Matthew Rocklin

CONTACT INFORMATION

E-mail: mrocklin@gmail.com *webpage:* matthewrocklin.com

RESEARCH INTERESTS

I build and maintain parallel computing infrastructure for Python's open source ecosystem. This is part of a broader effort to increase accessibility of numerical methods to science and policy practitioners.

EDUCATION

University of Chicago, Chicago, IL

Ph.D, Computer Science **2013**
M.S. Computer Science **2011**

University of California, Berkeley, Berkeley, CA

B.A., Physics, Mathematics, and Astronomy **May 2007**

PROFESSIONAL EXPERIENCE

NVIDIA
Systems Software Manager **2018 - 2020**
Manage a team of engineers to build Dask+GPU accelerated data science stack. Manage open source community relations.

Anaconda Inc
Computational Scientist **2014 - 2018**
Build software and communities to scale Python's data analytics ecosystem

Sandia National Laboratory - Livermore, CA
Postdoctoral Researcher **2013 - 2014**
Computation and analysis of large time evolving networks

UC Berkeley Physics Department - Berkeley, CA
Staff Research Assistant **2007 - 2008**
Wrote software to track movement of biological materials in cells. Also developed biophysics educational tools.

Berkeley Engineering and Research/4D Imaging - Berkeley, CA
Software Developer **2003 - 2005**
Designed and constructed 3d-scanner based on structured light techniques. Founding engineer of a startup engineering company.

SOFTWARE

I coordinate and maintain several libraries within Python's numeric computing ecosystem, particularly around efficient and scalable computing.

I am primarily known for my work on Dask, a library for scalable computing with dynamic task scheduling. Dask combines a high-speed task scheduler with parallel algorithms to scale exisitng Python libraries like Numpy, Pandas, Scikit-Learn, and many others.

More generally though I work with other core developers within the ecosystem to promote the general health and efficiency. I contribute to and maintain dozens of libraries. A more complete record of my contribution is available on GitHub: github.com/mrocklin.

PUBLICATIONS

Today I mostly publish on technical topics online at matthewrocklin.com/blog. Previously I engaged in traditional academic publishing. This page contains references to those works.

Theses

- M. Rocklin, Modular Generation of Scientific Software, 2013, a PhD dissertation.
- M. Rocklin, Uncertainty Quantification and Sensitivity Analysis in Dynamical Systems, 2011, a masters thesis

Papers

- Al-Rfou, Rami, et al *Theano: A Python framework for fast computation of mathematical expressions*, arXiv preprint arXiv:1605.02688 (2016).
- A. Meurer et al *SymPy: symbolic computing in Python*, PeerJ Computer Science 3 (2017): e103.
- M. Rocklin, A. Pinar *On Clustering on Graphs with Multiple Edge Types*, Internet Mathematics, 2012
- E. Constantinescu,V. Zavala, M. Rocklin, S. Lee, and M. Anitescu, *A Computational Framework for Uncertainty Quantification and Stochastic Optimization in Unit Commitment with Wind Power Generation.* IEEE Transactions on Power Systems, 2010.

Conference Proceedings

- M. Rocklin, *Dask: Parallel computation with blocked algorithms and task scheduling*, Proceedings of the 14th Python in Science Conference. 2015.
- M. Rocklin, *Uncertainty Modeling with SymPy Stats* SciPy 2012
- M. Rocklin, A. Pinar, *Computing an Aggregate Edge-Weight Function for Clustering Graphs with Multiple Edge Types.* Algorithms and Models for the Web-Graph, 2010
- M. Rocklin, A. Pinar, *Latent Clustering on Graphs with Multiple Edge Types* Algorithms and Models for the Web-Graph, 2011

Other

- M. Rocklin, A Pinar, *Spectral Generation and Latent Community Structure of Multiweighted Networks*, 2010
- M. Rocklin, E. Constantinescu , *Adjoint Sensitivity Analysis for Wind Power Generation*, 2009
- US Patent 7620209: *Method and apparatus for dynamic space-time imaging system*