From Matrix to Machine: GPU Hardware Design on FPGA for AI: Winter 2025

Why is it important to learn GPU Hardware Design for AI:

With the exponential growth of AI applications, there's an increasing need for specialized hardware solutions that can efficiently handle AI workloads. Understanding GPU hardware design on FPGA enables engineers to create custom, optimized solutions for specific AI applications, bridging the gap between theoretical knowledge and practical implementation. This course provides a unique opportunity for students to gain hands-on experience in both hardware design and AI applications, making them valuable assets in the rapidly evolving field of AI hardware acceleration.

What are we doing differently:

This first-of-its-kind course combines theoretical knowledge with practical implementation, allowing students to journey from understanding GPU software stacks to implementing actual GPU hardware on FPGAs. Instead of traditional theoretical approaches, students will work with industry-standard tools, HDL programming and use AWS EC2 F1 instances, gaining real-world experience in hardware design and implementation. The course culminates in a practical project where students design their own GPU for AI matrix multiplication operations.

Course Description:

This course takes students on a comprehensive journey from understanding GPU software stacks to implementing custom GPU hardware on FPGAs. Students will gain proficiency in Hardware Description Languages (HDLs), particularly Verilog, and learn to program FPGAs for specialized purposes. The curriculum includes case studies in healthcare and AI applications, providing practical insights into real-world implementations. The course also explores how the research community is collaborating to democratize custom hardware development.

Course Objectives:

- Master Hardware Description Languages, with a focus on Verilog
- Develop expertise in FPGA programming, simulation and synthesis
- · Learn to design processing elements for custom applications

- Gain comprehensive understanding of GPU architecture
- Create a functional GPU implementation on FPGA

Course Topics:

- 1. Digital Design
- 2. Hardware Description Languages, verification & Project Announcement/Tutorial
- 3. Basics of Computer Organization
- 4. Computer Architecture-2
- 5. GPUs- hardware & software
- 6. Review Lecture, Field Programmable Gate Arrays- how they work and case studies
- 7. More on Accelerator Design
- 8. Class Paper Presentation, Quantization on FPGAs for AI
- 9. Guest lecture from Prof. Yatish Turakhia, Paper Presentation
- 10. Online-Take Home Exam
- 11. Class Project Presentation

Assessment:

All assignments will be individually assessed and graded. The final project will include both implementation and documentation components. The final exam will cover both theoretical concepts and practical applications discussed throughout the course.

Prerequisites

- An ideal class consists of students with basic understanding of digital logic design.
- Some understanding of computer architecture concepts is required.
- Familiarity with any programming language is preferred.
- Prior experience with FPGAs or HDLs is helpful but not mandatory we cover these topics from basics.