Photonic Chips for Machine Learning

Mariah Smith-Jones

Boston University

Abstract

The purpose of this paper is to explore the use of photonics chips in machine learning. Similarly to how the terms "machine learning" and "AI" have become buzzwords that draw interest to any tech project, photonics is generating the same buzz in regards to chips. This surge in interest is not without reason, as the increasing demands for computational power for artificial intelligence has led researchers to turn towards photonics as a possible solution to overcome the capacity limits of current hardware. The main difference between how electronic and photonic chips operate is that electronic chips use electrons to carry out their calculations while photonics chips utilize photons. Because the optical fibers in photonic chips possess a greater bandwidth and faster speed than electrical wiring, this gives photonics chips a significant edge in performing matrix multiplication and expedites the process of machine learning/deep learning.[2] These optical fibers also have the added benefit of being more energy efficient/eco-friendly than their electric counterparts. The increase in efficiency provided by photonic chips will result in the ability to create more complex and powerful AI systems than those built with electronic chips.

Review of Literature

Demands of Machine Learning

Most chips that are currently in devices used for machine learning are electronic, but the demands of computational power for machine learning are steadily increasing and the current devices are reaching the limit of what they are capable of on their own. One of the most effective methods for deep learning is through use of a neural network, which requires processing vast quantities of data. In recent years these neural networks have become larger and more complex in an effort to improve the efficiency of machine learning, which in turn requires an increase in computational power in order to handle the larger load of data.[2] Another limitation encountered by neural networks is tied directly to the limitations of electronic devices.

> First, they are usually implemented using digital-clock-based platforms such as graphics processing units (GPUs), which limits their computation speed to the frequencies of the clocks—less than 3 gigahertz for most state-of-the-art GPUs. Second, conventional electronics separate memory and processing units. Shuttling data back and forth between these components wastes both time and energy. In addition, raw visual data usually needs to be converted to digital electronic signals, consuming time. Moreover, a large memory unit is often needed to store images and videos, raising potential privacy concerns.[3]

These concerns in combination with the looming limit of the number of transistors able to be fit on a single silicon chip requires companies to look at alternative sources for computing.[8] As a result, researchers and companies are turning towards photonics as a potential answer because of their many benefits.

Photonic Chips

The primary benefits to using photonic chips are their significant increase in computational speed and being more environmentally friendly. The increase in speed will allow the calculations for machine learning to be carried out faster, which means automated systems will be able to learn from the data faster. The benefits of this can already be seen on a photonic microchip that can classify images in less than a nanosecond. One of the most obvious areas that will benefit from this is autonomous vehicles, where learning from data faster will allow the vehicle to make decisions faster which is incredibly important when decisions made in a second mean the difference between safely avoiding an accident and death and destruction. Less heat means that the device with the photonic chips will be able to run for longer without risk of malfunctioning from overheating, and directly leads to the third benefit of it being more eco-friendly. One of the biggest issues with electronic chips is the amount of heat that they produce. This is caused by the combination of the way electronic chips function, and because of the number of resistors they have. The advances in technology in recent years have led to the ability to produce transistors for chips at smaller and smaller sizes, resulting in the ability to produce chips at a smaller size while simultaneously increasing its computing power. Unfortunately, more transistors also means more heat. This increase in temperature means that these devices can only be run for so long before becoming incredibly hot, and need some way to keep the device cool. This phenomenon is most easily seen in data centers, where a large portion of the expense in running the data center comes from the cost of keeping the electronics cool. This can also be observed in a much smaller scale with laptops where when the laptop is running a lot of resource intensive applications the laptop will become very hot. While photonics is a new technology relative to electronics, it has been in use since 1975, when optical fibers replaced a submarine copper communications cable.[6] As the cable has not been replaced with another electric cable, this demonstrates the use of photonic not without precedent, and is a logical successor to electronic technology.

Environmental considerations. While eco-consciousness was not at the forefront of consideration during the early days of computing, a larger push towards environmentally friendly solutions also leans in favor of photonic devices over electronic devices due to significant decrease in the amount of energy required to produce the same results. One study done on the environment effects from training a deep neural network to be used for natural-language processing on current devices produced five times the CO2 emissions typically associated with the amount produced by that of an automobile over the course of its lifetime.[2] Another environmental concern with electronic chips is their inability to be reprogrammed. In addition to the aforementioned heat issue that requires a large amount of energy for cooling and often replacement if pieces get to hot, because electronic chips can not be reprogrammed it requires entirely new chips whenever an upgrade is required contributing to the ever-growing amount of e-waste.

> Other chips are physically wired through metal, which makes them hard to rewire and redesign, so you'd need to make a new chip if you wanted to add any new function," says MIT postdoc Hyunseok Kim. "We replaced that physical wire connection with an optical communication system, which gives us the freedom to stack and add chips the way we want."[5] "When current computer chips process electrical signals they often run them through a Graphics Processing Unit, or GPU, which takes up space and energy," says Ashtiani. "Our chip does not need to store the information, eliminating the need for a large memory unit.""And, by eliminating the memory unit that stores images, we are also increasing data privacy," Aflatouni says. "With chips that read image data directly, there is no need for photo storage and thus, a data leak does not occur." A chip that reads information at light speed and provides a higher degree of cybersecurity would

undoubtedly have an impact in many fields; this is one of the reasons research into this technology has ramped up in the past several years.[1]

Today's trend of reducing carbon footprints collides with the growing AI models. Pure photonics seems to be the only large-scale sustainable solution to reconcile these trends.[6]

Drawbacks of photonics. Despite the many benefits of using photonics chips, they are not without their drawbacks. The biggest problem with photonic chips is that producing them is expensive. Even though the physical manufacturing has become easier in recent years, the main cause for the expense is the lack of standardization, which requires them to be be custom-built, and until standard libraries are set up.[8] Combined with the chip shortage during the past couple of years, the research into and production of them has been delayed further driving up the cost. Silicon chips reduce the space, time, and energy needed to conduct computations, but the Another drawback of photonics chips is that at this time they are not able to be produced at the same size as electronic chips. Because the optical components of a photonic chip are tens of micrometers in size, they can not be packed nearly as tightly as transistors which result in a much larger chip.[2] This is also due to the larger wavelength of photons.[6] This problem may prove to be a temporary one as demonstrated by the photonic chip developed by the University of Pennsylvania, which was able to perform calculations much faster than an electronic one while also being much smaller than most photonic chips that have been developed so far. optical deep neural network implemented on a 9.3 square millimeter chip.[1] Another drawback of photonic chips comes from the optical processors suffering from various sources of noise and because the digital-to-analog and analog-to-digital converters used to get the data in and out are of limited accuracy.[3] Another issue with photonics chips is spurious reflections. Another issue with photonic devices is that waveguides and fibers are harder to use than wire. While research is currently being conducted into alleviating some of these drawbacks, there is currently no solution for them. Despite their current problems and limitations, the obvious immediate benefits of incorporating them into existing devices has led to companies adopting a hybrid model to further pursue development of these chips.

Hybrid Models

Photonics chips are primarily being used by companies in devices that are a combination of electronic and photonic. This is observable in Lightelligence's hybrid photo-electronic Comet and PACE demonstrator chip. Pace is the successor chip of Comet, and has 1 million-fold speedup compared to their prototype from 2019[4] This large improvement of processing speed from their original chip to the current version in just two years is incredibly significant, especially when done in conjunction with the fact that it was also "the first time Lightelligence showed use cases beyond AI acceleration on its hardware. Pace can run algorithms for the NP-Complete class of problems, which are computationally extremely difficult, many times faster than existing accelerators. While not demonstrating optical superiority for all applications, it did execute the Ising problem 100 times faster than a typical GPU, even beating a system purpose-built for the Ising problem–Toshiba's simulated bifurcation machine, which runs on FPGAs–by a factor of 25."[4] Another such example Boston-based company Lightmatter, in combination with Harvard and Boston University received a grant from IARPA to develop a new Electro-Photonic Computing (EPiC) system for AI-based navigation in Autonomous Vehicles.[3] While it remains to be seen to what extent the photonic chip will improve the capacity of autonomous vehicles to react to obstacles and make decisions, current information we have on photonic chips point to at the very least a significant improvement in their performance. ", Lightelligence's vice president of engineering, said individual chips are hard to engineer, and integrating them is even tougher. "With photonic computing, it's really a class of analog computing. So a high-fidelity result requires a tremendous amount of circuit design, simulation, iteration and test chips," he said. Moreover, at 1 GHz, the system uses light pulses shorter than a nanosecond. Compared to a megahertz system, noise and electronic crosstalk are proportionally much larger."[4] ""The other [challenge] is the packaging architecture," Steinman said. "We are taking two chips built on different fabrication processes and stacking them up directly with thousands of connections between them. "One is powered by light, so we need to get a light source in there. The other needs electric current to power it… and heat removal. There are tremendous challenges we have to systematically attack to get all of that to come together," he added."[4] "Like competitor Lightmatter, Lightelligence uses a silicon photonics version of the Mach Zehnder Interferometer (MZI) as its computing element. However, where Lightmatter uses MEMS to change the physical shape of the waveguide in its MZI, Lightelligence instead injects electrons into the waveguide to modulate its photonic refractive index, modulating the optical signal passing through it.”[4]

Pure Photonics

Despite the hybrid models primarily used by companies, universities are still pursuing research on the applications of pure photonic chips. One such example is MIT’s reconfigurable artificial intelligence photonic chip, which holds promise for reducing the amount of electronic waste currently created by chips.

> The design comprises alternating layers of sensing and processing elements, along with light-emitting diodes (LED) that allow for the chip’s layers to communicate optically. Other modular chip designs employ conventional wiring to relay signals between layers. Such intricate connections are difficult if not impossible to sever and rewire, making such stackable designs not reconfigurable. The MIT design uses light, rather than physical wires, to transmit information through the chip. The chip can therefore be reconfigured, with layers that can be swapped out or stacked on, for instance to add new sensors or updated processors. “You can add as many computing layers and sensors as you want, a reconfigurable AI chip because it has unlimited expandability depending on the combination of layers.”[5]

Meanwhile researchers at Melbourne’s Monash University and RMIT have developed a photonic circuit they believe will advance research into various areas of artificial intelligence.

> While photonics opens up massive opportunities, it is not easy going to make devices which can be programmed and reprogrammed. This is because the manufacturing needs to be done with precision down to the scale of the wavelength of light – nanometres.“Our solution is to calibrate the chips after manufacturing, to tune them up in effect by using an on-chip reference, rather than by using external equipment,” explains Lowery. “We use the beauty of causality, effect following cause, which dictates that the optical delays of the paths through the chip can be uniquely deduced from the intensity versus wavelength, which is far easier to measure than precise time delays. We have added a strong reference

path to our chip and calibrated it. This gives us all the settings required to 'dial up' and desired switching function or spectral response." Instead of dialing in a setting manually, the chip is tuned in one step allowing data streams to be switched seamlessly. "As we integrate more and more pieces of bench-sized equipment onto fingernail-sized chips, it becomes more and more difficult to get them all working together to achieve the speed and function they did when they were bigger," says Dr Andy Boes, a collaborator from the University of Adelaide. "We overcame this challenge by creating a chip that was clever enough to calibrate itself so all the components could act at the speed they needed to in unison."[7]

Future Applications

There are many future applications for photonic chips in the realm of machine learning. The primary benefit of photonics chips is their ability to perform calculations faster. This in turn could lead to other developments not previously thought of as a result of being able to have an increase in ability to compute. The University of Pennsylvania plans to continue their research into making their chip smaller, as well using it to classify 3-D images in addition to 2-d images.[1] They have also begun experimenting with classifying video, as well as testing whether larger chips will be able to classify higher-resolution images.[3] Researchers have developed a photonic deep neural network that can directly analyze images without the need for a clock, sensor, or large memory modules. It can classify an image in less than 570 picoseconds, which is comparable with a single clock cycle in state-of-the-art microchips.[3] This significant increase in speed demonstrated by the photonic chip may be the key needed to push the capabilities of machine learning to the next level, even though the full extent of its applications may are not fully understood.[1] This combined with the smaller size may be the push needed to start the switch from photonic to electronic. MITresearchers also remain hopeful about the future applications of their reconfigurable chip, where they hope to implement different types of neural networks, such as for image or voice recognition, that can have different pieces swapped in as necessary.[5] I believe that not only will photonics replace electronics in devices for machine learning, they will eventually replace all electronic devices. While there is still a long way to go before the switch is made from electronic devices, the future is undoubtedly photonic.

RQ1: To what extent can electronic devices be replaced by photonic ones?

Method

To answer this question I would focus on the capabilities of purely photonic devices, in order to accurately examine the capability of current photonic devices and to determine how much more work needs to be done. Because of the vast array of possibilities for photonic devices, my research would focus solely on photonic devices in relation to machine learning. In order to determine the full current capabilities of these devices I would need to do a more in-depth review of publications of these devices, and if possible speak with the people who made them, and then compare them to current devices, such as GPUs, that are being utilized for machine learning.

References

1. Pappas, M. (2022, June 1). *Penn engineers create chip that can process and classify nearly two billion images per second*. Penn Engineering Blog. Retrieved September 18, 2022, from https://blog.seas.upenn.edu/penn-engineers-create-chip-that-can-process-and-classify-nearly-two-billion-images-per-second/
2. Hamerly, R. (2021, August 4). *The future of deep learning is photonic*. IEEE Spectrum. Retrieved September 18, 2022, from https://spectrum.ieee.org/the-future-of-deep-learning-is-photonic
3. Choi, C. Q. (2022, July 12). *Photonic chip performs image recognition at the speed of light*. IEEE Spectrum. Retrieved September 18, 2022, from https://spectrum.ieee.org/photonic-neural-network
4. Ward-Foxton, S. (2022, January 4). *Optical chip solves hardest math problems faster than gpus*. EETimes. Retrieved September 18, 2022, from https://www.eetimes.com/optical-computing-chip-runs-hardest-math-problems-100x-faster-than-gpus/
5. Chu, J. (n.d.). *Engineers build Lego-like Artificial Intelligence Chip*. MIT News | Massachusetts Institute of Technology. Retrieved September 18, 2022, from https://news.mit.edu/2022/stackable-artificial-intelligence-chip-0613
6. Pelé, A.-F., & Cohen, E. (2022, August 22). *Photonics computing converts decades-long skeptics*. EE Times Europe. Retrieved September 18, 2022, from https://www.eetimes.eu/photonics-computing-converts-decades-long-skeptics/

7. Yazgin, E. (2022, July 7). *World's first self-calibrating photonic chip*. Cosmos. Retrieved September 18, 2022, from https://cosmosmagazine.com/technology/self-calibrating-photonic-chip/
8. Gooding, M. (2022, June 22). *Silicon photonic chips could hold the key to faster and more efficient data centres*. Tech Monitor. Retrieved September 18, 2022, from https://techmonitor.ai/technology/silicon/silicon-photonic-chips-data-centre