6

INVENTING A NEW PHYSICS OF COMPUTING

More than other single technology, the microchip has driven the tremendous improvements in computing performance over the past sixty-plus years in everything from personal computing to consumer electronics to space exploration. Every two years, scientists develop techniques for putting twice as many transistors on a tiny chip of silicon, doubling the chip's capacity for storing or processing information. The microchip is arguably the most essential technology of the information age. Unfortunately, this rocket ship's ascent is slowing, and it's in danger of falling back toward earth. With each passing year, the expected performance gains become harder to achieve. And now scientists foresee a time, approximately a decade from now, when progress will slow to the point that it will begin to affect not only the computer industry but the whole of society.

The microchip combines two of the crucial scientific advances of the mid-twentieth century—the transistor and the integrated circuit. The transistor, invented in 1947 by John Bardeen, Walter Brattain, and William Shockley at AT&T's Bell Labs, provides the switching capability that's required for computation and data storage. The integrated circuit, invented separately by Jack Kilby at Texas Instruments and Robert Noyce at Fairchild Semiconductor, packs numerous transistors on a substrate of semiconducting material, enabling massive-scale production of a wide variety of tiny electronic devices. Over the years, silicon emerged as the semiconducting material of choice. In the 1960s, the microchip pioneer Gordon Moore set the tempo of an entire industry when he observed that the number of transistors on integrated circuits had been doubling roughly every year and predicted that the pace of progress would continue. Ultimately, he forecast that the density of transistors on ICs would double every two years—a statement that has been codified as Moore's Law.[1] Robert Dennard of IBM Research, who is best known for inventing the DRAM memory chip in 1967, also led the team that pioneered the concept of integrated circuit scaling. His rules provided a recipe for miniaturization that engineers have followed ever since. Today, after much innovation by a host of scientists, billions of transistors can be placed on a single chip. The work that Bob launched enabled Moore's Law to hold for decades.

But there's a problem: physics. One of the miracles of the miniaturization of silicon-based semiconductors has been that transistors actually perform better the smaller you make them. But because of the laws of physics, we have reached a point where transistor performance will increase only marginally or not at all with each new wave of miniaturization. Transistors have ceased to perform as perfect switches. Electrical charges "leak" from one part of the transistor to another when they're extremely small and the transistor does not perform well, which leads to a significant increase in the power consumption of current chips and might lead to more severe malfunctions in future chips. The densely packed chips can heat to near the point of failure—not a pleasant experience for the computer user.

Chip designers and manufacturers have resorted to a series of refinements and work-arounds to extend the life of Moore's Law and improve the performance of chips. In the past few years, designers have changed some of the materials they use to build transistors and have limited the clock rate—the speed at which electrical pulses race through chip circuitry. In order to make up for the loss of speed, they have divided elements of the processors into separate cores, where data processing is performed in parallel. Now, to counter the effects of the limits of physics, additional layers of silicon and metal are being added to create three-dimensional chips.

But as important as these advances are for enabling progress in microelectronics, they don't address the core problem: atoms don't "scale." No one can make traditional silicon devices whose essential elements approach the size of a few atoms. Also, in devices that are just a dozen or so atoms across, unpredictable things start to happen. They cross the border between classical physics and quantum mechanics. "We're very near the end of classical scaling.

It's a real discontinuity with huge implications," says Bob Dennard.[2]

What's required is a fundamental rethinking of microelectronics. The goal is to produce computer chips that are much more densely packed than today's and offer much more functionality, without being so power hungry. Toward that end, scientists are exploring a variety of approaches, ranging from developing new semiconductor materials that would replace the silicon transistor to designing new types of devices using the principles of quantum physics or new circuits and computing strategies that, for instance, take as their models the neural networks of the human brain. At IBM, we have a name for this endeavor: the search for a new physics of computing.

This journey will take computing from today's top microprocessors, with billions of transistors, to the nanosystems of the future, which will consume 100 to a 1,000 times less the power and be more efficient. Observers have predicted that some day we'll have chips with trillions of transistors, but IBM's top scientists believe that if we succeed with the most ambitious new approaches, it might actually require fewer switches to perform more work. When scientists achieve these breakthroughs, it will be as important a development as the invention of the transistor itself.[3]

This search for a new physics of computing is vital to the progress of electronics. We need to do more computing in smaller spaces using less electricity. But the effort is also necessary because of the new aspirations we have for computing. To make the most of the flood of big data, to enable massive interconnectivity between systems and people, and to produce the next advances in learning systems, we need vastly more capable computers. If we are to take on complex problems that involve the interplay of multiple social, technical, natural, and economic systems, we need computers that "think" differently at the nano level.

These efforts are all part of the quest to invent the switch that will replace the transistor. These are early days. Nobody knows for sure what the replacement will be. One approach that seems especially promising to researchers at IBM is a new semiconductor device category called piezotronics. The idea is to manipulate sound waves in a tiny piece of material that will serve as a switching element—the piezoresistive element, perhaps twenty nanometers thick. Applying voltages to those tiny elements changes their thickness. The change in resistance constitutes a change in state—the switch from zero to one that is the basis for computation. These devices, theoretically, could be packed densely on a chip and would require much less energy than today's transistors—about ten times less power.[4]

Taking a longer view, there's great potential in applying the principles of quantum mechanics to create a new kind of computational device. Quantum computing has been a Holy Grail for researchers ever since Nobel Prize–winning physicist Richard Feynman challenged the scientific community in 1981 to build computers based on quantum mechanics. A classical von Neumann–type computer makes use of bits, where each bit represents either a one or a zero. In contrast, a quantum bit, or qubit, can represent a one, a zero, or both at once. Therefore, two qubits can be in the states 00, 01, 10, and 11 at the same time. For each added qubit, the total number of potential states doubles. Hence, the use of qubits in certain kinds of computation could enable us to process exponentially larger quantities of data than is possible with the same number of conventional bits.[5]

For nearly two decades after Feynman issued his famous challenge, the field of quantum computing remained primarily theoretical. But in the 1990s and early 2000s, breakthroughs in algorithms and experimental techniques gave researchers hopes that what had seemed like one brilliant physicist's pipe dream could actually become reality. And now new achievements are being reported every few months. "The work we're doing shows it's no longer just a brute-force physics experiment. It's time to start creating computing systems based on this science," says Matthias Steffen, the manager of an experimental quantum computing project at IBM Research.[6]

Quantum computing has huge implications for the field of data encryption because quantum computers can, theoretically, factor large numbers like those used for making sensitive data indecipherable to prying eyes. So, the first application for quantum computing could be code breaking—just as it was for the first programmable computers in the 1940s. It would take the fastest contemporary supercomputers billions of years to factor numbers hundreds of digits long. But, theoretically, a quantum computer could perform these tasks in hours or even minutes.

While quantum machines seem unlikely to replace general-purpose computers any time soon, these techniques could affect other domains of computing in ways not yet foreseeable. An especially intriguing target is helping people to understand the interactions of complex systems of systems that underlie everything from the human body to cities to the global financial industry. One thing seems certain: only through fundamental breakthroughs in physics will we be able to deal with so much complexity and uncertainty on a planetary scale.

SCENARIO: DESIGNING PRODUCTS FROM THE MOLECULE UP

In 1935, scientists at the E. I. DuPont Co. invented a synthetic fabric, later named nylon, that became the first commercially successful synthetic polymer. Nylon was first used as a replacement for silk in women's stockings and parachutes but later came to be employed much more broadly in everything from clothing to gears in machines. The invention of nylon was a long and painstaking process that stretched for five years as a team of DuPont scientists experimented with one combination of molecules after another until they discovered a combo that had all the characteristics they desired in terms of usefulness, manufacturability, and affordability.[7] It was a triumph of ingenuity.

But what if scientists and engineers could use computers to help them develop new materials—indeed, whole products—from the molecule up? That way, they could much more quickly identify promising materials, mechanisms, and designs, potentially saving a tremendous amount of time and effort. If the computation, search, and learning capabilities of IBM's Watson could be applied not to searching for answers that are already known, as Watson does, but to a search for answers that are not yet known, the result would be a discovery machine.

It's possible that discovery machines like this will be commonplace in the new era of computing. Already, pharmaceutical companies use supercomputers to identify promising molecules in the drug-discovery process. And the U.S. Department of Energy's national laboratories use massively powerful racks of computers to simulate the behavior of fissionable materials in the design of nuclear weapons. But to be able to design products from the molecules up, computers will have to become much more capable and computation will need to become much more affordable than they are today.

In fact, these tasks will require several orders-of-magnitude improvements in processing power and energy efficiency. The machines will have to be 1,000 times more powerful than today's fastest supercomputers. If we were to try to achieve that level of performance using the chip designs and materials we have today, a computer would require about 100 megawatts of electricity to run, which would result in an annual electricity bill topping $100 million.[8] So the improvements we need can only be achieved through fundamental changes at computing's nanotechnology level.

But brute-force computation is just part of what will be needed to design products from the molecule up. Discovery machines will also require a combination of sophisticated simulation, machine learning, and analytics software to get the job done. The machines will be capable of ingesting all the scholarly articles, patent documents, and basic scientific knowledge pertaining to a specific area of enquiry. They'll learn to be expert chemists, materials scientists, and physicists. Based on what has been tried before and other knowledge they acquire, they'll simulate millions of potential combinations of molecules and make predictions about which combinations will achieve the desired results—then make recommendations to the real chemists, materials scientists, and physicists.

Consider a scenario from the energy industry. The photovoltaic cells in solar-energy panels are made from electronic circuitry combined with crystalline silicon. These cells are too expensive to enable the kind of mass deployment that will be necessary if solar energy is to play a major role in moving the world away from fossil fuels. But what if you could develop solar cells out of much cheaper organic materials? A discovery machine could help by coming up with a way to capture energy from light by using an organic dye like chlorophyll, which is the molecule that allows plants to absorb energy from light during photosynthesis. Because every place on earth has different solar characteristics, the machine would have to come up with recipes for thousands of organic dyes that could capture the maximum amount of energy in each location. It's mass invention at the nano-scale.

Jed Pitera, an IBM researcher who specializes in using computer simulation to address fundamental questions about biology and chemistry, has an interesting take on the potential for discovery machines. Because they aren't humans, he says, they won't be limited by the prejudices that human engineers and scientists are burdened with when the process of discovery takes them into the realm of nanotechnology. "At the nano-scale, a lot of what people think of as normal or intuitive tends to be wrong or incomplete," Jed says. "If we can build a cognitive computing system that starts with a blank slate, it could develop a better understanding of how things work in that world."[9]

So far, we've been talking about chemistry and physics at the molecular level, but there's another dimension of computer-aided discovery: invention at the macro-scale. Eventually, using a combination of high-performance computation and data analytics, we will be able to simulate the operations of entire physical systems, such as a jet engine or an entire aircraft. Products like these are made up of countless materials and have multiple simultaneous physical processes. But the capability to simulate the behavior of these systems while they're still on the drawing boards could cut years off the time it takes to develop a new aircraft, nuclear power plant, or submarine.[10]

When companies can design entire products from the molecules up, it will revolutionize innovation and production, remaking entire industries and roiling economies. But in order to be able to design products this way, scientists will first have to design new computing components from the *atoms* up.

JOURNEY OF DISCOVERY: THE NEW PHYSICS OF COMPUTING

On March 2, 2011, Andreas Heinrich and his team of scientists at IBM Research–Almaden accomplished a major breakthrough in nanotechnology. They produced the smallest device that can be used to reliably store a single bit of magnetic information—albeit at low temperature. (A bit is either a one or a zero.) It comprised just twelve atoms.[11] Today's disk drives use about one million atoms to store a bit. The team's breakthrough has the potential to make a big difference in the realm of data storage, making it possible, for instance, to store an entire movie and music collection on a charm-sized pendant. In addition, the techniques could be used by scientists who are exploring quantum computing

Andreas was thrilled by the discovery. Sitting in his team's tiny laboratory in a building perched on top of California's Santa Cruz Mountain range, he pressed the button on his computer mouse to reverse the magnetic states of the twelve atoms and stared at a small computer screen where the atoms showed up graphically as small bumps on a flat surface. When the structure received an electrical pulse from the tiny probe tip of the microscope, some atoms grew visibly taller while others shrank, based on their antiferromagnetic arrangement. Andreas switched the magnetic state of the atoms back and forth for nearly four hours. "It was basically, 'Wow, this works.' Do it again. 'Wow, this works,' " he recalls.[12] Such are the joys of a scientist working at the edge of the known physical universe.

Andreas and his work illustrate another key element of what we call the new culture of innovation: the willingness to invest a tremendous amount of time and energy in a project, to go deep on the science and see what it teaches us. This approach to research produced the transistor and many other advances that came from corporate laboratories in the twentieth century. But corporate labs have withered as one company after another cut funding for long-term, high-risk projects. At IBM, we believe that it's essential that corporations, working with their partners in government and universities, restore their commitment to fundamental scientific research—to boldness and persistence.

IBM has operated at the front edge of nanotechnology since the field's earliest days. The first breakthrough came at IBM Research–Zurich in 1978 after the lab hired the recent Ph.D. Gerd Binnig and teamed him with the senior researcher Heinrich Rohrer. Within a few months they got the idea for the scanning tunneling microscope, which can detect individual atoms by scanning over a surface with a tiny electronic probe. The instrument takes advantage of a phenomenon called tunneling, which causes electrons to jump from the atoms on the surface to the apex of a sharp tip placed in very close proximity. By measuring the amount of tunneling current that occurred, the scientists could plot a picture of the surface and the atoms positioned on top of it.[13] For the first time, scientists could "see" landscapes composed of individual atoms. This was the birth of nanotechnology. For their efforts, Gerd and Heinrich won the Nobel Prize for Physics in 1986.

Another of Gerd's inventions, the atomic force microscope, allowed scientists to observe atoms within insulator (nonconducting) materials and also, in a sense, to see inside molecules. IBM researcher Gerhard Meyer, also at the Zurich lab, is the senior member of a team that uses the AFM to directly observe the chemical structure of a molecule by imaging the chemical bonds and the electrical charge distribution within the molecule. With this capability, scientists are able to study chemistry on a single-molecule level, holding great promise for atomic-scale computing and energy storage.

Moving atoms came in 1989. A team of scientists at IBM Research–Almaden headed by Don Eigler made improvements to the STM, and Don became the first person to manipulate an individual atom. A few weeks later, in a marathon twenty-two-hour session in the lab, he arranged thirty-five xenon atoms to spell out "IBM." While the moving of a single atom was remarkable in itself, Don later commented that there was an even more important result: "The biggest change was it changed people's perspective. We showed that there was a world out there made of small things, which we had never had access to, but we now have direct access to. It changes the way we think about things."[14]

That world of small things now looms large on the human landscape. Advances like this don't usually come quickly or predictably. Andreas had been experimenting with the STM for ten years when the team achieved the big magnetic-storage breakthrough. His most important attribute, he says, is persistence. "Nobody tells us what to pursue. We can explore what we want. The atoms tell us what they want to do."[15]

Andreas's persistence originally won him his job at IBM. While studying for his Ph.D. in physics from Germany's Göttingen University, he was so determined to land a job on Don Eigler's team at the Almaden lab that he kept trying even after Don turned him down twice. Each time, he proposed a new experiment using the STM. He finally won Don over with a proposal to study the spin effects of atoms on surfaces. Spin is a quantum-mechanical property of particles such as electrons, a phenomenon that could prove useful in next-generation computing. Instead of switching the states of electronic components by applying electrical pulses, you might be able to do it by changing the spin, which would take much less energy.

In order to do these experiments, the team had to build another STM with new capabilities—a project that took nearly two years. Over time, Andreas's research focused began to focus on the smallest number of atoms that could retain the stable magnetic characteristics of classical physics. The creation of the smallest device that could store a bit of magnetized data was a byproduct of the team's study, not the goal. That's what Andreas means when he says the atoms talk to him.

The newest member of the STM team is Susanne Baumann, a twenty-five-year-old Swiss Ph.D. candidate. She has the same intensity and persistence as Andreas. Susanne recalls the day in second grade that she first became obsessed with science. She was reading a book that contained a description of atoms and a photograph of Don Eigler's tiny "IBM" logo. Susanne remembers running in shock to her mother and demanding to know: "Atoms are in the air and in me and in metal. How can this be?" Years later, when she was studying for her master's degree, she began pestering Don and, later, Andreas, about giving her a job in the lab. (After Don retired from IBM, Andreas succeeded him as head of the STM team.) "I even offered to find my own money for the stay," she says. Finally, Andreas relented—based on the quality of Susanne's proposal: to experiment with ways to make it easier for the STM to differentiate the atoms on different parts of the surface. This research, she hopes, will help the group make progress toward quantum computing.

That's the team's new quest. Now that they have discovered that as few as twelve atoms can combine and retain the effects of classical physics, they're moving in the opposite direction and asking how many atoms can be combined while still achieving the behavior of quantum mechanics. And they want to know how to make the quantum effects more pronounced and how to control them. The goal is to control a single quantum bit on a surface within three years. Once they can do that, they believe they'll be able to assemble devices from qubits that can potentially be used for computing.

Andreas and his team are in the early stages of their exploration. It could take them ten years or more to produce a working prototype for quantum computing. But since they know how to build structures from the atom up they have some advantages over scientists pursuing other approaches. If they get this right, they may be able to develop practical techniques that could be used for manufacturing massive quantities of quantum-mechanical devices.[16] Those ultra-tiny devices could become the equivalent of today's microchips for the next era of computing.