

HOW CAN WE READ GENETIC CODES IN A SINGLE PIECE OF DNA... QUICKLY?

GENE SEQUENCING IS AN INCREDIBLY USEFUL TOOL FOR READING THE VAST AMOUNTS OF INFORMATION DNA CONTAINS. WITH A CUTTING-EDGE TECHNIQUE CALLED 'NANOPORE GENE SEQUENCING,' THIS CAN BE DONE USING JUST A SINGLE PIECE OF DNA. HOWEVER, ONE CRUCIALLY IMPORTANT PART OF THIS PROCESS CAN ONLY BE DONE EXTREMELY SLOWLY. DR LEI JIANG, AT INDIANA UNIVERSITY BLOOMINGTON IN THE USA, AIMS TO SOLVE THIS CHALLENGE, USING THE LATEST TECHNIQUES IN COMPUTER SCIENCE

TALK LIKE A COMPUTER ENGINEER

DNA – long, stringy molecules containing the genetic instructions for all living organisms. They contain two ribbons of atoms, which coil around each other in a 'double-helix' pattern, and are bonded together by chemicals called 'nucleobases'

NUCLEOBASE – the building blocks of DNA molecules, which come in four possible types: cytosine, guanine, adenine and thymine. Within DNA molecules, they are always found in pairs

GENE SEQUENCING – a set of techniques that allows geneticists to find

out the order of nucleobase pairs within particular DNA molecules

BASE CALLING – an important part of the gene sequencing process, that involves translating the complex signals produced by gene sequencing into orders of nucleobase pairs

NEURAL NETWORK – computing systems that can be trained to recognise incredibly complex patterns, by mimicking exchanges of electrical signals within our own brains

In the world of nanoscience, things are done at an incredibly small scale. A nanometre is one thousand-millionth of a metre, and a nanopore is a very small hole measuring only a few of these nanometres. Amazingly, it is at this almost unimaginably small scale that researchers are making some huge advancements.

Nanopore gene sequencing is one of the latest approaches to gene sequencing and can be

used to read the genetic information carried by single molecules of DNA. Within a DNA strand, nucleobase building blocks always come in pairs (either cytosine with guanine, or adenine with thymine), and the information they carry is hidden in the order that these pairs appear within the DNA molecule. When studied using a piece of equipment called a 'sequencer,' the molecules can then generate distinctive patterns, which vary depending on these nucleobase orders.

In nanopore sequencing, this is done by stringing long strands of DNA through tiny holes, called 'nanopores', in a thin synthetic material. In the process, they generate tiny electrical currents in the material, which vary slightly depending on the order of nucleobase pairs that pass through the hole. By picking up these currents, geneticists can identify the unique genetic code contained in the DNA molecule, using a piece of computer code named a 'base caller'.

Compared with previous gene sequencing techniques, geneticists now hope that nanopore sequencing could soon be carried out using cheaper and more easily transportable equipment, which could produce results far more quickly. But before this can happen, there are still many improvements which need to be made, particularly to the base calling process. Dr Lei Jiang, a computer engineer at Indiana University Bloomington, aims to show how these improvements can be made.

HOW DOES THE BASE CALLER WORK?

In the latest methods, geneticists train neural networks to detect extremely subtle



DR LEI JIANG

Assistant Professor of Intelligent Systems Engineering, Indiana University
Bloomington, USA

FIELD OF RESEARCH

Computer Engineering

RESEARCH PROJECT

Developing a co-design of neural network algorithms and nanopore gene sequencing hardware, to speed up the base calling process

FUNDER

US National Science Foundation

patterns in the currents produced by nanopore sequencing. Once trained, they can be used to pick up patterns that would be all but invisible to even the most observant human scientists! In theory, this would make them ideal for identifying the unique orders of nucleobase pairs carried by DNA molecules – but today, the technology is still far from perfect.

Currently, even the smartest neural networks can only identify the order of about 1 million nucleobase pairs per second. That might sound fast, but since DNA molecules contain billions of nucleobase pairs, base calling is actually incredibly time-consuming. At this speed, it takes some 25 hours to analyse a single strand of human DNA (containing 3 billion nucleobase pairs) in enough detail to be useful to geneticists. Because of this, base calling takes far longer than any other part of the nanopore gene sequencing process and needs huge amounts of computing power to run.

HOW CAN BASE CALLING BECOME FASTER?

To solve such a challenging problem, Lei and his colleagues can make improvements to the neural network algorithms involved in base calling – but this is just one part of the solution. Alongside these improvements, they are also aiming to improve the physical equipment involved in collecting and analysing electrical currents – known as hardware. Through its research, Lei’s team is working towards a ‘co-design’ of new algorithms and hardware, which aims to make base calling more efficient, and analyse nucleobases at faster rates.

To achieve this, the researchers will aim to reduce the number of errors made during base calling measurements, while reducing the amount of power it requires to run. On top of this, the team also hopes to make nanopore genetic sequencing more accessible to geneticists – without the need for expensive equipment, or specialist training in how to use it. Ultimately, the team’s work is leading to base calling techniques that can identify nucleobase sequences far more quickly than ever before, allowing researchers to read genetic information whenever they need it.

WHO ELSE IS GETTING INVOLVED IN LEI’S PROJECT?

One of the most exciting and challenging aspects of Lei’s co-design project is that it involves two completely different fields of cutting-edge research. While most computer scientists probably do not know much about how gene sequencing works, neither will the average geneticist have the necessary skills to build and train neural networks. To bridge this gap, Lei is encouraging scientists working in the two different areas to work closely together and communicate their ideas clearly with each other.

An important part of the team’s goal is to make its results freely available online. This will allow scientists across the globe to benefit from the team’s discoveries. Lei and his colleagues are also looking out for new graduate and undergraduate students to join them and contribute to the research with their own fresh perspectives.

WHAT DOES THE FUTURE HOLD FOR THE PROJECT?

Already, Lei’s co-design approach has caused a stir in the gene sequencing community with companies including Pacific Biosciences, in the US, and Oxford Nanopore Technologies, in the UK, expressing a keen interest in his team’s work. With this backing, the researchers will have access to the funds, techniques and minds they need to improve base calling speeds even further.

Lei and his team now hope that their improvements to nanopore gene sequencing could soon be used to solve a diverse variety of real-world problems, including safer and more effective medicines that could be personalised to individual patients. In addition, they could lead to new farming techniques for growing crops that are more nutritious, and more resistant to diseases and natural disasters. This could improve food security for many millions of people in the developing world.

Even further, the team’s techniques could allow conservationists to better monitor the populations of animals, plants and fungi. This will help them to learn more about the threats that these wild species face, and the measures we will need to take to protect them in a rapidly changing world.

ABOUT COMPUTER ENGINEERING

WHY ARE COMPUTER ENGINEERS GETTING INVOLVED IN GENE SEQUENCING?

Since the two fields are so different, it might, at first, be hard to see how researchers in computer engineering and gene sequencing could come to work together. In his research, Lei used to focus completely on computer hardware, but like many other researchers in his field, he soon realised that he could not just focus on one aspect of computer engineering.

To develop better hardware, Lei saw that he also needed to study the software and algorithms that computers run, and how they are being used in real scientific research. This soon led his team to consider how the co-design of hardware and neural network

algorithms could be used to solve key challenges in gene sequencing.

WHAT IS REWARDING ABOUT A CAREER IN COMPUTER ENGINEERING?

For Lei, it is incredibly rewarding to guide his students, and to watch them as they master cutting-edge techniques in hardware design and ‘artificial intelligence’, and computer programs including neural networks, which can learn about the world around them by themselves and will likely be a centrally important part of future technologies.

Recently, Lei’s first student has joined the Samsung Artificial Intelligence Center – just one of many exciting opportunities that are now available in the ever-growing field of computer engineering.

WHAT ISSUES WILL THE NEXT GENERATION OF COMPUTER ENGINEERS FACE?

Until just a few years ago, computers seemed to follow a reliable trend called ‘Moore’s law’ – where they became roughly twice as small and twice as powerful every two years. But recently, it seems that this rapid progress has slowed down.

This presents an enormous problem for computer engineers, who are constantly aiming to build computers that consume lower amounts of power and perform better than devices that came before them. It will be up to the next generation of computer engineers to find new ways around this problem.

EXPLORE A CAREER IN COMPUTER ENGINEERING

- The Luddy School of Informatics, Computing, and Engineering at Indiana University Bloomington offers exciting opportunities for students from diverse backgrounds: luddy.indiana.edu/about/diversity/lodi.html
- Lei recommends you visit the Institute of Electrical and Electronics Engineers (IEEE) Spectrum website to find out about the exciting new directions that computer science is taking: spectrum.ieee.org/
- Indeed.com has a useful computer engineering guide: www.indeed.com/career-advice/careers/what-does-a-computer-engineer-do

LEI’S TOP TIP

Take STEM classes. STEM is cooler than NFL or the Premier League!

PATHWAY FROM SCHOOL TO COMPUTER ENGINEERING

- Lei says that mathematics – at school, college and university – is very important for all areas of computer science.
- Most computer engineers obtain a bachelor’s and many have a master’s degree.
- Leveragedu.com has useful information about the different degree-level subject areas related to computer engineering: leveragedu.com/blog/computer-engineering-subjects
- Find out more: www.learnhowtobecome.org/computer-engineer/

HOW DID LEI BECOME A COMPUTER ENGINEER?

WHO OR WHAT INSPIRED YOU TO BECOME A SCIENTIST?

I enjoyed reading when I was growing up, but it was my high school chemistry teacher who inspired me to become a scientist.

WHAT ATTRIBUTES HAVE MADE YOU SUCCESSFUL AS A SCIENTIST?

For me, being open-minded and giving students more freedom are two keys to

being successful. Sometimes, my students understand the details of a problem better than me!

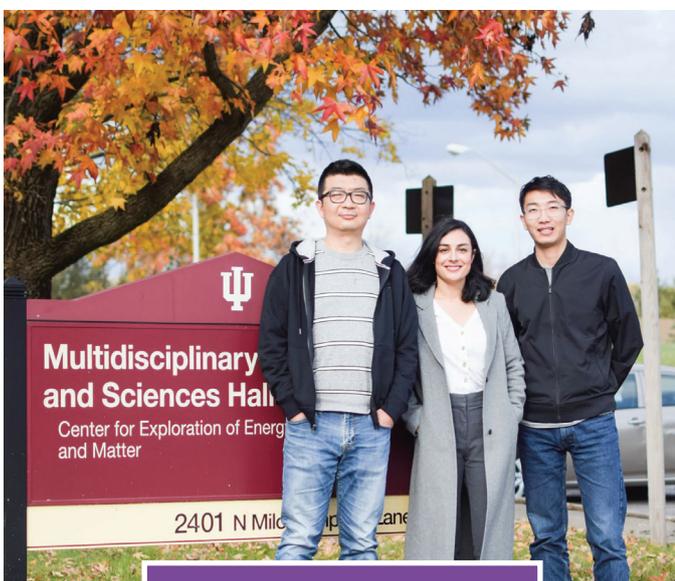
HOW DO YOU OVERCOME OBSTACLES IN YOUR WORK?

I solve a lot of problems by talking to senior colleagues, collaborating with peers, and interacting more with PhD students. Discussing ideas with a range of people has

helped me make many right decisions and tackle many challenges.

WHAT ARE YOUR PROUDEST CAREER ACHIEVEMENTS, SO FAR?

My first PhD student graduated last summer, which was a great career highlight for me. I now plan to build a larger research team and collaborate with more PhD students.



Lei with colleagues



Lei leading a presentation at Indiana University