

Course Logistics

Advanced Topics in VLSI: Deep Learning

Brad Quinton, Scott Chin

2026

Outline

- Introductions
- Course overview and philosophy
- Textbook, online resources
- Communications and content
- Mark distribution, midterm, final
- Lectures
- Rough timeline
- Assignments
- Prerequisite/background skills
- Questions, concerns, etc.

Introductions

- **Dr. Brad Quinton**

- Adjunct Professor, Department of Electrical and Computer Engineering, UBC
- Chairman and Advisor at Singulos Research Inc.
- Formerly Sr. Director, Technology at Qualcomm, Chief Architect at Tektronix, CEO at Singulos, CEO at Invionics, CTO at Veridae

- Who are you?

- EE? Comp Eng? Ph.D.? M.Sc.? Other?

Our day jobs....

- Scott and I are founders of local AR startup Singulos Research Inc. (<https://www.singulosresearch.com/>)
- At Singulos we develop fundamental IP on the leading edge of Deep Learning, computer vision and AI
- Our focus is on enabling new applications that are infeasible with current Deep Learning state-of-the-art

What is this course all about?

ELEC 502 Calendar Course Description:

Deep learning has emerged as an important technique for solving critical problems across a diverse set of applications, however VLSI architectures for implementing efficient learning systems remain an open question and active research area. In this course we will cover the **advanced VLSI topics** required when **designing deep learning architectures**, with a focus on VLSI design **trade-offs**, techniques and optimizations to create efficient and effective hardware acceleration of deep learning systems. Further, we will explore the latest optimized learning architectures and **open research topics** in this area.

What this course isn't

- This is **NOT** a course about applications of Deep Learning (image classification, NLP, *etc.*)
- This is **NOT** a course about using software tools and frameworks (TensorFlow, PyTorch, *etc.*) for Deep Learning
- This is **NOT** a course for learning VLSI design

Ok, so, what is it really about?

- This course is about building custom (“Advanced”) VLSI designs to *enable* Deep Learning techniques that would otherwise be infeasible on existing hardware (CPUs, GPUs, *etc.*)
- To get there we will dive deep into the core operations that underly the training and deployment of Deep Learning systems **AND** the fundamental trade-offs of VLSI design to support these operations
- **Our Goal:** Get you to where you can understand, and start to contribute to, the latest research on custom hardware for Deep Learning

Textbooks

- Ian Goodfellow, Yoshua Bengio and Aaron Courville, “Deep Learning”, MIT Press, 2016. Available online for free at: <https://www.deeplearningbook.org/>
- Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang and Joel S. Emer, “Efficient Processing of Deep Neural Networks”, Morgan & Claypool, 2020.
(<https://www.morganclaypool.com/doi/abs/10.2200/S01004ED1V01Y202004CAC050>)

Course communications and content

- All course material (slides, *etc.*) will be posted to **Canvas**
- **Canvas** discussions forms will be the primary path to getting help on assignments and answering questions in general

Mark Distribution

- Final exam: **30%**
- Mid-term: **25%**
- Assignments: **30%**
- Paper Presentation: **10%**
- Attendance/Participation: **5%**
- Notes:
 - Final exam will be cumulative
 - Midterm will be **Feb. 12, 2025** during class time
 - Assignments will be composed of 4-5 (I am working on a new assignment) programming assignments and 2 written paper reviews
 - Paper presentation will be “in-class”

Lectures

- Lectures will be delivered in-person except for 1-2 pre-recorded videos for complex math derivations that are easier to follow with a “pause” button
- We will **NOT** record lectures, the **expectation** is that you will attend the lectures/zoom sessions in person, and participate in class discussions

Lecture Participation

- Participation is a very important part of this course (and being a graduate student)
- I expect that you will ask questions, and provide your thoughts and insights during class
- A good way to be ready to participate in class is to spend time reviewing last lecture before class to be ready with thoughts/questions

Rough Timeline

- **Week 1:** Course Logistics, What is Deep Learning?
- **Week 2:** Logistic Regression, Neural Networks; **Assignment #1 due**
- **Week 3:** Neural Networks (continued), Backprop + Activation Functions; **Assignment #2 due**
- **Week 4:** Vectorized Backprop, CNNs; **Assignment #3 due**
- **Week 5:** CNN Backprop, CNN Arch; **Assignment #4 due**
- **Week 6:** CNN Arch (continued), **Midterm**
- **Week 7:** *Reading Break*
- **Week 8:** Custom Hardware for DNNs, Mapping NN Operations to CPUs/GPUs
- **Week 9:** RNNs, Attention and Transformers; **Paper Review #1 due**
- **Week 10:** DNN Accelerators, Dataflows, **Paper Review #2 due**
- **Week 11:** Flexible Interconnect, Buffering and Memory Architectures
- **Week 12:** Power Reduction/Estimation, Reducing Precision
- **Week 13:** Exploiting Sparsity, Efficiency of CNNs and Transformers
- **Week 14:** **Student paper reviews presentations (2 lectures)**

Assignments – Programming/Verilog

- There will be 4-5 programming assignments at the start of the year
- These will ensure you understand to inner workings of Deep Learning, before we look at hardware design
- Assignments are non-trivial, so please start early!
- Distributed as a Jupyter notebooks which will allow you to complete them from anywhere using a simple web browser (the login details will be available shortly)

Assignments – Paper Reviews

- There will be **2 written paper reviews** and **1 oral paper review/presentation** in the second part of the course
- These will help you get up-to-speed on the latest research in this area
- Expect these reviews to take a significant amount of work (especially the oral paper review)

Assignments – Getting Help

- There will be two primary means of getting help with assignments:
 1. Canvas.
 - Please use the Canvas discussion board as your first step (*i.e.* has someone already had this problem?)
 - Discussion board will be reviewed regularly
 - Please note direct e-mails about technical assignment details **will be ignored!**
 2. In-class
 - We want to have an interactive class! Feel free to interrupt during the lectures or even ask questions about the previous lecture at the start of the next.

Assignments / Paper Reviews – Plagiarism

- Please note all assignments are **individual** and anything you submit must reflect your personal understanding and implementation of the solution
- Make sure you **write your own code** as we will be monitoring and comparing submissions (remember we are AI researchers!)
- The consequences of academic misconduct big and the ownness is on the student to meet the standard of academic honesty:
 - <http://www.calendar.ubc.ca/vancouver/index.cfm?tree=3,286,0,0#15620>

What About AI?

- Can't I just use chatGPT to answer everything? Maybe...
- Why even take a course? Presumably, the goal is to learn.
- Evaluations (assignments, exams, reviews) are meant to provide motivation and feedback to facilitate learning (remember I already know the answers to the questions!)
- On the other hand, LLMs, are potentially valuable tools (not unlike the internet) that you will have access to in the future
- Therefore, rather than restrict usage, I will try to incorporate the use of AI tools into your exams and assignments (more to come as we get into the course)

Required Background

- Hardware Design
 - You **must** have completed a 4th year VLSI design course and understand hardware design
 - If you are not comfortable with hardware design concepts, including using HDLs, it will be very difficult to keep up in the second half of the course
- Python
 - You must be comfortable with Python
 - Lots of good on-line documents and tutorials if you need them, start now!
 - <https://www.python.org/doc/>
 - <https://pythonspot.com/>
- Linear algebra
 - As long as you remember the basics from your math course you will be fine
 - Textbook has a quick refresher
 - https://www.deeplearningbook.org/contents/linear_algebra.html

Final Thoughts

- This will be a fun, interesting and informative course, however, fair warning, there will be a lot of complex content and it will, therefore, take **a lot of work** to get through the assignments, paper reviews and exams

Questions, Comments, Concerns...

- Please log into Canvas to make sure it works and you are properly registered (<https://canvas.ubc.ca/>)
- My e-mail: bradq@ece.ubc.ca
- Questions?