



IN PARTNERSHIP WITH:
CNRS

**Institut polytechnique de
Grenoble**

**Université Joseph Fourier
(Grenoble)**

**Université Pierre Mendès-France
(Grenoble)**

Activity Report 2015

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

1. Members	1
2. Overall Objectives	2
2.1. Preamble	2
2.2. Objectives	2
3. Research Program	3
3.1. Scheduling	3
3.2. Adaptive Parallel and Distributed Algorithms Design	5
3.3. Interactivity	6
3.3.1. User-in-the-loop	6
3.3.2. Expert-in-the-loop	7
3.4. Adaptive middleware for code coupling and data movements	7
3.4.1. Application Programming Interface	7
3.4.2. Kernel for Asynchronous, Adaptive, Parallel and Interactive Application	8
4. New Software and Platforms	8
4.1. FlowVR	8
4.2. K'Star	8
4.3. KAAPI	9
4.4. KaStORS	9
4.5. LinBox	9
4.6. OAR	9
4.7. QuickCSG	10
4.8. Triva	10
4.9. XKaapi	10
4.10. TakTuk	10
4.11. Platforms	11
4.11.1. Multi-camera Platforms Grimage and Kinovis	11
4.11.2. HPC experimentation platforms: Digitalis, Grid'5000 and Ciment	11
5. New Results	11
5.1. TABARNAC: Tools for Analyzing Behavior of Applications Running on NUMA Architecture.	11
5.2. Computing the Rank Profile Matrix	11
5.3. Parallel Algebraic Linear Algebra Dedicated Interface	12
5.4. Communication models insights meet simulations	12
5.5. Adaptive Resource and Job Management for Limited Power Consumption	12
5.6. Lessons Learned from Building In Situ Coupling Frameworks	12
5.7. Design and analysis of scheduling strategies for multi-CPU and multi-GPU architectures	13
6. Bilateral Contracts and Grants with Industry	13
7. Partnerships and Cooperations	13
7.1. Regional Initiatives	13
7.2. National Initiatives	13
7.2.1. ANR	13
7.2.2. Competitivity Clusters	14
7.2.3. National ADT	14
7.2.4. Inria Project Lab	14
7.3. European Initiatives	15
7.3.1.1. HPC4E	15
7.3.1.2. VISIONAIR	15
7.3.1.3. VELAССo	16
7.3.1.4. GRAIN 2	17

7.4. International Initiatives	17
7.4.1. Inria International Labs	17
7.4.2. Inria Associate Teams not involved in an Inria International Labs	18
7.4.3. Participation In other International Programs	18
7.4.3.1. LICIA	18
7.4.3.2. CAPES/COFECUB StarShip	18
7.5. International Research Visitors	18
7.5.1. Visits of International Scientists	18
7.5.2. Visits to International Teams	19
8. Dissemination	19
8.1. Promoting Scientific Activities	19
8.1.1. Scientific events organisation	19
8.1.2. Scientific events selection	19
8.1.3. Journal	20
8.1.4. Invited talks	20
8.2. Teaching - Supervision - Juries	20
8.2.1. Teaching	20
8.2.2. Supervision	21
8.2.3. Juries	21
9. Bibliography	22

Project-Team MOAIS

Creation of the Project-Team: 2006 January 01, end of the Project-Team: 2015 December 31

Keywords:

Computer Science and Digital Science:

- 1.1.4. - High performance computing
- 1.1.5. - Exascale
- 2.2.3. - Run-time systems
- 2.4. - Reliability, certification
- 7.1. - Parallel and distributed algorithms

Other Research Topics and Application Domains:

- 1.1.2. - Molecular biology

1. Members

Research Scientists

Thierry Gautier [Inria, Researcher, moved to AVALON from June 2015]
Gia Toan Nguyen [Inria, Senior Researcher, until Nov 2015]
Bruno Raffin [Inria, Researcher, HdR]

Faculty Members

Jean-Louis Roch [Team leader, 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]
Denis Trystram [Grenoble-INP, Professor, HdR]
Frédéric Wagner [Grenoble-INP, Associate Professor]

Engineers

Pierre Neyron [CNRS]
Pierrick Brunet [Inria, until Aug 2015]
Youenn Corre [Inria]
Tristan Ezequel [Inria, from Feb 2015]
Benoit Lange [Inria, until Aug 2015]

PhD Students

Marcos Amaris Gonzalez [FAPESP, Univ Sao Paulo, from Oct 2015]
David Beniamine [Univ. Grenoble I]
Raphaël Bleuse [CNRS]
Estelle Dirand [CEA, from Nov 2015]
Matthieu Dreher [Inria, until Feb 2015, granted by ANR EXAVIZ project]
David Glessner [Bull Atos, granted by CIFRE]
Mohammed Khatiri [Univ Mohammed 1er, Oujda, Maroc, from May 2015]
Nicolas Kox [Univ. Grenoble I]
Alessandro Kraemer [Univ du Parana, Brazil]
Fernando Machado Mendonca [Bourse CAPES]
Millian Poquet [Univ. Grenoble I]
Maxime Puy [Univ. Grenoble I]

Valentin Reis [Grenoble-INP]
Adrien Roussel [IFP Energies nouvelles]
Marwa Sridi [CEA]
Abhinav Srivastav [Univ. Grenoble I]
Ziad Sultan [Univ. Grenoble I]
Julio Toss [UFRGS, Brazil]
Philippe Virouleau [Inria, granted by PIA ELCI]

Post-Doctoral Fellows

Damien Dosimont [Univ. Paris VII, until Oct 2015, granted by OSEO Innovation]
Giorgio Lucarelli [Grenoble-INP]

Visiting Scientists

Wafa Nafti [Univ de Tunis, Apr 2015]
Daouda Traore

Administrative Assistant

Annie Simon [Inria]

Other

Christine Guimet [CNRS]

2. Overall Objectives

2.1. Preamble

MOAIS is a project-team of Inria jointly with CNRS (LIG laboratory) and Grenoble Universities (Grenoble-INP, UJF, UPMF). It was created in 2006 as an offspring of the former APACHE project-team, together with MESCAL. Within the Inria theme *Distributed and High Performance Computing*, MOAIS research activities were positively evaluated by the Inria evaluation committee in December 2012 and extended for another 4 years period. During 2015, MOAIS prepared transition to new research projects.

2.2. Objectives

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 these 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 **XKaapi** runtime software that manages the scheduling of multithreaded computations with billions of threads on a virtual architecture with an arbitrary number of resources; XKaapi supports node additions and resilience. XKaapi 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.

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 compromises).

Uncertainties. 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 uncertainties on the scheduling algorithms. There are several ways for dealing with this problem: First, it is possible to design robust algorithms that can optimize 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 uncertain 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 uncertainties. Finally, we are 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 **XKaapi**.

3.2. Adaptive Parallel and Distributed Algorithms Design

Participants: 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 XKaapi (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 kinds 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 benchmarks 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 of 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.

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 who knows the limits of the program that is being executed and who can interact 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. Executing 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 expects.

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 specifies a read frequency on each of the incoming buffers. This approach ensures a strong coherency but has an 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: 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 semantics.

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, or 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 deferred 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, multi-threading, 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. New Software and Platforms

4.1. FlowVR

FUNCTIONAL DESCRIPTION

The goal of the FlowVR library is to provide 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 Web3D. FlowVR enforces a modular programming that leverages software engineering issues while enabling high performance executions on distributed and parallel architectures.

- Participants: Jérémie Allard, Valérie Gourantou, Jean Denis Lesage, Sébastien Limet, Emmanuel Melin, Clément Ménier, Bruno Raffin, Sophie Robert, Matthieu Dreher and Jérémy Jaussaud
- Contact: Bruno Raffin
- URL: <http://flowvr.sf.net>

4.2. K'Star

K'Star Action

KEYWORDS: Parallel computing - Task-based algorithm - Runtime system - Task scheduling - OpenMP - Source-to-source compiler - Data parallelism

FUNCTIONAL DESCRIPTION

The K'Star action supports the development of Klang-Omp, a source-to-source compiler that turns C and C++ programs with OpenMP pragmas to C programs with calls to either the StarPU or the XKaapi runtime system, as well as the development of the KaStORS benchmarks suite for experimenting with OpenMP dependent tasks.

- Participants: Philippe Virouleau, Pierrick Brunet, Thierry Gautier, Olivier Aumage, Samuel Thibault, Nathalie Furmento, Samuel Pitoiset and François Broquedis
- Contact: Thierry Gautier
- URL: <http://kstar.gforge.inria.fr/>

4.3. KAAPI

KAAPI – *Kernel for Adaptive, Asynchronous Parallel and Interactive programming* – is a C library that allows to execute fine/medium grain multithreaded computation with dynamic data flow synchronizations.

FUNCTIONAL DESCRIPTION

KAAPI means Kernel for Adaptive, Asynchronous Parallel and Interactive programming. It is a C library that allows to execute multithreaded computation with data flow synchronization between threads. The library is able to schedule fine/medium size grain programs on a multicore machine with several GPUs. The data flow graph is unfold at runtime.

- Contact: Thierry Gautier
- URL: <http://kaapi.gforge.inria.fr>

4.4. KaStORS

The KaStORS OpenMP Benchmark Suite

KEYWORDS: Benchmarking - HPC - Task-based algorithm - Task scheduling - OpenMP - Data parallelism

FUNCTIONAL DESCRIPTION

The KaStORS benchmarks suite has been designed to evaluate implementations of the OpenMP dependent task paradigm, introduced as part of the OpenMP 4.0 specification.

- Participants: Olivier Aumage, François Broquedis, Pierrick Brunet, Nathalie Furmento, Thierry Gautier, Samuel Thibault and Philippe Virouleau
- Contact: Thierry Gautier
- URL: <http://kastors.gforge.inria.fr/>

4.5. LinBox

FUNCTIONAL DESCRIPTION

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.

- Participants: Clément Pernet and Thierry Gautier
- Contact: Clément Pernet
- URL: <http://linalg.org/>

4.6. OAR

KEYWORDS: HPC - Cloud - Clusters - Resource manager - Light grid

SCIENTIFIC DESCRIPTION

This batch system is based on a database (PostgreSQL (preferred) or MySQL), a script language (Perl) and an optional scalable administrative tool (e.g. Taktuk). It is composed of modules which interact mainly via the database and are executed as independent programs. Therefore, formally, there is no API, the system interaction is completely defined by the database schema. This approach eases the development of specific modules. Indeed, each module (such as schedulers) may be developed in any language having a database access library.

FUNCTIONAL DESCRIPTION

OAR is a versatile resource and task manager (also called a batch scheduler) for HPC clusters, and other computing infrastructures (like distributed computing experimental testbeds where versatility is a key).

- Participants: Olivier Richard, Pierre Neyron, Salem Harrache and Bruno Bzeznik
- Partners: LIG - CNRS - Grid'5000 - CIMENT
- Contact: Olivier Richard
- URL: <http://oar.imag.fr>

4.7. QuickCSG

KEYWORDS: 3D modeling - CAD - 3D reconstruction - Geometric algorithms

FUNCTIONAL DESCRIPTION

QuickCSG is a library and command-line application that computes boolean operations between polyhedra.

- Participants: Matthys Douze, Jean-Sébastien Franco and Bruno Raffin
- Partner: Grenoble-INP
- Contact: Matthys Douze
- URL: <http://pascal.inrialpes.fr/data2/douze/QuickCSG>

4.8. Triva

FUNCTIONAL DESCRIPTION

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.

- Participant: Guillaume Huard
- Contact: Guillaume Huard
- URL: <http://triva.gforge.inria.fr/>

4.9. XKaapi

FUNCTIONAL DESCRIPTION

XKaapi is a library for high performance applications running on multi-cores/multi-processors with support for multi-GPUs. XKaapi provides ABI compliant implementations of libGOMP (GCC runtime for OpenMP) and was one of the target runtime of the K*Star compiler,

- Contact: Thierry Gautier
- URL: <http://kaapi.gforge.inria.fr>

4.10. TakTuk

TakTuk: Adaptive large scale remote executions deployment

KEYWORD: Deployment

FUNCTIONAL DESCRIPTION

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.

- Participants: Guillaume Huard, Pierre Neyron, Benoît Claudel, Johann Bourcier and Olivier Richard
- Partner: LIG
- Contact: Guillaume Huard
- URL: <http://taktuk.gforge.inria.fr/>

4.11. Platforms

4.11.1. Multi-camera Platforms *Grimage and Kinovis*

MOAIS has managed with the LJK-Inria Morpheo team the Grimage platform (<http://grimage.inrialpes.fr>) dedicated to off-line and on-line 3D modeling from multiple cameras and telepresence. In 2012, we received an Equipex funding, Kinovis (<http://kinovis.inrialpes.fr>), to renew this platform with an installation which consists of a significantly larger acquisition space with 68 cameras connected to a new computation cluster. Moais participated to the installation and setup of the new platform (in charge of the network, compute and storage infrastructures) which is operational since summer 2015. FlowVR is the software backbone of both platforms for live processing. MOAIS is participating to the FP7 infrastructure project Visionair to enable European research teams to experiment on both platforms.

4.11.2. HPC experimentation platforms: *Digitalis, Grid'5000 and Ciment*

MOAIS (with MESCAL) develops and operates the Digitalis experimental platform (<http://digitalis.inria.fr>) which hosts local experimental HPC machines as well as a site of the Grid'5000 national testbed and nodes of the CIMENT regional HPC center (mesocentre of the University of Grenoble). MOAIS with the Mescal and LIG-Erods team obtained in 2014 a grant (FAIRE from Grenoble-INP and LIG) to acquire new experimental machines. 2 ARM64 development boards and one last generation 4 CPU machine which can hosts up to 4 accelerators are now part of the platform.

5. New Results

5.1. TABARNAC: Tools for Analyzing Behavior of Applications Running on NUMA Architecture.

In modern parallel architectures, memory accesses represent a common bottleneck. We develop TABARNAC, a tool for analyzing the memory behavior of parallel applications with a focus on NUMA architectures. TABARNAC provides a new visualization of the memory access behavior, focusing on the distribution of accesses by thread and by structure. Such visualization allows the developer to easily understand why performance issues occur. Using TABARNAC, we explain why some applications do not benefit from data and thread mapping. Moreover, we propose several code modifications to improve the memory access behavior of several parallel applications [29].

5.2. Computing the Rank Profile Matrix

We propose the definition of a new matrix invariant, the rank profile matrix, summarizing all information on the row and column rank profiles of all the leading sub-matrices. We also explore the conditions for a Gaussian elimination algorithm to compute all or part of this invariant, through the corresponding PLUQ decomposition [12].

5.3. Parallel Algebraic Linear Algebra Dedicated Interface

We propose a domain specific language based on C/C++ macros, PALADIn (Parallel Algebraic Linear Algebra Dedicated Interface) [15]. This domain specific language allows the user to write C++ code and benefits from sequential and parallel executions on shared memory architectures. With a unique syntax, the user can switch between different parallel runtime systems such as OpenMP, TBB and xKaapi. This interface provides data and task parallelism and has been used for recursion-based parallelization of exact dense linear algebra routines[7].

5.4. Communication models insights meet simulations

It is well-known that taking into account communications while scheduling jobs in large scale parallel computing platforms is a crucial issue. In modern hierarchical platforms, communication times are highly different when occurring inside a cluster or between clusters. Thus, allocating the jobs taking into account locality constraints is a key factor for reaching good performances. However, several theoretical results prove that imposing such constraints reduces the solution space and thus, possibly degrades the performances. In practice, such constraints simplify implementations and most often lead to better results. Our aim in this work is to bridge theoretical and practical intuitions, and check the differences between constrained and unconstrained schedules (namely with respect to locality and node contiguity) through simulations. We have developed a generic tool, using SimGrid as the base simulator, enabling interactions with external batch schedulers to evaluate their scheduling policies. The results confirm that insights gained through theoretical models are ill-suited to current architectures and should be reevaluated [13].

5.5. Adaptive Resource and Job Management for Limited Power Consumption

The last decades have been characterized by an evergrowing requirement in terms of computing and storage resources. This tendency has recently put the pressure on the ability to efficiently manage the power required to operate the huge amount of electrical components associated with state-of-the-art high performance computing systems. The power consumption of a supercomputer needs to be adjusted based on varying power budget or electricity availabilities. As a consequence, Resource and Job Management Systems have to be adequately adapted in order to efficiently schedule jobs with optimized performance while limiting power usage whenever needed. We introduce in this paper a new scheduling strategy that can adapt the executed workload to a limited power budget. The originality of this approach relies upon a combination of speed scaling and node shutdown techniques for power reductions. It is implemented into the widely used resource and job management system SLURM. Finally, it is validated through large scale emulations using real production workload traces of the supercomputer Curie [17].

5.6. Lessons Learned from Building In Situ Coupling Frameworks

Over the past few years, the increasing amounts of data produced by large-scale simulations have motivated a shift from traditional offline data analysis to in situ analysis and visualization. In situ processing began as the coupling of a parallel simulation with an analysis or visualization library, motivated primarily by avoiding the high cost of accessing storage. Going beyond this simple pairwise tight coupling, complex analysis workflows today are graphs with one or more data sources and several interconnected analysis components. In this paper, we review four tools that we have developed to address the challenges of coupling simulations with visualization packages or analysis workflows: Damaris, Decaf, FlowVR and Swift. This self-critical inquiry aims to shed light not only on their potential, but most importantly on the forthcoming software challenges that these and other in situ analysis and visualization frameworks will face in order to move toward exascale [11]. Besides, focusing on asynchronous In Situ Processing with Gromacs, we have exhibited how to take Advantage of GPUs [25].

5.7. Design and analysis of scheduling strategies for multi-CPU and multi-GPU architectures

In [8], we present a comparison of scheduling strategies for heterogeneous multi-CPU and multi-GPU architectures. We designed and evaluated four scheduling strategies on top of XKaapi runtime: work stealing, data-aware work stealing, locality-aware work stealing, and Heterogeneous Earliest-Finish-Time (HEFT). On a heterogeneous architecture with 12 CPUs and 8 GPUs, we analysed our scheduling strategies with four benchmarks: a BLAS-1 AXPY vector operation, a Jacobi 2D iterative computation, and two linear algebra algorithms Cholesky and LU. We conclude that the use of work stealing may be efficient if task annotations are given along with a data locality strategy. Furthermore, our experimental results suggests that HEFT scheduling performs better on applications with very regular computations and low data locality.

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral Contracts with Industry

6.1.1. CEA

Thanks to past collaboration with CEA, XKaapi was used for multi-core version of EPX. We have a contract with CEA [2014-2015] to manage transition from XKaapi to OpenMP as well as specific loop scheduling among hierarchical NUMA architecture.

7. Partnerships and Cooperations

7.1. Regional Initiatives

- MOAIS participates to the creation of an Alpine Multidisciplinary Network on CYbersecurity Studies (AMNECYS). The academic teams and laboratories participating in this project have already developed great expertise on encryption technologies, vulnerabilities analysis, software engineering, protection of privacy and personal data, international & European aspects of cybersecurity. The first project proposal (ALPEPIC ALPs-Embedded security: Protecting IoT & Critical infrastructure) focuses on the protection of the Internet of Things (IoT) and Critical Infrastructure (CI). Leader: CESICE, UPMF (Théodore Cristakis). Partners: Inria/Privatics and LIG/Moais, Gipsa-lab, LJK, Institut Fourier, TIMA, Ve´rimag, LISTIC (Pole MSTIC)

7.2. National Initiatives

7.2.1. ANR

- **ANR grant MOEBIUS (2013-2015)**. Multi-objective scheduling for large computing platforms. Coordinator: Grenoble-INP (Moais team). Partners: Grenoble-INP, Inria, BULL.
- **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)

7.2.2. Competitvity Clusters

- **SoC-Trace**, Minalogic 2011-2015 contract. This project aims the development of tools for the monitoring and debug of munticore systems on chip. Leader: ST-Microelectronic. 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.
- **ARAMIS, PIA contract n°P3342-146798 (2014-2017)**: Architecture Robuste pour les Automates et Matériels des Infrastructures Sensibles. Coordinator: ATOS-WorldGrid; Partners: CEA, SecLab, UJF. The UJF gathers the folowing teams: LIG (Moais, Drakkar, Vasco); LJK (Casys); IF; Verimag (DCS). BPI funds UJF with 775 ke (funds 4 PhDs and 5 years eginers), among which 410ke for LIG. Moais co-advises two PhD Thesis: Nicolas Kox with LIG-VASCO team (Protocol firewall with security guarantees for control-command systems); Maxime Puy with VERIMAG-DCS (Generation of certified filters for control-command systems).
- **PIA ELCI (2014-2017)**. Environnement Logiciel pour le Calcul Intensif. Coordinator BULL. Partners: BULL, CEA, Inria, SAFRAB, UVSQ.

7.2.3. 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-CPU multi-GPUs by using XKaapi and StarPU runtimes.

7.2.4. Inria Project Lab

7.2.4.1. C2S@Exa - Computer and Computational Sciencs 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.

7.3. European Initiatives

7.3.1. FP7 & H2020 Projects

7.3.1.1. HPC4E

Title: HPC for Energy

Programm: H2020

Duration: 2015-2020

Coordinator: Barcelona Supercomputing Center

Inria contact: Stephane Lanteri

The main objective is to develop beyond-the-state-of-the-art high performance simulation tools that can help the energy industry to respond future energy demands and also to carbon-related environmental issues using the state-of-the-art HPC systems. The project also aims at improving the usage of energy using HPC tools by acting at many levels of the energy chain for different energy sources: Exploitation: In wind energy (respond to demand peaks, output prediction) Efficiency: In biomass-derived fuels (develop more efficient and renewable fuels, reduce green-house gas emissions, reduce hydrocarbon dependency and fuel cost) Exploration: In wind energy (resource assessment) and in hydrocarbons (improve available reserves, explore with less financial and environmental risk). Another objective is to improve the cooperation between energy industries from EU and Brazil. The project includes relevant energy industrial partners from Brazil (PETROBRAS) and EU (REPSOL and TOTAL as O&G industries), which will benefit from the project's results. They guarantee that TRL of the project technologies will be very high. A last objective is to improve the cooperation between the leading research centres in EU and Brazil in HPC applied to energy industry. This includes sharing supercomputing infrastructures between Brazil and EU. The cross-fertilization between energy-related problems and other scientific fields will be beneficial at both sides of the Atlantic.

7.3.1.2. VISIONAIR

Title: VISION ADVANCED INFRASTRUCTURE FOR RESEARCH

Programm: FP7

Duration: February 2011 - January 2015

Coordinator: Grenoble-INP

Partners:

Ecole Nationale Supérieure d'arts et Métiers (France)

Université d'Aix Marseille (France)

Consiglio Nazionale Delle Ricerche (Italy)

Cranfield University (United Kingdom)

Ecole Centrale de Nantes (France)

"European Manufacturing and Innovation Research Association, A Cluster Leading Excellence" (Belgium)

Fraunhofer-Gesellschaft Zur Foerderung Der Angewandten Forschung E.V (Germany)
Institut Polytechnique de Grenoble (France)
Inpg Entreprise (France)
Kungliga Tekniska Hoegskolan (Sweden)
Politecnico di Milano (Italy)
Instytut Chemii Bioorganicznej Polskiej Akademii Nauk (Poland)
Poznan University of Technology (Poland)
Rheinisch-Westfaelische Technische Hochschule Aachen (Germany)
Magyar Tudomanyos Akademia Szamitastechnikai Es Automatizalasi Kutato Intezet (Hungary)
Technion - Israel Institute of Technology. (Israel)
University College London (United Kingdom)
University of Essex (United Kingdom)
Technische Universitaet Kaiserslautern (Germany)
University of Patras (Greece)
Universitaet Stuttgart (Germany)
University of Bristol (United Kingdom)
The University Offord (United Kingdom)
Universite de la Mediterranee'd'aix-Marseille Ii (France)
Universiteit Twente (Netherlands)
"fundacio Privada I2cat, Internet I Innovacio Digital A Catalunya" (Spain)

Inria contact: G. Dumont

VISIONAIR is a project of creation of a European infrastructure that should be a unique, visible and attractive entry towards high level visualisation facilities. These facilities must be open to the access of a wide set of research communities. By integrating existing facilities, it will create a world-class research infrastructure enabling to conduct frontier research. This integration will provide a significant attractiveness and visibility of the European Research Area. Current scientific challenges concern climate evolution, environmental risks, health, energy, etc. and require the management of more and more complex information. The development of information technologies, the increasing complexity of the information to be handled and analysed, along with the increasing capacities in scientific and engineering simulations, call for the development of increasingly powerful visualisation tools and methods. The Europe Research Area must be able to compete with other big Research Areas when addressing the previously defined challenges. By integrating visualisation facilities with the VISIONAIR project, ERA will be able to answer integrated challenges out of the scope of usually disseminated research teams. Both, physical access and virtual services, will be provided by the infrastructure. A full access to visualisation dedicated software will be organised, while physical access on high level platforms, will be partially (about 20% of global usage) open for other scientists for free on behalf of excellence of submitted projects. The partners of this project propose to build a common infrastructure that would grant access to high level visualisation facilities and resources to researchers. Indeed, researchers from Europe and from around the world will be welcome to carry out research projects using the visualisation facilities provided by the infrastructure. Visibility and attractiveness of ERA will be increased by the invitation of external projects.

7.3.1.3. VELA^{SSCo}

Title: Visualization For Extremely Large-Scale Scientific Computing

Program: STREP (Specific Targeted Research Project)

Duration: January 2014 - December 2016

Coordinator: Centre Internacional de Metodes Numerics en Enginyeria (Spain)

Partners: JOTNE (No.), SINTEF (No.), Fraunhofer IGD (D), ATOS (SP), Univ. Edinburgh (UK)

Inria contact: Toan Nguyen, Bruno Raffin

Abstract: VELaSSCo aims at developing a new concept of integrated end-user visual analysis methods with advanced management and post-processing algorithms for engineering modelling applications, scalable for real-time petabyte level simulations [59]. The interface will enable real-time interrogation of simulation data, generating key information for analysis. Main concerns have to do with handling of large amounts of data of a very specific kind intrinsically linked to geometrical properties; how to store, access, simplify and manipulate billion of records to extract the relevant information; how to represent information in a feasible and flexible way; and how to visualise and interactively inspect the huge quantity of information they produce taking into account end-user's needs. VELaSSCo achieves this by putting together experts with relevant background in Big Data handling, advanced visualisation, engineering simulations, and a User Panel including research centres, SMEs and companies from key European industrial sectors such as aerospace, household products, chemical, pharmaceutical and civil engineering.

7.3.1.4. GRAIN 2

Type: Cooperation

Defi: Transport (incl. Aeronautics)

Instrument: Coordination and Support Action (CSA)

Duration: October 2013 - June 2016

Coordinator: Centre Internacional de Metodes Numerics en Enginyeria, Barcelone (Spain)

Partner: Airbus (Sp), Alenia (I), EADS-IW (F), Rolls-Royce (UK), Ingenia (Sp.), Numeca (B), U. Sheffield (UK), U. Birmingham (UK), CIRA (I), VKI (B), Airbone (NL), Leitat (Sp), Cerfacs (F), U. Cranfield (UK), CAE (CN), GTE (CN), ARI (CN), FAI (CN), ASRI (CN), SAERI (CN), BIAM (CN), ACTRI (CN), BUAA (CN), NPU (CN), PKU (CN), NUAU (CN), ZIU (CN)

Inria contact: Toan Nguyen

Abstract: The main objective of GRAIN2 is to focus its greening activities following the Flight Path 2050 Vision for Aircraft. GRAIN2 will in particular identify innovative R&D methods, tools and HPC environments (supercomputers and GPGPUs) according to the needs of major aeronautical industries to deeper understand the mechanism of engine exhaust emissions, to improve fuel efficiency and environmental performance.

7.4. International Initiatives

7.4.1. Inria International Labs

JLESC

Associate Team involved in the International Lab:

7.4.1.1. ANOMALIES@EXASCALE

Title: Anomalies Detection and Handling towards Exascale Platforms

International Partner (Institution - Laboratory - Researcher):

University of Chicago (United States) - Argonne National Laboratory (ANL) - Franck Cappello

Start year: 2014

See also: <http://anomalies.imag.fr>

The Anomalies@exascale project intends to prospect new scheduling solutions for very large parallel computing platforms. In particular, we consider the new problems related to fault tolerance raising with the developments of exascale platforms. We expect to define new ways to detect both execution failures and more transient performance anomalies. Information gathered from the detectors will then be taken into account by schedulers to implement corrective measures.

7.4.2. Inria Associate Teams not involved in an Inria International Labs

7.4.2.1. ExaSE

Title: Exascale Computing Scheduling and Energy

International Partner (Institution - Laboratory - Researcher):

UFRGS, PUC Minas and UPS (Brazil)

Duration: 2014 - 2016

See also: <https://team.inria.fr/exase/>

The main scientific context of this project is high performance computing on Exascale systems: large-scale machines with billions of processing cores and complex hierarchical structures. This project intends to explore the relationship between scheduling algorithms and techniques and the energy constraints present on such exascale systems.

7.4.3. Participation In other International Programs

7.4.3.1. LICIA

Title: International Laboratory in High Performance and Ubiquitous Computing

International Partner (Institution - Laboratory - Researcher):

UFRGS (Brazil)

Duration: 2011 - 2018

See also: <http://licia-lab.org/>

The LICIA is an International Laboratory and High Performance and Ubiquitous Computing born in 2011 from the common desire of members of Informatics Institute of the Federal University of Rio Grande do Sul and of Laboratoire d'Informatique de Grenoble to enhance and develop their scientific partnership that started by the end of the 1970. LICIA is an International Associated Lab of the CNRS, a public french research institution. It has support from several brazilian and french research funding agencies, such as CNRS, Inria, ANR, European Union (from the french side) and CAPES, CNPq, FAPERGS (from the Brazilian side). Moais is deeply involved in the creation and animation of LICIA. Bruno Raffin is LICIA associate director.

7.4.3.2. CAPES/COFECUB StarShip

Title: Scalable Tools and Algorithms para Resilient, Scalable, Hybrid Interactive Processing

International Partner (Institution - Laboratory - Researcher):

UFRGS (Brazil)

Duration: 2013 - 2016

7.5. International Research Visitors

7.5.1. Visits of International Scientists

- Daouda Traore, Director of Informatics Dept at Seguou Univ., Mali (oct-nov. 2015)

7.5.1.1. Internships

KHATIRI Mohammed

Date: Sep 2015 - Dec 2015

Institution: UMP (Morocco)

Supervisor: Denis Trystram

7.5.2. Visits to International Teams

7.5.2.1. Research stays abroad

- B. Raffin visited the Universidad A Coruña, Spain, from Sept 2015 to Dec 2015.

8. Dissemination

8.1. Promoting Scientific Activities

Moais has an intensive reviewing activity for a large variety of Conferences and Journals we do not detail here.

8.1.1. Scientific events organisation

8.1.1.1. Member of the organizing committee

- IEEE VR 2015 (IEEE Conference on Virtual Reality): Workshop Chair
- HCW'2015 (24th IEEE Heterogeneous Computing Workshop) may 23, Hyderabad, India: Program Chair

8.1.2. Scientific events selection

8.1.2.1. Member of the conference program committee

- LDAV 2015 (IEEE Symposium on Large-Scale Data Analysis and Visualization)
- SEARIS 2015 (Workshop on Software Engineering and Architectures for Realtime Interactive Systems)
- EGVE 2015 (Eurographics Symposium on Virtual Environments)
- EPGV 2015 (Eurographics Symposium on Parallel Rendering and Visualization)
- SVR 2015 ((Symposium on Virtual and Augmented Reality - Brazil)
- HPSC (6th IEEE Internat. conf on High performance Scientific Computing) march 16-20, 2015, Hanoi, Vietnam
- CCgrid (14th IEEE/ACM internat symposium on Cluster, Cloud and Grid Computing), may 4-7, Shenzhen, China
- IPDPS 2015 (26th IEEE International Parallel & Distributed Processing Symposium) may 24-28, Hyderabad, India
- HICOMB'2015 (14th IEEE Internat Workshop on High Performance Computational Biology) may 23, Hyderabad, India
- COMPAS, june 22-july 25, 2015, Lille, France
- ISPDC (14th Internat Symposium on Parallel and Distributed Computing) june 29-july 01, 2015, Limassol, Cyprus
- MISTA (7th conference multidisciplinary internat. Scheduling conference), august 25-28, 2015, Prague, Czech Republic
- Topic chair Euro-Par (Scheduling and load balancing) august 26-28, 2015, Vienna, Austria
- ParCo2015 (International Conference on Parallel Computing) sept. 1-4, Edinburgh, Scotland
- PPAM'15 (11th internat. conf on parallel processing and applied Maths) sept. 6-9, 2015, Krakow, Poland
- EuroMPI, sept. 21-25, 2015, Bordeaux, France
- 27th SBAC-PAD, october 18-21, 2015, Santa catarina, Brazil

- Edu-HPC (Workshop on Education for High-Performance Computing), nov 16, 2015, Austin, USA

8.1.3. Journal

8.1.3.1. Member of the editorial board

- Associate Editor of the Parallel Computing journal PARCO, D. Trystram.
- Member of the Editorial Board of JPDC, D. Trystram.
- Member of the Editorial Board of Computational Methods in Science and Technology, D. Trystram.
- Member of the Editorial Board of ARIMA (revue africaine de recherche en informatique et maths appliquées), D. Trystram
- Member of the Editorial Board of IEEE Trans. Parallel and Distributed Systems TPDS, D. Trystram.

8.1.4. Invited talks

- Jean-Louis Roch: Invited speaker at 7th International Workshop on Parallel Symbolic Computation (PASCO 2015), Bath, UK, July 10-11. Title of the talk "Interactive computations and outsourcing"

8.2. Teaching - Supervision - Juries

8.2.1. 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). It has been refunded in 2015 to build a new Master 2 program "Cybersecurit" that will open in Q3 2016.

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, Grenoble-INP.

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

Licence: Bruno Raffin. 38 hours per year. Engineering school ENSIMAG.

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).

8.2.2. Supervision

David Beniamine (since 2013). Parallelisation Patterns and Scheduling for Real-Time Physics Simulations. Co-advised by Guillaume Huard and Bruno Raffin.

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

Gilles Daviet (since 2013). Parallel Macroscopic Simulations of Fibrous Materials. Co-advised by Florence Bertails-Descoubes and Bruno Raffin.

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.

Adrien Roussel. Multi-CPU/Multi-GPU parallelization of numerical solvers. Co-advised by Jean-Marc Gratien (IFPEN) and Thierry Gautier.

Philippe Virouleau. Scalability of OpenMP runtime. Co-advised by Bruno Raffin, François Broquedis, Raymond Namyst and Thierry Gautier.

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.

Ziad Sultan (since 2012). High-performance algebraic computations. Co-advised Ph.D by Jean-Guillaume Dumas (LJK Lab) and Clément Pernet.

Stefano Mor (2011-2015). Analysis of work-stealing and adaptive algorithms. Join Ph.D with UFRGS, Brazil. Co-advised by Jean-Louis Roch, Nicolas Maillard and Bruno Raffin

Mathieu Dreher (2011-2015). In-Situ Visualization for Molecular Dynamics. Advised by Bruno Raffin.

David Glesser. Energy Aware Resource Management for HPC. Co-advised by Yianis Georgiou (BULL) and Denis Trystram.

Raphaël Bleuse. Affinity Scheduling. Co-advised by Gregory Mounié and Denis Trystram.

Millian Poquet. Enhanced Data Movements for HPC. Co-advised by Pierre-François Dutot and Denis Trystram.

Abhinav Srivastav. Persyval. Multi-objective Scheduling. Co-advised by Oded Maler (VERIMAG) and Denis Trystram.

Fernando Machado Mendonca. Locality Aware Scheduling. Co-advised by Frederic Wagner and Denis Trystram.

Alessandro Kraemer. Scheduling in the Cloud. Join Ph.D with UFPR, Brazil. Co-advised by Olivier Richard, and Denis Trystram.

8.2.3. Juries

- B Raffin: Reviewer of the PhD. of Julien Duchateau, 9th of December 2015, Université du Littoral Côte d'Opale. "Parallélisation de simulations physiques utilisant un modèle de Boltzmann multi-phases et multi-composants en vue d'un épandage de GNL sur sol."
- JL Roch: PhD of Marcelo Alejandro Forets Irurtia, 10th of December 2015, Université de Grenoble : "Marches quantiques et mécanique quantique relativiste".
- JL Roch: PhD of Bastien Violla, 14th of December 2015, Université de Montpellier : "Contributions à l'Algèbre Linéaire Exacte sur Corps Finis et au Chiffrement Homomorphe"
- Denis Trystram reviewer of the PhD of Samia Kouki (14 février 2015)
- Denis Trystram jury member of the HDR of Cécile Murat (12 mai 2015)

- Denis Trystram president of the PhD jury of Nathalie Herr (19 novembre 2015)
- Denis Trystram reviewer of the PhD of Julien Herman (25 novembre 2015)

9. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [1] D. DOSIMONT. *Spatiotemporal Aggregation for Execution Trace Visualization*, Grenoble Université, June 2015, <https://hal.inria.fr/tel-01176440>

Articles in International Peer-Reviewed Journals

- [2] J. BLAZEWICZ, E. PESCH, D. TRYSTRAM, G. ZHANG. *New perspectives in scheduling theory*, in "Journal of Scheduling", 2015, vol. 18, n^o 4, pp. 333–334 [DOI : 10.1007/s10951-014-0377-x], <https://hal.archives-ouvertes.fr/hal-01230289>
- [3] R. BLEUSE, S. KEDAD-SIDHOUM, F. MONNA, G. MOUNIÉ, D. TRYSTRAM. *Scheduling independent tasks on multi-cores with GPU accelerators*, in "Concurrency and Computation: Practice and Experience", April 2015, vol. 27, n^o 6, pp. 1625-1638 [DOI : 10.1002/CPE.3359], <https://hal.archives-ouvertes.fr/hal-01081625>
- [4] M. BOUGERET, P.-F. DUTOT, D. TRYSTRAM, K. JANSEN, C. ROBENEK. *Improved approximation algorithms for scheduling parallel jobs on identical clusters*, in "Journal of Theoretical Computer Science (TCS)", 2015, vol. 600, pp. 70–85 [DOI : 10.1016/J.TCS.2015.07.003], <https://hal.archives-ouvertes.fr/hal-01230293>
- [5] J. BŁAŻEWICZ, S. KEDAD-SIDHOUM, F. MONNA, G. MOUNIÉ, D. TRYSTRAM. *A study of scheduling problems with preemptions on multi-core computers with GPU accelerators*, in "Discrete Applied Mathematics", December 2015, vol. 196, pp. 72-82 [DOI : 10.1016/J.DAM.2015.04.009], <https://hal.archives-ouvertes.fr/hal-01208434>
- [6] J. COHEN, D. CORDEIRO, D. TRYSTRAM. *Coordination mechanisms for decentralized parallel systems*, in "Concurrency and Computation: Practice and Experience", May 2015, vol. 27, n^o 5, pp. 1255–1272 [DOI : 10.1002/CPE.3298], <https://hal.archives-ouvertes.fr/hal-01119320>
- [7] J.-G. DUMAS, T. GAUTIER, C. PERNET, J.-L. ROCH, Z. SULTAN. *Recursion based parallelization of exact dense linear algebra routines for Gaussian elimination*, in "Parallel Computing", November 2015, <https://hal.archives-ouvertes.fr/hal-01084238>
- [8] J. V. FERREIRA LIMA, T. GAUTIER, V. DANJEAN, B. RAFFIN, N. MAILLARD. *Design and analysis of scheduling strategies for multi-CPU and multi-GPU architectures*, in "Parallel Computing", March 2015, vol. 44, pp. 37-52 [DOI : 10.1016/J.PARCO.2015.03.001], <https://hal.inria.fr/hal-01132037>
- [9] L. STANISIC, A. LEGRAND, V. DANJEAN. *An Effective Git And Org-Mode Based Workflow For Reproducible Research*, in "Operating Systems Review", 2015, vol. 49, pp. 61 - 70 [DOI : 10.1145/2723872.2723881], <https://hal.inria.fr/hal-01112795>

International Conferences with Proceedings

- [10] D. BENIAMINE, M. DIENER, G. HUARD, P. O. A. NAVAU. *TABARNAC: Visualizing and Resolving Memory Access Issues on NUMA Architectures*, in "Visual Performance Analysis (VPA), 2015 Second Workshop on", Austin, Texas, United States, November 2015 [DOI : 10.1145/2835238.2835239], <https://hal.inria.fr/hal-01221146>
- [11] M. DORIER, M. DREHER, T. PETERKA, G. ANTONIU, B. RAFFIN, J. M. WOZNIAK. *Lessons Learned from Building In Situ Coupling Frameworks*, in "First Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization", Austin, United States, November 2015 [DOI : 10.1145/2828612.2828622], <https://hal.inria.fr/hal-01224846>
- [12] J.-G. DUMAS, C. PERNET, Z. SULTAN. *Computing the Rank Profile Matrix*, in "ISSAC", Bath, United Kingdom, K. YOKOYAMA (editor), ISSAC 2015, ACM, July 2015, pp. 146–153 [DOI : 10.1145/2755996.2756682], <https://hal.archives-ouvertes.fr/hal-01107722>
- [13] P.-F. DUTOT, M. POQUET, D. TRYSTRAM. *Communication models insights meet simulations*, in "Euro-Par workshops, HeteroPar - Algorithms, Models, and Tools for Parallel Computing on Heterogeneous Platforms, august 24–25", Vienna, Austria, 2015, <https://hal.archives-ouvertes.fr/hal-01230288>
- [14] P.-F. DUTOT, E. SAULE, A. SRIVASTAV, D. TRYSTRAM. *Online non-preemptive scheduling to optimize stretch*, in "MAPSP, 12th Workshop on Models and Algorithms for Planning and Scheduling Problems", La Roche-en-Ardenne, Belgium, 2015, <https://hal.archives-ouvertes.fr/hal-01230284>
- [15] T. GAUTIER, J.-L. ROCH, Z. SULTAN, B. VIALLA. *Parallel Algebraic Linear Algebra Dedicated Interface*, in "PASCO: International Workshop on Parallel Symbolic Computation", Bath, United Kingdom, July 2015 [DOI : 10.1145/2790282.2790286], <https://hal.inria.fr/hal-01221106>
- [16] Y. GEORGIU, D. GLESSER, K. RZADCA, D. TRYSTRAM. *A Scheduler-Level Incentive Mechanism for Energy Efficiency in HPC*, in "CCGrid 2015 - 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing", Shenzhen, China, May 2015, pp. 617–626 [DOI : 10.1109/CCGRID.2015.101], <https://hal.archives-ouvertes.fr/hal-01230295>
- [17] Y. GEORGIU, D. GLESSER, D. TRYSTRAM. *Adaptive Resource and Job Management for Limited Power Consumption*, in "IEEE International Parallel and Distributed Processing Symposium Workshop, IPDPS 2015, May 25–29", Hyderabad, India, 2015, pp. 863–870 [DOI : 10.1109/IPDPSW.2015.118], <https://hal.archives-ouvertes.fr/hal-01230292>
- [18] Y. GEORGIU, D. GLESSER, D. TRYSTRAM. *Improving Job Scheduling by using Machine Learning*, in "SLURM user group, Georges Washington Univ., USA, 15–16 sept", Washington, United States, 2015, <https://hal.archives-ouvertes.fr/hal-01230290>
- [19] S. KEDAD-SIDHOUM, F. MONNA, D. TRYSTRAM. *Scheduling Tasks with Precedence Constraints on Hybrid Multi-core Machines*, in "IPDPSW 2015 - IEEE International Parallel and Distributed Processing Symposium Workshop", Hyderabad, India, May 2015, pp. 27–33 [DOI : 10.1109/IPDPSW.2015.119], <https://hal.archives-ouvertes.fr/hal-01230286>
- [20] G. LUCARELLI, F. MACHADO MENDONCA, D. TRYSTRAM, F. WAGNER. *Contiguity and Locality in Backfilling Scheduling*, in "CCGrid 2015 - 15th IEEE/ACM International Symposium on Cluster, Cloud and

Grid Computing", Shenzhen, China, May 2015, pp. 586–595 [DOI : 10.1109/CCGRID.2015.143], <https://hal.archives-ouvertes.fr/hal-01230294>

- [21] G. LUCARELLI, D. TRYSTRAM, F. WAGNER. *A generic approach for heterogeneous scheduling*, in "ECCO 2015, the 28th Conference of the European Chapter on Combinatorial Optimization, May 28-30", Catania, Italy, 2015, <https://hal.archives-ouvertes.fr/hal-01230287>
- [22] D. TRYSTRAM. *Handling Heterogeneity for Efficient Implementations: A Case Study on Sequence Comparison*, in "IPDPSW 2015 - IEEE International Parallel and Distributed Processing Symposium Workshop", Hyderabad, India, May 2015, pp. 346–349 [DOI : 10.1109/IPDPSW.2015.123], <https://hal.archives-ouvertes.fr/hal-01230285>
- [23] D. TRYSTRAM. *Message from the HCW Program Committee Chair*, in "2015 IEEE International Parallel and Distributed Processing Symposium Workshop, IPDPS 2015, May 25-29", Hyderabad, India, 2015, 4 p. [DOI : 10.1109/IPDPSW.2015.155], <https://hal.archives-ouvertes.fr/hal-01230291>

Conferences without Proceedings

- [24] E. GAUSSIER, D. GLESSER, V. REIS, D. TRYSTRAM. *Improving Backfilling by using Machine Learning to predict Running Times*, in "SuperComputing 2015", Austin, TX, United States, November 2015 [DOI : 10.1145/2807591.2807646], <https://hal.archives-ouvertes.fr/hal-01221186>
- [25] M. L. HERNANDEZ ARIZA, M. DREHER, C. J. BARRIOS-HERNANDEZ, B. RAFFIN. *Asynchronous In Situ Processing with Gromacs: Taking Advantage of GPUs*, in "Latin America High Performance Computing Conference", Petropolis, Brazil, August 2015, <https://hal.inria.fr/hal-01180364>
- [26] V. ROCA, L. JACQUIN, S. FALL, J.-L. ROCH. *New Results for the PTB-PTS Attack on Tunneling Gateways*, in "GreHack 2015", Grenoble, France, Cédric Lauradoux, Florent Autréau, November 2015, <https://hal.inria.fr/hal-01245629>

Scientific Books (or Scientific Book chapters)

- [27] J.-G. DUMAS, J.-L. ROCH, E. TANNIER, S. VARRETTE. *Foundations of Coding: Compression, Encryption, Error-Correction*, Wiley, 2015, 376 p. , <https://hal.archives-ouvertes.fr/hal-00765802>

Research Reports

- [28] D. BENIAMINE, Y. CORRE, D. DOSIMONT, G. HUARD. *Memory Organisation Cartography & Analysis*, Inria Grenoble ; Inria, March 2015, n^o RR-8694, <https://hal.inria.fr/hal-01130478>
- [29] D. BENIAMINE, M. DIENER, G. HUARD, P. O. A. NAVAU. *TABARNAC: Tools for Analyzing Behavior of Applications Running on NUMA Architecture*, Inria Grenoble Rhône-Alpes, Université de Grenoble, October 2015, n^o 8774, 24 p. , <https://hal.inria.fr/hal-01202105>
- [30] M. DOUZE, J.-S. FRANCO, B. RAFFIN. *QuickCSG: Arbitrary and Faster Boolean Combinations of N Solids*, Inria - Research Centre Grenoble – Rhône-Alpes ; Inria, March 2015, n^o RR-8687, <https://hal.inria.fr/hal-01121419>

Other Publications

-
- [31] J.-G. DUMAS, C. PERNET, Z. SULTAN. *Fast Computation of the Rank Profile Matrix and the Generalized Bruhat Decomposition*, December 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01251223>
- [32] B. LANGE, T. NGUYEN. *A Hadoop use case for engineering data*, June 2015, working paper or preprint, <https://hal.inria.fr/hal-01167510>