

# SHILEI TIAN

✉ i@tianshilei.me · ☎ (+1) 631-590-0814 · in shiltian

## EDUCATION

---

- Stony Brook University (SBU)**, New York, United States 2019/08 – Present  
*Ph.D. Candidate* in Computer Science Advisor: Dr. Barbara Chapman
- Shanghai Jiao Tong University (SJTU)**, Shanghai, China 2015/09 – 2018/03  
*M.S.* in Computer Technology Advisor: Dr. Xiaofeng Gao
- Xi'An University of Posts and Telecommunications (XUPT)**, Shaanxi, China 2010/09 – 2014/07  
*B.S.* in Computer Science and Technology

## PUBLICATIONS

---

- **Shilei Tian**, Joseph Huber, Konstantinos Parasyris, Barbara Chapman, Johannes Doerfert, “Direct GPU Compilation and Execution for Host Applications with OpenMP Parallelism”, in *Workshop on the LLVM Compiler Infrastructure in HPC (LLVM-HPC)*, November 13th, 2022, Dallas, Texas, USA
- Wenbin Lu, **Shilei Tian**, Tony Curtis, and Barbara Chapman, “Extending OpenMP and OpenSHMEM for Efficient Heterogeneous Computing”, in *Parallel Applications Workshop, Alternatives To MPI+X (PAW-ATM)*, November 14th, 2022, Dallas, Texas, USA
- **Shilei Tian**, Joseph Huber, John Tramm, Barbara Chapman, Johannes Doerfert, “Just-in-Time Compilation and Link Time Optimization for OpenMP Target Offloading”, in *International Workshop on OpenMP (IWOMP)*, September 27th–30th, 2022, Chattanooga, Tennessee, USA
- Johannes Doerfert, Atmn Patel, Joseph Huber, **Shilei Tian**, Jose M Monsalve Diaz, Barbara Chapman, Giorgis Georgakoudis, “Co-Designing an OpenMP GPU Runtime and Optimizations for Near-Zero Overhead Execution”, in *IEEE International Parallel & Distributed Processing Symposium (IPDPS)*, May 30th–June 3rd, 2022, Lyon, France
- Joseph Huber, Melanie Cornelius, Giorgis Georgakoudis, **Shilei Tian**, Jose M Monsalve Diaz, Kuter Dinel, Barbara Chapman, Johannes Doerfert, “Efficient Execution of OpenMP on GPUs”, in *International Symposium on Code Generation and Optimization (CGO)*, April 2nd-6th, 2022, Seoul, South Korea
- Seonmyeong Bak, Colleen Bertoni, Swen Boehm, Reuben Budiardja, Barbara Chapman, Johannes Doerfert, Markus Eisenbach, Hal Finkel, Oscar Hernandez, Joseph Huber, Shintaro Iwasaki, Vivek Kale, Paul R.C. Kent, JaeHyuk Kwack, Meifeng Lin, Piotr Luszczek, Ye Luo, Buu Pham, Swaroop Pophale, Kiran Ravikumar, Vivek Sarkar, Thomas Scogland, **Shilei Tian**, P.K. Yeung, “OpenMP Application Experiences: Porting to Accelerated Nodes”, in *Parallel Computing*, October 23rd, 2021.
- Atmn Patel, **Shilei Tian**, Johannes Doerfert, Barbara Chapman, “A Virtual GPU as Developer-Friendly OpenMP Offload Target”, in *Workshop on LLVM in Parallel Processing (LLPP)*, August 9th, 2021, Chicago, IL, USA
- **Shilei Tian**, Jon Chesterfield, Johannes Doerfert, Barbara Chapman, “Experience Report: Writing A Portable GPU Runtime with OpenMP 5.1”, in *International Workshop on OpenMP (IWOMP)*, September 15th–17th, 2021, Bristol, UK
- **Shilei Tian**, Johannes Doerfert, Barbara Chapman, “Concurrent Execution of Deferred OpenMP Target Tasks with Hidden Helper Threads”, in *Workshop on Languages and Compilers for Parallel Computing (LCPC)*, October 14th–16th, 2020, Stony Brook, New York, USA
- **Shilei Tian**, Haotian Wang, Sha Li, Fan Wu, Guihai Chen, “Trajectory-Based Multi-Hop Relay Deployment in Wireless Networks”, in *International Conference on Combinatorial Optimization and Applications (COCOA)*, December 16th–18th, 2017, Shanghai, China
- Haotian Wang, **Shilei Tian**, Xiaofeng Gao, Lidong Wu, Guihai Chen, “Approximation Designs for Cooperative Relay Deployment in Wireless Networks”, in *International Conference on Distributed Computing Systems (ICDCS)*, June 5th–8th, 2017, Atlanta, GA, USA

## SELECTED PROJECTS

---

### LLVM OpenMP

2019/08–Present

*Key Contributor*

- Proposed and implemented *hidden helper task* to support concurrent kernel execution of OpenMP target tasks (asynchronous target offloading). A similar concept is being discussed by OpenMP committee and is expected to be added to OpenMP 6.0. Benchmark results show up to  $7\times$  performance improvement. Intel's latest OpenMP compiler has adopted the implementation.
- Optimized target memory allocation by introducing target memory manager. OpenMP version of GridMini can reach same performance as CUDA version.
- Proposed and implemented a new OpenMP offloading target, virtual GPU (VGPU), that enables target offloading to host by reusing same compilation passes, execution model, and device runtime libraries as a real GPU.
- Rewrote the OpenMP target device runtime with OpenMP 5.1.
- Co-designed and implemented a new pure C++ OpenMP device runtime.
- Implemented middle-end optimizations to specialize four runtime function calls in compile time.

### Data Parallel C++

2018/05–2019/08

*Software Engineer, Validation Leader*

- Led the validation team of DPC++ compiler responsible for quality assurance, test development, etc.
- Enabled all regular tests and analyzed thousands of product issues. Enabled Khronos SYCL conformance test suite with Intel SYCL and developed 10 new test cases for Intel specific extension subgroups.
- Implemented 10+ new features for DPC++ runtime library, such as `cl::sycl::stream`, host side of `cl::sycl::half`, asynchronous handler, kernel and device description for sub-group. Fixed 20+ product issues.

### TensorFlow (MKL-DNN Backend)

2017/12–2018/05

*Software Engineer Intern*

- Pinpointed and fixed a critical issue in MKL-DNN backend, which seriously affected convergence.
- Proposed a bug pinpoint tool, CoSim, which is an operator-wise comparison tool that can locate the operator that first introduces large computing error between the target and reference. The corresponding poster has been published in the Intel Validation Summit 2018. With CoSim, two issues were found and fixed in the operator `AddN` that potentially affected convergence. In addition, fixed a number of failures in unit test.
- Enabled TensorFlow (MKL-DNN backend) on macOS.
- Designed and implemented a test framework for the MKL-DNN backend, and correspondingly added hundreds of unit test cases. No integration issue found any more since then.

### Intel Chainer

2017/03–2017/12

*Software Engineer Intern*

- Optimized dropout and softmax, yielding  $20\times$  and  $8\times$  performance improvements respectively.
- Designed and implemented a performance evaluation tool, *layer performance comparison* (LPC), which is to measure the performance of different layers among different frameworks. A large performance gap was found in the layer of `sum` and `relu` with the help of LPC. Corresponding optimization gained 10% performance improvement in GoogLeNet.

## TALKS

---

- **Shilei Tian**, Joseph Huber, Johannes Doerfert, “LTO and JIT Support in LLVM OpenMP Target Offloading”, in *LLVM Developers' Meeting*, November 16th–19th, 2021

- Johannes Doerfert, Joseph Huber, Stefan Stipanovic, Giorgis Georgakoudis, Hamilton Tobon Mosquera, **Shilei Tian**, “(OpenMP) Parallelism Aware Optimizations”, in *LLVM Developers’ Meeting*, October 6th–8th, 2020
- **Shilei Tian**, Johannes Doerfert, Barbara Chapman, “Asynchronous OpenMP Offloading on NVIDIA GPUs”, in *LLVM Performance Workshop*, February 23rd, 2020, San Diego, USA

## SKILLS

---

- Programming Languages: Proficient in C/C++; conversant in Python, Perl, and Shell Script
- Tools: Git, SVN,  $\LaTeX$
- Platform: Linux