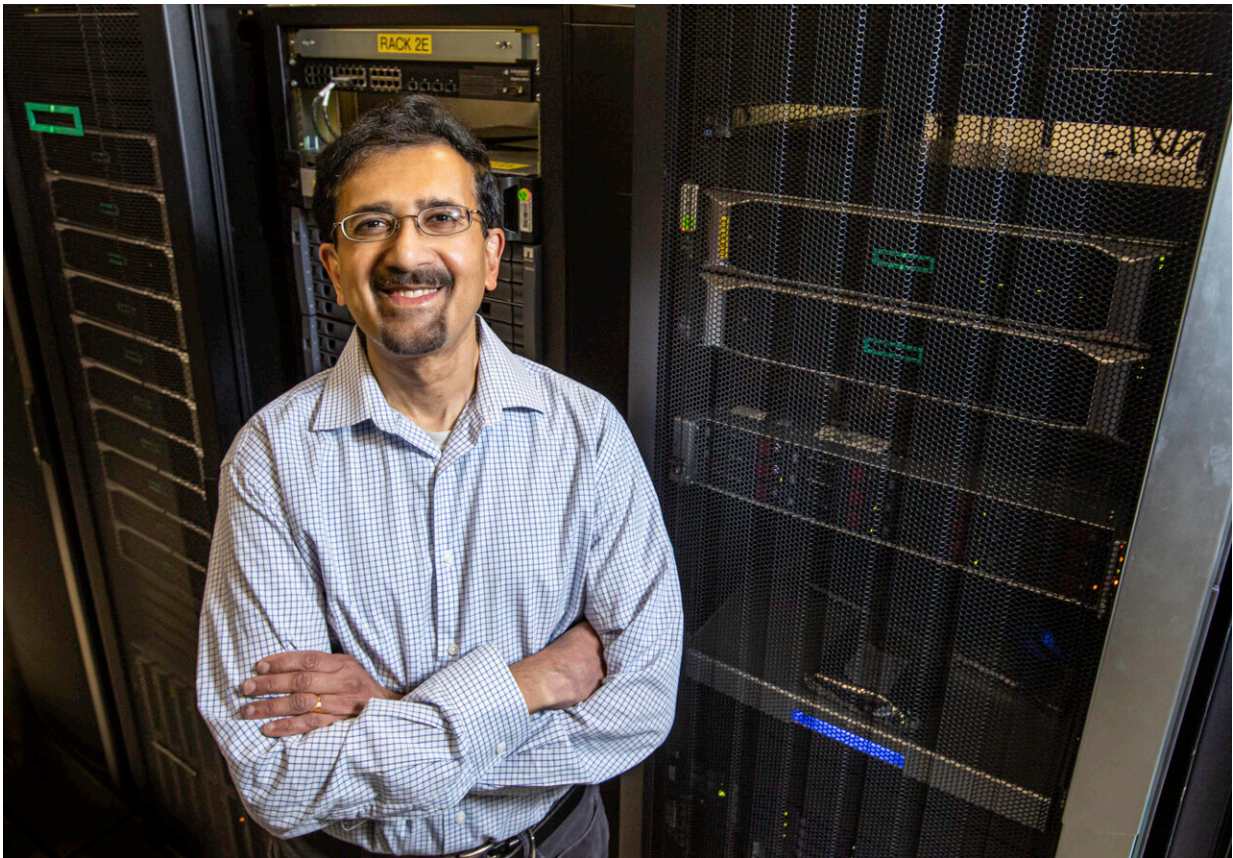# Q&A: How to make computing more sustainable

March 22 2023, by David Krause

SLAC researcher Sadasivan Shankar inside SSRL, where he works on improving the energy efficiency of computing. Credit: Jacqueline Ramseyer Orrell/SLAC National Accelerator Laboratory

Ask your computer or phone to translate a sentence from English to

Italian. No problem, right? But this task is not as easy as it appears. The software behind your screen had to learn how to process hundreds of billions of parameters, or tasks, before displaying the correct word—and doing those tasks takes energy.

Now, researchers at the Department of Energy's SLAC National Accelerator Laboratory and other public and private institutions are searching for ways to supply less energy to software and hardware systems while still accomplishing everyday tasks, like language translation, as well as solving increasingly difficult but socially-important problems like developing new cancer drugs, COVID-19 vaccines and self-driving cars.

This computing effort at SLAC is part of a larger DOE national initiative, led by the Advanced Materials and Manufacturing Technologies Office, called The Energy Efficiency Scaling for 2 Decades (EES2) that was announced in September 2022. The initiative involves several national labs as well as industry leaders, and will focus on increasing the energy efficiency of semiconductors by a factor of 1,000 over the next two decades, the initiatives says. By 2030, semiconductors could use almost 20 percent of the world's energy, meaning improving the efficiency of this sector is essential to help grow the economy and take on the climate crisis, the initiative says.

On paper, tailoring software models to reduce their energy usage is simple: just include a new design variable that accounts for the energy requirements of a model when designing its algorithm, said Sadasivan Shankar, a research technology manager at SLAC and adjunct professor at Stanford University. However, many software models, like those that rely upon machine learning, lack this energy design variable. Instead, they are often built with performance, not efficiency, as their driving force, he said.

In this Q&A, Shankar explains how his team's projects at SLAC will try to improve the energy efficiency of computing going forward.

Computing is a large, broad field. What parts are your team tackling?

We are currently looking at three main parts of computing: algorithms, architecture and hardware. For algorithms, we will study how to reduce the energy required by machine learning algorithms and models. One way to improve the energy efficiency of these models is to use tailored algorithms, which compute specific tasks for each unique application, whereas more general algorithms are designed and applied to complete a range of tasks.

The second part we are looking at is how to design software architecture and their algorithms together—called "co-design"—rather than designing them independently of one another. If these components are co-designed, they should need less energy to run. And third, we are looking at the fundamental level of materials, devices and interconnects that generate less heat.

To tackle these three areas, we are going to look at efficiencies in nature, like how our brain and molecular cells perform tasks, and try to apply these learnings to our design of computing systems.

Tell us about the importance of machine learning models in today's society. To which problems are we seeing the models being applied?

Machine learning models are being applied to more and more fields, from language processing tools, to biology and chemistry problems, to electric vehicles, and even to particle accelerator facilities, like at SLAC.

A specific example that we have looked at already is language learning models. A few natural language learning models have more than 170 billion parameters that need to be optimized when training the model. Machine learning models generally attempt to learn the patterns between a defined set of inputs and outputs in a large dataset. This part of building a model is called training, and it is incredibly energy intensive.

In our preliminary analysis, we found that a single language model (e.g. ChatGPT) on the lower bound required about as much electricity as the average monthly electricity usage by the city of Atlanta or Los Angeles in 2017. Therefore, if we can design more efficient training models, like by using specific, tailored algorithms, energy usage for training can go down. Our intent is to analyze these training energy needs systematically and use the learned principles to develop better solutions in applying AI.

Does computing require more energy today than a decade ago?

Computing today is more energy-efficient than a decade ago, but we are using many more computing tools today than a decade ago. So overall, the amount of energy required by computing has increased over time. We want to bend the energy usage trajectory curve down, so we can continue to grow computing across the world without adversely affecting the climate.

What is the most difficult challenge in your mind to reducing the energy requirements of computing?

Right now, I would say scaling new manufacturing technology is our most difficult challenge. Currently new generations of technology are below 10 nanometers and are approaching length scales on-par with spacing between the atoms. In addition, development of new technology

is expensive and needs several billion dollars in research and development.

Second to this challenge is changing architecture and [hardware](#), which is more difficult than changing software and algorithms. Hardware requires manufacturing at scale and many more players are involved. Let's say you came up with the most efficient algorithm on the most efficient device, but it takes twenty billion dollars to manufacture. In this case, the design failed because it is too expensive to build at scale. You have to look at manufacturing in conjunction with new architectures, software design and other factors. Otherwise, the whole effort becomes a moot academic discussion. We are hoping to map out several solutions for our research and industrial partners to build upon.

The third challenge is to develop algorithms and software that can keep up with our increasing dependencies on technology, but are energy-efficient.

What future research areas in computing energy efficiency are you most excited about?

The most exciting opportunity to me is to use artificial intelligence itself to solve our energy efficiency problem in computing. Let's use the positive aspect of AI to reduce our [energy usage](#).

The other exciting thing is in the future, computers will be more like brains with distributed sensors that require way less energy than today's devices in processing optimal information. These future computers can be inspired by the ways neurons are connected and may borrow principles from quantum computing but can do classical computing as well. This will get our machines to function more like nature—more efficiently.

Provided by SLAC National Accelerator Laboratory