



UNIVERSITÉ DE REIMS CHAMPAGNE-ARDENNE

INFORMATIQUE

École Doctorale Sciences Technologie Santé

The Way to Exascale: From Theorics to Applied Problems

Vers l'Exascale: Des Problèmes Théoriques aux Problèmes Appliqués

par **Julien Loiseau**

Thèse de doctorat d'Informatique

Sous la direction de : **M. Michaël KRAJECKI, Professeur des Universités**

JURY

| | | |
|-----------------------------|--|--------------|
| ?? | ?? | Rapporteur |
| ?? | ?? | Rapporteur |
| Benjamin Bergen | Docteur-Ingénieur au Los Alamos National Laboratory, USA | Examineur |
| Guillaume Colin de Verdière | Ingénieur au CEA | Examineur |
| Michaël KRAJECKI | Professeur à l'Université de Reims Champagne-Ardenne | Directeur |
| Christophe JAILLET | Maître de Conférence à l'Université de Reims Champagne-Ardenne | Co-directeur |
| François ALIN | Maître de Conférence à l'Université de Reims Champagne-Ardenne | Encadrant |

Acknowledgments

Thanks everyone !

Contents

| | |
|--|-----------|
| Acknowledgements | i |
| Contents | i |
| List of Figures | iv |
| List of Tables | v |
| Introduction | 1 |
| 1 HPC and Exascale | 3 |
| 1.1 Introduction | 3 |
| 1.2 Parallelism | 3 |
| 1.2.1 Flynn taxonomy | 3 |
| 1.2.2 Goals | 3 |
| 1.2.3 Bottlenecks | 3 |
| 1.2.4 Amdahl and Gustafson | 3 |
| 1.3 Hardware | 4 |
| 1.3.1 Classical CPU | 4 |
| 1.3.2 GPU | 4 |
| 1.3.3 FPGA and ASICS | 4 |
| 1.4 Clusters and Exascale | 4 |
| 1.4.1 Benchmarking | 4 |
| 1.4.2 Composition and usage | 4 |
| 1.4.3 Interconnection | 4 |
| 1.5 Languages | 4 |
| 1.5.1 Accelerators | 4 |
| 1.5.2 Runtimes | 4 |
| 1.6 Optimization | 4 |
| 1.6.1 CPU specifications | 4 |
| 1.6.2 GPUs specifications | 4 |
| 1.6.3 Communications | 4 |
| 1.7 Conclusion | 4 |
| 2 Complex systems | 5 |
| 2.1 Introduction | 5 |
| 2.2 Combinatorial problems | 5 |
| 2.2.1 Combinatorial search | 5 |
| 2.2.2 Combinatorial optimization | 5 |
| 2.3 A case study, the Langford problem | 5 |
| 2.4 Complex systems as a benchmark | 5 |
| 2.5 GRAPH500 | 5 |
| 2.6 Conclusion | 5 |
| 3 Application | 7 |
| Conclusion | 7 |
| Annexes | 9 |

| | |
|---------------------|-----------|
| Bibliography | 11 |
|---------------------|-----------|

List of Figures

List of Tables

Introduction

Chapter 1

HPC and Exascale

1.1 Introduction

We only consider homogenous cluster in that work.

1.2 Parallelism

1.2.1 Flynn taxonomy

The Flynn taxonomy presents a hierarchical organization of computation machine.

In this classification [Fly72b], Michael J. Flynn presents the SIMD, SISO, MISD and MIMD. Add table and present some examples of machines

1.2.2 Goals

Speedup

Speedup can be separated into two parts, Latency and Throughput.

1.2.3 Bottlenecks

1.2.4 Amdahl and Gustafson

The Amdahl's [Amd67] law is used to find the theoretical speedup in latency of a program. We can separate a program into two parts, the one that can be executed in parallel and the one that is sequential. And even if we reduce the parallel part to infinite the sequential part will reach 100% of the total time.

1.3 Hardware

1.3.1 Classical CPU

1.3.2 GPU

1.3.3 FPGA and ASICs

1.4 Clusters and Exascale

1.4.1 Benchmarking

TOP500

1.4.2 Composition and usage

1.4.3 Interconnection

1.5 Languages

1.5.1 Accelerators

1.5.2 Runtimes

1.6 Optimization

Memory locality

Vectorization

1.6.1 CPU specifications

1.6.2 GPUs specifications

1.6.3 Communications

1.7 Conclusion

Chapter 2

Complex systems

2.1 Introduction

2.2 Combinatorial problems

2.2.1 Combinatorial search

2.2.2 Combinatorial optimization

2.3 A case study, the Langford problem

Use the articles here for figures and a base report.

2.4 Complex systems as a benchmark

2.5 GRAPH500

Use the articles here too

2.6 Conclusion

Chapter 3

Application

Conclusion

‘

Annexes

Bibliography

- [ABL15] Ali Assarpour, Amotz Barnoy, and Ou Liu. Counting the number of langford skolem pairings. 2015.
- [AC14] Alejandro Arbelaez and Philippe Codognet. A gpu implementation of parallel constraint-based local search. In *Parallel, Distributed and Network-Based Processing (PDP), 2014 22nd Euromicro International Conference on*, pages 648–655. IEEE, 2014.
- [Amd67] Gene M Amdahl. Validity of the single processor approach to achieving large scale computing capabilities. In *Proceedings of the April 18-20, 1967, spring joint computer conference*, pages 483–485. ACM, 1967.
- [BAP13] Scott Beamer, Krste Asanović, and David Patterson. Direction-optimizing breadth-first search. *Scientific Programming*, 21(3-4):137–148, 2013.
- [BG97] Rick Beatson and Leslie Greengard. A short course on fast multipole methods. *Wavelets, multilevel methods and elliptic PDEs*, 1:1–37, 1997.
- [BM06] David A Bader and Kamesh Madduri. Parallel algorithms for evaluating centrality indices in real-world networks. In *Parallel Processing, 2006. ICPP 2006. International Conference on*, pages 539–550. IEEE, 2006.
- [BP07] Ulrik Brandes and Christian Pich. Centrality estimation in large networks. *International Journal of Bifurcation and Chaos*, 17(07):2303–2318, 2007.
- [Bra01] Ulrik Brandes. A faster algorithm for betweenness centrality*. *Journal of mathematical sociology*, 25(2):163–177, 2001.
- [CDPD⁺14] Federico Campeotto, Alessandro Dal Palu, Agostino Dovier, Ferdinando Fioretto, and Enrico Pontelli. Exploring the use of gpus in constraint solving. In *Practical Aspects of Declarative Languages*, pages 152–167. Springer, 2014.
- [CPW⁺12] F. Checconi, F. Petrini, J. Willcock, A. Lumsdaine, A. R. Choudhury, and Y. Sabharwal. Breaking the speed and scalability barriers for graph exploration on distributed-memory machines. In *High Performance Computing, Networking, Storage and Analysis (SC), 2012 International Conference for*, pages 1–12, Nov 2012.
- [DBG014] Andrew Davidson, Sean Baxter, Michael Garland, and John D Owens. Work-efficient parallel gpu methods for single-source shortest paths. In *Parallel and Distributed Processing Symposium, 2014 IEEE 28th International*, pages 349–359. IEEE, 2014.
- [Deh01] Walter Dehnen. Towards optimal softening in 3-D n-body codes: I. minimizing the force error. *Mon. Not. Roy. Astron. Soc.*, 324:273, 2001.
- [DG08] Jeffrey Dean and Sanjay Ghemawat. Mapreduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113, 2008.
- [DTC⁺14] H. Djidjev, S. Thulasidasan, G. Chapuis, R. Andonov, and D. Lavenier. Efficient Multi-GPU Computation of All-Pairs Shortest Paths. In *Parallel and Distributed Processing Symposium, 2014 IEEE 28th International*, pages 360–369, 2014.
- [FDB⁺14] Zhisong Fu, Harish Kumar Dasari, Bradley Bebee, Martin Berzins, and Bradley Thompson. Parallel breadth first search on gpu clusters. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 110–118. IEEE, 2014.

- [Fly72a] M. Flynn. Some computer organizations and their effectiveness. *Computers, IEEE Transactions on*, C-21(9):948–960, Sept 1972.
- [Fly72b] Michael J Flynn. Some computer organizations and their effectiveness. *IEEE transactions on computers*, 100(9):948–960, 1972.
- [Gar56] Martin Gardner. *Mathematics, magic and mystery*. Dover publication, 1956.
- [GJ79] M. R. Garey and D. S. Johnson. *Computer and Intractability*. Freeman, San Francisco, CA, USA, 1979.
- [GLG⁺12] Joseph E Gonzalez, Yucheng Low, Haijie Gu, Danny Bickson, and Carlos Guestrin. Powergraph: Distributed graph-parallel computation on natural graphs. In *Presented as part of the 10th USENIX Symposium on Operating Systems Design and Implementation (OSDI 12)*, pages 17–30, 2012.
- [GSS08] Robert Geisberger, Peter Sanders, and Dominik Schultes. Better approximation of betweenness centrality. In *Proceedings of the Meeting on Algorithm Engineering & Experiments*, pages 90–100. Society for Industrial and Applied Mathematics, 2008.
- [HKK07] Takahiro Harada, Seiichi Koshizuka, and Yoichiro Kawaguchi. Smoothed particle hydrodynamics on gpus. In *Computer Graphics International*, pages 63–70. SBC Petropolis, 2007.
- [HKS00] Z. Habbas, M. Krajecki, and D. Singer. Parallel resolution of csp with openmp. In *Proceedings of the second European Workshop on OpenMP*, pages 1–8, Edinburgh, Scotland, 2000.
- [HKS02] Z. Habbas, M. Krajecki, and D. Singer. Parallelizing Combinatorial Search in Shared Memory. In *Proceedings of the fourth European Workshop on OpenMP*, Roma, Italy, 2002.
- [HVN09] Pawan Harish, Vibhav Vineet, and P. J. Narayanan. Large graph algorithms for massively multi-threaded architectures. *International Institute of Information Technology Hyderabad, Tech. Rep. IIIT/TR/2009/74*, 2009.
- [IABT11] Markus Ihmsen, Nadir Akinci, Markus Becker, and Matthias Teschner. A parallel sph implementation on multi-core cpus. In *Computer Graphics Forum*, volume 30, pages 99–112. Wiley Online Library, 2011.
- [IOS⁺14] Markus Ihmsen, Jens Orthmann, Barbara Solenthaler, Andreas Kolb, and Matthias Teschner. Sph fluids in computer graphics. 2014.
- [Jai05] Christophe Jaillet. *In french: Résolution parallèle des problèmes combinatoires*. Phd, Université de Reims Champagne-Ardenne, France, December 2005.
- [JAO⁺11] John Jenkins, Isha Arkatkar, John D Owens, Alok Choudhary, and Nagiza F Samatova. Lessons learned from exploring the backtracking paradigm on the gpu. In *Euro-Par 2011 Parallel Processing*, pages 425–437. Springer, 2011.
- [JG03] Morris Jette and Mark Grondona. *SLURM : Simple Linux Utility for Resource Management*. U.S. Departement of Energy, June 23, 2003.
- [JHC⁺10] Shuangshuang Jin, Zhenyu Huang, Yousu Chen, Daniel Chavarría-Miranda, John Feo, and Pak Chung Wong. A novel application of parallel betweenness centrality to power grid contingency analysis. In *Parallel & Distributed Processing (IPDPS), 2010 IEEE International Symposium on*, pages 1–7. IEEE, 2010.
- [JK04] Christophe Jaillet and Michaël Krajecki. Solving the langford problem in parallel. In *International Symposium on Parallel and Distributed Computing*, pages 83–90, Cork, Ireland, July 2004. IEEE Computer Society.
- [KFM04] M. Krajecki, O. Flauzac, and P.-P. Merel. Focus on the communication scheme in the middleware confiit using xml-rpc. In *Parallel and Distributed Processing Symposium, 2004. Proceedings. 18th International*, pages 160–, April 2004.
- [Kra99] Michaël Krajecki. An object oriented environment to manage the parallelism of the fiit applications. In *Parallel Computing Technologies*, pages 229–235. Springer, 1999.
- [Kre02] Valdis E Krebs. Mapping networks of terrorist cells. *Connections*, 24(3):43–52, 2002.

- [KWm12] David B Kirk and W Hwu Wen-mei. *Programming massively parallel processors: a hands-on approach*. Newnes, 2012.
- [Lar09] J Larsen. Counting the number of skolem sequences using inclusion exclusion. 2009.
- [LCK⁺10] Jure Leskovec, Deepayan Chakrabarti, Jon Kleinberg, Christos Faloutsos, and Zoubin Ghahramani. Kronecker graphs: An approach to modeling networks. *The Journal of Machine Learning Research*, 11:985–1042, 2010.
- [LGK⁺14] Yucheng Low, Joseph E Gonzalez, Aapo Kyrola, Danny Bickson, Carlos E Guestrin, and Joseph Hellerstein. Graphlab: A new framework for parallel machine learning. *arXiv preprint arXiv:1408.2041*, 2014.
- [LGS⁺09] Christian Lauterbach, Michael Garland, Shubhabrata Sengupta, David Luebke, and Dinesh Manocha. Fast bvh construction on gpus. In *Computer Graphics Forum*, volume 28, pages 375–384. Wiley Online Library, 2009.
- [LLP⁺12] Min-Joong Lee, Jungmin Lee, Jaimie Yejean Park, Ryan Hyun Choi, and Chin-Wan Chung. Qube: a quick algorithm for updating betweenness centrality. In *Proceedings of the 21st international conference on World Wide Web*, pages 351–360. ACM, 2012.
- [MAB⁺10] Grzegorz Malewicz, Matthew H Austern, Aart JC Bik, James C Dehnert, Ilan Horn, Naty Leiser, and Grzegorz Czajkowski. Pregel: a system for large-scale graph processing. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pages 135–146. ACM, 2010.
- [MB14a] Adam McLaughlin and David A Bader. Revisiting edge and node parallelism for dynamic gpu graph analytics. In *Parallel & Distributed Processing Symposium Workshops (IPDPSW), 2014 IEEE International*, pages 1396–1406. IEEE, 2014.
- [MB14b] Adam McLaughlin and David A. Bader. Scalable and high performance betweenness centrality on the GPU. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 572–583. IEEE Press, 2014.
- [MGG15] Duane Merrill, Michael Garland, and Andrew Grimshaw. High-performance and scalable gpu graph traversal. *ACM Transactions on Parallel Computing*, 1(2):14, 2015.
- [Mil99] J.E. Miller. Langford’s problem: <http://dialectrix.com/langford.html>, 1999.
- [Mon74] U. Montanari. Networks of Constraints: Fundamental Properties and Applications to Pictures Processing. *Information Sciences*, 7:95–132, 1974.
- [ND10] John Nickolls and William J. Dally. The GPU Computing Era. *IEEE Micro*, 30(2):56–69, 2010.
- [NVI] NVIDIA. *CUDA Occupancy calculator*. http://developer.download.nvidia.com/compute/cuda/CUDA_Occupancy_calculator.xls.
- [Nvi07] CUDA Nvidia. Compute unified device architecture programming guide. 2007.
- [Nvi08] CUDA Nvidia. Programming guide, 2008.
- [NVI13] NVIDIA. *CUDA C PROGRAMMING GUIDE*, jul 2013. http://docs.nvidia.com/cuda/pdf/CUDA_C.Programming-Guide.pdf.
- [PB11] P Pande and David A Bader. Computing betweenness centrality for small world networks on a gpu. In *15th Annual High Performance Embedded Computing Workshop (HPEC)*, 2011.
- [Pro93] Patrick Prosser. Hybrid algorithms for the constraint satisfaction problem. *Computational intelligence*, 9(3):268–299, 1993.
- [Sim83] James E. Simpson. Langford sequences: perfect and hooked. *Discrete Math*, 44(1):97–104, 1983.
- [SKN10] Jyothish Soman, Kothapalli Kishore, and PJ Narayanan. A fast gpu algorithm for graph connectivity. 2010.
- [SKSQ13] Ahmet Erdem Sariyüce, Kamer Kaya, Erik Saule, and Ümit V Çatalyürek. Betweenness centrality on gpus and heterogeneous architectures. In *Proceedings of the 6th Workshop on General Purpose Processor Using Graphics Processing Units*, pages 76–85. ACM, 2013.

- [Smi00] B. Smith. Modelling a Permutation Problem. In *Proceedings of ECAI'2000, Workshop on Modelling and Solving Problems with Constraints, RR 2000.18*, Berlin, 2000.
- [SN11] J. Soman and A. Narang. Fast Community Detection Algorithm with GPUs and Multicore Architectures. In *Parallel Distributed Processing Symposium (IPDPS), 2011 IEEE International*, pages 568–579, 2011.
- [Spr05] Volker Springel. The cosmological simulation code gadget-2. *Monthly Notices of the Royal Astronomical Society*, 364(4):1105–1134, 2005.
- [SZ11] Zhiao Shi and Bing Zhang. Fast network centrality analysis using GPUs. *BMC Bioinformatics*, 12:149, 2011.
- [Vol10] Vasily Volkov. Better performance at lower occupancy. In *Proceedings of the GPU Technology Conference, GTC*, volume 10, page 16. San Jose, CA, 2010.
- [Wal01] T. Walsh. Permutation problems and channelling constraints. Technical Report APES-26-2001, APES Research Group, January 2001.
- [War13] Michael S Warren. 2hot: an improved parallel hashed oct-tree n-body algorithm for cosmological simulation. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, page 72. ACM, 2013.
- [WDP⁺15] Yangzihao Wang, Andrew Davidson, Yuechao Pan, Yuduo Wu, Andy Riffel, and John D Owens. Gunrock: A high-performance graph processing library on the gpu. In *ACM SIGPLAN Notices*, volume 50, pages 265–266. ACM, 2015.
- [WL15] J. Willcock and A. Lumsdaine. A Unifying Programming Model for Parallel Graph Algorithms. In *Parallel and Distributed Processing Symposium Workshop (IPDPSW), 2015 IEEE International*, pages 831–840, 2015.
- [WS95] Michael S Warren and John K Salmon. A portable parallel particle program. *Computer Physics Communications*, 87(1):266–290, 1995.