HUGE Breakthrough In Computing with Light

A transcription of a YouTube Video: https://youtu.be/b_PS8o8pi9A?si=xIi2aywO75g4myIC

This tiny microchip is only a few millimeters across. But if you zoom in, you would see hundreds of billions of transistors, so small they're just a few atoms wide. But we can't really shrink them much further.

At the same time, our ambitions for artificial intelligence keep skyrocketing, and it's demanding more computing power than ever before. Here's a big question. If today's chips can't keep up, what comes next? But maybe the answer isn't in electrons at all, but in light.

That would mean computing literally at the speed of light. This new microchip makes it real. **It's up to a thousand times faster than today's chips, using the same power as a single LED bulb.**

This new light-based processor is a big collaboration between top US universities, led by some of the biggest names in photonics. After years of trial and error, they've done something stunning. For the first time, light itself can not only compute, but also remember with incredible precision.

Something that seemed impossible just a few years ago. But hold on, how do you actually store light? Inside your smartphone, there are over 60 chips, built of more than 100 billion transistors. Think of transistors as tiny switches, so small you can't actually see them.

They do computing by flipping on and off billions of times every second. Every time it flips from 0 to 1, it's like cars driving through city traffic. They have to stop at every red light, start again, stop again, and it all takes time.

It slows computing down and wastes energy. With light, it works differently. Light is a wave, so you can process it while it's still moving, without ever stopping the data.

You can bend it, split it, or combine it, and it just keeps flowing. In other words, here you're computing on the fly. And that's the beauty of it.

And the best part, it uses way less energy, because you only need energy to send and to receive the light. At least, that's the theory. But the light alone has a fatal flaw.

Memory. Imagine you're worried about how much energy modern AI is burning, and you should be. And you decide to invent a completely new kind of computer.

To make it work, there are three things you have to figure out. First of all, compute. You have to find a way to manipulate signals, to add and to multiply them.

Second is interconnect, the ability to move data around on a chip. And the third one, which is usually forgotten, is memory. You need to be able to somehow store results in order to be able to use them later on.

With photonic chips, we do really well on the first two. We can compute with light, and we can send data across chips and even entire data centers with light. But storing results with light has always been a weak spot, because reliable photonic memory doesn't exist.

This means whenever we have to save data, we have to go back to electronics, basically to switch from light to regular transistors. That part slows everything down, and in many cases cancels out the very benefits of using light in the first place. And this is exactly the part scientists thought was impossible.

There is no way to store light. And this new processor, **for the first time, actually gives light a memory**. And this is a paradigm shift, because *now the computing happens right where the data is stored.*

They found a way to compute directly in memory. You could say this processor never forgets, because it's got a photographic memory.

Now, this computer can actually remember things, and that's where it starts to get really interesting. If we zoom in into the chip, it's built of special devices called resonators.

You can imagine it as a tiny ring that traps light. You can think of it like a wine glass. You tap it, and it rings at one special note.

In the same way, the resonator rings at one special color of light. By tuning the ring, you can decide how light goes through. And if you could build memory right there, it could be a game-changer for computing, because AI demand is exploding.

In 2025, over half of all companies are already using AI. Research shows that 40% of people worry AI will replace them at their job. But in reality, people using AI will replace those who don't.

And it's already happening right now. Microsoft, Google, Amazon are currently hiring people that understand AI, who know how to build with AI. If you build a startup like me, or create anything, or a working professional, AI isn't a threat to you.

It's a leverage. Because used right, it can help you to save hours of time and thousands of dollars in costs. And you need to learn how to use it now.

So, how do they actually store light on this computer chip? **They built a tiny ring on the chip and attached a special phase-change memory right on top of this ring.**

It's made of a crystal-like material. This layer can store numbers with very high precision, up to 12 bits, which means the chip can do calculations much more accurately, which is a big deal for photonics, because one of the biggest flaws of its analog nature is actually precision. **So, by putting these two parts together, ring and memory device, we can actually store data on the chip.**

And because we combined computer engine and memory at one place, we can do in-memory computing. So, when light passes through the ring, it doesn't just move, it does the math instantly right where the value is stored. And that's a big deal, because for the first time a light-based processor got the memory.

Finally, a chip with light memory. Why does this matter? Because our current processors, like in your laptop or your phone, move data back and forth between memory and processor. In fact, **roughly 80% of total energy is spent in this moving data rather than on compute itself.**

Imagine, that's like driving a Porsche 911 GT3 RS in rush hour traffic. What a waste. While this new photonic chip avoids this bottleneck.

The outcome? It can perform a quadrillion operations every second at low power. Imagine, it's a sort of running a supercomputer on the power of a night light. It's manufactured using conventional silicon photonic technology, nothing exotic, at 22 nanometers at GlobalFoundries, the chipmaker that some years ago span out of AMD.

The light part and the electronic part are stacked together vertically, kind of like the light matter chip I showed you before. And if you missed that episode, subscribe to the channel, and I will leave all the links in the description box below for you to catch up. Why this chip is so special? To understand this, you have to read this paper (The second attachment Leo will share) which is honestly not an easy read.

The best investment of my life is studying this stuff at university. What is special about this chip? That with light on a chip you can do things you can never do with electronics. And the genius of light is that it has many different colors, and you can encode data in many different colors at the same time, and process it all in parallel.

This new chip can already process data at 32 different colors of light in parallel, and it can be scaled to more. Here is how it works. Let's take an example with four colors.

We can encode one number into purple light, another into blue, the third into green, and another into red. That means four different numbers are written into four different colors of light, and we can compute all of them at once on just one device, this white ring. Basically, it allows us to process a lot of information using just one device.

In electronics, you would need a separate device for each number. That makes a big difference. Now, why do we care? Because the core math behind all the AI workloads is running billions of times per second, every second, every single day.

And it's done now on traditional chips, which takes one device per operation. Just imagine, it's like airport security. Passengers go through the scanner one at a time.

But on a light-based chip, it's like scanning the entire crowd all at once. On a GPU, this kind of operation takes roughly 1,000 steps, depending on the size of cores, for example. But on a photonic chip, we can do it all at once.

That means computing with light enables us to compute up to 1,000 times faster. Well, that's a big promise. But what does it enable? These chips are not built to replace the processor in your laptop, but they can do something normal chips can't, handle massive amounts of data ultra-fast with very little energy.

That makes them perfect for workloads like the 3 billion daily requests behind ChargerBT, which today burns more energy than entire stages. The reason why all the hyperscalers like Google, Microsoft, Amazon, everyone is building custom chips, because they all have massive AI workloads and looking for a cheaper way to run at its scale. As of today, saving energy isn't just about saving the Earth's resources.

It's actually a huge part of the operating costs of a data center. So if a light-based computer can deliver the same performance using at least a little bit less energy, everyone will want to buy it. But we see that the promise here is in fact much bigger.

Computing with light could speed up way more things than just AI inference workloads. For example, scientific simulations, which often take many months to finish. But before you get too excited, there's a catch.

Or three. The first one is scalability. This is a big one, because photonic components are much larger than electronic transistors, so you can't pack as many of them onto a chip, otherwise the chip will become huge.

Still, scaling it to real AI models like GPT-5 remains a big challenge. The second challenge is the material itself. And this is where it starts to get really tricky.

This special memory wears out if you use it too much. Like a battery that lasts just so many charges. Thousands of cycles are fine for demos, but data centers need billions of cycles.

And this part isn't solved yet. And then, as with any new technology, there's a challenge of integration with existing computers and systems that we're already using today. Everything today – CPUs, GPUs, the software stack, entire data centers – is built for electronic processors.

And light-based accelerators need new interfaces and rewriting the software, which takes time to build. This work shows that the potential is real. The question is who will make it real first, and who will make it to work at scale.

Here, competition is heating up. Startups like Lightmatter, Light Intelligence, Quant, and others are all racing to deliver.