





YANG YANG

 Website Elio-yang.github.io
 Official yangyang@virginia.edu
 GitHub Github.com/Elio-yang
 Address Rice Hall, 85 Engineer's Way, Charlottesville, VA, 22904

EDUCATION

 Jilin University, Changchun, China *Sept. 2019 – Jul. 2023*
 B.S. in Computer Science and Technology
GPA: 3.69/4.0
Rank: 9%
Thesis: *The Design and Implementation of Binary Code Analysis Framework for NVIDIA GPU.* [Score: 95/100]
Advisor: [Prof. Jingweijia Tan](#)

 University of Virginia, Charlottesville, USA *Aug. 2023 – Present*
 Ph.D. in Computer Science
Interests: GPU · Storage · Energy-Efficiency · Security
Advisor: [Prof. Adwait Jog](#)

PUBLICATION

Facilitating Profile Guided Compiler Optimization with Machine Learning.
Yang Yang, Xueying Wang, Guangli Li*
[SRC@CGO'23](#) [Poster]

- Achieving an average of $1.03\times$ and $1.95\times$ speedups on representative real-world applications and *Polybench* benchmark suite over the baseline (i.e., the programs without PGO), respectively.
- The performance of our machine learning-aided PGO is very close to the classic PGO ($1.05\times$ and $1.97\times$ speedups over the baseline) while reducing 58.3% and 94.8% optimization costs.

RESEARCH EXPERIENCE

[Insight Computer Architecture Lab](#) *Aug. 2023 – Present*
University of Virginia, Charlottesville, Virginia, USA
Advisor: [Prof. Adwait Jog](#)
Research on: GPU Memory and Storage & GPU Security
What We Do:



- Exploring the [memory and storage system](#) of GPUs by enabling direct GPU communication to [NVMe SSDs](#).
- Exploring how to enable [confidential computing](#) on GPU and proposed our flexible solutions.
- Implement and simulate the GPU memory safety system with the help of cryptography.
- Exploiting the opportunities to utilize [post-quantum cryptography](#) (e.g. FHE, LWE) in GPUs and how to make them faster.
- Exploiting the feasibility for using [CXL](#) memory and new GPU architecture (e.g. GH200).

[Emerging Technology Enabled Computer Architecture Lab](#) *Feb. 2022 – Jul. 2023*
Jilin University, Changchun, Jilin, P.R.China
Research Assistant, Advisor: [Prof. Jingweijia Tan](#)
Research on: GPU Architecture & Reliability & Energy Efficiency & Accelerator
What We Do:

- Explored the process variation of MCM-GPUs based on FinFET and state-of-the-art chiplet technology.
- Exploited the potential of FPGA for building open-sourced GPU like Vortex.
- Implemented a [Low-Level Analysis and Modeling](#) framework for NVIDIA Ampere GPU.
- Applied deep learning techniques for accurate power modeling.
- Examined the power-level effect of the instruction control flag when generating the SASS.

- Proposed a branch predictor using XGBoost based on static features.
- Explore the speedup sensibility of different programs towards different feature design.
- Utilize GNNs to build predictive profile-guided optimization framework and integrated it into LLVM.
- Released a new dataset for graph-related static analysis tasks.

SKILLS

Languages C/C++ · Assembly (x86, RISC-V) · Python · Go
 Frameworks CUDA · Pytorch · LLVM
 Software  LINUX ·  L^AT_EX · Markdown · GNU compiler (gcc, etc.) · GPGPU-Sim · Varius-TC · Z3 Solver

AWARDS


Undergraduate Academic Year Scholarship

- The First Class Fellowship Sept. 2020
- The Second Class Fellowship Sept. 2021
- The Third Class Fellowship Sept. 2022
- The Third Class Fellowship Jun. 2023

PROJECTS

MapReduce Engine is a Go language implementation of the paper¹. Apr. 2022

- Fault tolerance (failures like crash and communication-lose of workers) master and a worker cluster.
- Characterized cluster size and working functions (mapf & reducef).
- Communicate with the master through Remote Procedure Call.

This Engine is a basic component for building a large-scale distributed system.  Codes [here.](#)]


EOS is a 32bit *nix operating system developed in C language. Sept. 2021


- Basic bootloader, 2-level paging, 4GB memory management and kernel multithreads.
- Provide a set of traditional shell programs and multi-process mechaism.
- Follow the x86 ABI, so it's easy to port those x86 applications.

This project is still *active* and it will provide a library and compiler support in the future.  Codes [here.](#)]

WYZ-BAR is a bar management system developed in C language. Mar. 2020

- WYZ-BAR is a *collaborative project* (WYZ stands for 3 members and Y is for me) and I am the leader.
- Multi-process organization for effective system building.
- Re-implemented a simple sqlite style database.
- Used lots of parsing techniques for input checking.

WYZ-BAR is my *first* course project in the university.  Codes [here.](#)]

You can find more projects including course labs (like MIT 6.828), Android application (SmogDetector), CUDA operators (FFT) etc., from  [GitHub](#).

OTHER INFORMATION

Chinese · Native proficiency.

English · Professional proficiency.

¹ Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. *Communications of the ACM*. 2008 Jan 1;51(1):107-13.