Accelerated Computing is Sustainable Computing

Improve energy efficiency and reduce costs in the era of AI.

Over the past 60 years, the computing industry has experienced several significant shifts, from the introduction of PCs to the advent of mobile and cloud technologies. The industry is now at a pivotal moment where the CPU—the processing engine used universally— continues to advance, but at a cost. Today's CPU's generational performance gains are diminishing and increasingly require more power. At the same time, the need for more computational power is skyrocketing as **enterprises continue to innovate**.

The growing gap between increasing volumes of data and lagging CPU performance has led to what's known as computation inflation. This situation, where data growth is exponential but CPU performance isn't keeping pace, isn't sustainable, and it's leading to greater energy consumption as data centers scale to keep up. This growing gap necessitates a shift toward more efficient computing solutions, which can handle compute-heavy workloads with less energy consumption.

The long-standing prediction that the number of transistors per unit area will double about every two years, known as Moore's Law, is no longer physically achievable.

Boosting Performance and Energy Efficiency

NVIDIA has developed an architecture that pairs a GPU with a CPU, significantly enhancing processing speeds. By allowing these two processors to operate concurrently, each functioning autonomously and independently, performance and energy efficiency can be dramatically accelerated. Tasks that previously took 100 units of time can now be completed in just one unit of time. This translates into incredible savings of time, energy, and costs per task, enabling breakthroughs across scientific computing, healthcare, climate tech, and more.

In the PC industry, this approach is quite common. For instance, integrating a $500 **NVIDIA® GeForce™ GPU** into a $1,000 PC can significantly boost the PC's performance. Similarly, by transitioning from CPU-only operations to GPU-accelerated systems, high-performance computing and AI workloads can **save over 40 terawatt-hours of energy annually**, equivalent to the electricity needs of nearly 5 million U.S. homes.

Future-Ready Infrastructure

To accommodate new workloads like generative AI while working toward environmental, social, and governance (ESG) goals, investment in accelerated computing is critical. **Generative AI** is reshaping industries, creating new opportunities for innovation and growth. It represents a fundamental shift from traditional computing, which is primarily retrieval-based, to a model where computers generate contextually relevant content. This transition requires new hardware and software solutions and reevaluating of existing computing strategies. Accelerated computing offers a comprehensive suite of hardware, software, and services to, not only support future workloads like generative AI, but to power industrial digitalization, making it possible to deploy AI at industry scale.

Looking ahead, the implications of accelerated computing extend beyond technical enhancements: They herald a shift toward more sustainable and cost-effective computing practices. Many companies invest hundreds of millions of dollars to process data. Applying accelerated computing to this process could lead to significant financial savings or vastly increased computational capacity, while laying the foundation for new, innovative workloads.

General-purpose computing has long been burdened by computational inflation. By deciding to accelerate processes, businesses can recover a substantial amount of built-in loss and continue to innovate without being hindered by the physical limitations of hardware.

Ready to Get Started?

Learn more about NVIDIA and **accelerated computing.**

NVIDIA