

Deep learning compute system is billed as world's fastest

November 21 2019, by Nancy Cohen



A company called Cerebras offers up some dizzying superlatives that need some time to sift through. The world's biggest chip? And 56x larger than any other chip? At the heart of the fastest AI supercomputer in the world?

The computer is called the CS-1. The [chip](#)—actually, an entire wafer, is called the WSE (stands for wafer-scale engine).

Back in August, the WSE was introduced. "Normally, [silicon chips](#) are

carved out of processed 12-inch silicon wafers," [said](#) Dean Takahashi in *VentureBeat*, with many hundreds of chips on a wafer." Cerebras did something different—lots of little cores, repeated across an entire wafer sawed into a big rectangle.

Takahashi shared eye-catching numbers for how big is big, with more than 1.2 trillion transistors across all of the cores on one wafer, compared with a typical processor with 10 billion transistors on one chip.

The company now has a dramatic second chapter to its story; with the WSE news in August, the news that follows is the debut of the entire system, called CS-1. The CS-1 houses the WSE.

"You can't take a Ferrari engine and put it in a Volkswagen to get Ferrari performance," said Andrew Feldman, CEO, in his exchange with *VentureBeat*. "What you do is you move the bottlenecks if you want to get a 1,000 times performance gain."

So, now there is a CS-1 "powered by a 400,000-core, 1-trillion-transistor wafer-scale processor chip," [wrote](#) Samuel Moore in *IEEE Spectrum*.

The company said that, the CS-1 packs the performance of a room full of servers into a single unit the size of a dorm room mini-fridge. *ZDNet* [explained](#) it as a chassis measuring 15 standard rack units high (a little over two feet, and a foot and a half wide and three feet deep.)

"Given the performance, that's an incredibly compact package," [commented](#) Paul Alcorn in *Tom's Hardware*, and one might consider that "a single Cerebras chip has 78 times more cores, 3,000 times more memory, and 10,000 times more memory bandwidth than a single GPU. It also has 33,000 times more bandwidth (PB/s)."

Andrew Feldman said that "Depending on workload, the CS-1 delivers hundreds or thousands of times the performance of legacy alternatives at one-tenth the power draw and one-tenth the space per unit compute."

There are heat challenges in such a construct. Cerebras has developed an elaborate network of pipes to carry water to effectively irrigate the WSE, said Nick Farrell in [Fudzilla](#).

The company said the CS-1 is an internally water-cooled system. "Like a giant gaming PC, the CS-1 uses water to cool the WSE, and then uses air to cool the water. Water circulates through a closed loop internal to the system."

Likely customers? Cerebras describes itself as a company "dedicated to accelerating deep learning." That is your answer. Supercomputer buyers.

Argonne National Laboratory (ANL), one of the nation's premier research centers, is the first customer to deploy the Cerebras CS-1 system. Reports said that the CS-1 was being used for such aims as understanding cancer drug response rates and traumatic brain injury.

A news [release](#) said that "By pairing supercompute power with the CS-1's AI processing capabilities, Argonne can now accelerate research and development of deep learning models to solve science problems not achievable with existing systems."

More information: www.cerebras.net/introducing-tellintelligence-computer/

www.cerebras.net/

Citation: Deep learning compute system is billed as world's fastest (2019, November 21)
retrieved 3 May 2024 from <https://techxplore.com/news/2019-11-deep-billed-world-fastest.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.