Min Si

minsii.github.io | minsi.atwork@gmail.com | (Mobile) +1 630-880-4388
linkedin.com/in/minsi-atwork | github.com/minsii | Naperville, IL 60540
Research Scientist, Facebook

Highlights: Research interests include parallel programming models and HPC runtime systems. Lead research and development of one-sided communication models including MPI RMA and OpenSHMEM for HPC systems. Experienced in project management and staff supervision.

Education

Ph.D. in Computer Science 2012/10 – 2016/03
Department of Computer Science Tokyo, Japan
Graduate School of Information Science and Technology
The University of Tokyo

Master of Science 2010/10 – 2012/09
Department of Computer Science Tokyo, Japan
Graduate School of Information Science and Technology
The University of Tokyo

Bachelor of Arts 2003/09 – 2007/07
Department of Japanese Language and Culture Chongqing, China
College of Oriental Languages
Sichuan International Studies University

Professional Experience

Research Scientist 2021/06 – Current
Facebook, CA, USA
- HPC hardware/software co-design for AI.

Assistant Computer Scientist 2018/05 – 2021/06
Argonne National Laboratory, IL, USA
- OpenSHMEM implementation over MPI (PI).
- High performance MPI implementation (Co-PI).
- Data management and runtime optimizations for AI workflow (Senior Personnel).
- Supporting infrastructure for surrogate benchmarks on HPC systems (Senior Personnel).

Enrico Fermi Postdoctoral Scholar 2016/10 – 2018/04
Argonne National Laboratory, IL, USA
- Dynamic execution runtime (PI).
- High performance MPI implementation (Co-PI).

Postdoctoral Appointee 2016/05 – 2016/09
Argonne National Laboratory, IL, USA
- High performance MPI implementation.

Guest Graduate Student 2014/05 – 2016/03
Argonne National Laboratory, IL, USA
- MPI optimization on massively parallel multi-/many-core architectures.

Research Aide 2013/05 – 2013/09
Argonne National Laboratory, IL, USA
- Multithreaded MPI for many-core environments.

Summer Internship 2011/08 – 2011/11
NEC Corp, Tokyo, Japan
- InfiniBand driver modification for direct data transfer from/to FPGA board.

Part-time Programmer 2010/12 – 2011/12
Secioss Corp, Tokyo, Japan
- E-commerce web system development using PHP.

Full-time Engineer 2008/07 – 2010/09
Kinotrope Inc, Tokyo, Japan
- Web system (CMS, E-commerce, Auto-test) development using PHP.

Full-time Engineer 2007/07 – 2008/04
DGT Information Systems Ltd, Shanghai, China
- Credit card member management system development using Java.

Project Leadership

OpenSHMEM over MPI: Analyzing and Improving the Suitability of MPI as an OpenSHMEM Runtime.
- *PI. Period*: 01/2021 - present. *Technical Lead and Core Developer. Period*: 08/2018 to 12/2020.
- Leading research and development of OpenSHMEM over MPI implementation. Managing project milestones and reporting to DOD program managers. Participating at OpenSHMEM specification committee and contributing to specification revisions. Serving as the committee RMA working group chair since 2020/11.
- *Websites*: `http://www.mcs.anl.gov/project/oshmpi`, `https://github.com/pmodels/oshmpi`
- *Supervision (1 Postdoc)*:
  Huansong Fu: Postdoctoral Research Associate, period: 03/2019 to 08/2019. Predoctoral Research Associate, period: 08/2018 to 12/2018. Ph.D. student, Florida State University, USA, period: 05/2018 to 08/2018.

MPICH: A high-performance, portable implementation of the Message Passing Interface (MPI).
- *Co-PI and Core Developer. Period*: 10/2016 to present.
- Leading research and development of MPI RMA model and shared memory based communication. Driving new research directions such as optimizing runtime system for AI on HPC and using AI in communication system. Contributing to MPI RMA specification revision.
- *Websites*: `http://www.mpich.org`, `https://github.com/pmodels/mpich`
- *Supervision (4 Graduate students)*:
  Michael Wilkins: Ph.D. student, Northwestern University, USA. Period: 08/2020 to 11/2020.
  Li Cao: Master student, University of Chicago, USA. Period: 08/2020 to 12/2020.
  Subhadeep Bhattacharya: Ph.D. student, Florida State University, USA. Period: 05/2019 to 08/2019.
  Sarunya Pumma: Ph.D. student, Virginia Tech, USA. Period: 08/2016 to 01/2017.

Braid: Data Flow Automation for Scalable and FAIR Science.
- *Senior Personnel. Period*: 10/2020 to present.
- Leading research and development of AI/DL data management and optimization on HPC systems. Co-working with Dr. Bogdan Nicolae.
- *Supervision (1 Graduate student)*:
  Jie Liu: Ph.D. student, University of California, Merced, USA. Period: 01/2021 to present.

Beehive: A Dynamic Execution Environment for Performance, Power, and Resilience on Extreme-Scale Computing Systems.
- *PI and Core Developer. Period*: 10/2016 to 03/2020.
- Leading research and development of the Beehive eco-system. Managing project milestones and reporting to Argonne LDRD program office. Collaborating research with Intel, RIKEN Japan, and BSC Spain.
- *Website*: `http://www.mcs.anl.gov/project/casper`, `https://github.com/pmodels/casper`
- *Supervision (2 Graduate students)*:
  Kaiming Ouyang (joint project with MPICH): Ph.D. student, University of California, Riverside, USA. On PhD committee. Period: 10/2018 to present.
  Yanhao Chen: Ph.D. student, Rutgers University, USA. Period: 05/2018 to 08/2018.

Honors and Awards

Impact Argonne Award for Extraordinary Effort 2020/08
Argonne National Laboratory, USA

IEEE-CS Technical Consortium on High Performance Computing (TCHPC) *2018/11*
Early Career Researchers Award for Excellence in High Performance Computing. SC 2018, USA

Best Paper Award at the 27th ACM International Symposium on High-Performance Parallel and Distributed Computing (Top 1/112). HPDC 2018, USA *2018/06*

Enrico Fermi Fellowship for Postdoctoral Scientists 2016 – 2018
Argonne National Laboratory, USA

Dean's Award for Outstanding Achievement Doctoral Course 2016/03
The University of Tokyo, Japan

Fellowship Special Scholarship Program for International Students 2012 – 2015
The University of Tokyo, Japan

Computer Science Research Award for Young Scientists 2012
Information Processing Society of Japan (IPSJ), Japan

Selected Publications and Presentations

Refereed Journal Articles

1. S. Pumma, **M. Si**, W. Feng, and P. Balaji. Scalable Deep Learning via I/O Analysis and Optimization. In *ACM Transactions on Parallel Computing (TOPC)*, vol. 6, no. 2, pp. 1–34. June, 2019.
2. **M. Si**, A. J. Peña, J. Hammond, P. Balaji, M. Takagi, and Y. Ishikawa. Dynamic Adaptable Asynchronous Progress Model for MPI RMA Multiphase Applications. In *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, vol. 29, no. 9, pp. 1975–1989. September, 2018.

Research Papers in Refereed Conferences (peer-reviewed)

1. K. Ouyang, **M. Si**, A. Hori, Z. Chen, and P. Balaji. CAB-MPI: Exploring Interprocess Work-Stealing toward Balanced MPI Communication. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC 2020)*. IEEE Press, Article 36, 1–15. (*Acceptance Rate: 25.4%*)
2. A. Amer, C. Archer, [et al, including **M. Si**]. Software Combining to Mitigate Multithreaded MPI Contention. In *Proceedings of the 33rd ACM International Conference on Supercomputing (ICS 2019)*, pages 367–379. June 2019. (*Acceptance Rate: 23%*)
3. A. Hori, **M. Si (Joint First Co-Author)**, B. Gerofi, M. Takagi, J. Dayal, P. Balaji, and Y. Ishikawa. Process-in-Process: Techniques for Practical Address-Space Sharing. In *Proceedings of the 27th International Symposium on High-Performance Parallel and Distributed Computing (HPDC 2018)*, pages 131–143. June 2018. **Best Paper Award**. (*Acceptance Rate: 19.6%*)
4. Sarunya Pumma, **M. Si**, Wu-Chun Feng, and P. Balaji. Parallel I/O Optimizations for Scalable Deep Learning. In *Proceedings of 2017 IEEE 23rd International Conference on Parallel and Distributed Systems (ICPADS 2017)*, pages 720–729. December 2017.
5. **M. Si** and P. Balaji. Process-based Asynchronous Progress Model for MPI Point-To-Point Communication. In *Proceedings of 2017 IEEE 19th International Conference on High Performance Computing and Communications (HPCC 2017)*, pages 206–214. December 2017. (*Acceptance Rate: 38%*)
6. Sarunya Pumma, **M. Si**, Wu-Chun Feng, and P. Balaji. Towards Scalable Deep Learning via I/O Analysis and Optimization. In *Proceedings of 2017 IEEE 19th International Conference on High Performance Computing and Communications (HPCC 2017)*, pages 223–230. December 2017. (*Acceptance Rate: 38%*)
7. K. Raffenetti, A. Amer, [et al, including **M. Si**]. Why is MPI so Slow? Analyzing the Fundamental Limits in Implementing MPI-3.1. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC 2017)*, pages 62:1–62:12. November 2017. (*Acceptance Rate: 18%*)
8. **M. Si**, A. J Peña, J. Hammond, P. Balaji, M. Takagi, and Y. Ishikawa. Casper: An Asynchronous Progress Model for MPI RMA on Many-Core Architectures. In *Proceedings of the IEEE/ACM International Parallel and Distributed Processing Symposium (IPDPS 2015)*, pages 665–676, May 2015. (*Acceptance Rate: 21.8%*)
9. **M. Si**, A. J Peña, J. Hammond, P. Balaji, and Y. Ishikawa. Scaling NWChem with Efficient and Portable Asynchronous Communication in MPI RMA. In *Proceedings of 8th IEEE International Scalable Computing Challenge - Colocated with IEEE/ACM CCGrid 2015*, pages 811–816, May 2015. (*Acceptance Rate: 33%*)
10. **M. Si**, A. J. Peña, P. Balaji, M. Takagi, and Y. Ishikawa. MT-MPI: Multithreaded MPI for Many-core Environments. In *Proceedings of the 28th ACM International Conference on Supercomputing (ICS 2014)*, pages 125–134. May 2014. (*Acceptance Rate: 21%*)

Research Papers in Refereed Workshops (peer-reviewed)

1. **M. Si**, Y. Ishikawa, and M. Tatagi. Direct MPI Library for Intel Xeon Phi Co-Processors. In *Proceedings of the 2013 IEEE 27th International Parallel and Distributed Processing Symposium Workshops - PhD Forum (IPDPSW)*, pages 816-–824, May 2013.
2. **M. Si** and Y. Ishikawa. Design of Direct Communication Facility for Many-Core Based Accelerators. In *Proceedings of the 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops - PhD Forum (IPDPSW)*, pages 924-–929, May 2012.

Thesis

1. **Min Si**. Techniques For Enabling Highly Efficient Message Passing on Many-Core Architectures. Dissertation. Department of Computer Science, Graduate School of Information Science and Technology, The University of Tokyo, Japan. March 2016.

2. **Min Si**. Communication Facility in Manycore-based Cluster System. Master Thesis. Department of Computer Science, Graduate School of Information Science and Technology, The University of Tokyo, Japan. September 2012.

Invited Talks

General HPC Runtime Systems.
- Challenges and Opportunities in Co-Design for High-Performance Computing Software Systems *Invited talk at UMN Cloud Data Center and Edge (CDCE) workshop, 2021/02*
- Co-Design for High-Performance Computing Software Systems. *REBASE talk at ACM SIGPLAN conference on Systems, Programming, Languages, and Applications: Software for Humanity (SPLASH 2020).*
- AI@Edge: Software System Implications. *Microelectronics Workshop at Argonne National Laboratory*, October 2019.

OpenSHMEM.
- OpenSHMEM over MPI: A Performance Contender. *OpenSHMEM Birds of a Feather (unofficial) at SC20.*
- OpenSHMEM over MPI: Capabilities and Challenges. *OpenSHMEM Birds of a Feather at SC19.*
- Performance Analysis of MPI RMA in Supporting OpenSHMEM Runtime. *OpenSHMEM Birds of a Feather at SC18.*
- OpenSHMEM over Portable MPI RMA with Asynchronous Progress Support. *OpenSHMEM Birds of a Feather at SC17.*

Beehive and MPI.
- Towards Dynamic Communication Runtime for Scalable Irregular Parallel Computing. *Florida State University*, December 2018.
- Towards Portable and Adaptable Asynchronous Communication for One-Sided Applications. *Tenth International Workshop on Parallel Programming Models and Systems Software for High-End Computing (P2S2 2017).*
- Data Locality Challenges in Irregular Applications for Exascale Programing. *Fourth Workshop on Programming Abstractions for Data Locality (PADAL 2017).*

Professional Services and Activities

Editorial Affiliations

- IEEE Transactions on Parallel and Distributed Systems (IEEE TPDS). *Co-Editor* of Special Section on Parallel and Distributed Computing Techniques for AI, ML and DL in 2021, 2020.
- Software: Practice and Experience (SPE). *Co-Editor* of Special Issue on: New Trends in High Performance Computing: Software Systems and Applications in 2021.
- International Journal of Concurrency and Computation: Practice and Experience (CCPE). *Guest Co-Editor* of the Special Issue on Programming Models and Applications for Multicores and Manycores (PMAM) in 2021, 2020, 2019. *Guest Co-Editor* of the Special Issue on Parallel Programming Models and Systems Software (P2S2) in 2021-2020.
- Elsevier International Journal of Parallel Computing (PARCO). *Guest Co-Editor* of the Special Issue on Applications and System Software for Hybrid Exascale Systems (AsHES) in 2019, 2018. *Guest Co-Editor* of the Special Issue on Parallel Programming Models and Systems Software (P2S2) in 2019, 2018. *Guest Co-Editor* of the Special Issue on Programming Models and Applications for Multicores and Manycores (PMAM) in 2018.

Chair and Co-Chair - Conferences / Workshops

- IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC). *Program Track Vice-chair* of 2021.
- IEEE Symposium on High Performance Interconnects (HOTI). *Program Co-Chair* of 2021.
- International Workshop on Accelerators and Hybrid Emerging Systems (AsHES), held in conjunction with IEEE International Parallel and Distributed Processing Symposium (IPDPS). *General Chair* of 2021, 2020. *Program Co-Chair* of 2019, 2018.
- International Workshop on Parallel Programming Models and Systems Software for High-End Computing (P2S2), held in conjunction with International Conference on Parallel Processing (ICPP). *Program Co-Chair* of 2021, 2020, 2019, 2018. *Publicity Chair* of 2016.
- International Workshop on Programming Models and Applications for Multicores and Manycores (PMAM), held in conjunction with Principles and Practice of Parallel Programming (PPoPP). *Program Co-Chair* of 2021, 2020, 2019.
- International Conference for High Performance Computing, Networking, Storage, and Analysis (SC). *Panels Vice-Chair* of 2019.

- IEEE International Conference on Cluster Computing (Cluster). *Financial Chair* of 2022. *Virtual Arrangements Co-Chair* of 2021.
- European MPI Users Group Meeting (EuroMPI). *Financial Chair* of 2020. *Financial Co-Chair* of 2017. *Web Co-Chair* of 2018.
- International Conference on High Performance Computing in Asia-Pacific Region (HPC Asia). *Program Track Co-Chair* of 2021.

Committee Member - Conferences / Workshops

- International Conference for High Performance Computing, Networking, Storage and Analysis (SC). *TPC Member* of 2021, 2020. *Panels PC Member* of 2021.
- IEEE International Parallel and Distributed Processing Symposium (IPDPS). *TPC Member* of 2021. *Workshops Committee Member* of 2020, 2019, 2018, 2017.
- ACM International Conference on Supercomputing (ICS). *TPC Member* of 2021, 2020, 2018. *External Review Committee Member* of 2019.
- ACM SIGPLAN/SIGBED International Conference on Languages, Compilers, and Tools for Embedded Systems (LCTES). *TPC Member* of 2021.
- ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP). *TPC Member* of 2018. *External Review Committee Member* of 2021, 2020.
- IEEE/ACM International Symposium on Cluster,Cloud and Grid Computing (CCGrid). *TPC Member* of 2019, 2018, 2017.
- IEEE International Conference on Cluster Computing (Cluster). *TPC Member* of 2021, 2018, 2017.
- International Conference on Parallel Processing (ICPP). *TPC Member* of 2021, 2019.
- European MPI Users Group Meeting (EuroMPI). *TPC Member* of 2019, 2018, 2017.
- Workshop on OpenSHMEM and Related Technologies (OpenSHMEM). *TPC Member* of 2018, 2017.
- Supercomputing Asia conference (SCA). *TPC Member* of 2019, 2018.
- International Symposium on High Performance Interconnects (HOTI). *TPC Member* of 2017.
- IEEE International Conference on High Performance Computing and Communications (HPCC). *TPC Member* of 2017.

Journal Referee

- IEEE Transactions on Parallel and Distributed Systems (TPDS). 2020, 2019, 2018, 2017, 2014.
- IEEE Transactions on Computers (TC). 2020, 2019.
- IEEE Transactions on Cloud Computing (TCC). 2017.
- IEEE Transactions on Multi-Scale Computing Systems (TMSCS). 2018, 2017.
- Journal of Parallel and Distributed Computing (JPDC). 2020, 2019, 2018, 2017, 2016, 2015.
- Concurrency and Computation: Practice and Experience (CPE). 2017, 2016
- Elsevier International Journal of Parallel Computing (PARCO). 2020, 2018, 2017, 2016.
- International Journal of Parallel Programming (IJPP). 2018, 2016.