BACKCHANNEL

GEAR

SCIENCE

CADE METZ 05.19.16 05:21 PM

BUSINESS

Google's Making Its Own Chips Now. Time for Intel to Freak Out

CULTURE



custom-designed chip helps run what is surely the future of its vast online empire: artificial intelligence.

In building its own chip, Google has taken yet another step along a path that has already remade the tech industry in enormous ways. Over the past decade, the company has designed all sorts of new

The Internet's most powerful company sent a few shock waves

through the tech world yesterday when it revealed that a new

hardware for the massive data centers that underpin its myriad online services, including computer servers, networking gear, and more. As it created services of unprecedented scope and size, it needed a more efficient breed of hardware to run these services.

Over the years, so many other Internet giants have followed suit,

forcing a seismic shift in the worldwide hardware market.

With its new chip, Google's aim is the same: unprecedented efficiency. To take AI to new heights, it needs a chip that can do more in less time while consuming less power. But the effect of this chip extends well beyond the Google empire. It threatens the future of commercial chip makers like

Intel and nVidia—particularly when you consider Google's vision for

the future. According to Urs Hölzle, the man most responsible for the

No, Google will not sell its chips to other companies. It won't directly

compete with Intel or nVidia. But with its massive data centers,

Google is by far the largest potential customer for both of those

companies. At the same time, <u>as more and more businesses adopt</u>

the cloud computing services offered by Google, they'll be buying

fewer and fewer servers (and thus chips) of their own, eating even

global data center network that underpins the Google empire, this

new custom chip is just the first of many.

that no other company has.

Google's Need for Speed

built a chip that's even more efficient.

further into the chip market. Indeed, Google revealed its new chip as a way of promoting the cloud services that let businesses and coders tap into its AI engines and build them into their own applications. As Google tries to sell other companies on the power of its AI, it's claiming—in rather loud ways—that it boasts the best hardware for running this AI, hardware

Google's new chip is called the Tensor Processing Unit, or TPU.

That's because it helps run TensorFlow, the software engine that

data. Other tech giants typically run their deep neural nets with

graphics processing units, or GPUs—chips that were originally

applications. These are well-suited to running the types of

designed to render images for games and other graphics-heavy

calculations that drive deep neural networks. But Google says it has

drives the Google's deep neural networks, networks of hardware and

software that can learn particular tasks by analyzing vast amounts of

According to Google, it tailored the TPU specifically to machine learning so that it needs fewer transistors to run each operation. That means it can squeeze more operations into the chip with each passing second. For now, Google is using both TPUs and GPUs to run its neural

nets. Hölzle declined to go into

specifics on how exactly Google

was using its TPUs, except to say

that they handle "part of the

computation" needed to drive

voice recognition on Android

would be releasing a paper

describing the benefits of its

chip and that Google will

phones. But he said that Google

continue to design new chips that handle machine learning in other ways. Eventually, it seems, this will push GPUs out of the equation. "They're already going away a little," Hölzle says. "The GPU is too general for machine learning. It wasn't actually built for that." That's not something nVidia wants to hear. As the world's primary seller of GPUs, nVidia is now pushing to expand its own business into the AI realm. As Hölzle points out, the latest nVidia GPU offers a mode specifically for machine learning. But clearly, Google wants the change to happen faster. Much faster. **The Smartest Chip** In the meantime, other companies, most notably Microsoft, are exploring another breed of chip. The field-programmable gate array, or FPGA, is a chip you can re-program to perform specific tasks.

is "overkill," pointing out that such a chip takes at least six months to build—a long time in the incredibly competitive marketplace in which the biggest Internet companies compete.

anyone can use and modify.

RELATED VIDEO

an FPGA, Hölzle said: "It's just much faster."

FPGAs.

Core Business Hölzle also points out that Google's chip doesn't replace CPUs, the central processing units at the heart of every computer server. The search giant still needs these chips to run the tens of thousands of machines in its data centers, and CPUs are Intel's main business. Still, if Google is willing to build its own chips just for AI, you have to

wonder if it would go so far as to design its own CPUs as well.

are not solved," he says. In other words, CPUs are a mature

Google wants healthy competition in the chip market. In other

all, more competition means lower prices for Google. As Hölzle

OpenPower Foundation, which seeks to offer chip designs that

That's a powerful idea, and a potentially powerful threat to the

world's biggest chip makers. According to Shane Rau, an analyst with

research firm IDC, Google buys about 5 percent of all server CPUs

explains, expanding its options is why Google is working with the

Hölzle plays down the possibility. "You want to solve problems that

technology that pretty much works as it should. But he also said that

words, it wants to buy from many sellers—not just, say, Intel. After

Microsoft has tested FPGAs with machine learning, and Intel, seeing

where this market was going, recently acquired a company that sells

Some analysts think that's the smarter way to go. An FPGA provides

principal analyst at Moor Insights and Strategy, a firm that closely

follows the chip business. Moorhead wonders if the new Google TPU

But Google doesn't want that flexibility. More than anything, it wants

speed. Asked why Google built its chip from scratch rather than using

far more flexibility, says Patrick Moorhead, the president and

sold on Earth. Over a recent year-long period, he says, Google bought about 1.2 million chips. And most of those likely came from Intel. (In 2012, Intel exec Diane Bryant told WIRED that Google bought more server chips from Intel than all but five other companies—and those were all companies that **sell** servers.) Whatever its plans for the CPU, Google will continue to explore chips specifically suited to machine learning. It will be several years before we really know what works and what doesn't. After all, neural networks are constantly evolving as well. "We're learning all the time," he says. "It's not clear to me what the final answer is." And as it

TECHNOLOGY

it's innovations

Here's Everything New From Google

Google made several announcements at its annual developers conference. As

expected the tech giant's progress with Artificial Intelligence is at the core of many

OPINION **Trump's WeChat Ban May Alienate His Chinese Supporters** RONG XIAOQING

These Are a Few of Our **Favorite Things (for Under \$20)** ADRIENNE SO

CREATURE COMFORTS

HAPPY FEAT

DAMIAN CARRINGTON

Get unlimited WIRED access Subscribe

A Satellite Spots 11 New

Emperor Penguin Colonies

How to Use Android's Intel Is Investigating How Confidential Data Ended Up Online

The NSA's Tips to Keep Your Phone From Tracking You

LEAKED

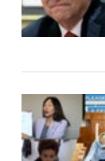
WIRED

SECURITY ROUNDUP

BRIAN BARRETT

Bill Gates on Covid: Most **US Tests Are 'Completely** Garbage'

Most Popular



IDEAS

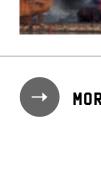
STEVEN LEVY



Hybrid Schooling May Be the Most Dangerous Option of All DAVID ZWEIG



BACKCHANNEL The Tragic Physics of the **Deadly Explosion in Beirut** RACHEL LANCE



MORE STORIES

learns, you can bet that the world's chip makers will be watching.

#AI #ARTIFICIAL INTELLIGENCE #ENTERPRISE #GOOGLE #INTEL #MICROCHIPS #ARTIFICIAL INTELLIGENCE

Nearby Share to Send Photos and Files JULIAN CHOKKATTU

DROID DROP

JIM SALTER, ARS TECHNICA AND DAN GOODIN, ARS TECHNICA

Sign In EXPAND ^