Mark Horeni

Linkedin Profile

Email : markerdasharker@gmail.com
Mobile : +1-331-481-9068

Education

- **University of Notre Dame** — Notre Dame, IN
  Doctorate of Philosophy Computer Science and Engineering (Degree Not Obtained, ABD) — *Aug 2018 – Jul 2024*
- **University of Notre Dame** — Notre Dame, IN
  Master's of Computer Science and Engineering — *Aug 2018 – Dec 2022*
- **Lewis University** — Romeoville, IL
  Bachelor's of Computer Science and Engineering and Bachelor's of Applied Math — *Aug 2016 – May 2018*

Experience

- **Outlier** — Remote
  AI Trainer — *Jul 2024 – Present*
  - **Rewriting, RLHF, SFT, and Prompt Engineering**: Created and evaluated prompts that caused Large Language Models (LLM) to fail, corrected the mistakes, and wrote detailed justifications of why the models were wrong. Included topics such as math, coding, and general knowledge.
- **University of Notre Dame** — Notre Dame, IN
  Research Assistant — *Aug 2018 – Jul 2024*
  - **Optimized Mappings**: Updated open source software from NVIDIA, Timeloop, and improved neural network accelerator latency by 50% with no added circuitry overhead.
  - **Discovered New Architecture Designs**: Investigated and evaluated different architectural designs, such as a hybrid digital and analog design that achieved 1.5× benefit over baseline.
  - **Improved On-device Learning**: Designed transfer learning methodology that recovered 94% of accuracy on ReRAM-based compute-in-memory systems.
  - **Evaluated Compute-in-Memory**: Engineered infrastructure that informed circuit-level design decreased energy-delay product by 50% over current digital designs for large language models.
  - **Developed Software**: Learned and implemented algorithms using Python packages for deep learning, including TensorFlow or PyTorch, and architected faster software with C/C++ when needed.
- **EMD Electronics** — San Jose, CA
  Architecture Intern — *Jun 2022 – Sep 2022*
  - **Assessed Future Technologies**: Collaborated with team members to evaluate impact and candidate devices targeting 10× storage density for neural network weights.
  - **Pipelined Execution**: Developed methodologies to pipeline neural networks for energy reduction, which led to 1.5× energy reduction on convolutional neural networks.
  - **Consulted**: Independently researched baseline accelerator architectures to compete against and modify for future designs and documented these decisions to team leaders.
  - **AWS**: Debugged and beta-tested new AWS (Amazon Web Services) cluster for usability.
- **University of Notre Dame** — Notre Dame, IN
  Teaching Assistant — *Aug 2018 – May 2019*
  - **Tutored**: Conducted weekly office hours for student support in Discrete Math.
  - **Graded**: Graded tests following instructor-created rubric with team members of other teaching assistants.

Publications

- Martin Schiemer, Clemens JS Schaefer, Jayden Parker Vap, **Mark Horeni**, Yu Emma Wang, Juan Ye, Siddharth Joshi, "Hadamard Domain Training with Integers for Class Incremental Quantized Learning", CoLLA, 2024
- **Mark Horeni**, Siddharth Joshi, 'Improvements in Interlayer Pipelining of CNN Accelerators Using Genetic Algorithms", arXiv, 2023
- **Mark Horeni**, Pooria Taheri, Po-An Tsai, Angshuman Parashar, Joel Emer, Siddharth Joshi, 'Ruby: Improving Hardware Efficiency for Tensor Algebra Accelerators Through Imperfect Factorization", ISPASS, 2022
- Clemens JS Schaefer, **Mark Horeni**, Pooria Taheri, Siddharth Joshi, 'LSTMs for Keyword Spotting with ReRAM-based Compute-In-Memory Architectures", ISCAS, 2021
- Piotr Szczurek, **Mark Horeni**, 'Using Link Analysis Algorithms to Study the Role of Neurons in the Worm Connectome", AINA 2018

Other Projects

- **Analytical Solver for Timeloop** *Nov 2022 – Jul 2024*
  - **Project Description**: Wrote code that would take in the same inputs as Timeloop but would analytically solve the optimization target to output the best tiling given a neural network layer and accelerator specification. Able to reduce wallclock time (from hours or minutes to seconds) for optimizing performance of an arbitrary accelerator by continuously pruning loopnests that couldn't be optimal.
  - **Project Status**: Abandoned due to dropping out of PhD.
- **Quantization and Noise Aware Training** *Aug 2020 – Jul 2024*
  - **Project Description**: Coded and implemented PyTorch and Tensorflow custom layers to simulate hardware quantization and noise of eNVM devices such as ReRAM.
  - **Project Status**: Enveloped into other projects.
- **Neural Network Optimization of Wireless Side-channel Attacks** *Jun 2019 – Aug 2023*
  - **Project Description**: Attempted to generalize previous work on more realistic environments and limited training data. Able to identify the key within the Top-5 over 80% of the time.
  - **Collected Data**: Learn how to use an oscilloscope to capture power traces and an SDR to capture Bluetooth signal data programmatically.
  - **Project Status**: Presented as a Poster at Notre Dame's Annual Poster Contest and won "Best Pitch"
- **L2 Race Workshop at Telluride 2020** *Aug 2020 – Sept 2020*
  - **Project Description**: Develop an algorithm to drive a car around a track using limited data, and no underlying access to the dynamics model. Helped design with a team recurrent neural network architecture that trained using user gathered data to be deployed.
  - **Investigated Reinforcement Learning Methods**: Learned about different current RL methods, implementation, and other control theory fundamentals such as closed-loop vs open-loop experiments, environmental modeling, and agent behavior.
  - **Project Status**: Published on GitHub.
- **Biological Plausible Learning** *Oct 2018 – Dec 2019*
  - **Project Description**: Attempting to find a more biologically plausible and energy efficient method of training artificial neural networks than backpropagation. Developed *LDTP* (Linear Difference Target Prop) to simplify computation by approximating targets linearly instead of inversing achieving a faster convergence time.
  - **Researched Current Methods**: Researched and understood different biologically plausible and alternative learning methodologies such as Hebbian learning, Feedback Alignment, Equilibrium Propagation, and Difference Target Propagation.
  - **Project Status**: Did not seem fruitful to continue as current neural network architectures have other optimizations that can optimized to greater degree as well.

Programming Skills

- **Languages**: Python, C/C++, SQL
- **Technologies**: PyTorch, TensorFlow, Matlab, LaTeX, Pandas, Numpy, Matplotlib, Seaborn