The world today is very different from that of just a decade ago, thanks to our ability to readily access enormous quantities of information. Tools that we take for granted—social networks, Internet search engines, online maps with point-to-point directions, and online libraries of songs, movies, books and photographs—were unavailable just a few years ago. We owe the arrival of this information age to the rapid development of remarkable technologies in high-speed communications, data processing and—perhaps most important of all but least appreciated—digital data storage.

Each type of data storage has its Achilles’ heel, however, which is why computers use several types for different purposes. Most digital data today, such as the information that makes up the Internet, resides in vast farms of magnetic hard disk drives (HDDs) and in the HDDs of individual computers. Yet these drives, with their rotating disks and moving read/write heads, are unreliable and slow. Loss of data because of so-called head crashes occurs relatively often. Regarding speed, it can take up to 10 milliseconds to read the first bit of some requested data. In computers, 10 milliseconds is an eon—a modern processor can perform 20 million operations in that time.

That is why computers use a second type of storage, solid-state memory, for their computational operations. Solid-state memories read and write data with great speed, enabling swift processing. High-performance versions, such as static and dynamic random-access memory (SRAM and DRAM, respectively), use the electronic state of transistors and capacitors to store data bits. These chips lose their data, however, when the computer powers down—or crashes.

A few computers use nonvolatile chips, which retain data when the power is off, as a solid-state drive in place of an HDD. The now ubiquitous smart cell phones and other handheld devices also use nonvolatile memory, but there is a trade-off between cost and performance. The cheapest nonvolatile memory is a kind called flash memory, which, among other uses, is the basis of the little flash drives that some people have hanging from their key rings. Flash memory, however, is slow and unreliable in comparison with other memory chips. Each time the high-voltage pulse (the “flash” of the name) writes a memory cell, the cell is damaged; it becomes unusable after only perhaps 10,000 writing operations. Nevertheless, because of its low cost, flash memory has become a dominant memory technology, particularly for applications in which the data will not be changed very often.

The computing world is thus crying out for a memory chip with high data density that is also cheap, fast, reliable and nonvolatile. Research groups around the world are investigating several approaches to meet this demand, including systems based on new electronic components called memristors [see box on page 81] and others making use of spintronics, in which the spin, or magnetism, of electrons plays a key role [see “Spintronics,” by David D. Awschalom, Michael E. Flatté and Nitin Samarth; SCIENTIFIC AMERICAN, June 2002].

The answer may lie in a new kind of spintronic chip called racetrack memory (RM), which I
BITS ZOOMING along nanoscopic racetracks are the basis of a proposed new kind of solid-state memory that would combine all the best features of hard disk drives and present-day memory chips and none of the bad.
Racetrack memory stores data in the form of magnetized regions, or domains (red and blue), in nanoscopic wires on a silicon substrate. The nanowires may lie horizontally (below) or stand as vertical columns (right). Pulses of electric current move the magnetic domains rapidly along the racetracks past heads for reading and writing data.

Drawbacks of Disk Drives

The basic structure of an HDD has not changed since its inception in the 1950s, although the technology of individual components has altered enormously, in particular shrinking by many orders of magnitude. An HDD stores data as the directions of magnetization of tiny regions in an ultrathin layer of magnetic material coating the surface of a highly polished glass disk. The disk rotates at high speed (commonly 7,200 revolutions per minute in computers currently sold) under a recording head on a moving arm that reads and writes the magnetic bits.

In the early decades HDDs were refrigerator-size devices and the cost per stored bit was very high. The figure of merit for a disk technology is its areal density: the number of data bits reliably stored per unit area of the magnetic surface. At first areal densities of disk platters improved by only about 25 percent each year, but beginning in the late 1980s HDDs rapidly morphed into much more compact and capacious machines.

An important milestone in this evolution was the development of read heads exploiting spintronics, or what I like to call spin-engineered materials. My research in the period of 1988 to 1991 into the fundamental properties of materials constructed of multiple magnetic nanolayers led to development of the spin-valve magnetoresistive sensor. This sensor detects tiny magnetic fields as a change in its resistance, and at the time of its invention it was the most sensitive detector of such fields at ambient temperatures.

The first use of spin-valve sensors in HDD read heads came in 1997 with IBM’s Deskstar 16GP “Titan.” Within five years HDD storage capacities had increased 1,000-fold, the rapidest advance in the half-century-long history of HDDs. Today the collective storage capacity of all HDDs manufactured in one month exceeds 200 exabytes, or $2 \times 10^{20}$ bytes—enough to store all the extant analog data in the world, that is, all the data on paper, film and videotape.

The spin-valve sensor was the first spintronic nanodevice, and knowing some spintronics is essential to understanding how RM works. Spin is a fundamental quantum property of electrons. Imagine each electron as a tiny spinning ball of electric charge, with a magnetic field pointing along the axis of the spin. The spin axis of an
electron in an ambient magnetic field lines up either parallel or antiparallel to the field. It is said to have either “spin up” or “spin down,” with respect to the local magnetic field.

When electrons travel through a magnetized metal, the spin-up electrons travel more easily, resulting in a spin-polarized current or spin current—one in which most of the moving electrons carry a specific spin. In contrast, an ordinary current, such as one traveling along copper wire, involves electrons whose spins point randomly in all directions. Permalloy, a strongly magnetic alloy of nickel and iron, can produce as much as 90 percent spin polarization in a current.

The spin-valve sensor consists of a nanosandwich, a layer of nonmagnetic metal between two magnetic layers. The first magnetic layer spin-polarizes the current in a specific direction. The second magnetic layer changes its magnetism back and forth to match the field coming from each passing magnetic domain representing a 0 or a 1 on a disk. When the two magnetic layers of the sensor are parallel, the spin-polarized current flows through relatively easily. When the layers are antiparallel, the polarized electrons are impeded. The changing resistance of the device is known as giant magnetoresistance, a phenomenon independently discovered in 1988 by the groups of Albert Fert of the University of Paris–South and Peter Grünberg of the Jülich Research Center in Germany. Giant magnetoresistance allows read heads to detect much weaker fields, which in turn allows magnetic domains on a disk to be much smaller and more tightly packed.

Yet the era of the spin-valve sensor has lasted no more than a decade. A newer spintronic technology known as magnetic tunnel junctions has already replaced it in the HDDs manufactured today. Magnetic tunnel junctions exploit an effect called tunneling magnetoresistance to achieve even greater sensitivity to small magnetic fields than spin-valve devices [see “Magnetic Field Nanosensors,” by Stuart A. Solin; SCIENTIFIC AMERICAN, July 2004].

Although spintronic read heads have enabled vast increases in the storage capacity of HDDs and helped to bring the cost of storing data down to about 10 cents per gigabyte, the basic mechanical nature of an HDD’s rotating disk and moving read head remains, leading to two major deficiencies. First, a head crash occurs when the recording head accidentally strikes the magnetic layer, thereby damaging it, and can result in loss of all data in the HDD. Second, it takes a lot of energy to spin a glass disk at 7,200 rpm, and even at such a pace, rotating the disk to the data of interest takes millions of times longer than accessing data from volatile memory. As a result, HDDs are very inefficient for many quite ordinary applications, such as recording transactions to bank accounts—each transaction may involve a very small amount of data, yet it takes time to rotate the disk and move the write head to the correct location, and backup copies must be made in case of a head crash.

Persistence of Memory

Researchers have devoted much time and effort in recent years toward developing types of nonvolatile memories that might combine the good features of HDDs and silicon chips while avoiding all the bad. In 1995, for example, my colleagues at IBM and I proposed building a spintronic memory based on magnetic tunnel junctions. Data are stored in the magnetic state of the magnetic tunnel junction and can be read using the tunneling magnetoresistance of the device. These magnetic random-access memories, or MRAMs, went on sale in 2006 from Freescale Semiconductor, a spin-off of Motorola.

Many other proposed memory devices involve a component whose resistance changes for one reason or another. All of them, however, require a transistor connected in series with every resistive memory element to access each selected bit. The transistor size largely determines the memories’ cost. Despite tremendous advances, the cheapest solid-state memory, flash, remains 20 to 100 times as expensive per bit as HDD.

The trade-off in the cost and performance of HDDs and the various types of solid-state memories mean that it makes sense to build computers that use many different technologies for storing digital data. Consequently, volatile RAM holds data in active use by programs, and HDDs serve as temporary stores of excess data that will not fit in the RAM, as well as for long-term storage of files and programs for when the computer is turned off or crashes. Nonvolatile and read-only memories are also used for special purposes.

This assortment of technologies makes computers and related devices more complex and bulkier than they need be, as well as much more energy-consuming. A memory storage device that had the nonvolatility and the low cost of an HDD, along with the fast reading and writing and the high reliability of conventional solid-state memories, would be a game-changing technology. My RM design can be that device.

Each racetrack consists of a nanoscopic wire
Domain walls can move **150 nanometers** in a nanosecond, allowing access millions of times faster than hard disk drives.

A racetrack stores bits as a series of magnetized domains along its length. The domains may point one way along the wire to represent 0s and the other way to represent 1s. Just as with an HDD, such domains retain their state when the power is off.

Unlike an HDD, however, the magnetic medium never has to move. Instead the bits themselves travel back and forth along their racetrack, passing a nanoscopic read head and write head along the way. Thus, the hundreds of bits in each racetrack require only a few transistors, instead of a transistor for every bit as in conventional solid-state memory designs.

This idea of moving magnetic data through a medium instead of moving the medium itself is in some respects an old one. Bubble memory, which had its heyday in the 1970s, also involved the motion of small magnetic domains (the “bubbles”), but disk drives and solid-state memories outcompeted it by relentlessly shrinking and becoming faster. The bubbles were moved by a complicated system of magnetic fields. RM uses a much simpler, spintronic motive force.

The key to moving magnetic domains is the domain wall that exists wherever two domains with different directions of magnetization meet. In RM nanowires, a domain wall exists anywhere that a 0 is next to a 1. The customary way to move domain walls involves applying a magnetic field. The magnetization in each domain actually comes about because the atoms in the domain have their individual magnetism aligned. In a sufficiently strong external field aligned with one of the domains, the antiparallel atoms at the domain wall tend to flip around to line up with the applied field—and so the position of the wall shifts. Unfortunately, this process does not move the data bits along the nanowire. Consider a 0 sitting between a pair of 1s, with the applied field pointing in the 1 direction. The two domain walls will move to increase the size of the 1 domains, eventually wiping out the 0 altogether.

The spintronic trick to moving the domain walls uniformly along the racetrack is as simple as sending an electric current along the nanowire. Again consider the 1-0-1 arrangement of domains. The electrons flowing through the first domain will be spin-polarized with their own magnetism aligned in the 1 direction. As each electron crosses the 1-0 domain wall, its magnetism tends to flip to the 0 direction. But the electron’s magnetism is tied to its spin, which is a quantity of angular momentum. As with energy and ordinary momentum, angular momentum is a conserved quantity. For the electron to flip from 1 to 0, something else must flip from 0 to 1, and that something else is an atom just on the 0 side of the domain wall. As the spin-polarized electrons flow through the domain wall, they move the domain wall along the nanowire one atom at a time.

Now consider what happens when one of these same electrons arrives at the 0-1 domain and crosses it. The same reasoning shows that it flips from 0 back to 1, which flips an atom from 1 to 0, again moving the domain wall a tiny amount along the wire in the direction the electron is flowing. With both its domain walls moving along the wire in lockstep, the 0 bit itself travels without expanding or shrinking. So far my group has demonstrated in the lab that nanosecond-long pulses of spin-polarized current indeed cause a series of as many as six domain walls to move in lockstep along magnetic nanowires. The domain walls can move 150 nanometers in a nanosecond, allowing access times of nanoseconds, millions of times faster than HDDs and comparable to volatile memories.

The domain walls could easily drift out of po-
sition, however, perhaps propelled by small stray currents or magnetic fields or because the controlling pulses are not exactly the correct magnitude and duration. This hazard can be averted by building small notches in the racetrack’s sides, spaced at the intended size of the bits. The domain walls tend to become pinned at these notches because they will have the smallest area and thus the least energy when sitting at a notch. Tiny stray currents will not suffice to move a domain wall away from a notch, and slightly imperfect control pulses will still move a wall from one notch to the next one, no more and no less.

A few different kinds of domain wall may occur. So-called transverse walls are relatively simple and much like the kind of wall you may be imagining based on my description above. Vortex walls, in contrast, have a complicated, swirling pattern of magnetization, including a little “core” in their center. A smaller current suffices to move a vortex wall because moving the core pulls the whole wall along. If the 0 and 1 domains are magnetized perpendicular to the racetrack instead of along it, only one relatively simple kind of domain wall occurs. In principle, this kind of magnetization should have advantages because the race-track can be narrower and the domain walls should require less current to move.

With racetracks arranged as forests of vertical columns rising above the surface of a silicon wafer, the memory becomes three-dimensional, greatly increasing the density of data storage compared with the horizontal design. In contrast, both HDDs and silicon-based microelectronic memory (and logic) devices are fundamentally two-dimensional in nature, and many analysts predict the pace of their evolution to smaller sizes will run into fundamental physical problems in as little as a decade from now.

My group has built some vertical racetracks, but we have not yet integrated devices on the same chips for writing or reading domain walls. Constructing robust vertical racetracks on silicon chips at a commercially viable cost may be a significant challenge. Most of my group’s experiments are with nanowires placed horizontally on the silicon surface. The storage capacity of this mode of RM would be merely comparable to flash memory, though with significant advantages: the memory would be much faster than flash, would use less energy and would not wear out. We are currently working on a prototype that will exploit our ability to move half a dozen domain walls along each racetrack.

RM faces stiff competition from storage-class memories under development that use more traditional approaches [see box above]. Research groups are trying to turn their chosen approach into a superior commercially viable product. Stay tuned to see which technology crosses the finish line at a store near you.

Emerging Memories

Researchers are working on several emerging technologies to pursue storage-class memories with a more traditional design than that of racetrack memory—placing the bits in horizontal arrays on a silicon wafer. Achieving capacities comparable to vertical RM or hard disk drives would require stacks of these arrays.

RESISTIVE RANDOM-ACCESS MEMORY (RRAM) uses materials that can be switched between two or more distinct resistance states. Many companies are investigating metal oxide nanolayers switched by voltage pulses. Researchers generally think that the pulses’ electric fields produce conducting filaments through the insulating oxide. HP Labs plans to release prototype chips this year based on “memristors,” in which migrating oxygen atoms change the resistance.

PHASE-CHANGE RAM (PRAM) has memory elements made of chalcogenide glass, the same class of material as used in rewritable CDs and DVDs. The glass has a crystalline state with low resistance and an amorphous state with high resistance. A current pulse sets the state by heating the chalcogenide and then allowing it to cool either rapidly or slowly. In 2006 BAE Systems introduced a 512-kilobyte PRAM radiation-hardened for space applications. Numonyx, a spin-off of Intel Corporation and STMicroelectronics, began commercial sales of its 16-megabyte “Alverstone” chip in late 2008.

SPIN TORQUE TRANSFER RAM (STT-RAM) is a new kind of magnetic RAM (MRAM). MRAM stores data as the magnetization direction of each bit. Nanoscopic magnetic fields set the bits in conventional MRAM, but STT-RAM uses spin-polarized currents, enabling smaller and less energy-consuming bits. Companies developing STT-RAM include EverSpin, Grandis, Hynix, IBM, Samsung, TDK and Toshiba.

MORE TO EXPLORE

