# **Run Wang**

- ♦ Phone: +41765849967 ♦ Url: https://run-wang.netlify.app ♦
- ◊ Email: samanthawangdl@gmail.com ◊ Nationality: Chinese ◊
- ♦ Address: 8406, Zurich, Switzerland ♦

# **EDUCATION**

ETH Zurich, Switzerland Master of Science in Electrical Engineering Fudan University, China Bachelor of Engineering in Electrical Engineering (Honours)

# **PUBLICATIONS**

Run Wang, Gamze Islamoglu, Andrea Belano, Viviane Potocnik, Francesco Conti, Angelo Garofalo, Luca Benini. VEXP: A Low-Cost RISC-V ISA Extension for Accelerated Softmax Computation in Transformers, IEEE International Symposium on Computer Arithmetic (ARITH), 2025.

Xiaotian Zhou, Run Wang, Wei Li, Zhongzhi Zhang. Maximizing the Smallest Eigenvalue of Grounded Laplacian Matrix, Journal of Global Optimization, 2025/2/26, pp. 1-22.

Run Wang, Ke Xu, Hui Feng, and Wei Chen. Hybrid RNN-ANN Based Deep Physiological Network for Pain Recognition, Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2020.

# **RESEARCH EXPERIENCE**

### 0. ETH Integrated Systems Laboratory (IIS), Zurich, Switzerland

Advisor: Luca Benini, Professor at the Department of ITET of ETH Zurich Topic: On-Device Learning on Heterogeneous SoCs with Software-Managed Caches

Description: Designed and implemented an end-to-end flow for fine-tuning vision transformers on ultra-low-power SoCs with software-managed memory hierarchies.

- ► Designed a lightweight training engine for on-device learning on heterogeneous PULP SoCs with software-managed caches.
- ► Achieved end-to-end deployment of Vision Transformer (ViT) models for both inference and fine-tuning directly on PULP clusters.
- ► Integrated RedMulE GEMM accelerator into the GVSoC model and Deeploy stack to enable fast matrix kernel training.
- ► Enabled both full fine-tuning and LoRA-based parameter-efficient fine-tuning of ViT models on ultra-low-power SoCs.

Keywords: On-Device Learning, PULP, FP Kernels, ONNX, Software-Managed Caches, Training Acceleration, RedMulE, Deeploy, Heterogeneous SoC

1. ETH Integrated Systems Laboratory (IIS), Zurich, Switzerland

Advisor: Luca Benini, Professor at the Department of ITET of ETH Zurich

**Topic: Transformer Hardware Software Codesign for NeuroSoC** 

Description: Implemented hardware and software optimizations to enhance the transformer algorithm performance on NeuroSoC, ensuring high efficiency and rapid processing capabilities.

► Implemented hardware and software optimizations to enhance the transformer algorithm performance on NeuroSoC, ensuring high efficiency and rapid processing capabilities.



Sep. 2022 - Present

Sep. 2018 - Jun. 2022

Feb. 2024 - Dec.2024

Jan. 2025 - Present

- ► Developed and integrated fast, precise floating-point exponential algorithm based on the improved Schraudolph Expf algorithm, integrating the new exponential algorithm into FPU and implementing RISC-V exponential instruction extensions in a Snitch-based SoC architecture.
- ► Accelerated Softmax, Attention and LLM kernel computations using new exponential instructions, SIMD parallelism and stream semantic registers Extensions (SSR), significantly boosting the processing speed and throughput of transformer models.
- ► Utilized GF12 technology and Fusion Compiler for comprehensive synthesis and layout.

Keywords: RTL, Architectural Design, Kernel Programming, RISCV, Transformer, C, Assembly, Sythesis, Backend

## 2. KUL EAST-MICAS, Leuven, Belgium

Advisor: Georges Gielen, Department Chair, Department of Electrical Engineering, KU Leuven (KUL) **Topic: Neuromorphic Electronic Skin System for Prosthetic Sensory Integration** 

Description: Designed an electronic system for an electronic skin (e-skin), incorporating custom tactile sensors and a neuromorphic SNN processor.

- ► Designed an electronic skin (e-skin) glove, including PCB design and RTOS-based embedded development for real-time data processing from five fingers, utilizing custom analog tactile sensors and a custom SNN processor.
- ► Developed and applied advanced quantization and model compression techniques for the SNN processor, training a compact SNN network for material classification with int8 QAT.
- ► Demonstrated real-time touch sensing and accurate material classification, highlighting the system's advanced functionality and precision.

Keywords: ML Edge Deployment, PCB Design, Embedded System, RTOS, FPGA

### 3. Intelligent Complex Systems Lab, Fudan University

Advisor: Zhongzhi Zhang, Professor at the Computer Science Department, Fudan University Topic: Maximizing the Smallest Eigenvalue of Grounded Laplacian Matrix by node removel

▶ Proved NP-hardness and non-submodularity of a combinatorial optimization problem, proposed a nearly-linear time heuristic algorithm with proven approximation, and evaluated its effectiveness on large-scale real-world graphs.

### 4. Center for Intelligent Medical Electronics, Fudan University

Advisor: Wei Chen, Professor, Information Science and Engineering, Fudan University Topic: Hybrid RNN-ANN Based Deep Physiological Network for Pain Recognition

- ► Initiated a pain research project addressing real-world clinical challenges in hospice care.
- ► Developed a hybrid RNN-ANN model for pain level classification using the Biovid Heat and EmoPain databases.
- ► Achieved state-of-the-art performance in accuracy and clinical applicability.

# **OPEN SOURCE EXPERIENCE**

### **Buddy Compiler**

Topic: Development of a Performance Test Suite for LLVM MLIR Compiler on RISC-V Vector Platform for LLM Keywords: MLIR, LLVM, C++, LLM, RVV

Apache@Alibaba RocketMQ Open Source Project **Role**: Documentation Manager Keywords: Java, HTML, CSS, Javascript

Jun. 2024 - Oct. 2024

Jun. 2022 - Sep. 2022

Jun. 2023 - Sep. 2023

Sep. 2020 - May.2021

Jun. 2019 - Feb. 2020