Thierry Moreau

moreau@uw.edu - (206) 883 1723 - http://homes.cs.washington.edu/~moreau/ - github: tmoreau89

EDUCATION

University of Washington – Computer Science and Engineering *2012-present*

Master of Science (2015), Ph.D. candidate, Advisor: Luis Ceze

- Research Focus: FPGA Acceleration, Deep Learning Systems, Compilers, Approximate Computing

University of Toronto – Computer Engineering *2007-2012*

Bachelor of Applied Science (BASc) with Honors (GPA 3.89/4.0)

RESEARCH COMMUNITY INVOLVEMENT HIGHLIGHTS

- TVM open source deep learning compiler **code committer** [github.com/dmlc/tvm]
- VTA open source deep learning accelerator **project lead** [tvm.ai/vta]
- REQUEST workshop **co-founder**, co-located with ASPLOS 2019 [cknowledge.org/request.html]

WORK EXPERIENCE

Altera Corporation, OpenCL compiler group - *Software Engineer Intern* *Fall 2014*

- Optimized and improved maintainability and testability of OpenCL conformant IPs instantiated by the compiler. Designs that made use of floating point division had an average reduction of 7% in DSPs from re-architecting the floating-point divider *[LLVM, C++, OpenCL]*

Altera Corporation, Memory IP group - *Software Engineer Intern* *May 2010-August 2011*

- Pushed Memory Interface IP support in Altera Quartus II compiler for 28-nm and 45-nm FPGA devices *[C++]*
- Wrote simulation libraries used in IP core modeling and design verification *[Verilog, VHDL, VCS]*
- Took ownership over regression testing infrastructure to track compilation and timing closure results *[Perl, TCL]*

University of Toronto - *Undergraduate Researcher - advised by Professor Natalie Enright Jerger* *Summer 2009*

- Profiled and analyzed memory access patterns in multithreaded applications *[Pintool]*
- Wrote a configurable memory system simulator for 64-core architectures *[C++]*

Total S.A. France, IT group – *Software Intern* *Summer 2008*

- Designed an interactive email attachment server storage service for corporate use *[Visual Basic, SharePoint]*

CONFERENCE & WORKSHOP PAPERS

- "Learning to Optimize Tensor Programs", Tianqi Chen, Lianmin Zheng, Eddie Yan, Ziheng Jiang, **Thierry Moreau**, Luis Ceze, Carlos Guestrin, Arvind Krishnamurthy. At NIPS 2018.
- "Towards Reproducible and Reusable Deep Learning Systems Research Artifact Evaluation", **Thierry Moreau**, Anton Lokhmotov, Grigori Fursin. At MLOSS 2018 (co-located with NIPS).
- "TVM: An Automated End-to-End Optimizing Compiler for Deep Learning", Tianqi Chen, **Thierry Moreau**, Ziheng Jiang, Lianmin Zheng, Eddie Yan, Haichen Shen, Meghan Cowan, Leyuan Wang, Yuwei Hu, Luis Ceze, Carlos Guestrin, Arvind Krishnamurthy. At OSDI 2018.
- "MATIC: Learning Around Errors for Efficient Low-Voltage Neural Network Accelerators", Sung Kim, Patrick Howe, **Thierry Moreau**, Armin Alaghi, Luis Ceze, Visvesh Sathe. At DATE 2018 (application track best paper).
- "TVM: End-to-End Optimization Stack for Deep Learning", Tianqi Chen, **Thierry Moreau**, Ziheng Jiang, Haichen Shen, Eddie Yan, Leyuan Wang, Yuwei Hu, Luis Ceze, Carlos Guestrin, Arvind Krishnamurthy. At SysML 2018 (one of six contributed talks).
- "Exploring Quality-Energy Tradeoffs with Arbitrary Quantization", **Thierry Moreau**, Felipe Augusto, Patrick Howe, Armin Alaghi, Luis Ceze. At CODES+ISSS 2017 (special session).
- "Exploring Computation-Communication Tradeoffs in Camera Systems", Amrita Mazumdar, **Thierry Moreau**, Sung Kim, Meghan Cowan, Armin Alaghi, Luis Ceze, Mark Oskin, Visvesh Sathe. At IISWC 2017.

- "Approximating to the Last Bit", **Thierry Moreau**, Adrian Sampson, Luis Ceze, Mark Oskin. At WAX 2016 (co-located with ASPLOS).
- "REACT: A Framework for Rapid Exploration of Approximate Computing Techniques", Mark Wyse, Andre Baixo, **Thierry Moreau**, Bill Zorn, James Bornhsolt, Adrian Sampson, Luis Ceze and Mark Oskin. At WAX 2015 (co-located with PLDI).
- "SNNAP: Approximate Computing on Programmable SoCs Via Neural Acceleration", **Thierry Moreau**, Mark Wyse, Jacob Nelson, Adrian Sampson, Hadi Esmaeilzadeh, Luis Ceze and Mark Oskin. At HPCA 2015.

TECHNICAL REPORTS & OTHER

- "Automating Generation of Low Precision Deep Learning Operators", Meghan Cowan, **Thierry Moreau**, Tianqi Chen, Luis Ceze. ArXiv:1810.11066.
- "Exploiting Errors for Efficiency: A Survey from Circuits to Algorithms", Phillip Stanley-Marbell, Armin Alaghi, Michael Carbin, Eva Darulova, Lara Dolecek, Andreas Gerstlauer, Ghayoor Gillani, Djordje Jevdjic, **Thierry Moreau**, Mattia Cacciotti, Alexandros Daglis, Natalie Enright Jerger, Babak Falsafi, Sasa Misailovic, Adrian Sampson, Damien Zufferey. ArXiv:1809.05859.
- "VTA: An Open Hardware-Software Stack for Deep Learning", **Thierry Moreau**, Tianqi Chen, Ziheng Jiang, Luis Ceze, Carlos Guestrin, Arvind Krishnamurthy. ArXiv:1807.04188.
- "QAPPA: A Framework for Navigating Quality-Energy Tradeoffs with Arbitrary Quantization", **Thierry Moreau**, Felipe Augusto, Patrick Howe, Armin Alaghi, Luis Ceze. UW-CSE-17-03-02.
- "Compilation and Hardware Support for Approximate Acceleration", **Thierry Moreau**, Adrian Sampson, Andre Baixo, Mark Wyse, Ben Ransford, Jacob Nelson, Luis Ceze and Mark Oskin. At TECHCON 2015.
- "ACCEPT: A Programmer-Guided Compiler Framework for Practical Approximate Computing", Adrian Sampson, Andre Baixo, Benjamin Ransford, **Thierry Moreau**, Joshua Yip, Luis Ceze, Mark Oskin. UW-CSE-15-01-01.

JOURNAL ARTICLES

- "Energy-Efficient Neural Network Acceleration in the Presence of Bit-Level Memory Errors", Sung Kim, Patrick Howe, **Thierry Moreau**, Armin Alaghi, Luis Ceze, Visvesh Sathe. In IEEE Transactions on Circuits and Systems, Dec. 2018.
- "A Taxonomy of Approximate Computing Techniques", **Thierry Moreau**, Joshua San Miguel, Mark Wyse, James Bornholt, Armin Alaghi, Luis Ceze, Natalie Enright Jerger, Adrian Sampson. In IEEE Embedded Systems Letter, Oct. 2017.
- "Approximate Computing: Making Mobile Systems More Efficient", **Thierry Moreau**, Adrian Sampson, and Luis Ceze. In IEEE-Pervasive Computing, April 2015

POSTER PRESENTATIONS AND TALKS

- VTA: Open, Modular, Customizable Deep Learning Acceleration Hardware/Software Stack, talk and poster at CSE Industry Affiliates Day 2018
- Bringing Custom Hardware Acceleration to the TVM Stack, talk and poster at ASPLUW (UW funding retreat) 2018
- VTA: Open Hardware-Software Co-Design Stack for Deep Learning Systems Research, talk at ReQuEST 2018, co-located with ASPLOS 2018
- TVM: End-to-End IR Stack for Deep Learning Systems, talk and poster at ASPLUW (UW funding retreat) 2017
- Exploring Quality-Energy Tradeoffs with Arbitrary Quantization, talk at CODES+ISSS 2017
- An End-to-End Approximate Computing Demonstration, 2nd prize winner demo at C-FAR 2016
- Approximating to the Last Bit, talk at WAX 2016 co-located with ASPLOS 2016
- Compilation and Hardware Support for Approximate Acceleration, talk and poster session at TECHCON 2015
- SNNAP: Approximate computing on Programmable SoCs Via Neural Acceleration, talk at HPCA 2015
- Approximate Computing on SoCs via Neural Acceleration, poster presentation at C-FAR 2014
- Approximate Acceleration, talk and poster session at the Qualcomm Innovation Fellowship Winners Day 2014
- Approximate Acceleration, talk at the University of Washington SANE retreat 2014
- Approximate Acceleration, talk and poster at the Qualcomm Finalist Competition 2013

AWARDS AND DISTINCTIONS

- C-FAR Semi-Annual Workshop Demo – 2nd Prize Winner *2016*
- Qualcomm Innovation Fellowship 2013 ($100K in research funding for a team of two) *2013-2014*
- Weil Research Fellowship in Computer Science and Engineering *2012-2013*
- NSERC PostGraduate Scholarship *2012-2013*
- University of Toronto Scholarship *2009-2010*
- NSERC Undergraduate Student Research Award Summer *2009*

TEACHING/TUTORING/ADVISING

(University of Washington) **Co-Instructor for Hardware-Software Co-Design for Deep Learning (CSE599S)** *Spring 2018*
(University of Washington) **Head T.A. for Computer Architecture (CSE548)** *Spring 2017*
(University of Washington) **Head T.A. for Hardware Design and Implementation (CSE352)** *Spring 2013*
(University of Washington) **Undergraduate Researcher Advisor** *2014-2017*

- Mark Wyse (Winter-Summer 2014) – evaluating the effectiveness of high level design synthesis tools on a set of software kernels
- Sung Min-Kim (Spring-Summer 2015) – porting a neural network accelerator FPGA design to a low-power ASIC
- Yufang Sun (Spring-Summer 2015) – evaluating approximation and specialization techniques on vision algorithms
- Felipe Augusto (Summer 2016) – building piece-wise polynomial approximation libraries to approximate general purpose code
- Wyatt Muntean (Winter-Spring 2017) – building a stochastic neural network compiler for FPGAs

(University of Washington) **Mentoring Undergraduates for Hardware/Software Interface (CSE351)** *2014-2017*

LEADERSHIP AND EXTRACURRICULARS

(University of Washington) ***Shotokan Karate Club,*** *President* *2016-2018*
(University of Washington) ***Visit Days Party Coordinator****, Prospective Student Committee* *2014-2015*
(University of Toronto) **Case Competition Director**, *Sustainable Engineers Association* *2011-2012*

- Lead a team (planning, marketing, finance and logistics) to organize a first-of-its-kind case competition on sustainability, technology and entrepreneurship with 40 contestants and over 100 attendees. Organized a jury composed of entrepreneurs, CEOs, sustainability researchers, and finance experts.