

Jonathan C. Beard

CRAZY AWESOME MAD SCIENTIST · COMPUTER SYSTEMS ARCHITECT

✉ jonathan.c.beard@gmail.com | 🌐 www.jonathanbeard.io

“My goal is to make high performance computing easy to use and portable, it’s time to make the hardware work for the software and not the other way around.”

Summary

Jonathan is an experienced researcher, leader, and technical manager. He has a broad educational background and a history of successfully leading diverse teams to accomplish extremely difficult, complex technical tasks.

Education

Washington University in St. Louis

PH.D. COMPUTER SCIENCE

St. Louis, Missouri

Aug. 2010 - Aug. 2015

The Johns Hopkins University

M.S. BIOINFORMATICS

Baltimore, Maryland

May 2006 - May 2010

Louisiana State University

B.S. IN BIOLOGICAL SCIENCES

- *minor(s): history*

Baton Rouge, Louisiana

Aug. 2001 - Dec. 2005

Louisiana State University

B.A. IN INTERNATIONAL STUDIES, CENTRAL ASIAN FOCUS

Baton Rouge, Louisiana

Aug. 2001 - Dec. 2005

Work Experience

Google LLC

TECHNICAL LEAD MANAGER, STAFF PERFORMANCE SYSTEM ARCHITECT

Austin, Texas, USA

Sept. 2022 - current

- Staff system architect on Google’s internal system-on-chip program (gSoC).
- Technical Lead Manager (TLM) for New Platform Introduction - Performance Team. This team is responsible for Google’s forward looking Arm-based system roadmap assessment. This team aligns architecture, uArchitecture, and software expertise to enable rapid evaluation and introduction of new technologies into Google’s fleet.
- Technical Lead Manager of gSoC Perf/TCO optimization team. This team is responsible for the hardware-software Perf/TCO code-sign process, leveraging hardware engineers, software engineers, and commodities managers’ inputs along with multi-variate models to explore optima in Perf/TCO for Google, designing systems from the cores up to the racks for multiple product verticals.
- Technical Lead Manager for Google’s gSoC ML-guided tuning efforts. This team is responsible for designing, implementing, and executing the infrastructure needed to tune SoC parameters for internal customer needs (e.g., core registers, memory controller settings, mesh QoS settings, etc.).
- Technical Lead of Google Specific (name redacted) accelerator IP for gSoC.

Arm Inc.*Austin, Texas, USA*

PRINCIPAL SYSTEM ARCHITECT

May 2021 - Sept. 2022

- Tech lead for scalable processor communication system architecture.
- Senior technical staff for scalable memory system architecture, responsible for forward looking far-memory and interconnect technology development.
- Developed enhancements for Arm's CHI, including those for memory copy operations.
- Responsible for interfacing with major hyperscale customers, incorporating their needs into Arm's IP development.

Arm Inc.*Austin, Texas, USA*

PRINCIPAL RESEARCH ENGINEER

Apr. 2020 - May 2021

- Lead system architecture researcher for accelerator rich, heterogeneous systems. With a particular focus on system architecture for disaggregated memory systems coupled with application accelerators.
- Jonathan directs cross-over research between high-performance computing and data-center applications (e.g., *memcached*, *redis*, packet processing [P4], GROMACS, NAMD, etc.).
- Lead of Project-38 contract at Arm (A joint DoD/DoE/Arm project looking at late-Moore architectural concepts).
- Lead of architectural simulation project for scalable, multi-fidelity simulation. This is a collaborative project across industry and academia including the developers of *gem5* and *SST*
- Mentor to (and funding source of) several academics, fostering industry leading pre-competitive research in scalable systems, specifically: dataflow, queue management, and disaggregated accelerator systems.
- Developed Arm's LS64 architecture, co-authored the gathering-hint instruction.
- Responsible for interfacing with major hyperscale customers, incorporating their needs into Arm's IP development.

Arm Inc.*Austin, Texas, USA*

STAFF RESEARCH ENGINEER

Apr. 2017 - Apr. 2020

- Research lead for system architecture for heterogeneous systems, including early versions of the AArch64 8.7/9.2 accelerator extensions (e.g., *st64* and variants). Member of GenZ-CXL sub-committee. Lead researcher for scalable systems academic engagement, managing researchers from Barcelona Supercomputing Center, The University of Texas, and Georgia Tech through funded student/professor engagements including 12+ funded students, 5 professors.

Arm Inc.*Austin, Texas, USA*

SENIOR RESEARCH ENGINEER

Feb. 2015 - Apr. 2017

- Technical lead for the research of next generation systems architectures focused on scalability and reduced data movement. This work includes virtual memory systems, accelerator integration, low overhead communications, and improving programmability of heterogeneous systems. In previous roles Jonathan has investigated hardware and workloads for compute near data as part of the DOE FastForward-2 Project, as well as programming models for massively parallel distributed computation. Arm representative to Sandia National Labs led Data Movement Dominates DOE Project, developing methodologies to reduce the cost of data movement, improve the utilization of memory relative to computation through a variety of innovative technologies. Developing simulation techniques to deal with massive scale-out issues that arise from the need to simulate high throughput traffic on hardware that doesn't yet exist.

FastData*Los Angeles, CA*

ADVISOR

Nov. 2016 - Dec. 2020

- Advisor and technical consultant.

Washington University in St. Louis*St. Louis, MO, USA*

RESEARCH ASSISTANT

July 2010 - Apr. 2015

- Ph.D. Student at Washington University in St. Louis. I worked as a research assistant under the direction of Dr. Roger Chamberlain in the Stream Based Supercomputing Lab. My research involved using machine learning, queuing theory, signal processing, and control theory to improve the understanding and application of mathematical models for the optimization of stream compute systems.

Washington University in St. Louis*St. Louis, MO, USA*

LECTURER

Jan. 2012 - May 2012

- Co-Instructed CSE 462M (Computer Systems Design), introduction to modern design practices and the use of FPGA hardware prototyping. Students used a commercial CAE/CAD system for VHDL-based design and simulation while designing a computation system. Students focused on the hardware / software co-design of a novel bio-sequence assembly algorithm deployed on a field programmable gate array and multiple multi-core processors.
- course webpage: <http://goo.gl/dPEdC9>

Extracurricular Open Source Software

RaftLib C++ DSL

PRIMARY AUTHOR, MAINTAINER, EVANGELIST

International

May. 2013 - current

- **969 stars, 127 forks, avg. of 46 daily clones from unique users**
- Designed and authored C++ DSL.
- Built asynchronous FIFOs.
- Designed novel instrumentation approaches.
- Applied machine learning, queueing theory, flow-network theory, and heuristics to auto-tuning dataflow/streaming applications.
- Evangelized the benefits of RaftLib, did two presentations at C++Now (2016/2017), **RaftLib has since been to the top of Y Combinator's Hacker News twice.**
- Provide on-demand support for open-source code, provide pay-support on request (e.g., custom FPGA plug-ins), and continue to evolve the DSL / framework for the open-source community.
- <http://raftlib.io>

Additional open source works

PRIMARY AUTHOR

International

2010-current

- constexpr version of Google's Highway Hash algorithm
- LeigNet: Prototype Simulation Composition Framework
- Raft Language
- SHM Inter-process Communication Library, IPC (Linux and Mac compatible)
- List of all open-source contributions

Thesis

Beard, J. C. (2015). *Online Modeling and Tuning of Parallel Stream Processing Systems*. PhD thesis, Department of Computer Science and Engineering, Washington University in St. Louis, link to thesis: <https://bit.ly/387cGgY>

Journal Publications

Li, P., Beard, J. C., and Buhler, J. D. (2017). Deadlock-free buffer configuration for stream computing. *The International Journal of High Performance Computing Applications*, 31(5):441–450

Beard, J. C., Li, P., and Chamberlain, R. D. (2017). RaftLib: a C++ template library for high performance stream parallel processing. *The International Journal of High Performance Computing Applications*, 31(5):391–404

Conference Publications

Wu, Q., Li, R., Beard, J., and John, L. (2024). Blq: Light-weight locality-aware runtime for blocking-less queuing. In *Proceedings of the 33rd ACM SIGPLAN International Conference on Compiler Construction*, CC 2024, page 100–112, New York, NY, USA. Association for Computing Machinery

Wu, Q., Ekanayake, A., Li, R., Beard, J., and John, L. (2022). Spamer: Speculative push for anticipated

- message requests in multi-core systems. In *Proceedings of the 51st International Conference on Parallel Processing (ICPP '22)*, pages 1–12
- Barredo, A., Armejach, A., Beard, J., and Moreto, M. (2021). Planar: a programmable accelerator for near-memory data rearrangement. In *Proceedings of the ACM International Conference on Supercomputing*, pages 164–176
- Lavin, P., Young, J., Vuduc, R., and Beard, J. (2021). Online model swapping for architectural simulation. In *Proceedings of the 18th ACM International Conference on Computing Frontiers*, pages 102–112
- Wu, Q., Beard, J. C., Ekanayake, A., and John, L. (2021). Virtual-Link: A Scalable Multi-Producer, Multi-Consumer Message Queue Architecture for Cross-Core Communication. In *2021 IEEE International Parallel & Distributed Processing Symposium*. IEEE
- Asri, M., Dunham, C., Rusitoru, R., Gerstlauer, A., and Beard, J. (2020). The non-uniform compute device (nucd) architecture for lightweight accelerator offload. In *2020 28th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP)*, pages 38–45. IEEE
- Cabrera, A. M., Chamberlain, R. D., and Beard, J. C. (2019). Multi-spectral Reuse Distance: Divining Spatial Information from Temporal Data. In *The IEEE High Performance Extreme Computing Conference 2019*, HPEC2019
- Barredo, A., Beard, J. C., and Moreto, M. (2019). SPiDRE: Accelerating Sparse Memory Access Patterns. In *Proceedings of the 28th International Conference on Parallel Architectures and Compilation Techniques*, PACT2019. ACM
- Beard, J. C. (2017). The Sparse Data Reduction Engine (SPiDRE): Chopping Sparse Data One Byte at a Time. In *Proceedings of the Second International Symposium on Memory Systems*. ACM
- Beard, J. C., Epstein, C., and Chamberlain, R. D. (2015b). Online Automated Reliability Classification of Queueing Models for Streaming Processing using Support Vector Machines. In *Proceedings of EuroPar 2015 Parallel Processing*, pages 82–93. Springer
- Beard, J. C. and Chamberlain, R. D. (2015a). Run Time Approximation of Non-blocking Service Rates for Streaming Systems. In *Proceedings of the 17th IEEE International Conference on High Performance and Communications*, pages 792–797. IEEE
- Beard, J. C., Epstein, C., and Chamberlain, R. D. (2015a). Automated Reliability Classification of Queueing Models for Streaming Computation using Support Vector Machines. In *Proceedings of the 6th ACM/SPEC international conference on Performance engineering*, ICPE 2015, pages 325–328, New York, NY, USA. ACM
- Beard, J. C. and Chamberlain, R. D. (2013a). Analysis of a Simple Approach to Modeling Performance for Streaming Data Applications. In *Proceedings of the IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*, pages 345–349
- Beard, J. C. and Chamberlain, R. D. (2013b). Use of Simple Analytic Performance Models of Streaming Data Applications Deployed on Diverse Architectures. In *Proceedings of the International Symposium on Performance Analysis of Systems and Software*, pages 138–139

Lancaster, J. M., Wingbermuehle, J. G., Beard, J. C., and Chamberlain, R. D. (2011). Crossing Boundaries in TimeTrial: Monitoring Communications Across Architecturally Diverse Computing Platforms. In *Proceedings of 9th IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*, pages 280–287

Workshop Publications

Dunham, C. and Beard, J. C. (2018). This Architecture Tastes Like Microarchitecture. In *Online Proceedings of the 2nd Workshop on Pioneering Processor Paradigms*, WP3. The 2nd Workshop on Pioneering Processor Paradigms

Beard, J. C. and Randall, J. (2017). Eliminating Dark Bandwidth: a data-centric view of scalable, efficient performance, post-Moore. In *Proceedings of the High Performance Computing Post-Moore (HCPM'17) Workshop*. Lecture Notes in Computer Science

Li, P., Beard, J. C., and Buhler, J. (2015). Deadlock-free Buffer Configuration for Stream Computing. In *Proceedings of Programming Models and Applications on Multicores and Manycores*, PMAM 2015, pages 164–169, New York, NY, USA. ACM

Beard, J. C., Li, P., and Chamberlain, R. D. (2015c). Raftlib: A C++ template library for high performance stream parallel processing. In *Proceedings of Programming Models and Applications on Multicores and Manycores*, PMAM 2015, pages 96–105, New York, NY, USA. ACM

Beard, J. C. and Chamberlain, R. D. (2014). Use of a Levy Distribution for Modeling Best Case Execution Time Variation. In Horvath, A. and Wolter, K., editors, *Computer Performance Engineering*, volume 8721 of *Lecture Notes in Computer Science*, pages 74–88. Springer International Publishing

ArXiv

Beard, J. C. and Chamberlain, R. D. (2015b). Run Time Approximation of Non-blocking Service Rates for Streaming Systems. *arXiv preprint arXiv:1504.00591v2*

Presentations

9th Workshop on Irregular Applications: Architectures and Algorithms

Denver, Colorado

PANEL MEMBER FOR DEBATE ON: "EASY EXPRESSION CAN GO WITH HIGH PERFORMANCE"

Nov. 2019

- An argument against the proposition that abstractions and performance cannot go together.
- link to slides: <https://bit.ly/2RmqaiN>

The International Symposium on Memory Systems

Washington, DC

PANEL MEMBER FOR NEW AND COOL MEMORY TECHNOLOGIES

Oct. 2019

- Introduced several critical problems for accelerator rich systems, system efficiency, and programmability.
- link to slides: <https://bit.ly/2YhCTEV>

Smoky Mountains Computational Science and Engineering Conference

Kingsport, TN

INVITED TALK ON EDGE COMPUTING

Aug. 2019

- What is the edge, where's the edge, and why not talk about the edge with the rest of the system?
- Systems-level concepts that could enable more portable edge computing.
- link to slides: <https://tinyurl.com/qs5xp4u>

Workshop on Memory-Centric High Performance Computing (MCHPC-2018)

Dallas, Texas, USA

PANELIST ON CHALLENGES FOR MEMORY CENTRIC COMPUTING

Nov. 2018

- Introduction to near-memory compute as just another form of an accelerator rich system.
- Introduction to key challenges and some potential avenues of research to solve them
- link to slides: <https://bit.ly/2LnGWdr>

University Corporation for Atmospheric Research Software Engineering Assembly

Boulder, Colorado

SPEAKER: REDUCING DARK BANDWIDTH THROUGH DATA REDUCTION NEAR MEMORY

April 2018

- Dark bandwidth and some ways to get rid of it through near memory gather/scatter.
- link to slides: <https://bit.ly/2reuE0h>

University Corporation for Atmospheric Research Software Engineering Assembly

Boulder, Colorado

INVITED TALK: A VISION FOR DESTRUCTION OF POST-MOORE DISRUPTION

Apr. 2018

- My views on architecture from a biological evolution perspective.
- Technical debt is evolutionary pressure.
- All novel hardware has a huge chance of going extinct, unless we reduce impact of failing.
- link to slides: <https://bit.ly/33RJNli>

The International Symposium on Memory Systems

Washington, DC

PANELIST ON NEW AND COOL MEMORY TECHNOLOGIES

Oct. 2017

- A panel discussion on how to use new memory technologies, challenges, and in general what is new and cool in the field.
- link to slides: <https://goo.gl/rqkAmQ>

CPPNow 2017

Aspen, Colorado

ACCEPTED TALK: RAFTLIB TUTORIAL

May 2017

- Tutorial on a simple, easy to use stream computation library for C++.
- link to slides: <https://bit.ly/33VgOwW>
- link to talk: <https://youtu.be/liQ787fJgmU>

CPPNow 2017

Aspen, Colorado

ACCEPTED TALK: GOOD FIFOs MAKE GOOD NEIGHBORS

May 2017

- FIFO Optimization for Software Engineers
- link to slides: <https://goo.gl/TIRQFG>
- link to talk: <https://youtu.be/pvp-QQD0Xx0>

Supercomputing 2016

Salt Lake City, Utah

INVITED PANELIST FOR FUTURE OF MEMORY TECHNOLOGY FOR EXASCALE AND BEYOND IV

Nov. 2016

- Panel discussion on the future of memory technology for exascale. My take: instead of focusing just on new technologies, lets focus on systems.
- link to slides: <https://goo.gl/mKM0jK>

CPPNow 2016

Aspen, Colorado

ACCEPTED TALK: RAFTLIB TUTORIAL

May 2016

- Tutorial on a simple, easy to use stream computation library for C++.
- link to slides: <https://bit.ly/2OSdyyg>
- link to talk: <https://youtu.be/gghGLDDaROw>

Patents

Message channels

CO-INVENTOR

US 18/446,570
Filed 9 September 2023

Translation hints

CO-INVENTOR

US 12,007,905
Granted 16 April 2024

Multi-channel Q-monitor

CO-INVENTOR

US 11,960,945
Granted 16 April 2024

Checkpoint Saving

CO-INVENTOR

US 11,934,272
Granted 19 March 2024

Virtual stashing

CO-INVENTOR

US 11,841,800
Granted 12 December 2023

Translation for blocks with tightly-coupled interfaces

CO-INVENTOR

US 11,550,585
Granted 10 January 2023

Insert operation - Low contention hardware-accelerated multi-producer single consumer queue

CO-INVENTOR

US 11,614,985
Granted 28 March 2023

Write Clean Conditional

CO-INVENTOR

US 11,445,020
Granted 13 September 2022

Cache stash relay

CO-INVENTOR

U.S. 11,314,645
Granted 26 April 2022

Snoop Logging

CO-INVENTOR

U.S. 11,176,042
Granted 16 November 2021

Fine granular protections

CO-INVENTOR

U.S. 10,909,045
Granted 2 February 2021

Mapping a Distributed Heap onto a Hierarchical Memory System to Enable Efficient Inter-Process Communication

CO-INVENTOR

U.S. 10,901,691
Granted 26 January 2021

Fault tolerant memory system

CO-INVENTOR

U.S. 10,884,850
Granted 5 January 2021

Dynamic SVE Vectorization of Scalar Operations using Dataflow Vectorization

CO-INVENTOR

U.S. 10,620,954
Granted 14 April 2020

An efficient method for Scalable Range Based Coherence Modification	U.S. 10,592,424
CO-INVENTOR	Granted 17 March 2020
The Memory Storm Fabric for Hardware Accelerated, Scalable Virtual Shared Memory	U.S. 10,467,159
CO-INVENTOR	Granted 26 November 2019
Apparatus and method for predicting a redundancy period	U.S. 10,423,510
CO-INVENTOR	Granted 24 September 2019
Method and Apparatus for Two-Layer Copy-on-Write	U.S. 10,565,126
CO-INVENTOR	Granted 18 February 2020
Memory Node Controller	U.S. 10,467,159
CO-INVENTOR	Granted 5 November 2019
Method and Apparatus for Fast Context Cloning in a Data Processing System	U.S. 10,353,826
CO-INVENTOR	Granted 16 July 2019
Memory Address Translation	U.S. 10,489,304
CO-INVENTOR	Granted 26 November 2019
Cache-Based Communication Between Execution Threads of a Data Processing System	U.S. 10,474,575
CO-INVENTOR	Granted 12 November 2019
Fast Address Translation for Virtual Machines	U.S. 10,613,989
CO-INVENTOR	Granted 7 April 2020
Virtual Context Table for Fast Heterogeneous Context Migration	U.S. 10,423,466
CO-INVENTOR	Granted 24 September 2019
Smart Sparse Data Movement Engine for Increasing Utilization of Bandwidth and Cache Lines	U.S. 10,353,601
CO-INVENTOR	Granted 16 July 2019
Efficient Lazy Migration of Virtual Compute Contexts	U.S. 10,552,212
CO-INVENTOR	Granted 4 February 2020
Virtual Context Format for Fast Heterogeneous State Migration	U.S. 10,671,426
CO-INVENTOR	Granted 2 June 2020
Method and Apparatus for Reordering in a Non-Uniform Compute Device	U.S. 10,445,094
CO-INVENTOR	Granted 15 October 2019
Method and Apparatus for Scheduling in a Non-Uniform Compute Device	U.S. 10,552,152
CO-INVENTOR	Granted 4 February 2020

Memory Synchronization Filter

CO-INVENTOR

U.S. 10,067,708

Granted 4 September 2018

Program Committees

2022	Extended Review Committee , 28th IEEE International Symposium on High-Performance Computer Architecture (HPCA)	<i>online</i>
2021	Technical Program Committee Member , IEEE/ACM International Symposium on Computer Architecture (ISCA)	<i>online</i>
2021	Technical Program Committee Member , IEEE/ACM International Symposium on Microarchitecture (MICRO54)	<i>Greece</i>
2020	Technical Program Committee Member , IEEE/ACM International Symposium on Microarchitecture (MICRO53)	<i>Greece</i>
2021	Committee Member , 2021 Smoky Mountains Computational Sciences & Engineering Conference	<i>United States</i>
2020	Committee Member , 2020 Smoky Mountains Computational Sciences & Engineering Conference	<i>United States</i>
2022	Architecture and Networks Member , Supercomputing 2022 Technical Program Committee	<i>United States</i>
2021	Speed Mentorship Chair , Supercomputing 2021 Committee	<i>United States</i>
2021	Architecture and Networks Member , Supercomputing 2021 Technical Program Committee	<i>United States</i>
2020	Architecture and Networks Member , Supercomputing 2020 Technical Program Committee	<i>United States</i>
2019	Architecture and Networks Co-Chair , Supercomputing 2019 Technical Program Committee	<i>United States</i>
2018	Architecture and Networks Member , Supercomputing 2018 Technical Program Committee	<i>United States</i>
2017	Architecture and Networks Member , Supercomputing 2017 Technical Program Committee	<i>United States</i>
2016	Architecture and Networks Member , Supercomputing 2016 Technical Program Committee	<i>United States</i>
2019	General Chair , Arm Research Summit	<i>United States</i>
2018	Organizing Committee Member , Arm Research Summit	<i>United Kingdom</i>
2017	Organizing Committee Member , Arm Research Summit	<i>United Kingdom</i>
2022	Organizing Committee Member , The International Conference on Memory Systems (MEMSYS)	<i>United States</i>
2021	Organizing Committee Member , The International Conference on Memory Systems (MEMSYS)	<i>United States</i>
2021	Organizing Committee Member , The International Conference on Memory Systems (MEMSYS)	<i>United States</i>
2020	Organizing Committee Member , The International Conference on Memory Systems (MEMSYS)	<i>United States</i>
2018	Organizing Committee Member , The International Conference on Memory Systems (MEMSYS)	<i>United States</i>
2017	Organizing Committee Member , The International Conference on Memory Systems (MEMSYS)	<i>United States</i>
2018	Report co-author , US Department of Energy Extreme Heterogeneity Workshop	<i>United States</i>
2022	Technical Program Committee , 9 th Workshop on Irregular Applications: Architectures and Algorithms	<i>United States</i>
2019	Technical Program Committee , 9 th Workshop on Irregular Applications: Architectures and Algorithms	<i>United States</i>
2020	Committee Member , US Department of Energy Performance, Portability, and Productivity (P3HPC)	<i>United States</i>
2019	Committee Member , US Department of Energy Performance, Portability, and Productivity (P3HPC)	<i>United States</i>
2022	Architecture Technical Program Committee , 51 st International Conference on Parallel Processing (ICPP)	<i>United States</i>
2018	Architecture Technical Program Committee , 47 th International Conference on Parallel Processing (ICPP)	<i>United States</i>
2020	Architectures, Networks & Infrastructure Technical Program Committee Member , ISC 2020	<i>Germany</i>
2019	Architectures, Networks & Infrastructure Technical Program Committee , ISC 2019	<i>Germany</i>
2018	Co-general Chair , GoingArm 2018	<i>Germany</i>
2017	Co-general Chair , GoingArm 2017	<i>Germany</i>