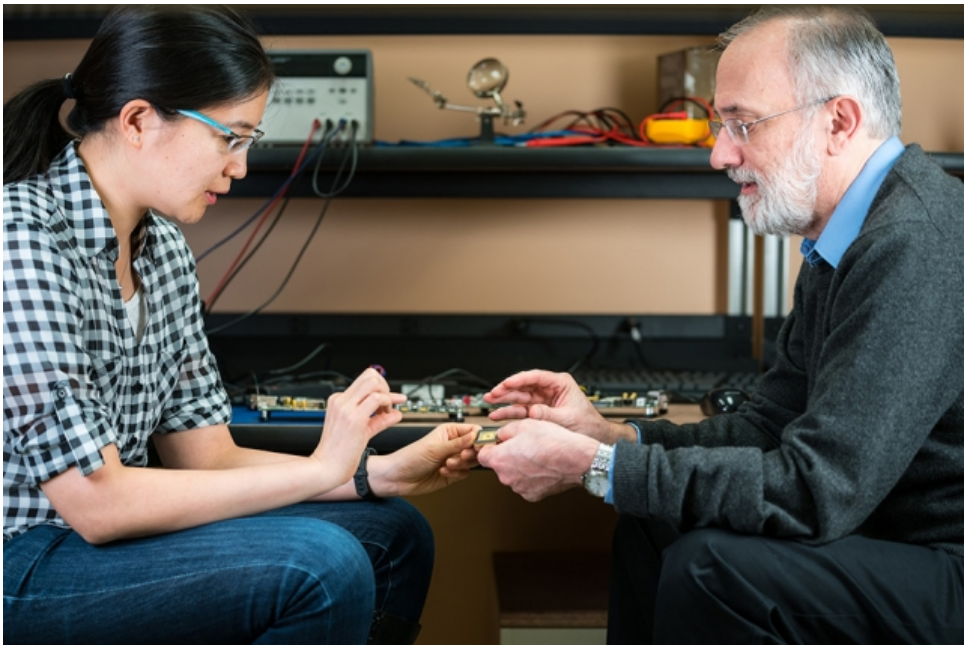
MIT News
ON CAMPUS AND AROUND THE WORLD



Vivienne Sze and Joel Emer teach Hardware Architecture for Deep Learning, a class in MIT's Department of Electrical Engineering and Computer Science that focuses on building specialized hardware for AI.

Image: Lillie Paquette/MIT School of Engineering

Building the hardware for the next generation of artificial intelligence

Class taught by Vivienne Sze and Joel Emer brings together traditionally separate disciplines for advances in deep learning.

Meg Murphy | School of Engineering
November 30, 2017

On a recent Monday morning, Vivienne Sze, an associate professor of electrical engineering and computer science at MIT, spoke with enthusiasm about network architecture design. Her students nodded slowly, as if on the verge of comprehension. When the material clicked, the nods grew in speed and confidence. "Everything crystal clear?" she asked with a brief pause and a return nod before diving back in.

This new course, 6.S082/6.888 (Hardware Architecture for Deep Learning), is modest in size — capped at 25 for now — compared to the bursting lecture halls characteristic of other MIT classes focused on machine learning and artificial intelligence. But this course is a little different. With a long list of prerequisites and a heavy base of assumed knowledge, students are jumping into deep water quickly. They blaze through algorithmic design in a few weeks, cover the terrain of computer hardware design in a similar period, then get down to the real work: how to think about making these two fields work together.

The goal of the class is to teach students the interplay between two traditionally separate disciplines, Sze says. "How can you write algorithms that map well onto hardware so they can run faster? And how can you design hardware to better support the algorithm?" she asks rhetorically. "It's one thing to design algorithms, but to deploy them in the real world you have to consider speed and energy consumption."

"We are beginning to see tremendous student interest in the hardware side of deep learning," says Joel Emer, who co-teaches the course with Sze. A professor of the practice in MIT's Department of Electrical Engineering and Computer Science, and a senior distinguished research scientist at the chip manufacturer NVidia, Emer has partnered with Sze before.

RELATED

Department of Electrical Engineering and Computer Science

School of Engineering

MIT Professional Education: Designing Efficient Deep Learning Systems

ARCHIVES


Vivienne Sze shares Engineering Emmy Award with colleagues


Bringing neural networks to cellphones


Energy-friendly chip can perform powerful artificial-intelligence tasks

Together they wrote a journal article that provides a comprehensive tutorial and survey coverage of recent advances toward enabling efficient processing of deep neural networks. It is used as the main reference for the course.

In 2016, their group unveiled a new, energy-efficient computer chip optimized for neural networks, which could enable powerful artificial-intelligence systems to run locally on mobile devices. The groundbreaking chip, called “Eyeriss,” could also help usher in the internet of things.

“I’ve been in this field for more than four decades. I’ve never seen an area with so much excitement and promise in all that time,” Emer says. “The opportunity to have an original impact through building important and specialized architecture is larger than anything I’ve seen before.”

Hardware at the heart of deep learning

Deep learning is a new name for an approach to artificial intelligence called neural networks, a means of doing machine learning in which a computer learns to perform some tasks by analyzing training examples. Today, popular applications of deep learning are everywhere, Emer says. The technique drives image recognition, self-driving cars, medical image analysis, surveillance and transportation systems, and language translation, for instance.

The value of the hardware at the heart of deep learning is often overlooked, says Emer. Practical and efficient neural networks, which computer scientists have researched off and on for 60 years, were infeasible without hardware to support deep learning algorithms. “Many AI accomplishments were made possible because of advances in hardware,” he says. “Hardware is the foundation of everything you can do in software.”

Deep learning techniques are evolving very rapidly, Emer says. “There is a direct need for this sort of hardware. Some of the students coming out of the class might be able to contribute to that hardware revolution.”

Meanwhile, traditional software companies like Google and Microsoft are taking notice, and investing in more custom hardware to speed up the processing for deep learning, according to Sze.

“People are recognizing the importance of having efficient hardware to support deep learning,” she says. “And specialized hardware to drive the research forward. One of the greatest limitations of progress in deep learning is the amount of computation available.”

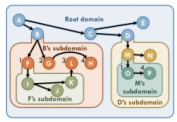
New hardware architectures

Real-world deployment is key for Skanda Koppula, a graduate student in electrical engineering and computer science. He is a member of the MIT Formula SAE Race Car Electronics Team.

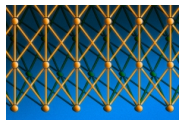
“We plan to apply some of these ideas in building the perception systems for a driverless Formula student race car,” he says. “And in the longer term, I see myself working toward a doctorate in related fields.”

Valerie Sarge, also a graduate student in electrical engineering and computer science, is taking the course in prepration for a career that involves creating hardware for machine learning applications.

“Deep learning is a quickly growing field, and better hardware architectures have the potential to make a big impact on researchers' ability to effectively train networks,” she says. “Through



Practical parallelism


Explained: Neural networks

this class, I'm gaining some of the skills I need to contribute to designing these architectures."

Topics: **Electrical Engineering & Computer Science (eecs)**

School of Engineering **Artificial intelligence** **Classes and programs**

Education, teaching, academics **Machine learning**

Computer science and technology **Algorithms**
