Wafer-Scale Engine: The Largest Chip Ever Built

The Wafer-Scale Engine (WSE-2), which powers the Cerebras CS-2 system, is the largest chip ever built. The WSE-2 is 56 times larger than the largest GPU, has 123 times more compute cores, and 1000 times more high-performance on-chip memory. The only wafer scale processor ever produced, it contains 2.6 trillion transistors, 850,000 AI-optimized cores, and 40 gigabytes of high performance on-wafer memory all at accelerating your AI work.

Cerebras Wafer-Scale Engine

Fabrication process
7nm

Silicon area
46,225mm²

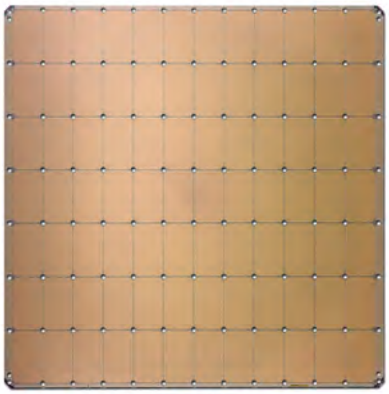
Transistors
2.6 Trillion

AI-optimized cores
850,000

Memory (on-chip)
40GB

Memory bandwidth
20PB/s

Fabric bandwidth
220Pb/s



Cerebras WSE-2
2.6 Trillion Transistors
46,225 mm² Silicon

Largest GPU
54.2 Billion Transistors
826 mm² Silicon

Compute Designed for AI

Each core on the WSE-2 is independently programmable and optimized for the tensor-based, sparse linear algebra operations that underpin neural network training and inference for deep learning. The WSE-2 empowers teams to train and run AI models at unprecedented speed and scale, without the complex distributed programming techniques required to use a GPU cluster.

Cluster-Scale in a Single Chip

Unlike traditional devices with tiny amounts of on-chip cache memory and limited communication bandwidth, the WSE-2 features 40GB of on-chip SRAM, spread evenly across the entire surface of the chip, providing every core with single-clock-cycle access to fast memory at an extremely high bandwidth of 20PB/s. This is 1,000x more capacity and 9,800x greater bandwidth than the leading GPU.

High Bandwidth, Low Latency

The WSE-2 on-wafer interconnect eliminates the communication slowdown and inefficiencies of connecting hundreds of small devices via wires and cables. It delivers an astonishing 220 Pb/s interconnect bandwidth between cores. That's more than 45,000x the bandwidth delivered between graphics processors. The result is faster, more efficient execution for your deep learning work at a fraction of the power draw of traditional GPU clusters..

For more information about the Cerebras CS-2 system click here.

CEREBRAS SYSTEMS, INC.
1237 E. ARQUES AVE, SUNNYVALE, CA 94085 USA
CEREBRAS.NET

