PSC

PSC Neocortex

Neocortex is an innovative NSF-funded computing resource that will accelerate time-to-science by orders or magnitude by vastly shortening the time required for deep learning training. The system will do this by exploring a revolutionary combination of Cerebras Wafer Scale Engine (WSE) processors, which are designed specifically to accelerate AI, and an extremely large-memory HPE Superdome Flex system for massive data handling capability. This balanced system will democratize access for researchers to game-changing compute power for training, the most time-consuming step of AI, to be much faster, even interactive.

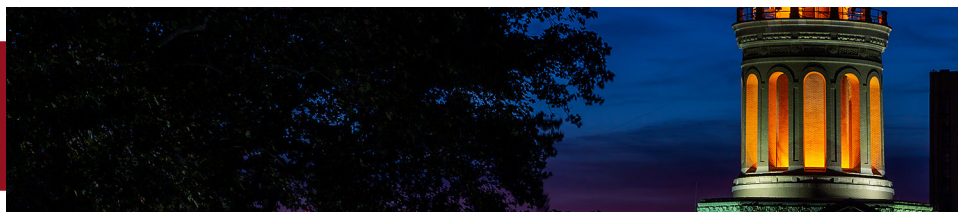
How is Neocortex unique?

Neocortex is the first of its kind. It will be the first open science system to introduce CS-1 servers and the Wafer Scale Engine processor developed by Cerebras Systems. Neocortex will explore scaling to more than one of them. This system is expected to offer unprecedented and game changing capabilities to researchers and engineers tackling the most challenging problems with complex AI models. The system's unique AI capabilities for advanced research and development, and its speed, will remove barriers to AI innovation, supporting the Executive Order on Maintaining American Leadership in Artificial Intelligence and the American Artificial Intelligence Initiative.

Access to Neocortex will start with the Early User Program. A representative group of applications in diverse domains including AI algorithms, bioinformatics, neurophysiology, materials science, and COVID-19 drug discovery, will have the opportunity to benefit and test the impressive capabilities of the Neocortex system.

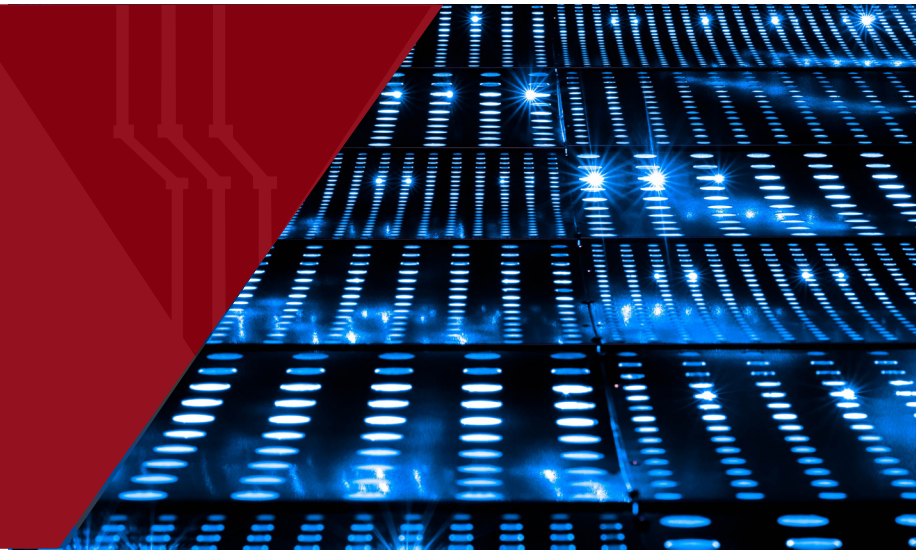
The system is designed to power and accelerate the most ambitious deep learning networks and algorithms including sequential models, graph neural networks, models with induced sparsity or with separable convolutions, and models where model parallelism is desirable.

Neocortex is funded by the NSF through award NSF *2005597* and is planned to be delivered late 2020.



PSC

PSC Neocortex



Architecture

The novel Neocortex architecture will transform deep learning powered research by coupling two exceptionally powerful Cerebras CS-1 AI servers with an extreme shared-memory HPE Superdome Flex server to achieve unprecedented AI training capability with excellent system balance.

Each ***Cerebras CS-1*** is powered by one Cerebras Wafer Scale Engine (WSE) processor, a revolutionary high-performance processor designed specifically to accelerate deep learning training and inferencing. The Cerebras WSE is the largest computer chip ever built, containing 400,000 AI-optimized cores implemented on a 46,225 square millimeter wafer with 1.2 trillion transistors, compared to only billions of transistors in high end CPUs and GPUs.

The Cerebras CS-1 server is the first available system featuring the Cerebras Wafer Scale Engine (WSE) processor which is the largest chip ever built.

Neocortex will use the ***HPE Superdome Flex***, an extremely powerful, user-friendly front-end high-performance computing (HPC) solution for the Cerebras CS-1 servers. This will enable flexible pre- and post-processing of data flowing in and out of the attached WSEs, preventing bottlenecks and taking full advantage of the WSE capability. HPE Superdome Flex will be robustly provisioned with 24 terabytes of memory, 205 terabytes of high-performance flash storage, 32 powerful Intel Xeon CPUs, and 24 network interface cards for 1.2 terabits per second of data bandwidth to each Cerebras CS-1.

A precisely provisioned HPE Superdome Flex server will be the front-end to the two CS-1 systems.

Neocortex will be integrated with Bridges-2 which will yield great benefits to the user community including: 1) access to the Bridges-2 filesystem for management of persistent data; 2) general-purpose computing for data preprocessing and complementary, traditional machine learning; 3) interoperation with data-intensive projects that will already be using Bridges-2; and 4) high-bandwidth external network connectivity to other service providers, campus, labs and clouds.

For more information and to find out how you can leverage Neocortex for your research needs visit: https://www.cmu.edu/psc/aibd/neocortex/index.html