

Mohamed S. Abdelfattah

Assistant Professor
Cornell Tech - Bloomberg Center 257
2 West Loop Road, New York, NY 10044

mohamed@cornell.edu
+1 (929) 866-7480
mohsaied.com

Professional Experience

Cornell University, New York, NY, USA

Since 2022 Assistant Professor of Electrical and Computer Engineering

Mako, New York, NY, USA

Since 2024 Co-Founder and Chief Science Officer

Samsung AI Center, Cambridge, UK

2022 - 2023 Visiting Professor, On-Device AI

2020 - 2022 Principal Scientist and Manager, Automated Machine Learning

2019 - 2020 Senior Scientist, Embedded Artificial Intelligence

Intel, Toronto, ON, Canada

2015 - 2018 Member of Technical Staff, Machine Learning Acceleration

Altera, Toronto, ON, Canada

Winter 2014 Research Intern, OpenCL Compiler

Philips Healthcare, Böblingen, Germany

Winter 2011 Research Intern, Patient Monitoring Hardware

Mentor Graphics, Cairo, Egypt

Summer 2008 Research Intern, Design and Simulation

Education

2011 - 2016 **University of Toronto**, Canada

PhD in Electrical and Computer Engineering

Dissertation: Architecture and CAD for Embedded Networks-on-Chip on FPGAs

Advisor: Prof. Vaughn Betz

2009 - 2011 **University of Stuttgart**, Germany

MSc in Information Technology

Thesis: Evaluation of Advanced Techniques for Structural FPGA Self-Test

Advisor: Prof. Hans-Joachim Wunderlich

2005 - 2009 **German University of Cairo**, Egypt

BSc in Electronics Engineering

Thesis: Design of an RF Transmitter for RFID Tags with Ultra-thin Silicon Substrates
Advisor: Prof. Manfred Berroth

Honours & Awards

- 2024 Nvidia AI Research Award
- 2024 NSF CAREER Award
- 2022 Meta Faculty Award in Networking for AI
- 2022 (Runner-up) Best Paper Award at the FPGA 2022 Symposium
- 2021 Samsung Best Paper Award Gold Medal (one gold medal is awarded annually in AI research across Samsung)
- 2016 (Runner-up) Michel Servit Best Paper Award at the FPL 2016 Conference
- 2016 University of Toronto TATP (University-wide) Teaching Award
- 2016 University of Toronto Faculty of Applied Science Teaching Award
- 2015 FPL 2013 paper selected as one of the most significant papers in the first 25 years of the conference
- 2015 Best Paper award at the FPGA 2015 Symposium
- 2015 University of Toronto Department of Electrical & Computer Engineering Teaching Award
- 2014 (Runner-up) Adel Sedra Distinguished Graduate Award
- 2014 University of Toronto Alumni Association Graduate Award
- 2013 - 2016 Vanier Canada Graduate Scholarship (Canada's highest doctoral award)
- 2013 Stamatis Vassiliadis Best Paper Award at the FPL 2013 Conference
- 2012, 2013 Right Track CAD Graduate Scholarship (twice)
- 2011 Connaught Doctoral Scholarship Award
- 2009 DAAD Master's Scholarship Award
- 2007 GUC Gerhard Schröder Scholarship Award

Academic Service

CONFERENCE ORGANIZING COMMITTEES

- 2024 International Conference on Automated Machine Learning (AutoML), Local Chair
- 2023 International Conference on Automated Machine Learning (AutoML), Online Experience Chair
- 2022 International Symposium On Computer Architecture (ISCA), Local Arrangements Chair
- 2022 International Symposium On Field-Programmable Custom Computing Machines (FCCM), Local Arrangements Chair

CONFERENCE TECHNICAL PROGRAM COMMITTEES

- Since 2021 International Conference on Learning Representations (ICLR)
- Since 2021 Conference on Neural Information Processing Systems (NeurIPS)
- Since 2019 International Symposium on Field-Programmable Gate-Arrays (FPGA)
- Since 2017 International Conference on Field Programmable Logic and Applications (FPL)
- 2023 Design Automation Conference (DAC)
- 2022 Design and Test Europe (DATE)
- 2018–2020 Artificial Intelligence Circuits and Systems (AICAS)

- 2017 International Conference on Field-Programmable Technology (FPT)
- 2016 International Workshop on Network on Chip Architectures (NocArc)

JOURNAL REVIEW

- Since 2023 Transactions on Machine Learning Research (TMLR)
- Since 2021 IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- Since 2018 IEEE Transactions on Computer-Aided Design (TCAD)
- Since 2016 IEEE Transactions on Computers (TCOMP)
- Since 2015 IEEE Transactions on Very Large Scale Integrated Circuits (TVLSI)
- Since 2015 ACM Transactions on Reconfigurable Technology and Systems (TRETs)

GRANT ADMINISTRATION

- 2017 - 2018 Grant Research Program Committee Member (Intel Representative)
NSERC Computing Hardware for Emerging Intelligent Sensing Applications (COHESA) Network.

Teaching

- 2022 - 2024 **ECE5545 Machine Learning Hardware and Systems**, Instructor, Cornell University
Enrollment: 24 (2022), 23 (2023), 40 (2024)
Course evaluation: 4.58/5 (2022), 4.38/5 (2023), 4.5/5 (2024)
Instructor evaluation: 4.7/5 (2022), 4.63/5 (2023), 4.64/5 (2023)
- 2024 **ECE6950 Special Topics on Deep Learning Efficiency**, Instructor, Cornell University
Enrollment: 8
Course evaluation: 4.4/5
Instructor evaluation: 4.75/5
- 2023 **ECE6950 Reconfigurable Computing and FPGA Architecture**, Instructor, Cornell University
Enrollment: 10
Course evaluation: 4.5/5
Instructor evaluation: 4.5/5
- 2015, 2016 **ECE297 Communication and Design**, Head Teaching Assistant, University of Toronto
Enrollment: 350 (both years)
Course evaluation: 4.4/5 (2015), 4.3/5 (2016)
My TA evaluation: 7/7 (both years)
- 2012 - 2014 **ECE241 Digital Logic Design**, Teaching Assistant, University of Toronto
- 2012 - 2014 **ECE243 Computer Organization**, Teaching Assistant, University of Toronto

Publications

CONFERENCE PROCEEDINGS

- 2025 **Palu: Compressing KV-Cache with Low-Rank Projection**
Chi-Chih Chang, Wei-Cheng Lin, Chien-Yu Lin, Chong-Yan Chen, Yu-Fang Hu, Pei-Shuo Wang, Ning-Chi Huang, Luis Ceze, Mohamed S. Abdelfattah, Kai-Chiang Wu
International Conference on Learning Representations (ICLR)

- 2025 **BitMoD: Bit-serial Mixture-of-Datatype LLM Acceleration**
Yuzong Chen, Ahmed F. AbouElhamayed, Xilai Dai, Yang Wang, Marta Andronic, George A. Constantinides, Mohamed S. Abdelfattah
IEEE International Symposium on High-Performance Computer Architecture (HPCA)
- 2024 **ShadowLLM: Predictor-based Contextual Sparsity for Large Language Models**
Yash Akhauri, Ahmed F. AbouElhamayed, Jordan Dotzel, Zhiru Zhang, Alexander M. Rush, Safeen Huda, Mohamed S. Abdelfattah
Conference on Empirical Methods in Natural Language Processing (EMNLP)
- 2024 **BBS: Bi-directional Bit-level Sparsity for Deep Learning Acceleration**
Yuzong Chen, Jian Meng, Jae-sun Seo, Mohamed S. Abdelfattah
IEEE/ACM International Symposium on Microarchitecture (MICRO)
- 2024 **FLIQS: One-Shot Mixed-Precision Floating-Point and Integer Quantization Search**
Jordan Dotzel, Gang Wu, Andrew Li, Muhammad Umar, Yun Ni, Mohamed S. Abdelfattah, Zhiru Zhang, Liqun Cheng, Martin G. Dixon, Norman P. Jouppi, Quoc V. Le, Sheng Li
International Conference on Automated Machine Learning (AutoML)
- 2024 **Kratos: An FPGA Benchmark for Unrolled DNNs with Fine-Grained Sparsity and Mixed Precision**
Xilai Dai, Yuzong Chen, Mohamed S. Abdelfattah
International Conference on Field-Programmable Logic and Applications (FPL)
- 2024 **Encodings for Prediction-based Neural Architecture Search**
Yash Akhauri, Mohamed S. Abdelfattah
International Conference on Machine Learning (ICML)
- 2024 **Learning from Students: Applying t-Distributions to Explore Accurate and Efficient Formats for LLMs**
Jordan Dotzel, Yuzong Chen, Bahaa Kotb, Sushma Prasad, Gang Wu, Sheng Li, Mohamed S. Abdelfattah, Zhiru Zhang
International Conference on Machine Learning (ICML)
- 2024 **Towards Neural Architecture Search through Hierarchical Generative Modeling**
Lichuan Xiang, Łukasz Dudziak, Mohamed S. Abdelfattah, Abhinav Mehrotra, Nicholas D. Lane, Hongkai Wen
International Conference on Machine Learning (ICML)
- 2024 **Beyond Inference: Performance Analysis of DNN Server Overheads for Computer Vision**
Ahmed F. AbouElhamayed, Susanne Balle, Deshanand Singh, Mohamed S. Abdelfattah
Design Automation Conference (DAC)
- 2024 **On Latency Predictors for Neural Architecture Search**
Yash Akhauri, Mohamed S. Abdelfattah
International Conference on Machine Learning and Systems (MLSYS)
- 2023 **M4BRAM: Mixed-Precision Matrix-Matrix Multiplication in FPGA Block RAMs**
Yuzong Chen, Jordan Dotzel, Mohamed S. Abdelfattah
International Conference on Field-Programmable Technology (FPT)
- 2023 **DiviML: A Module-based Heuristic for Mapping Neural Networks onto Heterogeneous Platforms**
Yassine Ghannane, Mohamed S. Abdelfattah
International Conference on Computer-Aided Design (ICCAD)
- 2023 **Multi-Predict: Few Shot Predictors for Efficient Neural Architecture Search**
Yash Akhauri, Mohamed S. Abdelfattah
International Conference on Automated Machine Learning (AutoML)

- 2023 **BRAMAC: Compute-in-BRAM Architectures for Multiply-Accumulate on FPGAs**
Yuzong Chen, Mohamed S. Abdelfattah
International Symposium On Field-Programmable Custom Computing Machines (FCCM)
- 2023 **Zero-Cost Operation Scoring in Differentiable Architecture Search**
Lichuan Xiang, Łukasz Dudziak, Mohamed S. Abdelfattah, Thomas Chau, Nicholas D. Lane, Hongkai Wen
The AAAI Conference on Artificial Intelligence (AAAI)
- 2022 **BLOX: Macro Neural Architecture Search Benchmark and Algorithms**
Thomas Chau, Łukasz Dudziak, Hongkai Wen, Nicholas D. Lane, Mohamed S. Abdelfattah
Conference on Neural Information Processing Systems (NeurIPS)
- 2022 **Adaptable Butterfly Accelerator for Attention-based NNs via Hardware and Algorithm Co-design**
Hongxiang Fan, Thomas Chau, Stylianos Venieris, Royson Lee, Alexandros Kouris, Wayne Luk, Nicholas D. Lane, Mohamed S. Abdelfattah
IEEE/ACM International Symposium on Microarchitecture (MICRO)
- 2022 **Logic Shrinkage: Learned FPGA Netlist Sparsity for Efficient Neural Network Inference**
Erwei Wang, James Davis, Georgios Stavrou, Peter Cheung, George Constantinides, Mohamed S. Abdelfattah
International Symposium on Field-Programmable Gate Arrays (FPGA)
- 2021 **Temporal Kernel Estimation for Blind Video Super-Resolution**
Lichuan Chang, Royson Lee, Hongkai Wen, Mohamed S. Abdelfattah, Nicholas D. Lane
International Conference on Computer Vision (ICCV) Workshop
- 2021 **Zero-Cost Proxies for Lightweight NAS**
Mohamed S. Abdelfattah, Abhinav Mehrotra, Łukasz Dudziak, Nicholas D. Lane
International Conference on Learning Representations (ICLR)
- 2021 **NAS-Bench-ASR: Reproducible Neural Architecture Search for Speech Recognition**
Abhinav Mehrotra, Alberto Gil C. P. Ramos, Sourav Bhattacharya, Łukasz Dudziak, Ravichander Vipperla, Thomas Chau, Mohamed S. Abdelfattah, Samin Ishtiaq, Nicholas D. Lane
International Conference on Learning Representations (ICLR)
- 2020 **BRP-NAS: Prediction-based NAS using GCNs**
Łukasz Dudziak, Thomas Chau, Mohamed S. Abdelfattah, Royson Lee, Hyeji Kim, Nicholas D. Lane
Conference on Neural Information Processing Systems (NeurIPS)
- 2020 **Iterative Compression of End-to-End ASR Model Using Reinforcement learning**
Abhinav Mehrotra, Łukasz Dudziak, Jinsu Yeo, Younyoon Lee, Ravichander Vipperla, Mohamed S. Abdelfattah, Sangjeong Lee, Daehyun Kim, Nicholas D. Lane
Conference of the International Speech Communication Association (INTERSPEECH)
- 2020 **Journey Towards Tiny Perceptual Super-Resolution**
Royson Lee, Łukasz Dudziak, Mohamed S. Abdelfattah, Hyeji Kim, Stylianos Veneris, Hongkai Wen, Nicholas D. Lane
European Conference on Computer Vision (ECCV)
- 2020 **Best of Both Worlds: AutoML Codesign of a CNN and its Hardware Accelerator**
Mohamed S. Abdelfattah, Łukasz Dudziak, Thomas Chau, Royson Lee, Hyeji Kim, Nicholas D. Lane
Design Automation Conference (DAC)
- 2019 **ShrinkML: End-to-End ASR Model Compression Using Reinforcement Learning**
Łukasz Dudziak*, Mohamed S. Abdelfattah*, Ravichander Vipperla, Stefanos Laskaridis, Nicholas D. Lane
Conference of the International Speech Communication Association (INTERSPEECH)
- 2018 **DLA: Compiler and FPGA Overlay for Neural Network Inference Acceleration**
Mohamed S. Abdelfattah, David Han, Andrew Bitar, Roberto DiCecco, Shane O'Connell, Nitika Shanker,

- Joseph Chu, Ian Prins, Joshua Fender, Andrew C. Ling and Gordon R. Chiu
International Conference on Field-Programmable Logic and Applications (FPL)
- 2018 **Harnessing Numerical Flexibility for Deep Learning on FPGAs**
Andrew C. Ling, Mohamed S. Abdelfattah, Shane O'Connell, Andrew Bitar, David Han, Roberto Dicecco, Suchit Subhaschandra, Chris N Johnson, Dmitry Denisenko, Josh Fender, Gordon R. Chiu
International Symposium on Highly-Efficient Accelerators and Reconfigurable Technologies (HEART)
- 2018 **Flexibility: FPGAs and CAD in Deep Learning Acceleration**
Gordon R. Chiu, Andrew C. Ling, Davor Capalija, Andrew Bitar, Mohamed S. Abdelfattah
International Symposium on Physical Design (ISPD)
- 2016 **LYNX: CAD for FPGA-based networks-on-chip**
Mohamed S. Abdelfattah, Vaughn Betz
International Conference on Field-Programmable Logic and Applications (FPL)
- 2015 **Bringing programmability to the data plane: Packet processing with a NoC-enhanced FPGA**
Andrew Bitar, Mohamed S. Abdelfattah, Vaughn Betz
International Conference on Field-Programmable Technology (FPT)
- 2015 **Design and simulation tools for Embedded NOCs on FPGAs**
Mohamed S. Abdelfattah, Andrew Bitar, Ange Yaghi, Vaughn Betz
International Conference on Field-Programmable Logic and Applications (FPL)
- 2015 **Take the highway: Design for embedded NoCs on FPGAs**
Mohamed S. Abdelfattah, Andrew Bitar, Vaughn Betz
International Symposium on Field-Programmable Gate Arrays (FPGA)
- 2014 **Gzip on a chip: High performance lossless data compression on fpgas using openc1**
Mohamed S. Abdelfattah, Andrei Hagiescu, Deshanand Singh
International Workshop on OpenCL (IWOCL)
- 2013 **Augmenting FPGAs with Embedded Networks on Chip**
Mohamed S. Abdelfattah, Vaughn Betz
Workshop on the Intersection of Computer Architecture and Reconfigurable Logic (CARL)
- 2013 **The Power of Communication: Energy-Efficient NoCs for FPGAs**
Mohamed S. Abdelfattah, Vaughn Betz
International Conference on Field-Programmable Logic and Applications (FPL)
- 2012 **Design tradeoffs for hard and soft FPGA-based Networks-on-Chip**
Mohamed S. Abdelfattah, Vaughn Betz
International Conference on Field-Programmable Technology (FPT)
- 2012 **Transparent Structural Online Test for Reconfigurable Systems**
Mohamed S. Abdelfattah, Lars Bauer, Claus Braun, Michael E Imhof, Michael A Kochte, Hongyan Zhang, Jörg Henkel, Hans-Joachim Wunderlich
International Symposium on On-Line Testing and Robust System Design (IOLTS)
- 2011 **2.2 GHz LC VCO for RFID on a 0.5- μ m digital gate-array designed for ultra-thin silicon substrates**
Mohamed S. Abdelfattah, Damir Ferenci, Markus Grözing, Manfred Berroth, Cor Scherjon, Joachim Burghartz
German Microwave Conference (GeMiC)
- 2009 **Design of a RF Transmitter for RFID Tags in a New Technology with Ultra Thin Silicon Substrates**
Mohamed S. Abdelfattah, Damir Ferenci, Markus Grözing, Manfred Berroth, Cor Scherjon, Joachim N Burghartz
Workshop on Circuit Design and Digital Signal Processing (ProRISC)

JOURNAL ARTICLES

- 2024 **PQA: Exploring the Potential of Product Quantization in DNN Hardware Acceleration**
Ahmed F. AbouElhamayed, Angela Cui, Javier Fernandez-Marques, Nicholas D. Lane, Mohamed S. Abdelfattah
ACM Transactions on Reconfigurable Technology and Systems (TRETs)
- 2023 **Learned Connectivity Sparsification for LUT-based Neural Networks**
Erwei Wang, Georgios Stavrou, Peter Cheung, George Constantinides, Mohamed S. Abdelfattah, James Davis
ACM Transactions on Reconfigurable Technology and Systems (TRETs)
- 2016 **Design and applications for embedded networks-on-chip on FPGAs**
Mohamed S. Abdelfattah, Andrew Bitar, Vaughn Betz
IEEE Transactions on Computers (TCOMP)
- 2015 **Power analysis of embedded NoCs on FPGAs and comparison with custom buses**
Mohamed S. Abdelfattah, Vaughn Betz
IEEE Transactions on Very Large-Scale Integration Systems (TVLSI)
- 2014 **Networks-on-Chip for FPGAs: Hard, Soft or Mixed?**
Mohamed S. Abdelfattah, Vaughn Betz
ACM Transactions on Reconfigurable Technology and Systems (TRETs)
- 2014 **The Case for Embedded Networks-on-Chip on FPGAs**
Mohamed S. Abdelfattah, Vaughn Betz
IEEE Micro Magazine

BOOK CHAPTERS

- 2015 **Embedded Networks-on-Chip for FPGAs**
Mohamed S. Abdelfattah, Vaughn Betz
Reconfigurable Logic: Architecture, Tools and Applications

PATENTS

- 2020 **Method and Apparatus for Neural Architecture Search**
Mohamed S. Abdelfattah, Łukasz Dudziak, Abhinav Mehrotra
Application Number: UK2015231.0
- 2020 **Method and Apparatus for Analysing Neural Network Performance**
Thomas Chau, Łukasz Dudziak, Mohamed S. Abdelfattah, Royson Lee, Nicholas D. Lane
Application Number: UK20199106.4
- 2019 **Method for Designing Accelerator Hardware**
Mohamed S. Abdelfattah, Łukasz Dudziak, Thomas Chau, Royson Lee, Hyeji Kim, Sourav Battacharaya
Application Number: UK1913353.7
- 2015 **Field Programmable Gate-Array with Embedded Network-on-Chip Hardware and Design Flow**
Mohamed S. Abdelfattah, Vaughn Betz
Application Number: US14060253