Naifeng Zhang

Department of Electrical and Computer Engineering
College of Engineering
Carnegie Mellon University

naifengz@cmu.edu
+1 323 868 5267
naifengz.com

EDUCATION

Ph.D.   Electrical and Computer Engineering, Carnegie Mellon University, 2026
         Advisor: Prof. Franz Franchetti

B.S.   Computer Science, University of Southern California, 2021
         Advisor: Prof. Viktor K. Prasanna
         Thesis: Lightweight Augmented Neural Network For Performance Prediction and Its Applications
         W.V.T. Rusch Undergraduate Engineering Honors Program

B.S.   Mathematics, University of Southern California, 2021
         Departmental Honors Program

AWARDS

2024   Best Poster Runner-up
         Together with S. Fu (Lead Student) and F. Franchetti
         PRISM Annual Review, Systems & Software track

2024   First Place, ACM Student Research Competition
         Together with S. Fu (Lead Student) and F. Franchetti
         International Conference on Parallel Architecture and Compilation Techniques

2023   Outstanding Short Paper Award
         Together with P. Brinich, A. Ebel, F. Franchetti, and J. Johnson.
         IEEE High Performance Extreme Computing Conference

2023   Second Place, ACM Student Research Competition
         Together with F. Franchetti
         International Symposium on Code Generation and Optimization

2021   Discovery Scholar Distinction
         University of Southern California

2018–21 Academic Achievement Award
         University of Southern California

FELLOWSHIPS

2021–22 Carnegie Institute of Technology Dean's Fellowship

2019–21 University of Southern California Provost's Research Fellowship

GRANTS

2023-   *High-Performance Code Generation for Homomorphic Encryption on GPUs using SPIRAL*
Tuned and benchmarked SPIRAL-generated number theoretic transform (NTT) implementations for homomorphic encryption (HE) applications on start-of-the-art GPUs.
N. Zhang (PI), F. Franchetti (Co-PI)
200,000 ACCESS Credits
NSF

RESEARCH EXPERIENCE

2024-   *LLM Cerberus: Guardrails for Generative AI in High-Performance Math Kernels*
Extended SPIRAL with symbolic execution and theorem proving to derive semantics and provide correctness guarantees for LLM-generated math kernels.
NSF

2023-   *Code Synthesis for the PRISM Architecture*
Extended SPIRAL to target processing-in-memory (PIM) kernels on PRISM architectures.
SRC JUMP 2.0

2022-   *Neocortex: SPIRAL Code Generation for Wafer-Scale Engine*
Extended SPIRAL to target Cerebras second-generation Wafer-Scale Engine (WSE-2).
NSF

2023   *SciDAC: Simulation of the Response of Structural Metals in Molten Salt Environment*
Applied SPIRAL's formalism to optimize elasto-viscoplastic fast Fourier transform (EVPFFT) for materials science simulations.
DoE

2021-23   *Trebuchet: NTTX for OpenFHE*
Developed SPIRAL NTTX package to automatically generate high-performance vectorized number theoretic transform (NTT) code for fully homomorphic encryption (FHE) applications.
DARPA DPRIVE

2022–23   *Performance Analysis and Optimization of Quantum Library*
Conducted real-world performance analysis and optimizations of IBM's quantum library Qiskit for Shor's Algorithm.

2020–21   *Compiler Abstractions Supporting High Performance on Extreme-scale Resources (CASPER)*
Developed a compiler-oriented autotuner that automatically profiles a kernel and performs tuning guided by performance prediction.
DARPA PAPPA

2019   *Dynamic Data-Aware Reconfiguration, INtegration and Generation (DDARING)*
Developed a lightweight augmented neural network for performance prediction.
DARPA SDH

PUBLICATIONS

Conference Proceedings

2024   N. Zhang, F. Franchetti. "Code Generation for Cryptographic Kernels Using Multi-word Modular Arithmetic on GPU." International Symposium on Code Generation and Optimization (CGO). *Accepted with Shepherding.*

2023   N. Zhang, A. Ebel, N. Neda, P. Brinich, B. Reynwar, A. G. Schmidt, M. Franusich, J. Johnson, B. Reagen, F. Franchetti. "Generating High-Performance Number Theoretic Transform Implementations for Vector Architectures." IEEE High Performance Extreme Computing Conference (HPEC).

2023   D. Sun, N. Zhang, F. Franchetti. "Optimization and Performance Analysis of Shor's Algorithm in Qiskit." IEEE High Performance Extreme Computing Conference (HPEC).

2023   D. Soni, N. Neda, N. Zhang, B. Reynwar, H. Gamil, B. Heyman, M. N. T. Moopan, A. Al Badawi, Y. Polyakov, K. Canida, M. Pedram, M. Maniatakos, D. B. Cousins, F. Franchetti, M. French, A. Schmidt, B. Reagen. "RPU: The Ring Processing Unit." IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS).

2021   N. Zhang, A. Srivastava, R. Kannan, V. K. Prasanna. "GenMAT: A General-Purpose Machine Learning-Driven Auto-Tunerfor Heterogeneous Platforms." Workshop on Programming Environments for Heterogeneous Computing (PEHC), in conjunction with the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC).

2020   A. Srivastava*, N. Zhang*, R. Kannan, V. K. Prasanna. "Towards High Performance, Portability, and Productivity: Lightweight Augmented Neural Networks for Performance Prediction." International Conference on High Performance Computing, Data, and Analytics (HiPC). *Equal contribution*.

2020   C. Imes, A. Colin, N. Zhang, A. Srivastava, V. K. Prasanna, J. P. Walters. "Compiler Abstractions and Runtime for Extreme-scale SAR and CFD Workloads." Workshop on Extreme Scale Programming Models and Middleware (ESPM2), in conjunction with the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC).

Other Conference Papers, Technical Reports, Extended Abstracts, and Posters

2024   S. Fu, N. Zhang, F. Franchetti. "Accelerating High-Precision Number Theoretic Transforms using Intel AVX-512." The International Conference on Parallel Architectures and Compilation Techniques (PACT), Extended abstract with poster. **First Place, ACM Student Research Competition**. **Best Poster Runner-up** at PRISM Annual Review, Systems & Software track.

2024   Y. Eum, N. Zhang, L. Tang, F. Franchetti. "Towards a RISC-V Instruction Set Extension for Multi-word Arithmetic." IEEE High Performance Extreme Computing Conference (HPEC), Extended abstract with poster.

2024   N. Zhang, S. McAleer, T. Sandholm. "Faster Game Solving via Hyperparameter Schedules." arXiv, Preprint.

2023   P. Brinich, N. Zhang, A. Ebel, F. Franchetti, J. Johnson. "Twiddle Factor Generation for a Vectorized Number Theoretic Transform." IEEE High Performance Extreme Computing Conference (HPEC), Extended abstract with poster. **Outstanding Short Paper Award**.

2023   H. Mankad, A. Rovinelli, M. Zecevic, P. McCorquodale, F. Franchetti, N. Zhang, S. Rao, R. A. Lebensohn, L. Capolungo "EVPFFTX: A First Look at FFTX Applications in Material Science." IEEE High Performance Extreme Computing Conference (HPEC), Extended abstract with poster.

2023   D. B. Cousins, Y. Polyakov, A. Al Badawi, M. French, A. Schmidt, A. Jacob, B. Reynwar, K. Canida, A. Jaiswal, C. Mathew, H. Gamil, N. Neda, D. Soni, M. Maniatakos, B. Reagen, N. Zhang, F. Franchetti, P. Brinich, J. Johnson, P. Broderick, M. Franusich B. Zhang, Z. Cheng, M. Pedram. "TREBUCHET: Fully Homomorphic Encryption Accelerator for Deep Computation." Government Microcircuit Applications and Critical Technology Conference (GOMACTech), Preprint.

2023   N. Zhang, F. Franchetti. "Generating Number Theoretic Transforms for Multi-Word Integer Data Types." International Symposium on Code Generation and Optimization (CGO), Extended abstract with poster. **Second Place, ACM Student Research Competition**.

2022   N. Zhang, H. Gamil, P. Brinich, B. Reynwar, A. Al Badawi, N. Neda, D. Soni, K. Canida, Y. Polyakov, P. Broderick, M. Maniatakos, A. G. Schmidt, M. Franusich, J. Johnson, B. Reagen, D. B. Cousins, F.

Franchetti. "Towards Full-Stack Acceleration for Fully Homomorphic Encryption." IEEE High Performance Extreme Computing Conference (HPEC), Extended abstract with presentation.

2022   I. Grosof, N. Zhang, M. Heule. "Towards the shortest DRAT proof of the Pigeonhole Principle." Pragmatics of SAT International Workshop (PoS), Preprint.

TALKS

Conference Presentations

2023   "Generating High-Performance Number Theoretic Transform Implementations for Vector Architectures." IEEE High Performance Extreme Computing Conference. Virtual. Sep 29.

2023   "Generating Number Theoretic Transforms for Multi-Word Integer Data Types." IEEE/ACM International Symposium on Code Generation and Optimization. Montreal, Canada. Feb 28.

2022   "Towards Full-Stack Acceleration for Fully Homomorphic Encryption." IEEE High Performance Extreme Computing Conference. Virtual. Sep 23.

2021   "GenMAT: A General-Purpose Machine Learning-Driven Auto-Tuner for Heterogeneous Platforms." Workshop on Programming Environments for Heterogeneous Computing. Virtual. Nov 19.

2020   "Towards High Performance, Portability, and Productivity: Lightweight Augmented Neural Networks for Performance Prediction." International Conference on High Performance Computing, Data, and Analytics. Virtual. Dec 16.

Tutorials Given

2024   "Open Source SPIRAL 8.5 Tutorial." IEEE High Performance Extreme Computing Conference. Virtual. Sep 25. Together with F. Franchetti, M. Franusich.

2023   "Open Source SPIRAL 8.5 Tutorial." IEEE High Performance Extreme Computing Conference. Virtual. Sep 27. Together with F. Franchetti, M. Franusich, P. Broderick.

TEACHING EXPERIENCE

Carnegie Mellon Univeristy

Teaching Assistant

24 Fall   Mathematical Foundations of Electrical Engineering

23 Spring   Computational Problem Solving for Engineers

University of Southern California

Undergraduate Teaching Assistant

21 Spring   Special Topics - Accelerated Computing Using FPGAs

20 Fall   Parallel and Distributed Computation

20 Spring   Special Topics - Accelerated Computing Using FPGAs
Discrete Methods in Computer Science

19 Fall   Parallel and Distributed Computation
Discrete Methods in Computer Science

MENTORING

Master's

2024-	Yujun Lee
2023	Kofi Poku
2022–23	Dewang Sun
2022	Hongbo Sun

Undergraduate

2024-	Zubin Narayan
2024-	Misho Alexandrov
2024-	Sophia Fu
2024-	Youngjin Eum
2024-	Govind Malasani
2024	Steven Lee
2023	Gordon Xu
2022–23	Matt Ngaw
2022–23	Jimmy Zhou

SERVICE

Conference Peer Review

IEEE International Conference on Big Data (BigData)

IEEE High Performance Extreme Computing Conference (HPEC)

Service to the University

CMU ECE Faculty Hiring Student Council, 2022-

Outreach

CMU College of Engineering Graduate Student Outreach Committee, 2023-

Updated November 2024