Adam O'Brien, PhD

Mountain View, CA · (603) 266-7012 · obrienadam89@gmail.com
github.com/obrienadam · linkedin.com/in/obrien-adam

Summary

Software Engineer and Computational Scientist with a PhD in Applied Mathematics and 7+ years of experience designing high-performance computing systems, low-latency runtime components, and massive-scale simulation engines. Experienced in **JAX/XLA runtime and ML infrastructure** (data pipelines, embedding serving, distributed training at cross-datacenter scale), **low-latency C++ systems and GPU acceleration** (CUDA, XLA), and large-scale numerical solvers. Proven track record of optimizing core compute infrastructure to drive significant cost savings and throughput increases in production environments.

Work Experience

Google LLC Sunnyvale, CA
Software Engineer – Machine Learning Infrastructure Apr. 2023 – Present

- **Data Infrastructure (PyGrain & JAX Data):** Maintain and scale PyGrain and JAX Data, the foundational data pipeline libraries handling ingestion and processing for JAX-based training workloads globally across Google. Designed and built the PyGrain autotuner, a closed-loop online controller using Universal Scalability Law (USL) and Markovian queueing models to dynamically tune worker concurrency, prefetch depth, and batch parameters, significantly reducing the time to build high-performance data pipelines.
- **JAX CPU Embedding Library:** Designed and built Google's high-performance native C++ embedding library from the ground up, featuring cache-optimized embedding table layouts, lock-free lookup structures, and lookup coalescing, delivering a 2x to 4x throughput improvement over the incumbent TensorFlow-based solution for embedding serving.
- **JAX Parameter Server Training (PST):** Engineered CPU-based distributed training architectures optimized for massive-scale recommendation and ranking models. Optimized parameter server infrastructure to deliver up to a **3x throughput improvement** in dense worker/server topographies by implementing copy-eliding RPC handlers, custom cache-friendly C++ data structures, and low-overhead threading models. Developed a scalable native C++ optimizer library from scratch with horizontally scalable routines tailored for these topologies, and partnered with cross-functional stakeholders (Ads, YouTube, Play, Geo) to successfully deploy PST into production.
- **Megascale XLA:** Contributed to the Megascale XLA team, implementing topology-aware collective communication primitives within the XLA runtime, including hierarchical tree-based collective planning across intra- and inter-datacenter network tiers to enable efficient distributed training at cross-metro scale. Received the internal **"Gold Perfy" Award** for this contribution.

Aurora Innovation Mountain View, CA
Senior Software Engineer – Motion Planning Simulation Oct. 2021 – Apr. 2023

- Architected core high-performance software components in C++ for a highly distributed, massive-scale motion planning simulation engine executing hundreds of thousands of autonomous vehicle scenarios daily.
- Developed dynamic agent behavior routing frameworks, introducing adversarial interaction profiles and real-time lane-changing mechanics to evaluate safety-critical edge cases.
- Created automated validation tools for Foreign Object Debris (FOD) simulations, engineering workflows that dynamically combined real-world sensor logs with synthetic environmental augmentations.
- Discovered and executed sweeping runtime optimizations within the mapping and routing engine, generating audited infrastructure savings of **~$130,000 per month** in cloud compute expenditures.

Siemens PLM Software Lebanon, NH
Advanced Software Engineer – Physics Solvers Jul. 2019 – Oct. 2021

- Maintained and optimized parallel fluid dynamics solvers within the flagship industrial simulation suite **STAR-CCM+**.

– Researched and implemented highly parallelized computational methods (**MPI, OpenMP**) to resolve large-scale, complex partial differential equations on multi-node cluster topographies.

– Researched and developed GPU offloading strategies for core Algebraic Multi-Grid (AMG) linear solvers using **CUDA**.

Education

University of Toronto Toronto, ON, Canada
PhD in Computational Science / Applied Mathematics Sep. 2014 – Jul. 2019

– Thesis: Immersed boundary methods for moving bodies at fluid-fluid-interfaces.

– Authored multiple peer-reviewed publications in numerical and computational journals (˜95 academic citations).

Technical Skills

Languages: C++ (C++11/14/17/20), CUDA, Python, Bash
HPC & ML Infrastructure: JAX, XLA (Accelerated Linear Algebra), PyGrain, TPU/GPU Hardware Acceleration, MPI, Multi-threading
Numerical Software: PETSc, Trilinos, Eigen, BLAS/LAPACK, HYPRE, NumPy, SciPy