



3D POINT CLOUDS ON EMBEDDED PLATFORMS



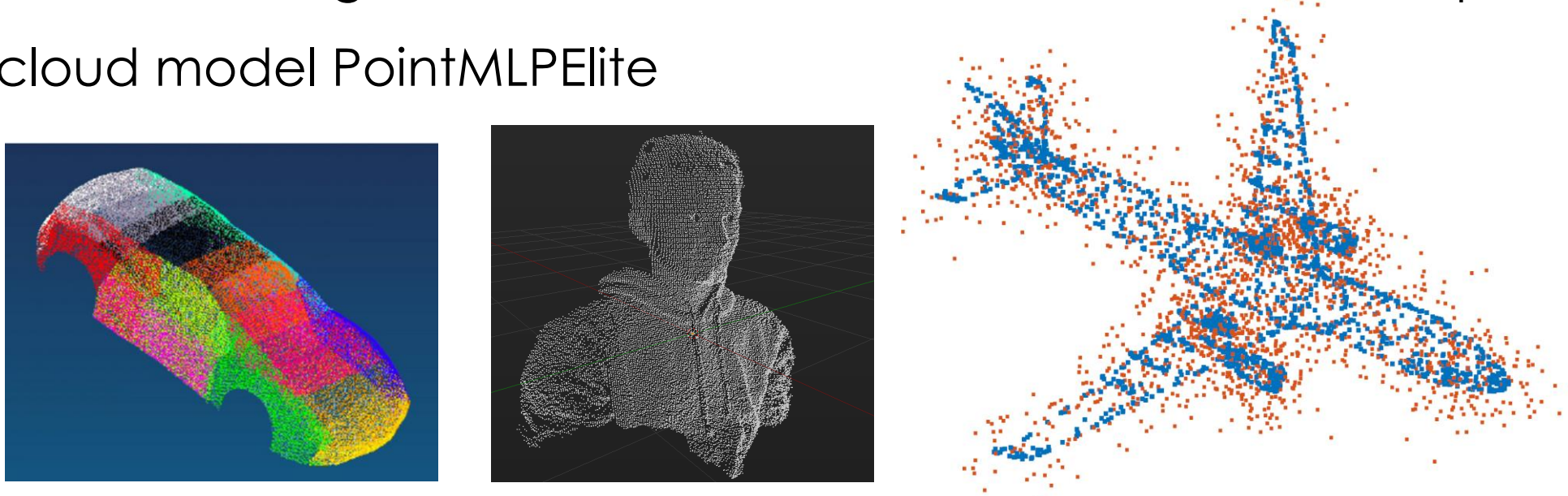
Amur Pal — Sannan Abbasi — Junaid Ali

Supervisors: Dr. Faisal Shafait, Dr. Adnan Ul Hasan

Introduction

AI accelerators via FPGA are fueling deep learning on 3D point clouds, revolutionizing LIDAR-powered autonomous vehicles and emerging AR/VR fields. This project focuses on:

- Building a framework for implementing 3D point cloud DL models on FPGAs
- Building a high-level synthesis library with the supported algorithms
- Demonstrating novel hardware results with SOTA 3D point cloud model PointMLPElite



Methodology

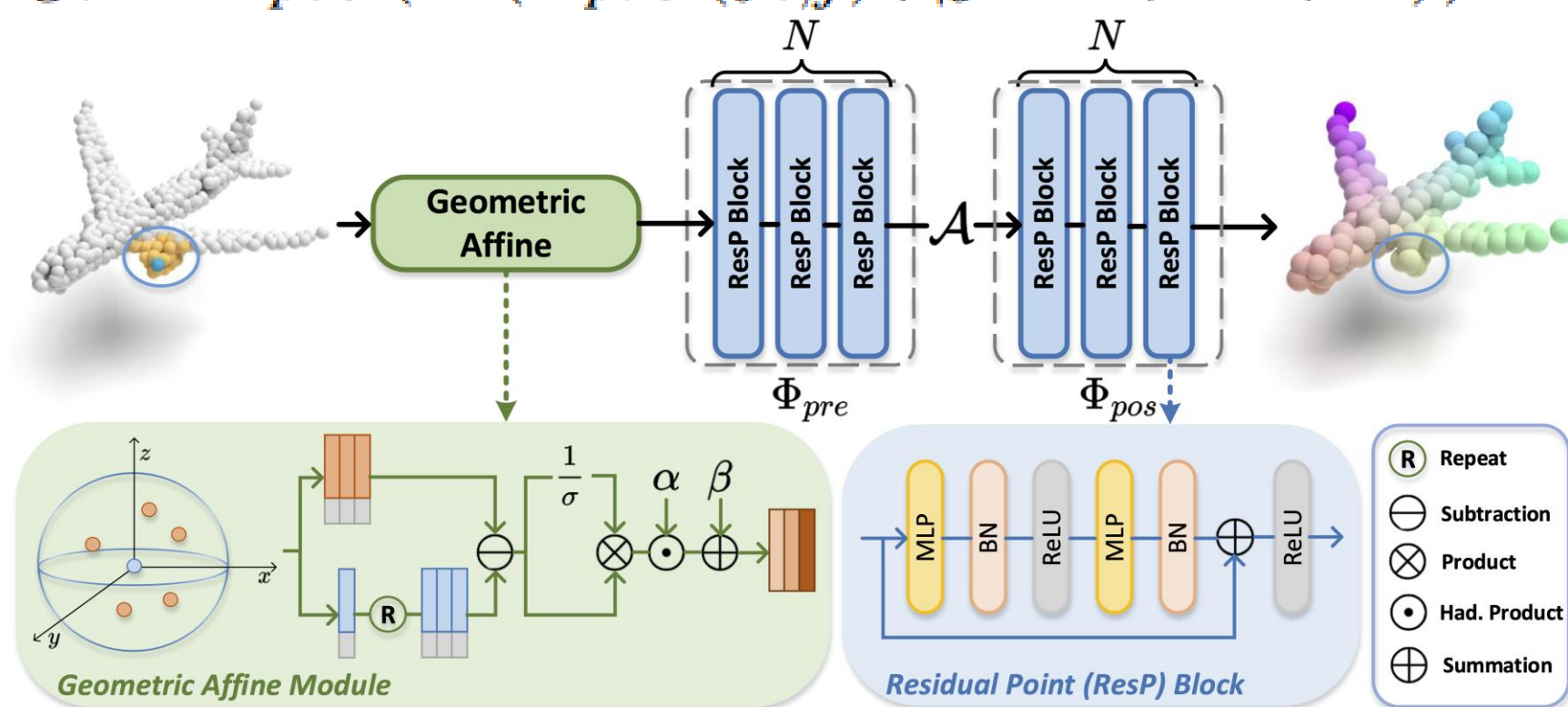
PointMLPElite

PointMLP is the state of the art 3D point cloud deep learning model. Some of its features are:

- Lightweight geometric affine module
- Novel normalization layer with learnable parameters that improve performance drastically
- Simple MLP structure with reconfigurable depth

PointMLP has the following kernel function:

$$g_i = \Phi_{pos}(\mathcal{A}(\Phi_{pre}(f_{i,j}), |j = 1, \dots, K))$$



Software Optimization

Next, we compress & optimize the model at software level by:

- Quantizing its weights, biases, activations, & geometric parameters
- Fusing CONV-BN layers, reducing number of input points

Hardware Implementation

We use high-level synthesis to create the RTL for our model. For this purpose, the optimized algorithms are written in HLS C++ using Xilinx Vivado HLS 2018.3. This implementation includes:

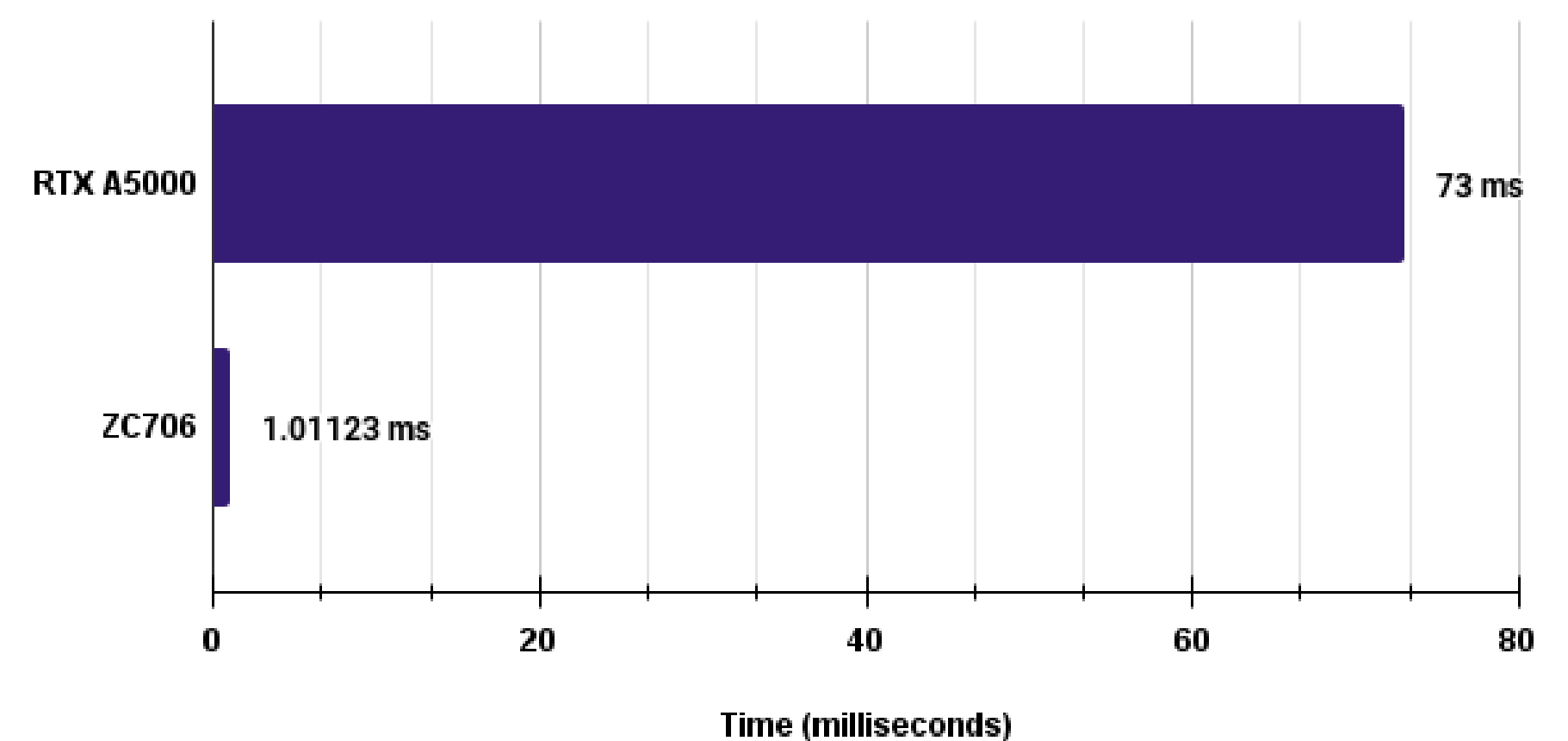
- Linear Feedback Shift Registers (LFSRs)
- Streaming k-Nearest Neighbors, Normalization

Results & Discussion

Xilinx ZC706 is used to accelerate **PointMLPElite** on **ModelNet40** dataset. Significant performance gain is observed:

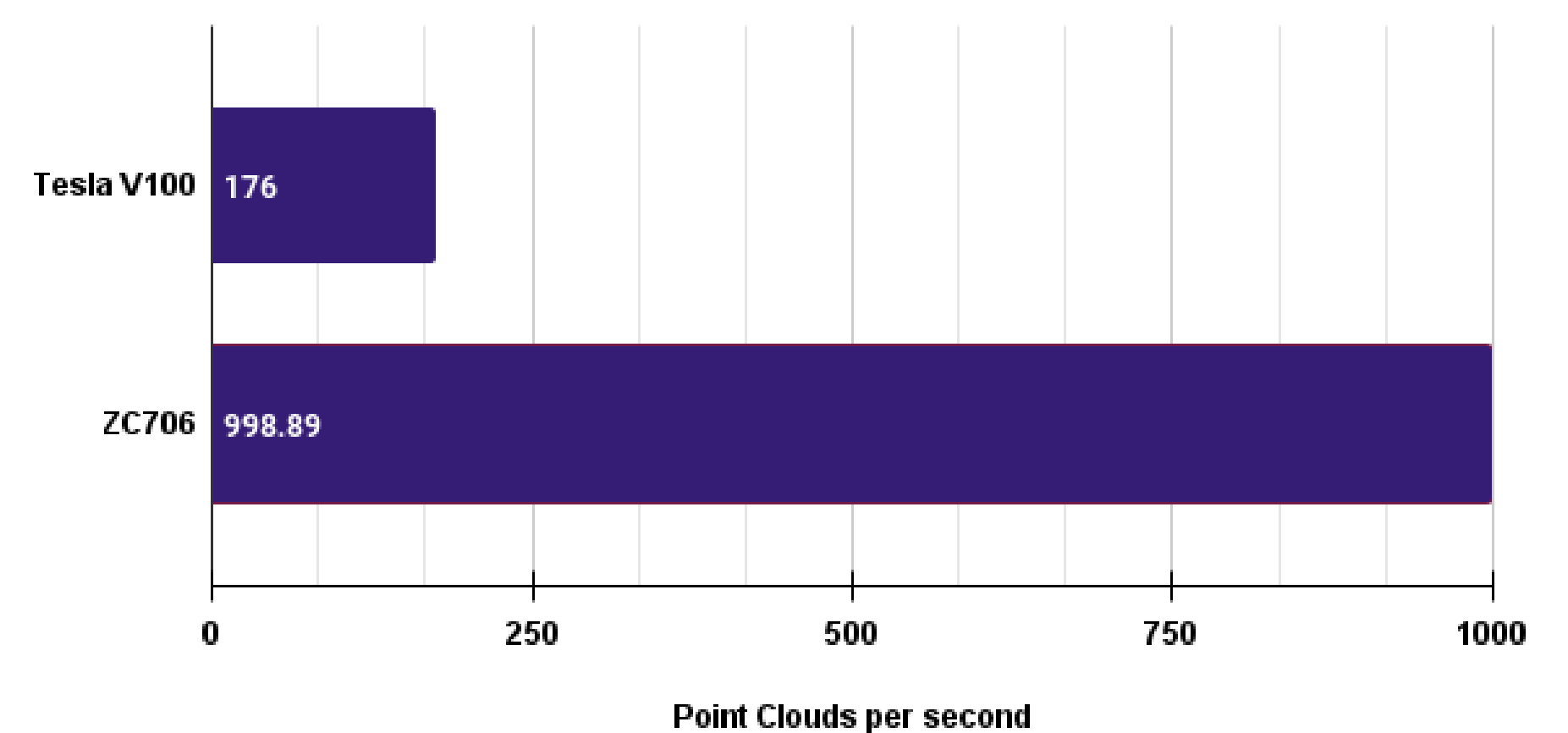
Latency Comparison

RTX A5000 vs Xilinx ZC706

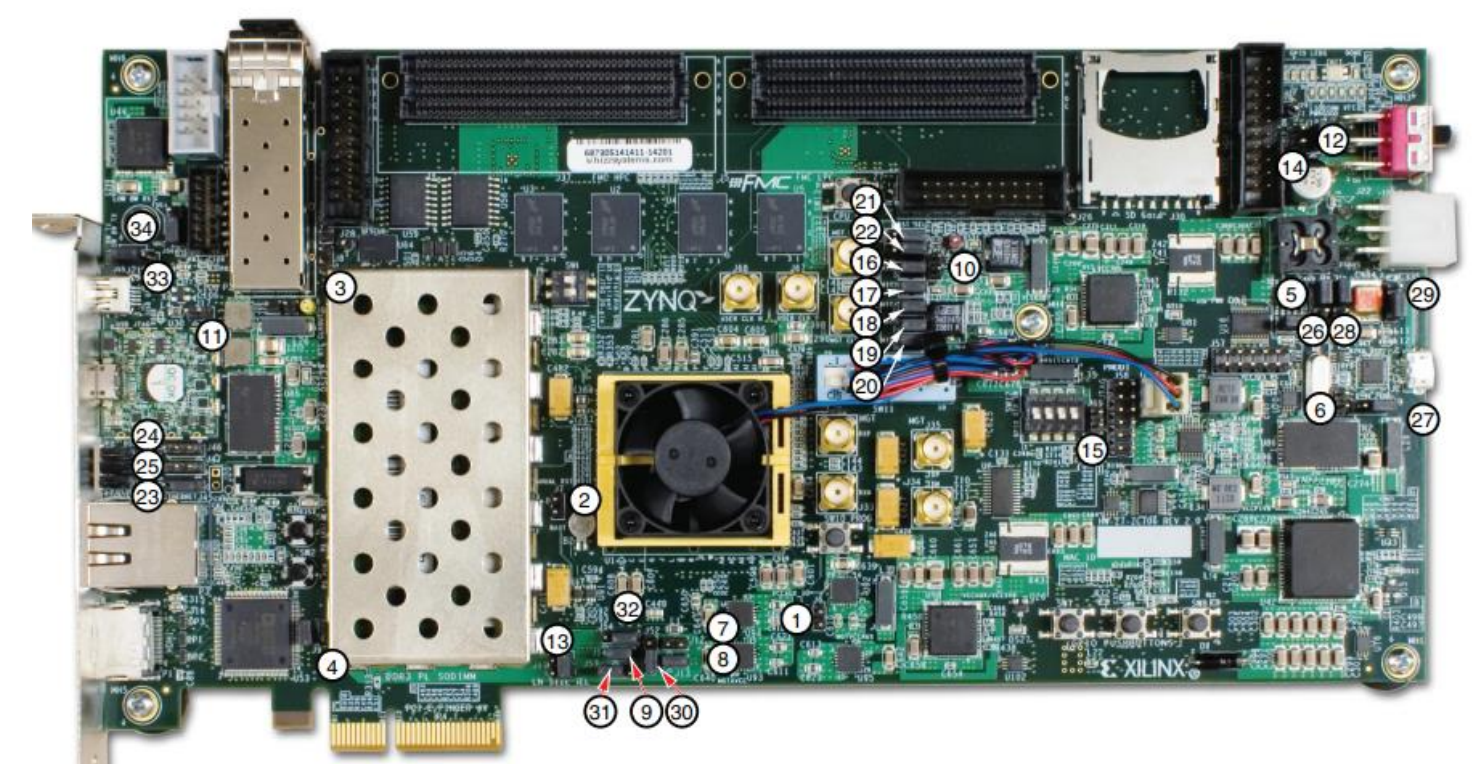


Throughput Comparison

Tesla V100 vs Xilinx ZC706



- 73x improvement in latency vs. NVIDIA RTX A5000
- 5.61x increase in throughput vs. NVIDIA Tesla V100



- Low power consumption of **0.575W**

On-Chip Power

