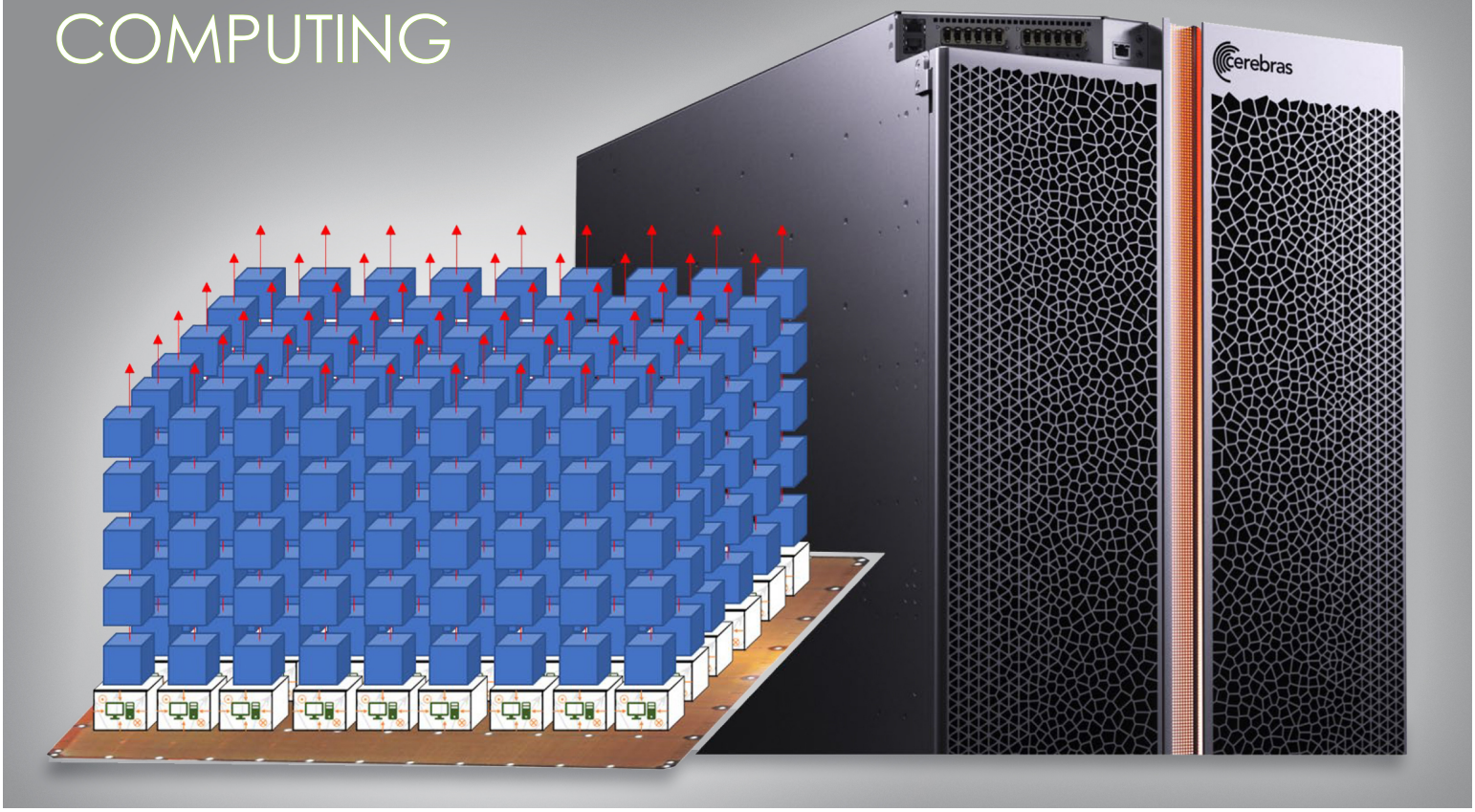
THE WAFER-SCALE ENGINE FIELD EQUATION APPLICATION PROGRAMMING INTERFACE: HIGH-SPEED SOLUTIONS FOR NET-ZERO COMPUTING



NETL

NATIONAL ENERGY TECHNOLOGY LABORATORY

BACKGROUND

Computers have become faster, smaller and more efficient every year, but this evolution is reaching its inevitable conclusion due to fundamental physical limitations of silicon computer chips. Complex computations, such as those needed to model energy systems, traditionally require large high-performance computing (HPC) systems comprising thousands of processing nodes that are each roughly the size of a desktop computer. Consequently, these modern supercomputers are roughly the size of a house, and the physical distance between the computing components significantly reduces achievable bandwidth and increases latency.

An artificial intelligence hardware manufacturer, Cerebras Systems, has created the largest single computer chip to date — the Wafer-Scale Engine (WSE) — which has essentially taken the power of a supercomputer and reduced it down to a single chip. The equivalent of each computational node has been shrunk down to just a couple hundred microns in size. The significant reduction in size allows maximum sustained bandwidths in silicon while latencies

remain single cycle. The distance in traditional HPC significantly limits achievable bandwidths and latencies. NETL has been working to tap into this transformative computing power by developing a simple user interface for solving field equations. This work has the potential to revolutionize computing while greatly reducing energy consumption and carbon emissions, which will help reach critical decarbonization goals.

INTRODUCTION

The most current WSE consists of 850,000 cores manufactured on a single 12-inch wafer with more than 2.6 trillion transistors. It is packaged in a system less than a meter high and as wide as a single rack that houses processor nodes in a supercomputer. The WSE requires less than 1% of the space and power of NETL's supercomputer Joule 2.0, yet it has comparable peak performance and much higher algorithmic performance.

"Tiles" on the WSE minimize the fundamental components of a traditional HPC node (e.g., processor, memory, etc.). Every tile can act independently with its own program code or as a group for collective operations. Tiles are a few hundred microns in size. This small size allows each processor to access its own memory within a single clock cycle and a neighboring processor within a single fabric hop (also single cycle).

Minimizing data movement energy is incredibly important, as scientific simulations move enormous amounts of data at these computational speeds. A prominent multinational digital communications corporation recently estimated total internet traffic to be approximately 0.7 petabytes per second. The total data movement on the WSE during simulations in NETL's WFA software run can be as much as 200 times the entire world's internet traffic volume.

PROGRAM DESCRIPTION AND BENEFITS

NETL's Wafer-Scale Engine Field Equation Application Programming Interface (WFA) unlocks the power of the WSE for scientific computing by providing a simple user interface, resulting in 470 times faster processing than traditional HPC resources such as the lab's Joule 2.0 supercomputer. This increase of speed translates into the ability to run a model that would normally take approximately one year in less than a day. Even more impressively, the WSE achieves these step-change increases while using far less energy, resulting in an estimated two to three orders of magnitude lower carbon emissions.

This paradigm shift in computing enabled by NETL's WFA has the potential to revolutionize much of the lab's critical decarbonization research. Currently, the research team is working to use the WFA to develop a finite volume computational fluid dynamics solver, which will help design future energy systems in far less time and at reduced monetary and energy costs. In the future, the WFA could branch out into areas such as geodynamics, geomechanics, molecular dynamics and structural dynamics.

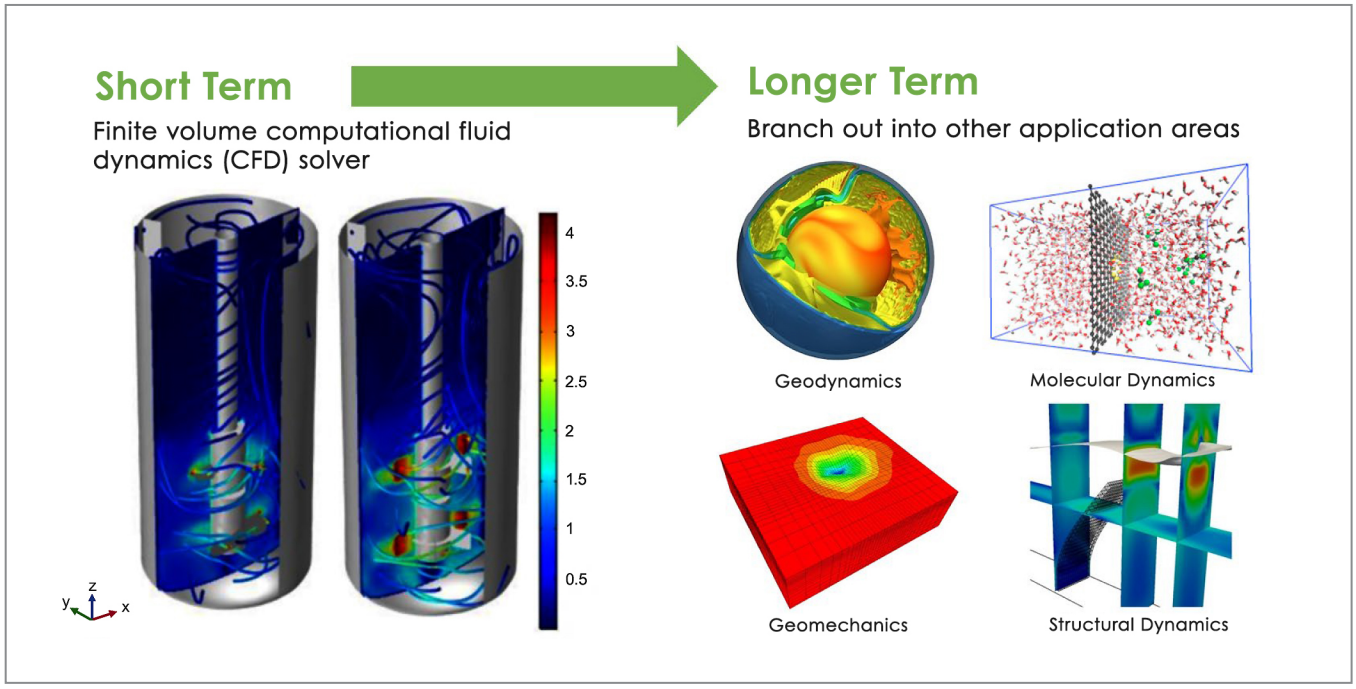
NETL and Cerebras have signed an agreement in which NETL is developing the WFA and Cerebras is running tests on their hardware cluster pictured here.

THE WAFER-SCALE ENGINE FIELD EQUATION APPLICATION PROGRAMMING INTERFACES: HIGH-SPEED SOLUTIONS FOR NET-ZERO COMPUTING

Research computing represents about half of NETL's total energy consumption. Moving even just 50% of the lab's projected research computing to the WSE could lead to approximately 30% reduction of total emissions by 2025. This move would be the largest and least expensive single-source emission reduction NETL could make and will help support DOE's Net-Zero Labs pilot initiative — an effort to begin decarbonizing NETL and three other National Laboratories in support of the Administration's goal to reach net-zero greenhouse gas emissions no later than 2050.

This research, designed to bring scientific modeling to the WSE through the WFA, is also supporting DOE's Energy Earthshots, which is accelerating breakthroughs of more abundant, affordable and reliable clean energy solutions within the decade by driving innovation and creating jobs in the new clean energy sector.

NETL's WFA is matching a once-in-a-lifetime need with a once-in-a-lifetime solution.



NETL's initial focus is on computational fluid dynamics. However, the framework being developed is also valuable for many other scientific areas of study, including geodynamics, geomechanics, molecular dynamics and structural dynamics, among many others.

PARTNERSHIPS

Primary access to the WSE for WFA development comes directly from Cerebras Systems through a signed agreement (AGMT-1067), in which NETL is treated like internal engineer and granted unique access.

NETL is also helping the National Center for Atmospheric Research (NCAR) unlock the potential of the WSE to perform critical climate modeling that could lead to better climate change predictions. The NETL Pittsburgh Super Computing Center and NCAR are using the WFA to conduct the research using Carnegie Mellon University's cutting-edge Neocortex high-performance computing resource, which is powered by the WSE.

A number of nterested partners have stepped forward and expressed interest in using the WFA to help advance their own computational research. Other potential partnerships will help push the WFA into the broader usership if adopted by field equation modeling software such as OpenFOAM®, COMSOL, Star-CCM+, MFiX, VASP and Ansys.

NETL is a U.S. Department of Energy national laboratory that drives innovation and delivers technological solutions for an environmentally sustainable and prosperous energy future. By leveraging its world-class talent and research facilities, NETL is ensuring affordable, abundant and reliable energy that drives a robust economy and national security, while developing technologies to manage carbon across the full life cycle, enabling environmental sustainability for all Americans.

Contact

Dirk Van Essendelft, Ph.D.
Principal Investigator
Dirk.VanEssendelft@netl.doe.gov

Chris Guenther, Ph.D.
Supervisor
Advanced Computing & Artificial Intelligence Team
Chris.Guenther@netl.doe.gov