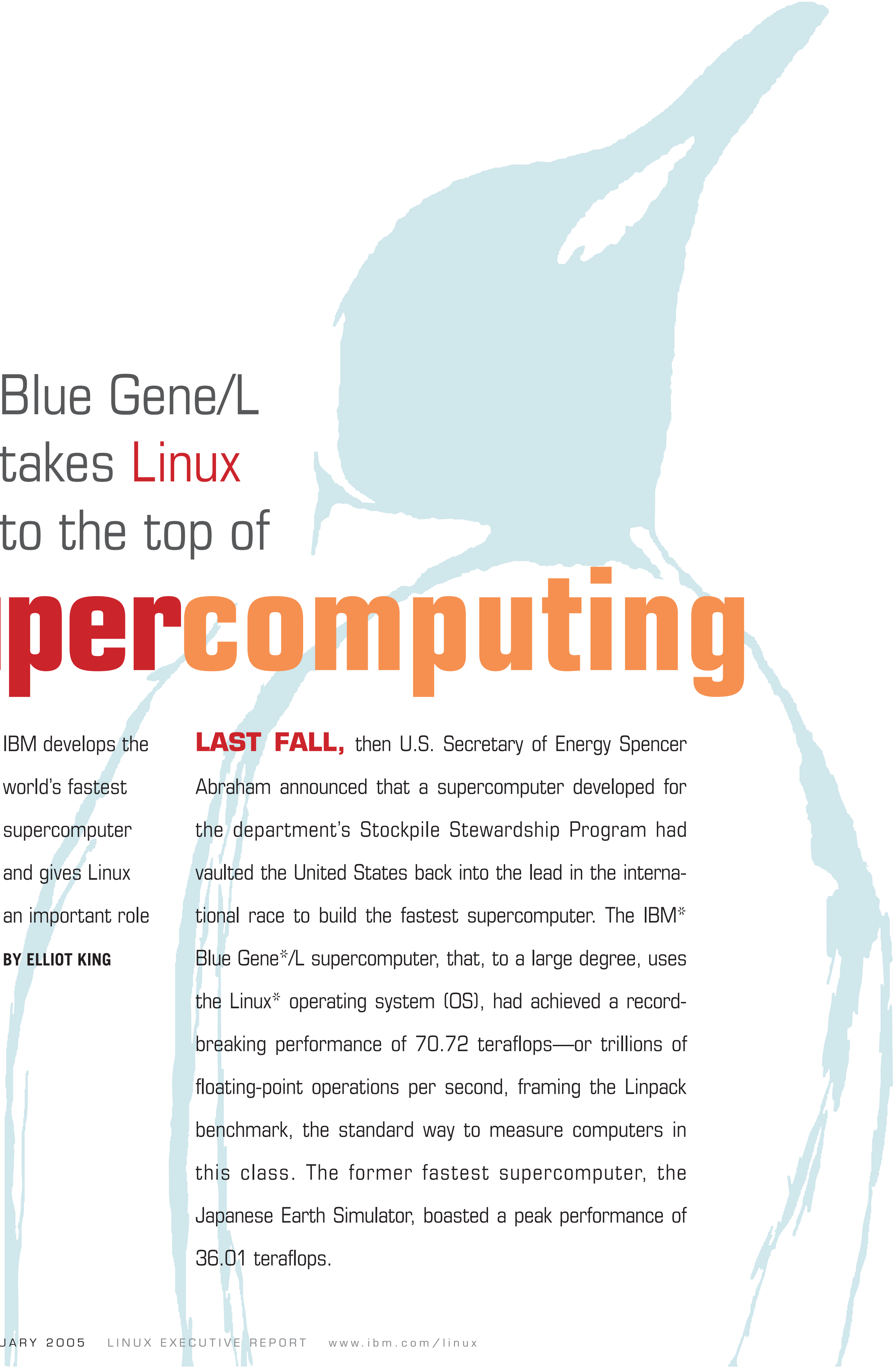


Blue Gene/L takes Linux to the top of supercomputing

IBM develops the world's fastest supercomputer and gives Linux an important role

BY ELLIOT KING

LAST FALL, then U.S. Secretary of Energy Spencer Abraham announced that a supercomputer developed for the department's Stockpile Stewardship Program had vaulted the United States back into the lead in the international race to build the fastest supercomputer. The IBM* Blue Gene*/L supercomputer, that, to a large degree, uses the Linux* operating system (OS), had achieved a record-breaking performance of 70.72 teraflops—or trillions of floating-point operations per second, framing the Linpack benchmark, the standard way to measure computers in this class. The former fastest supercomputer, the Japanese Earth Simulator, boasted a peak performance of 36.01 teraflops.

The performance breakthrough represented by Blue Gene/L is even more dramatic than dry statistical comparisons indicate. Two earlier Blue Gene/L prototypes had already cracked the Top 10 list, and when Abraham made the announcement, he noted that a computation that would take the number three supercomputer on 2003's Top 500 list 30 days to complete could be completed on the current iteration of Blue Gene/L, which is one quarter of the final size, in three days. The final Blue Gene/L system scheduled for delivery to Lawrence Livermore National Laboratory will exceed the performance of the Japanese Earth simulator by a factor of seven. The goal of the project, says David Turek, vice president, Deep Computing at IBM, wasn't to simply make an incremental step in the marketplace, but to make a major step forward. And it has.

Making a Supercomputer

Interestingly, according to Turek, the Blue Gene project was launched with just a tad of deception. In 1999, IBM announced its plans to create a machine to tackle what was then described as one of the grand challenges of science—an effort to simulate protein folding. Grand challenges in science had been the organizing theme for much of supercomputings development through the 1990s. The idea behind the grand challenges was to

ILLUSTRATION BY JOANNA SZACHOWSKA

marry breakthroughs in high-performance computing (HPC) with specific applications in the life sciences, geophysics and other compute-intensive applications

While sequencing the human genome has received the lion's share of publicity concerning the application of HPC and biology, protein folding is a more complex problem. Simply put, genes express themselves through proteins and protein folding controls the body's physical processes. Abnormal protein folding causes diseases. By simulating protein folding, Blue Gene would be able to help determine the cause of diseases. It could also be used to assist in the development of new drugs and other means to cure disease.

Nevertheless, IBM never really intended for Blue Gene to only be a "protein-folding machine," says Turek. "That was the target because we needed an application that everybody could understand and interest in the life sciences was extraordinarily high." And, it was an application that would stretch the architectural design of supercomputing.

The first criterion for the new machine was speed. "We wanted an application that genuinely needed a petaflop of computing," Turek says. IBM also wanted to create a reliable system that didn't regularly fail. And, it wanted to build something that would have a more general use.

In fact, says Turek, "For the past two-and-a-half to three years, we've been looking at the applicability of a wide-ranging set of applications for the platform. We've used the insights garnered from those investigations to shape the design of the systems since then." Last summer, the Blue Gene effort at IBM made the transition from research and development to product-development status.

The Many Uses of Blue Gene

IBM anticipates that supercomputers based on Blue Gene technology ultimately will be used in commercial applications. After all, notes Turek, past generations of HPC technologies have been implemented by financial-services companies for risk-management applications. Product design, oil exploration and logistics can also require massive amounts of compute power.

As expected, however, the initial users of the Blue Gene technology come from the academic-and government-scientific research community. "We're interested in Blue Gene/L for two reasons," says Steve Louis, assistant department head for Integrated Computing and Communications at the Lawrence Livermore National Laboratory, which is one of IBM's key partners in the Blue Gene development effort. "First, its design is well-matched to some of our most important applications and second, it addresses critical issues on the path to petaflop computing." Those issues are, he says: power, floor space, cost, single-processor performance and network scalability.

Each is significant in its own right and Blue Gene/L has made strides in each area. Ironically, it's accomplished these goals in an unorthodox manner. The hype surrounding most new high-performance computers usually focuses on higher performance CPUs, advanced interconnects and more memory. Blue Gene used slower CPUs and less memory per node to go with its innovative interconnect technology.

Scalability is perhaps the most important aspect of the Blue Gene/L project. In many ways, Blue Gene represents what can be understood as the return of the concept of massively parallel computing. When the complete machine is delivered this year, it will be able to scale to 65,536 (64K) processing nodes with two CPU cones per node.

The idea of massively parallel computing isn't new. In the 1980s, innovative companies such as Thinking Machines built early generations of the technology. At that time, however, the technology didn't live up to its promise. "For awhile in the mid-1990s, massively parallel dropped from 64,000 (nodes) to 64," observes Richard Loft, formerly an engineer at Thinking Machines and now the associate director of the Scientific Computing division at the National Center for Atmospheric Research (NCAR), where he's also in charge of the computational science section. The scientific computing division has traditionally been responsible for all of the supercomputer assets that NCAR employs for simulating the world's climate and meteorological applications, as well as space, weather and oceanographic applications. "In people's thinking, the focus shifted from large parallel computers in processor count to smaller computers," Loft says.

> **"For the past two-and-a-half to three years, we've been looking at the applicability of a wide-ranging set of applications for the platform."** —David Turek, vice president, IBM Deep Computing

Spreading the Word

With Blue Gene, the thinking has shifted back to larger scales. For example, Louis says the laboratory wants to focus on the physics and materials part of the Advanced Simulation and Computing (ASC) program operated by the National Nuclear Security Administration of the U.S. Department of Energy. "There is a major program element in Advanced Simulation and Computing that focuses on physics and materials science," said Louis. "That's one of the focuses we wanted to have for this machine."

The applications initially targeted included compute-intensive problems in classical and first-principles quantum molecular dynamics, as well as dislocation dynamics for material strength, research hydro dynamics, instability and turbulence. Those applications were selected because they were important to ASC and they had enthusiastic support among the scientists at the laboratory, according to Louis. He adds, "They had very good potential for code scaling. Blue Gene/L has a particular kind of architecture and we tried hard to target the kinds of codes that we thought would be able to best take advantage of the way this machine was designed."

Scientists at Lawrence Livermore have also spread the word about Blue Gene/L's capabilities. For example, they encouraged a team of scientists at the University of Chicago's

> **"We're interested in Blue Gene/L for two reasons. First, its design is well-matched to some of our most important applications and second, it addresses critical issues on the path to petaflop computing."**
>
> —Steve Louis, assistant department head,
> Integrated Computing and Communications,
> Lawrence Livermore National Laboratory

Flash Center to think about ways to use the Blue Gene/L prototype and urged the scientists at Chicago to explore the opportunity. The Flash Center is an ASC Academic Strategic Alliances Program (ASAP) center and the Flash code project was launched by the Department of Energy to study thermonuclear flashes on the surfaces of supernovas. “We’re trying to figure out how stars explode; how they go supernova,” says Katherine Riley, a senior applications engineer at Argonne National Laboratory, who has worked on the Flash project for the past seven years.

The scientists associated with the Flash project study what are known as Type 1A stars, which are very dense and steal mass from neighboring stars, which are generally younger and larger. As the star steals mass, it builds up material on its surface, which results in nuclear explosions on the star. “There are really large nuclear explosions similar to the mechanisms of nuclear weapons,” Riley says.

While these phenomena have been observed, the processes that lead to the giant explosions that light up the night sky are still unknown. The Flash project scientists run simulations to try to determine the mechanism. “We’ve come up with some new and promising results,” says Riley.

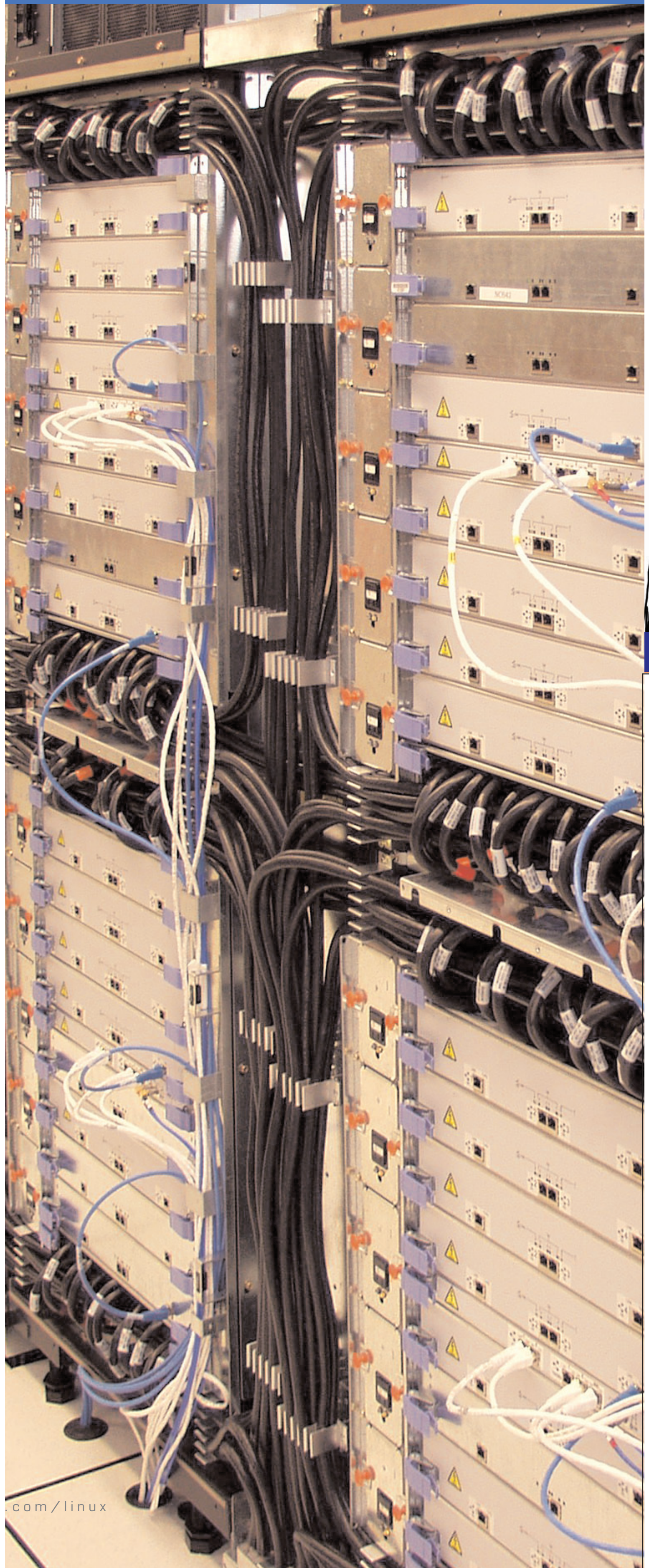
In past simulations, the scientists would let the star burn for a period of time. Then they would manually program an explosion to see the results. “What Flash has been able to do is form first principles to have a star explode on its own,” she says. “It’s a very cool result and this is what Flash was designed to do, study stars and make them explode.”

Over the past several years, the Flash team has had access to the largest supercomputing systems in the world. “Flash is an interesting problem because of the problem size,” Riley notes. “The amount of data is so large that you need to make sure your code can deal with all of it.”

To run the star simulation, the computing must occur on a very fine grid, which changes while the simulation is underway. The greater the resolution for problems like this, the more interesting the physics taking place. Indeed, in physics there’s a well-known scale that ranges from kilometers to the nuclear. To deal with problems that must be resolved across such a wide range requires the cutting edge of computing power.

“With such a large number of nodes that can scale and an interconnect that allows you to use all of the nodes effectively, Blue Gene will be able to do simulations with a resolution and study physics; that hasn’t been able to be done before—period,” Riley says.

Simulations will be conducted along a much broader range of physical scales. In the past, Riley says, the machines have simply not been able to manage the problem. But with 64,000 processors, or even 16,000 processors, “we’re going to be able to attack these problems and bring them to finer detail than we’ve ever been able to before,” she says.



The IBM Blue Gene/L supercomputer has achieved record-breaking performance.

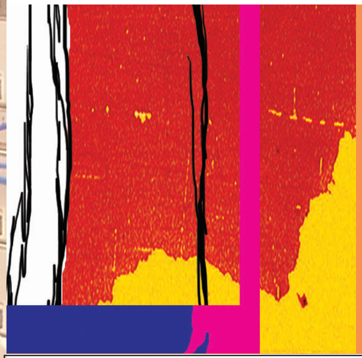
The Linux Factor

Blue Gene/L is much more than just a large Linux cluster. Two of the primary challenges that Blue Gene/L had to overcome to achieve its degree of scalability are power management and size. “The problem with large super computer systems is that they consume a lot of power and space. And they are relatively expensive,” says NCAR’s Loft. “The facility’s costs have become obvious to us as a hidden cost, sitting on top of cost of the rest of the system.”

Blue Gene/L has solved many of those problems. It’s a low-power technology that delivers a lot of “fuel efficiency,” Loft says. “We measure fuel efficiency in terms of sustained flops-per-watt. That’s how you efficiently turn electrical energy into computations.” By that metric, Blue Gene/L is 10 times more efficient than a common Linux cluster.

While people haven’t traditionally tracked computing fuel efficiency, they’ve been caught by surprise by the leap in operating costs as computers have gotten larger. IBM’s Turek estimates that if the approach used in the NEC Earth Simulator computer was pursued to achieve petaflop computing, the operating costs would run in the hundreds of millions of dollars and it would require a huge amount of space, perhaps as much as two football fields. The electrical bill alone could be $150 million a year, he says. On the other hand, with Blue Gene/L, petaflop computing will cost only about $2 million a year in electricity and fit in the space of, perhaps, a tennis court. While that isn’t cheap or small, it’s manageable.

Within the scientific community, many observers believe that Blue Gene/L will have a dramatic impact. “People are



> **“Blue Gene will be able to do simulations with a resolution and study physics; that hasn’t been able to be done before—period.”**
>
> —Katherine Riley, senior applications engineer, Argonne National Laboratory

rethinking their codes," says Claudio Rebbi, professor of physics and director for the Center for Computational Science at Boston University. Running on Linux, Turek notes, should facilitate the porting of codes.

The Next Steps

Of course, IBM has grand ambitions for a technology that started its life being promoted for use with a single application. The next steps in the product roadmap call for IBM to deliver a machine that can achieve 360 teraflops/second to Lawrence Livermore National Laboratory within the next several months. It also plans to deliver a developer edition to aid scientists who wish to develop code for the machine.

And, over time, IBM officials believe that Blue Gene/L will win market acceptance as a commercial product. In the spring of 2004, for example, the Dutch astronomy group Astron announced a joint project with IBM that will harness the 34-teraflop supercomputing capability of IBM's Blue Gene/L to peer back billions of years, deep into the history and even the birth of the universe. Argonne National Laboratory is also reportedly considering the purchase of a smaller version of Blue Gene/L.

In any case, with Blue Gene/L, IBM has clearly achieved at least three major goals. It's on the road to petaflop computing. It's demonstrated its commitment to this initiative. "In 1999, we committed $100 million to this and we delivered," says Turek. And, it's given Linux an essential role in the most advanced HPC platform in the world.

Elliot King *is an associate professor at Loyola College in Maryland where he specializes in new communication technology. Dr. King has written five books and several hundred articles about the emergence and use of new computing and communication technology.*