WILLIAM MALLARD

wjm@llard.net

EDUCATION

Harvard University — *Cambridge, MA*
PhD, Biochemistry — *8/2016 – 5/2025*

University of California, Berkeley — *Berkeley, CA*
BSc, Engineering Physics — *1/2002 – 5/2007*
Minors in Computer Science and Mathematics

SKILLS

Languages: Python, C, Assembly, Java, Matlab, Simulink, Unix text processing (sed, awk, grep, etc)
Systems: Linux, Xilinx FPGAs, HPC/SLURM, Real-time Signal Processing
Domains: Machine Learning, Computer Vision, Genomics, Scientific Computing

EXPERIENCE

Harvard University — *Cambridge, MA*
Graduate Student Researcher, Molecular and Cellular Biology — *8/2016 – 5/2025*

Led a multi-year research project from conception through publication, discovering a novel regulatory mechanism in bacterial cell division (PLoS One, 2025).

- Designed an AlphaFold pipeline for batch structural modeling, with automated feature extraction and conformational clustering
- Built computer vision pipelines: cell segmentation, mathematical morphology, and computational geometry for automated morphometric analysis
- Developed video analysis pipelines: frame registration, particle tracking, and motion analysis
- Independently authored the manuscript and shepherded it through peer review

Tools: Python, NumPy, Pandas, SciPy, scikit-image, SLURM, Bash

Broad Institute of MIT and Harvard — *Cambridge, MA*
Computational Biologist, Rinn Lab — *2/2013 – 7/2016*

Contributed functional genomics analyses to 13 peer-reviewed studies in stem cell biology and developmental epigenetics.

- Built end-to-end data processing pipelines, handling ∼10TB of data across 20+ projects
- Developed custom toolchains integrating diverse genomics data modalities
- Managed workload scheduling across heterogeneous compute resources, routing jobs based on queue congestion and hardware availability
- Wrote a Python C extension for automated job monitoring and queue management on LSF clusters
- Partnered with domain researchers to translate experimental questions into computational workflows and deliver publication-ready analyses

Tools: Python, NumPy, Pandas, SciPy, scikit-learn, C, LSF, Unix pipelines

Broad Institute of MIT and Harvard — *Cambridge, MA*
Software Engineer, The Cancer Genome Atlas — *4/2012 – 2/2013*

Maintained and extended production genomics pipelines for the Cancer Genome Atlas, contributing to 4 consortium publications (Nature/Cell, 2013–2014).

- Rewrote memory-intensive R scripts into streaming Python, enabling pipelines to scale to thousands of samples
- Identified a Python standard library limitation blocking large-dataset processing; contributed fix to CPython 3.4 in zipfile and shutil (issues #17189, #17201)

Tools: Python, NumPy, R, Bash

Space Sciences Laboratory *Berkeley, CA*
Systems Engineer, Townes Lab *9/2009 – 8/2011*

Designed the digital backend for the Infrared Spatial Interferometer at Mount Wilson Observatory.

- Built a real-time signal processing system handling 6 GSPS data rates on Xilinx Virtex-5
- Implemented synchronized cross-correlation across three FPGAs via 10 Gbps interconnects
- Extended CASPER's DSP library with DSP48E support for FFTs and polyphase filter banks, enabling higher clock rates across the platform
- Wrote placement constraint tooling to meet timing at near-maximum clock rates
- Developed an end-to-end data acquisition pipeline from FPGA fabric through embedded Linux to networked storage

Tools: Simulink, MATLAB, C, Python

Center for Astronomy Signal Processing and Electronics Research *Berkeley, CA*
Systems Engineer, Werthimer Lab *4/2008 – 9/2009*

Contributed to CASPER, an open-source FPGA toolkit for radio astronomy signal processing.

- Developed a 10 Gbps UDP receiver with a synchronized ring buffer; core architecture adopted into PySPEAD, the data transport layer for the Square Kilometre Array
- Extended DSP libraries for real-time spectrometry and correlation
- Deployed an instrument at Arecibo Observatory

Tools: Simulink, MATLAB, C, Python

MusicianLink, Inc. *San Jose, CA*
Software Engineer *6/2007 – 3/2008*

Developed network software for JamLink, a real-time music collaboration device.

- Designed the binary protocol: packet format, session negotiation, master election
- Built a multi-threaded connection manager in Java using raw TCP sockets

Tools: Java

UC Berkeley Clustered Computing *Berkeley, CA*
Unix Systems Administrator *1/2006 – 8/2006*

Administered Unix systems and job schedulers for scientific HPC clusters.

Massachusetts Institute of Technology *Cambridge, MA*
Research Assistant, Chuang Lab *6/2005 – 8/2005*

Built instrumentation to characterize planar ion traps for quantum computation.

- Wrote a programmable pulse sequencer in 1k lines of Scenix microcontroller assembly
- Designed a serial protocol for laptop-to-microcontroller communication
- Fabricated a custom NIM enclosure housing microcontroller and buffer electronics

Tools: Assembly, C, Python

University of California, Berkeley *Berkeley, CA*
Research Assistant, Holzapfel Lab *6/2004 – 5/2005*

Wrote software to thermocycle a helium dilution fridge to cool millimeter-wave detectors to 250mK.

- Rewrote LabView control system from scratch in C++ for robustness and remote deployment
- Interfaced with Linux Comedi drivers for ADC thermometer readout and DIO heater control

Tools: C++, Linux, Comedi