Machine Learning Training on FPGA

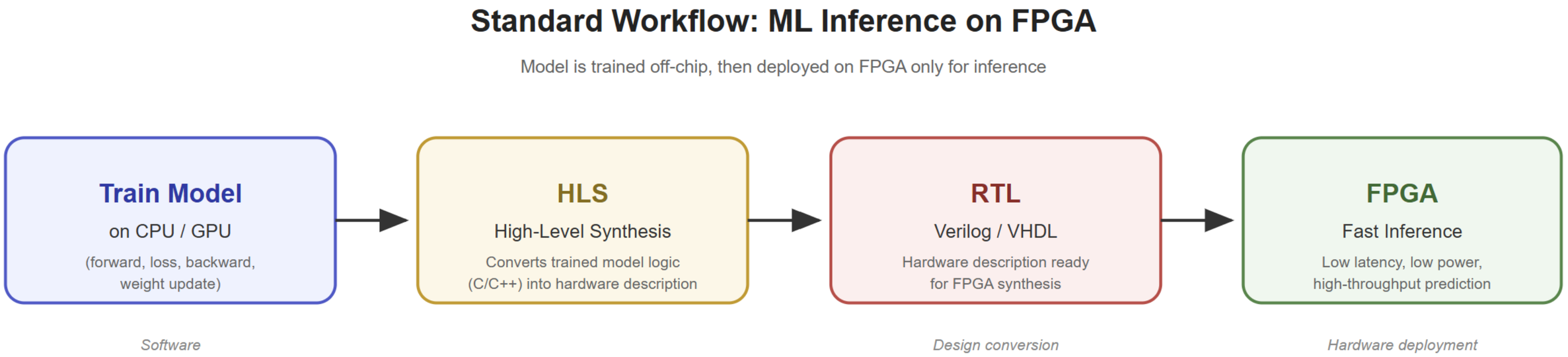
Nguyen Tran Duc anh

Supervisor: Mr. Yun-Tsung Lai

About me

- Viet Nam
- HUST-Hanoi University of Science and Technology
- Third year student
- Microelectronics and Nanotechnology Engineering
- Main research: FPGA
- Machine learning on FPGA

Machine Learning Training on FPGA



But what if the FPGA didn't just run a trained model — what if it could train itself?

→ This is the question my research explores

Machine Learning Training on FPGA

THE IDEA

"What if FPGA could train itself — not just run a trained model?"

- Full training loop runs directly on-chip:
  Forward → Loss → Backward → Update
- No CPU / GPU needed — model learns where it runs
- Enables real-time, on-device learning

CURRENT CHALLENGES

Limited Resources
Training needs more memory than inference — must store activations for backward pass

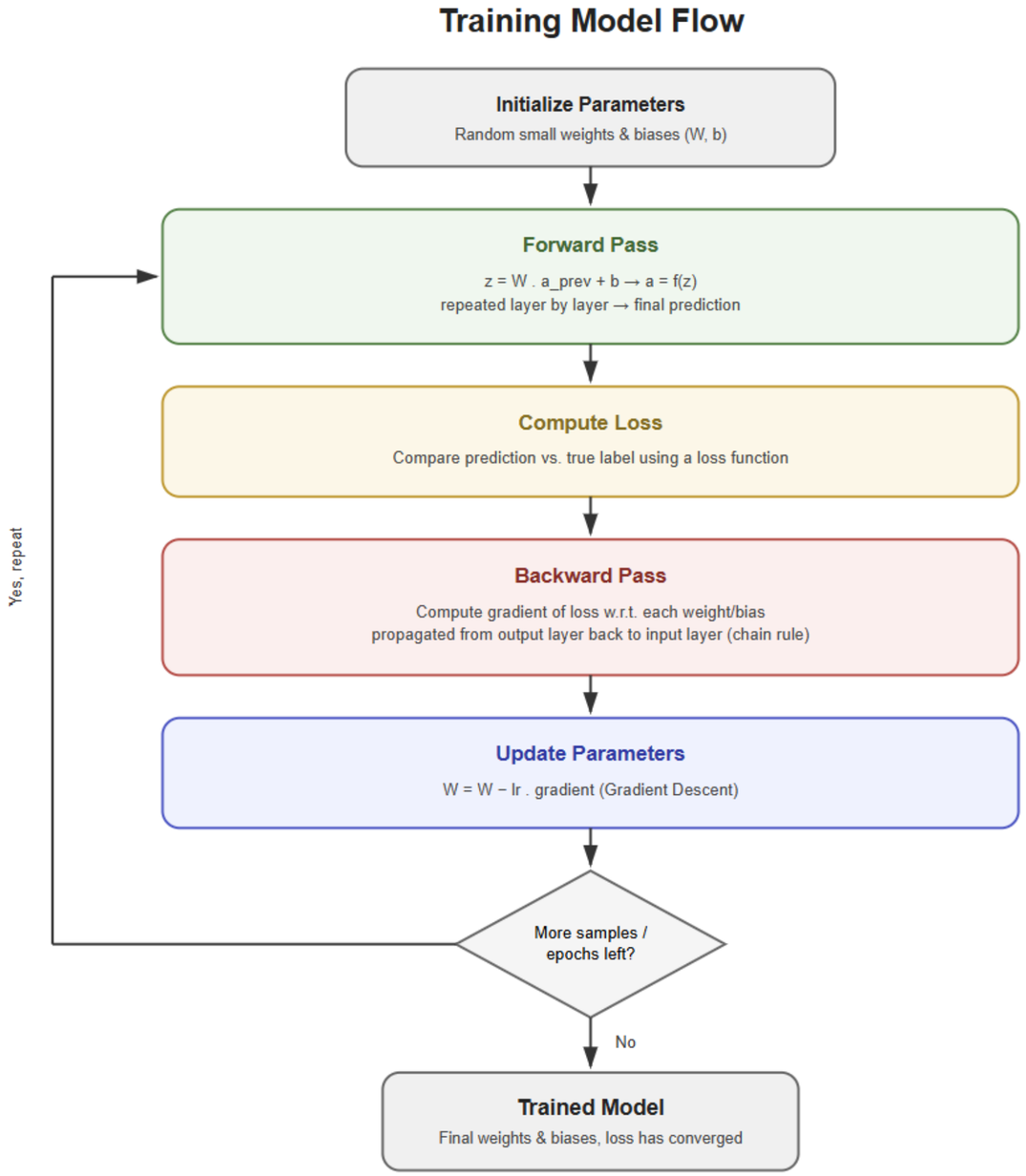
Numerical Precision
Floating-point is costly on FPGA; fixed-point introduces precision/stability issues

Sequential Dependencies
Backward pass depends on forward values; update must finish before next step — hard to pipeline

Complex Control Flow
Looping over samples/epochs is harder for HLS than inference's simple, static dataflow

- Build Software Reference Model
- Convert to Fixed-Point
- Rewrite in C/C++ for HLS
- HLS Generates RTL
- Simulate & Compare with Reference
- Deploy on Real FPGA
- Verify Final Results

In my first week



- **Forward and Loss** are straightforward — data flows in one direction

- **Backward** is where the real complexity lies: gradients must flow backward through every layer before anything can be updated

- **Update** only happens once every gradient has been computed — this sequential dependency is what makes training harder to parallelize than inference

My core mission

- My core mission: make the entire training loop — Forward → Loss → Backward → Update — run directly on FPGA hardware
- Not just inference — the model learns and adapts on-chip, without relying on CPU/GPU
- This matters because it removes the dependency on external hardware for training, and opens the door to real-time, on-device learning
- Long-term vision: a system that can keep improving itself in the field, without sending data back to a server or GPU cluster

Thank You!