Tom Cornebize

R&D Engineer in High Performance Computing

tom.cornebize@gmail.com
cornebize.net

Education

2017 – 2021 | Grenoble (FR)

PhD in Computer Science — **Université Grenoble Alpes**

Under the supervision of Arnaud Legrand.
High Performance Computing: Towards Better Performance Predictions and Experiments

- I present a methodology for predicting the performance of parallel non-trivial applications through simulation. I describe several models for communications and computations, with an increasing complexity. I compare these models through an extensive validation by matching the predictions with real experiments. This validation shows that modeling the spatial and temporal variability of the platform is essential for faithful predictions. As a consequence, predictions require careful sensibility analysis accounting for the uncertainty on the resource models, which I illustrate through several case studies.
- I present the lessons learned while making the numerous experiments required in the first part and how I improved the methodology. I show that measurements can suffer from multiple experimental biases and I explain how some of these biases can be overcome. I also present how I implemented systematic performance non-regression testing, which allowed to detect many significant changes of the platform throughout this thesis.
- Main results published at Cluster'19 conference (CORE **A**) [CLH19] and currently under review for publication in JPDC journal (CORE **A**) [CL21].

2015 – 2017 | Grenoble (FR)

Master's & Engineering Degrees in Computer Science (M1-M2) — **Ensimag**

Graduate specialization in parallel and distributed systems.
Obtained a Master of Science, with the highest honor, ranked $2^{nd}/88$.

2013 – 2015 | Lyon (FR)

Bachelor's Degree in Theoretical Computer Science (L3-M1) — **ENS Lyon**

Undergraduate and postgraduate intensive program in theoretical computer science.
Obtained a Bachelor of Science, with great honor.

2011 – 2013 | Grenoble (FR)

Undergraduate program (L1-L2) — **Université Grenoble Alpes**

Undergraduate program in computer science and mathematics.

Experience

Dec/21 – now | Remote (FR)

HPC R&D Engineer — **Intel**

Work on the performance prediction of MPI application.

Oct/17 – Dec/17 | Chicago (US)

Performance Variability in Supercomputers — **Argonne Laboratory**

Under the supervision of Swann Perarnau.

- Carried experiments and statistical analyses to characterize performance variability.

Feb/17 – Jul/17 | Grenoble (FR)

Efficient Simulation of Large-Scale MPI Applications — **Inria**

Under the supervision of Arnaud Legrand.

- Modified the simulator and the simulated application to enable large scale simulations.
- Outcome: simulate executions several orders of magnitude larger.
- Obtained results for the large scale experiment of a conference article [Hei+17].

May/16 – Aug/16 | Walldorf (DE)

Multicast Communication in SAP HANA — **SAP**

Under the supervision of Norman May.

- Analyzed several multicast algorithms.
- Implemented these algorithms in C++, using HANA codebase.

May/15 – Aug/15 | Grenoble (FR)

Job Isolation in Fat Tree Topologies — **Bull**

Under the supervision of Matthieu Perotin.

- Designed several algorithms to prevent the leak of sensitive information in a cluster.
- Results published in a conference article [PC17].

Jun/13 – Jul/13 | Grenoble (FR)

Monitoring of Distributed Systems — **Université Grenoble Alpes**

Under the supervision of Yliès Falcone.

- Designed an algorithm for decentralized monitoring of distributed systems.
- Results published in a conference article [FCF14].

Teaching

Jan/20 – Mar/20 Grenoble (FR)	**Statistical Modeling and Literate Programming** 20 hours – 3rd year students (L3) in business informatics. Data analysis and visualization in R.	**Université Grenoble Alpes**
Sep/19 – Dec/19 Grenoble (FR)	**Algorithmics and Imperative Programming** 30 hours – 3rd year students (L3) in software engineer school. Computational complexity, correctness proof, data structures.	**Polytech Grenoble**
Sep/18 – Dec/18 Grenoble (FR)	**Principles of Operating Systems** 30 hours – 4th year students (M1) in computer science. Memory allocation, multithreading, synchronization, buffered I/O, performance evaluation.	**Université Grenoble Alpes**
Sep/18 – Dec/18 Grenoble (FR)	**Software Development Basis** 30 hours – 2nd year students (L2) in computer science. Functional & robustness testing, modularization, type abstraction.	**Université Grenoble Alpes**
Jan/18 – May/18 Grenoble (FR)	**Introduction to Python** 64 hours – 1st year students (L1) in earth science. Variables and types, control flow statements, data structures, files.	**Université Grenoble Alpes**

Publications

Conference Articles

[CLH19] Fast and Faithful Performance Prediction of MPI Applications: the HPL Case Study
Cornebize, T.; Legrand, A., and Heinrich, F. C.
URL: `https://hal.inria.fr/hal-02096571`
2019 IEEE International Conference on Cluster Computing (CLUSTER), 2019

[Hei+17] Predicting the Energy Consumption of MPI Applications at Scale Using a Single Node
Heinrich, F. C.; Cornebize, T.; Degomme, A.; Legrand, A.; Carpen-Amarie, A.; Hunold, S.; Orgerie, A.-C., and Quinson, M.
URL: `https://hal.inria.fr/hal-01523608`
2017 IEEE International Conference on Cluster Computing (CLUSTER), 2017

[PC17] Isolating Jobs for Security on High-Performance Fabrics
Perotin, M. and Cornebize, T.
2017 IEEE 3rd International Workshop on High-Performance Interconnection Networks in the Exascale and Big-Data Era (HiPINEB), 2017

[FCF14] Efficient and Generalized Decentralized Monitoring of Regular Languages
Falcone, Y.; Cornebize, T., and Fernandez, J.-C.
URL: `https://hal.archives-ouvertes.fr/hal-00972559`
34th Formal Techniques for Networked and Distributed Systems (FORTE), 2014

Preprints

[CL21] Simulation-based Optimization and Sensibility Analysis of MPI Applications: Variability Matters
Cornebize, T. and Legrand, A.(submitted at JPDC, currently under review)
URL: `https://hal.inria.fr/hal-03141988`

[CL19] DGEMM performance is data-dependent
Cornebize, T. and Legrand, A.
URL: `https://hal.inria.fr/hal-02401760`

Thesis

[Cor21] High Performance Computing : towards better Performance Predictions and Experiments
Cornebize, T.
URL: `https://tel.archives-ouvertes.fr/tel-03328956`
June 2021

[Cor17] Capacity Planning of Supercomputers: Simulating MPI Applications at Scale
Cornebize, T.
URL: `https://hal.inria.fr/hal-01544827`
June 2017

Software Projects

Jul/19 – May/21 **G5K tests** **cornebize.net/g5k_test**

Performance tests on Grid'5000.

- On a near-daily basis, run micro-benchmarks on hundreds of Grid'5000 nodes.
- Statistical tests on several observed metrics (e.g. CPU performance, frequency, power consumption, temperature).
- Detected several events that affected significantly the nodes performance and went unnoticed by both the staff and the users (e.g. cooling issue, faulty memory, power instability, BIOS upgrade).

Aug/18 – May/21 **Peanut** **github.com/Ezibenroc/peanut**

Experiment engine for Grid'5000.

- Experiments written as readable and concise Python scripts.
- Common tasks (e.g. node deployment, resource monitoring) factorized in the engine.
- Experiment data and meta-data automatically collected in an archive.

Mar/18 – May/21 **Pycewise** **github.com/Ezibenroc/pycewise**

Python module to compute a segmented linear regression.
Implemented using a two-phase greedy algorithm. Several objective functions are available:

- Ordinary Least Square,
- Weighted Least Square,
- Logarithmic Least Square.

Feb/17 – now **Contribution to Simgrid** **simgrid.org**

Simulation of distributed computer systems.
Contributed to SMPI, the simulator of MPI applications.

- Implemented several features to improve the simulation efficiency.
- Greatly improved the calibration procedure for modeling an existing platform.

May/16 – now **Contribution to Roaring bitmap** **roaringbitmap.org**

Fast and lightweight set for unsigned 32 bits integers.

- Contributed to CRoaring, the C library.
  - Implemented range constructor, selection and subset queries.
  - Fixed several bugs.
  - Repository: github.com/roaringBitmap/CRoaring
- Developed PyRoaring, a Python wrapper for the C library.
  - Similar API than the built-in Python set, but several orders of magnitude faster.
  - Used the Cython programming language.
  - Extensive tests caught several bugs of the C library.
  - Repository: github.com/Ezibenroc/PyRoaringBitMap

Sep/14 – Dec/14 **Platypus** **askplatyp.us**

Modular and open source question answering framework.

- Developed a question parsing module in Python, with a grammatical approach (Stanford CoreNLP and NLTK libraries).
- Framework currently used and valorized by Lexistems SAS.

Invited Talks

May/21 **Practical Approach to Experiment Management**
Online Datamove Seminar, Inria Grenoble

Oct/20 **Measuring the Duration of a Matrix Product is Harder than you Think**
Online XUG meeting, Inria Rennes

Sep/20 **Performance Prediction of High Performance Linpack: a Long Journey**
Online PC2/HIT Seminar, Paderborn University

Sep/20 **DGEMM Performance is Data-Dependent**
Online 11^{th} JLESC Workshop

May/20 **Detecting Changes on Grid'5000 Machines**
Online Polaris Seminar, Inria Grenoble

Nov/19 **Quality Assurance of a HPC Cluster: Testing for Performance non-Regression**
Grenoble (FR) Journée des utilisateurs GRICAD