

PhD and Engineer in Computer Science and Applied Mathematics

Sebastien Mondet

seb@mondet.org — PGP: ABDE 8FF0 8479 CA0C — <http://seb.mondet.org>.

References available upon request.

Abstract

PhD and Engineer specialized in Statically-typed Functional Programming for large-scale distributed software development.

- ▷ Developed large-scale distributed platforms (“Big Data Genomics”, web-applications, HPC).
- ▷ Used code generation and formal methods for security purposes (post-doc work).
- ▷ Developed qualified avionics software (*DO-178B* standard).
- ▷ Did a cross-domain PhD and postdoc while co-advising several students (3D Geometry, Compression, Networking, Multimedia, Mobile development, Security, etc.).
- ▷ Quick and eager to learn both math and computer science (currently studying [Coq](#)).
- ▷ Proficient in OCaml; very familiar and interested in Haskell, F#, and Scala; a lot of experience with C, C++ and Java.
- ▷ Pragmatic, well organised, flexible, and team-worker; good problem solving skills; strong aptitude for both software development and technical writing.

Work Experience

Apr. 2014 – Present: Computer Scientist at Mount Sinai Health System

Icahn School of Medicine, Department of Genetics and Genomic Science

Software Engineer at the [Hammer Lab](#).

- ▷ *Computational Workflow Management*: Developed software to manage heavy and complex computational workflows for biomedical applications.
 - ▷ Made 99% “in the open” (Apache 2.0 license).
 - ▷ Written in OCaml including a WebUI based-on [js_of_ocaml](#) and TyXML.
 - ▷ Used for the [NCT02721043](#) clinical trial and various other studies.
 - ▷ Includes [Ketrew](#), [Coclobas](#), [Biokepi](#), [Epidisco](#), and related smaller projects.
- ▷ *Development Operations & System Administration*: Developed tools for users to manage their own deployments on various platforms as well as general maintenance of local computing resources.
 - ▷ Enabled the Lab’s work on Google Cloud (Compute and Container *Engines*), AWS (EC2, ECS, S3, Batch), a local Hadoop cluster, and a local LSF [cluster](#).
 - ▷ Wrote the [stratocumulus](#), and then [Secotrec](#) suite of tools, and the more generic [Genspio](#) library.
- ▷ *Outreach*:
 - ▷ Presented the projects and advances at the OCaml 2015 and 2017 conferences and at the 2015 and 2017 Compose conferences.
 - ▷ Attended biology/bioinformatics conferences (e.g. ISMB, the [Bioinformatics Open Source Conference](#) and [Hackathon](#)).
 - ▷ Contributed to the Lab’s [blog](#).
 - ▷ (In progress) write-up of a more formal research paper on the Lab’s computational workflow management stack.
- ▷ *Technology Watch*: Kept up to date on computer science research, formal methods in industry/open-source, and general cryptography and security. Shared with the team through regular written reports, and periodic presentations a.k.a. “lunch & learn” talks.

Sept. 2011 – Mar. 2014: Software Engineer for NYU

Center for Genomics and Systems Biology, Biology Department, New York University

Software engineer (Associate Research Scientist) responsible for all computational aspects of the [Genomics Sequencing Core Facility \(GenCore\)](#).

- ▷ Architected, developed, deployed, documented, and maintained *HITSCORE*: production-quality, fault-tolerant, high-performance [laboratory information management system](#) and preliminary analysis pipeline for Next Generation DNA sequencing.
- ▷ Full software platform, dealing with jobs running on HPC clusters, servers, tracking meta-data about samples and the facility, managing the genomic data of the sequencers (a few expensive terabytes per week); while providing a dynamic web-application for administration, monitoring, and delivering results to the clients ([source](#)).
- ▷ Based on discussions with bioinformaticians and users, HITSCORE was a key contributor to the facility's CSPRO Certification by Illumina. The system has been in production for more than 2 years, with no major bug, and no data loss.
- ▷ Applied *type-theory* and functional programming advanced techniques with OCaml, PBS/Torque, PostgreSQL, Jane St Core suite, the [Ocigen](#) web-framework (with [Js_of_ocaml](#)).
- ▷ Maintained Linux-based servers (Puppet, CentOS).
- ▷ Participated, initiated, and maintained open-source projects (see section "Software Projects").
- ▷ Assist bioinformaticians/biologists with Unix and HPC matters.
- ▷ Attended conferences (IFCP 2012, OCaml CUFP 2012 and 2013, IBM Programming Languages Day 2012) and the 2013 International Summer School on HPC Challenges in Computational Sciences.

Sept. 2009 – Jun. 2011: Post-doc at the University of Oslo

Distributed Multimedia Systems (DMMS) group, University of Oslo, Norway

SIRIUS Project: *Sensing, Adapting and Protecting Pervasive Information Spaces.*

- ▷ Co-advised PhD and Master students on Quality of Information, Distributed Complex Events Processing, and Anomaly Detection, within Sparse Mobile Ad-Hoc Networks, and Resource-Constrained Devices.
- ▷ Worked on protection middleware with focus on *safety and security of implementations* through meta-programming and formal methods (see Sec'2011 article and the Promiwag project).

Oct. 2006 – Jun. 2009: PhD in Computer Science

IRIT (Computer Science Research Institute of Toulouse), University of Toulouse, France

Simulation of large 3D natural scenes: modeling and adaptive streaming.

- ▷ *Supervision:* Prof. Mathias Paulin, Geraldine Morin, Romulus Grigoras (Vortex group).
- ▷ *Research focus:* Server resources optimization, multi-resolution content packetization, compression and progressive modeling of plant models, network measurements, mobile computing, distributed systems.
- ▷ *Software realizations:* *Wadis*, *LibGenCyl*, and *OMAN* (c.f. "Software Projects"). Also involved in the development of "NatSim" a visualization tool for natural scenes (Python, OpenGL/GLSL).
- ▷ *Co-Advising:* Master and Engineering students working on 3D streaming for mobile devices.
- ▷ *Internship:* Three months (2008) at the National University of Singapore, under the supervision of Dr. Wei Tsang Ooi.
- ▷ *Teaching:* Assistant at INP-ENSEEIH (the "Monitorat" French program), labs in C Programming, Geometric Modeling, 3D Rendering, Operating Systems, Data-Bases, Multimedia.
- ▷ *Training:* Communication, Advanced English, Basic First Aid Techniques.
- ▷ *Dissertation:* *Adaptive Modeling and Distribution of Large Natural Scenes*, PhD thesis reviewed by Pr. Stefanie Hahmann and Pr. Eckehard Steinbach, and defended on June 8th, 2009.
- ▷ The thesis received the [Léopold Escande Award 2009](#) of the University of Toulouse.

Jul. 2005 – Sept. 2006: Embedded Software Engineer

Avionics Department, Atos Origin Integration (Toulouse, France)

- ▷ Developed for Airbus (EYY) of embedded air/ground communication software qualified under the DO-178B standard (HOOD design, ANSI C, LynxOS, RTRT).
- ▷ Developed for Airbus (EYT) of avionic networks testing software (ARINC 429, AFDX, UML, C++, wxWidgets).

Feb. – Jun. 2005: Master Internship

Computer Vision Team, IRT - UMR 5505 (Toulouse, France)

Streaming of large point-based 3D scenes, adaptation to resources and navigation.

- ▷ Implemented a streaming client-server system over HTTP, TCP and DCCP; C++ with Qt/OpenGL on GNU/Linux.
- ▷ *Keywords:* Point based 3D, Compression, Adaptive Streaming.
- ▷ *Advisors:* Geraldine Morin and Romulus Grigoras.

Jun. – Jul. 2004: Engineering Internship

Dassault Aviation, (Biarritz, France)

Processing and visualization module for numerical data measured during polymerization in autoclaves.

- ▷ Wrote technical specifications.
- ▷ Developed a C++ application for MS-Windows, and Shell/C scripts for AIX/RS6000.

Publications

Peer-reviewed Articles

- ▷ S. Mondet, I. Alberdi, and T. Plagemann; *Generating Optimised and Formally Checked Packet Parsing Code*. IFIP SEC, 2011 [URL].
- ▷ M. Zhu, S. Mondet, G. Morin, W. T. Ooi, and W. Cheng; *Towards peer-assisted rendering in networked virtual environments*. ACM MM'11, 2011 [URL].
- ▷ W. Cheng, W. T. Ooi, S. Mondet, G. Morin, and R. Grigoras; *Modeling Progressive Mesh Streaming: Does Data Dependency Matter?* ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP) Volume 7, Issue 2, 2011 [URL].
- ▷ P. Kamisiński, S. Mondet, V. Goebel, and T. Plagemann; *Resource-Aware Complex Event Processing for Mobile Ubiquitous Environments*. UbiComp'10; OPPORTUNITY Workshop, 2010 [URL].
- ▷ W. Cheng, S. Mondet, W. T. Ooi, R. Grigoras, and G. Morin; *Network-Aware Streaming of Partially Ordered Media*. IEEE COMSOC MMTC E-letter Volume 5, Number 6, 2010 [URL].
- ▷ A. Doran, S. Mondet, R. Grigoras, G. Morin, W. T. Ooi, and F. Boudon; *A demonstration of MobiTree: progressive 3D tree models streaming on mobile clients*. ACM Multimedia (Technical Demonstration), 2009 [URL].
- ▷ S. Mondet, W. Cheng, G. Morin, R. Grigoras, F. Boudon, and W. T. Ooi; *Compact and progressive plant models for streaming in networked virtual environments*. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP) Volume 5, Issue 3, 2009 [URL].
- ▷ S. Mondet, W. Cheng, G. Morin, R. Grigoras, F. Boudon, and W. T. Ooi; *Streaming of Plants in Distributed Virtual Environments*. 16th ACM international conference on Multimedia, 2008 (**Best Paper Award**) [URL].
- ▷ W. Cheng, W. T. Ooi, S. Mondet, G. Morin, and R. Grigoras; *An Analytical Model for Progressive Mesh Streaming*. 15th ACM international conference on Multimedia, 2007 [URL].

Pre-prints

- ▷ A. Rubinsteyn, J. Kodysh, I. Hodes, S. Mondet, B. A. Aksoy, J. P. Finnigan, N. Bhardwaj, and J. Hammerbacher; *Computational pipeline for the PGV-001 neoantigen vaccine trial*. BioRxiv Preprint, 2017 [URL].

PhD Thesis

- ▷ S. Mondet; *Adaptive Modeling and Distribution of Large Natural Scenes*. PhD Thesis of the University of Toulouse, 2009 (Defended on June 8, 2009; awarded of the Léopold Escande Price 2009) [URL].

Master Thesis

- ▷ S. Mondet; *Mise en ligne de modèles 3D échelonables basés points*. Master Thesis of the INP Toulouse, 2005.

Research Activities

Reviewed for various high-impact computer-science journals and conferences including the ACM Multimedia 2009, 2010, 2011; the ACM Transactions on Multimedia Computing, Communications and Applications; NOSSDAV 2010 (Network and Operating Systems Support for Digital Audio and Video); and the Springer/ACM Multi-Media Systems Journal. Also reviewed grant applications for The Polish Science Foundation and was part of PhD recruitment committees at the University of Oslo.

Education

2006 – 2009: Philosophiæ Doctor in Computer Science

University of Toulouse, France

Thesis: “*Adaptive Modeling and Distribution of Large Natural Scenes*”

2002 – 2005: Master Degree in Computer Science and Applied Mathematics

ENSEEIH (National Polytechnic Institute of Engineering in Electrotechnics, Electronics, Computer Science, Hydraulics and Telecommunications), Toulouse, France

- ▷ Engineer Diploma (French system).
- ▷ Research-oriented Master’s degree on Software Safety and High-Performance Computing.

2000 – 2002: CPGE (“Classes Préparatoires aux Grandes Écoles”, previously known as “Math sup/spé”)

CPGE Louis Barthou, Pau, France

Undergraduate 2 years prestigious program for competitive entrance exams into national engineering schools; *speciality* “Mathematics and Physics”.

Human Languages

- ▷ *French*: native speaker
- ▷ *Spanish*: native speaker
- ▷ *English*: very fluent
- ▷ *German and Norwegian*: basic knowledge

Software Projects

Computer Science Research:

- ▷ **Promiwag**: A code-generation library specialised in packet-parsing code. It generates C or OCaml code on which safety/security properties are *formally proved*. It uses **Why** and **Alt-Ergo** for automatic formal proofs.
- ▷ **Wadis**: *Walk-through DIstant Scenes* is an experimental testbed for Client-Server streaming of 3D scenes. It implements streaming over TCP, UDP, DCCP; uses OpenGL, **SDL**, GNU Triangulated Surfaces Library, 3DS Max file format.
- ▷ **LibGenCyl**: A library for manipulating 3D models of plants represented by Generalized Cylinders. It provides efficient progressive (de)compression, export (SVG, VRML, **OpenAlea**), and OpenGL rendering.
- ▷ **Oman**: A toolkit for traffic generation, measurements, and tunneling toolkit, for networking experiments over TCP, UDP and DCCP. It provides an UDP tunneling system for DCCP on WAN experiments.
- ▷ **Master Thesis Project**: A C++ client-server system for streaming point-based (a.k.a. “splat-based”) 3D scenes. It streamed over HTTP (Apache with CGI), TCP and DCCP. The *Visualization* client was based on PointShop3D’s render engine.

Computational Workflow Automation:

- ▷ **Ketrew**: A workflow engine specialized in complex and convoluted computational workflows.

- ▷ **Coclobas**: A job-scheduler for container-based HPC-like jobs with various backends: Google Container Engine, AWS Batch, and (local) Docker.
- ▷ **OCaml-PBS**: A helper library for dealing with the PBS/Torque scheduler from OCaml.

Bioinformatics Pipelines / Computational Biology:

- ▷ **Biocaml**: A standard library for solving Bioinformatics problems with OCaml.
- ▷ **Biokepi**: A library of Ketrew workflow “nodes” which wrap bioinformatics tools in order to build bigger workflows.
- ▷ **Epidisco**: The “flagship” Biokepi workflow; it is computational pipeline for personalized cancer epitope discovery and peptide vaccine prediction.
- ▷ **Plawireg**: An experimental library for exploring graph-based Genome representations.

DevOps / Resource Deployment:

- ▷ **Secotrec**: A family of tools (i.e. a library and pre-assembled example applications) of automated deployment of computing infrastructure *by the users* as well as related dev-ops tooling (e.g. generation and build of docker containers).
- ▷ **Genspio**: An EDSL to generate very portable POSIX shell scripts/one-liners from OCaml while enjoying a type system and proper syntax rules.

Music:

- ▷ **Vecosek**: An extremely controllable/programmable MIDI sequencer designed for live performances; music & interaction “scenes” are created with an EDSL. The software features a `js_of_ocaml`/TyXML-based GUI.
- ▷ **Misuja**: A low-level “sequencer thread” implemented in C with which one can send/receive MIDI events from an OCaml API (communicating through ring-buffers provided by the Jack API).
- ▷ **Stamifi**: A pure-OCaml library to parse/print Standard MIDI files.
- ▷ **Locoseq**: A real-time midi-sequencer designed for live performance (Jack Audio Connection Kit, LablGTK, MIT License). Project abandoned and *resurrected* as Vecosek.

Demos and Toys:

- ▷ **WebPDB**: A basic protein visualizer based on WebGL, developed as a `js_of_ocaml` demo for the NYC OCaml Meetup.
- ▷ **Habust**: A tool to build and extract software artifacts inside arbitrary QEMU virtual machines; built as a demo/stress-test of Genspio.
- ▷ **Pewolio**: A simple encoder/decoder of arbitrary byte arrays into pronounceable sequences of English words using the PGP Word List.

General Usage Libraries:

- ▷ **Sosa**: A set of APIs (module types) that define what a string of characters should be, and a set of modules and functors that implement them.
- ▷ **Trakeva**: A unified “key/value + transactions” API on top of database libraries (PostgreSQL, Sqlite, or any user provided plugin) with dynamic backend loading.
- ▷ **Pvem**: A module providing simple handling of an error monad type based on polymorphic variants.
- ▷ **Pvem_lwt_unix**: A library high-level operating-system library focussing meaningful abstractions comprehensive error handling.
- ▷ **Bufx**: A module implementing “passive buffered pipes”, in the style of the `Cryptokit`.
- ▷ **Simple_pam**: Extremely simple C binding to check a user’s password on a PAM-enabled operating system.
- ▷ **Atd2cconv**: Code generator from ATD descriptions to OCaml using modules at the `CConv` library.

Personal Activities

Music

Classical/Electric/Bass guitars and drums.

Have played in and/or initiated various bands, in various styles: *Rock*, *Blues*, *Hard rock*, *Funk*, *Electro-jazz*, and *Tribal Grind Core*.

Currently playing in *Cheia De Soul*.

Sports

Taekwondo (*3rd Keup*), Running, Cross-country skiing.

Hobbies

Juggling, Digital Photography, Cinema.