Life After Silicon: Ultrascale Computing

FEATURE

By Nancy Forbes

Researchers explore exotic alternatives to semiconductor chips for computers of the future

Trying to obtain the highest storage density on computer chips at the lowest cost per logic function or memory bit poses problems that increasingly raise concerns about how far the industry can advance semiconductor technology. As demand increases for higher performance, greater speed, higher storage density, and lower power and cost, extrapolations from existing technologies run up against formidable physical obstacles. For instance, in shrinking transistor size in order to pack more on a chip, other parameters must scale accordingly. The Semiconductor Industry Association's National Technology Roadmap for Semiconductors, which plots and predicts continued improvements in chip performance, extends only to the year 2010, reflecting the uncertainty about how and when industry will solve these problems. There is, however, an updated road map currently in preparation that will extend beyond the year 2010.

The Defense Advanced Research Projects Agency (DARPA), the R&D arm of the U.S. Department of Defense, addresses the issue by supporting innovative "high risk–high payoff" research on alternative computing technologies for military and dual-use applications. DARPA funds researchers to pursue an idea from its inception, through the feasibility stage, to a demonstrated prototype, with the goal of accelerating a technology's development beyond its normal evolutionary pace. The agency worries about the future of computing because of the military's vast computing needs, including signal processing capabilities for sonar and radar, computational prototyping of new systems in the area of acquisition and program development, information systems for command, control, communications, and intelligence (C^{3}I), and distributed interactive simulations to facilitate training, policy, and strategic analysis.

In 1996, DARPA inaugurated the Ultrascale Computing Program to explore novel methods of information processing ("ultrascale" meaning beyond the scaleable). The 5-year program currently has under contract more than 20 scientists and engineers from many leading academic and private-sector research institutions.

This information processing program focuses on physical phenomena from disciplines other than electrical engineering, namely chemistry, biology, and physics. In some cases, it pairs new concepts with the appropriate engineering discipline to create a hybrid computing architecture. The program's mission is to assess the fundamental limits of existing computer technologies through predictive modeling techniques, devise new concepts in architecture and modes of computing, and devise new physical mechanisms with which to implement these modes. It also aims to establish the theoretical basis for the new mechanisms, validate them through experimentation, and, finally, link researchers to these fledgling technologies by attracting a highly multidisciplinary research community.

Some ultrascale projects that have dual military and commercial promise include continuum computer architecture; amorphous computing; cellular engineering; the use of DNA for computing; cultivated neural networks, and quantum computing.

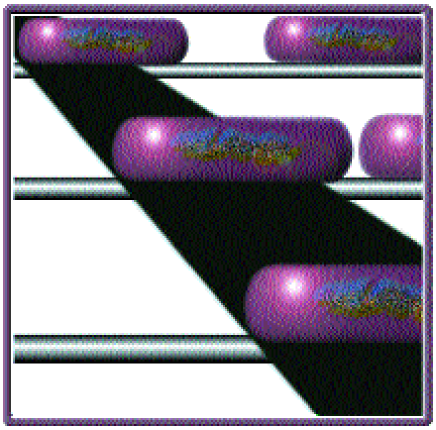
CCA blocks

Continuum Computer Architecture (CCA), a project at the California Institute of Technology (Caltech), seeks to establish a new, symbiotic relationship between computer hardware and software. The concept calls for an architecture based on simple fine-grained elements called CCA blocks. These function locally and as an ensemble, the latter mode providing a general structure for program execution. CCA blocks combine the functionality of logic, storage, and nearest-neighbor communication, completely eliminating the distinction between processing, storage, and execution elements found in massively parallel computer hardware and eradicating the resulting bottlenecks. All the data is related in the form of abstract structures embodied in two- and three-dimensional arrays of CCA blocks. The local operations of the individual blocks are in symbiosis with the higher-level, aggregate functioning and produce a global computing model capable of executing general parallel computer programs. "At this stage,

industry hasn't yet shown much interest in this technology,” says Paul Messina, director of Caltech’s Center for Advanced Computing Research. “I personally believe that if CCA were built and used, there would be industrial interest. However, short of that, it will be difficult to get commercial users interested.”

Nevertheless, Messina sees much promise in the CCA technology. “I can seriously imagine some CCA concepts being incorporated into commercial products, much like the progression we have seen with very long instruction word (VLIW) machines, for example, Multiflow or Cydrome,” he says. VLIW processors have more than one functional unit for tasks such as load/store, floating-point add/multiply, and integer arithmetic operations. VLIW devices were once commercialized, but their makers went out of business. However, today’s hottest microprocessors use VLIW techniques.

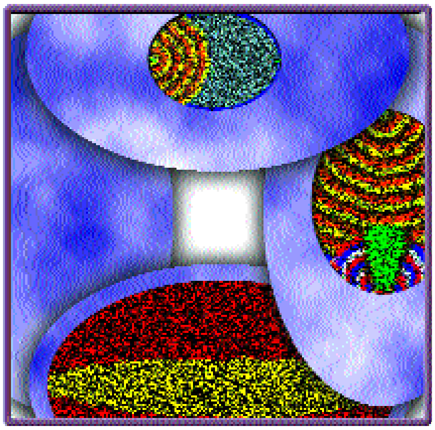
CCA will make use of future semiconductor feature sizes of 0.1 μm for its implementation (current feature sizes are about 0.25 μm). It will operate at a rate of 100 teraflops (1 teraflop=10^{12} floating point operations per second) with substantially reduced design costs.



Amorphous computing

Amorphous computing, a project at the Massachusetts Institute of Technology (MIT), is essentially an enabling technology, designed to help meet the challenge of learning how to build and organize information processing systems comprised of huge numbers of molecular-scale components. Working from an engineering standpoint, it aims to explore the basic principles that, for instance, underlie a swarm of bees cooperating to form a coherent hive, or living cells comprising a functioning multicellular organism whose behavior is directed by a common genetic program. Amorphous computing will then translate these principles into the technological foundation for a system architecture and its component algorithms.

The principles behind amorphous computing are those that enable large numbers of unreliable elements to connect in unknown, irregular, and erratic ways to form coherent group behavior. “Our goal is to identify engineering principles for organizing and instructing myriad programmable entities to cooperate in order to achieve pre-established goals, creating amorphous computing systems for intelligent materials,” explains Gerald Sussman of MIT. “With some experience, we can expect to extract the common patterns and organize them into an engineering language with primitive elements—means of combination and means of abstraction. Amorphous computing has fundamental implications for software and system design, in addition to immediate application to programmable materials.”



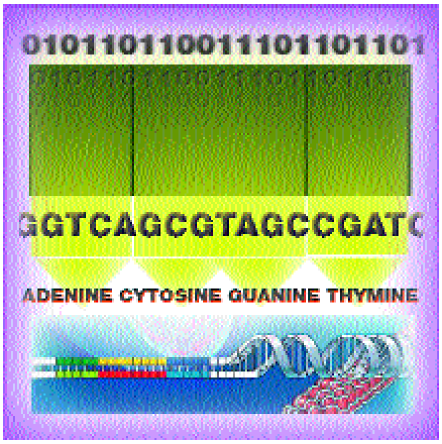
Amorphous computing involves a set of elements that themselves change in the course of the computation. It is a powerful tool, Sussman contends, for problems formulated so that the process of reaching a solution depends only on the local interactions imposed on the interconnect geometry of the computing elements. The MIT scientists hope to identify an industrial partner who will work with them in designing an intelligent material on this basis.

Cellular engineering

Cellular engineering is a project that is closely related to amorphous computing and is also being developed at MIT. It seeks to understand and use the enormous number of biological system components that function on a molecular level to devise new technologies—most important, extremely efficient and compact computing mechanisms. By using biologists’ understanding of the genetic codes that determine an organism’s structure and function, it may be possible to utilize these biological codes as a basis for fabricating novel materials and structures on a molecular scale. One can also invent new computer codes while devising automated means to understand, construct, and debug them. Sussman, Hal Abelson, and Tom Knight lead this work at MIT’s Artificial Intelligence Laboratory. “We have recently worked out a biochemically plausible mechanism for constructing digital logic signals and gates of significant complexity within living cells,” Sussman says. “The resulting logic technology, although slow, allows us to engineer the chemical behavior of cells for use as sensors and effectors. One promising use of such technology is the control of fabrication processes at the molecular scale. We are now actively working with wet-lab biologists on testing these ideas and building cells that have circuits of our design in them.”

DNA computing

Yet another ultrascale effort involves a foray into the biological world for the purpose of DNA computing. DARPA funds several investigators in this area, including ones at the University of Southern California (USC) and Princeton University. The field is still emerging. Leonard



Adleman of USC performed the first experiments in 1994.

A DNA computer processes information by making and breaking the chemical bonds between separate strands of DNA. The binary code of 0’s and 1’s, which represent information in conventional computers by the flow of electrons through logic gates, is expressed in terms of the base pairs of DNA (formed by the nucleotide combinations cytosine–guanine and adenine–thymine). While ordinary computers perform calculations by using a program that instructs the electrons to travel particular paths, DNA calculations require synthesizing particular sequences of DNA and letting them react. The bonding that occurs between selected base pairs in the reaction is chosen to represent a certain algorithm, and thus serves as the information processing step.

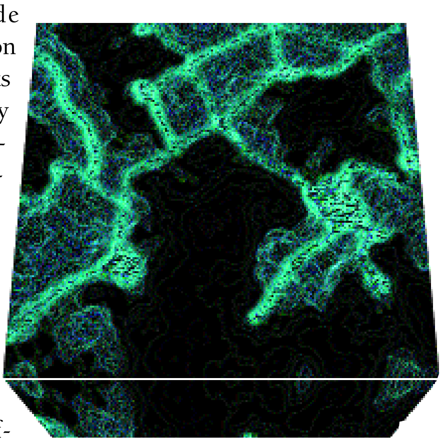
“You can imagine a DNA computation arranged as a series of test tubes containing DNA,” explains Princeton’s Richard Lipton. “Each tube is created from earlier ones by performing some biological operation, such as separating the strands, pouring one test tube into another, extracting those strands with a given pattern, heating or cooling the test tube, etc. The series of test tubes form a ‘single instruction multiple data’ computation, performed in parallel on the DNA. There is also an encoding method that maps binary strings of 0’s and 1’s into sequences of DNA base pairs.” For example, Lipton has devised the command “and” by separating DNA strands according to their sequences, while the command “or” is performed by pouring together DNA solutions containing specific sequences.

DNA promises computers that are a billion times as energy efficient as conventional computers, use a trillionth of the space to store information (DNA stores memory at a density of about 1 bit/nm^3, about a trillion times as efficient as videotape), and are massively parallel. Billions of molecules would undergo chemical reactions simultaneously, each reaction representing one operation. Some claim DNA can calculate at about 10^{14} operations per second, about 10 times faster than existing supercomputers. However, error-correcting codes are needed for these computations because biological operations are imperfect and information encoded in DNA decays at a finite rate. Applications include cryptography, computer-aided design, verifying the correctness of circuits or protocols, and factoring large numbers.

Cultured neural networks

Biology and electronics combine in a hybrid architecture for cultured neural networks, an ultrascale project at Science Applications International Corporation (SAIC). This initiative involves the controlled growth of mammalian neurons on silicon or other patterned substrates in order to study their suitability as logic circuits, either alone or interfaced with electronic circuits.

Both conventional computers and the brain process information via electrical signals that travel along a given pathway, which includes numerous points of interconnection known as logic gates in computers and synapses in the brain. By culturing living neurons on patterned substrates made from either silicon or glass, scientists at SAIC will study how closely a living neurobiological system can approximate an integrated circuit. The patterns for directing the neurons’ growth are prepared on self-assembled monolayers of organic molecules laid down upon substrates. The resulting cultured neurons will last long enough (at least several months) to enable control, positioning, and examination of their synaptic junctions. The overall goal is to control formation of synapses in such a way that they simulate several different kinds of electronic logic gates.

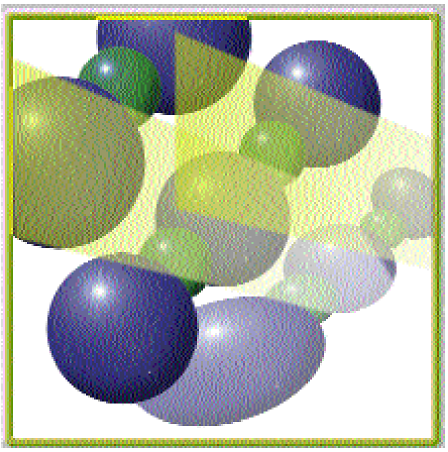


Quantum computing

Finally, the DARPA program looks to develop a practical quantum computer that processes information via transitions in quantum states of matter.

In the theory of quantum computing, the conventional representation of bits of information as a Boolean code of 0’s and 1’s is replaced by the quantum bit, called a qubit. Qubits represent the quantum states of matter of 0’s and 1’s associated with some simple atomic system, such as spin states (up or down) or the internal energy states of an atom. The state of a quantum computer at any given time is described by a wavefunction in Hilbert space. Information processing occurs as the quantum states undergo various changes known as unitary transformations. Other properties that differentiate quantum computing from classical computing are the superposition principle, whereby a quantum computer can exist in an arbitrary, complex, linear combination of classical Boolean states; “entanglement,” which means that information can be encoded in entangled (nonlocalized) quantum states; and decoherence, where the quantum state collapses when interacting with the environment.

DARPA funds several groups in this area. The Quantum Information and Computation Consortium (QUIC) of Caltech, USC, and MIT conducts both theoretical and



experimental work, along with simulations. Its quantum computer consists of an optical resonator where individual atoms become the qubits—with information encoded in their internal atomic states—while the interactions of the atoms with photons in the cavity form the basis for building an optical quantum logic gate.

A group at Harvard Medical School has built a quantum computer using nuclear magnetic resonance (NMR) spectroscopy, where the qubits are weakly polarized spin states in liquid samples and the unitary transformations occur via the standard NMR pulses. A team at Stanford University uses NMR with bulk samples for its quantum computer. Researchers at North Carolina State University (NCSU) are exploring the use of strongly coupled semiconductor quantum dots to make a quantum computer. According to NCSU's Bill Holton, "The use of quantum dots offers the possibility of designing a quantum computer that will be performance optimized and manufacturable in a modem integrated circuit factory."

QUIC theorist John Preskill of Caltech believes that, in principle, a commercial quantum computer can be built but will not be unless broad and important applications are found. According to Preskill, "a quantum computer does not have to have an 'exponential' speed advantage over a classical computer to be a useful device. The most important application we know of now is Grover's algorithm for searching an unsorted database. Unlike Shor's factoring algorithm, it provides only a relative speedup over classical algorithms, but is important because of its broad applications."

Preskill believes that "spectacular progress" has been made in quantum-error correction methods over the past 2 years (methods that correct errors in quantum information processing arising from the inherent instability of analog devices and decoherence) and that "these methods will impact technology in ways we cannot clearly anticipate at present."

Some industrial activity in quantum computing has occurred at IBM's Thomas Watson Research Center, AT&T Bell Laboratories, and Hewlett-Packard Laboratories in Bristol, England. "IBM's research policy is that it wishes to have 10 to 15% of its work in areas that have no impact on the bottom line in the foreseeable future, but which represent important areas for the future of information technologies. Quantum computing fits comfortably in that profile," explains IBM's David DiVincenzo, a leading researcher in the field.

These few corporate efforts consist primarily of theoretical work. To date, no company has made the investment needed to develop the experimental hardware for a quantum computer, although IBM's Almaden Research Center (San Jose, CA) does have a few exploratory projects. DiVincenzo believes that all current approaches to a practical quantum computer are viable. "They are all hard, and they will all struggle very slowly towards their goal," he predicts.

Developing a practical quantum computer will ultimately depend on two things—overcoming the decoherence problem and finding the "killer" application—says Richard Hughes, who heads a quantum computing group at Los Alamos National Laboratory. The more bits or qubits, the more one must worry about decoherence, or the collapse of the quantum states from contact with the environment. "Will the very large quantum states of multiple bits hold up for long enough to do anything interesting, from the computational point of view?" Hughes asks. "If this is achievable, then the second issue driving the development of a practical quantum computer is its applications. Right now cryptography looks to be a big driver, but there may also be algorithms developed for airline routing and network optimization."

Future prospects

Some in industry question the long-term potential for these technologies to ultimately replace the silicon chip. Consultant Robert Burger is a former vice president for research and chief scientist for the Semiconductor Research Corporation (SRC), a not-for-profit consortium created by the Semiconductor Industry Association to strengthen the competitiveness of the U.S. semiconductor industry. "None of these technologies [in the DARPA program] have demonstrated any more than an ability to replicate the functions of simple transistor circuits," he says. "An educated guess might be that they will not be competitive enough to replace any of today's integrated circuits for at least 20 years, and who knows where silicon technology will have advanced to in that period."

Whether any of the technologies in the Ultrascale Computing Program will take root in industry, even for very specialized uses, still remains an open question. However, one might recall Michael Faraday's now-famous remark regarding the utility of induced currents when asked by a well-meaning matron what good these discoveries were. Faraday answered: "Madam, one might as well ask of what good is a newborn babe."

BIOGRAPHY

Physicist Nancy Forbes focuses on information processing technology at Litton PRC, Inc., McLean, Virginia (forbes@bs1.prc.com).