What Is Green Computing?

Green computing, also called sustainable computing, aims to maximize energy efficiency and minimize environmental impact in the ways computer chips, systems and software are designed and used.

Author: Rick Merritt

Everyone wants green computing.

Mobile users demand maximum performance and battery life. Businesses and governments increasingly require systems that are powerful yet environmentally friendly. And cloud services must respond to global demands without making the grid stutter.

For these reasons and more, green computing has evolved rapidly over the past three decades, and it's here to stay.

Green computing, or sustainable computing , is the practice of maximizing energy efficiency and minimizing environmental impact in the ways computer chips, systems and software are designed and used.

Also called green information technology, green IT or sustainable IT, green computing spans concerns across the supply chain, from the raw materials used to make computers to how systems get recycled.

In their working lives, green computers must deliver the most work for the least energy, typically measured by performance per watt.

Green computing is a significant tool to combat climate change, the existential threat of our time.

Global temperatures have risen about 1.2°C over the last century. As a result, ice caps are melting, causing sea levels to rise about 20 centimeters and increasing the number and severity of extreme weather events.

The rising use of electricity is one of the causes of global warming. Data centers represent a small fraction of total electricity use, about 1% or 200 terawatt-hours per year, but they're a growing factor that demands attention.

Powerful, energy-efficient computers are part of the solution. They're advancing science and our quality of life, including the ways we understand and respond to climate change.

Engineers know green computing is a holistic discipline.

"Energy efficiency is a full-stack issue, from the software down to the chips," said Sachin Idgunji, co-chair of the power working group for the industry's MLPerf AI benchmark and a distinguished engineer working on performance analysis at NVIDIA.

For example, in one analysis he found NVIDIA DGX A100 systems delivered a nearly 5x improvement in energy efficiency in scale-out AI training benchmarks compared to the prior generation.

"My primary role is analyzing and improving energy efficiency of AI applications at everything from the GPU and the system node to the full data center scale," he said.

Idgunji's work is a job description for a growing cadre of engineers building products from smartphones to supercomputers.

Green computing hit the public spotlight in 1992, when the U.S. Environmental Protection Agency launched Energy Star , a program for identifying consumer electronics that met standards in energy efficiency.

A 2017 report found nearly 100 government and industry programs across 22 countries promoting what it called green ICTs, sustainable information and communication technologies.

One such organization, the Green Electronics Council , provides the Electronic Product Environmental Assessment Tool , a registry of systems and their energy-efficiency levels. The council claims it's saved nearly 400 million megawatt-hours of electricity through use of 1.5 billion green products it's recommended to date.

Work on green computing continues across the industry at every level.

For example, some large data centers use liquid-cooling while others locate data centers where they can use cool ambient air. Schneider Electric recently released a whitepaper recommending 23 metrics for determining the sustainability level of data centers.

Wu Feng, a computer science professor at Virginia Tech, built a career pushing the limits of green computing. It started out of necessity while he was working at the Los Alamos National Laboratory.

A computer cluster for open science research he maintained in an external warehouse had twice as many failures in summers versus winters. So, he built a lower-power system that wouldn't generate as much heat.

He demoed the system, dubbed Green Destiny, at the Supercomputing conference in 2001. Covered by the BBC , CNN and the New York Times , among others, it sparked years of talks and debates in the HPC community about the potential reliability as well as efficiency of green computing.

Interest rose as supercomputers and data centers grew, pushing their boundaries in power consumption . In November 2007, after working with some 30 HPC luminaries and gathering community feedback, Feng launched the first Green500 List , the industry's benchmark for energy-efficient supercomputing.

The Green500 became a rallying point for a community that needed to reign in power consumption while taking performance to new heights.

"Energy efficiency increased exponentially, flops per watt doubled about every year and a half for the greenest supercomputer at the top of the list," said Feng.

By some measures, the results showed the energy efficiency of the world's greenest systems increased two orders of magnitude in the last 14 years.

Feng attributes the gains mainly to the use of accelerators such as GPUs, now common among the world's fastest systems.

"Accelerators added the capability to execute code in a massively parallel way without a lot of overhead — they let us run blazingly fast," he said.

He cited two generations of the Tsubame supercomputers in Japan as early examples. They used NVIDIA Kepler and Pascal architecture GPUs to lead the Green500 list in 2014 and 2017, part of a procession of GPU-accelerated systems on the list.

"Accelerators have had a huge impact throughout the list," said Feng, who will receive an award for his green supercomputing work at the Supercomputing event in November.

"Notably, NVIDIA was fantastic in its engagement and support of the Green500 by ensuring its energy-efficiency numbers were reported, thus helping energy efficiency become a first-class citizen in how supercomputers are designed today," he added.

Today, GPUs and data processing units (DPUs) are bringing greater energy efficiency to AI and networking tasks, as well as HPC jobs like simulations run on supercomputers and enterprise data centers.

AI, the most powerful technology of our time, will become a part of every business. McKinsey & Co. estimates AI will add a staggering $13 trillion to global GDP by 2030 as deployments grow.

NVIDIA estimates data centers could save a whopping 19 terawatt-hours of electricity a year if all AI, HPC and networking offloads were run on GPU and DPU accelerators (see the charts below). That's the equivalent of the energy consumption of 2.9 million passenger cars driven for a year.

It's an eye-popping measure of the potential for energy efficiency with accelerated computing .

Because AI represents a growing part of enterprise workloads, the MLPerf industry benchmarks for AI have been measuring performance per watt on submissions for data center and edge inference since February 2021.

"The next frontier for us is to measure energy efficiency for AI on larger distributed systems, for HPC workloads and for AI training — it's similar to the Green500 work," said Idgunji, whose power group at MLPerf includes members from six other chip and systems companies.

The public results motivate participants to make significant improvements with each product generation. They also help engineers and developers understand ways to balance performance and efficiency across the major AI workloads that MLPerf tests.

"Software optimizations are a big part of work because they can lead to large impacts in energy efficiency, and if your system is energy efficient, it's more reliable, too," Idgunji said.

In PCs and laptops, "we've been investing in efficiency for a long time because it's the right thing to do," said Narayan Kulshrestha, a GPU power architect at NVIDIA who's been working in the field nearly two decades.

For example, Dynamic Boost 2.0 uses deep learning to automatically direct power to a CPU, a GPU or a GPU's memory to increase system efficiency. In addition, NVIDIA created a system-level design for laptops, called Max-Q , to optimize and balance energy efficiency and performance.

When a user replaces a system, the standard practice in green computing is that the old system gets broken down and recycled. But Matt Hull sees better possibilities.

"Our vision is a cyclical economy that enables everyone with AI at a variety of price points," said Hull, the vice president of sales for data center AI products at NVIDIA.

So he aims to find the system a new home with users in developing countries who find it useful and affordable. It's a work in progress seeking the right partner and writing a new chapter in an existing lifecycle management process .

Energy-efficient computers are among the sharpest tools fighting climate change.

Scientists in government labs and universities have long used GPUs to model climate scenarios and predict weather patterns. Recent advances in AI, driven by NVIDIA GPUs, can now help model weather forecasting 100,000x quicker than traditional models. Watch the following video for details:

In an effort to accelerate climate science, NVIDIA announced plans to build Earth-2 , an AI supercomputer dedicated to predicting the impacts of climate change. It will use NVIDIA Omniverse , a 3D design collaboration and simulation platform, to build a digital twin of Earth so scientists can model climates in ultra-high resolution.

In addition, NVIDIA is working with the United Nations Satellite Centre to accelerate climate-disaster management and train data scientists across the globe in using AI to improve flood detection.

Meanwhile, utilities are embracing machine learning to move toward a green, resilient and smart grid. Power plants are using digital twins to predict costly maintenance and model new energy sources, such as fusion-reactor designs.

Feng sees the core technology behind green computing moving forward on multiple fronts.

In the short term, he's working on what's called energy proportionality, that is, ways to make sure systems get peak power when they need peak performance and scale gracefully down to zero power as they slow to an idle, like a modern car engine that slows its RPMs and then shuts down at a red light.

Long term, he's exploring ways to minimize data movement inside and between computer chips to reduce their energy consumption. And he's among many researchers studying the promise of quantum computing to deliver new kinds of acceleration.

It's all part of the ongoing work of green computing, delivering ever more performance at ever greater efficiency.

Discover how AI is powering the future of clean energy .

Original URL: https://blogs.nvidia.com/blog/2022/10/12/what-is-green-computing/