YANG YANG

Homepage Elio-yang.github.io
Official xqg5sq@virginia.edu or yangyang@email.virginia.edu
Personal jluelioyang2001@gmail.com
GitHub Github.com/Elio-yang
Address School of Engineering and Applied Science, 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*
Advisor: Prof. Jingweijia Tan

University of Virginia, Charlottesville, USA *Aug. 2023 – Present*
Ph.D. in Computer Science **Interests**: GPU · FPGA for Accelerator Design · Reliability · Compiler
Advisor: Prof. Adwait Jog

PUBLICATION

Yang Yang, Xueying Wang, Guangli Li*. Facilitating Profile Guided Compiler Optimization with Machine Learning. *In Student Research Competition of the 21st IEEE/ACM International Symposium on Code Generation and Optimization*.[Poster]

- Achieving an average of 1.03× and 1.95× 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× and 1.97× speedups over the baseline) while reducing 58.3% and 94.8% optimization costs.

RESEARCH EXPERIENCE

Emerging Technology Enabled Computer Architecture Lab *Feb. 2022 – Present*
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**.

Project: **LLAM: A Low-Level Power Modeling and Prediction Framework for NVIDIA Ampere GPU**

- Implemented a **L**ow-**L**evel **A**nalysis and **M**odeling 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.

State Key Laboratory of Processor *Jul. 2022 – Present*
Institute of Computing Technology, Chinese Academy of Science, Beijing, P.R.China
Research Assistant, Advisor: Prof. Guangli Li
Research on: Compiler & Programming Systems & Deep Learning
What We Do:

- Improved the **optimization** ability of compilers based on application's **run-time** characteristics.
- Using **machine learning** methods to guide the **LLVM** compiler for better machine code **generation**.

Project: Facilitating Profile Guided Compiler Optimization with Machine Learning

- Proposed a **branch predictor** using **XGBoost** based on **static** features.
- Explore the speedup sensibility of different programs towards different feature design.
- Using GNNs to predict branch behavior in programs.
- Released a new dataset for graph-related static program analysis tasks.

SKILLS

Languages C/C++ · Assembly (x86, RISC-V) · Python · Go
Frameworks CUDA · Pytorch · LLVM
Hardware Verilog · Vivado · FPGA
Software LINUX · LaTeX · Markdown · GNU compiler (gcc, *etc.*) · gpgpu-sim · Varius-TC

AWARDS

Undergraduate Academic Year Scholarship

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

PROJECTS

MapReduce Engine is a **Go** language implementation of the paper[1]. *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.
- Follow the **x86 ABI**, so it's easy to port those x86 applications.
- 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.*, in 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.