Gwangho Lee

101, 26-6, Seongsan-ro 24-gil, Seodaemun-gu, Seoul, 03759, Republic of Korea
(+82) 10-2615-1505 | gwangho.btb@gmail.com | gwangho-lee.github.io

Research Interests

Computer Architecture, Hardware/Software Co-Design, Scalable Computer Systems, Multicore Computing, Domain-Specific Architecture, Datacenter Networking, Chiplet Design

Education

B.S. in Electrical and Computer Engineering & B.A. in Economics August 2023 (Expected)
Seoul National University (SNU)
· GPA : 4.16 / 4.30
· Courses : Computer Organization and Design, Multicore Computing, Data Communication Networks, Digital Systems Design and Experiments, Computer Organization, Digital Integrated Circuits, Computer Architectures for AI, Digital Logic Design, Data Structures, Compilers, Quantum Computing and Information, Machine Learning Fundamentals

Honors and Awards

KFAS Overseas PhD Scholarship, *KFAS (Korea Foundation for Advanced Studies)* Up to five years from 2023
Full tuition, insurance, and stipend of $20,000USD (Annually 4 students in Electrical Engineering are selected in Korea)

National Science & Technology Scholarship, *Korea Student Aid Foundation* Spring 2015 - Fall 2020
Full tuition (National Scholarship for excellent students in science and engineering)

Publications

Conference

Jongmin Kim*, **Gwangho Lee***, Sangpyo Kim, Gina Sohn, John Kim, Minsoo Rhu, Jung Ho Ahn, "ARK: Fully Homomorphic Encryption Accelerator with Runtime Data Generation and Inter-Operation Key Reuse", *55th IEEE/ACM International Symposium on Microarchitecture (**MICRO**)*, 2022 [Paper]

Journal

Gwangho Lee, Sunwoo Lee, Dongsuk Jeon, "Dynamic Block-Wise Local Learning Algorithm for Efficient Neural Network Training", *IEEE Transactions on Very Large Scale Integration Systems (**TVLSI**)*, 2021 [Paper]

*equally contributed

Research Experiences

Scalable Computer Architecture (SCALE) Lab, SNU January 2022 - August 2022
Advisor : Prof. Jung Ho Ahn Undergraduate Research Intern

· ARK : Fully Homomorphic Encryption (FHE) Accelerator
· Algorithm-architecture co-designed accelerator targeting bootstrappable parameters for practical FHE workloads.
· Analyzed the underutilization of functional units during bootstrapping in prior works due to severe memory bottleneck.
· Proposed inter-operation evaluation key reuse and on-the-fly plaintext generation, which enables higher arithmetic intensity of FHE operations by dramatically reducing the excessive off-chip memory access.
· Designed ARK microarchitecture, effectively utilizing the increased arithmetic intensity with tailored functional units, alternating data distributions, and dataflow-aware organization.
· Outperformed the state-of-the-art CPU implementations in execution time by 18,000× on ResNet-20 inference (0.125s), by 11,000× on sorting (1.990s), and on other existing FHE workloads.

· Rethinking Various Parameters to Further Improve the Performance of FHE
· Practical set of prime numbers and parameters for the CKKS scheme over the conjugate-invariant ring to minimize data puff-up of ciphertexts while ensuring a certain level of precision for machine learning workloads.
· Alternating data distributions and HRot hoisting that can be adapted to multi-GPU implementations.

High Performance Computer System (HPCS) Lab, SNU December 2020 - June 2021
Advisor : Prof. Jangwoo Kim Bachelor's Thesis Project

· Reproduced a Work on Extending gem5 GCN3 to Support Multi-GPU Systems
- · Extended gem5 GCN3 modeling recent AMD APUs with multiple GPU cores, which enables the simulation of data partitioning and workload scheduling in multi-GPU systems.
- · Replicated GPU components by assigning each GPU a unique id and doorbell region.
- · Modified the kernel using a hash table and offset of the doorbell region to maintain the mapping between software queues and the corresponding GPUs.
- · Refined the cache write policy of cache coherence protocols to effectively support multi-GPU systems.
- · Evaluated the execution time on several workloads with different numbers of GPUs.

Mobile Multimedia Systems (MMS) Lab, SNU January 2020 - November 2020
Advisor : Prof. Dongsuk Jeon Undergraduate Research Intern

· Hardware-Friendly CNN Training Algorithm for Low-Power ASIC Design
- · CNN training scheme for VGGNets based on the local learning algorithm that does not require backpropagation of error signals during training, enabling additional parallelism and reducing off-chip memory access.
- · Analyzed the limitations of prior works that relied on compute-heavy auxiliary networks and proposed an algorithm that performs local learning on a block-by-block basis while dynamically changing block boundaries during training.
- · Achieved up to 15% and 81% reduction in multiply-accumulate (MAC) operations and off-chip memory access while matching the test accuracy of the backpropagation algorithm.
- · 20% faster training than the baseline on the ASIC chip having limited on-chip memory designed for edge devices.

· Fixed-Point Quantization for Efficient Machine Learning
- · Implemented fixed-point quantization schemes on the PyTorch framework for various neural networks and evaluated their effectiveness in terms of performance and accuracy degradation.

Undergraduate Projects

Accelerating CNN Inference with multi-GPUs
Multicore Computing
· Improved CNN inference throughput using 16 GPUs in 4 nodes with OpenCL, MPI, and kernel optimizations.

Accelerating CNN Inference using FPGA
Digital Systems and Design
· Implemented an accelerator for CNN inference using FPGA that communicates through AXI-Stream protocol.

Accelerating FFT in GPGPU-Sim
Computer Organization and Design
· Minimizing the number of cycles required for FFT in GPGPU-Sim using CUDA.

Kernel Synchronization / P2P File-Sharing Protocol / TCP Packet-Level Simulator
Data Communication Networks
· Implemented kernel synchronization in Raspberry Pi.
· Torrent-esque P2P file-sharing system with functionality validation.
· Implemented TCP Packet-Level Simulator and observed the performance of a given network environment.

Binary Classifier for Machine Learning
Digital Integrated Circuits
· Minimized Energy-Delay Product by tuning supply voltage, clock frequency, logic style, and transistor sizing.

Compulsory Military Service

Republic of Korea Air Force Headquarters July 2017 - July 2019
Data System Management Group
· Database management, query optimization, web server back-end management, various vulnerability fix

Teaching Experiences

Digital Systems and Design September 2022 - Present
Peer Tutor
· Helped fellow students learn digital systems and design and complete the final project

Skills

Languages **C++, Python, Verilog, CUDA, OpenCL**, x86/64 assembly, Ocaml, Java, MySQL
Tools **Pytorch, Unix, Synopsys Design Compiler, gem5**, GPGPU-Sim, Vivado, LTSpice