What's Up? Watts Down — More Science, Less Energy

Applications on one of the world's largest supercomputers demonstrate NVIDIA GPUs lead in energy efficiency.

Author: Dion Harris

People agree: accelerated computing is energy-efficient computing.

The National Energy Research Scientific Computing Center (NERSC), the U.S. Department of Energy's lead facility for open science, measured results across four of its key high performance computing and AI applications.

They clocked how fast the applications ran and how much energy they consumed on CPU-only and GPU-accelerated nodes on Perlmutter , one of the world's largest supercomputers using NVIDIA GPUs.

The results were clear. Accelerated with NVIDIA A100 Tensor Core GPUs , energy efficiency rose 5x on average. An application for weather forecasting logged gains of 9.8x.

On a server with four A100 GPUs, NERSC got up to 12x speedups over a dual-socket x86 server.

That means, at the same performance level, the GPU-accelerated system would consume 588 megawatt-hours less energy per month than a CPU-only system. Running the same workload on a four-way NVIDIA A100 cloud instance for a month, researchers could save more than $4 million compared to a CPU-only instance.

The results are significant because they're based on measurements of real-world applications, not synthetic benchmarks.

The gains mean that the 8,000+ scientists using Perlmutter can tackle bigger challenges, opening the door to more breakthroughs.

Among the many use cases for the more than 7,100 A100 GPUs on Perlmutter, scientists are probing subatomic interactions to find new green energy sources.

The applications NERSC tested span molecular dynamics, material science and weather forecasting.

For example, MILC simulates the fundamental forces that hold particles together in an atom. It's used to advance quantum computing, study dark matter and search for the origins of the universe.

BerkeleyGW helps simulate and predict optical properties of materials and nanostructures, a key step toward developing more efficient batteries and electronic devices.

EXAALT, which got an 8.5x efficiency gain on A100 GPUs, solves a fundamental challenge in molecular dynamics. It lets researchers simulate the equivalent of short videos of atomic movements rather than the sequences of snapshots other tools provide.

The fourth application in the tests, DeepCAM, is used to detect hurricanes and atmospheric rivers in climate data. It got a 9.8x gain in energy efficiency when accelerated with A100 GPUs.

The NERSC results echo earlier calculations of the potential savings with accelerated computing. For example, in a separate analysis NVIDIA conducted, GPUs delivered 42x better energy efficiency on AI inference than CPUs.

That means switching all the CPU-only servers running AI worldwide to GPU-accelerated systems could save a whopping 10 trillion watt-hours of energy a year. That's like saving the energy 1.4 million

homes consume in a year.

You don't have to be a scientist to get gains in energy efficiency with accelerated computing.

Pharmaceutical companies are using GPU-accelerated simulation and AI to speed the process of drug discovery . Carmakers like BMW Group are using it to model entire factories .

They're among the growing ranks of enterprises at the forefront of what NVIDIA founder and CEO Jensen Huang calls an industrial HPC revolution , fueled by accelerated computing and AI.

Discover how AI is powering the future of clean energy .

Original URL: https://blogs.nvidia.com/blog/2023/05/21/gpu-energy-efficiency-nersc/