

Science Will Win | Season 4
Part 2 – How Video Games Helped Pave the Way for AI in Medicine
Episode 2 Script | FINAL

JEREMIAH:

We're going to start this episode at an unexpected place: The movies.

Imagine, it's the early 1990s. You're headed to the movies with your best friend, decked out in matching windbreakers. A portable tape player weighs down your left pocket. You sit down in the creaky theatre seat — you're there to watch the new blockbuster that came out that weekend. The lights dim, and the theatre is quiet except for some rustling here and there, some popcorn crunching.

The screen comes to life — a giant velociraptor darts across, looking as realistic as any that you've ever seen.

Now, why am I bringing you into a dark theatre? That's certainly not where new drugs are discovered. But the technology that made that velociraptor come to life was a precursor to another piece of technology that has helped the artificial intelligence revolution take root in drug discovery.

Because the hardware to create a realistic looking velociraptor, for example, was so bulky and expensive, it was something mainly only done for big-budget films. But over time, there was an increased appetite for high-definition, realistic, 3D graphics in other mediums as well.

TOR:

These GPUs started, uh, showing up in gaming consoles as well.

JEREMIAH:

That's Tor Aamodt. He's a professor at the University of British Columbia, where his research focuses on the architecture of GPUs — graphics processing units — and machine learning.

When developers first made the GPUs as we know them today in the late 1990s, they had one express purpose:

TOR:

They were there to accelerate graphics primarily for, for real time rendering for video games. Um, so you could interact with the game and it be more realistic

JEREMIAH:

But the same technology that could make it feel like you're really there in a scene — driving a car, running away from a villain, exploring a new land — can also help scientists and researchers more efficiently and effectively create models, run experiments, and analyze data.

This episode, we're tracing the hardware evolution from its theoretical and entertainment origins to its future potential in accelerating drug discovery.

[theme music starts]

JEREMIAH:

Welcome to Science Will Win. I'm your host, Jeremiah Owyang. I'm an entrepreneur, investor, and tech industry analyst. I'm passionate about emerging technologies and the ways they can shape our world.

Last episode, we talked about data, and how large swaths of data are crucial in assisting drug development. Well, more data requires A LOT of storage. To make use of big data and all its possibilities, scientists need new tools at their disposal.

Today we're focusing on the breakout stars in hardware that have had huge implications for the advancement of medicine and AI: CPUs and GPUs.

[theme music fades out]

When you're sick, you probably look for a medicine that might make you feel better. You might swallow a pill with a gulp of water, allowing it to travel down into your stomach, and dissolve.

Every pill you take is a bundle of carefully selected and designed chemicals. When some pills dissolve, they release those chemicals, which bind to receptors on cells in your body. That then triggers a response inside the cell.

And increasingly, artificial intelligence is helping researchers identify and test new potential drugs or therapies.

ENOCH:

My name is Enoch Huang. I've been with Pfizer, it's my 25th year, and I lead a group called Machine Learning and Computational Sciences

JEREMIAH:

Enoch works on the chemistry side of drug development and discovery. When his team is testing a potential new drug, there are a few different properties that it must have.¹ One of them is high affinity, which means a strong attraction between the molecules in the drug and the molecules in the human cell that it's trying to affect.

High affinity is good! Scientists want the molecules they're designing to have high affinity. It means the interactions are strong. When designing molecules with high affinity, researchers need to be aware of molecular strain. If a molecule is bonded to another at an odd angle, its internal energy increases and becomes unstable. Too much molecular strain is bad.

ENOCH:

At the risk of anthropomorphizing too much, like a molecule, to fit in the binding pocket, gets unhappy if it has to adopt the conformation where it feels strained.

JEREMIAH:

There are thousands of ways for any given molecule to attach to another molecule. Think of a car looking for parking in a busy downtown area: there's parallel parking on the street, lots designated for specific businesses, and pay-by-the-hour garages. So many options, each with their own downsides, or levels of strain.

Enoch and his team were running an experiment to calculate conformational strain on a particular molecule, or the level of instability in a molecule when parked in particular spots.

ENOCH:

It turns out that conformational strain is something that is difficult to predict, unless you do an expensive calculation using quantum mechanics. And so to do this requires a lot of computational horsepower and, and expertise. And so that's why if you can avoid doing the expensive calculation, um, it speeds up the discovery process. So what the team did was say, okay, what if we used a massively parallel computational system?

JEREMIAH:

A massively parallel computation system is basically any computer that can do multiple calculations at once, rather than one at a time. Like Enoch said, quantum mechanical, or QM,

¹ <https://pmc.ncbi.nlm.nih.gov/articles/PMC8654478/>

calculations are *expensive* and take a *long* time. Every one of those potential configurations of molecule attaching to a spot on another molecule requires a separate calculation. With a little help from parallel computing, and artificial intelligence, Enoch and his team were able to cut down on that time drastically.

ENOCH:

Actually many technologies came together.

ENOCH:

My team had, again, a moment of inspiration saying, you know, why don't we build machine learning models on QM calculations, use the QM calculations as ground truth, and becomes the training set for machine learning models so that you don't have to do the QM calculation.

JEREMIAH:

By training artificial intelligence on QM data, Enoch and his team were able to predict the molecular strain of different molecule positions. Then based on those results, they could move onto the next step of development and test the molecules with the lowest strain, and discard the rest.

ENOCH:

And the benefit there is before you actually make and test a molecule, you have a pretty good estimate of its binding affinity. So that way you can decide which of the molecules you actually wanna ~~make a~~ test because you have a, you know, um, a good prediction of, of its eventual affinity. So in the vein of even the model's not perfect, if it's generally predictive, it allows you to do your experiments more productively than just brute force trying every single possibility

ENOCH:

So that was a wonderful innovation because it brought together cloud computing, machine learning, and quantum chemistry.

ENOCH:

20 years ago, we wouldn't have attempted the experiment...Even if you did crank away, because we're talking about millions of compounds, you'd be waiting forever for that experiment to finish.

JEREMIAH:

So, before even making a molecule, Enoch and his colleagues were able to predict its level of strain, thanks to a wealth of QM data and strong enough computational power to sift through all of it — and fast. This is just one way the drug discovery and development process has become more efficient thanks to innovations in hardware. But it wasn't always this way.

[music transition]

JEREMIAH:

When you're using the Internet on your smart phone or laptop, it feels like you have infinite access to as much information as you'd like, magically stored in a slim piece of metal, glass, and plastic. You might picture a cloud hanging above you, simmering with electric energy, passing weightless information like puffs of vapor through the sky.

The digital and the physical seem like two different things. But everything digital needs to be physically stored somewhere. Behind every search engine query or social media post or online shopping listing or function put into a spreadsheet is a tangible piece of hardware that makes it possible.

There are seven major components of hardware that make up a computer: central processing units (CPUs), random access memory (RAM), motherboards, computer data storage, graphics cards, sound cards, and computer cases. We're focusing on two of those pieces of hardware today.

A CPU, or central processing unit, is the brain of a computer. It's responsible for a wide range of essential tasks such as web browsing or running software.²

A *GPU*, or graphics processing unit, is an electronic circuit board that belongs inside graphics cards. You could hold one in the palm of your hand, though you'll usually find it slotted into a computer's motherboard. GPUs perform rapid mathematical calculations based on data and turn that data into visual information. Here's Tor again.

TOR:

these, uh, devices have billions of transistors in them, and lots of things are going on at the same time. Uh, a lot of it is just calculations, arithmetic calculations, like adding

² https://www.splunk.com/en_us/blog/learn/cpu-vs-gpu.html#:~:text=CPU%20manages%20general%20computing%20tasks,enhancing%20the%20system's%20overall%20performance.

numbers, multiplying numbers. And then, uh, moving that data around is, is primarily what's actually happening.

JEREMIAH:

Before GPUs came onto the market in 1999, graphics cards relied on CPUs to process data into visuals.³ But, while CPUs are designed to work quickly, they aren't great at multi-tasking, and can only do things in sequence. A CPU can typically work on 4-12 threads, or processes, at a time. A GPU, however, can work on *thousands* of threads at a time.⁴

TOR:

inside, you know, the way you, uh, program the device is you have software and it's describing the operation of what we call threads, which are just, uh, think of a, think of a program as like a recipe. So each, um, thread is like a individual recipe. inside a graphics processor you're, you're basically trying to, uh, cook up a bunch of things at the same time.

JEREMIAH:

Once GPUs came onto the market, computers had the ability to cook up many recipes at once. Data was processed much more quickly. As a consequence, video games transformed. With just an additional circuit board installed inside your personal computer or gaming console, landscapes, objects, and characters were rendered faster and with more detail than ever before. Games went from boxy, pixelated characters moving around along on 2D planes to 3D cinematic experiences.

At this point in the early 2000s, GPUs were fix-functioned, meaning their processes were pre-defined and could not be customized by a programmer. But gaming companies wanted them to be more programmable and customizable.

TOR:

the idea was that the games, uh, would have different lighting effects. So maybe you want to like a special fog or something, ~~you know,~~ in the scene, in your video game if you can write that up as software, then you can in, in that sense, you can configure the hardware you can think of it as for every little pixel on your screen, there's a little program that's running to generate the color for that pixel is what, what's happening Um, and so, uh, that's the kind of the flexibility that the, the game companies wanted. And so

³ <https://aws.amazon.com/what-is/gpu/#:~:text=Origin%20of%20the%20GPU,all%20on%20a%20programmable%20chip.>

⁴ <https://www.sciencedirect.com/science/article/abs/pii/S1361841513000820>

now that you've got that flexibility and the hardware, it's like, well, what else can I do with that?

JEREMIAH:

Soon, computer scientists started taking interest in what else the high power of GPUs could do outside of rendering video game visuals. And in 2007, the first general purpose GPU was released. It turns out, rather than making really realistic images, GPUs could also help computers complete tasks much more efficiently.

TOR:

there were a lot of people in academia who saw the potential of this. it's giving them this, new platform that they can use to explore And try to make things faster.

TOR:

And so a lot of universities started to run courses. And at some point it was like hundreds of courses where we're teaching this how to use the GPU for something ~~o~~ other than other than graphics.

TOR:

and then things started to evolve. So today people are using them for things that they're not really graphics, um, still called GPUs but like for machine learning, for example, they're running, uh, something where they're getting a bunch of numbers out of it, really. And, and so, ~~uh~~, so it's become something much more than just the graphics

JEREMIAH:

Programmers were interested in building GPUs that could enhance CPUs and cook up many recipes at once. But the companies who build and sell hardware were skeptical that GPUs were a worthwhile investment.

TOR:

People like to use the phrase, um, you know, what's the killer app for GPU? Um, and, and there was a lot of more traditional computer systems folks who are questioning whether the GPU would be good at anything important.

TOR:

you'd hear financial analysts asking the How much money is that earning you? And, um, and the answer was like, if we want to grow beyond graphics, we have to, to do these

things. And so initially, like when I started doing research um, looking into applications, this was a question, so what, what is this good for?

JEREMIAH:

Initially, as Tor tells it, there weren't very many uses for general purpose GPUs.

TOR:

oil and gas companies in, in Alberta, they would take, you know, seismic data and try to figure out if there's oil.

JEREMIAH:

But there was still doubt among computer scientists and those invested in the hardware business about how far GPUs could go. What was the killer application? If supercomputers or oil and gas exploration were low-budget indie films, GPUs needed a blockbuster.

TOR:

there's a, um, a famous sort of, law, we call it a law called Amdahl's Law, which basically says, when you take a, uh, what, what most people might call a regular program where it does one thing after the other, and you try to run it in, in what we call in parallel on multiple different cores, how much of a improvement you could get.

JEREMIAH:

Basically, when a part of a system is improved, the overall system improvement will be proportional to how much that part makes up of the system. Naysayers claimed that Amdahl's Law applied to GPUs. If a GPU — which is only ONE tiny part of a graphics card, which is only one tiny part of a computer — is improved, the GPU's potential to improve the computer overall is limited.

TOR:

*and they would always point to this law and say, say, well, this limits the upside, so much. Then why would you wanna do this? so the answer is that if you write new applications that just have lots of parallelism, uh, and, and it turns out there's this super important one called **machine learning**.*

[music transition for effect]

JEREMIAH:

There it is. The killer application. Artificial intelligence has to be trained on massive amounts of data, which requires hefty hardware to process. GPUs, and their ability to run multiple processes at once, makes them primed to handle the task. According to Tor, the moment when the scientific community came to understand GPUs potential to assist with machine learning came in 2012.

TOR:

There's this, uh, paper, it was at, it was at this conference in 2012. Um, uh, people refer to it as Alex net because that's, uh, one of the author, first author on the paper, Alex Krizhevsky

JEREMIAH:

Alex and his colleagues were trying to train a computer to look at images in a database known as ImageNet, and categorize them. Could it take an image of a dog and identify it as a dog, or take an image of a cat and identify it as a cat?⁵ This team of scientists found a very effective way to do it.

TOR:

And a key part of their solution happened to be using a, a graphics processor unit. And, um, and that it improved the accuracy on this problem by, you know, about, I think, uh, roughly 10 times as much as the prior year increment. So it really got people's attention it was getting to the point where people could see machine learning and neural network not just as a novelty thing, but a thing that would actually be useful to people. And so then you see a lot of, uh, uh, investment starting to happen.

JEREMIAH:

Which brings us back to drug discovery and development.

On average, it takes more than 12 years to bring a new drug to market.⁶ One of the factors that makes this process so time consuming is parsing through and making sense of large amounts of data, whether it be imaging, genomics, or molecular data. Data can be empowering, but only if you actually have the time to look through all of it, and draw connections. GPUs help make this data analysis possible — and more efficient. So it makes sense that pharmaceutical companies are seeking out GPUs to add to their stock of hardware.

⁵ <https://www.mathworks.com/help/deeplearning/ref/alexnet.html>

⁶ <https://www.forbes.com/sites/nvidia/2018/07/27/how-gpus-are-impacting-drug-research-and-development-for-pharma/>

DANIEL:

So there are only very few companies in the world that gen that, that, um, manufacture GPUs and so the price of GPUs, even though it's such a mass market, um, is pretty high right now. So there is this scarcity of GPUs, so it's actually not easy to even buy them. And because so many companies, not only pharmaceutical companies, um, want to buy these GPUs, it's really hard to obtain them.

JEREMIAH:

That's Daniel Ziemek. He is Vice President of Integrative Biology and Systems Immunology at Pfizer. He oversees teams of computational biologists who work on creating new medicines. He talked about the current boom in using GPUs in drug discovery and development.

DANIEL:

basically any deep learning you wanna do, you need GPUs or you'll wait for years to get your results. And where you would only wait, say seconds for A GPU.

DANIEL:

these days, oftentimes even laptops have GPUs. But the larger scale you go, you need these bigger GPUs that are really hard to get and that live in these big server farms, um, that, that are expensive to maintain. Uh, but you do need them very quickly if you wanna do any of these things at scale.

JEREMIAH

One area in which artificial intelligence and GPUs can be helpful is spatial omics. Last episode, we talked about omics: the umbrella term for different types of biological study. *Spatial* omics refers to biological technology that can examine molecules within the context of where they are located within a tissue.⁷

DANIEL:

And so what the spatial omics allows you to do is understand which cells are close to each other and which structures they form in the tissue. And that's yet another layer of information, uh, that can help to understand how, what goes wrong in disease and how can you try to intervene to fix it. Why that is maybe important in the context of AI is that, a lot of the AI innovation came from the field of image recognition, right?

DANIEL:

⁷ <https://www.mosaic-research.com/spatial-omics#:~:text=Spatial%20omics%20is%20a%20broad,native%20location%20within%20a%20tissue.>

Again, the famous example of Googling show me cat images, which was completely impossible a while ago, by now, it's probably 20 years ago, but before that, it was not possible at all. And then with these deep learning algorithms, this became possible.

JEREMIAH:

But there are still challenges to these new innovations. As Daniel mentioned, GPUs aren't always available. That brings us back to our story from the beginning of the episode. Pfizer's Enoch Huang and his team were performing an experiment. It was meant to predict molecular strain in different molecules that could potentially be turned into new medicines.

We've spent a lot of time talking about GPUs, but CPUs can be just as useful when used in the right way. Enoch and his team were able to replicate the way that a GPU can perform a number of calculations at the same time. They used cloud computing to stack multiple CPUs on top of each other. Here's Enoch again.

ENOCH:

each molecule is a separate calculation

ENOCH:

by having more CPUs, you can each give a separate task

JEREMIAH:

So the CPUs were cooking up multiple recipes, as Tor called them, at once. This was possible due to advances in hardware across the board at Pfizer. In the past, it wouldn't have been possible to use ~~up~~ that much computing power for an experiment like that. And if they did, they would have been waiting a *long* time for the computer to finish calculating molecular strain for each molecule, one after another.

ENOCH:

20 years ago you had maybe four processors or 12 processors CPUs, and they were probably doing other things that were business critical. So you couldn't say, I'm gonna take these 12 and crank away.

JEREMIAH:

Enoch's experiment is just one example of the MASS of potential that innovations in hardware have to assist in drug discovery and development.

Enoch:

now it's a resource that our, our teams can benefit on, it's an effort that you make not to help one individual project move faster, but all projects that potentially would face this issue down the road.

[outro music starts]

Next time on Science Will Win:

With a greater-than-ever ability to process data quickly, new applications for AI are on the horizon. And behind every piece of generative AI is software.

BETH:

So far we're in this golden age where what most, what mostly people have been able to use deep learning for really well is automate the stuff that was tedious and painful and nobody really wanted to do it and it was critical to get right, and get to actually thinking about science.

JEREMIAH:

Science Will Win is created by Pfizer and hosted by me, Jeremiah Owyang. It's produced by Wonder Media Network. Please take a minute to rate, review and follow Science Will Win wherever you get your podcasts. It helps new listeners to find the show.

Special thanks to all of our guests and the Pfizer research and development teams. And thank you for listening!