Andrew Boutros

Curriculum Vitae

Personal Information

Employment Authorization: Permanent Resident of Canada
Email: andrew.m.boutros@gmail.com
Phone Number: +1 (647) 818-3248
Web Page: https://www.aboutros.info/

Education

Ph.D. in Electrical and Computer Engineering *Sep 2019 - Present (Expected Sep 2024)*
University of Toronto Ontario, Canada
Thesis: Reconfigurable Computing for Deep Learning
Advisor: Vaughn Betz
Grade-point Average: 4.0/4.0

M.A.Sc. in Electrical and Computer Engineering *Sep 2016 - Aug 2018*
University of Toronto Ontario, Canada
Thesis: Enhancing FPGA Architecture for Efficient Deep Learning Inference
Advisor: Vaughn Betz
Grade-point Average: 4.0/4.0

B.Sc. in Electronics Engineering *Sep 2011 - Jul 2016*
German University in Cario Cairo, Egypt
Thesis: Pedestrian Detection Based on the HOG Algorithm on a Xilinx Zynq FPGA
Advisor: Diana Göhringer @ Ruhr University Bochum
Grade-point Average: 4.0/4.0 (0.72 on the German grading scale)

Academic Experience

Research Assistant *Sep 2016 – Aug 2018, Sep 2019 – Present*
University of Toronto Ontario, Canada

- Designed efficient deep learning accelerators on current FPGA architectures. The neural processing unit (NPU) accelerator I developed is now being productized by Intel's Programmable Solutions Group.
- Enhanced the FPGA fabric for deep learning by proposing micro-architecture improvements to the FPGA's logic and digital signal processing (DSP) blocks. Similar DSP block enhancements were incorporated in the Intel Agilex FPGA family.
- Developed and open-sourced the RAD-Sim and RAD-Gen tools for architecture exploration of next-generation reconfigurable acceleration devices (RADs) that incorporate conventional FPGAs, specialized accelerator cores, and networks-on-chip for system-level communication.
- Curated and released the Koios deep-learning-targeted benchmark suite for FPGA CAD and architecture research.
- Led the architecture research vector of the Intel/VMWare Crossroads center for 3D FPGAs, working with more than 10 PhD/MASc students from Carnegie Mellon University, the University of Texas at Austin, and the University of Toronto.
- Mentored 1 PhD, 2 MASc, and 12 undergraduate students on different research/capstone projects and internships.
- Selected as one of 32 students from all Canadian universities as a postgraduate affiliate of the Vector Institute for AI.
- Affiliated to the International Centre for Spatial Computational Learning.

Research Intern *Jul 2016 – Aug 2016*
German University in Cairo & Cairo University Cairo, Egypt

- Implemented new hardware-friendly chaos-based image encryption algorithms and accelerated them using FPGAs.
- Developed VPR architecture description files and bitstream generation for a research prototype FPGA chip.

Visiting Research Student *Mar 2015 – Jul 2015*
Ruhr University Bochum Bochum, Germany

- Implemented an FPGA-based real-time pedestrian detection system for autonomous vehicles.

Industrial Experience

Machine Learning Systems Engineer *Aug 2022 – Present*
MangoBoost Inc. Ontario, Canada

- Developed and optimized FPGA-based data processing unit (DPU) IPs for accelerating datacenter infrastructure operations such as networking, storage, and deep learning inference pre-processing.
- Architected a CAD toolflow for automated DPU customization and generation. The toolflow takes a high-level description of the required DPU IP composition as an input and produces an optimized FPGA bitstream for it.

Research Scientist *Nov 2019 – Aug 2022*
CTO Office, Programmable Solutions Group, Intel Corp. Ontario, Canada

- Developed and maintained the Intel neural processing unit (NPU) FPGA overlay architecture for accelerating deep learning sequence model inference (RNNs, GRUs, LSTMs). This research prototype achieved an order of magnitude higher performance than same-generation Nvidia GPUs and is now being productized by Intel.
- Guided academic research collaborations between Intel and different universities (e.g. Carnegie Mellon University, Seoul National University, Imperial College London, University of California Los Angeles, University of Michigan, University of Texas at Austin) on topics related to deep learning acceleration and FPGA fabric architecture enhancements.

Research Scientist *Aug 2018 – Jun 2019*
Accelerator Architecture Lab, Intel Labs, Intel Corp. Oregon, USA

- Performed path-finding research for Intel's next-generation reconfigurable architectures. I was part of a team responsible for defining the architecture and evaluating its performance and efficiency for deep learning workloads.

Teaching Experience

Tutorial & Lab Instructor for ECE297: Software Communication and Design *Winter 2017-2018, 2020-2022*
University of Toronto

- Developed and presented weekly tutorials on algorithms, data structures, efficient C++ coding, and graphical user interfaces (for the 2022 appointment).
- Mentored 10 teams of second-year ECE students per semester through implementing a Google-Maps-like navigation software using C++ and the EasyGL graphics library.
- Graded project milestone submissions and gave detailed feedback on coding style, demos, and final presentations.
- Received teaching evaluations well above the department average (6.5 out of 7.0 average across 51 anonymous reviews).

Teaching Assistant for ECE1756: Reconfigurable Computing & FPGA Architecture *Fall 2019-2021*
University of Toronto

- Added new content to the course lectures and redesigned its assignments to target the deep learning acceleration domain.
- Implemented a procedure to enable live streaming and recording the lectures at a high quality with the instructor, chalk board explanations, and lecture slides all clearly visible for remote students during the Covid-19 pandemic. This made the course more accessible and was continued even after the pandemic ended.
- Graded weekly reading questionnaires and several FPGA design/CAD assignments for a class of 30-35 graduate students.

Lab Instructor for ECE244: Programming Fundamentals *Fall 2019-2021*
University of Toronto

- Ran 2 weekly lab sessions on the fundamentals of C++ programming and data structures for second-year ECE students.

Teaching Assistant for ECE1508: Introduction to Statistical Learning *Fall 2017*
University of Toronto

- Graded weekly problem sets, programming assignments, midterm/final exams, and final course projects for a class of 47 graduate students.

Lab Instructor for ECE241: Digital Systems *Fall 2017*
University of Toronto

- Ran a weekly lab session for second-year ECE students on the fundamentals of digital logic and FPGA design.
- Mentored 6 teams through final course projects of implementing full digital systems of their choice on DE1-SoC boards.

Junior Teaching Assistant for ELCT707: Microcomputer Applications *Winter 2015*
German University in Cairo

- Selected as a top senior-year undergraduate student to be appointed as a junior teaching assistant.
- Developed and presented tutorials on basics of RTL design for third-year electronics engineering students.
- Mentored 5 teams through the final course project of implementing a 5-stage MIPS processor.

Scholarships, Honors & Awards

2023	**Best Paper Award** at the FPT 2023 conference
2018, 2023	University of Toronto **Right Track CAD Graduate Scholarship** valued at $4,500 (twice)
2021	**Best Paper Award** at the ICM 2021 conference
2021	Project selected as **one of the best projects** presented at Intel's DTTC 2021 conference
2017, 2019	**Vector Institute Student Affiliation** with an honorarium valued at $6,000 (twice)
2018	**Stamatis Vassiliadis Best Paper Award** at the FPL 2018 conference
2015	**Best Paper Award** at the ReConfig 2015 conference
2011, 2013, 2015	German University in Cairo **Top-Ranked Student Award** valued at $2,500 (3 times)
2011, 2013	German University in Cairo **Academic Excellence Award** valued at $2,000 (twice)
2011-2015	German University in Cairo **High-School Excellence Award** valued at $3,600 per year

Publications[1]

Summary[2]

Book Chapters	1
Refereed Journal Articles	10
Refereed Conference Papers	18
Best Paper Awards	4
Best Paper Nominations	7
Pending U.S. Patent Applications	2
Conference Workshops & Tutorials	6
Invited Talks	6
Collaborators & Co-authors	82

Citation Count Summary[3]

h-Index	15
i10-Index	17
Total Citations	829

Book Chapters

[1] **A. Boutros**, V. Betz, "Field-Programmable Gate Array Architecture", *Handbook of Computer Architecture, Springer*, 2021

Refereed Journal Articles

[2] Anupreetham, M. Ibrahim, M. Hall, **A. Boutros**, A. Kuzhively, A. Mohanty, E. Nurvitadhi, V. Betz, Y. Cao, and J. Seo, "High Throughput FPGA-Based Object Detection via Algorithm-Hardware Co-Design", *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, 2023

[3] A. Arora, **A. Boutros**, S. A. Damghani, K. Mathur, V. Mohanty, T. Anand, M. A. Elgammal, K. B. Kent, V. Betz, and L. K. John, "Koios 2.0: Open-Source Deep Learning Benchmarks for FPGA Architecture and CAD Research", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 2023

[4] S. Hur, S. Na, D. Kwon, J. Kim, J. Kim, **A. Boutros**, and E. Nurvitadhi, "A Fast and Flexible FPGA-based Accelerator for Natural Language Processing Neural Networks", *ACM Transactions on Architecture and Code Optimization (TACO)*, 2022

[5] **A. Boutros**, E. Nurvitadhi, and V. Betz, "Architecture and Application Co-Design for Beyond-FPGA Reconfigurable Acceleration Devices", *IEEE Access*, 2022

[6] R. Ma, E. Georganas, A. Heinecke, **A. Boutros**, and E. Nurvitadhi, "FPGA-based AI Smart NICs for Scalable Distributed AI Training Systems", *IEEE Computer Architecture Letters (CAL)*, 2022

[7] Z. Que, H. Nakahara, E. Nurvitadhi, **A. Boutros**, H. Fan, C. Zeng, J. Meng, K. H. Tsoi, X. Niu, and W. Luk, "Recurrent Neural Networks with Column-wise Matrix-Vector Multiplication on FPGAs", *IEEE Transactions on Very Large Scale Integration (VLSI)*, 2021

[8] **A. Boutros** and V. Betz, "FPGA Architecture: Principles and Progression", *IEEE Circuits and Systems Magazine (CAS-M)*, 2021

[9] M. Eldafrawy, **A. Boutros**, S. Yazdanshenas, and V. Betz, "FPGA Logic Block Architectures for Efficient Deep Learning Inference", *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, 2020

[10] **A. Boutros**, S. Yazdanshenas, and V. Betz, "You Can't Improve What You Don't Measure: FPGA vs. ASIC Efficiency Gaps for CNN Inference", *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*, 2018

[1] The symbol * indicates equal contribution.
[2] Counts include published, accepted, and submitted works.
[3] From Google Scholar (last updated August 7, 2024)

[11] J. Rettkowski, **A. Boutros**, and D. Goehringer, "A HW/SW Co-Design of the HOG Algorithm on Xilinx Zynq SoC", *Springer Journal of Parallel and Distributed Computing (JPDC)*, 2017

Refereed Conference Papers

[12] T. Zhang, **A. Boutros**, S. Gribok, K. Boateng, and V. Betz, "A Software-Programmable Neural Processing Unit for Graph Neural Network Inference on FPGAs", *International Conference on Field Programmable Logic and Applications (FPL)*, 2024 **(Accepted)**

[13] **A. Boutros***, F. Mahmoudi*, A. Mohaghegh*, S. More*, and V. Betz, "Into the Third Dimension: Architecture Exploration Tools for 3D Reconfigurable Acceleration Devices", *International Conference on Field Programmable Technology (FPT)*, 2023 **(Acceptance Rate 25%)** **Received Best Paper Award**

[14] **A. Boutros**, S. More, and V. Betz, "A Whole New World: How to Architect Beyond-FPGA Reconfigurable Acceleration Devices?", *International Symposium on Field-Programmable Logic and Applications (FPL)*, 2023 **(Acceptance Rate 36%)**

[15] S. Srinivasan*, **A. Boutros***, F. Mahmoudi, and V. Betz, "Placement Optimization for NoC-Enhanced FPGAs", *International Symposium on Field-Programmable Custom Computing Machines (FCCM)*, 2023 **(Acceptance Rate 26%)** **Nominated for Best Paper Award**

[16] **A. Boutros**, E. Nurvitadhi, and V. Betz, "RAD-Sim: Rapid Architecture Exploration for Novel Reconfigurable Acceleration Devices", *International Symposium on Field-Programmable Logic and Applications (FPL)*, 2022

[17] **A. Boutros**, E. Nurvitadhi, and V. Betz, "Specializing for Efficiency: Customizing AI Inference Processors on FPGAs", *International Conference on Microelectronics (ICM)*, 2021 **Received Best Paper Award**

[18] A. Arora, **A. Boutros**, D. Rauch, A. Rajen, A. Borda, S. A. Damghani, S. Mehta, S. Kate, P. Patel, K. B. Kent, V. Betz, and L. K. John, "Koios: A Deep Learning Benchmark Suite for FPGA Architecture and CAD Research", *International Symposium on Field-Programmable Logic and Applications (FPL)*, 2021 **(Acceptance Rate 22%)**

[19] Anupreetham, M. Ibrahim, M. Hall, **A. Boutros**, A. Kuzhively, A. Mohanty, E. Nurvitadhi, V. Betz, Y. Cao, and J. Seo, "End-to-End FPGA-based Object Detection Using Pipelined CNN and Non-Maximum Suppression", *International Symposium on Field-Programmable Logic and Applications (FPL)*, 2021 **(Acceptance Rate 22%)** **Nominated for Best Paper Award**

[20] X. Wang, V. Goyal, J. Yu, V. Bertacco, **A. Boutros**, E. Nurvitadhi, C. Augustine, R. Iyer, and R. Das, "Compute-Capable Block RAMs for Efficient Deep Learning Acceleration on FPGAs", *International Symposium on Field-Programmable Custom Computing Machines (FCCM)*, 2021 **(Acceptance Rate 22%)**

[21] **A. Boutros**, E. Nurvitadhi, R. Ma, S. Gribok, Z. Zhao, J. C. Hoe, V. Betz, and M. Langhammer, "Beyond Peak Performance: Comparing the Real Performance of AI-Optimized FPGAs and GPUs", *International Conference on Field-Programmable Technology (FPT)*, 2020 **(Acceptance Rate 25%)** **Featured on Intel Stratix 10 NX official webpage and Intel whitepaper**

[22] **A. Boutros**, M. Hall, N. Papernot, and V. Betz, "Neighbors From Hell: Voltage Attacks Against Deep Learning Accelerators on Multi-Tenant FPGAs", *International Conference on Field-Programmable Technology (FPT)*, 2020 **(Acceptance Rate 25%)** **Nominated for Best Paper Award**

[23] E. Nurvitadhi, **A. Boutros**, P. Budhkar, A. Jafari, D. Kwon, D. Sheffield, A. Prabhakaran, K. Gururaj, P. Appana, and M. Naik, "Scalable Low-Latency Persistent Neural Machine Translation on CPU Server with Multiple FPGAs", *International Conference on Field-Programmable Technology (FPT)*, 2019

[24] E. Nurvitadhi, D. Kwon, A. Jafari, **A. Boutros**, J. Sim, P. Tomson, H. Sumbul, G. Chen, P. Knag, R. Kumar, R. Krishnamurthy, S. Gribok, B. Pasca, M. Langhammer, D. Marr, and A. Dasu, "Why Compete When You Can Work Together: FPGA-ASIC Integration for Persistent RNNs", *International Symposium on Field-Programmable Custom Computing Machines (FCCM)*, 2019 **(Acceptance Rate 26%)**

[25] **A. Boutros***, M. Eldafrawy*, S. Yazdanshenas, and V. Betz, "Math Doesn't Have to be Hard: Logic Block Architectures to Enhance Low-Precision Multiply-Accumulate on FPGAs", *International Symposium on Field-Programmable Gate Arrays (FPGA)*, 2019 **(Acceptance Rate 19%)**

[26] **A. Boutros**, S. Yazdanshenas, and V. Betz, "Embracing Diversity: Enhanced DSP Blocks for Low-Precision Deep Learning on FPGAs", *International Conference on Field-Programmable Logic and Applications (FPL)*, 2018 **(Acceptance Rate 17%)** **Received Best Paper Award**

[27] **A. Boutros**, B. Grady, M. Abbas, and P. Chow, "Build Fast, Trade Fast: FPGA-based High-Frequency Trading using High-Level Synthesis", *International Conference on Reconfigurable Computing and FPGAs (ReConFig)*, 2017

[28] **A. Boutros**, S. Hesham, B. Georgey, and M. A. Elghany, "Hardware Acceleration of Novel Chaos-Based Image Encryption for IoT Applications", *International Conference on Microelectronics (ICM)*, 2017

[29] J. Rettkowski, **A. Boutros**, and D. Goehringer, "Real-time Pedestrian Detection on a Xilinx Zynq using the HOG Algorithm", *International Conference on Reconfigurable Computing and FPGAs (ReConFig)*, 2015 **Received Best Paper Award**

Conference Demos & Posters

[29] M. Hall, **A. Boutros**, E. Nurvitadhi, and V. Betz, "Programmable Hardware for Efficient Deep Learning Inference", *The Vector Institute Symposium on Evolution of Deep Learning*, 2019

[30] E. Nurvitadhi, D. Kwon, A. Jafari, **A. Boutros**, J. Sim, P. Tomson, H. Sumbul, G. Chen, P. Knag, R. Kumar, R. Krishnamurthy, S. Gribok, B. Pasca, M. Langhammer, D. Marr, and A. Dasu, "Evaluating and Enhancing Intel Stratix 10 FPGAs for Persistent Real-Time Artificial Intelligence", *International Symposium on Field-Programmable Gate Arrays (FPGA)*, 2019

Conference Workshops & Tutorials

[31] **A. Boutros** and V. Betz, "FPGA Architecture for Deep Learning", *International Symposium on Field-Programmable Custom Computing Machines (FCCM)*, 2023

[32] **A. Boutros**, A. Arora, S. Rasoulinezhad, V. Betz, L. K. John and P. H. Leong, "FPGA Architecture for Deep Learning", *International Symposium on Field-Programmable Gate Arrays (FPGA)*, 2023

[33] A. Arora, **A. Boutros**, S. Rasoulinezhad, V. Betz, L. K. John and P. H. Leong, "FPGA Architecture for Deep Learning", *International Symposium on Microarchitecture (MICRO)*, 2022

[34] E. Nurvitadhi, **A. Boutros**, R. D'Souza, and T. Vanderhoek, "AI-Optimized Intel Stratix 10 NX FPGA", *International Conference on Field-Programmable Logic and Applications (FPL)*, 2021

[35] E. Nurvitadhi, **A. Boutros**, R. D'Souza, and T. Vanderhoek, "AI-Optimized Intel Stratix 10 NX FPGA", *International Symposium on Field-Programmable Custom Computing Machines (FCCM)*, 2021

[36] E. Nurvitadhi, **A. Boutros**, R. D'Souza, and T. Vanderhoek, "AI-Optimized Intel Stratix 10 NX FPGA", *International Symposium on Field-Programmable Gate Arrays (FPGA)*, 2021

Patent Applications

[37] E. Nurvitadhi, M. Langhammer, **A. Boutros**, "Systems and Methods for Processor Circuits", *US Patent App. 17/561,534*, 2022

[38] E. Nurvitadhi, R. Poornachandran, A. Davare, N. Jain, C. Lacewell, A. Bhiwandiwalla, J. P. Munoz, **A. Boutros**, Y. Akhauri, "Apparatus, Articles of Manufacture & Methods for Composable ML Compute Nodes", *US Patent App. 17/558,284*, 2022

Invited Talks

[39] **A. Boutros**, "CAD and Architecture Exploration Tools for Next-Generation Reconfigurable Acceleration Devices", *Intel/VMWare Crossroads 3D FPGA Research Center Seminar Series*, Oct 2023

[40] **A. Boutros**, "Reconfigurable Computing for Deep Learning: Current, Future & Beyond FPGAs", *SpatialML Annual Meeting*, Feb 2023 (33 attendees)

[41] **A. Boutros**, "RAD-Sim: Architecture Exploration for Novel Reconfigurable Acceleration Devices", *Intel/VMWare Crossroads 3D FPGA Research Center Seminar Series*, Oct 2022 (25 attendees)

[42] M. Elgammal and **A. Boutros**, "Computer Hardware for AI", *Information Technology Institute in Egypt*, Jul 2022 (106 attendees)

[43] **A. Boutros**, "AI Inference on FPGAs: Design Approaches & Architecture Enhancements", *Open Source FPGA Foundation Online Webinar*, Sep 2021 (193 attendees)

[44] **A. Boutros**, "Neighbors From Hell: Voltage Attacks Against Deep Learning Accelerators", *Intel Labs*, Mar 2021

Academic Service

- Organizer of the International Centre for Spatial Computational Learning monthly seminars (May 2023 - Present)
- Reviewer for IEEE Transactions on Circuits and Systems (8 reviews)
- Reviewer for ACM Transactions on Reconfigurable Technology and Systems (4 reviews)
- Reviewer for ACM Transactions on Architecture and Code Optimization (4 reviews)
- Reviewer for ACM Computing Surveys (2 reviews)
- Reviewer for IEEE Access (1 review)
- Secondary reviewer for all four top-tier FPGA conferences (FCCM'21, FPT'21, FPT'22, FPGA'23, FCCM'23, FPL'23, FPT'23, FPGA'24)

Mentorship

List of undergraduate and graduate students I mentored under the supervision of Prof. Vaughn Betz at the University of Toronto during my M.A.Sc and Ph.D. research:

Name(s)	Degree	Topic
Zach Zheng	Research Intern (2018)	Implementing an FPGA Accelerator for LSTMs
Helen Dai	Research Intern (2020)	Creating VTR-Compatible FPGA Deep Learning Benchmarks
Jack Berezny	EngSci. Thesis (2022)	Adding HBM Support for the NPU Overlay

James Jiang Xun Deng Yue Fei Youngjo Kin	ECE Capstone (2022)	Enhancing the NPU Overlay for CNN Acceleration
Kai Qin Patrick Wang Taikun Zhang Atharva Datar	ECE Capstone (2022)	Enhancing the NPU Overlay for Transformer Acceleration (Certificate of Distinction as one of the top projects of this year)
George Trieu	Research Intern (2023)	RTL/SystemC co-simulation in RAD-Sim
Stephen More	M.A.Sc. (2023 - Present)	CAD Tools for RAD Physical Design
Taikun Zhang	M.A.Sc. (2023 - Present)	Accelerating Graph Neural Networks on FPGAs
Hang Yan	Ph.D. (2023 - Present)	AI Engines Architecture Exploration in Future Reconfigurable Acceleration Devices (RADs)
Yan Zhu	EngSci. Thesis (2023 - Present)	Characterization of NoC-Attached HBM in RADs
Abnash Bassi	EngSci. Thesis (2023 - Present)	Simulation and Modeling of Network-Attached RAD Clusters