

R306, 140 Fenway, Boston, MA 02115

☐ (+1) 716-868-2480 | **S** yifansun@coe.neu.edu | **A** syifan.github.io | 😯 syifan I ★ syifan

Education

Northeastern University

Boston, MA Ph.D. in Computer Engineering Sep. 2013 - Present

University at Buffalo

Buffalo, NY

M.S. in Electrical Engineering

Sep. 2011 - Jun. 2013

Huazhong University of Science and Technology

Wuhan, China Sep. 2007 - Jun. 2011

B.S. in Electrical Engineering

Wuhan University

Wuhan, China

B.A. in Business Administration

Sep. 2008 - Jun. 2011

Industry Experience _____

AMD, Software Engineer (Co-op)

Boxborough, MA

Performance modeling and graphics simulator development for AMD Navi GPU

Jul. 2018 - Dec.2018

Dell EMC, Software Engineer (Co-op)

Hopkinton, MA

Cloud-based GPU-as-a-service system design, development, and deployment

Jul. 2016 - Dec.2016

Awards

2019	Outstanding Graduate Student in Experiential Learning, Northeastern University	Boston, MA
2019	Teaching Award , Northeastern University College of Engineering	Boston, MA
2018	Best Paper Award, ICPE	Berlin, Germany
2013	Best Student Paper Award, WUWNET	Kaohsiung, Taiwan
2016	Best Paper Candidate, IISWC	Providence, RI
2019	ACM Student Travel Grant, ISCA	Phoenix, AZ
2016	NSF Student Travel Grant, IISWC	Providence, RI

Publications

Peer-reviewed Conference Publications

- 1. Trinayan Baruah, Yifan Sun, Ali Tolga Dinçer, Saiful A. Mojumder, José Luis Abellán, Yash Ukidave, Ajay Joshi, Normal Rubin, John Kim, David Kaeli. 2020. Griffin: Hardware-Software Support for Efficient Page Migration in Multi-GPU Systems. In Proceedings of the 26th IEEE International Symposium on High-Performance Computer Architecture (HPCA '20). IEEE. [Acceptance rate \approx 19.4%]
- 2. Yifan Sun, Trinayan Baruah, Saiful A. Mojumder, Shi Dong, Xiang Gong, Shane Treadway, Yuhui Bao, Spencer Hance, Carter McCardwell, Vincent Zhao, Harrison Barclay, Amir Kavyan Ziabari, Zhongliang Chen, Rafael Ubal, José L. Abellán, John Kim, Ajay Joshi, and David Kaeli. 2019. MGPUSim: Enabling multi-GPUPerformance Modeling and Optimization. In Proceedings of the 46th International Symposium on Computer Architecture (ISCA '19). ACM, New York, NY, USA, 197-209. [Acceptance rate $\approx 17.0\%$]
- 3. Mohammad Khavari Tavana, Yifan Sun, Nicolas Bohm Agostini, and David Kaeli. 2019. Exploiting Adaptive Data Compression to Improve Performance and Energy-Efficiency of Compute Workloads in Multi-GPU Systems. In

- Proceedings of the 33rd IEEE International Parallel and Distributed Processing Symposium (IPDPS '19). IEEE, Rio de Janeiro, Brazil, 664-674 [Acceptance rate $\approx 27.7\%$]
- 4. Saiful A Mojumder, Marcia S Louis, <u>Yifan Sun</u>, Amir Kavyan Ziabari, José L Abellán, John Kim, David Kaeli, and Ajay Joshi. 2018. **Profiling DNN Workloads ona Volta-based DGX-1 System**. In Proceedings of the 2018 IEEE International Symposium on Workload Characterization (IISWC '18). IEEE, Raleigh, North Carolina, USA, 122-133. [Acceptance rate ≈ 36.2%]
- 5. Rozhin Doroudi, Rana Azghandi, Zlatan Feric, Omic Mohaddesi, <u>Yifan Sun</u>, Jacqueline Griffin, Ozlem Ergun, David Kaeli, Pedro Sequeira, Stacy Marsella, and Casper Harteveld. 2018. **An Integrated Simulation Framework for Examining Resiliency in Pharmaceutical Supply Chains Considering Human Behavior**. In Proceedings of the 2018 Winter Simulation Conference (WSC '18). ACM, Gothenburg, Sweden, 88-99. [Acceptance rate ≈ 70.4%]
- 6. <u>Yifan Sun</u>, Saoni Mukherjee, Trinayan Baruah, Shi Dong, Julian Gutierrez, Prannoy Mohan, and David Kaeli. 2018. **Evaluating Performance Tradeoffs on the Radeon Open Compute Platform**. In Proceedings of the 2018 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS '18). IEEE, Belfast, Northern Ireland, United Kingdom, 209-218. [Acceptance rate ≈ 31.3%]
- 7. [Best Paper Award] Shi Dong, Gong Xiang, <u>Yifan Sun</u>, Trinayan Baruah, and David Kaeli. 2018. Characterizing the Microarchitectural Implications of a Convolutional Neural Network (CNN) on GPUs. In Proceedings of the 2018 ACM International Conference on Performance Engineering (ICPE '18). ACM, Berlin, Germany, 96-106. [Acceptance rate = 24.0%]
- 8. Trinayan Baruah, <u>Yifan Sun</u>, Shi Dong, David Kaeli, and Norm Rubin. 2018. **Airavat: Improving Energy Efficiency of Heterogeneous Applications**. In Proceedings of the 2018 Design, Automation & Test in Europe Conference & Exhibition (DATE '18). IEEE, Dresden, Germany, 731-736. [Acceptance rate ≈ 24.2%]
- 9. Leiming Yu, Xun Gong, <u>Yifan Sun</u>, Qianqian Fang, Norm Rubin, and David Kaeli. 2017. **Moka: Model-based Concurrent Kernel Analysis**. In Proceedings of the 2017 IEEE International Symposium on Workload Characterization (IISWC '17). IEEE, Seattle, Washington, USA, 197-206. [Acceptance rate ≈ 27.7%]
- 10. [Best Paper Candidate] Yifan Sun, Xiang Gong, Amir Kavyan Ziabari, Leiming Yu, Xiangyu Li, Saoni Mukherjee, Carter McCardwell, Alejandro Villegas, and David Kaeli. 2016. Hetero-Mark, a Benchmark Suite for CPU-GPU Collaborative Computing. In Proceedings of the 2016 IEEE International Symposium on Workload Characterization (IISWC '16). IEEE, Providence, Rhode Island, USA, 1-10. [Acceptance rate ≈ 30.4%]
- 11. <u>Yifan Sun</u>, Chisheng Liang, Steven Sutherland, Casper Harteveld, and David Kaeli. 2016. **Modeling Player Decisions** in a Supply Chain Game. In Proceedings of the 2016 IEEE Conference on Computational Intelligence and Games (CIG '16). IEEE, Santorini, Greece, 1-8. [Acceptance rate unknown]
- 12. Saoni Mukherjee, <u>Yifan Sun</u>, Paul Blinzer, Amir Kavyan Ziabari, and David Kaeli. 2016. **A Comprehensive Performance Analysis of HSA and OpenCL 2.0**. In Proceedings of the 2016 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS '16). IEEE, Uppsala, Sweden, 183-193. [Acceptance rate ≈ 35.1%]
- 13. Jithin Jagannath, Anu Saji, Hovannes Kulhandjian, <u>Yifan Sun</u>, Emrecan Demirors, and Tommaso Melodia. 2013. **A Hybrid MAC Protocol with Channel-Dependent Optimized Scheduling for Clustered Underwater Acoustic Sensor Networks**. In Proceedings of the 8th ACM International Conference on Underwater Networks and Systems (WUWNET).
 ACM, Kaohsiung, Taiwan, Article 3, 8 pages. [Acceptance rate = 20.0%]
- 14. [Best Student Paper Award] <u>Yifan Sun</u>, and Tommaso Melodia. The internet underwater: An IP-compatible Protocol Stack for Commercial Undersea Modems. In Proceedings of the 8th ACM International Conference on Underwater Networks and Systems (WUWNET). ACM, Kaohsiung, Taiwan, Article 37, 8 pages. [Acceptance rate = 20.0%]

Journal Publications

1. Amir Kavyan Ziabari, <u>Yifan Sun</u>, Yenai Ma, Dana Schaa, José L. Abellán, Rafael Ubal, John Kim, Ajay Joshi, and David Kaeli. 2016. **UMH: A Hardware-based Unified Memory Hierarchy for Systems with Multiple Discrete GPUs**. ACM

Transactions on Architecture and Code Optimization (TACO) 13, 4 Article 35 (December 2016), 25 pages.

- Abdulla K. Al-Ali, <u>Yifan Sun</u>, Marco Di Felice, Jarkko Paavola, and Kaushik R. Chowdhury. 2015. Accessing Spectrum
 Databases using Interference Alignment in Vehicular Cognitive Radio Networks. IEEE Transactions on Vehicular
 Technology 64, 1 (2014), 263-272
- 3. <u>Yifan Sun</u>, and Kaushik R. Chowdhury. 2015. **Enabling Emergency Communication through a Cognitive Radio Vehicular Network**. IEEE Communications Magazine 52, 10 (2014), 68-75.

Book Chapters

1. Shih-Hao Hung, Thomas B. Jablin, <u>Yifan Sun</u>, Rafael Ubal, and David Kaeli. 2015. **HSA Simulators**. A book chapter in Heterogeneous System Architecture: Practical Applications for Industry, 1st edition, Elsevier Nov. 2015.

Patents

- 1. <u>Yifan Sun</u>, Layne Peng, Robert A. Lincourt JR., John Cardente, and Junping Zhao. (Jun. 2019). **Managing access to a resource pool of graphics processing units under fine grain control**. Patent No. US 10,262,390, Filed Apr. 14th., 2017, Issued Jun. 27th., 2019.
- 2. Junping Zhao, Layne Peng, Jie Bao, Kun Wang, and <u>Yifan Sun</u>. (Apr. 2019). **Checkpointing for GPU-as-a-Service in Cloud Computing Environment**, Patent No. US 10,275,851, Filed Apr. 25th., 2017, Issued Apr. 30th., 2019.
- 3. <u>Yifan Sun</u>, Layne Peng, Robert A. Lincourt JR., John Cardente, John S Harwood. (Oct. 2018). **Queue-based GPU Virtualization and Management System**. Patent No. US 10,109,030, Filed Dec. 27th., 2016, Issued Oct. 23rd., 2018.

Workshop Publications

1. <u>Yifan Sun</u>, Trinayan Baruah, Shi Dong, and David Kaeli. 2019. **MGSim: A Flexible High-Performance Simulator for Multi-GPU Systems**. International Workshop on OpenCL (IWOCL).

Preprints

Yifan Sun, Trinayan Baruah, Saiful A Mojumder, Shi Dong, Rafael Ubal, Xiang Gong, Shane Treadway, Yuhui Bao, Vincent Zhao, José Luis Abellán, John Kim, Ajay Joshi, and David Kaeli. 2019. MGSim+MGMark: A Framework for Multi-GPU System Research. arXiv preprint arXiv:1811.02884.

Open-Source Software _____

MGPUSim (https://gitlab.com/akita/gcn3)

Multi-GPU system simulator based on AMD GCN3 GPUs

Akita (https://gitlab.com/akita/akita)

High-flexibility, high-performance, parallel computer architecture simulation framework

Hetero-Mark (https://github.com/NUCAR-DEV/Hetero-Mark)

Benchmark suite for CPU-GPU collaborative computing

Drug Supply Chain Simulator (https://gitlab.com/syifan/crisp)

Human-in-the-loop logistics simulator for the U.S. drug supply chain

VistaLights (https://github.com/syifan/VistaLights)

Strategic game for maritime traffic management and disaster relieving

Talks and Tutorials

1. **Tutorial on the Akita Simulator Framework and MGPUSim**. With Trinayan Baruah, Shi Dong, and David Kaeli. To be presented at HPCA 2020.

- 2. Research in the NUCAR Laboratory at Northeastern University. FutureWei. With David Kaeli. July 2019.
- 3. MGPUSim: a Flexible High-Performance Simulator for Multi-GPU Systems. International Workshop on OpenCL (IWOCL). May 2019.
- 4. AKITA: A Go-Based Computer Architecture Simulator Framework. Google. May 2019.
- 5. **Enabling Multi-GPU High Performance Computing with Memory System Design**. Lighting talk at Boston University Red Hat Collaboratory. Feb. 2019.
- 6. Benchmarking the New Unified Memory of CUDA 8. With Frank Zhao. GTC 2017 San Jose. Aug. 2017.
- 7. Multi2Sim 5.0 Tutorial at IISWC. Sep. 2016.

Teaching

Northeastern University, Boston, MA

Fundamental Digital Design and Computer Organization

Fall 2019

Co-instructor. With Dr. Pereira da Silva Aloizio

Intermediate-level Undergraduate Course (3rd year)

Fundamentals of Engineering Algorithms

Spring 2018

Instructor

Intermediate-level Undergraduate Course (3rd year)

Redesigned the course ("I have learned a lot in this course": 4.7 out of 5)

Instructor Effectiveness 4.4 out of 5

Embedded Design Enabling Robotics

Fall 2017

Instructor

Intermediate-level Undergraduate Course (2nd year)

Instructor Effectiveness 4.6 out of 5

Student Mentees _____

Ali Mosallaei	High school student research	2019
Ali Tolga Dinçer	Research internship from Istanbul Technical University, Turkey	2019
Xin Li	REU student from Bunker Hill Community Colledge	2019
Nicholas Fresneda	Undergraduate research	2019
Anton Lazarev	Undergraduate research	2019
Prannoy Mohan	Undergraduate research (first job: Software Engineer at Amazon)	2017 - 2019
Yuhui Bao	Graduate research (now: Ph.D Student at Northeastern University)	2018
Shane Treadway	Undergraduate research	2017 - 2018
Spencer Hance	Undergraduate research (first job: Software Engineer at Google)	2018
Carter McCardwell	Undergraduate research (first job: Software Engineer at Airbnb)	2016 - 2018
Vincent Zhao	Undergraduate research	2017 - 2018
Joseph Moore	Undergraduate research	2017
Benhamin Logan	Undergraduate research	2016
Harrison Barclay	Undergraduate research	2015 - 2018
Chisheng Liang	Undergraduate research	2014 - 2016

Selected Media Coverage _____

HiPEAC info 58 MGPUsim announced at ISCA 2019

News@Northeastern A Student Went off to Do a Co-op at a Major Tech Firm. He Came Back With a Patent.

WIRED Finally, the Underwater We've All Been Waiting For

NBC News Deep-sea Internet to Detect Tsunamis, Spy on Smugglers, and Discover Oil

Service _____

Program Committee	13th Workshop on General Purpose Processing Using GPU (GPGPU)	2019
Web Chair	9th Workshop on General Purpose Processing Using GPU (GPGPU)	2016