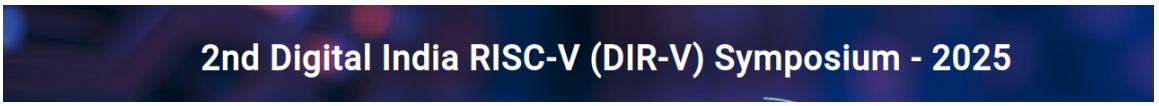

DIR-V Symposium Hackathon

Track: Core/SoC Design Enhancements

Objective:

To enable participants to design and explore hardware accelerators for machine learning. To promote innovation and enhance performance in ML applications through hardware-software co-design.

How: Hands-on Online

Participants will

- Learn to design and implement a systolic-array-based matrix multiplication engine in TL-Verilog, supporting configurable dataflows (output/weight/input stationary).
- Understand and apply hardware-software co-design principles by integrating a RISC-V CPU with a custom ML accelerator on an FPGA platform.
- Compare workload execution cycles on CPU-alone versus CPU+FPGA systems, identifying bottlenecks and exploring optimization strategies.
- Develop and deploy a simple C/C++ CNN application on a heterogeneous system to evaluate hardware acceleration benefits for machine learning.

Who can Participate:

Students, researchers, developers, and open-source hardware enthusiasts passionate about machine learning hardware accelerators.

Problem Statement:

The growing compute and memory demands of machine learning applications underscore the need for heterogeneous systems integrating CPUs and accelerators. Design a systolic-array-based matrix multiplication engine in TL-Verilog, supporting popular dataflows (output/weight/input stationary) and visualize with Makerchip's VIZ feature. Develop a simple C/C++ application for a single CNN layer computation, and implement the ML accelerator on the VSDSquadronFPGA. Use a simulated RISC-V CPU and hybrid simulation with Renode to run the application on a CPU+FPGA (ML Accelerator) system. Compare the simulated cycles for workload execution on the CPU alone versus the CPU+FPGA system, identify compute or memory bottlenecks, and propose strategies for performance improvement.

More details about the problem statement:

Matrix multiplications, involving numerous multiplications and accumulations, are the core of machine learning workloads. Efficiently accelerating these operations is key to optimizing ML performance with custom hardware. This project focuses on designing a systolic-array-based hardware accelerator for matrix multiplication.

The accelerator generator will support configurable systolic array shapes and multiple dataflows—output-stationary, weight-stationary, and input-stationary—and will be implemented using TL-Verilog for enhanced design flexibility. Dataflow visualization will be developed using the VIZ framework.

To demonstrate real-world applicability, a simple C/C++ application simulating a CNN layer or simple neural network will serve as the workload. The design will be deployed on the VSDSquadronFPGA platform and evaluated in a hybrid simulation environment with a simulated RISC-V CPU and FPGA accelerator using Renode. This setup enables execution of the workload on both the CPU alone and the CPU+FPGA system, allowing a detailed performance comparison.

The results will identify bottlenecks—whether compute- or memory-bound—and guide optimizations such as improved memory access, enhanced data reuse, or scaling the systolic array. This project advances understanding of hardware-software co-design and paves the way for accelerating complex ML workloads with RISC-V CPU based heterogeneous systems.