Michael **LeBeane**

Computer Architect · Software Engineer

mlebeane@gmail.com | www.mlebeane.com | mlebeane | Michael LeBeane

Summary

I'm a computer architecture Ph.D. with 8+ years of combined industry experience. My most recent work is creating performance models for upcoming GPU graphics and compute specifications. I also have experience performing fundamental research in the areas of GPUs for general purpose computation, off-chip computer networks, and simulation frameworks. I'm always interested in exploring other areas with exciting new problems to solve.

Education

University of Texas at Austin

Austin, TX

Ph.D. and M.S. in Electrical and Computer Engineering **(GPA 3.97/4.0)** *Summer 2018, Spring 2015*

- Microelectronics and Computer Development Fellow

Ph.D. Dissertation

Optimizing Communication in GPU Clusters

With the impending slow-down of Moore's Law, high performance computing has turned to accelerators to continue pushing the performance and power trends of the last 50 years. Chief among proposed accelerator architectures are GPUs, which have already found a comfortable home in datacenter and high-performance computing ecosystems. However, despite rapid adoption, communication between networks of GPUs remains cumbersome, requiring tight coordination with a host CPU to communicate with other systems.

My dissertation describes several techniques to optimize GPU networking in a distributed memory environment. I evaluate hardware and software modifications to both network interface controllers and GPUs to allow end-to-end, user-space communication between networks of GPUs, avoiding critical path CPU interference when possible. These techniques increase both performance and programmability across many important multi-node primitives and critical emerging workloads.

Washington University in St. Louis

St. Louis, MO

B.S. in Computer Engineering and B.S in Computer Science **(GPA 3.95/4.0)** *Spring 2012*

- Graduated Summa Cum Laude
- Served as Teaching Assistant for Computer Science intro courses
- Member of Computer Science Student Advisory Board

Work Experience

Qualcomm

San Diego, CA

Staff Engineer (Graphics Research) *Fall 2020 - Now*

- Developing performance models for upcoming GPU graphics features.
- Contributing to in-development GPU graphics specifications and standards bodies.
- Designing hardware and software solutions to support emerging programming models for Adreno GPUs.
- Providing technical leadership and mentoring to junior engineers and interns.

Advanced Micro Devices

Austin, TX

Member of Technical Staff Design Engineer (Research) *Summer 2019 - Fall 2020*

Senior Engineer (Research) *Winter 2016 - Summer 2019*

Post Graduate Engineer (Research) *Summer 2015 - Winter 2016*

- Researched GPU networking strategies for multiple government-funded projects.
- Wrote open-source GPU networking runtime for AMD's ROCm software stack.
- Contributed to external funding proposals to improve the breadth of AMD's research portfolio.
- Drove networking insights gained through research into AMD's product roadmap.
- Contributed new features and performance optimizations to AMD's event-driven, cycle-level CPU/GPU simulator.
- Mentored several intern projects and new hires.
- Interviewed candidates for positions in several technical areas.
- Authored and presented multiple publications at domestic and international conferences.
- Wrote 10+ patent applications to protect AMD's competitive intellectual property.

Graduate Intern (Research) *Summer 2014*

- Researched novel techniques to reduce GPU remote kernel launch latency accross high performance networks.
- Modeled and evaluated ideas in open-source GPU simulation software.

Graduate Intern (Client SoC) *Summer 2013*

- Contributed custom code and debugged in-house, trace-driven SoC simulator.
- Developed and documented work-flow to collect HyperTransport memory bus traces.
- Conducted GPU/CPU sensitivity studies on strategically important OpenCL applications.

University of Texas at Austin *Austin, TX*

Graduate Research Assistant (Advisor: Lizy K. John) *Fall 2013 – Summer 2015*

- Worked on graph partitioning algorithms for heterogeneous data centers using GraphLab.
- Explored the use of McPAT for modeling modern microprocessors using performance counters.
- Performed performance evaluation on Hadoop Map/Reduce framework.
- Prototyped microarchitecture changes on several research simulation platforms.

Intel *Hillsboro, Oregon*

Graduate Intern (Storage Technology Group) *Summer 2012*

- Designed hardware in Verilog for a prototype of Intel's Volume Management Device (VMD) technology.
- Assisted software development team with debugging of prototype Linux driver for VMD.

Undergraduate Intern (Storage Technology Group) *Summer 2011*

- Delivered key insights into Storage Area Network (SAN) caching techniques and optimizations.
- Provided quantitative measurements detailing strengths and weaknesses of SAN topologies.

Washington University in St. Louis *St. Louis Missouri*

Network Security Analyst *Spring 2011*

- Prototyped Snort open-source packet-analyzer for deployment on live networks.
- Wrote PHP and Bash scripts for automation of network maintenance tasks.

Announce Media *St. Louis Missouri*

Operations Intern *Summer 2010*

- Designed and implemented automated solutions for backing up critical company resources to cloud storage.
- Engineered automated deployment process of company resources to new machines.

Technical Skills

- Knowledgeable of modern CPU and GPU microarchitectures
- Knowledgeable of the fundamentals of systems software and operating systems
- Experience with MPI, PGAS languages, and high performance RDMA networking stacks
- Experience with workload characterization and hardware/software performance analysis
- Proficient with C/C++, CUDA, OpenCL, and HIP programming languages
- Proficient with git, Gerrit, and Jenkins for version control, continuous integration, and regression testing
- Proficient with Bash and Python for scripting
- Proficient with event-driven architectural simulators, such as gem5

Awards and Service

2019-20	**Spotlight Award (awarded 3 times),** AMD	*Austin, TX*
2015	**Best Paper Runner-Up,** International Conference on Parallel Processing (ICPP)	*Beijing, China*
2015	**Best in Session,** Semiconductor Research Corporation (SRC) TECHCON	*Austin, TX*
2012-16	**Microelectronics and Computer Development Fellow,** The University of Texas at Austin	*Austin, TX*
2012	**Outstanding Undergraduate Student,** Washington University in St. Louis	*St. Louis, MO*
2020	**Program Committee Member,** Workshop on General Purpose Processing Using GPU (GPGPU)	*San Diego, CA*
2021	**External Review Committee Member,** International Symposium on Computer Architecture (ISCA)	*Online Only*
2021	**External Review Committee Member,** International Symposium on Microarchitecture (MICRO)	*Online Only*
2022	**External Review Committee Member,** International Symposium on Computer Architecture (ISCA)	*New York City, USA*

Publications

Increasing GPU Translation Reach by Leveraging Under-Utilized On-Chip Resources
Jagadish Kotra, Michael LeBeane, Mahmut Kandemir, Gabriel Loh

International Symposium on Microarchitecture (MICRO). October 2021.

GPU Initiated OpenSHMEM: Correct and Efficient Intra-Kernel Networking for dGPUs
Khaled Hamidouche, Michael LeBeane
Principles and Practice of Parallel Programming (PPoPP). February 2020.

Optimizing GPU Cache Policies for MI Workloads
Johnathan Alsop, Matthew D. Sinclair, Anthony Gutierrez, Srikant Bharadwaj, Xianwei Zhang, Bradford Beckmann, Alexandru Dutu, Onur Kayiran, Michael LeBeane, Brandon Potter, Sooraj Puthoor, Tsung Tai Yeh
International Symposium on Workload Characterization (IISWC) (Short Paper). November 2019.

ComP-Net: Command Processor Networking for Efficient Intra-Kernel Communications on GPUs
Michael LeBeane, Khaled Hamidouche, Brad Benton, Mauricio Breternitz, Steven K. Reinhardt, Lizy K. John
International Conference on Parallel Architectures and Compilation Techniques (PACT). November 2018.

Neighborhood-Aware Address Translation for Irregular GPU Applications
Seunghee Shin, Michael LeBeane, Yan Solihin, Arkaprava Basu
International Symposium on Microarchitecture (MICRO). October 2018.

Case Study of Process Variation-Based Domain Partitioning of GPGPUs
Shomit Das, Michael LeBeane, Bradford Beckmann, Greg Sadowski
International Symposium on Asynchronous Circuits and Systems (ASYNC). May 2018.

Lost in Abstraction: Pitfalls of Analyzing GPUs at the Intermediate Language Level
Anthony Gutierrez, Bradford M Beckmann, Alexandru Dutu, Joseph Gross, Michael LeBeane, John Kalamatianos, Onur Kayiran, Matthew Poremba, Brandon Potter, Sooraj Puthoor, Matthew D Sinclair, Mark Wyse, Jieming Yin, Xianwei Zhang, Akshay Jain, Timothy Rogers
International Symposium on High Performance Computer Architecture (HPCA). Industrial Session. February 2018.

GPU Triggered Networking for Intra-Kernel Communications
Michael LeBeane, Khaled Hamidouche, Brad Benton, Mauricio Breternitz, Steven K. Reinhardt, Lizy K. John
International Conference for High Performance Computing, Networking, Storage, and Analysis (SC). November 2017.

Extended Task Queuing: Active Messages for Heterogeneous Systems
Michael LeBeane, Brandon Potter, Abhisek Pan, Alexandru Dutu, Vinay Agarwala, Wonchan Lee, Deepak Majeti, Bibek Ghimire, Eric Van Tassell, Samuel Wasmundt, Brad Benton, Mauricio Breternitz, Michael L. Chu, Mithuna Thottethodi, Lizy K. John, and Steven K. Reinhardt.
International Conference for High Performance Computing, Networking, Storage, and Analysis (SC). November 2016.

Proxy-Guided Load Balancing of Graph Processing Workloads on Heterogeneous Clusters
Shuang Song, Meng Li, Xinnian Zheng, Jee Ho Ryoo, Reena Panda, Michael LeBeane, Andreas Gerstlauer, and Lizy K. John.
International Conference on Parallel Processing (ICPP). August 2016.

Genesys: Automatically Generating Representative Training-sets
Reena Panda, Xinnian Zheng, Jee Ho Ryoo, Michael LeBeane, Shuang Song, Andreas Gerstlauer, and Lizy K. John.
International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS). July 2016.

Data Partitioning Strategies for Graph Workloads on Heterogeneous Clusters
Michael LeBeane, Shuang Song, Reena Panda, Jee Ho Ryoo, and Lizy K. John.
International Conference for High Performance Computing, Networking, Storage and Analysis (SC). November 2015.

Performance Characterization of Modern Databases on Out-of-order CPUs
Reena Panda, Christopher Erb, Michael LeBeane, Jee Ho Ryoo, and Lizy K. John.
International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD). October 2015.

WattWatcher: Fine-Grained Power Estimation for Emerging Workloads
Michael LeBeane, Jee Ho Ryoo, Reena Panda, and Lizy K. John.
International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD). October 2015.

WattWatcher: Fine-Grained Power Estimation on Live Multicore Systems Using Configurable Models (Best in Session)
Michael LeBeane, Jee Ho Ryoo, Reena Panda, and Lizy K. John.
(SRC TECHCON). September 2015.

GPGPU Benchmark Suites: How Well Do They Sample the Performance Spectrum (Best Paper Runner-Up)
Jee Ho Ryoo, Saddam Quirem, Michael LeBeane, Reena Panda, Shuang Song, and Lizy K. John.
International Conference on Parallel Processing (ICPP). September 2015.

Patents

Khaled Hamidouche, Michael LeBeane, Hari Thangirala. **Systems and methods for reducing instruction code memory footprint for multiple processes executed at a coprocessor.** Advanced Micro Devices Incorporated, assignee. Application Number 16/719076.

Michael LeBeane, Khaled Hamidouche, Hari Thangirala, Brandon Potter. **Efficient memory-semantic networking using scoped memory models.** Advanced Micro Devices Incorporated, assignee. Application Number 17/033170.

Jagadish Kotra, Michael LeBeane. **Techniques to improve translation lookaside buffer reach by leveraging idle resources.** Advanced Micro Devices Incorporated, assignee. Application Number 17/008435.

Michael LeBeane, Khaled Hamidouche, Brandon Potter. **Network Command Coalescing on GPUs.** Advanced Micro Devices Incorporated, assignee. Application Number 16/993150.

Michael LeBeane, Seunghee Shin. **Apparatus and method for neighborhood-aware virtual to physical address translations.** Advanced Micro Devices Incorporated, assignee. Patent Number 10684957.

Michael LeBeane, Khaled Hamidouche, Brad Beckmann. **Network-related performance for gpus.** Advanced Micro Devices Incorporated, assignee. Application Number 16/049216.

Khaled Hamidouche, Michael LeBeane, Nicholas Malaya, Joseph Greathouse. **Optimized and scalable sparse triangular linear systems on networks of accelerators.** Advanced Micro Devices Incorporated, assignee. Patent Number 10936697.

Khaled Hamidouche, Michael LeBeane, Brad Benton. **Network packet templating for gpu-initiated communication.** Advanced Micro Devices Incorporated, assignee. Patent Number 10740163.

Arkaprava Basu, Michael LeBeane, Eric Van Tassell. **Quality of service for input/output memory management unit.** Advanced Micro Devices Incorporated, assignee. Patent Number 11144473B2.

Khaled Hamidouche, Michael LeBeane, Brad Benton, Michael Chu. **Optimized asynchronous training of neural networks using a distributed parameter server with eager updates.** Advanced Micro Devices Incorporated, assignee. Application Number 15/898433.

Michael LeBeane, Khaled Hamidouche, Brad Benton. **Gpu networking using an integrated command processor.** Advanced Micro Devices Incorporated, assignee. Application Number 15/815043.

Michael LeBeane, Brad Benton, and Vinay Agarwala. **Network cache injection for coherent GPUs.** Advanced Micro Devices Incorporated, assignee. Application Number 15/498076.

Michael LeBeane and Steve Reinhardt. **GPU remote communication with triggered operations.** Advanced Micro Devices Incorporated, assignee. Patent Number 10936533.

Michael LeBeane, Abhisek Pan, and Steve Reinhardt. **Network interface controller-based scheduling of processing tasks in a distributed computing system.** Advanced Micro Devices Incorporated, assignee. Patent Number 10963309.

Mauricio Breternitz, Deepak Majeti, and Michael LeBeane. **Power-aware Work Stealing.** Advanced Micro Devices Incorporated, assignee. Patent Number 10089155.