by Kaapi (default work-stealing schedule based on work-first principle).

However, this adaptation is restrictive. More generally, the algorithm should adapt itself at runtime to improve its performance by decreasing the overheads induced by parallelism, namely the arithmetic operations and communications. This motivates the development of new parallel algorithmic schemes that enable the scheduler to control the distribution between computation and communication (grain) in the application to find the good balance between parallelism and synchronizations. MOAIS has exhibited several techniques to manage adaptivity from an algorithmic point of view:

- amortization of the number of global synchronizations required in an iteration (for the evaluation of a stopping criterion);
- adaptive deployment of an application based on on-line discovery and performance measurements of communication links;
- generic recursive cascading of two kind of algorithms: a sequential one, to provide efficient executions on the local resource, and a parallel one that enables an idle resource to extract parallelism to dynamically suit the degree of parallelism to the available resources.

The generic underlying approach consists in finding a good mix of various algorithms, what is often called a "poly-algorithm". Particular instances of this approach are Atlas library (performance benchmark are used to decide at compile time the best block size and instruction interleaving for sequential matrix product) and FFTW library (at run time, the best recursive splitting of the FFT butterfly scheme is precomputed by dynamic programming). Both cases rely on pre-benchmarking of the algorithms. Our approach is more general in the sense that it also enables to tune the granularity at any time during execution. The objective is to develop processor oblivious algorithms: similarly to cache oblivious algorithms, we define a parallel algorithm as *processor-oblivious* if no program variable that depends on architecture parameters, such as the number or processors or their respective speeds, needs to be tuned to minimize the algorithm runtime.

We have applied this technique to develop processor oblivious algorithms for several applications with provable performance: iterated and prefix sum (partial sums) computations, stream computations (cipher and hd-video transformation), 3D image reconstruction (based on the concurrent usage of multi-core and GPU), loop computations with early termination.

By optimizing the work-stealing to our adaptive algorithm scheme, the non-blocking (wait-free) implementation of Kaapi has been designed and leads to the C library X-kaapi.

Extensions concern the development of algorithms that are both cache and processor oblivious on heterogeneous processors. The processor algorithms proposed for prefix sums and segmentation of an array are cache oblivious too.

3.3. Interactivity

Participants: Vincent Danjean, Pierre-François Dutot, Thierry Gautier, Bruno Raffin, Jean-Louis Roch.

The goal of this theme is to develop approaches to tackle interactivity in the context of large scale distributed applications.

We distinguish two types of interactions. A user can interact with an application having only little insight about the internal details of the program running. This is typically the case for a virtual reality application where the user just manipulates 3D objects. We have a "user-in-the-loop". In opposite, we have an "expert -in-the-loop" if the user is an expert that knows the limits of the progam that is being executed and that he can interacts with it to steer the execution. This is the case for instance when the user can change some parameters during the execution to improve the convergence of a computation.

3.3.1. User-in-the-loop

Some applications, like virtual reality applications, must comply with interactivity constraints. The user should be able to observe and interact with the application with an acceptable reaction delay. To reach this goal the user is often ready to accept a lower level of details. To execute such application on a distributed architecture requires to balance the workload and activation frequency of the different tasks. The goal is to optimize CPU and network resource use to get as close as possible to the reactivity/level of detail the user expect.

Virtual reality environments significantly improve the quality of the interaction by providing advanced interfaces. The display surface provided by multiple projectors in CAVE -like systems for instance, allows a high resolution rendering on a large surface. Stereoscopic visualization gives an information of depth. Sound and haptic systems (force feedback) can provide extra information in addition to visualized data. However driving such an environment requires an important computation power and raises difficult issues of synchronization to maintain the overall application coherent while guaranteeing a good latency, bandwidth (or refresh rate) and level of details. We define the coherency as the fact that the information provided to the different user senses at a given moment are related to the same simulated time.

Today’s availability of high performance commodity components including networks, CPUs as well as graphics or sound cards make it possible to build large clusters or grid environments providing the necessary resources to enlarge the class of applications that can aspire to an interactive execution. However the approaches usually used for mid size parallel machines are not adapted. Typically, there exist two different approaches to handle data exchange between the processes (or threads). The synchronous (or FIFO) approach ensures all messages sent are received in the order they were sent. In this case, a process cannot compute a new state if all incoming buffers do not store at least one message each. As a consequence, the application refresh rate is driven by the slowest process. This can be improved if the user knows the relative speed of each module and specify a read frequency on each of the incoming buffers. This approach ensures a strong coherency but impact on latency. This is the approach commonly used to ensure the global coherency of the images displayed in multi-projector environments.The other approach, the asynchronous one, comes from sampling systems. The producer updates data in a shared buffer asynchronously read by the consumer. Some updates may be lost if the consumer is slower than the producer. The process refresh rates are therefore totally independent. Latency is improved as produced data are consumed as soon as possible, but no coherency is ensured. This approach is commonly used when coupling haptic and visualization systems. A fine tuning of the application usually leads to satisfactory results where the user does not experience major incoherences. However, in both cases, increasing the number of computing nodes quickly makes infeasible hand tuning to keep coherency and good performance.

We propose to develop techniques to manage a distributed interactive application regarding the following criteria :

- latency (the application reactivity);
- refresh rate (the application continuity);
- coherency (between the different components);
- level of detail (the precision of computations).

We developed a programming environment, called FlowVR, that enables the expression and realization of loosen but controlled coherency policies between data flows. The goal is to give users the possibility to express a large variety of coherency policies from a strong coherency based on a synchronous approach to an uncontrolled coherency based on an asynchronous approach. It enables the user to loosen coherency where it is acceptable, to improve asynchronism and thus performance. This approach maximizes the refresh rate and minimizes the latency given the coherency policy and a fixed level of details. It still requires the user to tune many parameters. In a second step, we are planning to explore auto-adaptive techniques that enable to decrease the number of parameters that must be user tuned. The goal is to take into account (possibly dynamically) user specified high level parameters like target latencies, bandwidths and levels of details, and to have the system automatically adapt to reach a trade-off given the user wishes and the resources available. Issues include multi-criterion optimizations, adaptive algorithmic schemes, distributed decision making, global stability and balance of the regulation effort.

3.3.2. Expert-in-the-loop

Some applications can be interactively guided by an expert who may give advices or answer specific questions to hasten a problem resolution. A theoretical framework has been developed in the last decade to define precisely the complexity of a problem when interactions with an expert is allowed. We are studying these interactive proof systems and interactive complexity classes in order to define efficient interactive algorithms dedicated to scheduling problems. This, in particular, applies to load-balancing of interactive simulations when a user interaction can generate a sudden surge of imbalance which could be easily predicted by an operator.

3.4. Adaptive middleware for code coupling and data movements

Participants: François Broquedis, Vincent Danjean, Thierry Gautier, Clément Pernet, Bruno Raffin, Jean-Louis Roch, Frédéric Wagner.

This theme deals with the design and implementation of programming interfaces in order to achieve an efficient coupling of distributed components.

The implementation of interactive simulation application requires to assemble together various software components and to ensure a semantic on the displayed result. To take into account functional aspects of the computation (inputs, outputs) as well as non functional aspects (bandwidth, latency, persistence), elementary actions (method invocation, communication) have to be coordinated in order to meet some performance objective (precision, quality, fluidity, *etc*). In such a context the scheduling algorithm plays an important role to adapt the computational power of a cluster architecture to the dynamic behavior due to the interactivity. Whatever the scheduling algorithm is, it is fundamental to enable the control of the simulation. The purpose of this research theme is to specify the semantics of the operators that perform components assembling and to develop a prototype to experiment our proposals on real architectures and applications.

3.4.1. Application Programming Interface

The specification of an API to compose interactive simulation application requires to characterize the components and the interaction between components.The respect of causality between elementary events ensures, at the application level, that a reader will see the *last* write with respect to an order. Such a consistency should be defined at the level of the application to control the events ordered by a chain of causality. For instance, one of the result of Athapascan was to prove that a data flow consistency is more efficient than other ones because it generates fewer messages. Beyond causality based interactions, new models of interaction should be studied to capture non predictable events (delay of communication, capture of image) while ensuring a semantic.

Our methodology is based on the characterization of interactions required between components in the context of an interactive simulation application. For instance, criteria could be coherency of visualization, degree of interactivity. Beyond such characterization we hope to provide an operational semantic of interactions (at least well suited and understood by usage) and a cost model. Moreover they should be preserved by composition to predict the cost of an execution for part of the application.

The main result relies on a computable representation of the future of an execution; representations such as macro data flow are well suited because they explicit which data are required by a task. Such a representation can be built at runtime by an interpretation technique: the execution of a function call is differed by computing beforehand at runtime a graph of tasks that represents the (future) calls to execute.

3.4.2. Kernel for Asynchronous, Adaptive, Parallel and Interactive Application

Managing the complexity related to fine grain components and reaching high efficiency on a cluster architecture require to consider a dynamic behavior. Also, the runtime kernel is based on a representation of the execution: data flow graph with attributes for each node and efficient operators will be the basis for our software. This kernel has to be specialized for the considered applications. The low layer of the kernel has features to transfer data and to perform remote signalization efficiently. Well known techniques and legacy code have to be reused. For instance, multithreading, asynchronous invocation, overlapping of latency by computing, parallel communication and parallel algorithms for collective operations are fundamental techniques to reach performance. Because the choice of the scheduling algorithm depends on the application and the architecture, the kernel will provide an *causally connected representation* of the system that is running. This allows to specialize the computation of a good schedule of the data flow graph by providing algorithms (scheduling algorithms for instance) that compute on this (causally connected) representation: any modification of the representation is turned into a modification on the system (the parallel program under execution). Moreover, the kernel provides a set of basic operators to manipulate the graph (*e.g.* computes a partition from a schedule, remapping tasks, ...) to allow to control a distributed execution.

4. Application Domains

4.1. Virtual Reality

Participants: Thierry Gautier, Bruno Raffin, Jean-Louis Roch.

We are pursuing and extending existing collaborations to develop virtual reality applications on PC clusters and grid environments:

- Real time 3D modeling. An on-going collaboration with the MORPHEO project focuses on developing solutions to enable real time 3D modeling from multiple cameras using a PC cluster. An operational code base was transferred to the 4DViews Start-up in September 2007. 4DViews is now selling turn key solutions for real-time 3D modeling. Recent developments take two main directions:
  - Using a HMD (Head Mounted Display) and a Head Mounted Camera to provide the user a high level of interaction and immersion in the mixed reality environment. Having a mobile camera raises several concerns. The camera position and orientation need to be precisely known at anytime, requiring to develop on-line calibration approaches. The background subtraction cannot anymore be based on a static background learning for the mobile camera, required here too new algorithms.
  - Distributed collaboration across distant sites. In the context of the ANR DALIA we are developing a collaborative application where a user at Bordeaux (iParla project-team) using a real time 3D modeling platform can meet in a virtual world with a user in Grenoble also using a similar platform. We rely on the Grid'5000 dedicated 10 Gbits/s network to enable a low latency. The main issues are related to data transfers that need to be carefully managed to ensure a good latency while keeping a good quality, and the development of new interaction paradigms.

  On these issues, Benjamin Petit started a Ph.D. in October 2007, co-advised by Edmond Boyer (PERCEPTION) and Bruno Raffin.
- Real time physical simulation. We are collaborating with the EVASION project on the SOFA simulation framework. Everton Hermann, a Ph.D. co-advised by François Faure (EVASION) and Bruno Raffin, works on parallelizing SOFA using the KAAPI programming environment. The challenge is to provide SOFA with a parallelization that is efficient (real-time) while not being invasive for SOFA programmers (usually not parallel programmer). We developed a first version using the Kaapi environment for SMP machines that relies on a mix of work-stealing and dependency graph analysis and partitioning. A second version targets machines with multiples CPUs and multiple GPUs. We extended the initial framework to support a work stealing based load balancing between CPUs and GPUs. It required to extend Kaapi to support heterogeneous tasks (GPU and CPU ones) and to adapt the work stealing strategy to limit data transfers between CPUs and GPUs (the main bottleneck for GPU computing).
- Distant collaborative work. We conduct experiments using FlowVR for running applications on Grid environments. Two kinds of experiments will be considered: collaborative work by coupling two or more distant VR environments ; large scale interactive simulation using computing resources from the grid. For these experiments, we are collaborating with the LIFO and the LABRI.
- Parallel cache-oblivious algorithms for scientific visualization. In collaboration with the CEA DAM, we have developed a cache-oblivious algorithm with provable performance for irregulars meshes. Based on this work, we are studying parallel algorithms that take advantage of the shared cache usually encountered on multi-core architectures (L3 shared cache). The goal is to have the cores collaborating to efficiently share the L3 cache for a better performance than with a more traditional approach that leads to split the L3 cache between the cores. We are obtaining good performance gains with a parallel iso-surface extraction algorithm. This work is the main focus of Marc Tchiboukdjian Ph.D.

4.2. Code Coupling and Parallel Programming

Participants: Thierry Gautier, Jean-Louis Roch, Vincent Danjean, Frédéric Wagner.

Code coupling aim is to assemble component to build distributed applications by reusing legacy code. The objective here is to build high performance applications for multi-cores, cluster or grid infrastructures.

- **Parallel programming model and runtime support.** Programming parallel applications is a challenging problem. The MOAIS Team has a strong knowledge in parallel algorithms and develop a runtime support for scheduling parallel program written in a very high level interface. The parallelism from recursive divide and conquer applications and those from iterative simulation are studied. Scheduling heuristics are based on online work stealing for the former class of applications, and on hierarchical partitioning for the latter. The runtime support provides capabilities to hide latency by computation thanks to a non-blocking one-side communication protocol and by re-ordering computational tasks.
- **Grid application deployment.** To test grid applications, we need to deploy and start programs on all used computers. This can become difficult if the real topology involves several clusters with firewall, different runtime environments, etc. The MOAIS Team designed and implemented a new tool called `karun` that allows a user to easily deploy a parallel application wrote with the KAAPI software. This KAAPI tool relies on the `TakTuk` software to quickly launch programs on all nodes. The user only needs to describe the hierarchical networks/clusters involved in the experiment with their firewall if any.
- **Visualization of grid applications execution.** The analysis of applications execution on the grid is challenging both because of the large scale of the platform and because of the heterogeneous topology of the interconnections. To help users to understand their application behavior and to detect potential bottleneck or load unbalance, the MOAIS team designed and implemented a tool named **Triva**. This tool proposes a new three dimensional visualization model that combines topological information to space time data collected during the execution. It also proposes an aggregation mechanism that eases the detection of application load unbalance.

4.3. Safe Distributed Computations

Participants: Vincent Danjean, Thierry Gautier, Clément Pernet, Jean-Louis Roch.

Large scale distributed platforms, such as the GRID and Peer-to-Peer computing systems, gather thousands of nodes for computing parallel applications. At this scale, component failures, disconnections (fail-stop faults) or results modifications (malicious faults) are part of operation, and applications have to deal directly with repeated failures during program runs. Indeed, since failure rate in such platform is proportional to the number of involved resources, the mean time between failure is dramatically decreased on very large size architectures. Moreover, even if a middleware is used to secure the communications and to manage the resources, the computational nodes operate in an unbounded environment and are subject to a wide range of attacks able to break confidentiality or to alter the resources or the computed results. Beyond fault-tolerancy, yet the possibility of massive attacks resulting in an error rate larger than tolerable by the application has to be considered. Such massive attacks are especially of concern due to Distributed Denial of Service, virus or Trojan attacks, and more generally orchestrated attacks against widespread vulnerabilities of a specific operating system that may result in the corruption of a large number of resources. The challenge is then to provide confidence to the parties about the use of such an unbound infrastructure. The MOAIS team addresses two issues:

- fault tolerance (node failures and disconnections): based on a global distributed consistent state , for the sake of scalability;
- security aspects: confidentiality, authentication and integrity of the computations.

Our approach to solve those problems is based on the efficient checkpointing of the dataflow that described the computation at coarse-grain. This distributed checkpoint, based on the local stack of each work-stealer process, provides a causally linked representation of the state. It is used for a scalable checkpoint/restart protocol and for probabilistic detection of massive attacks.

Moreover, we study the scalability of security protocols on large scale infrastructures. One goal is trusting the usage of remote-platforms (such as high-performance cluster or cloud infrastructure) by provding quantified guarantees on integrity, accountability and confidentiality. Within the global competitiveness cluster Minalogic, and in collaboration with Privatics team and industrial partners, we have developed a high-rate systematic ciphering architecture that provides red-black segregation on an Intenret network based on the coupling of a multicore architecture with security components (FPGA and smart card).

5. Software and Platforms

5.1. KAAPI

Participants: Thierry Gautier [correspondant], Vincent Danjean, François Broquedis, Joao Ferreira Lima.

- ACM: D.1.3
- License: CeCILL
- OS/Middelware: Unix (Linux, MacOSX, ...)
- Programming language: C/C++, Fortran
- Characterization of Software : A-3 / SO-4 / SM-3 / EM-3 / SDL-4
- Own Contribution: DA-4 / CD-4 / MS-4 / TPM-4
- Additional information:

  Kaapi (http://kaapi.gforge.inria.fr, coordinator T. Gautier) is a middleware for high performance applications running on multi-cores/multi- processors as well as cluster or computational grid. Kaapi provides 1/ a very high level API based on macro data flow language; 2/ several scheduling algorithms for multi-threaded computations as well as for iterative applications for numerical simulation on multi-CPUs / multi-GPUs; 3/ fault-tolerant protocols. Publicly available at http://kaapi.gforge.inria.fr under CeCILL licence. Kaapi has won the 2008 Plugtest organized by Grid@Works. Kaapi provides ABI compliant implementations of Quark (PLASMA, Linear Algebra, Univ. of Tennesse) and libGOMP (GCC runtime for OpenMP). Direct competitors with 1/: Quark (UTK), OMPSs (UPC, BSC), OpenMP. Direct competitors with 2/: StarSs, StarPU (Inria RUNTIME). Direct competitors providing 3/: Charm++, MPI.

5.2. FlowVR

Participants: Bruno Raffin [correspondant MOAIS], Matthieu Dreher, Jérémy Jaussaud.

- ACM: D.1.3
- License: GPL and LGPL
- OS/Middelware: Unix (Linux, MacOSX, ...)
- Programming language: C/C++
- Characterization of Software : A-3 / SO-4 / SM-3 / EM-3 / SDL-4
- Own Contribution: DA-4 / CD-3 / MS-3 / TPM-4
- Additional information: FlowVR (http://flowvr.sf.net, coordinator B. Raffin) provides users with the necessary tools to develop and run high performance interactive applications on PC clusters and Grids. The main target applications include virtual reality, scientific visualization and in situ analytics. FlowVR enforces a modular programming that leverages software engineering issues while enabling high performance executions on distribued and parallel architectures. FlowVR is the reference backbone for Grimage. See also the web page http://flowvr.sf.net.

5.3. TakTuk - Adaptive large scale remote execution deployment

Participants: Guillaume Huard [correspondant], Pierre Neyron.

- Characterization of Software : A-2 / SO-3 / SM-5 / EM-3 / SDL-4
- Own Contribution: DA-4 / CD-4 / MS-4 / TPM-4
- Additional information:
  - web site: http://taktuk.gforge.inria.fr, Coordinator G. Huard
  - Objective of the software: TakTuk is a tool for deploying parallel remote executions of commands to a potentially large set of remote nodes. It spreads itself using an adaptive algorithm and sets up an interconnection network to transport commands and perform I/Os multiplexing/demultiplexing. The TakTuk mechanics dynamically adapt to environment (machine performance and current load, network contention) by using a reactive work-stealing algorithm that mixes local parallelization and work distribution.
  - Users community: TakTuk is a research open source project available in the Debian GNU/Linux distribution (package taktuk) used in lower levels of Grid5000 software architectures (nodes monitoring in OAR, environment diffusion in Kadeploy). The community is small : developers and administrators for large scale distributed platforms, but active.
  - Positioning: main competing tools are pdsh (but uses linear deployment) and gexec (not fault tolerant, requires installation), for more details : B. Claudel, G. Huard and O. Richard. TakTuk, Adaptive Deployment of Remote Executions. In Proceedings of the International Symposium on High Performance Distributed Computing (HPDC), 2009. TakTuk is the only tool to provide to deployed processes a communication layer (just like an MPIrun, but not tied to a specific environment) and synchronization capabilities.

5.4. Triva

Participant: Guillaume Huard [correspondant].

- Characterization of Software : A-2 / SO-4 / SM-5 / EM-3 / SDL-3
- Own Contribution: DA-4 / CD-3 / MS-3 / TPM-3
- Additional information:
  - web site: http://triva.gforge.inria.fr/, Coordinator, Lucas Schnorr
  - Objective of the software: Triva is an open-source tool used to analyze traces (in the pajé format) registered during the execution of parallel applications. The tool serves also as a sandbox to the development of new visualization techniques.
  - Users community: Research open source project, applications developers, especially parallel applications.
  - Positioning: Main competing tools are Vampir (classical 2D Gantt charts) and Tau (less advanced agregation techniques), more details in : A Hierarchical Aggregation Model to achieve Visualization Scalability in the analysis of Parallel Applications. Lucas Mello Schnorr, Guillaume Huard, Philippe Olivier Alexandre Navaux. Parallel Computing. Volume 38, Issue 3, March 2012.

5.5. OAR

Participants: Pierre Neyron [correspondant MOAIS], Grégory Mounié.

- Characterization of Software : A-5 / SO-3 / SM-4 / EM-4 / SDL-5
- Own Contribution: DA-3 / CD-2 / MS-1 / TPM-1
- Additional information: OAR (http://oar.imag.fr, Coordinator O. Richard, Inria MESCAL) is a batch scheduler. The MOAIS team develops the central automata and the scheduling module that includes successive evolutions and improvements of the policy.OAR is used to schedule jobs both on the CiGri (Grenoble region) and Grid50000 (France) grids. CiGri is a production grid that federates about 500 heterogeneous resources of various Grenoble laboratories to perform computations in physics. MOAIS has also developed the distributed authentication for access to Grid5000.

5.6. SOFA

Participant: Bruno Raffin [correspondant].

- ACM: D.1.3
- Programming language: C/C++
- Characterization of Software : A-5 / SO-4 / SM-4 / EM-4 / SDL-5
- Own Contribution: DA-2 / CD-2 / MS-1 / TPM-1
- Additional information: SOFA (http://www.sofa-framework.org/, Coordinator F. Faure, Inria IMAGINE) is an Open Source framework primarily targeted at real-time simulation, with an emphasis on medical simulation. It is mostly intended for the research community to help develop newer algorithms, but can also be used as an efficient prototyping tool. Moais contributes to parallelization of kernel algorithms used in the simulation.

5.7. LinBox

Participants: Clément Pernet [correspondant], Thierry Gautier.

- Characterization of Software : A-3 / SO-4 / SM-2 / EM-3 / SDL-5
- Own Contribution: DA-4 / CD-3 / MS-3 / TPM-4
- Additional information:
  - web site: http://linalg.org
  - Objective of the software: LinBox is an open-source C++ template library for exact, high-performance linear algebra computations. It is considered as the reference library for numerous computations (such as linear system solving, rank, characteristic polynomial, Smith normal forms,...) over finite fields and integers with dense, sparse, and structured matrices.
  - The LinBox group is an international collaboration (USA: NCSU, UDel; Canada: U Waterloo, U Calgary; France: LIP, LIRMM, LJK and LIG). Articles related to the library have been published in the main Conferences of the area: ISSAC, ICMS. MOAIS contributes to its development and more specifically to its parallelization in the context of ANR HPAC project. It is currently experiencing a major change of design, to better integrate parallelism.
  - Users community: mostly researchers doing computational mathematics (number theory, cryptology, group theory, persistent homology. They use the library by either linking against it directly (the library is packaged in Debian, Fedora, etc) or withing the general purpose math software Sage (sagemath.org very broad diffusion) which includes LinBox as a kernel for exact linear algebra.

6. New Results

6.1. Distributed Art Performance

Moais collaborated with partners fromI2cat, Barcelona, Psnc, Poznan and Grenoble-INP to setup a live distributed art performance for the ICT 2013 conference at Vilnius. This distributed performance gathered musicians located a Poznan, Barcelona and Vilnius, as well as a dancer modeled in 3D on the Grimage platform at Inria Grenoble. Though physically present in different cities these artists performed together for a numerical dance and music performance numerically assembled and transmitted in real-time at Vilnius. This joint effort relies on the FlowVR framework from Moais and the UltraGrid software from CESNET. This event received a significant attention from the medias (In France: FR3 and Tele-grenoble, France inter, etc.). A video is available at http://cyan1.grenet.fr/podcastmedia/Visionair/ICT2013_promo.m4v.

6.2. VTK Parallelization Framework

Moais developed a framework for the parallelization of scientific visualization algorithms based on on-demand task extraction and work stealing techniques. This work is developed for the VTK software and supports the OpenMP, Intel TBB and Kaapi runtime environments. Mathias Ettinger visited the Kitware company, NY, for two months to prepare the integration of this work in the next release of VTK. This work is performed in collaboration with the EDF company.

6.3. Parallel Sorting Algorithm

We developped a novel adaptive sorting algorithm, called PaVO, relying on a Packed Mermory Array data structure. Maintaining gaps in the array of elements enable to reduce the span of modifications needed when reordering elements. This is particularly relevant in a parallel context to reduce the data dependencies. Performance results on a NUMA architecute show that PaVo outperfoms standard parallel sorting algorithms even for a large amount of disorder.

6.4. High bandwidth IPSec gateways and ICMP

Internet Control Message Protocol (ICMP)i is essential for performance aspects in particular for Path Maximum Transmission Unit discovery but is also known to be a cause of attacks. In collaboration with Planet, we demonstrate, through a real exploit on a testbed, that an external attacker having eavesdropping and traffic injection capabilities in the black untrusted network, without any access to clear-text (thesis of Ludovic Jacquin). This impacts out current research on trusted outsourced computations.

6.5. Efficient Parallel multi-GPUs execution

We developped a novel scheduling algorithm in Kaapi to perform multi-GPUs execution of task' based program [19]. Performance results on Cholesky factorization on up to 8-GPUs shows that Kaapi outperfoms similar runtime systems and even hand code parallel version.

6.6. Porting Kaapi for Native Mode on Intel Xeon Phi

Kaapi was ported natively on Intel Xeon Phi co-processor. Specific memory hierarchy was managed transparently to the applicaiton by the development of specific hierarchical work stealing scheduler. Experimentations on dense linear algebra kernels (Cholesky, LU and QR factorization) shows a very promising gain compared to the standard parallel implementation available in the Intel MKL [16]. Extension of theses resultats are under publication process.

6.7. Adaptive loop scheduling in GCC OpenMP runtime library

We port an adaptive loop scheduler from Kaapi into the OpenMP runtime library of GCC called libGOMP [12]. The loop scheduler is consencious of the bloc data mapping to improve locality of computation.

6.8. Kaapi in EPX standard distribution

Kaapi software developped by the MOAIS team was included in the standard EPX distribution. EPX has won the 2013 Grand Prix SFEN (http://www-epx.cea.fr).

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

- Contract with Bull (2013–2016). Multiobjective scheduling on supercomputer towards exascale. Associated to a CIFRE PhD grant (David Glesser, started in 4/2013). Partners: Inria - LIG Moais, Bull

7.2. Bilateral Grants with Industry

- Contract with EDF (2010-2013). High performance scientific visualization. Funds 1 postdoc and 1 PhD (Mathias Ettinger). Partners: Inria (MOAIS and EVASION), EDF R&D
- CEA: Collaboration with CEA (2012): Europlexus Parallelization with KAAPI. Partners: Inria Rhônes-Alpes and CEA Saclay (CEA funds the PhD of Marwa Sridi started in 4/2013).

8. Partnerships and Cooperations

8.1. Regional Initiatives

- Action OCA : Combinatorial Optimization on Accelerator. Funding by Labex PERSYVAL Grenoble.

8.2. National Initiatives

8.2.1. ANR

- **ANR grant REPDYN (2010-2013).** High performance computing for structure and fluid computing. Partners: Inria Rhône-Alpes, CEA, ONERA, EDF, LaMSID lab from CNRS and LaMCoS lab from INSA Lyon.
- **ANR/JST grant PETAFLOW (2010-2013).** France/Japan international program. Peta-scale data intensive computing with transnational high-speed networking: application to upper airway flow. Inria Rhône-Alpes, Gipsa-lab from UJF, NITC (Japan), Cyber Center of Osaka, DITS (Osaka) and the Visualization Lab of Kyoto.
- **ANR grant EXAVIZ (2011-2015).** Large Scale Interactive Visual Analysis for Life Science. Partners: Inria Rhône-Alpes, Université d'Orléans, the LBT lab from IBPC, the LIMSI from Université d'Orsay, and the CEMHTI labs from CNRS.
- **ANR HPAC (2012-2015)**. High Performance Algebraic Computing. Coordinator: UJF (LJK/CASYS team). Partners: project-team MOAIS (Grenoble), project-team ARENAIRE (LIP, Lyon), project-team SALSA (LIP6, Paris), the ARITH group (LIRMM lab, Montpellier).
- **Equipex Kinovis (2012-2017)**. 2.6 Meuros. Large scale multi-camera platform (extension of the Grimage platform to 60 cameras, depth and X-ray cameras). Coordinator E Boyer, LJK Inria MORPHEO team. Partners: Inria Rhône-Alpes and the LJK, LIG, LADAF and GIPSA labs.
- **ANR-11-LABX-0025 PERSYVAL-Lab** funds the following PhD in collaboration with other labs:
  - in collaboration with Verimag: Multi-objective optimization for resource management on multicore systems, (PhD Abhinav Srivastav, since 9/2012)
  - In collaboration with Gipsa-lab and Inria BiBop: Simulations of Fibrous Materials. (PhD Gilles Daviet, since 9/2013)
  - in collaboration with Inria Privatics and Verimag: Secure Outsourcing (PhD Amrit Kumar, since 11/2013)

8.2.2. Competitivity Clusters

- SoC-Trace, Minalogic 2011-2014 contract. This project aims the development of tools for the monitoring and debug of mumticore systems on chip. Leader: ST-Microelectonic. Partners: Inria (Mescal, Moais); UJF (TIMA, LIG/Hadas); Magilem, ProBayes. Moais contributes with technics and tools for visual aggregation of application traces. The contract funds 1 phD thesis (Damien Dosimont) and 1 year engineer.

8.2.3. National ADT

- ADT K'STAR with cooperation between EPIs MOAIS and RUNTIME (Bordeaux). Coordinator: T. Gautier. https://gforge.inria.fr/projects/kstar. The main objective is to provide OpenMP-3.1 with some extension from OpenMP-4.0 standard to perform OpenMP programs on multi-CPUs multi-GPUs by using Kaapi and StarPU runtimes.

8.2.4. Inria Project Lab

8.2.4.1. C2S@Exa - Computer and Computational Scienecs at Exascale

Participants: Olivier Aumage [RUNTIME project-team, Inria Bordeaux - Sud-Ouest], Jocelyne Erhel [SAGE project-team, Inria Rennes - Bretagne Atlantique], Philippe Helluy [TONUS project-team, Inria Nancy - Grand-Est], Laura Grigori [ALPINE project-team, Inria Saclay - Île-de-France], Jean-Yves L'excellent [ROMA project-team, Inria Grenoble - Rhône-Alpes], Thierry Gautier [MOAIS project-team, Inria Grenoble - Rhône-Alpes], Luc Giraud [HIEPACS project-team, Inria Bordeaux - Sud-Ouest], Michel Kern [POMDAPI project-team, Inria Paris - Rocquencourt], Stéphane Lanteri [Coordinator of the project], François Pellegrini [BACCHUS project-team, Inria Bordeaux - Sud-Ouest], Christian Perez [AVALON project-team, Inria Grenoble - Rhône-Alpes], Frédéric Vivien [ROMA project-team, Inria Grenoble - Rhône-Alpes].

Since January 2013, the team is participating to the C2S@Exa http://www-sop.inria.fr/c2s_at_exa Inria Project Lab (IPL). This national initiative aims at the development of numerical modeling methodologies that fully exploit the processing capabilities of modern massively parallel architectures in the context of a number of selected applications related to important scientific and technological challenges for the quality and the security of life in our society. At the current state of the art in technologies and methodologies, a multidisciplinary approach is required to overcome the challenges raised by the development of highly scalable numerical simulation software that can exploit computing platforms offering several hundreds of thousands of cores. Hence, the main objective of C2S@Exa is the establishment of a continuum of expertise in the computer science and numerical mathematics domains, by gathering researchers from Inria project-teams whose research and development activities are tightly linked to high performance computing issues in these domains. More precisely, this collaborative effort involves computer scientists that are experts of programming models, environments and tools for harnessing massively parallel systems, algorithmists that propose algorithms and contribute to generic libraries and core solvers in order to take benefit from all the parallelism levels with the main goal of optimal scaling on very large numbers of computing entities and, numerical mathematicians that are studying numerical schemes and scalable solvers for systems of partial differential equations in view of the simulation of very large-scale problems.

T. Gautier is coordinator of the Pole 4: Programming Models.

8.3. European Initiatives

8.3.1. FP7 Projects

8.3.1.1. VISIONAIR

Type: CAPACITIES

Defi: INFRA-2010-1.1.29

Instrument: Combination of COLLABORATIVE PROJECTS and COORDINATION and SUPPORT ACTIONS

Objectif: NC

Duration: February 2011 - January 2015

Coordinator: Frédéric Noël (Inpg)

Partner: Inria (France)

Inria contact: G. Dumont

Abstract: Visionair calls for the creation of a European infrastructure for high level visualisation facilities that are open to research communities across Europe and around the world. By integrating existing facilities, Visionair aims to create a world-class research infrastructure for conducting state-of-the-art research in visualisation, thus significantly enhancing the attractiveness and visibility of the European Research Area (ERA). With over 20 members across Europe participating, VISION-AIR offers facilities for Virtual Reality, Scientific Visualisation, Ultra High Definition, Augmented Reality and Virtual Services.

8.4. International Initiatives

8.4.1. Participation In other International Programs

- The MOAIS EPI is activiley participating to the international LICIA lab supporting collaborations with the UFRGG, Brazil. Bruno Raffin is member of the LICIA scientific comittee.
- Moais is also leading a CAPES/COFECUB program (2013-2014) with UFRGS, Brazil.
- Moais is also leading a CAPES/COFECUB program (2013-2014) with USP, Brazil.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

- Jacek Blazewicz (Professor, Polish Academy of Sciences, Poznań), invited prof INP (2 months)
- Afredo Goldman (Professor, USP Sao Paulo) (1 month)
- Daniel Cordeiro (Postdoc, USP Sao Paulo) (1 month)
- Mario Cesar Lopez Loces (UFRGS) (1 month)
- Adel Essafi (ISIG Kairouan, Tunis) (2 month)

8.5.2. Visits to International Teams

- Damien Dosimont, Oct-Dec 2013, UFRGS, Brazil
- Clement Pernet, sabbatical, Sept-Dec 2013, LIP Lyon, Aric Team

9. Dissemination

9.1. Scientific Animation

- Chair, Steering board:
  - Thierry Gautier, Chair of ComPAS'2013, Track Parallelism, located in Inria Grenoble, January 2013.
  - Bruno Raffin, Paper co-chair of JVRC 2013 (EGVE - EuroVR Joint Virtual Reality Conference).
  - Bruno Raffin, Member of the Steering Board of EGPGV (Eurographics Symposium on Parallel Rendering and Visualization)
  - Bruno Raffin, Member of the Steering Board of the journée visualisation scientifique.

  - Denis Trystram, Chair of the track Algorithms HiPC (20th Int'l Conf. on High Performance Computing), dec. 16-20, Hyderabad, India
- Program committees:
  - ICCS 2013 (International Conference on Computational Science).
  - GPUCOMP 2013 (third Workshop on GPU Computing).
  - IEEE VR 2013.
  - ISVC 2013 (International Symposium on Visual Computing).
  - SEARIS 2013 (Workshop on Software Engineering and Architectures for Realtime Interactive Systems).
  - COMPASS, january 15-18 2013, Grenoble, France
  - IPDPS 2013 (24th IEEE International Parallel & Distributed Processing Symposium) may 20-24 2013, Boston USA
  - HCW'2013 (22nd IEEE Heterogeneous Computing Workshop) may, 2013, Boston USA
  - ISPDC (12th Internat Symposium on Parallel and Distributed Computing) june 27-30, 2013, Bucharest, Romania
  - OPTIM'13 (Workshop on Opt. Issues in Energy Efficient Distributed Systems) july 1-5, 2013, Helsinki, Finland
  - 6th MISTA, august 27-29 2013, Ghent, Belgium
  - 10th PPAM 2013, september 9-12, 2011, Warsaw, Poland
  - ParCo'2013, sept. 10-13, 2013, Munich, Germany
  - IC3 2012 (5th International Conference of Contemporary Computing) august 6-8, 2012, Noida, India
  - EuroMPI (20th internat conference) sept. 15-18, 2013, Madrid, Spain
  - ESA (21st European Symposium on Algorithms) sept. 2-6, 2013, Sophia Antipolia, France
  - 25th SBAC-PAD, october 23-26, 2013, Pernambuco, Brazil

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master: J-L. Roch co-director (Grenoble-INP) with P Elbaz-Vincent (Université Joseph Fourier, Math. Dept) of the Master "SCCI Security, Cryptology and Coding of Information Systems" (M2) joined between UJF and INP Grenoble Universities. This Master, started in 2001, is taught in English from sept 2007 (international Master).

Master: C. Pernet and Denis Trystram are responsible of the first year (M1) of the international Master of Science in Informatics at Grenoble (MOSIG-M1).

Master: V Danjean: course "Parallel Programming" (M2), Grenoble University,.

Master: J-L. Roch, "Security models" 24h (M2), Grenoble University

Master: D. Trystram, P.-F. Dutot, J.-L. Roch, "Complexity, approximation theory and randomization" master course (M2) at Grenoble University

Master: François Broquedis. 192 hours per year. 192 hours per year. Engineering school Grenoble-INP/Ensimag, 1st year/L3 and Master (M1/2nd year and M2/3rd year).

Master: Vincent Danjean. 242 hours per year. Licence (third year) and Master (first and second year) at Joseph Fourier University. First to third year of engineering school at Polytech' Grenoble.

Master: Pierre-François Dutot. 226 hours per year. Licence (first and second year) at IUT2/UPMF (Institut Universitaire Technologique de l'Université Pierre Mendès-France) and 9 hours Master M2R-ISC Informatique-Systèmes-Communication at Joseph Fourier University.

Master: Guillaume Huard. 242 hours per year. Licence (first and third year) and Master (first year) at Joseph Fourier University.

Master: Grégory Mounié. 242 hours per year. Master (first year) and Computer Science for Non Computer Scientist Post-Master at Engineering school ENSIMAG and Dept TELECOM, Grenoble-INP.

Master: Clement Pernet. 210 hours per year. University J. Fourier. Master (first year and second year) and Licence (3rd year).

Master: Bruno Raffin. 22 hours per year. Master at Université d'Orléans and Polytech'Grenoble.

Master: Jean-Louis Roch. 242 hours per year. Engineering school Grenoble-INP/Ensimag and Master (M1/2nd year and M2/3rd year)

Master: Denis Trystram. 200 hours per year in average, mainly at first level of Engineering School.

Master: Frédéric Wagner. 220 hours per year. Engineering school ENSIMAG, Grenoble-INP (M1/2nd year and M2/3rd year) (190h) ; Master DESS/M2-P SCCI Security (30h).

9.2.2. Supervision

PhD in progress : **David Beniamine (since 2013).** Parallelisation Patterns and Scheduling for Real-Time Physics Simulations. Co-advised Guillaume Huard and Bruno Raffin.

PhD in progress : **Amrit Kumar (since 2013)**. Analysis of work-stealing and adpative algorithms. Labex-Persyval co-advised PhD by Pascal Lafourcade (Verimag lab), Cedric Lauradoux (Inria Privatics team) and Jean-Louis Roch.

PhD in progress : **Gilles Daviet (since 2013).** Parallel Macroscopic Simulations of Fibrous Materials. Co-advised by Florence Bertails-Descoubes, Pierre Saramito and Bruno Raffin.

PhD in progress : **Julio Toss (since 2013).** Parallel Algorithms and Data Structures for Physically Based Simulation of Deformable Objects. Join Ph.D with UFRGS, Brazil. Co-advised by Joao Comba and Bruno Raffin.

PhD in progress : **Marwa Sridi (since 2013).** Parallel Algorithm Composition for Transient Mechanic Simulations. Join Ph.D with CEA, France. Co-advised by Vincent Faucher, Thierry Gautier and Bruno Raffin.

PhD in progress : **Ziad Sultan (since 2012)**. Hig-performance algebraic computations. Co-advised Ph.D by Jean-Guillaume Dumas (LJK Lab) and Clément Pernet.

PhD in progress : **Mathias Ettinger (since 2011).** Cache Efficient Parallel Adaptive Algorithms for Scientific Visualization. Co-advised by Bruno Raffin and François Broquedis

PhD in progress : **Stefano Mor (since 2011)**. Analysis of work-stealing and adpative algorithms. Join Ph.D with UFRGS, Brazil. Co-advised by Jean-Louis Roch, Nicolas Maillard and Bruno Raffin

PhD in progress : **Mathieu Dreher (since 2011)**. In-Situ Visualization for Molecular Dynamics. Advised by Bruno Raffin.

PhD in progress : **Joao Ferreira Lima (since 2010)**. Work Stealing on GPUs. Join Ph.D with UFRGS, Brazil. Co-advised by Vincent Danjean, Thierry Gautier, Nicolas Maillard and Bruno Raffin

i

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] L. JACQUIN. , *Compromis performance/sécurité des passerelles très haut débit pour Internet.*, Université de Grenoble, November 2013, http://hal.inria.fr/tel-00911075

Articles in International Peer-Reviewed Journals

[2] M. Bougeret, P.-F. Dutot, D. Trystram. *Moderately exponential approximation for makespan minimization on related machines*, in "Theoretical Computer Science", 2013, In press [*DOI :* 10.1016/j.tcs.2013.03.020], http://hal.inria.fr/lirmm-00838717

[3] M. Dreher, J. Prevoteau-Jonquet, M. Trellet, M. Piuzzi, M. Baaden, B. Raffin, N. Férey, S. Robert, S. Limet. *ExaViz: a Flexible Framework to Analyse, Steer and Interact with Molecular Dynamics Simulations*, in "Faraday Discussions of the Chemical Society", May 2014 [*DOI :* 10.1039/C3FD00142C], http://hal.inria.fr/hal-00942627

[4] C.-P. Jeannerod, C. Pernet, A. Storjohann. *Rank-profile revealing Gaussian elimination and the CUP matrix decomposition*, in "Journal of Symbolic Computation", 2013, vol. 56, pp. 46-68 [*DOI :* 10.1016/j.jsc.2013.04.004], http://hal.inria.fr/hal-00841300

International Conferences with Proceedings

[5] A. Ancel, I. Assenmacher, K. Baba, J. Cisonni, Y. Fujiso, P. Goncalves, M. Imbert, K. Koyamada, P. Neyron, K. Nozaki, H. Ohsaki, A. Orgerie, X. Pelorson, B. Raffin, N. Sakamoto, E. Sakane, S. Wada, S. Shimojo, A. Van Hirtum. *PetaFlow - An example of communication and computational technologies with social impact*, in "HSNCE 2013- The 4th IEEE International Workshop on High-Speed Network and Computing Environment - 2013", Kyoto, Japan, 2013, http://hal.inria.fr/hal-00849026

[6] A. Ancel, B. Raffin, I. Assenmacher, A. Van Hirtum, Y. Fujiso, K. Nozaki. *Parallel caching for remote visualisation*, in "HSNCE 2013- The 4th IEEE International Workshop on High-Speed Network and Computing Environment - 2013", Kyoto, Japan, 2013, http://hal.inria.fr/hal-00849027

[7] J. Blazewicz, P. Bouvry, J. Musial, J. Pecero, D. Trystram. *Dual Discounting Functions for Internet Shopping Optimization Problem*, in "Proceedings of the 6th Multidisciplinary International Scheduling Conference: Theory and Applications MISTA", Ghent, Belgium, 2013, http://hal.inria.fr/hal-00921360

[8] J. Blazewicz, S. Kedad-Sidhoum, F. Monna, G. Mounié, D. Trystram. *Preemptive scheduling of independent tasks on multi-cores with GPU*, in "ECCO XXVI: the 26th European Chapter on Combinatorial Optimization", Paris, France, 2013, http://hal.inria.fr/hal-00921355

[9] M. Dreher, M. Piuzzi, T. Ahmed, C. Matthieu, M. Baaden, N. Férey, S. Limet, B. Raffin, S. Robert. *Interactive Molecular Dynamics: Scaling up to Large Systems*, in "International Conference on Computational Science, ICCS 2013", Barcelone, Spain, Elsevier, June 2013, http://hal.inria.fr/hal-00809024

[10] M. Dreher, B. Raffin. *A Flexible Framework for Asynchronous In Situ and In Transit Analytics for Scientific Simulations*, in "CCGrid - International Symposium on Cluster, Cloud and Grid Computing", Chicago, IL, USA, IEEE Computer Science Press, May 2014, http://hal.inria.fr/hal-00941413

[11] J.-G. Dumas, C. Pernet, Z. Sultan. *Simultaneous computation of the row and column rank profiles*, in "ISSAC 2013 - 38th International Symposium on Symbolic and Algebraic Computation", Boston, MA, United States, M. Kauers (editor), ACM, June 2013, pp. 181-188 [*DOI :* 10.1145/2465506.2465517], http://hal.inria.fr/hal-00778136

[12] M. DURAND, F. BROQUEDIS, T. GAUTIER, B. RAFFIN. *An Efficient OpenMP Loop Scheduler for Irregular Applications on Large-Scale NUMA Machines*, in "International Workshop on OpenMP (IWOMP)", Canberra, Australia, A. P. RENDELL, B. M. CHAPMAN, M. S. MÜLLER (editors), Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, vol. 8122, pp. 141-155 [*DOI :* 10.1007/978-3-642-40698-0_11], http://hal.inria.fr/hal-00867438

[13] P.-F. DUTOT, K. JANSEN, C. ROBENEK, D. TRYSTRAM. *A (2+ϵ)-Approximation for Scheduling Parallel Jobs in Platforms*, in "the 19th International Euro-Par Conference", Aachen, Germany, Lecture Notes in Computer Science, Springer, 2013, vol. 8097, pp. 78-89, http://hal.inria.fr/hal-00921354

[14] J. EMERAS, V. PINHEIRO, K. RZADCA, D. TRYSTRAM. *OStrich: Fair Scheduling for Multiple Submissions*, in "PPAM'2013", Warsaw, Poland, Springer, 2013, http://hal.inria.fr/hal-00918374

[15] M. ETTINGER, F. BROQUEDIS, T. GAUTIER, S. PLOIX, B. RAFFIN. *VtkSMP: Task-based Parallel Operators for VTK Filters*, in "EGPGV - 13th Eurographics Symposium on Parallel Graphics and Visualization", Girona, Spain, Eurographics Association, May 2013, http://hal.inria.fr/hal-00926457

[16] J. V. FERREIRA LIMA, F. BROQUEDIS, T. GAUTIER, B. RAFFIN. *Preliminary Experiments with XKaapi on Intel Xeon Phi Coprocessor*, in "25th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)", Porto de Galinhas, Brazil, October 2013 [*DOI :* 10.1109/SBAC-PAD.2013.28], http://hal.inria.fr/hal-00878325

[17] *Best Paper*
T. FRANKE, V. SETTGAS, J. BEHR, B. RAFFIN. *VCoRE: a web resource oriented architecture for efficient data exchange*, in "18th International Conference on 3D Web", San Sebastian, Spain, ACM, June 2013, pp. 71-78, Best Long Paper - Second Prize [*DOI :* 10.1145/2466533.2466545], http://hal.inria.fr/hal-00919977.

[18] T. GAUTIER, J. V. FERREIRA LIMA, N. MAILLARD, B. RAFFIN. *Locality-Aware Work Stealing on Multi-CPU and Multi-GPU Architectures*, in "6th Workshop on Programmability Issues for Heterogeneous Multicores (MULTIPROG)", Berlin, Germany, January 2013, http://hal.inria.fr/hal-00780890

[19] T. GAUTIER, J. V. FERREIRA LIMA, N. MAILLARD, B. RAFFIN. *XKaapi: A Runtime System for Data-Flow Task Programming on Heterogeneous Architectures*, in "27th IEEE International Parallel & Distributed Processing Symposium (IPDPS)", Boston, Massachusetts, United States, May 2013, http://hal.inria.fr/hal-00799904

[20] T. GAUTIER, F. LEMENTEC, V. FAUCHER, B. RAFFIN. *X-Kaapi: a Multi Paradigm Runtime for Multicore Architectures*, in "Workshop P2S2 in conjunction of ICPP", Lyon, France, October 2013, 16 p. , http://hal.inria.fr/hal-00727827

[21] A. GOLDMAN, J. LEPPING, Y. NGOKO, D. TRYSTRAM. *Combining Parallel Algorithms Solving the Same Application: What is the Best Approach?*, in "ParLearning, IPDPS Workshop", Boston, United States, 2013, pp. 1859-1868, http://hal.inria.fr/hal-00921358

[22] S. KEDAD-SIDHOUM, F. MONNA, G. MOUNIÉ, D. TRYSTRAM. *Approximation Algorithms for a Scheduling Problem on Multi-Cores with GPUs*, in "Proceedings of the 11th workshop on Models and Algorithms for Planning and Scheduling Problems MAPSP", Pont-à-Mousson, France, 2013, http://hal.inria.fr/hal-00921356

[23] *Best Paper*
S. KEDAD-SIDHOUM, F. MONNA, G. MOUNIÉ, D. TRYSTRAM. *Scheduling Independent Tasks on Multi-Cores with GPU Accelerators*, in "Proceedings of the 11th HeteroPar workshop (Algorithms, Models and Tools for Parallel Computing and Heterogeneous Platforms)", Aachen, Germany, 2013, http://hal.inria.fr/hal-00921357.

[24] A. KUMAR, J.-L. ROCH, C. PERNET. *Secured Outsourced Linear Algebra*, in "Safecomp 2013 FastAbstract", Toulouse, France, M.-O. KILLIJIAN (editor), September 2013, http://hal.inria.fr/hal-00926445

[25] J. LEPPING, P. MERTIKOPOULOS, D. TRYSTRAM. *Accelerating Population-Based Search Heuristics by Adaptive Resource Allocation*, in "GECCO '13: Proceedings of the 15th ACM Annual Conference on Genetic and Evolutionary Computation", Amsterdam, Netherlands, 2013, pp. 1165-1172, http://hal.inria.fr/hal-00912839

[26] G. PAGANO, D. DOSIMONT, G. HUARD, V. MARANGOZOVA-MARTIN, J.-M. VINCENT. *Trace Management and Analysis for Embedded Systems*, in "Ieee 7th International Symposium on Embedded Multicore/Many-core SoCs", Tokyo, Japan, 2013, http://hal.inria.fr/hal-00918439

[27] D. TRYSTRAM. *Are we expecting too much from GPUs?*, in "PPAM, the 10th International Conference on Parallel Processing and Applied Mathematics", Warsaw, Poland, 2013, Keynote talk, http://hal.inria.fr/hal-00921365

National Conferences with Proceedings

[28] D. DOSIMONT, G. HUARD, J.-M. VINCENT. *La visualisation de traces, support à l'analyse, déverminage et optimisation d'applications de calcul haute performance*, in "Actes de l'atelier Visualisation d'informations, interaction et fouille de données (VIF) de la 13e Conférence Francophone sur l'Extraction et la Gestion des Connaissances (EGC'2013)", Toulouse, France, 2013, pp. 55–66, http://hal.inria.fr/hal-00918438

[29] L. EYRAUD-DUBOIS, A. LÈBRE, P. MARTINEAU, A. SOUKHAL, V. T'KINDT, D. TRYSTRAM. *A Server Consolidation Problem: Definition and Model*, in "Proceedings of the 14th conference ROADEF", Troyes, France, 2013, http://hal.inria.fr/hal-00921363

Conferences without Proceedings

[30] A. KUMAR, J.-L. ROCH, C. PERNET. *Secured Outsourced Linear Algebra*, in "SAFECOMP'2013, The 32nd International Conference on Computer Safety, Reliability and Security", Toulouse, France, Jérémie Guiochet (LAAS-CNRS, Univ. Toulouse, FR), August 2013, http://hal.inria.fr/hal-00926723

Scientific Books (or Scientific Book chapters)

[31] A. CASAMAYOU, G. CONNAN, T. DUMONT, L. FOUSSE, F. MALTEY, M. MEULIEN, M. MEZZAROBBA, C. PERNET, N. M. THIÉRY, P. ZIMMERMANN. , *Calcul mathématique avec Sage*, Amazon, 2013, 468 p. , electronic version available under Creative Commons license, http://hal.inria.fr/inria-00540485

[32] J.-G. DUMAS, J.-L. ROCH, E. TANNIER, S. VARRETTE. , *Théorie des Codes : compression, cryptage, correction, 2e edition*, Sciences Sup, Dunod, August 2013, 384 p. , http://hal.inria.fr/hal-00932851

[33] J.-G. DUMAS, J.-L. ROCH, E. TANNIER, S. VARRETTE. , *Foundations of Coding: Compression, Encryption, Error-Correction*, 2014, A paraître, http://hal.inria.fr/hal-00765802

[34] J.-G. DUMAS, C. PERNET. *Computational linear algebra over finite fields*, in "Handbook of Finite Fields", G. L. MULLEN, D. PANARIO (editors), Discrete Mathematics and Its Applications, Chapman & Hall / CRC, June 2013, pp. 514-528, http://hal.inria.fr/hal-00688254

[35] C. GRIMME, M. KEMMERLING, J. LEPPING. *On the Integration of Theoretical Single-Objective Scheduling Results for Multi-objective Problems*, in "EVOLVE - A bridge between probability, set oriented numerics and evolutionary computation", E. TANTAR, A.-A. TANTAR, P. BOUVRY, P. DEL MORAL, P. LEGRAND, C. A. COELLO COELLO, S. OLIVER (editors), Studies in Computational Intelligence, Springer Berlin Heidelberg, 2013, vol. 447, pp. 333-363, to appear, http://hal.inria.fr/hal-00921362

Books or Proceedings Editing

[36] M. BENDER, J. BLAZEWICZ, E. PESCH, D. TRYSTRAM, G. ZHANG (editors). , *Guest editorial: special issue of the workshop "New trends in scheduling", Frejus sept. 2010*, Springer, 2013, vol. 16, pp. 347-348, http://hal.inria.fr/hal-00921364

Research Reports

[37] J.-G. DUMAS, T. GAUTIER, C. PERNET, Z. SULTAN. , *Parallel computation of echelon forms*, February 2014, http://hal.inria.fr/hal-00947013

[38] M. ETTINGER, F. BROQUEDIS, T. GAUTIER, S. PLOIX, B. RAFFIN. , *VtkSMP: Task-based Parallel Operators for VTK Filters*, Inria, February 2013, n° RR-8245, 19 p. , http://hal.inria.fr/hal-00789814

[39] G. PAGANO, D. DOSIMONT, G. HUARD, V. MARANGOZOVA-MARTIN, J.-M. VINCENT. , *Trace Management and Analysis for Embedded Systems*, Inria, May 2013, n° RR-8304, 21 p. , http://hal.inria.fr/hal-00821907

Other Publications

[40] D. BENIAMINE. , *Impact of communication times on mixed CPU/GPU applications scheduling using KAAPI*, Université Joseph Fourrier, Grenoble IGrenoble, June 2013, 36 p. , http://hal.inria.fr/hal-00924020

[41] R. BLEUSE. , *Utilisation efficace des accélérateurs GPU – Ordonnancement sur machines hybrides*, École Nationale Supérieure d'Informatique et de Mathématiques Appliquées de GrenobleGrenoble, June 2013, 40 p. , http://hal.inria.fr/hal-00858233

[42] S. CURT. , *Calcul haute performance avec OpenCL sur GPU*, Université Joseph Fourier - Grenoble I, June 2013, http://hal.inria.fr/hal-00839341

[43] L. JACQUIN, V. ROCA, J.-L. ROCH. , *ICMP: an Attack Vector against IPsec Gateways*, 2013, http://hal.inria.fr/hal-00879997

[44] A. KUMAR, J.-L. ROCH. , *Algorithm-Based Secure and Fault Tolerant Outsourcing of Matrix Computations*, 2013, page number : 7 , Extended abstract, http://hal.inria.fr/hal-00876156