



MODERNISING MACHINE MEMORY

OUR DEVICES STORE CERTAIN DATA TEMPORARILY IN HARDWARE CACHES, WHICH ALLOW DATA TO BE ACCESSED FASTER. HOWEVER, THE SYSTEM DOES NOT WORK WELL FOR EMERGING APPLICATIONS AND PROFESSOR XIAOCHEN GUO, OF LEHIGH UNIVERSITY IN THE US, BELIEVES IT HAS THE POTENTIAL TO BE RADICALLY IMPROVED. SHE IS INVESTIGATING THE INTERPLAY BETWEEN SOFTWARE AND HARDWARE TO FULFIL THIS

TALK LIKE A COMPUTER ENGINEER

CACHE: A computer component that stores data, so that future requests for data can be fulfilled faster.

GRANULARITY: The breaking down of larger tasks or data into smaller tasks or data.

LOCALITY: The tendency of a processor to access data from a neighbouring memory location (i.e., spatial locality) or to repeatedly access the same data in a short time frame (i.e., temporal locality).

MACHINE LEARNING: Computer algorithms that improve themselves automatically through identifying patterns for increased efficiency.

MAIN MEMORY: A computer component that stores data, which can be addressed by the processor. Data have to be placed in main memory to be used by programs. Main memory can hold more data than hardware caches.

METADATA: Data that gives information about other data.

Leaps forward in data access are perhaps the defining technological advancement of the 21st century so far. From social media to streaming, we are able to access vast quantities of data at the tap of a touchscreen. Such wealth of content was barely imaginable even a couple of decades ago. As well as directly improving our daily lives,

this means that researchers have access to larger and more complex data sets than ever before, which unleashes new opportunities for science.

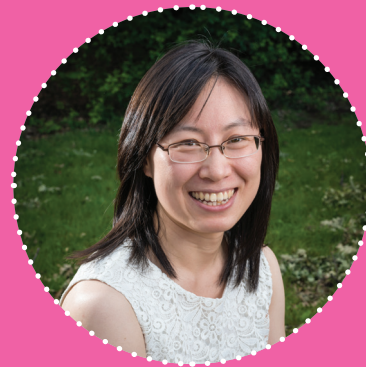
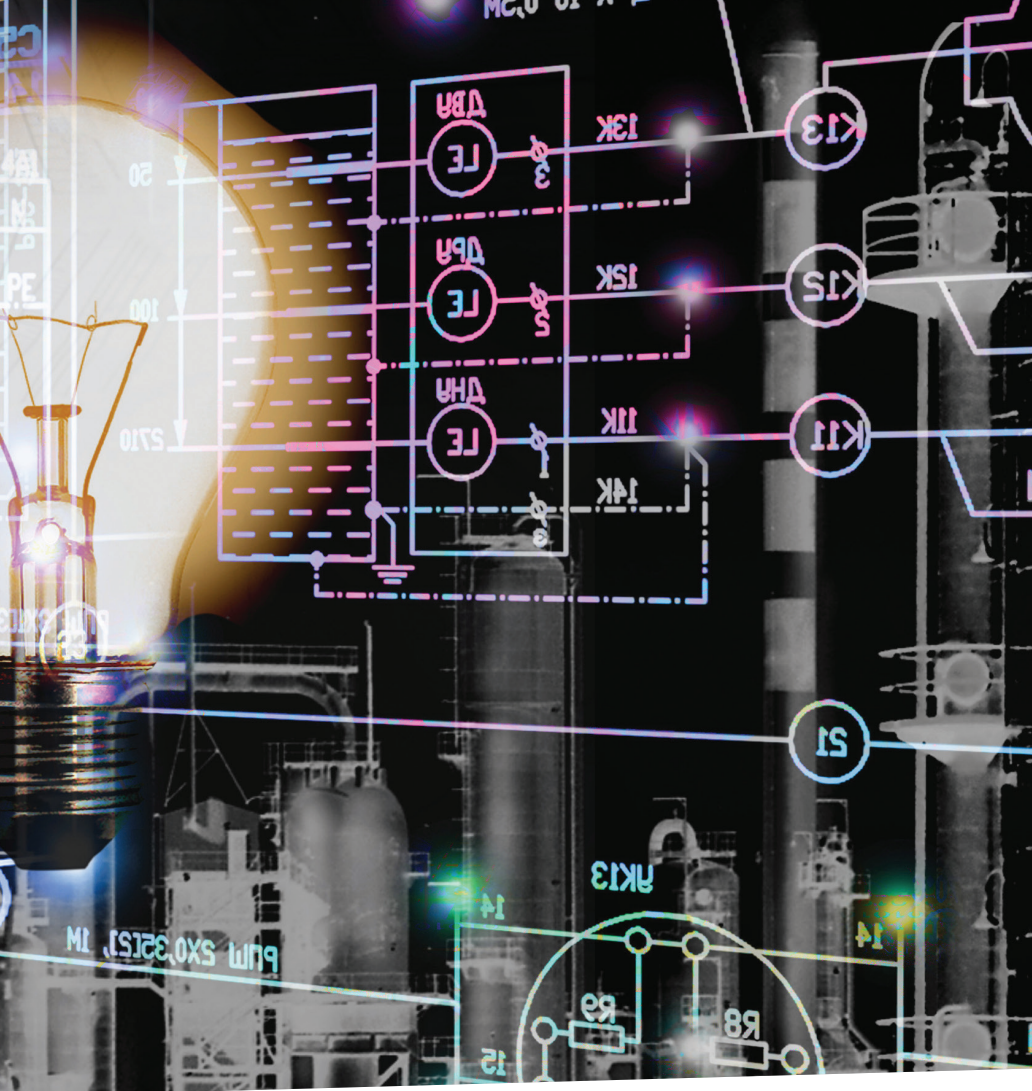
However, the technology is still relatively new, and there is plenty of room for improvement. Dr Xiaochen Guo works at the P. C. Rossin College

of Engineering and Applied Science at Lehigh University, which is situated in Bethlehem in the US. She and her team are busy revamping the way that memory systems work, to make them work better and more efficiently for big data applications.

MEMORY TROUBLE

“Current memory systems are designed for programs with good locality,” says Xiaochen. “This means programs where neighbouring memory locations are accessed repeatedly in a short period of time.” These data will be stored in the system’s cache. The cache is a hardware structure placed directly on the device and means that the device does not have to repeatedly access the same data from the main memory or the internet. It is designed to speed things up and reduce the processing power required. For instance, if you watch a video of a cat falling over on Instagram, the second time you watch it will be from your device’s cache. This is why when the WiFi cuts out when you have an app open, you can often still access a few photos and videos that had already been stored in the cache.

However, there is an issue with this. Although we sometimes like to watch the same video over and over again, modern-day browsing habits are more



PROF XIAOCHEN GUO

Lehigh University, P.C. Rossin College
of Engineering and Applied Science
Bethlehem, USA

.....

FIELD OF RESEARCH

Computer Engineering

.....

RESEARCH PROJECT

Revamping computer memory systems for
efficient data movement

.....

FUNDERS

National Science Foundation, Lehigh
University, US Department of Energy

likely to involve flitting between different apps and accessing a wide variety of content. Storing all this in the cache is pointless if it is never going to be accessed again. "Moving these useless data wastes energy and bandwidth, meaning that current memory systems need to be redesigned, especially as programs advance in complexity," says Xiaochen.

CACHE-22

"We are designing memory and cache architectures that are intelligent enough to avoid moving and storing useless data," says Xiaochen. Traditionally, this involves building systems that can support smaller access granularities – in other words, breaking down the data block received into smaller and more accessible pieces. Then, the hardware can decide which bits to keep and which to discard, rather than keeping the whole chunk. However, to do this involves introducing metadata, which is the 'overseer' used to tag which small pieces are in the cache. When using smaller access granularities, the proportion of metadata with respect to the actual data will increase, and this means that the total amount of data and metadata might not be reduced compared to a larger granularity.

Solving this conundrum involves minimising both the amount of useless data in the cache

and its associated metadata. The best way to reduce useless data is to anticipate what data will be useful. And to reduce metadata, common patterns among different metadata can be identified to compress the information. Both can be achieved by analysing the user's past habits.

ANTICIPATING NEEDS

"It can be difficult for hardware to anticipate which data are going to be requested next," says Xiaochen. "This is because hardware processes instructions one by one without knowing the high-level intention of the program." Her team's designs allow the programmer to pass on more information about potential user behaviour to the hardware, such that the hardware can predict better. This not only increases performance and available space in the cache, but also saves on energy expenditure.

"We envision this type of expressive hardware/software interface to be more common in the future," says Xiaochen. "This will involve large changes to programming languages, which will allow programmers to better optimise software."

CHALLENGES

Currently, as might be expected, machine learning applications take up a lot of data. It is a complex task that they are doing, after

all. Cleaning up this process by eliminating ineffective tasks is a key challenge for tech companies interested in this area.

"This project is pushing me out of my comfort zone," says Xiaochen. "We had to think outside the box to make our new programming interface able to make accurate predictions."



Xiaochen and her students
at Bethlehem, Pennsylvania.

ABOUT ELECTRICAL AND COMPUTER ENGINEERING

As its name suggests, electrical engineering is the branch of engineering that tackles the technology of electricity. As so much of modern society involves electricity, this has rapidly become a very broad field, covering everything from miniscule gadget components to the inner workings of power stations.

Computer engineering is a related field that covers the intersection between electrical engineering and computer science. Specifically, it is involved in the development of computer hardware and software, such as microprocessors, sensors and circuits.

THE DRAWS OF THE DISCIPLINE

"I was inspired to work in this field by the past success of the computing industry," says Xiaochen. "The capabilities of

computing have increased exponentially, which has led to tremendous social and economic impacts. I am motivated to contribute to this industry to further increase these capabilities, helping others to solve big problems."

Xiaochen received the coveted National Science Foundation CAREER award for this project. "I am honoured to receive this award, and to be recognised by the community," says Xiaochen. "It is an award for early career researchers and is a great way to help people like myself progress in the field."

INTO THE FUTURE

"I plan to continue to innovate on improving energy efficiency for computer systems," says Xiaochen. "I am also going to investigate

new computing paradigms. One particularly intriguing project involves using living neurons as computers!"

The future is bright for the discipline, which is ever-expanding as technology becomes ever-more integrated into our daily lives. "Attracting young people into the area is key," says Xiaochen. "Electrical and computer engineering tends to be more abstract compared to other fields of engineering, and often more challenging too, given how diverse it can be – but it's always extremely worthwhile."

HOW TO BECOME AN ELECTRICAL ENGINEER OR COMPUTER ENGINEER

- According to Top Universities, the best universities in the world for electrical and computer engineering are MIT, Stanford, University of California-Berkeley, Cambridge and ETH Zurich.
- A number of organisations offer apprenticeships in electrical and computer engineering. These can include companies designing computers and those working on smart phones, self-driving cars or robotics.
- Lehigh University, where Xiaochen works, has a CHOICES programme to introduce engineering and science to female middle school students. They also host the Lehigh Valley Science and Engineering Research Fair, which aims to increase awareness about STEM subjects amongst high schoolers.
- According to Indeed, the average salary for a computer engineer in the USA is \$93,000 per year.

PATHWAY TO BECOME AN ELECTRICAL ENGINEER OR COMPUTER ENGINEER

- Xiaochen recommends physics and maths as the most important subjects to take at school to pursue a career in electrical or computer engineering.
- Xiaochen also recommends learning programming early on, a skill that tends to be crucial for her field.
- Electrical engineering is a more common degree than computer engineering. Both of these, or a degree in related fields (such as engineering or computer science), can lead to a career in Xiaochen's field.



Students nominated Prof Guo for a Mentorship Appreciation Award.

HOW DID PROF XIAOCHEN GUO BECOME A COMPUTER ENGINEER?

WHO INSPIRED YOUR CAREER?

I have been very lucky to have the support of many people, including my family, friends, teachers and mentors. I still remember one of my grade school teachers who told me I should become a scientist, which gave me the confidence to pursue engineering. Later on, my PhD advisor said that I had the characteristics needed to become a good college professor, which encouraged me to pursue an academic career. Being recognised you can do something is very powerful!

WHAT MAKES A SUCCESSFUL ENGINEER?

Logical thinking, resilience, and the ability to collaborate with others have given me some advantages. Engineers need to be analytical, good at making connections and able to draw logical conclusions. It's important

to remember that many things can go wrong in engineering projects! Figuring out how to turn a failure into success is key. Additionally, being able to communicate with people with differing areas of expertise will open many more opportunities.

WHAT HAVE BEEN YOUR PROUDEST ACHIEVEMENTS TO DATE?

My students are a fantastic legacy. I have seen them solve difficult problems I did not know the answers to, come up with great ideas and get internships and jobs at well-known companies.

HOW DO YOU SPEND TIME OUTSIDE OF WORK?

I have always enjoyed sports. Nowadays, I am an endurance runner. I run almost every day and it is a good way to 'switch off'.

XIAOCHEN'S TOP TIPS FOR STUDENTS

- 01 It's fine to not know what you want to do, but you have to keep searching until you find it. I didn't know I would enjoy computer engineering and found it through exploring different subjects.
- 02 Being different is not a problem. Always be true to yourself. Doing a PhD set me aside from my peers, but I enjoyed focusing on what I found important and not feeling I had to fit in.
- 03 Learn to program as early as you can. As well as its directly applicable uses, learning to program is great training for logical thinking and problem solving.



Xiaochen at Lehigh University.