ILLUSTRATIONS: GAVIN POTENZA

THE IEEE TECHNOLOGY TIME MACHINE

WHAT'S AHEAD AT HONG KONG, JUNE 1–3, 2011

INTRODUCTION A custom has evolved around the world in all places where lakes freeze over in winter and thaw in the spring—from New England to Idaho, northern Scandinavia to Alaska, in the Andes and the Himalayas—of placing an object out on the ice during the winter and taking bets on when spring will arrive, as announced by the ice's melting and the thing's dropping into the water. Something of the same spirit will infuse the small, elite meeting that will convene in Hong Kong next month, with the purpose of exploring new directions in technology.

When will artificial objects start talking with each other and with us as well, just as we humans all chat with each other today on the Internet? Is cloud computing already at a hockey-stick inflection point and about to sweep everything in its path? When will any health-care provider, anywhere on the globe, be able to instantly access our complete medical histories (with our authorization, of course)? Can silicon microelectronics keep up with all the new demands being made on it? Will we have the bandwidth and storage capacity to move and permanently preserve all the data we care about most?

The purpose of the Hong Kong IEEE Technology Time Machine (IEEE TTM) meeting is to assemble people who are betting their corporate and national futures on the answers to such questions, so that they can compare notes and see how competing visions interact, complement, or contradict each other. The focus is on where key technologies will be 10, 15, and 20 years out.

A secondary objective of the conference, which has been organized by the Future Directions Committee, part of the IEEE Technical Activities Board, is to provide IEEE itself with perspectives on what's ahead. The expectation is that cross-disciplinary and technology-bending areas will be identified in which the Institute of Electrical and Electronics Engineers can undertake new initiatives.

The participants in TTM will be CEOs, CTOs, and CIOs, top government officials and managers, and leading scholars. Participation is deliberately being kept small, so that all participants will have ample opportunity to discuss whatever's on their minds with just the people they want to discuss it with.

For similar reasons, the number of topics addressed in plenary talks and subject sessions has been kept to under a dozen, partly—to be frank—so as to guarantee that every single talk will be delivered by a recognized top expert and address a truly critical subject. In the course of the conference, we are confident it will become apparent that some topics apparently absent from the agenda are actually included. There are no scheduled talks, for example, about smart buildings or smart vehicular transportation, despite the pronounced currency of those topics. Yet much of what will be said under the headings of the smart grid, electric vehicles, cloud computing, and the "Internet of Things" will bear directly on green architecture and accident-avoiding vehicles.

We certainly do not wish to pronounce the results of what is sure to be an exciting meeting prematurely. But we are completely confident of this: In 2030, all those heavy objects sitting during the winter on frozen lakes around the world will be communicating via wireless local networks and over the Internet with each other and with us in real time. So we'll be betting not just on when the archaic sport utility vehicle sitting in the middle of our local lake will drop to the lake bottom, to be preserved as a curious relic of a bygone era, but on which particular SUV of all the SUVs sitting on lakes drops first. ■

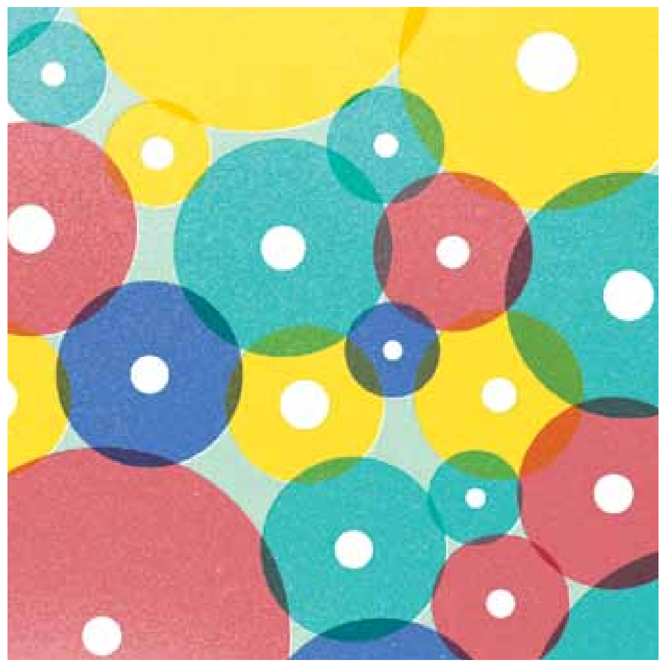
The IEEE TTM event is sponsored by IEEE Future Directions through Technical Activities. The organizing committee comprises Executive TTM Chairman Roberto de Marca, a professor at the Catholic University Rio de Janeiro; the general TTM chairs, Khaled Ben Letaief, dean of engineering at Hong Kong University of Science and Technology, and Nim Cheung, the CEO of ASTRI; and TTM Technical Program Chair Yrjö Neuvo, the retired CTO of Nokia and current professor/research director at Aalto University. Robert Hebner is current chairman of the Future Directions committee and is also the director of the Center for Electromechanics at the University of Texas at Austin.

THE INTERNET OF THINGS

BY ROBERTO SARACCO

The Internet of Things (IoT) refers to the vision that in the next 20 years, a revolution in device-to-device communication will take place that will be comparable to the revolution in person-to-person communication that erupted in the last two decades with the Internet and World Wide Web. We believe the vision is credible—that the second revolution will in fact occur and is already beginning before our eyes.

The IoT is going to be more than devices talking for one reason or another among themselves. Things equipped with sensors and actuators will become part of the Internet, just as today information and services are part of the Web. Thus, we will be able to browse for "things" just as today we search for information. We will be able to create environments out of things, just as today we can mash up services and information.



Examples are abundant, from the sublime to the mundane. In South Korea, a wireless network of 663 sensors constantly monitors conditions on the Jindo Bridge, which connects the country's southernmost tip with Jindo Island. Developed by teams at the University of Illinois, the University of Tokyo, and Korea's Advanced Institute of Science and Technology, the network employs sensors that cost less than $100 each and run on batteries that have to be replaced only every three years.

Go to your local hardware store, and you'll find a nifty little gizmo that you've wished for all your life but that only recently came into existence. For years you've been complaining about those ball-link and string pull cords on lamps and ceiling fans, which are always snapping and often are a real pain in the neck to fix. Now, you can just leave the pull cord in the on position and buy a wireless switch for the fan or light that you can put in any convenient spot.

In fact, if you also happen to be tired of hearing family members and houseguests complain about where the wired switches in your house happen to be, you might also be able to find wireless switches that just slap onto the wall and can be moved about as your kids get bigger and their grandparents get smaller.

What those situations have in common are simple, inexpensive devices linked in wireless networks, which might or might not involve some human intervention. That is, the IoT depends on networked sensors and actuators, which can be more or less sophisticated. To the extent there is embedded intelligence, an embodiment of the IoT can function quite autonomously, making decisions and taking actions that would normally require human activity. But some embodiments assume that data and decision making will be centralized somewhere and that, ultimately, people will take charge and maneuver the network to their advantage.

The current IoT incarnation most frequently mentioned is the RFID network. In many early applications, such as inventory management in big stores, RFIDs are essentially passive devices, read by machines that communicate and act centrally. But RFID devices can also be active, so that they can communicate directly over the Internet with other relevant things and act autonomously as a group without necessarily requiring people to be present.

Since the IoT vision is rather elastic, lists of enabling technologies can be more or less extensive, depending on how expansive the definition is. Just about any list, however, includes the following key ingredients:

• The cloud. Underlying the vision of wireless sensor and actuator networks with embedded intelligence is the assumption that computing resources can be accessed widely, so that information gleaned from the network can be processed economically and directions fed back to the network efficiently.

• Inexpensive things. To be virtually ubiquitous, which is the IoT vision, sensors, actuators, and similar devices must be cheap. What can make them so is the development of MEMS (microelectromechanical systems)—that is to say, the incorporation of active elements into microelectronics devices made with standard semiconductor manufacturing techniques.

• Small energy-harvesting devices. Efforts are taking place on a broad front to develop the means of generating energy from vibrations, temperature gradients, piezoelectric and photovoltaic effects, and radio frequencies, among others. Some of them will pan out.

• Infinite Internet addresses. IPv6, introduced incrementally starting a decade ago, provides all the addresses we could ever ask for. With its longer address space of 128 bits, the new Internet Protocol can generate something like 3.4 times 10 to the 38th addresses, enough to provide 1,500 addresses for each square meter of the earth's surface.

• Suitable wireless networks. The main network standard, though not the only one, is ZigBee 1.0, which was conceived and born in IEEE's 802 family, of which by far the best-known and most successful member is Wi-Fi (802.11). ZigBee, a product of the 802.11.4 personal-area network standards group, provides for a mesh networking stack that scales and self-heals well, along with an inexpensive short-range radio system. A mediating standard called 6LoWPAN, developed by the Internet Engineering Task Force, adapts the prescribed IPv6 packet size to the ZigBee stack.

According to Bob Heile, who has chaired the ZigBee effort from its inception, there are already about 100 million ZigBee-enabled sensor devices out there. As he sees it, ZigBee is virtually synonymous with the IoT. But ZigBee is not the only way to go, and not everybody agrees it's destined to carry the day. The Illinois Structural Health Monitoring Project system used in the Jindo Bridge uses a Chipcon CC2420 radio that implements the IEEE 802.15.4 standard but does not use ZigBee as such. Its developers built a data transport protocol on top of TinyOS protocols.

There are a number of top tech companies and tech groups eying large-scale employment of networked devices, with various purposes in mind. HP Labs thinks of its Central Nervous System for the Earth as somewhat more encompassing than what's usually meant by IoT, mainly because people are seen as important elements. One big application area is health, where sensors and actuators are attached to people, communicating with care providers by means of telemetry. HP's main current project under the nervous system heading is in oil exploration, with Shell Oil. The general idea is to employ intelligent sensor packages in geologic surveys and to vertically process the information obtained.

IBM's A Smarter Planet, with 10 major areas of business opportunity in view, seeks to employ distributed sensing and intelligence to bring objectives—among them smart grid and smart transportation implementations—into line with resources. The emphasis on the program seems to be mainly on computerized analytics, not primarily on what's thought of as IoT as such.

Things outfitted with sensors and actuators are front and center, on the other hand, in the IP for Smart Objects (IPSO) Alliance, in which companies like Bosch, Cisco, Ericsson, SAP, and Texas Instruments are leading players. A document on Cisco's website says, "Intended to complement the efforts of entities such as the Internet Engineering Task Force (IETF) and the Institute of Electrical and Electronics Engineers (IEEE), which develop and ratify technical standards in the Internet community, the IPSO Alliance will perform interoperability tests, document the use of new IP-based technologies, conduct marketing activities, and serve as an information repository for users seeking to understand the role of IP in networks of physical objects."

How pervasive and successful will such efforts be in the next two decades? This much seems sure: Wireless networks of sensors and actuators, however autonomous and intelligent, are going to transform buildings and most other engineered structures, including vehicles. We humans will be able to interact with any physical object in our environment that is equipped with a URL, which we will be able to click on to obtain associated information and services. The impact on businesses and economies will be significant. ■

Roberto Saracco has been director of the Future Centre, in Venice, since 2000, reporting directly to the chief technology officer of Telecom Italia. From 1996 to 1997 he chaired the European Union's Visionary Group on Super-Intelligent Networks, and he has initiated major World Bank projects to foster entrepreneurship in information technology. In the 1980s, he led research in telecommunications management in CSELT, the top Italian telecom R&D center, in Turin, and he participated in standardization activities at CCITT, the ITU's standards organization, as well as at OSI, ETSI, and T1M1.

CLOUD COMPUTING

BY STEPHEN L. DIAMOND, ALEXANDER PASIK AND JEFFREY VOAS

First of all, it's not really a cloud, it's clouds: Rather than meet all their computing needs with dedicated, on-the-premises computers—be they the mainframes of the '60s and '70s or the client-server systems of the '80s and '90s—organizations build their own private clouds, rent computing resources on a cloud service provider's systems to meet some or all of their needs, or do both. These various requirement sets will likely result in the development of clouds with dynamically flexible boundaries.

The IT outsourcing trend as such is not new. ADP has made a fortune in recent decades by supplying outsourced applications to handle things like payroll and human resources management. Those somewhat standardized systems could be run, in principle, on an organization's own computers. But the trend has gathered pace steadily and now appears to be at an inflection point, with cloud computing sure to soar in this decade. Small and medium-size businesses have scored critical successes, and large private-sector organizations and governments are starting to move significant fractions of their computing to the cloud. And with mobile applications spreading like wildfire on smart phones and tablets, consumers are not far behind.

In a sense, we're all already doing some of our computing in the clouds

Can anything hold back cloud computing's tidal flood?

on a daily basis. As individuals, we use e-mail systems hosted by Yahoo and Google. Small and medium-size businesses rely on Intuit for accounting services, and large enterprises use Salesforce.com for customer relations management. It can be argued that some of these are merely Web and not cloud applications, but the difference is blurring.

What has made the adoption of cloud computing practical is a trend in computer architecture called virtualization. Instead of having separate servers for every operating system, each with dedicated memory and processing capacity (involving horrendous redundancies), organizations can distribute many operating systems and applications across just a few servers, with shared resources. Virtualization hides the physical characteristics of computer resources behind an abstract model, so that, for example, many alternative operating systems and applications can run simultaneously on the same hardware. The first commercial computer with virtual memory dates to 1960. Now entire computers are virtualized, including CPUs, storage, and networking, creating a virtual machine.

Besides clearing the way for a radical streamlining of IT infrastructure and yielding huge economies in terms of reduced footprint, energy use, and total cost, virtualization has also greatly facilitated cloud computing. As it became possible to upload almost any operating system or application to the cloud to meet any organizational requirement and to scale operational solutions to almost any level, in theory just about any computer setup could be customized on the fly to do computing for any client.

How far can such developments ago? It's not inconceivable that by the end of the decade, computing resources will be sold as commodities, just like pork bellies, megawatt-hours, or carbon credits, which will be listed in daily papers (if daily papers still exist) and traded on spot and futures markets.

Amazon, which had to build enough capacity to handle peak traffic for its e-commerce on Black Friday (the day after Thanksgiving) and thus found itself with huge excess capacity for most of the year, pioneered commodification: rentable storage units, "instances" (slices of computing time), and "compute units."

Though cloud computing till now has been largely a private-sector preserve, cash-strapped governments around the world are hungrily eyeing it. In the United States, the White House's chief information officer, Vivek Kundra, has enunciated a "cloud first" policy for the whole government, which spends an estimated $80 billion a year on information technology, making it the world's biggest single consumer of IT services.

In Europe, where concerns about data privacy and cross-boundary data security are particularly acute, an EU technology/standards organization issued a report early this year assessing the cloud's net benefits in highly positive terms. "We have concluded that the cloud computing service delivery model satisfies…most of the needs of public administrations…since it offers scalability, elasticity, high performance, resilience, and security," the EU's European Network and Information Security Agency wrote. "Private and community clouds [clouds customized to serve the needs of particular user groups] appear to be the solution that currently best fits the needs of public administrators, since they [community clouds] offer the highest level of governance, control, and visibility."

The U.S. National Institute of Standards and Technology is doing what it can to expedite adoption of all related technologies, seeking to be an authoritative source for best practices and reference systems, though it is not setting standards or proposing policies as such. It distinguishes between three flavors of cloud computing: software as a service, platforms for developers of cloud software, and computing infrastructure as a service.

IEEE has initiated cloud standards development work, with a focus partly on interoperability and portability profiles and partly on protocols for cloud-to-cloud federations. With many relevant technical strengths and existing cloud-specific conferences, IEEE will step up educational activities such as seminars and workshops and will launch cloud-specific publications.

Can anything hold back cloud computing's tidal flood? Computing infrastructure is superabundant at present and probably not a medium-term constraint, even taking microchip and server farm energy requirements into account. Bandwidth also does not seem to be a serious constraint. Cloud computing is not as communications-intense as it might appear: Limited sets of input and output data are communicated, in the main, while databases reside "in the cloud."

At the top of everybody's list of limiting factors are concerns about privacy, security reliability, and standardization. And they can take different forms depending on where you are, complicating the development of universal solutions. On the Eurasian continent, where fascism and totalitarian communism have left searing memories, people worry about governments getting sensitive personal information of the kind that police states love. In North America, concerns about identify theft are more acute.

Data security is not a trivial matter, but when you get into specifics, it may not be quite what it seemed at first. Are your data necessarily more secure in your own data center than they would be stored on Amazon's servers? There will always be people who feel their money is safer under the mattress than it would be in a bank. But most prefer the bank.

In time, and probably not much time—with adoption of best practices, standards, and perhaps some regulation—most IT customers will find their computing needs are best handled by the data utility of the future. ■

Stephen L. Diamond is managing director of Picosoft, a Silicon Valley consulting firm specializing in high-tech management, strategic planning, marketing, business development, and standards. Previously he worked in executive positions at Equator Technologies, Tycho Networks, Sun Microsystems, National Semiconductor, and Honeywell/Synertek. Alexander Pasik is chief information officer for IEEE and teaches as Columbia University, in New York. Previously, besides serving as CIO at the Solomon R. Guggenheim Foundation, he founded and managed multimillion-dollar research departments at Gartner, Lazard Freres, and Ernst & Young. Jeffrey Voas is a computer scientist at the National Institute of Standards and Technology (NIST), in Gaithersburg, Md. Previously, as an entrepreneur, he cofounded Cigital in 1992. After more than a dozen years with Cigital, he became a director at Science Applications International Corp. and was an SAIC Technical Fellow from 2005 to 2009.

MOBILE SERVICES AND PAYMENTS

BY YRJÖ NEUVO AND TERO OJANPERÄ

With mobile data traffic doubling yearly, we are all used to having smart phones that let us browse the Internet, manage our e-mail, get spoken, real-time driving directions from wherever we happen to be to wherever we want to go, and play with all manner of other apps—thousands in all. And we are all used to doing so using layered screens and obtaining sound, as desired, from loudspeakers. You can sit in the passenger seat of a car and watch a soccer match on a cell phone while your spouse, driving, listens to it. Alternatively, you can have the phone feed the HDTV-quality video of the game into an onboard television display.

Today, I can carry all my business applications and presentations in my smart phone. I can be plugged into the "cloud," an infrastructure where data and apps exist outside of the mobile device, available anywhere and anytime.

Then there are location- or context-aware services, such as mobile navigation (an area in which Nokia with its Ovi Maps is at the forefront of such developments). Navigation and mapping sound data-intensive, but with today's technology, using vector maps, you can have the whole world literally in your hand—all in about 6 gigabytes of data (about 2 GB each for North America, Europe, and Asia). Given that today's high-end smart phones have a built-in capacity of 16 GB, it is quite feasible to have a global street map with the finest details permanently stored in the phone.

Maps can easily be preloaded to your mobile phone via home Internet or Wi-Fi. You can have a seamless free-of-charge navigation experience right away—not only in Finland, where we are from, but in almost 100 other countries around the globe, in 46 languages. The so-called hybrid or connected-PND (personal navigation device)" approach lets consumers experience the service both on- and offline. Users have a full navigation solution—including map data, free navigation, routing, and third-party content like TripAdvisor—preinstalled, with no need to go online. That saves users money and reduces bandwidth requirements for operators.

Getting mobile customers from point A to point B is only the beginning. Mobile services today are aggregating social networks like Facebook and Twitter, offering "best buy" opportunities depending on your location, and letting you know if you need to hurry up to catch your flight because of traffic. In the future, your mobile phone will become the nerve center of your life and your personal concierge. There's a Finnish bank (and doubtless others in other countries) that provides all its banking services from a single mobile app. Throughout Europe, the mobile phone, or "handy," is rapidly replacing the smart card as the payment medium of choice for public transportation. The phone can also handle airline boarding passes and bar-coded entertainment tickets, register and maintain lists of all RFID-tagged transactions, and of course buy that morning cup of java.

Increasingly, the mobile phone is in effect a mobile wallet—a kind of Swiss Army knife that equips the user for just about every aspect of daily life.

Two decades ago, much of that would have seemed the stuff of science fiction, not the science of the near future. So what will the next 10, 15, and 20 years bring?

People—that is to say, users—will decide what the killer applications of the future will be in mobile communications. The word convenience describes one major direction of development. A TV remote control and wireless car keys are good examples. Neither one solves a really serious or generally recognized problem; they merely increase convenience.

Some mobile applications require a substantial new infrastructure before they can be fully adopted as well as adding the corresponding functionality in the phones. This creates a kind of chicken-and-egg problem, where infrastructure investment waits for the phones to support the new functionality and the phone manufacturers wait for the infrastructure to appear.

This chicken-and-egg problem has been delaying development, notably in mobile payments. For good reasons, handling financial transactions is one of the most conservative areas. Security cannot be sacrificed. Here, we also need good cooperation with the banking world and the mobile carriers. Of course, there has to be a provision for canceling accounts if the phone is lost or stolen and a backup means of access.

Bank transfers are already possible from your mobile phone in the same way they can be conducted from a laptop, the difference being that this particular "checkbook" is always with you. Japan's NTT DoCoMo has been pioneering an "e-wallet" service called FeliCa. Some airlines accept boarding passes in the form of an SMS message or a two-dimensional bar code. In Helsinki, You can buy tram and train tickets over the Internet by texting from your cell phone. Near-field communication (NFC) devices—those equipped for very short-range, very low-energy communication—allow for integration of a contactless credit and debit card on your mobile device. The wireless handheld cash registers widely used in stores and taxis are pointing the way.

The smart phone is obviously destined to be a natural and fully integral part of people's browsing habits. The device is always with you and for that matter always on, and so it is the natural interface for accessing all manner of cloud computing services. Audiovisual expression, input capabilities, and data speeds will improve continuously. Functionality will increase dramatically for a given-size phone, and this trend will push innovations in audiovisual expression, as getting more elaborate output from an ever tinier device is not easy. We already have a low-energy version of the Bluetooth standard available for cell phones, allowing them to be your car and home keys. The same technology will let your phone connect to and locate anything equipped with a wireless tag: the tennis or golf shoes you put away somewhere for the winter, the liquor bottles you're trying to hide from your teenaged children. Interactive devices could let you read the outdoor temperature on your phone or activate lights and heating or cooling from your phone.

The SIM card contained in all GSM mobile phones—the European standard, widely used elsewhere in the world as well—can, in principle, be inserted into any kind of mobile equipment to provide 3G or 4G wireless connectivity. So, for example, if adding a cell phone connection to your lawn mower makes it easier to maintain and operate the machine, such connectivity will be easily affordable.

As world maps in even the finest detail have become free and the phones have integrated GPS, you can expect a fleet of location-aware applications. Making use of the phone's camera and high-speed data functionalities, augmented reality applications become feasible. The phone will thus be able to answer the question, "What is this?" And not only that. To the extent information about your particular location is enhanced with imagery from other users' phones and archived data, you can be told about things you haven't yet noticed in your immediate surroundings. A number of widely used applications—such as Layar and Nokia's Point & Find—are already delivering such capabilities.

In certain kinds of work, such as mineral exploration and in the world's most advanced militaries, organizations have for some years had the capability of seeing from the opposite side of the world where people are, what they are doing, and how they might operate more effectively. With camera phones, those capabilities are becoming available to all. As imagery is integrated with photorealistic 3-D mapping services, almost anything you're doing in the real world can be mirrored in the virtual world.

In principle, mobile phones can be equipped with just about any kind of sensor. Already they typically have GPS, a compass, and a camera, and the microphone picks up ambient sound, in addition to the voice it's designed for. But phones could also measure temperature and humidity, wind strength, the user's motion, or even the presence of toxics or radioactivity in the atmosphere.

As information from many such phone sensors is collected and processed, a kind of collective intelligence develops that can be harnessed to feed guidance back to the user or the environment in which the user is operating. For example, as numerous users obtain driving instructions and follow them, the carrier can determine local traffic conditions and advise the user how to modify routes. In addition, such information could be fed into a system of smart traffic lights.

Atmospheric readings obtained from many phones could yield much more fine-grained estimates and forecasting of weather conditions, so that users and the general public can be better alerted about what to wear and whether to drive or bike.

Every day, we see instances of individuals voluntarily taking video of accidents and other emergency conditions, which is then fed to broadcasting networks and first responders to improve understanding and management. That is but a foretaste of how the public will be served in the coming decades, as information is leveraged from crowds of mobile-phone users. ■

Yrjö Neuvo, a professor and technology advisor, is the former executive vice president of Nokia Corp., where he was responsible for mobile phones product development and CTO. He served on Nokia's group executive board from 1993 to 2005. Prior to joining Nokia he had an academic career as a professor of signal processing at the University of California, Santa Barbara, and the Tampere University of Technology. He is a Life Fellow of the IEEE. Tero Ojanperä is executive vice president and acting head of services and developer experience at Nokia. He joined Nokia in 1990 and has served as executive vice president, chief technology officer, and chief strategy officer. He was head of the Nokia Research Center in 2003–2004.

SILICON OUTLOOK

BY PAUL FARRAR AND WILLIAM TONTI

The main story in microelectronics for the next 10 years and beyond will continue to be silicon, but it won't be the old familiar Moorish tale of transistor density and chip performance scaling with time. That story is already obsolete. For several generations of chips, gains in performance have come increasingly from innovation in design and materials and ever less from cramming more and more switches on a wafer.

Keeping Moore's law on course has represented an impressive technical achievement, but it has come at costs—both economic and technical—that are becoming glaring. As the microelectronics industry has had to put more and more money into chip design, market growth and revenues have leveled off, so that there's less income to support future innovation. At the same time, even as chips approach atomistic and quantum-mechanical limits, obtaining enhanced performance from miniaturization is difficult or almost impossible without drastically added energy consumption. That is a concern in the computer industry as a whole and for everybody who depends on that industry—which is to say, just about everybody.

Those are the dilemmas that the four big players—Intel, TSMC, Samsung (in DRAM), and the IBM logic chip alliance—will be wrestling with. But there's another dilemma more immediately at hand, the much delayed cost-effective arrival of extreme ultraviolet lithography (EUV), the next-generation technique needed to etch smaller features on still more densely packed chips.

At present, most chips in mass production are made by means of deep-ultraviolet optical lithography, which has a wavelength of 193 nanometers. The chips typically have a feature size of 90 nm; some have nodes as small as 65 or 45 nm. By stretching the capabilities of standard optical lithography—mainly by immersing the imaging system in a liquid—the most advanced chips have feature sizes of 32 nm, and researchers are developing the 22-nm node. But without EUV, which has been demonstrated technically but not economically, it is not clear how chips with features of 11 or 8 nm will be made.

The basic problem is throughput, as lithography is by far the most expensive part of the semiconductor manufacturing process, in terms of floor space as well as money. We know how to make the wafers from which chips are cut at a rate of about one or two per hour using EUV, but for the process to be profitable, that rate has to be improved about a hundredfold. So it's possible—but by no means a given—that EUV won't happen at all. There is no obvious alternative. Absence of viability will certainly delay introduction of chips with 15-nm nodes (that is, with a half pitch of 15 nm) and stretch development time from this generation to the next from, say, two years to perhaps three years. You'd also have to change design to much more regular patterns: just one orientation to gate, one orientation to pitch (the distance between circuit lines on the bottom layer of the chip), and so on.

Silicon will nevertheless remain the workhorse of computer chips for the next 10 years and will still be dominant for a decade after that. We can see solutions down to sub-10 or sub-7 nm. There are going to be economic challenges, and the time needed to get from one generation to the next will likely lengthen, but there isn't big money chasing any other solution.

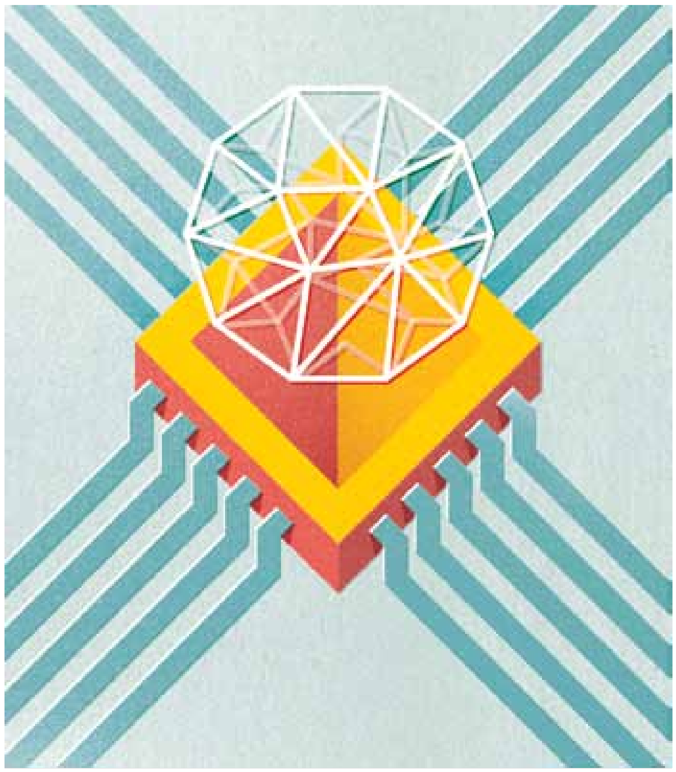
The likelihood of momentum carrying you to success using methods you know is greater than you probably think, and you tend to underestimate the challenges that will arise by doing things fundamentally differently. Within living memory, there were people saying we'd never get silicon below 1 micrometer.

The transition from oxide/polysilicon gate to high-k and metal gate stacks has shown the way. The higher-k materials now being introduced reduce electron leakage through gates a hundredfold. So chips with the higher-k gates scale down much better. Intel has started to ship chips incorporating them, and smart phones sold in the next few years will depend on them. Other improvements in 45-nm and 32-nm chips have come from embedding DRAM memory in processors, the use of lead-free copper, and advanced wiring interconnects.

Further gains will come from use of graphene and carbon nanowires, with lower resistance and higher conductivity; we'll see some hybrid structures using silicon and either carbon or graphene. There will be a lot of innovation in materials.

We'll be seeing three-dimensional devices built from FinFETs, transistors in which the gate rises vertically like a shark fin instead of being laid out horizontally. We'll find ways of tucking things in the crevices between planes. Much more complex and versatile circuitry will connect chip layers, and the number of such layers will increase from 10 to 12 at present to 14 to 16.

The holy grail, at least 5 to 10 years out, is wafer-on-wafer (or 3-D) technology, permitting better integration of logic and memory but requiring development of ultraprecise alignment and sure bonding techniques. If ways can be found to layer one wafer right on top of another, then chip-to-chip electrical connections can be radically shortened, yielding big gains in performance and energy consumption.



Optimizing the integration of processing and memory with silicon-on-silicon might buy us the equivalent of a generation or a generation and a half in feature reduction.

A technique already used in production of the most advanced chips today and that is sure to become increasingly important involves a kind of adaptive optics: a technique that goes by the name of source-mask optimization. Etching chips with features that are roughly one-third the length of the incident light leaves a lot of rough edges and tiny defects. So the chipmakers are employing teams of Ph.D. mathematicians around the world to devise ways of predicting where those defects will appear and modifying the light to fix them as they're cut.

Another holy grail, at least a decade out, is spintronics—the development of transistors and chips that exploit an individual electron's spin rather than electron flow to switch and amplify signals. Since an electron's spin (really a quantum-mechanical property somewhat analogous to rotation) is proportional to its magnetic momentum, in principle, when spin-polarized electrical currents flow through different types of magnetized metal, resistance changes could be exploited to store information. Even more interesting would be a microprocessor that encoded information using the orientation of electrons: It could handle data thousands of times as fast as present-day processors that rely only on charge.

A promising way to economize on energy consumption is programming chips to put sectors to sleep that aren't actively in use. The introduction of high-k dielectrics shows the way.

But the history of the metallic gate also illustrates the time needed to take a promising idea from the laboratory to a consumer product. High-k dielectrics began to be evaluated in the late 1990s, and by 2001 hafnium had been identified as a top candidate. The first SRAM array incorporating such gates was made only in 2006, however, and the industrial infrastructure was not in place for a rapid ramp-up of production until 2008.

There are big challenges in power, performance, and reliability, all of which have to be addressed effectively and in combination. In CMOS, we're kind of stuck at 1-volt operation, and this limits scaling the power budget. The threshold characteristic of a device to distinguish between on and off states is 300 to 700 millivolts, so when we run a 1-V power supply, it doesn't give us much headroom to operate the device when it's active. What's more, with CMOS, we've reached the point where bipolar was 20 years ago, where standby power—the power that is dissipated when a device is on but not in use—is a very significant quantity. That makes it a far from ideal switch. So the design box is very tight. Unless you boost the power supply, you can't get the frequency you want—but then you run into thermal and cooling limits.

About six or seven years ago, single-core CPUs ran into a brick wall because of those constraints. The solution generally adopted was the multicore CPU running parallel threads, each at lower frequencies, requiring lower active power. But where, when, and how will that solution path run out? It's not clear. At some point in the next 20 years, though, it's not going to be possible to move forward just by multiplying the number of cores. Semiconductor engineers are already talking about "dark silicon": cores that can't be activated in a processing task because of power limits—limitations that are generally application-driven.

Power versus performance also has reliability implications. As operating temperatures rise, the degradation of chip materials accelerates.

A radically different approach starts with the question of whether all the elements in a processor have to be transistors, considering that much of the time all the device is doing is moving data around. A promising device is the so-called cross-point switch, a programmable "crossroads" made with semiconductor manufacturing technology. Similar development paths can be expected for it and for resistive random-access memory elements (RRAM). Many companies (TI, IBM, Toshiba, and others) are developing RRAM switches, which are lower-power and have very high on-off ratios: When they're off they have very high resistance, and when they're on, very low resistance. The main developmental challenge with them is durability: With switching, the on-off ratio deteriorates. So as we approach 2020, the semiconductor industry is faced with many fundamental challenges to address and solve. Conventional scaling faces limits of size and basic physics. New technologies have to address concerns about device reliability and lifetimes and overcome or circumvent probable disruptions in the design flow that standard silicon has up to now done so well with. ■

Paul A. Farrar Jr. is vice president, Albany Expansion & Strategic Initiatives, Systems Technology Group, at IBM in Hopewell Junction, N.Y. He previously was vice president for process development at IBM and was responsible for unit process, lithography, and back-end-of-line development and characterization. He also managed IBM's R&D partnerships at Albany Nanotech and was instrumental in shaping IBM's collaborative development model. William R. Tonti is the Director of Future Directions at the IEEE. Retired from IBM, he held the positions of IBM senior technical staff member and senior manager. He holds over 250 U.S. patents and is a Fellow of the IEEE.

FUTURE DIGITAL MEDIA

BY NAOHISA OHTA

The digitization of creative and work flow processes—whether we are talking about the making and distribution of a high-definition film, architectural blueprints and medical procedures, or industrial processes—is already nearly complete. By 2020, work flows created in and preserved in digital media will be universal and invisible, taken for granted. But for that to happen, wide-ranging innovations will be required in imagers and sensors, digital processing of images and sound, high-speed data transmission, distributed file systems, content management, human-device interactivity, telepresence, displays, and storage, just to name the most obvious and important areas.

To look at present and future digital media in terms of bandwidth requirements, in the last few years we have learned to handle 4K cinema (4,096 pixels by 2,048 pixels), which is equivalent to 35-mm film and requires a bandwidth of about 1 gigabit per second. I personally was involved in the first transpacific 4K motion-picture transmission, from Tokyo to San Diego; other such demonstrations have taken place in the meantime. By 2015, we'll be capable of transmitting the digital equivalent of 70-mm film at close to 10 Gb/s. And by 2020, we'll know how to do 100 Gb/s, for transmission of the Imax films displayed on curved walls, holographic displays, and planetarium presentations of the kind shown on domed ceilings.

It used to be the case, not so long ago, that the medium for domed platform displays was still the analog film. Now the medium is digital, with resolution approaching 4,000 by 4,000 pixels. Such content will soon be shareable over networks.

To do that, however, will require "green" routers based entirely on optical switching and techniques for encoding and decoding very-high-quality imagery at rates of 10 to 100 Gb/s. To put that number range in perspective, the MPEG-2 moving picture standard, which emerged a decade ago as the heart of high-definition television systems the world over, compresses an initial bit stream of about 165 megabits per second to 12 to 20 Mb/s for broadcasting.

So the transmission data rates we're anticipating in the next two decades represent an enormous challenge. I have characterized its draw as pulling us toward a much more accurate representation and transmittal of reality. Another major challenge, which we might characterize in terms of the desire for eternity, is the preservation of digital data for 100 years or more, without fear of its degradation.

At present, the integrity of digital content cannot generally be guaranteed in any medium for more than a decade. Information stored using Sony's Blu-ray system is said by some to be good for a century or more, but nobody really knows what effect moisture and other quirks of the atmosphere may have.

Because of such uncertainties, large archives like the U.S. Library of Congress have to keep copying their voluminous files onto new tapes or optical disks, a wasteful, time-consuming, and somewhat accident-prone process. Were they to use hard disk drives, which is normally done only when it is needed to provide access to the stored material on an ongoing basis (as in data centers), then power must be fed continuously, which is even worse.

The problem is of almost equal interest to film producers. The Academy of Motion Picture Arts and Sciences, in the United States (best known for the Oscars), has a group working on the problem.

Meanwhile, without giving it much thought, we consumers of digital storage services are increasingly storing our most treasured files as attachments sent to ourselves in e-mail messages on services hosted by companies like Google and Yahoo. Apple Computer, having peremptorily informed its customers a decade ago that the floppy disk was on its way to the dustbin of history, is now cutting back onboard storage capacity on its devices, on the assumption that even the most slow-witted consumers will soon grasp that the place to save is the cloud.

But this assumes that the companies maintaining the huge server farms can guarantee the existence of stored data for as long as we may want to access it, which, strictly speaking, they cannot do at present. Google may be constantly replacing and replenishing its hard drives, but the preservation of the information on them is not guaranteed. To assure its preservation, we will need low-power, highly robust servers of a kind yet to be invented. ■

Naohisa Ohta, a professor at Keio University's Keio Media Design (KMD), directs projects to engender new experiences and business using high-quality digital media and networks. He has worked at NTT Laboratories, where he has researched and developed algorithms for high-quality audio-video communications and highly parallel processing systems, and he directed R&D on high-quality AV transmission for real-time networked applications at Sony.

E-HEALTH AND BIOCOMPUTING

BY YONGMIN KIM AND GUDRUN ZAHLMANN

Today's reality in health care is that a lot of records and prescriptions are still written by hand, often barely legibly, and medical histories are often not readily available except to those who compiled them in the first place—assuming the doctor's office or clinic has not in the meantime suffered a fire, flood, or some other natural disaster. Much of that will change dramatically over the next 10 to 20 years, and yet the change will not be total or universal.

It seems safe to say that by 2020 all medical records and doctor's orders will be created electronically, at least in the advanced industrial and rapidly industrializing countries. Much more medical information for individuals will be integrated in computer files and readily accessible just about anywhere and at any time, once security access queries are satisfied.

But much of the health delivery system will still be fragmented, and many countries, clinics, and clinical systems will still handle their records in ways that prevent them from being instantly retrievable, regardless of what system a patient is in. It will be at least another decade until medical records are fully integrated and made universally accessible throughout and among all developed countries. This will take even longer among the less well-off nations, where basic health concerns take higher priority.

Two models for electronic record keeping have emerged in recent years. In the more familiar one, a hospital or clinic, a network of health-care providers, or even a national health-care system keeps uniform records that are fully accessible internally in the system. Sensitivities about personal information and data security naturally lead to restrictions on access. Yet these systems have amply proven their worth, both in terms of the efficiency and effectiveness with which medical services are provided and in terms of preventing medical errors.

In Europe, the Scandinavian countries and the Netherlands have pioneered such systems. Sweden uses the same unique personal identifier that is used for tax records. Norway has made strides incorporating distance medicine and telemetry into its system because of its scattered population in the country's far north. Though the United States lags far behind in this area, organizations that have digital record keeping—notably the United States Department of Veterans Affairs—have markedly improved the quality of health care.

An alternative model, which Microsoft and others have been exploring, developing, and promoting, is client centered. In this kind of system, it is the individual who is ultimately responsible for maintaining an integrated case history, which can be more or less comprehensive. (Some may want to include information, for instance, about a particular massage technique they have found to be helpful, while others may prefer not to include such details. If the patient's visit to a psychiatrist is to be covered, the insurance company will need to know about it, but does the patient's employer need to know as well?) One obvious advantage of the patient-managed system is that the individual can take full control of all issues concerning confidentiality and security. But it also puts more responsibility on the patient than some care to (or can) assume.

Because of such considerations, the integrated electronic health-care systems that evolve over the next two decades will surely contain elements of both the clinic-based and personalized approaches. Perhaps the health-care provider should offer patients the opportunity to take the patient-managed approach, but it's not to be expected that all or even most patients will do so. Naturally, preferences vary not only personally but culturally and regionally. In Germany, for example, it proved impossible to introduce a standardized smart card with information concerning matters like emergency care, because of intractable disagreements about what specifically should be on the card. In developing its patient-centric HealthVault product, Microsoft has initiated research in Asia to find out how preferences about what belongs in a "comprehensive" health record differ from those in Europe and North America.

Ironically, those who would benefit most from the personalized approach—the elderly—are least well equipped with the computer skills to manage it. Though younger people have the savvy and can be expected to take responsibility for their records in time—but a long time!—it will probably take decades for patient-managed record keeping to become widespread. Meanwhile, however, mobile health apps—networked body sensors, emergency alarm systems, and so on—will become increasingly a part of our lives. Those systems will somehow be integrated into electronic health-care record keeping.

To a great extent, the fully computerized health-record-keeping system of the near future, however it is organized, will involve the wide application of existing technology. But the role of computers will not be confined merely to record keeping and the accurate transmittal of commands. As much more information is compiled and standardized, there will also be room for computers to perform analytic tasks autonomously, make suggestions, and offer physicians diagnostic alternatives.

The words used to describe conditions, procedures, and outcomes differ widely, however. So standardizing best practices and agreeing on common terminology is a nontrivial task. Even compiling accurate and complete tables of medical synonyms would not be a simple job.

Once much more information is compiled, integrated, and made universally available, however, computers can take on still more challenging tasks. They can comb the information for unsuspected connections, alert doctors to diagnostic possibilities, and warn them of hidden consequences. Some sense of this can already be seen in e-prescription systems, where computers are able to spot harmful drug interactions, detect histories of overprescription, and check the plausibility of dosage levels.

In addition to Microsoft's efforts, Google, IBM (with its "information-based medicine" brand), and Oracle are exploring how informatics and cloud computing can be brought to bear on health-care delivery problems. All the major IT companies are getting involved.

The drug discovery process will also benefit from computerized techniques and cloud computing. The process requires the storage and transmission of massive amounts of data. In the not very old days, pharmaceutical companies would literally deliver truckloads of data to regulatory agencies like the U.S. Food and Drug Administration. Now the agencies require all such data to be stored and submitted electronically. Cloud computing resources can be summoned not only to store data but to do searches of existing medical histories to determine which individuals might benefit from specific therapies or drugs. And applications can be found in the cloud to perform image analysis, a promising technique in molecule searching, which in turn is a big part of drug discovery.

Take the breast cancer known as HER2-positive, which can be stopped in its tracks if receptors on the exterior membranes of the cancer cells are somehow blocked by the right molecules. There are an immense number of candidate molecules to do the job, so why not look for them by means of massively parallel computing, available in the cloud? Such work, which used to take a year, may now be completed in hours. ■

Yongmin Kim is a professor in the departments of bioengineering and electrical engineering at the University of Washington, in Seattle. He is director of the university's Image Computing Systems Laboratory. He has chaired SPIE medical imaging conferences and has served EMBS in a variety of leadership roles. Gudrun Zahlmann is director of imaging infrastructure at Roche Pharmaceuticals, in Basel, Switzerland. She previously directed pharmaceutical imaging at Siemens. She has a doctorate in biomedical engineering and has been a leader in IEEE's Engineering in Medicine and Biology Society (EMBS).

SMART GRID

BY GEORGE W. ARNOLD AND WANDA K. REDER

The smart grid means using information and communications to make electric power delivery systems more efficient, flexible, and dynamic, to save energy, and to accommodate intermittent renewable sources of energy as well as electric vehicles (EVs) and plug-in hybrids (PHEVs).

There's a stark contrast between the grid as it exists today and other sectors such as telecommunications. In their essential features, power delivery systems are not much different than they were a hundred years ago. What is more, the information and communications technology that's starting to be applied in the grid is to a great extent similar to what was done 30 years ago to telecom infrastructure.

In a nutshell, the smart grid represents the convergence of technologies that control the flow of electrons on the one hand and bits on the other. Electrons, in comparison with pieces of information, are much less tolerant of impairments and manipulation—delays, latency, compression, coding, and so on. To remain stable, the grid must be balanced in real time at all times.

Smart grid technologies are being introduced at a time when power systems are in urgent need of renewal and expansion anyway. In the advanced industrial countries, the average age of transformers and substations is typically about 40 years. In some of the rapidly industrializing countries, such as India, electric infrastructure is inadequate and unreliable. And in some of the world's poorest countries, there's no infrastructure at all. About a billion and a half people have no access to electricity, according to a United Nations agency.

In the next 20 years, as systems are incrementally improved or built out, techniques will be borrowed from control system theory and IT to improve performance in transmission and distribution, which will tend to converge somewhat. Intelligence will be embedded in the grid, supported by a great many sensors, so that the system can self-correct and negotiate its way around problems. This is not trivial: Using such intelligence as the basis for modifying the behavior of the system is like changing the operation of a Boeing 747 in midair. It's going to take time to demonstrate that we actually know how to reliably control a stochastic system.

Engaging the consumer usefully in the workings of the smart grid is another major challenge. Smart meters are being installed by the millions around the world (in the United States, the 2009 stimulus bill provided billions of dollars for smart grid programs). But for the customer and society as a whole to benefit, consumers must have visibility into their energy use and the tools to manage it. So far, many utilities and energy providers do not seem to have kept this priority front and center in their planning.

Though a lot of innovation will take place at the granular level as smart grid technologies are introduced, the general features of the grid will not change radically in the next two decades; we will not see the kind of architectural sea change we've witnessed in telecom during the last three decades.

When there is a sea change in electric power, if there is one, it will be from centrally generated power to ubiquitous distributed power. Every home will have a solar array, personal windmill, or fuel cell, and each such system will feed energy into the grid at times as well as draw power from it. The grid as a whole will be an enormously dynamic federation of microgrids, in which supply and demand are constantly being rebalanced in response to price signals and physical constraints. This vision is more likely to be realized 30 or 40 years from now than in the next two decades.

That said, there are several potentially disruptive technologies that could accelerate the evolution of the smart grid. One would be cost-effective storage. With EV and PHEV development programs driving a lot of R&D in batteries, some manufacturers are starting to offer grid-scale arrays, drawing on new technology developed for cars.

The theoretical potential is huge. About half the power grid's generating capacity, built to handle peak loads, is unused on average, and automobiles are parked 90 percent of the time. So if car batteries could store energy for the grid and feed it back as needed, ubiquitously, it would be a revolution in power. What may be needed most is not so much technical innovation in the narrow sense but innovation in business models and regulation—think of Google and Facebook, where business ingenuity has been at least as important as engineering.

Batteries are not the only possible disrupter in energy storage. Fuel cells, tapping the thermal energy stored in buildings, pumped hydro, and compressed air are other candidates.

Whenever you graft existing technology onto another technology where reliability is absolutely paramount, it takes time, money, effort, and talent to innovate. This is the main constraint on the smart grid. The U.S. utility industry at present spends about 0.3 percent of its revenues on R&D, literally less than the dog food industry spends for research. The Center for Energy Workforce Development has found that in the next five years, about half the U.S. workforce in electric power will be lost to attrition. The situation appears to be similar in other advanced industrial countries, from Australia to Europe.

To begin remedying the situation in the United States, a 2007 National Science Foundation workshop was launched: the IEEE/U.S. Power and Energy Engineering Workforce Collaborative. The group's April 2009 report recommended doubling the rate at which degrees are conferred in power engineering, hiring 80 new university faculty members in the next five years, and increasing federal research funding. The current acute shortage of seasoned experts affects not just implementation of the smart grid but regulation of it as well. Supervisory authorities at all levels lack the talent they need. And although standards setting has gone rather well, there is a reluctance to give up proprietary approaches and adopt procedures that will make systems fully interoperable with others.

Cultural and subcultural barriers will have to be overcome. Power engineering, communications, computing, and IT all have their separate vocabularies, their special ways of talking. So there's a social aspect of getting people with diverse skills and backgrounds to work together meaningfully.

Traditionally, utilities take a very conservative and cautious view of new technology. Equipment depreciates over 20 to 30 years, and amortization rates are key to rates of return. So you don't make rash decisions, and robustness and reliability are still paramount.

In communications, conversely, there's been a trade-off of reliability for functionality in recent decades: Because there's so much redundancy and so many enticing new features, we're willing to put up with impairments that would have been intolerable when a five-nines philosophy ruled (99.999 percent reliability). But now telecom engineers are entering power system control, where five-nines still prevails. ■

George W. Arnold is national coordinator for smart grid interoperability at the National Institute of Standards and Technology. He joined NIST in September 2006 as deputy director of technology services after a 33-year career in the telecommunications and information technology industry. Arnold has served as chairman of the board of the American National Standards Institute (ANSI) and as president of the IEEE Standards Association; he currently is vice president for policy with the International Organization for Standardization (ISO). Wanda K. Reder is vice president of the power systems services division at S&C Electric Company in Chicago. She is the immediate past president of the IEEE Power & Energy Society (PES) and has served on the IEEE/PES governing board since 2002. She leads the IEEE/PES Scholarship Plus Initiative. She is also the chairperson of IEEE Smart Grid and a member of Energy Secretary Steven Chu's Electricity Advisory Committee.

ELECTRIFIED VEHICLES

BY MICHEAL AUSTIN AND RUSS LEFEVRE

Changes in vehicular transportation will be driven in the next 10, 15, and 20 years by escalating fossil fuel prices. BYD Automobile Co., of China, and Toyota predict that gasoline will be selling for more than US $5.50 gallon (or roughly 1 euro per liter) in 2015, with no way back to lower costs. Strains on global fuel supplies will originate mainly in China and India, where burgeoning middle classes are eagerly seeking car ownership.

If every middle-class family in China—an estimated 30 percent of the 1.3 billion Chinese are classified as middle class—were to buy a car, that would be equivalent to at least 100 million sales. It would take 10 years to sell that many cars in the United States, at current projections. In fact, private vehicle sales in China exceeded U.S. sales for the first time in 2009; in 2010, there were about 18 million automotive sales in China, versus 11.5 million in the United States. In 2015, the Chinese are expected to expand production capacity to 25 million cars, and their government proposes that 10 percent of the vehicles should be electric. By then, annual domestic oil production will be 500 million tons of oil short of consumption.

Governments around the world have responded to the emergent fuel crisis with concerted programs to encourage higher automotive fuel efficiency and accelerated conversion of vehicles to electrified or partially electrified propulsion systems. The European Commission has a Green Cars Initiative, one of three private-public partnerships with total funding of 5 billion euros. The Obama administration allocated $3 billion in the 2009 stimulus bill to automotive electrification and battery development and has set a target of 1 million electric vehicles on the road by 2015. China has set some of the world's most stringent automotive fuel efficiency standards.

The case for electric vehicles is compelling in terms of engineering and the environment. In principle, the "tank to wheels" efficiency from electric propulsion is three times that of an internal combustion engine (ICE). At a per-gallon gasoline price of $3.00 and a per-kilowatt-hour electricity price of 10 cents, the operating cost of an EV beats that of an ICE-powered car by a factor of four. At those prices, replacing a standard car with a hybrid saves the owner $10,000 over the automobile's lifetime.

The main immediate reasons countries are fostering conversion to EVs and hybrids are to save on oil imports, to boost what's almost universally seen as a critical industry of the future, and to bolster national security. In the longer run, saving on greenhouse-gas emissions and public health will also be important considerations.

Of course, electrification of vehicle fleets is not the only possible answer to tight global oil markets and dependence on unreliable suppliers. Biofuel production and flex-fuel vehicles can also do the job. But increased biofuel production is already putting undesirable pressure on world food prices, and biofueled automotive transport—in contrast to electrified transport—does not yield impressive cuts in carbon emissions. Natural gas could in principle substitute for gasoline, whether burned directly in ICEs or used as feedstock for fuel cells—but really, there's not much of a future for any finite energy source.

A major U.S. study conducted by the Electric Power Research Institute and the Natural Resources Defense Council examined nine U.S. plug-in hybrid scenarios, based on different rates of vehicle electrification and various electricity-generating mixes. All nine scenarios produced significant reductions in greenhouse-gas emissions, across all the nation's regions. The decrease in U.S. annual emissions could be as high as 612 million metric tons by 2050, and the cumulative cut in the four decades to 2050 could exceed 10 billion metric tons.

The main issue, looking at the two decades immediately ahead, is how fast vehicle electrification will take place and, in particular, when the transition will take off. The crucial factors here are progress in the development of car batteries or other storage devices, the capacity to provide low- or zero-carbon electrical energy, the build-out of a ubiquitous rapid-charging infrastructure, and the modification of power grids to accommodate charging at home, at the office, and on the road.

Even taking into account the dramatic improvements seen in cell phone batteries in the 15 years or so since mobile telephony took off, progress with car batteries is likely to be incremental, absent revolutionary innovations. Battery developments certainly have not followed a Moore's law curve, and so it remains to be seen whether progress will be sufficient in the next 10 years to stimulate an EV takeoff. Yet even so, today's car batteries already are at the boundary of making the EV and plug-in hybrid a viable economic proposition.

Let's say you have a 2-hour commute from home to work and back: Your 100-kilometer drive will require 12 to 14 kilowatt-hours of energy. Well, BYD's e6 already comes with a battery capable of storing 60 kWh, and using fast-charging systems being introduced in China, that car can be ready to go in under 30 minutes, with a range of over 250 km.

So, arguably, car batteries already are where they need to be in terms of storage capacity. The real issue is price. Analysts at Lux Research predict that the average cost of battery energy storage will drop from $900/kWh today to $500/kWh in 2015. BYD, which also makes batteries, is already selling battery systems at $500/kWh; in 10 years, the energy storage cost will be $350/kWh. The iron-phosphate battery BYD puts in its own vehicles, though heavy (it has 85 percent of the energy density of the more commonly used lithium-ion-cobalt types), has significant advantages in terms of safety and durability.

The optimal generating system for charging car batteries would consist of homes and businesses custom-built to collect solar energy, which can be fed to the grid when the vehicle is not charging, and with the car serving as a storage device that can also feed power to the grid on demand. KB Home, in Los Angeles, has developed 2,000-square-foot houses that are equipped with 18 polycrystalline photovoltaic panels capable of generating 3.9 kW and that sell for about the same price as comparable conventional homes—about $275,000.

Still, except perhaps for the unusually sunny regions—the Iberian peninsula, the U.S. Southwest, and so on—photovoltaic electricity will not be in a position to really take off as a major source of vehicular energy for close to another 10 years. In the meantime, the major zero- and low-carbon generators will be wind, natural gas, and nuclear power, delivered via the grid.

The power system itself is a major enabler for electric cars. Many cities around the world are building out charging stations, mostly Level 1 (110 to 120 volts, up to 30 amperes) or Level 2 (220 to 240 V, 30 to 70 A). Though even more rapid charging is desirable, there is no consensus as to what the ultimate configuration should be—AC or DC, what voltage level, what power level, and so on. Japan has developed a DC fast-charging system (CHAdeMO) that is quite popular but has not yet been picked by the United States.

China has adopted an even more ambitious target for faster charging (less than 20 minutes at 480 V and up to 600 A) and is building a network of such stations to service its electric taxi and e-bus fleets. Beijing is slated to spend $15 billion on electric vehicle infrastructure, and Shenzhen—the booming high-tech metropolis created out of nothing, between Hong Kong and Guangzhou—intends to have 22,000 stations by 2012.

To work optimally, an electrified vehicular system will depend on a highly flexible and adaptive power grid. So, in a word, the biggest single set of EV enablers is the smart grid in all its dimensions. The main problem specific to electric vehicles is the distribution system transformer. Several studies have shown that just two cars charging simultaneously in the evening could blow out the local distribution circuit; a particular problem is that many distribution transformers are designed to cool during off-peak nighttime hours.

As electric and hybrid electric cars become widespread (which will surely happen by 2020) and then ubiquitous (perhaps by 2030), there may have to be wholesale replacement of distribution of transformers and a redesign of distribution networks. But with the distinction between distribution and transmission disappearing as the grid becomes more communicative and interactive, much of that will happen anyway. ■

Micheal Austin is vice president, BYD America. He previously was at Motorola for 15 years, where he was responsible for engineering and program management, technical operations manager, new business development manager, and senior director for ODM procurement. Russ Lefevre is a former lead engineer at Hughes Aircraft and R&D leader at Technology Service Corp. He was an IEEE-USA Congressional Fellow in 2001, advising Senator Jay Rockefeller, and he is a past president of IEEE-USA.