FAST, National University of Computing and Emerging Sciences, Islamabad, Pakistan

ICD Integrated Circuit Design Lab

# Energy Efficient Accelerator Design for AI on Edge Applications

## Team Members


- Ahmad Shaban
- Kousar Gul
- Ume Kalsoom
- Awab Younas
- Muhammad Mirza
- Dr. Khurram Javed
- Eeman
- Osama Mirza
- Dr. Hassan Saif
- Fajar Waseem
- Salman Qazi
- Dr. Rashid Ramzan
- Humail Nawaz
- Sana Nazir

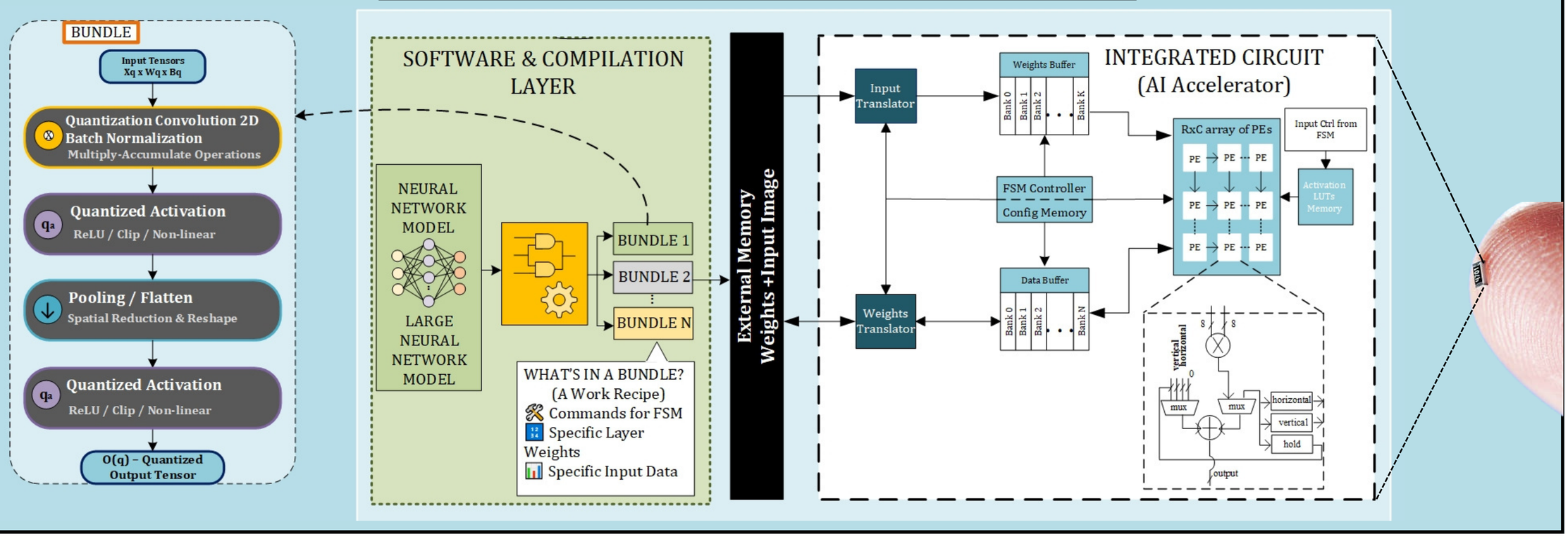

## Project Objectives

- Efficient Reconfigurable CNN Accelerator
- Optimizing PE using a Modified Booth multiplier and compression reduction tree for parallel MAC operations
- A parameterizable R×C PE array supporting sub-8-bit quantized convolution execution utilizing maximum data reuse strategy
- Reconfigurable CNN Accelerator deployment on FPGA/ASIC

## Abstract

This work presents an energy-efficient CNN accelerator that optimizes the Processing Element using a Modified Booth multiplier and reduction tree to reduce delay, power, and area. An FSM controller manages a reconfigurable R×CR times CR×C PE array to perform parallel MAC operations, while a LUT applies quantized activation functions to the outputs.

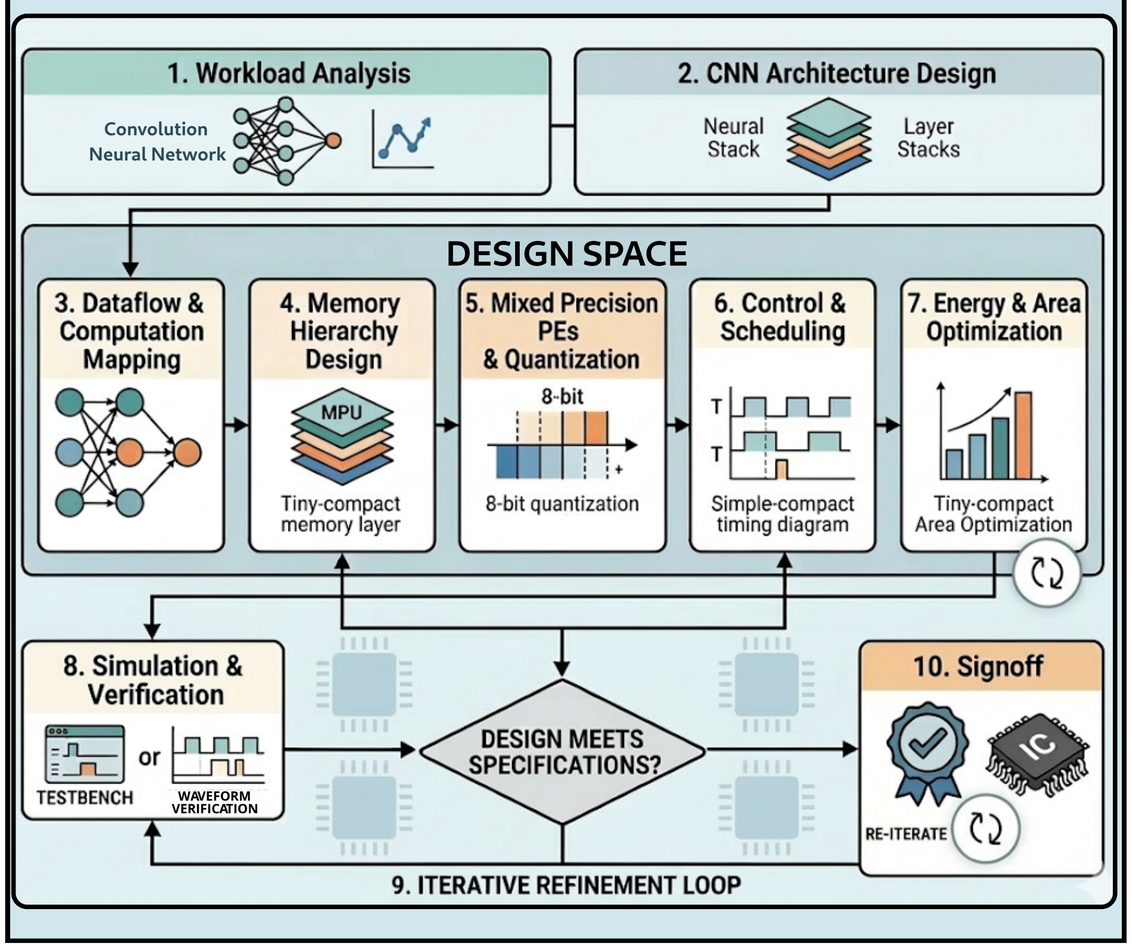
## System-Level Execution Flow




## Results

| Multiplier Architecture | Power (µW) | Area (µm²) |
|---|---|---|
| WTM (AND + CSA + CPA) | 50.2 | 74317.56 |
| MBE-R2 Multiplier | 22.2 | 67719.6 |
| MBE-R4 + GPC | 47.93 | 61121.64 |
| MBE-R4 Multiplier | 32.8 | 54523.68 |
| Approximate Multiplier | 47.9 | 47925.72 |
| MBE-R8 + Compressor + CLA | 419 | 41327.76 |
| MBE-R8 + GPC | 233.25 | 34729.8 |
| MBE-R8 Multiplier | 196.7 | 28131.84 |
| MBE-R4 (PP Generation) | 11.4 | 21533.88 |
| MBE-R8 (PP Generation) | 9.47 | 14935.92 |
| MBE-R4 + Compressor + CPA | 264 | 8337.96 |
| WTM (Basic Adders: AND + CSA) | 65 | 1740 |

## Methodology




## Future Work

- Integrate the optimised PE into a full R×C CGRA fabric and evaluate end-to-end CNN inference (ResNet-50, VGG) on FPGA with measured throughput and energy.
- Target ASIC tapeout on TSMC 65 nm and benchmark power, performance and area against commercial accelerators.