

Dian-Lun Lin

<https://dian-lun-lin.github.io> | <https://www.linkedin.com/in/dian-lun-lin-0ba721117/> | dianlun.lin@wisc.edu | (+1) 657-657-3479

EDUCATION

University of Wisconsin-Madison

US; July. 2024

PhD in Computer Engineering

- Thesis: Task-parallel Heterogeneous Programming System for Logic Simulation, Advisor: Tsung-Wei Huang

National Taiwan University

Taiwan; 2017 - 2019

MS in Electrical Engineering

- Thesis: On the Analysis of Network Creation Game with Imperfect Monitoring, Advisor: Ho-Lin Chen
- Teaching: Algorithms (2017 Fall & 2019 Spring)

National Cheng Kung University

Taiwan; 2013 - 2017

BS in Electrical Engineering

RESEARCH INTERESTS/SKILLS

Explore high-performance computing techniques to accelerate machine learning, electronic design automation, and other scientific computing applications using modern C++ and CUDA.

WORK EXPERIENCE

NVIDIA

Remote, US; May. 2022 – Aug. 2022

Research intern at Design Automation Research

- Topic: GenFuzz: A GPU-accelerated Hardware Fuzzer using Genetic Algorithm
- Built GenFuzz, a GPU-accelerated hardware fuzzer. We showed that GenFuzz running on a single A6000 GPU achieves 80× runtime speed-up when compared to state-of-the-art hardware fuzzers.

NVIDIA

Remote, US; May. 2021 – Nov. 2021

Research intern at Design Automation Research

- Topic: RTLflow: A GPU Acceleration Flow for RTL Simulation
- Built RTLflow, a GPU-accelerated RTL simulator. We showed that RTLflow running on a single A6000 GPU and eight CPU cores can achieve a 40× runtime speed-up when compared to an 80-thread multi-core CPU baseline.

OPEN-SOURCE PROJECTS

- Taro: Task-based Asynchronous Programming System using C++ Coroutine
 - <https://github.com/dian-lun-lin/taro>
 - I built Taro which allows users to write coroutines in a task graph while abstracting away complex coroutine management. I also developed a coroutine-aware work-stealing algorithm to support the programming model while minimizing CPU migration overhead. I presented Taro in CppCon 2023.
- Taskflow: A General-purpose Parallel and Heterogeneous Task Programming System
 - <https://github.com/taskflow/taskflow> (Over 9K stars)
 - I am one of the core developers. I developed cudaFlow, a significant feature within Taskflow. cudaFlow enables users to harness heterogeneous parallelism by constructing a CPU-GPU task graph. Currently, my focus lies on enhancing Taskflow's capabilities by integrating C++ Coroutines to facilitate multitasking. Additionally, I am actively engaged in optimizing Taskflow's runtime performance by leveraging modern C++20 concurrency.
- RTLflow: From RTL to CUDA - A GPU Acceleration Flow for RTL simulation with Multiple Inputs
 - https://github.com/dian-lun-lin/verilator_rtlflow
 - This is a one-year project cooperated with NVIDIA Design Automation Research. During the collaboration, I organized meetings with NVIDIA and effectively communicated the progress of the work. RTLflow demonstrated over 40 times speedup compared to the existing solutions, making it the most efficient open-source RTL simulator available. Through this project, I received second place at the PACT Student Research Competition 2022.
- SNIG: Accelerated Large Sparse Neural Network Inference using Task Graph Parallelism
 - <https://github.com/dian-lun-lin/SNIG>
 - I developed a powerful inference engine tailored for large sparse neural networks. The implementation outperformed the state-of-the-art baseline by up to 2.3 times on a machine equipped with 4 GPUs. Through this project, I received the champion of the 2020 IEEE HPEC Neural Network Challenge.

SELECTED AWARDS

- ACM/IEEE DAC Young Student Fellowship, 2020 & 2021 & 2023 & 2024
- Second place in ACM/PACT Student Research Competition (SRC), 2022
- ACM ISPD Wafer-Scale Physics Modeling Contest – Honorable Mention, 2021
- Champion of the IEEE/MIT/Amazon HPEC Large Sparse Neural Network Challenge, 2020
- Best Master Thesis Nomination, Department of Electrical Engineering, National Taiwan University, 2019
- Presidential Award, Department of Electrical Engineering, National Cheng Kung University, Fall 2015

SELECTED PAPERS/JOURNALS

- TaroRTL: Accelerating RTL Simulation using Coroutine-based Heterogeneous Task Graph Scheduling
 - **Dian-Lun Lin**, Tsung-Wei Huang, Joshua San Miguel, and Umit Ogras *Euro-Par*, 2024
- G-PASTA: GPU Accelerated Partitioning Algorithm for Static Timing Analysis
 - **Dian-Lun Lin** (co-first), Boyang Zhang, Che Chang, Cheng-Hsiang Chiu, Bojue Wang, Wan Luan Lee, Chih-Chun Chang, Donghao Fang, and Tsung-Wei Huang *DAC*, 2024
- GenFuzz: GPU-accelerated Hardware Fuzzing using Genetic Algorithm with Multiple Inputs
 - **Dian-Lun Lin**, Yanqing Zhang, Haoxing Ren, Shih-Hsin Wang, Brucek Khailany, and Tsung-Wei Huang *DAC*, 2023
- From RTL to CUDA: A GPU Acceleration Flow for RTL Simulation with Multiple Testbenches
 - **Dian-Lun Lin**, Haoxing Ren, Yanqing Zhang, Brucek Khailany and Tsung-Wei Huang *ICPP*, 2022
- Accelerating Large Sparse Neural Network Inference using GPU Task Graph Parallelism
 - **Dian-Lun Lin** and Tsung-Wei Huang *TPDS*, 2022
- Taskflow: A Lightweight Parallel and Heterogeneous Task Graph Computing System
 - Tsung-Wei Huang, **Dian-Lun Lin**, Chun-Xun Lin, and Yibo Lin *TPDS*, 2022
- Enabling Efficient GPU Computation using Task Graph Parallelism
 - **Dian-Lun Lin** and Tsung-Wei Huang *Euro-Par*, 2021
- A Novel Inference Algorithm for Large Sparse Neural Network using Task Graph Parallelism
 - **Dian-Lun Lin** and Tsung-Wei Huang **(Champion award)** *HPEC*, 2020

TALKS

- NERSC - GPUs for Science Day
 - Title: A Task Graph-based Programming System for CPU-GPU Heterogeneous Computing *California, US*; 2023
- The C++ Conference
 - Title: Taro: Task Graph-based Asynchronous Programming Using C++ Coroutines *Colorado, US*; 2023
- Berkeley National Lab
 - Title: An Introduction to C++ Coroutines Through a Thread Scheduling Demonstration *Remote, US*; 2023
- The C++ Now Conference
 - Title: An Introduction to C++ Coroutines Through a Thread Scheduling Demonstration *Colorado, US*; 2023
- MediaTek Research
 - Title: Accelerating Hardware Design Verification: Exploring Simultaneous Execution with RTLflow and GenFuzz *Remote, US*; 2023
- NVIDIA Research
 - Title: G-Fuzz: GPU-accelerated hardware fuzzing *Remote, US*; 2022
- The C++ Conference
 - Title: cudaFlow: A Modern C++ Programming Model for GPU Task Graph Parallelism *Colorado, US*; 2021
- NVIDIA Research
 - Title: RTLflow: A GPU acceleration flow for parallel RTL simulation *Remote, US*; 2021

ACADEMIC SERVICE

- Invited reviewer of *International Journal of Machine Learning and Cybernetics*, 2024
- Invited reviewer of *Concurrency and Computation: Practice and Experience*, 2024
- Program Committee in CppNow, 2023 & 2024
- Invited reviewer of *IEEE Access Journal*, 2023
- Invited reviewer of *The Journal of Supercomputing*, 2023
- Program Committee in CppCon, 2022 & 2023

INVITED POSTS

- A Concise Introduction to Coroutines
 - <https://www.modernescpp.com/index.php/a-concise-introduction-to-coroutines-by-dian-lun-li/>
- Coroutines: A Scheduler for Tasks
 - <https://www.modernescpp.com/index.php/coroutines-a-scheduler-for-tasks-by-dian-lun-li/>

STUDENTS MENTORED

- Wan Luan Lee, second-year PhD student at the University of Wisconsin-Madison, 2023
 - G-kway: Multilevel GPU-Accelerated k-way Graph Partitioner (Accepted by DAC 2024)
- Boyang Zhang, first-year PhD student at the University of Wisconsin-Madison, 2023
 - G-PASTA: GPU Accelerated Partitioning Algorithm for Static Timing Analysis (Accepted by DAC 2024)
- Mckay Mower, master's student at the University of Utah, 2022
- Elmir Dzaka, bachelor's student at the University of Utah, 2022
 - Parallel And-Inverter Graph Simulation Using a Task-graph Computing System (Accepted by IPDPSW 2023)