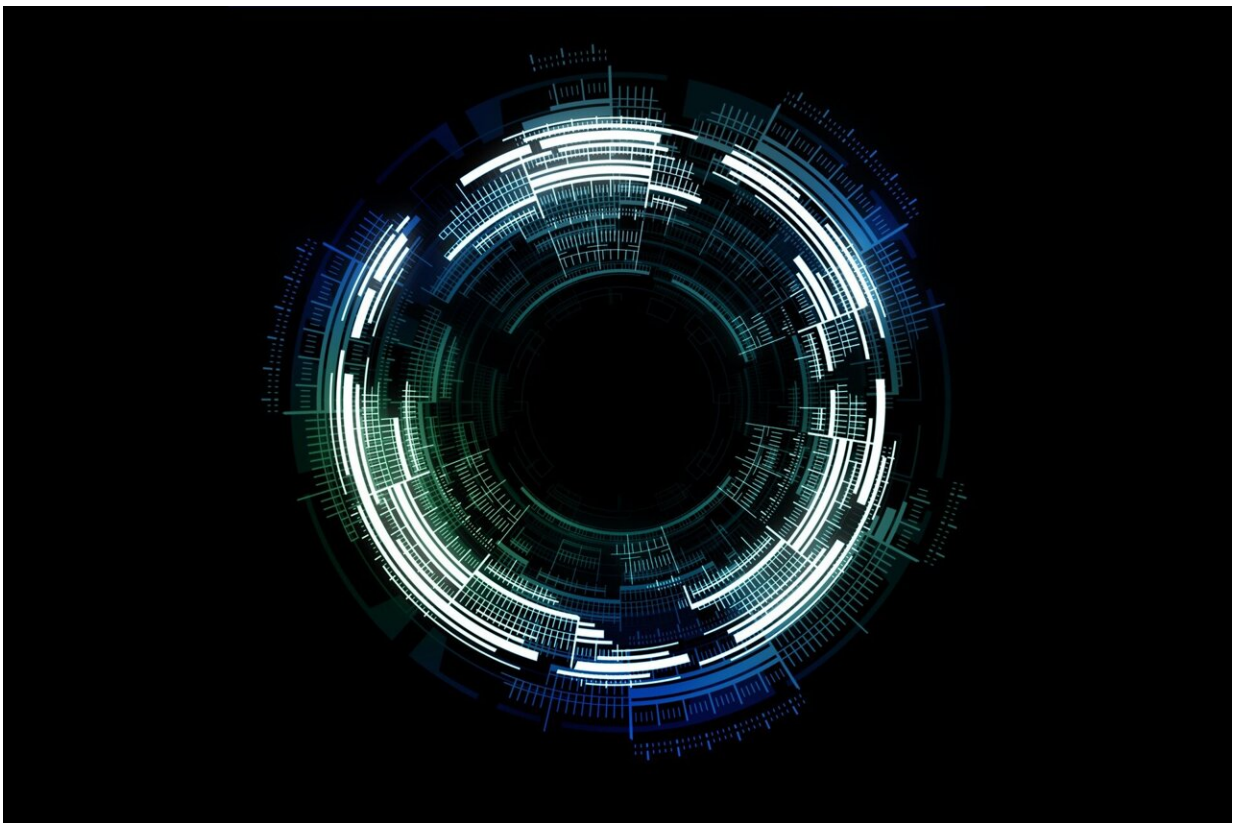


Breaking the scaling limits of analog computing

November 29 2022, by Adam Zewe



Credit: Pixabay/CC0 Public Domain

As machine-learning models become larger and more complex, they require faster and more energy-efficient hardware to perform computations. Conventional digital computers are struggling to keep up.

An analog optical neural [network](#) could perform the same tasks as a digital one, such as image classification or speech recognition, but because computations are performed using [light](#) instead of [electrical signals](#), optical neural networks can run many times faster while consuming less energy.

However, these analog devices are prone to hardware errors that can make computations less precise. Microscopic imperfections in [hardware components](#) are one cause of these errors. In an optical neural network that has many connected components, errors can quickly accumulate.

Even with [error](#)-correction techniques, due to fundamental properties of the devices that make up an optical neural network, some amount of error is unavoidable. A network that is large enough to be implemented in the [real world](#) would be far too imprecise to be effective.

MIT researchers have overcome this hurdle and found a way to effectively scale an optical neural network. By adding a tiny hardware component to the optical switches that form the network's architecture, they can reduce even the uncorrectable errors that would otherwise accumulate in the device.

Their work could enable a super-fast, energy-efficient, analog neural network that can function with the same accuracy as a digital one. With this technique, as an optical circuit becomes larger, the amount of error in its computations actually decreases.

"This is remarkable, as it runs counter to the intuition of analog systems, where larger circuits are supposed to have higher errors, so that errors set a limit on scalability. This present paper allows us to address the scalability question of these systems with an unambiguous 'yes,'" says lead author Ryan Hamerly, a visiting scientist in the MIT Research Laboratory for Electronics (RLE) and Quantum Photonics Laboratory

and senior scientist at NTT Research.

Hamerly's co-authors are graduate student Saumil Bandyopadhyay and senior author Dirk Englund, an associate professor in the MIT Department of Electrical Engineering and Computer Science (EECS), leader of the Quantum Photonics Laboratory, and member of the RLE. The research is published in *Nature Communications*.

Multiplying with light

An optical neural network is composed of many connected components that function like reprogrammable, tunable mirrors. These tunable mirrors are called Mach-Zehnder Interferometers (MZI). Neural network data are encoded into light, which is fired into the optical neural network from a laser.

A typical MZI contains two mirrors and two beam splitters. Light enters the top of an MZI, where it is split into two parts which interfere with each other before being recombined by the second beam splitter and then reflected out the bottom to the next MZI in the array. Researchers can leverage the interference of these optical signals to perform complex linear algebra operations, known as matrix multiplication, which is how neural networks process data.

But errors that can occur in each MZI quickly accumulate as light moves from one device to the next. One can avoid some errors by identifying them in advance and tuning the MZIs so earlier errors are cancelled out by later devices in the array.

"It is a very simple algorithm if you know what the errors are. But these errors are notoriously difficult to ascertain because you only have access to the inputs and outputs of your chip," says Hamerly. "This motivated us to look at whether it is possible to create calibration-free error

correction."

Hamerly and his collaborators previously [demonstrated a mathematical technique](#) that went a step further. They could successfully infer the errors and correctly tune the MZIs accordingly, but even this didn't remove all the error.

Due to the fundamental nature of an MZI, there are instances where it is impossible to tune a device so all light flows out the bottom port to the next MZI. If the device loses a fraction of light at each step and the array is very large, by the end there will only be a tiny bit of power left.

"Even with error correction, there is a fundamental limit to how good a chip can be. MZIs are physically unable to realize certain settings they need to be configured to," he says.

So, the team developed a new type of MZI. The researchers added an additional beam splitter to the end of the device, calling it a 3-MZI because it has three beam splitters instead of two. Due to the way this additional beam splitter mixes the light, it becomes much easier for an MZI to reach the setting it needs to send all light from out through its bottom port.

Importantly, the additional beam splitter is only a few micrometers in size and is a passive component, so it doesn't require any extra wiring. Adding additional beam splitters doesn't significantly change the size of the chip.

Bigger chip, fewer errors

When the researchers conducted simulations to test their architecture, they found that it can eliminate much of the uncorrectable error that hampers accuracy. And as the optical neural network becomes larger, the

amount of error in the device actually drops—the opposite of what happens in a device with standard MZIs.

Using 3-MZIs, they could potentially create a [device](#) big enough for commercial uses with error that has been reduced by a factor of 20, Hamerly says.

The researchers also developed a variant of the MZI design specifically for correlated errors. These occur due to manufacturing imperfections—if the thickness of a chip is slightly wrong, the MZIs may all be off by about the same amount, so the errors are all about the same. They found a way to change the configuration of an MZI to make it robust to these types of errors. This technique also increased the bandwidth of the optical neural network so it can run three times faster.

Now that they have showcased these techniques using simulations, Hamerly and his collaborators plan to test these approaches on physical hardware and continue driving toward an optical [neural network](#) they can effectively deploy in the real world.

More information: "Asymptotically Fault-Tolerant Programmable Photonics", *Nature Communications* (2022).

[dx.doi.org/10.1038/s41467-022-34308-3](https://doi.org/10.1038/s41467-022-34308-3)

Provided by Massachusetts Institute of Technology

Citation: Breaking the scaling limits of analog computing (2022, November 29) retrieved 19 May 2024 from <https://techxplore.com/news/2022-11-scaling-limits-analog.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is

provided for information purposes only.