Michael LeBeane

mlebeane@gmail.com
www.mlebeane.com

Computer architecture Ph.D. with 5+ years industry experience.

EDUCATION

University of Texas at Austin
GPA 3.97/4.0
Ph.D. in Electrical and Computer Engineering
M.S. in Electrical and Computer Engineering
Graduation: August 2018

Washington University in St. Louis
GPA 3.95/4.0 (*Summa Cum Laude)*
Bachelor of Science in Computer Engineering
Bachelor of Science in Computer Science
Graduation: May 2012

Ph.D. Dissertation
Optimizing Communication in GPU Clusters

With the impending end of Dennard scaling, 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 HPC 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 many 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 strategies are evaluated across many important multi-node primitives and critical emerging workloads, such as machine learning.

WORK AND RESEARCH EXPERIENCE

Advanced Micro Devices, Inc.
Austin, Texas
MTS Silicon Design Engineer **(Spring 2019 - Current)**
Senior Design Engineer **(Summer 2014, Summer 2015 - Spring 2019)**

- Researching novel, RDMA-based communications hardware for GPUs
- Prototyping and evaluating APIs, runtimes, and programming models for efficient and performant GPU/NIC interaction with minimal host CPU involvement
- Outlining high performance networking strategies for AMD GPUs as a part of the US Department of Energy's PathForward program
- Driving networking insights gained through research into AMD's product roadmap and ROCm software stack
- Contributing new features and performance optimizations to AMD's event-driven, cycle-level CPU/GPU simulator

Graduate Intern **(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, Texas
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 Corporation
Hillsboro, Oregon
Graduate Intern **(Summer 2012)**

- Designed and verified architecture and RTL for prototype hardware to be included in future Intel chipsets
- Assisted software development team with debugging of prototype Linux driver

Undergraduate Intern **(Summer 2011)**

- Delivered key insights into 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, and OpenCL programming languages
Proficient with Bash and Python for scripting
Proficient with event-driven architectural simulators

LEADERSHIP AND TEACHING

Washington University in St. Louis
Computer Science Student Advisory Board **(2008 - 2012)**

- Advised on proposed computer science curriculum changes
- Provided conduit for students to voice suggestions and concerns to department leadership

*Computer Science Teaching Assistant (***Spring 2010)**

- Aided students with understanding of Java and basic object-oriented design principles
- Held office hours and private tutoring sessions for students

HONORS AND AWARDS

Microelectronics and Computer Development Fellow **(2012 - 2016)**
Outstanding Undergraduate Student Award **(2012)**

PUBLICATIONS

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

Arkaprava Basu, Michael LeBeane, Eric Van Tassell. Quality of service for input/output memory management unit. Advanced Micro Devices Incorporated, assignee. Filling Number 16/007027.

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. Filing Number 15/898433.

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

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

Michael LeBeane and Steve Reinhardt. Efficient GPU remote communication with triggered operations. Advanced Micro Devices Incorporated, assignee. Filling Number 15/297079.

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. Filing Number 15/267936.

Mauricio Breternitz, Deepak Majeti, and Michael LeBeane. Power-aware Work Stealing. Advanced Micro Devices Incorporated, assignee. Filling Number 14/862038.