The power and promise of parallel computing

by I. Wladawsky-Berger

The use of parallel computing is reaching into both the technical and the commercial computing environments. The success of IBM in this broad area is based on the introduction of flexible and general-purpose families of scalable processors. This essay highlights the history of one of those families that includes the IBM scalable POWERparallel™ SP1™ and SP2™ systems. I also describe how these systems were rapidly developed and moved to market, and illustrate how some customers are using them in their businesses.

In little over a decade, we have progressed from a rather primitive understanding of how to lash processors together to a deep appreciation of today's agile architectures. We have watched the industry embrace parallel computers as the preferred architecture for high- performance computing, and at IBM we have delivered two large-system parallel families, the System/390* enterprise servers and the SP family of Scalable POWERparallel Systems*.

We are also seeing the application of parallel computing broaden beyond the supercomputing centers into the mainstream of business. And we can look into the future and anticipate that the power and affordability of parallel computing will generate revolutionary new applications.

This essay highlights the history of the SP1* and SP2* POWERparallel* machines, briefly discusses the approach used to quickly create and deliver these machines, and illustrates how some customers are using them in their businesses.

Parallel processing at IBM

IBM's program in parallel computing squarely aims at moving parallel processing into the mainstream of science and business. IBM's System/390 Division and the POWER Parallel Division have introduced two lines of powerful yet affordable parallel computers. Both are built with the kind of inexpensive, CMOS (complementary metal oxide semiconductor) technology used in the microprocessors that run personal computers and workstations. And both essentially redefine the world of large systems computing.

One family is the System/390 line of parallel servers, introduced for the first time in April 1994. The System/390 servers run enhanced versions of the Multiple Virtual Storage (MVS) and our other mainframe operating systems. The CMOS microprocessors that run these machines are inscribed with a System/390 "personality" that allows users to run their existing mainframe applications (in which IBM customers have invested $1 trillion over the years), without recoding.

Aimed primarily at our traditional mainframe customers, these new machines allow users to gain the performance advantages of parallel computing and participate in open, client/server networks.

The second family of IBM parallel products is again built with CMOS chips, but with the "personality" of the chips found in the popular RISC System/6000* line of workstations. These scalable POWERparal-

0018-8670/95/$3.00 © 1995 IBM

lel (SP) machines use the same AIX* operating system that powers the RISC System/6000 workstations, and they run almost all of the 10 000 or so applications written for those machines.

Together, IBM's two families of parallel computers offer an affordable way to scale to nearly unlimited computing power. And, regardless of the customer's choice of operating system, they offer access to open, distributed computing and a wide variety of applications.

A brief yet distinguished history

With this issue of the *IBM Systems Journal* devoted to the details of the POWERparallel SP machines, it seems a good time to reflect on the principles guiding our march into the marketplace, and on a little bit of the history that brought us to where we are today.

We made the decision to launch what became our POWERparallel business in the fall of 1991, because, at that point, a number of factors led us to believe the time was right.

First, the RISC System/6000 workstation, which had been announced the year before, was already becoming a big success. In a very short time, the RISC System/6000 workstation established itself as a superb, attractively priced system with the ability to run many jobs that previously had required a much more expensive supercomputer.

People inside and outside IBM were connecting their RISC System/6000 workstations together with local area networks into "clusters" and using them in place of much more expensive vector supercomputers. Looking at future plans for the RISC System/6000 microprocessors, it was clear that the micros, already very powerful, would over time approach the performance of the fastest vector machines, at a fraction of the cost. We became convinced that, before long, the entire area of scientific and technical computing would move in the direction of RISC-based parallel machines. And the success of the RISC System/6000 family gave us a strong marketing base to build on.

Second, investigations into parallelism conducted by IBM's Research Division were advancing rapidly. Our scientists had begun projects in parallel computing—projects called the Yorktown Simulation Engine (YSE),[1] the Engineering Verification Engine (EVE),[2] RP3,[3] GF11,[4] and Vulcan[5] —in the early 1980s. By 1991 they had developed lots of software, plus very powerful switch and interconnect mechanisms to allow microprocessors to work efficiently together. These technologies would form the cornerstone of our new products.[6–8]

And, third, United States federal government support of parallelism was advancing through the High-Performance Computation and Communication Initiative, which aimed to advance the adoption of highly parallel supercomputers, high-bandwidth networks, and other advanced technologies. We were strongly encouraged to participate, both by IBM groups and by potential partners, including the Cornell Theory Center and Argonne National Laboratory.

In December 1991, we successfully made the case to establish our new supercomputing organization. As our first official action, we established our development facility, which we called the Highly Parallel Supercomputing Systems Laboratory. A group of experts from around IBM was quickly assembled to recommend our product and market directions. After several weeks of hard work, the task force members, representing research, development, and marketing organizations, set the direction we are still following today. As a group we made four important recommendations.

First, because technology and products change so rapidly, time-to-market was critical. We decided that whatever products we developed should be made available to the marketplace in the shortest possible time, in no more than 12 to 18 months.

Second, we recommended that our parallel products should be positioned as an integral part of the RISC System/6000 family. We should use as many common components as possible, including the microprocessors, the AIX operating system, and others.

Third, we felt our parallel computer should be based on the architecture and other technologies from our Research Division's Vulcan project. We adopted the project's distributed memory architecture,[6] which was highly flexible and elegant in its simplicity, and its high-performance interconnect mechanisms.

Finally, we concluded that our AIX-based parallel system should be a general-purpose machine. And,

although we would initially aim the system at scientific and technical applications, it should evolve over time to support a wider variety of applications.

These recommendations became the principles that guided—and still guide—our parallel development. And they form the basis of a very effective strategy for running our business.

Speeding technology to market

Time-to-market guided every decision we made. It was clear that we just did not have time to design every component. Consequently, we drew on the resources of the IBM Research Division, the product division producing the RISC System/6000, and the overall technical community inside and outside IBM.

In February 1992, we set out a significant challenge to our development team: we asked them to bring our first product to market in no more than 18 months. In February 1993, exactly one year after we put the team together, we announced our first product, the SP1. Then, by April of that year we shipped the first SP1s to our development partners, the Cornell Theory Center and Argonne National Laboratory. We began volume shipments of the SP1 in September 1993.

These first, and subsequent, development partnerships have proven invaluable. They advise us on what works and what does not, on which applications are essential, and which are not. These partnerships are with many of the nation's preeminent research centers, including the Cornell Theory Center, Argonne National Laboratory, Fermilab, the Department of Defense (DoD) Supercomputing Center at Maui, and most recently, the National Aeronautics and Space Administration (NASA) laboratory at Ames and at Langley. These partnerships also include the CERN high-energy physics laboratory in Switzerland and the National Cancer Center in Japan. They have all installed our powerful supercomputers, and they are giving back to us information that has proved invaluable in serving the scientific and technical markets.

And the breathtaking pace continued. We announced our follow-on SP2 machine in April 1994, began shipping it in July and, last October, announced further enhancements to the hardware and software. By the end of March 1995, we had installed more than 350 machines in customer locations around the globe, including universities, industries, and businesses.

Part of the RISC System/6000 family

UNIX** workstations and servers had been quite distinct from parallel systems. There was no compatibility between the two and there was a very large price gap. By positioning our SP1 machines as an integral part of the RISC System/6000 family, and by designing them to run all existing RISC System/6000 applications, we "bridged the gulf" between workstations and parallel systems.

Now customers could invest in one architecture and one set of applications for workstations, servers, distributed cluster systems, and parallel systems. And they could invest in an SP1 parallel server with just a few processors (at a very attractive entry price) and later scale up their computing operations by adding many more processors to the parallel server. Essentially, we created a new segment within the global information technology market—scalable parallel computing.

Our parallel machines are part of IBM's strategy to offer a "palmtops to teraflops" family, ranging from the smallest hand-held computer to massively parallel systems with hundreds of processors. All those machines use our RISC-based POWER Architecture* and PowerPC Architecture* microprocessors.

Using the distributed memory architecture

The distributed memory architecture we use in our machines seems almost elegant in its simplicity and flexibility, but choosing that architecture was anything but simple. In fact, selecting the architecture was the most difficult decision we made.

Parallel architecture has been a subject of great interest in the computer science community for many years now. Many different kinds of parallel architectures, switches, programming models, algorithms, and the like have been investigated in universities and research laboratories around the world. Not surprisingly, different individuals and groups had widely different—and strongly held—opinions on how to best develop a parallel computer product.

For quite a while, the big debate was whether to build a single-instruction, multiple-data (SIMD)

stream or multiple-instruction, multiple-data (MIMD) stream computer, two widely different approaches. SIMD is only applicable to certain classes of applications, but for those applications, it is easier to program. The MIMD approach applies to a very large number of applications, but it is more complex to program. The debates raged for several years, until the early 1990s, when the SIMD approach faded and MIMD was accepted as the direction to go.

Another major debate, one that still rages, is whether to use a distributed memory approach, a shared-memory approach, or one of several variations in-between. Shared-memory scalable architectures make the programming easier, but severely limit the scalability of the system. Distributed memory architectures make scaling up the system much easier (since nothing beyond the switch is shared among all the processors), but the programming is harder. That is because software from shared-memory systems, like uniprocessors or symmetric multiprocessors, has to be rewritten or restructured. Sometimes this restructuring is easy, but sometimes it is very difficult. Many commercial applications written for client/server configurations (including those applications written for the RISC System/6000 servers) work well in distributed memory parallel systems, since they have already been broken up to work across distributed processors.

Inside IBM we have pursued a variety of parallel computing research projects in a number of laboratories. Scientists at the Research Division's Thomas J. Watson Research Center organized a first-rate effort in parallel computing, and over the years had explored most major architectures in projects like RP3, GF11, and Vulcan. Most of the Research Division experts strongly recommended that we use the distributed memory architecture of the Vulcan project as we built our scalable parallel family of products. Because the Vulcan project scientists had designed and started development of a number of hardware and software components, adopting the Vulcan approach would help us immensely in our time-to-market objective. This was an added bonus.

In the end, we adopted the Vulcan distributed memory architecture as well as a number of its hardware and software technologies for our SP product. While other architectures offer advantages in one area or another, overall, the SP distributed memory architecture has served us very well, for two main reasons: its simplicity and flexibility.

The simple, elegant architecture of our SP system made it possible for us to bring our first product to market in record time. And it has allowed us to keep this intense focus on time-to-market ever since, by making it possible for us to quickly incorporate the breathtaking technology advances going on in the industry. For example, we are able to introduce the latest microprocessors into our products in just a few months of their first appearance. And, by being relatively simple, our architecture helps us considerably in being able to offer a very attractive price and performance.

The Research Division experts recommended that we use the distributed memory architecture.

In addition, distributed memory architecture has allowed us to build a very flexible system, easily adapted to a variety of applications and configurations. This flexibility is very important. The only certainty about the future of the computer industry is that things will keep changing at a rapid pace, and our products will be called upon to do things we had not anticipated, or even dreamed of.

A system for every purpose

From the very beginning, one of our key objectives was to gain as much market presence as possible. We wanted our systems to be used in the research environments where so much experimentation was going on, but we also wanted them to quickly move to the production environments necessary to achieve wide acceptance and business success. We wanted to support as many parallel applications as possible, including the very large, "grand challenge" applications—problems of national importance so complex that they command the computing capacity of hundreds of processors. But we also

wanted to be able to support the mixed workloads found in most computing centers—even in most supercomputing centers—consisting of interactive users and batch jobs, scalar and parallel applications.

We therefore decided to build a general-purpose, scalable AIX system, capable of handling a variety of workloads and applications. Even for our initial scientific and technical customers, we knew that we had to satisfy a wide variety of requirements, for the SP family to become a very useful, and, thus, successful supercomputer system. Therefore, while we have invested considerable effort in tools and features specifically aimed at parallel applications, we have also invested a lot in systems management, workload managers, file systems, and other features needed to build a successful production-worthy, general-purpose computer.

We decided to build a general-purpose, scalable AIX system.

Our general-purpose base served us very well when we decided to expand beyond technical applications and support a variety of data-intensive applications, as well as commercial applications of all sorts. To do this, we made sure that we significantly enhanced the system's input/output capabilities; we expanded the capability to include relational databases and transaction monitors; and we worked to enable a variety of applications of interest to commercial customers. In addition, we continued to enhance the scientific and technical capabilities of the SP machines, and also the general-interest functions like systems management.

Today our SP2 machines are working around the globe, solving technical and commercial problems. We are already seeing evidence that parallel computing is revolutionizing the way our customers use their computers, primarily because the affordability of our machines now makes it practical to apply large processing power to all kinds of problems. For example, at Dow Chemical Co. in Midland, Michigan, chemists are using a 24-processor SP2 to determine whether a chemical compound they create in the lab can "dock" with disease agents like enzymes, viruses, and bacteria to stop their harmful action. The calculations involved are complex, partly because the interaction of these compounds is not governed by strict laws of nature.

At Western Geophysical in the United Kingdom, explorationists are using their SP machine to create three-dimensional models of underground oil reservoirs by analyzing billions of pieces of data. Those data are collected in seismic tests in a technique where sound waves are bounced off the underground contours of the earth. Computer models are used to predict where oil is likely to lie hidden before incurring the expense of drilling a well.

Hyundai Electronics Industries Co., Ltd. in Korea uses its SP machine to perform crash simulation and analysis and several American auto makers have purchased SP machines for the same purpose. The automobile manufacturers can now predict the crash-worthiness of new designs while the cars are still on the drawing board, saving time and money.

And John Alden Life Insurance Company in Miami is using its SP machine to find the hidden patterns in mounds of health care data. It then advises its clients about which health care plans offer the most reasonable rates, and which hospitals discharge patients in the most reasonable period of time.

Now we are expanding our partnerships with commercial customers, companies like Citibank (a subsidiary of Citicorp), Revlon, Inc., and Morgan Stanley Group Inc., to gain information about what works and what does not work. In conducting their own businesses, these companies are using our SP systems as information servers that will help them attain a real competitive advantage.

Summary

IBM is firmly committed to parallel computing, both with our SP computers based on the POWER Architecture and the AIX operating system, and the companion program that is reworking traditional mainframes into a new line of parallel System/390 computers that run the MVS operating system,

while protecting the huge investment in System/390 applications the world has generated during the last 25 years.

As we look into the future, we see whole new classes of applications for our scalable parallel systems. Multimedia and video-on-demand will leverage the vast input/output and storage potential of the SP machines. Networking applications of all sorts will reach out to provide information and computing to vast numbers of people around the world. Object-oriented technologies will help us develop and modify complex applications with huge productivities. Virtual reality and other advanced user interfaces will help us create whole new classes of applications and will help us reach far more people than ever before.

As we turn the promise of parallel computing into a reality, it will indeed be a "giant leap" that will benefit us all.

*Trademark or registered trademark of International Business Machines Corporation.

**Trademark or registered trademark of X/Open Co. Ltd.

Cited references

1. G. F. Pfister, "The IBM Yorktown Simulation Engine," *Proceedings of the IEEE* **74** (1986), pp. 850–860.
2. D. K. Beece, G. P. Papp, and F. Villante, "The IBM Engineering Verification Engine," *Proceedings of the 25th Design Automation Conference* (June 1988).
3. G. F. Pfister et al., "An Introduction to RP3," *Expermental Parallel Computing Architectures*, J. J. Dongarra, Editor, North-Holland Publishers, Amsterdam, Netherlands (1987), pp. 123–140.
4. J. Beetem, M. Denneau, and D. Weingarten, "GF11—A Supercomputer for Scientific Applications," *Proceedings of the 12th International Symposium on Computer Architecture*, Boston, IEEE Computer Society (1985).
5. C. B. Stunkel, D. G. Shea, B. Abali, M. G. Atkins, C. A. Bender, D. G. Grice, P. Hochschild, D. J. Joseph, B. J. Nathanson, R. A. Swetz, R. F. Stucke, M. Tsao, and P. R. Varker, "The SP2 High-Performance Switch," *IBM Systems Journal* **34**, No. 2, 185–204 (1995, this issue).
6. T. Agerwala, J. L. Martin, J. H. Mirza, D. C. Sadler, D. M. Dias, and M. Snir, "SP2 System Architecture," *IBM Systems Journal* **34**, No. 2, 152–184 (1995, this issue).
7. M. Snir, P. Hochschild, D. D. Frye, and K. J. Gildea, "The Communication Software and Parallel Environment of the IBM SP2," *IBM Systems Journal* **34**, No. 2, 205–221 (1995, this issue).
8. P. F. Corbett, D. G. Feitelson, J.-P. Prost, G. S. Almasi, S. J. Baylor, A. S. Bolmarcich, Y. Hsu, J. Satran, M. Snir, R. Colao, B. D. Herr, J. Kavaky, T. R. Morgan, and A. Zlotek, "Parallel File Systems for the IBM SP Computers," *IBM Systems Journal* **34**, No. 2, 222–248 (1995, this issue).

Accepted for publication November 8, 1994.

Irving Wladawsky-Berger *IBM Corporation, Route 100, P.O. Box 100, Somers, New York 10589.* Dr. Wladawsky-Berger is currently the General Manager of the IBM POWER Parallel Division. He joined IBM in 1970 as a research staff member at the Thomas J. Watson Research Center. During 15 years at the Center, he served in a variety of key assignments, including that of Vice-President of Systems with overall responsibility for computer science research at IBM. One of his main accomplishments in this position was the establishment of major technology transfer programs, in a number of system areas, between IBM's Research Division and the company's product laboratories. In 1985 Dr. Wladawsky-Berger joined IBM's Large Systems Product Division. He held a number of senior management positions and was responsible for launching IBM's System/370-based supercomputing efforts, as well as for directing the company's System/370 operating systems software business. He was deeply involved in initiating IBM's System/390 microprocessor and parallel technology efforts, and in helping to lead the transition of the company's mainframe business in this new direction. In 1992 he helped organize and launch IBM's RISC-based parallel processing business for scientific and commercial applications, and was named General Manager of IBM's POWERparallel Systems business unit. In recognition of the enterprise's rapid growth, IBM commissioned the business unit a formal division in January 1995, where he continues today. Dr. Wladawsky-Berger holds his master's and doctoral degrees from the University of Chicago. He is a member of the Fermilab Board of Overseers and has served on the Computer Sciences and Telecommunications Board of the National Research Council.

Reprint Order No. G321-5562.