THE HIGH-*k* SOLUTION

MICROPROCESSORS COMING OUT THIS FALL ARE THE RESULT OF THE FIRST BIG REDESIGN IN CMOS TRANSISTORS SINCE THE LATE 1960S

BY MARK T. BOHR, ROBERT S. CHAU, TAHIR GHANI & KAIZAD MISTRY



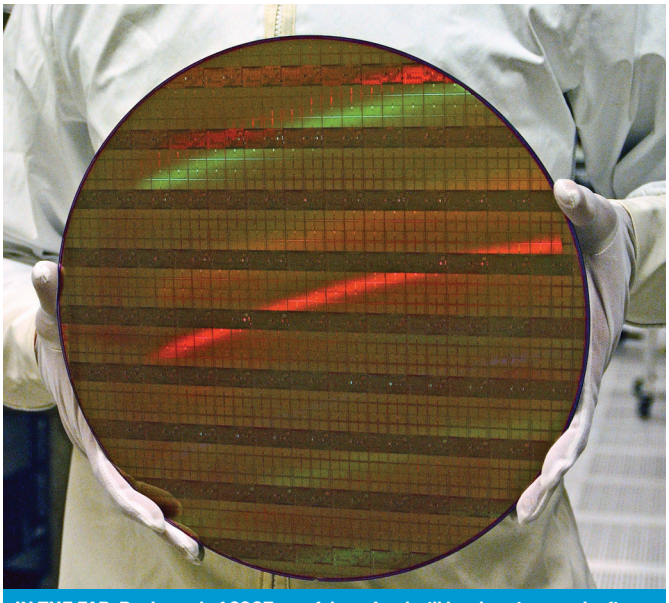
From left: Ghani, Mistry, Chau, and Bohr of Intel with a wafer of 45-nanometer microprocessors

As you read this, two of our most advanced fabs here at Intel are gearing up for the commercial production of the latest Core 2 microprocessors, code-named Penryn, due to start rolling off the lines before the year is up. The chips, based on our latest 45-nanometer CMOS process technology will have more transistors and run faster and cooler than microprocessors fabricated with the previous, 65-nm process generation. For compute-intensive music, video, and gaming applications, users will see a hefty performance increase over the best chips they are now using.

A welcome development but hardly big news, right? After all, the density of transistors on chips has been periodically doubling, as predicted by Moore's Law, for more than 40 years. The initial Penryn chips will be either dual-core processors with more than 400 million transistors or quad-core processors with more than 800 million transistors. You might think these chips don't represent anything other than yet another checkpoint in the inexorable march of Moore's Law.

But you'd be wrong. The chips would not have been possible without a major breakthrough in the way we construct a key component of the infinitesimal transistors on those chips, called the gate stack. The basic problem we had to overcome was that a few years ago we ran out of atoms. Literally.

To keep on the Moore's Law curve, we need to halve the size of our transistors every 24 months or so. The physics dictates that the smallest parts of those transistors have to be diminished by a factor of 0.7. But there's one critical part of the transistor that we found we couldn't shrink anymore. It's the thin layer of silicon dioxide (SiO_2) insulation that electrically isolates the transistor's gate from the channel through which current flows when the transistor is on. That insulating layer has been slimmed and shrunk with each new generation, about tenfold since the mid-1990s alone. Two generations before Penryn, that insulation had become a scant five atoms thick.



IN THE FAB: By the end of 2007, two fabs at Intel will be churning out the first commercial microprocessors made up of transistors fundamentally redesigned using new materials.

We couldn't shave off even one more tenth of a nanometer—a single silicon atom is 0.26 nm in diameter. More important, at a thickness of five atoms, the insulation was already a problem, wasting power by letting electrons rain through it. Without a significant innovation, the semiconductor industry was in danger of encountering the dreaded "showstopper," the long-awaited insurmountable problem that ends the Moore's Law era of periodic exponential performance gains in memories, microprocessors, and other chips—and the very good times that have gone with it.

The solution to this latest crisis involved thickening the insulator with more atoms, but of a different type, to give it better electrical properties. This new insulator works well enough to halt the power-sucking hail of electrons that's plagued advanced chips for the past four years. If Moore's Law crumbles in the foreseeable future, it won't be because of inadequate gate insulation. Intel cofounder Gordon Moore, of Moore's Law fame, called the alterations we made in introducing this latest generation of chips "the biggest change in transistor technology" since the late 1960s.

As difficult as finding the new insulator was, that was only half the battle. The point of the insulator is to separate the transistor's silicon gate from the rest of the device. The trouble is, a silicon gate didn't work with the new insulator material. The initial transistors made with them performed worse than older transistors. The answer was to add yet another new material to the mix, swapping the silicon gate for one made of metal.

It may not seem like such a big deal to change the materials used in a transistor, but it was. The industry went through a major upheaval several years ago when it switched from aluminum interconnects to copper ones and—at the same time—from SiO_2 cladding for those interconnects to chemically similar "low-*k*" dielectrics. And those changes had nothing to do with the transistor itself. A fundamental change to the composition of the transistor is pretty much unheard of. The combination of the gate and the insulator, the gate stack, hasn't changed significantly since Moore, Andrew S. Grove, and others described it in this magazine back in October 1969!

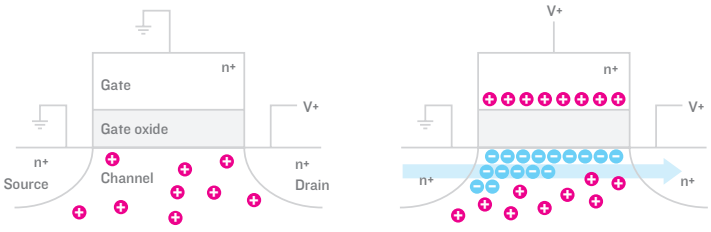
So when you boot up your next machine and you're surprised by how fast it rips through some video coding, remember: there's more new under its hood than in any computer you've ever owned.

The story of how we and our co-workers solved the gate-insulation problem may seem esoteric, and in a literal way it is. But it is also emblematic of how Moore's Law, the defining paradigm of the global semiconductor industry, is being sustained against often-daunting odds by the swift application of enormous intellectual and material resources to problems that, increasingly, are forcing engineers to struggle in realms until recently occupied only by physicists.

THE PROBLEM, ULTIMATELY, IS ONE OF POWER. At five atoms, that sliver of SiO_2 insulation was so thin that it had begun to lose its insulating properties. Starting with the generation of chips fabricated in 2001, electrons had begun to trickle through it. In the processors made just two years later, that trickle became some 100 times as intense.

All that current was a drain on power and a source of unwanted heat. Laptops were heating up too much and draining their batteries too quickly. Servers were driving up their owners' electric bills and taxing their air conditioners. Even before we

PREVIOUS PAGE: AARON HEWITT; LEFT: INTEL; ILLUSTRATIONS: BRYAN CHRISTIE DESIGN



THE TRANSISTOR: A positive voltage on the gate of an NMOS transistor drives positive charge in the channel away from the insulating gate oxide and attracts electrons, forming a path for electrons to flow.

ran out of atoms, designers had devised some tricks to throttle back on the power without losing speed. But without a way to stanch the unwanted flow of electrons through that sliver of insulation, the battle to make ever more powerful processors would soon be lost.

To understand why, you need a quick lesson (or refresher) in semiconductor basics. The type of transistor that is chained together by the hundreds of millions to make up today's microprocessors, memory, and other chips is called a metal-oxide-semiconductor field effect transistor, or MOSFET. Basically, it is a switch. A voltage on one terminal, known as the gate, turns on or off a flow of current between the two other terminals, the source and the drain [see illustration, "The Transistor"].

MOSFETs come in two varieties: N (for *n*-type) MOS and P (for *p*-type) MOS. The difference is in the chemical makeup of the source, drain, and gate. Integrated circuits contain both NMOS and PMOS transistors. The transistors are formed on single-crystal silicon wafers; the source and drain are built by doping the silicon with impurities such as arsenic, phosphorus, or boron. Doping with boron adds positive charge carriers, called holes, to the silicon crystal, making it *p*-type, while doping with arsenic or phosphorus adds electrons, making it *n*-type.

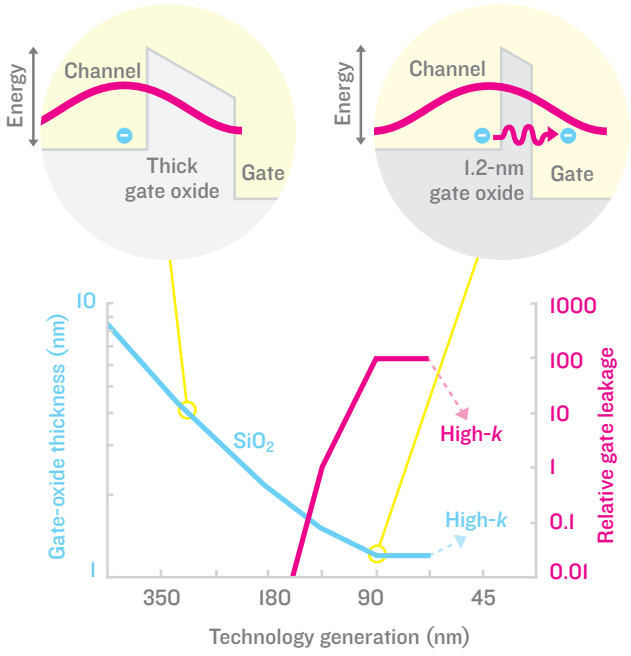
Taking an NMOS transistor as an example, the shallow source and drain regions are made of highly doped *n*-type silicon. Between them lies a lightly doped *p*-type region, called the transistor channel—where current flows. On top of the channel lies that thin layer of SiO_2 insulation, usually just called the gate oxide, which is the cause of the chip industry's most recent technological headaches.

Overlying that oxide layer is the gate electrode, which is made of partially ordered, or polycrystalline, silicon. In the case of an NMOS device it is also *n*-type. (The silicon gates replaced aluminum gates—the metal in "metal-oxide semiconductor"—in work described in the 1969 *IEEE Spectrum* article. But the "MOS" acronym has nevertheless lived on.)

The NMOS transistor works like this: a positive voltage on the gate sets up an electric field across the oxide layer. The electric field repels the holes and attracts electrons to form an electron-conducting channel between the source and the drain.

A PMOS transistor is just the complement of NMOS. The source and drain are *p*-type; the channel, *n*-type; and the gate, *p*-type. It works in the opposite manner as well: a positive voltage on the gate (as measured between the gate and source) cuts off the flow of current.

In logic devices, PMOS and NMOS transistors are arranged so that their actions complement each other, hence the term CMOS for complementary metal-oxide semiconductor. The arrangement of CMOS circuits is such that they are designed to draw power only when the transistors are switching on or off. That's the idea, anyway.



RUNNING OUT OF ATOMS: The transistor's gate oxide thinned with each new technology generation until it reached just 5 atoms (1.2 nm) thick. At that scale, the wave describing the probable location of an electron [red curve, top] is broader than the gate oxide, and the electron can simply appear on the other side of the gate oxide, having tunneled through the insulation. This so-called gate leakage increased 100-fold in the last three generations of transistors. A switch to a new gate oxide, a high-*k* dielectric, was needed to plug the leak.

Although the basic features and materials of the MOS transistor have stayed pretty much the same since the late 1960s, the dimensions have scaled dramatically. The transistor's minimum layout dimensions were about 10 micrometers 40 years ago, and are less than 50 nm now, smaller by a factor of more than 200. Suppose a 1960s transistor was as big as a three-bedroom house and that it shrank by the same factor. You could hold the house in the palm of your hand today.

In the Penryn processors that we recently began fabricating, most of their transistors' features measure around 45 nm, though one is as small as 35 nm. It's the first commercial microprocessor to have features this small; all other top-of-the-line microprocessors in production as this article is being written have 65-nm features. In other words, Penryn is the first of the 45-nm generation of microprocessors. Many more will soon follow.

The thickness of the SiO_2 insulation on the transistor's gate has scaled from about 100 nm down to 1.2 nm on state-of-the-art microprocessors. The rate at which the thickness decreased was steady for years but started to slow at the 90-nm generation, which went into production in 2003. It was then that the oxide hit its five-atom limit. The insulator thickness shrank no further from the 90-nm to the 65-nm generation still common today.

The reason the gate oxide was shrunk no further is that it began to leak current [see illustration, "Running Out of Atoms"]. This leakage arises from quantum effects. At 1.2 nm, the quantum nature of particles starts to play a big role. We're used to thinking of electrons in terms of classical physics, and we like to imagine an electron as a ball and the insulation as a tall and narrow hill. The height of the hill represents how much energy you'd need to provide the electron to get it to the other side. Give it a sufficient push and—sure enough—you could get it over the hill, busting through the insulation in the process.

But when the hill (the oxide layer) is so narrow that you are counting individual atoms of thickness, the electron looks less like a ball and more like a wave. Specifically, it's a wave that defines the probability of finding the electron in a particular location. The trouble is that the wave is actually broader than the hill, extending all the way to the other side and beyond. That means there is a distinct probability that an electron that should be on the gate side of the oxide can simply appear on the channel side, having "tunneled" through the energy barrier posed by the insulation rather than going over it.

IN THE MID-1990S, WE AT INTEL and other major chip makers recognized that we were fast approaching the day when we would no longer be able to keep squeezing atoms out of the SiO_2 gate insulator. So we all launched research programs to come up with a better solution. The goal was to identify a gate dielectric material as a replacement for SiO_2 and also to demonstrate transistor prototypes that leaked less while at the same time driving plenty of current across the transistor channel. We needed a gate insulator that was thick enough to keep electrons from tunneling through it and yet permeable enough to let the gate's electric field into the channel so that it could turn on the transistor. In other words, the material had to be physically thick but electrically thin.

The technical term for such a material is a "high-*k*" dielectric; *k*, the dielectric constant, is a term that refers to a material's ability to concentrate an electric field. Having a higher dielectric constant means the insulator can provide increased capacitance between two conducting plates—storing more charge—for the same thickness of insulator. Or if you prefer, it can provide the same capacitance with a thicker insulator [see illustration, "The High-*k* Way"]. SiO_2 typically has a *k* of around 4, while air and a vacuum have values of about 1. The *k*-value is related to how much a material can be polarized. When placed in an electric field, the charges in a dielectric's atoms or molecules will reorient themselves in the direction of the field. These internal charges are more responsive in high-*k* dielectrics than in low-*k* ones.

Incidentally, back in 2000, leading semiconductor firms began to change the material used to insulate the metal wires that connect transistors to each other from SiO_2 to low-*k* dielectrics. In the case of interconnects, you do not want the electric field from one wire to be felt in other nearby wires, because it creates a capacitor between the wires and can interfere with or slow down the signals on them. A low-*k* dielectric prevents the problem.

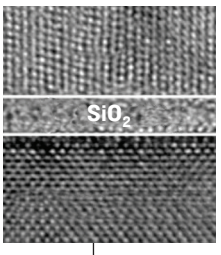
We set about studying a veritable alphabet soup of high-*k* dielectric candidates, including aluminum oxide (Al_2O_3), titanium dioxide (TiO_2), tantalum pentoxide (Ta_2O_5), hafnium dioxide (HfO_2), hafnium silicate ($HfSiO_4$), zirconium oxide (ZrO_2), zirconium silicate ($ZrSiO_4$), and lanthanum oxide (La_2O_3). We were trying to identify such things as the material's dielectric constant, how electrically stable it was, and its compatibility with silicon. For quick turnaround, we experimented with simple capacitor structures, building a sandwich consisting of titanium nitride electrodes, the high-*k* dielectric, and a silicon gate electrode. We then charged them up and discharged them again and again, watching to see how much the relationship between capacitance and voltage changed with each cycle.

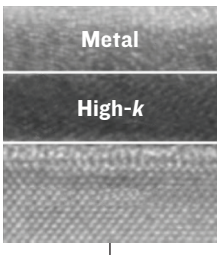
But for the first two years, all the dielectrics we tried worked poorly. We found that charges got trapped at the interface between the gate electrode and the dielectric. This accumulated charge within the capacitor altered the voltage level needed to store the same amount of energy in the capacitor from one charge-discharge cycle to the next. You want a transistor to operate exactly the same way every time it switches, but these gate-stack structures behaved differently each time they were charged up. The results were very discouraging, but eventually our team got an important break.

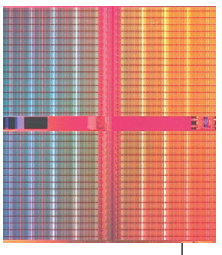
It turned out that the problem lay in how we constructed the test capacitor. To make the dielectric layer, we were using one of two different semiconductor-manufacturing techniques: reactive sputtering and metal organic chemical vapor deposition. Unfortunately, both processes produce surfaces that, though remarkably smooth by most standards, were nevertheless uneven enough to leave some gaps and pockets in which charges could get stuck.

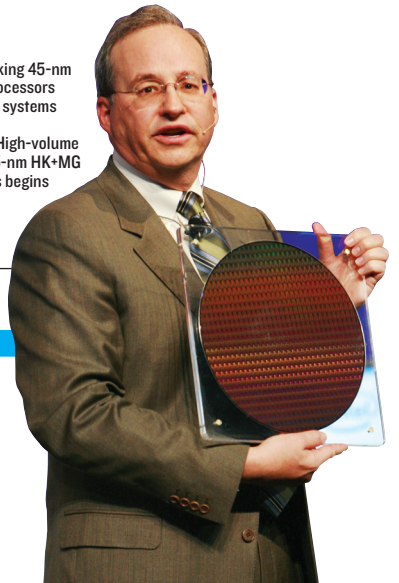
We needed something even smoother—as smooth as a single layer of atoms, actually. So we turned to a technology called atomic layer deposition, so new that its debut in CMOS chip production comes only this year with our new high-*k* chips. Atomic layer deposition lets you build up a material one layer of atoms at a time. In this process, you introduce a gas that reacts with the surface of the silicon wafer, leaving the whole substrate coated in a single layer of atoms. Then, because there is no more surface to react with, the deposition stops. The gas is evacuated from the chamber and replaced with a second gas, one that chemically reacts with the layer of atoms just deposited. It too lays down one layer of atoms and then stops. You can repeat the process as many times as you want, to produce layered materials whose total thickness is controllable down to the width of a single atom.

Deposited in this manner, both the hafnium- and zirconium-based high-*k* dielectrics we studied showed much more stable









1990 **Late 1990s** SiO_2's limits realized. Search begins for a high-*k* dielectric to replace it

2003

2004 **Nov. 2003** Working high-*k* plus metal gate (HK+MG) research transistors

2005

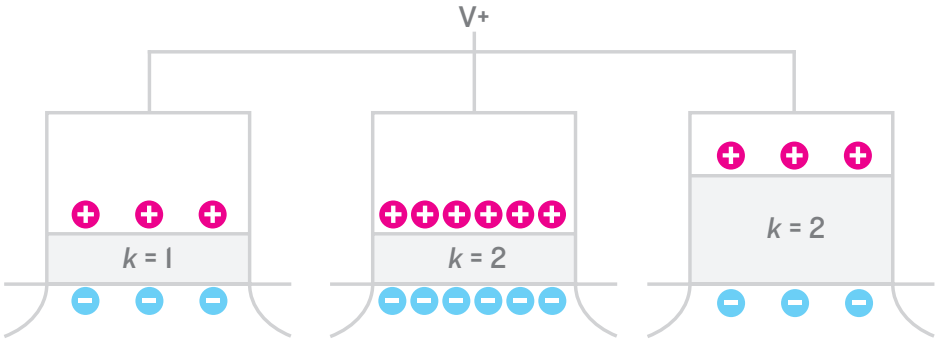
2006 **Jan. 2006** Fully functional 1-billion-transistor 45-nm HK+MG memory chips

2007 **Jan. 2007** Working 45-nm HK+MG microprocessors demonstrated in systems

2nd half 2007 High-volume production of 45-nm HK+MG microprocessors begins

electrical characteristics in comparison with the ones formed by sputtering or chemical vapors. The trapped-charge problem seemed to have been smoothed out.

WITH TWO CANDIDATE MATERIALS IDENTIFIED, we started making NMOS and PMOS transistors out of them. Then came the next snag. These transistors, pretty much identical to our existing transistors except for the different dielectric, had a few problems. For one thing, it took more voltage to turn them on than it should have—what's called Fermi-level pinning. For another, once the transistors were on, the charges moved sluggishly through them—slowing the device's switching speed. This problem is known as low charge-carrier mobility.



THE HIGH-*K* WAY: The dielectric constant, *k*, is a measure of an insulator's ability to concentrate an electric field. If one gate oxide has twice the dielectric constant of another, a given voltage will draw twice as much charge into the transistor channel. Or, the same amount of charge will accumulate, if the higher-*k* dielectric is made twice as thick.

We weren't the only ones encountering these problems; just about everybody else was struggling with them, too. With the countdown in progress for the next generation predicted by Moore's Law, understanding why the high-*k* dielectric transistors performed so poorly and finding a solution became an urgent task. Using a combination of experimental work and physics-based models, we began to figure out what had gone wrong. The source of the trouble, ultimately, came down to the interaction between the polysilicon gate electrode and the new high-*k* dielectrics.

Why this is so has a complicated explanation. The dielectric layer is made up of dipoles—objects with a positive pole and a negative one. This is the very aspect that gives the high-*k* dielectric such a high dielectric constant. These dipoles vibrate like a taut rubber band and lead to strong vibrations in a semiconductor's crystal lattice, called phonons [see illustration, "Bumpy Ride"]. These phonons knock around passing electrons, slowing them down and reducing the speed at which the transistor can switch. But theoretical studies and computer simulations performed by us and others showed a way out. The simulations indicated that the influence of dipole vibrations on the channel electrons can be screened out by significantly increasing the density of electrons in the gate electrode. One way to do that would be to switch from a polysilicon gate to a metal one. As a conductor, metal can pack in hundreds of times more electrons than silicon. Experiments and further computer simulations confirmed that metal gates would do the trick, screening out the phonons and letting current flow smoothly through the transistor channel.

What's more, the bond between the high-*k* dielectric and the metal gate would be so much better than that between the dielectric and the silicon gate that our other problem, Fermi-level pinning, would be solved by a metal gate as well.

NOW OUR ENGINEERS HAD A NEW MAJOR CHORE: find a metal they could use for the gate electrode that would combine well with the new high-*k* dielectric. Because the electrical characteristics of the gates of NMOS and PMOS transistors are different, they actually needed not one metal but two—one for NMOS and one for PMOS.

Just as standard MOS transistors use *n*-type and *p*-type polysilicon gates for NMOS and PMOS transistors, high-*k* transistors would need metal gate electrode materials with a key property similar to polysilicon's. This key property is known as the work function. In this context, work function refers to the energy of an electron in the gate electrode relative to that of an electron in the lightly doped silicon channel. The energy difference sets up an electric field that can modulate to the amount of voltage needed to begin to turn the transistor on, the threshold voltage. Unless the gate's work function is chosen well, the threshold voltage will be too high, and the transistor will not turn on easily enough.

We analyzed, modeled, and experimented with many types of metals, some with work functions that more closely matched highly doped silicon than others. But by themselves, none had exactly the work function of the doped silicon, so we had to learn to change the work function of metals to suit our needs. Eventually, the research group identified NMOS and PMOS metals by first building capacitors out of them and then transistors. We cannot disclose the exact makeup of our metal layers, because after all, the IC industry is very competitive!

We built our first NMOS and PMOS high-*k* and metal gate transistors in mid-2003 in Intel's Hillsboro, Ore., development fab. We started out using Intel's 130-nm technology, which was about three years old at the time and was used in high-volume production. The transistors, with a hafnium-based oxide and metal gate electrodes, had everything we needed: they turned on at the right voltage, leaked little current through the gate oxide, and passed a large amount of current through the channel for a given voltage. And that current moved quickly. In fact, for a given off-state current, these first transistors drove more current than any transistor reported at the time.

OF COURSE, WE WEREN'T ALONE. And there were still plenty of unknowns. By 2003, researchers in university labs and other semiconductor firms around the world had zeroed in on hafnium-based materials as the gate dielectric. A variety of them were under earnest study: hafnium oxides, hafnium silicates, and hafnium oxides containing nitrogen. The method of forming the high-*k* film, too, was unsettled, with different groups trying sputtering, chemical vapor deposition, and atomic layer deposition, which we eventually settled on. But the biggest unknowns at the time were what metal gate materials to use and how to fit them into the transistor-manufacturing process.

The normal fabrication method is known as "gate first." As the name implies, the gate dielectric and gate electrodes are constructed first. Then the dopants for the source and drain are implanted into the silicon on either side of the gate. Finally, the silicon is annealed to repair the damage from the implantation process. That procedure requires that the gate electrode material

INTEL (4)

be able to withstand the high temperatures used in the annealing step—not a problem for polycrystalline silicon but potentially a big one for some metals.

To make a long story short, the search for gate electrode materials with both the right work function and tolerance to high-temperature processing was very difficult and full of dead ends. Especially for the PMOS transistor.

Another transistor process sequence, dubbed "gate last," circumvents the thermal annealing requirement by depositing the gate electrode materials after the source and drain are formed. However, many of our peers saw the gate-last process, which we ultimately adopted, as too much of a departure and too challenging.

Meanwhile, a third approach remarkable in its simplicity emerged. Called fully silicided gates, it lets you follow the normal gate-first process but then lets you turn the polysilicon gate into a metal-silicide gate, essentially replacing every other silicon atom with metal (usually nickel). Then, by doping the nickel silicide, you can alter its work function for use in either an NMOS device or a PMOS one. By late 2006, though, nearly everyone, including us, had given up on the fully silicided gates approach. No one could move the silicide's work function quite close enough to where it needed to be.

Nevertheless, the search goes on at other major chip makers to find the materials with the right work function that could survive high temperatures and enable the industry standard gate-first process flow.

HAVING BUILT WELL-FUNCTIONING TRANSISTORS

using old technology, in the second half of 2003 it was time to move from research to development of high-*k* dielectric plus metal gate transistors, as we called them. Engineers began working to determine whether these early transistors could be scaled to the upcoming 45-nm dimensions and still meet the rigorous performance, reliability, and manufacturability requirements of an advanced microprocessor technology.

It was no cakewalk. The research group engineers had provided a critical lead in identifying promising high-*k* and metal gate materials, but the NMOS and PMOS transistors had not yet been combined on one wafer as they would be in a microprocessor, using a manufacturing process that could make both. What's more, there were hard questions still to be answered about how many good chips we could expect for every bad one (yield) and how reliable those chips would be.

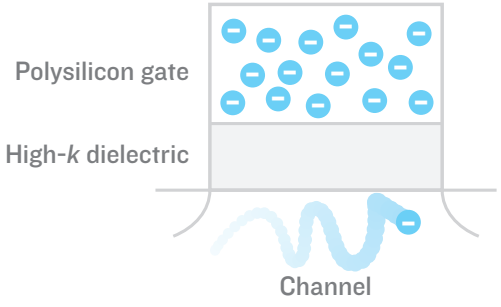
During the months that followed, the team cracked one problem after another—making changes to materials, chemical recipes, and manufacturing processes. It wasn't until late 2004 that the team felt it had enough convincing data that the new transistors could be made to work on our 45-nm technology. At that point, there was no turning back. Intel was now committed to making a high-*k* dielectric plus metal gate transistor structure using the gate-last process flow. It was a gutsy call. Our team knew it was committing all of Intel's next generation of microprocessors to the biggest change in transistor technology in 40 years.

The next key milestone was to demonstrate working test chips using the final scaled dimensions combined with the new transistor features. The traditional chip to test a new technology on is static random access memory, or SRAM, which is the type of memory collocated on the same chip with the microprocessor. Typically, microprocessor makers have designs for SRAM that are a year or more ahead of their processor designs. SRAM is a very regular array of memory cells, each of which consists of six densely packed and interconnected transistors. Because of their density and regularity, SRAM chips provide good data on how many defects a manufacturing process produces.

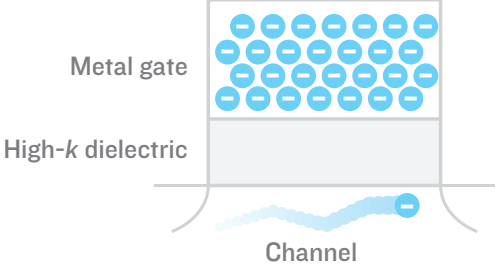
Our first fully functional test SRAM chips with the new transistors came off the line in January 2006. They were of a 153-megabit design consisting of more than 1 billion transistors. Each six-transistor memory cell in the chip occupied little more than one-third of a square micrometer. This test chip had all the features needed to build a 45-nm microprocessor, including the high-*k* plus metal gate transistors and nine layers of copper interconnects. Considering how new and radically different the transistor and manufacturing process were, it was a surprise even to some of the engineers in the development group that it all worked together so well. Even so, the development team still had a lot ahead of it to bring the performance, reliability, and yield of the process up to the level needed for manufacturing microprocessors.

The new gate stack worked wonders in battling leakage through the gate, reducing it by more than a factor of 10. But the gate oxide is not the only source of transistor leakage chip makers have to worry about. The other significant leak is called source-to-drain or subthreshold leakage. It's a trickle of current seen even when the transistor is intended to be in the "off" state. Making transistors smaller has also meant steadily lowering the amount of voltage needed to turn them on, the threshold voltage. Unfortunately, steadily lowering the threshold voltage lets more current slip through. For many years, each new generation of transistor would increase drive current (and improve performance) by about 30 percent but would pay the price of about a threefold increase in subthreshold leakage. Leakage currents have reached levels high enough to be a noticeable portion of total microprocessor power consumption.

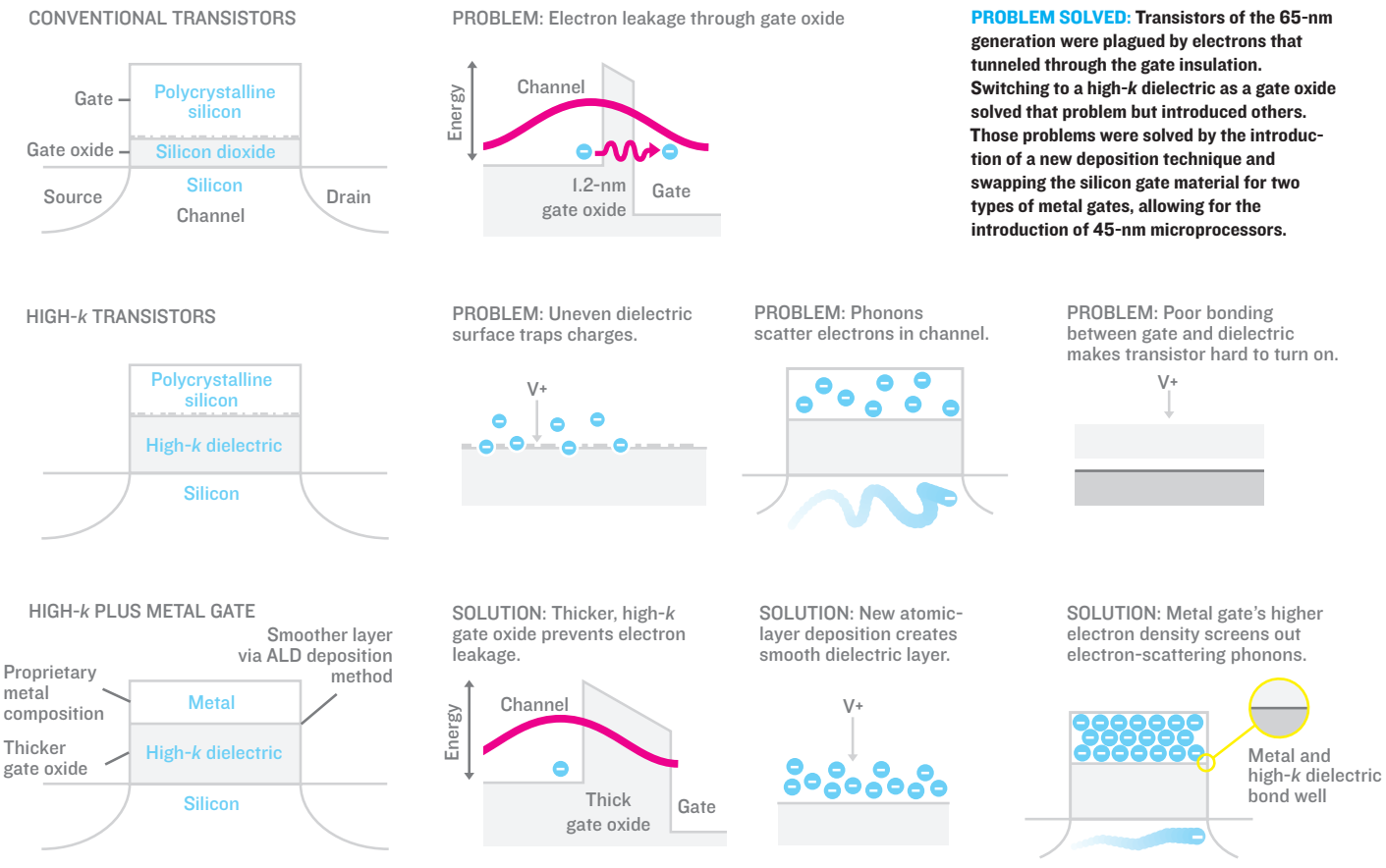
The industry is now in an era where power efficiency and low leakage are more important than raw speed increases. But a transistor can be designed to operate to favor either priority by adjusting the channel length or adjusting the threshold voltage. A shorter channel leaks more but allows for a higher drive current. A higher threshold voltage pinches off the leak but also throttles the drive current. Adjusting the threshold



BUMPY RIDE: **The particular density of electrons in a traditional polysilicon gate allowed inherent vibrations in the high-*k* dielectric to move into the transistor channel and disrupt the flow of current.**



The higher density of electrons in a metal screened out the vibrations, allowing current to flow more smoothly.



voltage is where the high-*k* dielectric comes into play. A thicker dielectric reduces the gate's ability to open a conductive channel between the source and the drain, increasing the threshold voltage. A thinner dielectric layer has the opposite effect. Compared with the previous 65-nm transistors, 45-nm high-*k* plus metal gate transistors provide either a 25 percent increase in drive current at the same subthreshold leakage or more than a fivefold reduction in leakage at the same drive current, or anywhere between those values. We can make the choice on a product-by-product basis, or different circuits on the same microprocessor chip can use different transistors to optimize for performance or power.

IN JANUARY 2007, Intel made the first working 45-nm microprocessors using these revolutionary high-*k* plus metal gate transistors. One was the Penryn dual-core microprocessor, which has 410 million transistors. Different versions of Penryn will be optimized for mobile, desktop, workstation, and server applications. The quad-core version of this product will have 820 million transistors. Penryn was followed a few months later by Silverthorne, a single-core microprocessor with 47 million transistors that is designed for low-power applications, including mobile Internet devices and ultramobile PCs. There are more than 15 new chips under development at Intel using our new technology. Production of Penryn and Silverthorne will start later this year at Intel plants in Oregon and Arizona. Next year, we'll start up the process at two other high-volume manufacturing fabs, in New Mexico and Israel.

The invention of high-*k* plus metal gate transistors was an important breakthrough. Although we could have continued to shrink transistors to fit the dimensions needed for the 45-nm generation without this breakthrough, those transistors would not have worked much better than their predecessors, and they certainly would have expended more watts. We're confident this new transistor can be scaled further, and development is already well under way on our next-generation 32-nm transistors using an improved version of high-*k* plus metal gate technology. Whether this type of transistor structure will continue to scale to the next two generations—22 nm and 16 nm—is a question for the future. Will we need new materials and new structures again?

Nobody knows for sure. But that is what makes integrated circuit research and development so exciting. ■

ABOUT THE AUTHORS

MARK T. BOHR, an IEEE Fellow, is the director of process architecture and integration at Intel. ROBERT S. CHAU, an IEEE Fellow, is the director of transistor research and nanotechnology. TAHIR GHANI, an IEEE member, is the director of transistor technology and integration. KAIZAD MISTRY, an IEEE senior member, manages the development of Intel's 45-nanometer CMOS technology in the logic and technology development group.

TO PROBE FURTHER

Robert S. Chau and colleagues explained the problem that led to the use of a metal gate in detail in "High-*k*/Metal-Gate Stack and Its MOSFET Characteristics," *IEEE Electron Device Letters*, June 2004.

Intel and others will be presenting the latest high-*k* dielectric and metal gate transistor research at IEEE's 2007 International Electron Devices Meeting, in Washington, D.C., from 10 to 12 December.