INRIA International program
Associate Team Final Report 2011

Bruno Raffin

September 12th, 2011

Please name this file "acronymeoftheassociateteam_FinalReport_2011.pdf"
Submit your final report online before 2011, September 30th, on https://international-programs.inria.fr

Name of the Associate Team: DIODE-A

URL of the Associate Team website: http://diodea.imag.fr

A. Scientific report

The Diode-A associated team started in 2006 and was renewed in 2009. In this repport we only focus on the second period (2009-2011).

A1.Did the goals shifted along the completion of the project?

Summarize the initial scientific objectives of the project and explain, if it occurred, the shifts in method and/or goals over the three years

The DIODE-A project associate the INRIA project-teams Mescal and Moais, as well as the Brazilian University UFRGS. This collaboration was initiated more than 13 years ago. The DIODE-A project enabled to pursue and enforce this strong collaboration focused on 3 main topics:

- Performance evaluation and deployment of large scale distributed environments ;
- Adaptive parallel programming ;
- Monitoring and visualization of parallel applications.

This work relies on the experimental grid platforms the partners work on (CIMENT and Gird'5000 on the French side, Clumssy and GBRAMS on the Brazilian side), but also include more theoretical research on scheduling, adaptive parallel algorithms and multi-criteria optimization problems, as well as software like the Kaapi parallel programing based on work-stealing, the grid simulator Sim-Grid, the OAR batch scheduler or the Pajè trace visualization.

The collaboration did not shift from its initial goals, but has been adapted to the changing context of parallel computing. The more significative evolution is the growing importance of GPU for high performance computing. While one PhD student (Brazilian preparing his PhD at Moais) was working on that topic when Diode-A was renewed in 2009, the importance of this topic grew with a master student work this summer 2011 (Julio Toss, Brafitec exchange program between ENSIMAG and UFRGS) and a sandwich PhD, Joao Lima, who started in 2010 after an initial venue as a master student in 2009 in the context of the Diode-A team.

A2. State the scientific results of the Associate Team

- *Advancement beyond the state-of-the-art*
- *Software, technology development and transfer*
- *Knowledge sharing among the partners, training of young researchers*

The Diode-A associated team main outcomes is the training of young researchers and more specifically double degree PhD students (see section B2 and C1 for details).

Everton Hermann (UFRGS student that prepared his PhD at MOAIS) studied the use of work stealing to balance the work on machines with heterogeneous resources, mainly CPUs and GPUs. His work mainly focused on a specific applications: real-time physics simulation in the context of the SOFA software developed at INRIA. Task scheduling is supported by the KAAPI library developed by the MOAIS team. Everton Hermann proposed a parallelization scheme for dynamically balancing work load between multiple CPUs and GPUs. Most tasks have a CPU and GPU implementation, so they can be executed on any processing unit. He relied on a two level scheduling associating a traditional task graph partitioning and a work stealing guided by processor affinity and heterogeneity. These criteria are intended to limit inefficient task migrations between GPUs, the cost of memory transfers being high, and to favor mapping small tasks on CPUs and large ones on GPUs to take advantage of heterogeneity. Experiments show that we can reach speedups of 22 with 4 GPUs and 29 with 4 CPU cores and 4 GPUs. CPUs unload GPUs from small tasks making these GPUs more efficient, leading to a cooperative speedup greater than the sum of the speedups separatly obtained on 4 GPUs and 4 CPUs.

These results set the base for new developments in the context of heterogeneous computing. Joao Lima (UFRGS student, double degree PhD between UFRGS and MOAIS) is currently studying how to extend the approach developed by Everton Hermann to support more general applications. In particular Joao Lima is focusing on the STL algorithms for which we are seeking efficient hybrid CPU/GPU paralllisation. Such algorithms will require to develop new steal policies for this hybrid execution context.

So far the program loaded on the GPU was seen as a black box by the work stealing scheduler. Julio Toss performed his Master internship at MOAIS and studied how to use work stealing to schedule tasks inside the GPU. It relies on the Nvidia GPU architecture that enables task parallelism at the level of the GPU multi-processors. Julio Toss is expected to start a double degree PhD. His work will focused on work stealing in the GPU with a particular attention given to new emerging architectures, the AMD fusion APU and the Intel Mic GPU. Beyond these 2 PhD works, we envision a generalized work stealing scheduler responsible for work assignment to all available cores, the CPU ones as well as the GPU ones and other specialized computing cores that may appear in the future. The Kaapi framework offer the required level of flexibility to experiment such approaches.

Stfano Drimon K. Mr started in 2011 a double degree PhD advised by Jean-Louis Roach and Nicolas Maillard. The goal of this work is to develop a theoretical framework for performance proving of alternative steal policies.

Jean-Franois Méhaut (Mescal) collaborates with Alexandre Carissimi (UFRGS) and Philippe Navaux (UFRGS) on affinity issues. In particular, Jean-Franois Méhaut (Mescal) collaborates with Alexandre Carissimi co-advised the PhD. work of Christiane Pousa. Multi-core platforms with non-uniform memory access (NUMA) design are now a common resource in High Performance Computing. In such platforms, the shared memory is organized in an hierarchical memory subsystem in which the shared memory is physically distributed into several memory

banks. Additionally, these platforms feature several levels of cache memories. Because of such hierarchy, memory access latencies may vary depending on the distance between cores and memories. Furthermore, since the number of cores is considerably high in these machines, concurrent accesses to the same memory banks are performed, degrading bandwidth usage. Memory affinity is a relationship between threads and data of application that describes how threads access data. In order to keep memory affinity a compromise between data and thread placement is then necessary. Christiane Pousa implemented memory affinity mechanisms that can be easily adapted and used in different parallel systems. The approach takes into account the different data structures used in High Performance Scientific Numerical workloads, to provide solutions that can be used in different contexts. All the ideas developed in this research work are implemented within a Framework named Minas (Memory affInity maNAgement Software). This mechanisms was evaluated with OpenMP, Charm++ and OpenSkel. Minas performance was also used and evaluated using several benchmarks and two real world earthquakes simulations (ONDES3D/BRGM, SPECFEM3D/Magique3D).

B. Outcomes of the Associate Team

B1. List the joint papers published by the participants within the realm of the Associate Team

References

[1] Francieli Zanon Boito, Rodrigo Virote Kassick, Philippe Navaux, and Yves Denneulin. A survey on applications' i/o characterization. In *proceeding os the IX Workshop de Processamento Paralelo e Distribuído*, Porto Alegre, 2011.

[2] Francieli Zanon Boito, Rodrigo Virote Kassick, Laércio L. Pilla, Norton Barbieri, Cláudio Schepke, Philippe Navaux, Nicolas Maillard, Yves Denneulin, Carla Osthoff, Pablo Grunmann, Pedro Dias, and Jairo Panetta. I/o performance of a large atmospheric model using pvfs. In *Renpar 20*, 2011.

[3] Leonardo Brenner. *Rseaux d'Automates Stochastiques : Analyse transitoire en temps continu et Algbre tensorielle pour une smantique en temps discret.* PhD thesis, Institut National Polytechnique de Grenoble, September 2009.

[4] Leonardo Brenner, Paulo Fernandes, Jean-Michel Fourneau, and Brigitte Plateau. Modelling Grid5000 point availability with SAN. *Electronic Notes In Theoretical Computer Science*, 232:165–178, March 2009.

[5] Márcio Castro, Luiz Gustavo Fernandes, Christiane Pousa, Jean-Franćois Méhaut, and Marilton S. de Aguiar. NUMA-ICTM: A Parallel Version of ICTM Exploiting Memory Placement Strategies for NUMA Machines. In *PDSEC '09: Proceedings of the 23rd IEEE International Parallel and Distributed Processing Symposium - IPDPS*, Rome, Italy, 2009. IEEE Computer Society.

[6] Márcia C. Cera, Yiannis Georgiou, Olivier Richard, Nicolas Maillard, and Philippe Olivier Alexandre Navaux. Supporting malleability in parallel architectures with dynamic cpusetsmapping and dynamic mpi. In *ICDCN*, pages 242–257, 2010.

[7] Marcia Cristina Cera. *Providing Adaptability to MPI Applications on Current Parallel Architectures.* PhD thesis, UFRGS, 2011.

[8] Daniel Cordeiro, Denis Trystram, and Frdric Wagner. Analysis of multi-organization scheduling algorithms. In *International Conference on Parallel Computing (Euro-Par)*, 2010.

[9] E. Cruz, C. Pousa, M. Alves, A. Carissimi, P. Navaux, and J-F. Méhaut. Using Memory Access Traces to Map Threads and Data on Hierarchical Multi-core Platforms. In *Workshop on Advances on Parallel and Distributed Processing Symposium (APDCM 2011)*, Urbana Champaign, USA, 2011.

[10] E. Cruz, C. Pousa, M. Alves, A. Carissimi, P. Navaux, and J-F. Méhaut. Using Memory Access Traces to Map Threads and Data on Hierarchical Multi-core Platforms. *International Journal on Networking and Computing*, To be published. Extended version of [9].

[11] Afonso Corra de Sales. *Rseaux d'Automates Stochastiques : Gnration de l'espace d'tats atteignables et Multiplication vecteur-descripteur pour une smantique en temps discret.* PhD thesis, Institut National Polytechnique de Grenoble, September 2009.

[12] Bruno Donassolo, Henri Casanova, Arnaud Legrand, and Pedro Velho. Fast and scalable simulation of volunteer computing systems using simgrid. In *Workshop on Large-Scale System and Application Performance (LSAP)*, 2010.

[13] Bruno Donassolo, Arnaud Legrand, and Claudio Geyer. Non-Cooperative Scheduling Considered Harmful in Collaborative Volunteer Computing Environments. In *Proceedings of the 11th IEEE International Symposium on Cluster Computing and the Grid (CCGrid'11)*. IEEE Computer Society Press, may 2011.

[14] Fabrice Dupros, Christiane Pousa, Alexandre Carissimi, and Jean-Franćois Méhaut. Parallel Simulations of Seismic Wave Propagation on NUMA Architectures. In *ParCo'09: International Conference on Parallel Computing (to appear)*, Lyon, France, 2009.

[15] Everton Hermann. *Interactive Physical Simulation on Multi-Core and Multi-GPU Architectures.* PhD thesis, Grenoble INP, June 2010.

[16] Everton Hermann, Bruno Raffin, Franoise Faure, Thierry Gautier, and Jrmie Allard. Multi-GPU and Multi-CPU Parallelization for Interactive Physics Simulations. In *Europar 2010*, September 2010.

[17] Rodrigo Kassick, Francieli Zanon, Nicolas Maillard, Philippe Navaux, Roberto Souto, Haroldo Velho, Bruno Bzeznik, Olivier Richard, Guillermo Berri, and Obidio Rubio-Mercedes. Gbrams-amsud: Latin-american grid for climatology. In *CLCAR*, 2010.

[18] Rodrigo Virote Kassick, Carla Osthoff, Philippe Navaux, Francieli Zanon Boito, Cláudio Schepke, Nicolas Maillard, Matthias Diener, and Yves Denneulin. Trace-based visualization as a tool to understand applications i/o performance. In *proceeding of the SBAC-PAD 2011 - WAMCA 2011 workshop*, 2011.

[19] Philippe Olivier Alexandre Navaux Lucas Mello Schnorr, Guillaume Huard. Towards visualization scalability through time intervals and hierarchical organization of monitoring data. In *CCGRID '09: Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*, pages 428–435, 2009.

[20] Philippe Olivier Alexandre Navaux Lucas Mello Schnorr, Guillaume Huard. Visual mapping of program components to resources representation: a 3d analysis of grid parallel

applications. In *SBAC-PAD '09: Proceedings of the 21st International Symposium on Computer Architecture and High Performance Computing*, 2009.

[21] Brigitte Plateau and A. Sales. Reachable state space generation for structured models which use functional transitions. In *Proceedings of the 6th International Conference on the Quantitative Evaluation of Systems (QEST'09)*, Budapest, Hungary, September 2009. IEEE Computer Society.

[22] C. R. Pousa, M. Castro, J-F. Mhaut, and A. Carissimi. Improving memory affinity of geophysics applications on numa platforms using minas. In *9th International Meeting High Performance Computing for Computational Science (VecPar)*, 2010.

[23] C. R. Pousa, N. Maillard, I. Stangherlini, and J-F. Mhaut. Compiling openmp applications to enhance memory affinity on hierarchical multi-core machines (poster). In *23rd International Workshop on Languages and Compilers for Parallel Computing*, 2010.

[24] C. R. Pousa, J-F. Mhaut, and A.Carissimi. Memory affinity management for numerical scientific applications over multi-core multiprocessors with hierarchical memory. In *PhD Forum of 24th IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2010.

[25] Christiane Pousa, Márcio Castro, Luiz Gustavo Fernandes, Alexandre Carissimi, and Jean-Franćois Méhaut. Memory Affinity for Hierarchical Shared Memory Multiprocessors. In *21st International Symposium on Computer Architecture and High Performance Computing - SBAC-PAD*, São Paulo, Brazil, 2009. IEEE.

[26] Christiane Pousa, Márcio Castro, Luiz Gustavo Fernandes, Fabrice Dupros, Alexandre Carissimi, and Jean-Franćois Méhaut. High Performance Applications on Hierarchical Shared Memory Multiprocessors. In *Colloque d'Informatique: Brésil / INRIA, Coopérations, Avancés et Défis*, São Bento, Brazil, 2009. SBC.

[27] Christiane Pousa Ribeiro. *Contributions on Memory Affinity Management for Hierarchical Shared Memory Multi-core Platforms*. PhD thesis, Grenoble Universit, July 2011.

[28] Lucas Mello Schnorr. *Some Visualization Models applied to the Analysis of Parallel Applications*. PhD thesis, Institut National Polytechnique de Grenoble, October 2009.

[29] Lucas Mello Schnorr, Guillaume Huard, and Philippe O.A. Navaux. Triva: Interactive 3d visualization for performance analysis of parallel applications. *Future Generation Computer Systems*, 26(3):348 – 358, 2010.

[30] Pedro Velho. *Accurate and Fast Simulations of Large-Scale Distributed Computing Systems*. PhD thesis, Grenoble Universit, July 2011.

[31] Thais Webber. *Reducing the Impact of State Space Explosion in Stochastic Automata Networks*. PhD thesis, Pontifcia Universidade do Rio Grande do Sul (PUCRS), March 2009.

B2. List the thesis jointly supervised within the realm of the Associate Team

We have currently five students who are engaged in joint, double degree PhD (co-tutelle). All these students are funded by Brazil (usually CAPES) :

- Joo Vicente Ferreira Lima (advisers: Bruno Raffin, Nicolas Maillard),

- Larcio Pilla (Advisers: Jean-Franois Mhaut, Philippe Navaux),
- Rodrigo Virote Kassick (advisers: Yves Denneulin, Philippe Navaux),
- Francieli Boito Zanon (advisers: Yves Denneulin, Philippe Navaux),
- Stfano Drimon K. Mr (advisers: Jean-Louis Roch, Nicolas Maillard).

Five Brazilian PhD. students defended since 2008. They are coming from UFRGS or co-advised with an UFRGS professor and prepared their Ph.D. in the MOAIS or MESCAL team:

- Christiane Pousa Ribeiro, defended in 2011, advised by Jean Franois Mhaut et Alexandre Carissimi, CAPES grant
- Pedro Velho, defended in 2011, advised by Arnaud Legrand, CAPES grant
- Marcia Cristina Cera, defended in 2011, advised by Philippe Navaux, Nicolas Maillard and Olivier Richard. Spent one year at Mescal with a sandwich grant from CAPES.
- Everton Hermann, defended in 2010, advised by Bruno Raffin et Franois Faure, INRIA Cordi
- Lucas Mello Schnorr, defended in 2009, advised by Guillaume Huard, CAPES grant.

B3. List the conferences or events organized in continuity of the Associate Team

The main event is probably the class that Jean-Marc Vincent (Mescal) gives every year since 2008 at UFRGS. This class, related to performance evaluation, (15h) is part of the UFRGS master program.

C. Assessment of the collaboration

C1. What is your assessment of the collaboration, and its added value for the research conducted within your Inria project-team?

This collaboration is for us an excellent way to have access to well-trained PhD. students. The way these students are involved in the collaboration shifted from a PhD prepared at 100% in Grenoble to double PhD degrees between Grenoble University and UFRGS and students splitting their time between Porto Alegre and Grenoble.

The flexible budget from EA enabled us to have Brazilian master students coming for a 1 or 2 month visit. This proved a very good way to motivate students to pursue double degree PhDs. The following students have been funded by the EA Diode-A in the last years, still during their Master studies, to prepare a future PhD stay: Mrcia Cristina Cer (2007), Rodrigo V. Kassick (2009), Larcio Pilla (2010), Joo V. F. Lima (2010), Stfano D. K. Mr (2011). Of those 5 students, four are now in joint degree PhD, and the 5th (Mrcia C. Cer) has stayed one year in Grenoble in 2009, during her PhD at the UFRGS. Additionally, Rodrigo Kassick, Larcio Pilla and Joo Lima also ca me to Grenoble funded by other mobility programs (Erasmus Mundus, or third-party research projects).

We are also actively involved in the Brafitec exchange program between INP Grenoble and INPG. This program enables the exchange of undergraduate students(about 3-6 students involved every year). Some of these students made their internships in the MOAIS/MESCAL teams:

- Bruno Donassolo has made a first internship 3 years ago, to complete his double degree (ENSIMAG/UFRGS), with Arnaud Legrand, and has continued with a Master degree, co-advised by the same researcher. During his master, he came back to Grenoble, funded by an INRIA internship.

- Jlio Toss has made an internship in the MOAIS project in 2011, with Thierry Gautier, to complete his double degree (ENSIMAG/UFRGS), and is now back in Porto Alegre, where he is advised by Nicolas Maillard. He is a potential candidate for further graduate studies.

- Thiago Presa has made an internship in the MESCAL project with Dereck Condo in 2010, to complete his Bachelor degree at the UFRGS, funded both by the EA and by an INRIA internship.

This long term collaboration enabled us to extend our contacts and collaborations well beyond the Rio Grande do Sul state. We have collaborations with universities from Sao Paulo, Rio or Minas Gerais for instance.

The collaboration is also a good way to enforce the diffusion of our softwares, get people involved in their developments (most of our softwares are open source). It includes contributions to KAAPI, OAR, or Sofa for instance. In particular, a cluster at UFRGS integrated in 2008 the Grid'5000 grid, adopting the Grid'5000 software layer the Mescal team strongly contributed to.

We always tried to maintain different sources of funding for this collaboration. Over the 2006-2008 period, beside the associated team grant we received funding from the CAPES/COFECUB program (2006-2009), the INRIA-CNPQ program (2008-2010) and the STIC-Amsud/CAPES program (2008-2010 and 2011-2014) that involves Argentina, Uruguay, Peru and Chile, the CNRS/Cnpq program (2011-2013).

Two important actions emerged this year (2011):

- The Laboratoire International Associ du CNRS LICIA between the LIG (Laboratoire d'Informatique de Grenoble) and UFRGS created en 2011. The LICIA also involves the UJF, UPMF , INRIA and INP Grenoble. High performance computing is one of the main research axis of this joint laboratory. Yves Denneulin (MESCAL), Jean-Marc Vincent (MESCAL), Philippe Navaux (UFRGS) and Nicolas Maillard (UFRGS) are the LICIA scientific leaders. Bruno Raffin (MOAIS) is part of the scientific committee.

- The European project HPC-GA (FP7 Marie Curie, IRSES) associating INRIA teams from Bordeaux and Grenoble (MOAIS, MESCAL), the BCAM (Spain), UFRGS, UNAM (Mexico) and as associated partners the BRGM, UJF and Genci. This 3 year project (2012-2014) is headed by Jean-Franois Mhaut (Mescal) and Alison Piastri (INRIA Bordeaux).

Also note that our collaboration with UFRGS enabled UFRGS to be involved in the Joint Laboratory for Petascale Computing between INRIA and Urbana-Champaign.

C2. Do you foresee further developments to this collaboration?

We are definitively working to maintain and develop this collaboration that is very important for both the MOAIS and MESCAL teams. We will propose this year a new associated team with UFRGS. The flexibility of the funding is very important in particular to easily involve students. Vincius Garcia Pinto, Daniel Oliveira and Jlio Toss are currently UFRGS Master students that are strong candidates for a future joint PhD. We still have several actions active for the coming years (CNRS/Cnpq, STIC-Amsud/CAPES) and expect the LICIA to become

an important tool to further enforce the collaboration. We are also participating to the answer to the Cnpq/INRIA call "High performance computing and data management driven by highly demanding applications" headed on the French side by Stphane Lanteri.

C3. Additional information and remarks