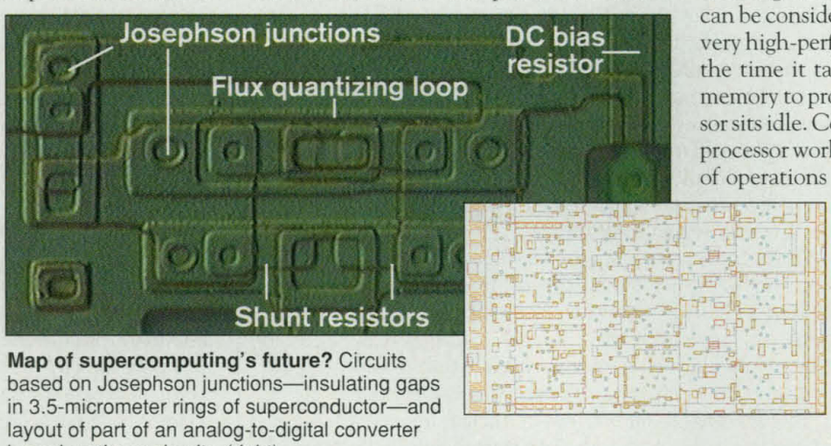
Redefining the Supercomputer

The government's high-performance computing program has an audacious new goal: a computer more than 1000 times faster than today's fastest machine. The technologies are already taking shape

The word is petaflops, computer jargon for 1000 trillion computations per second. Think of it as a year's labor for a powerful workstation compressed into 30 seconds. Think of it, also, as 1000 times the speed of the current computing benchmark, a trillion operations a second—teraflops—which is on the verge of becoming a reality at Sandia National Laboratories after 5 years of effort. Now the federal government's high-performance computing program is aiming for a petaflops, and researchers are exploring new technologies, sketching new architectures, and pondering the software challenge of harnessing this staggering computational power.



RSFQ LABORATORY, SUNY STONY BROOK

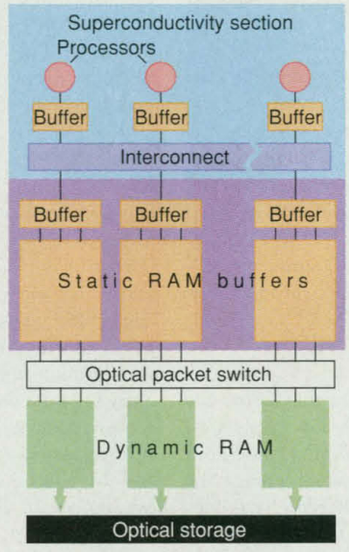
Map of supercomputing's future? Circuits based on Josephson junctions—insulating gaps in 3.5-micrometer rings of superconductor—and layout of part of an analog-to-digital converter based on these circuits *(right)*.

The effort illustrates a truism of supercomputer development: In fields such as climate modeling and computational chemistry, each jump in speed builds demand for the next one (see box on p. 1656). Indeed, it was only 2 years into the teraflops program, in 1993, when Dan Goldin, head of NASA, suggested that a teraflops might be short-sighted and requested a federal effort to look into "very high-performance computing," by which he meant petaflops speeds. Now NASA, the Department of Energy, the National Science Foundation (NSF), the National Security Agency, and the Defense Advanced Research Projects Agency have responded by convening what, for lack of a better name, has come to be called the Interagency Petaflops Initiative Computing Group.

Over the past 2 years, this group of computer designers and high-performance computer users has held four workshops, the last of them in June in Bodega, California, to begin laying the groundwork for a petaflops supercomputer that could be built within 10 years. In May, as part of the new initiative, NSF funded eight small research projects "to determine the viable approaches, ferret out those that might not work, and identify the key technological challenges," says Caltech computer scientist Thomas Sterling, a member of the petaflops group.

The projects—which total some $1 million—are a response to the first petaflops workshop, in January 1994. "Its principal finding," says Rick Stevens, a computer scientist at Argonne National Laboratory in Illinois, "was that we believe it is possible to build a machine of that performance level." But the advances that led to a teraflops machine—faster silicon processors and parallel designs that yoke together thousands of processors—may peter out well before a petaflops, participants noted. Says Stevens, "We will have to look at more exotic hardware technology."

He and others in the exploratory projects are studying a spectrum of technologies and processors. At one extreme are designs relying on several thousand superfast superconducting processors. In the middle are drastic extensions of current supercomputer architectures, incorporating roughly 100,000 conventional silicon processors in small clusters. At the other extreme are machines with a million simple processors in which, says Stevens, "the distinction between memory and processor is very blurry." No one can say yet whether any of these architectures—or others now in gestation—will deliver the needed speedup in a mere decade. But after decades in which the same basic silicon technology has sustained an exponential increase in computer speed, the quest for a petaflops could mark the beginning of something new.



T. STERLING

Hybrid strategy. A petaflops computer might combine superconducting processors with an optical memory based on holograms.

Any design will have to handle the twin challenges of latency and concurrency, which can be considered the Scylla and Charybdis of very high-performance computing. Latency is the time it takes a piece of data to get from memory to processor—time when the processor sits idle. Concurrency is the need to have a processor working on more than one sequence of operations simultaneously so it can fill its latent periods.

Both problems intensify as processors get faster. For a processor that does an operation every 10 picoseconds (trillionths of a second)—the outer limit of performance imaginable right now for any technology in the next 10 years—eliminating latency would require that all the pertinent hardware fit within a diameter of 3 millimeters, the maximum distance light or anything else can travel in 10 picoseconds. But there are limits to how small a petaflops computer can be. One is set by heat buildup: A petaflops machine built with conventional silicon technology would consume tens or hundreds of megawatts of power, says Sterling—the output of a modest power plant—and produce intense waste heat. "If we're not careful," says University of Notre Dame computer scientist Peter Kogge, "we're going to end up with computers that literally glow in the dark."

As a result, any design for a petaflops machine will have serious latency problems and will have to supply plenty of concurrency to beat them. "No matter what technology I'm using," says Stevens, "at least some memory is going to be many clock cycles away from the processor." To keep those 10-picosecond processors from sitting idle waiting for access to data, each processor would have to have on the order of 100

How Great a Need for Speed?

At 1000 times faster than the fastest computer today, does a petaflops computer constitute speed for speed's sake? Is it really worth the enormous challenge of designing processors that work on time scales of trillionths of a second and software that can orchestrate perhaps a million parallel operations (see main text)? In some areas, the payoff will be surprisingly modest. Rick Stevens of Argonne National Laboratory points out that a 1000-fold speedup buys only a sixfold increase in spatial and temporal resolution in climate models and other three-dimensional simulations. But a petaflops also marks a threshold where other kinds of computation will become feasible or practical for the first time.

■ *Time-critical problems*. Examples include real-time computational support for surgery, predicting the flow of contaminants from a toxic waste spill, or modeling a stock market anomaly that requires a quick decision. "With petaflops speed," says Stevens, "you could do complex analysis in about a second that would take 2 weeks on a workstation."

■ *Multidisciplinary applications*. These combine different kinds of existing simulations. Imagine an aircraft design computation that not only calculates the aerodynamics of the structure but also maximizes manufacturability and reliability while minimizing cost and noise. "Each of those today would require a different type of modeling regime," says Stevens, "but with petaflops, you could do it simultaneously."

■ *Exotic applications*. One example is what petaflops researchers call a "mirworld" —a three-dimensional simulation space in which multiple users could run simultaneous experiments. A mirworld of Earth, for instance, might allow researchers to simultaneously manipulate the biosphere, geosphere, atmosphere, and oceans, as well as agricultural systems and markets, exploring how the effects of each change ripple through the interacting systems.

And then there are national security applications, which have traditionally driven the quest for the highest performance computing. Petaflops speed should make it possible for the first time to create reasonable simulations of nuclear bomb explosions, a high priority for the Department of Energy since the government officially stopped underground nuclear testing. The National Security Agency is a petaflops enthusiast as well, says computer scientist Geoffrey Fox of Syracuse University in New York, but "we're not allowed to think about their applications." –G.T.

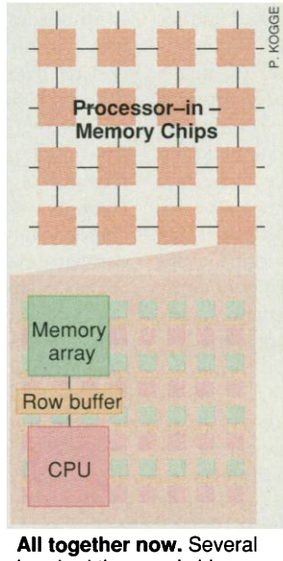
tasks outstanding at any given time. Supplying 100 tasks at a time to 10,000 processors—the number of 10-picosecond processors in a petaflops machine—means providing what Stevens calls "million-way parallelism," an unprecedented challenge for hardware and software designers.

That is about as low as the number can get, he says. Slow down the processors, and the latency problem gets better, which means less concurrency per processor, but more processors to reach a petaflops. Speed up the processors, and each processor has to keep more balls in the air at once. "There is no way to escape this millionfold concurrency," Stevens says. "You can push it around, but you can't get away from it."

Extreme measures

One way to juggle the numbers is to opt for comparatively few, blindingly fast processors. That is the strategy behind RSFQ, for Rapid Single-Flux Quantum logic, which enlists Superconducting Quantum Interference Devices (SQUIDs) as the circuit elements. Each SQUID is a ring of low-temperature superconductor cut by a thin gap. The ring can hold a single quantum of magnetic flux, and the presence or absence of the flux constitutes the 1 or 0 of binary logic. A weak current can switch the SQUID from one state to the other (*Science*, 24 September 1993, p. 1670). "The lay version of why a superconducting circuit can switch faster than a room-temperature circuit," says Stevens, "is because the amount of energy it takes to switch is much lower."

Konstantin Likharev, a physicist at the State University of New York, Stony Brook, who developed much of the technology, says that last spring he, Vasili Semenov, and their colleagues tested the first RSFQ system, an array of 15 logic elements that converts analog signals to digital output. The speed was comparable to that of the fastest silicon transistors, says Likharev, and he predicts that RSFQ processors should eventually be able to switch states in a few picoseconds, or at least 200 times faster than the silicon technology projected for 10 years from now.



All together now. Several hundred thousand chips combining processors and memory could be combined in a petaflops computer.

At that rate, a petaflops machine would need between 1000 and 10,000 RSFQ processors and would consume so little power—"between 30 and 300 watts," estimates Likharev—that it need be no larger than a PC. Given those dimensions, says Likharev, latency would be about 1 nanosecond, requiring 100-fold concurrency per processor, "which the computer science people we're speaking with say is doable." But RSFQ is a largely untried scheme, notes Stevens, and it will be a challenge to make it coexist comfortably with other critical parts of a supercomputer. "You can't put the whole universe inside a cryostat, or close to absolute zero," he says. "Somehow the machine has to talk to users and to memory, and those can't run at cryogenic temperatures."

A less exotic alternative, called processor-in-memory (PIM) technology, would attain petaflops speeds with vast numbers of simpler, slower processors. This fine-grained parallelism is the basic idea behind some of today's fastest computers, but PIM adds a twist to the concept. By placing between 16 and 64 processors on the same chip as the memory they call on, PIM technology not only minimizes latency; it also eliminates a bottleneck that plagues current computer design. When a processor in today's computers asks for data from memory, Notre Dame's Kogge explains, the data have to be sent in a narrow stream. "Inside the memory chip you have incredible bandwidth," explains Kogge, who is pursing PIM designs. "But because of the way we design the parts separately, as soon as you cross chip boundaries, you throw all that bandwidth away."

In a PIM design, however, the memory feeds directly into the processors, so there is no need to "serialize" it to go through external wires. Kogge estimates that a PIM design might require anywhere from 10,000 to several hundred thousand chips, each with dozens of processors and accompanying memory. While that sounds complex, he says, putting memory next to the processors reduces latency and concurrency, "and you can go back to Computer Science 101 kinds of [thinking] and do things a lot more simply."

Between the two extremes of RSFQ and PIM sits an extension of the cluster-based architecture that will be delivering teraflops performance at Sandia this fall. Supercomputer-makers are now exploring architectures in which the building blocks are clusters of 16 or 32 of the fastest processors silicon technology can offer, all tightly networked together and sharing the same memory, says Paul Smith, who

heads the Department of Energy's petaflops effort. Thousands of those clusters would be loosely linked to produce a petaflops computer. The design will require extraordinarily fast communications between the clusters, however. "We're going to have to work through and overcome all these interconnection problems," says Smith. "That's not trivial."

Petaperipherals, petaprogramming

All three of these concepts, adds Smith, are only "candidate or example architectures, and we don't expect to be locked in by those three." Among the other NSF-funded projects, for example, is a University of Illinois design, in which the software can actually manipulate the hardware during operation, in effect continuously rewiring the computer to match its hardware structure to the algorithm in use at the time, thereby reducing latency. Another candidate comes from a Purdue University group, which is working on an architecture called variable granularity processor hierarchy, which relies on multiple small processors for most processing tasks but calls on a few very powerful but power-hungry processors when needed.

The NSF is also studying exotic technologies to store and retrieve the oceans of data a petaflops machine would generate. Caltech's Sterling, Likharev, and collaborators, for example, are exploring a hybrid machine that would link superconducting logic to a primary memory (the equivalent of RAM in a desktop computer) based on optical holographic storage. The memory would capture data in a medium that records interference patterns generated by intersecting laser beams. Ten to 100 terabytes of data could be encoded in those patterns and packed into a cubic centimeter of the storage medium. And when a "read" beam shines on the storage material, adds Sterling, "you can get all that information in big shots, megabytes per access."

The downside is that lab prototypes of holographic memory suggest it will probably take microseconds to deliver the data—100 times longer than the best silicon memories of the next decade. So the computer would need a hierarchy of different kinds of memories, from fast silicon to slower holographic ones. "The trick is to have the memory system itself bundle up large bunches of data and send it all at once to the processors," says Sterling.

Then there is the problem of displaying the output of a petaflops computer to the user, which could mean somehow communicating 100 terabytes of data. Ian Foster of Argonne describes this as a "problem of trying to provide a thick pipe from the supercomputer to the outside world," with the most likely solution being some kind of three-dimensional display, either virtual reality or what is known as a cave, in which the operator is actually inside a room-sized display device.

If all that isn't trouble enough, says Stevens, "At least half of the work, and probably 70%, is going to be software." A petaflops computer will need what by today's standards is an enormous degree of control over the internal data movement. "This is the hardest possible problem in parallel computing," says Stevens, "an efficient way of managing data movement throughout the machine." Part of the challenge is orchestrating the million-way parallelism of such a machine, but the other part is moving data between what is likely to be six to eight levels of memory, each one getting progressively faster as it gets closer and closer to the processors. "We want to start software development in the next couple of years," says Foster. "There's no point having great hardware unless you have pretty great software."

All of this will take another step toward reality next month, when the NSF-funded architecture projects will present their findings at the Frontiers of Computing 1996 meeting in Annapolis, Maryland. At that point, says John Toole, director of the National Coordination Office for High-Performance Computing, "they'll bang each of these ideas up against computer architects, software people, applications people, to really get some synergy about what the real meat of problems will be for the long term." Then it will be up to the funding agencies to decide whether to support a full-scale petaflops project—an effort that is likely to cost $400 million each year for a decade, twice the spending on supercomputers of the current high-performance computing initiative.

If the government decides against it, industry isn't likely to step in on its own, say the petaflops researchers. Many supercomputer-makers have fared poorly in recent years. And the kind of focused effort necessary to achieve a petaflops within 10 years, says computer scientist Geoffrey Fox of Syracuse University in New York, is not the kind of effort that is likely to be driven by market forces. "Rather it's like the atom bomb or the space program," he says. "Something outside of an industrial endeavor, but of importance to the nation."

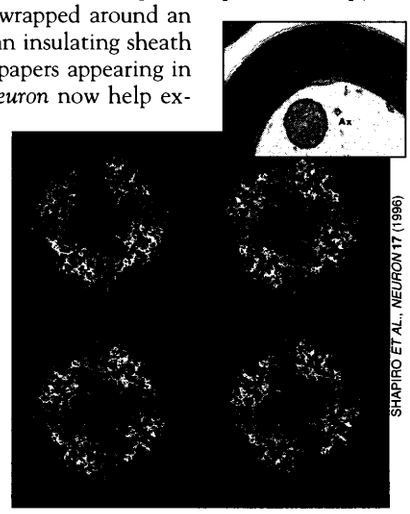
–Gary Taubes

NEUROBIOLOGY

Glimpsing Myelin's Protein Glue

A nerve signal can zip from your spinal cord to the tip of your toe in less than 25 milliseconds. But such rapid nerve transmission is only possible because the axons, the long neuronal projections that carry the signals, have very good electrical insulation. Layers of tightly packed cell membranes wrap the axons much like gauze wrapped around an injured finger, forming an insulating sheath known as myelin. Two papers appearing in this month's issue of *Neuron* now help explain how these tightly wrapped layers are glued together in peripheral nerves and how some mutations weaken the glue, leading to neurological disease.

By determining the crystal structure of part of the glue, a myelin protein known as P_0, a team including Lawrence Shapiro and Wayne Hendrickson of Columbia University in New York City and David Colman of Mount Sinai Medical School, also in New York, has found that the protein molecules apparently interlock to form a sort of molecular Velcro between the myelin membranes. And in work described in an accompanying paper, Laura Warner and James Lupski at Baylor College of Medicine in Houston and their colleagues use the new structure to analyze the effects of some of the 29 known disease-causing mutations in P_0, including five new ones they discovered. While the work doesn't open immediate avenues to improved therapy for the diseases, it does provide insights into how the P_0 mutations can cause symptoms ranging from poor coordination to paralysis.



Sticky structure. The myelin surrounding nerve fibers *(inset)* is held together by interlocking P_0 tetramers.

"Here is an example in which nature has already done the function part of a structure-function analysis for you," says neuroscientist Greg Lemke of the Salk Institute. "There are very few examples like that where you have a large panel of mutations that have differing effects, and you have a very nice, high-resolution structure you can map them onto."

Healthy myelin appears under the electron microscope as "beautiful perfect spirals with dozens of wraps of membrane, each with the exact same spacing," Shapiro says. Researchers have known for several years that P_0 seals the membranes together with that perfect spacing. Without it,