JOEL S. BIRNBAUM
Special Technical Assistant to the Chairman and CEO
HEWLETT-PACKARD COMPANY

Joel S. Birnbaum is Special Technical Assistant to the Chairman and CEO of the Hewlett-Packard Company. In this newly created position, he reports directly to HP CEO and Chairman Carly Fiorina. Birnbaum's role is to continue to help the company shape its technology strategy and to communicate this strategy to the marketplace. He is located in Palo Alto, Calif.

Before assuming his current role, Birnbaum was senior vice president for research and development (R&D) and director of HP Laboratories ?a role he retired from, at the age of 61, in February 1999. At that time, HP Labs, Hewlett-Packard Company's central research and development facility, had 1,300 employees, with headquarters in Palo Alto, Calif., and additional laboratories in Bristol, England; Cambridge, Mass; Tokyo, Japan; and Haifa, Israel. He was also responsible for the coordination of worldwide activities in R&D and served as the company's CTO.

Birnbaum joined HP in 1980 after 15 years at IBM Corp.'s Thomas J. Watson Research Center in Yorktown Heights, N.Y., where he had last served as director of computer sciences. At IBM he supervised the development of the first RISC computer, robotics, speech recognition, Query By Example, voice mail, time-shared data acquisition and control systems, and many other efforts that became products. His first assignment at HP was as the founding director of the Computer Research Center within HP Labs, which conducted research into new directions in computer architecture, hardware and software, as well as some novel applications. One of the technologies developed was the precursor of HP Precision Architecture, the basis for all of HP's RISC computers; among many others were technologies that led to new businesses in network management and software development tools. A world-wide research effort in pervasive computing has been underway since 1983.

In 1984, Birnbaum was named director of HP Labs and an HP vice president. In 1986, Birnbaum became general manager of the Information Technology Group, responsible for the development of all core hardware platforms and systems software for the Precision Architecture product line. After the first successful shipment of these systems in 1988, he was named general manager of the new Information Architecture Group, which developed systems architectures for cooperative-computing environments. These became the basis for HP's open client server system products. In 1991, he was elected senior vice president of R&D and director of HP Laboratories. In this role he was responsible for coordinating HP's global research and development, directing the central research laboratories, and acting as the technology spokesman for the company.

Birnbaum was born in the Bronx, NY. He holds a bachelor's degree in engineering physics from Cornell University and master's and doctoral degrees in nuclear physics from Yale University.

He was elected to the National Academy of Engineering in 1989, and to the American Academy of Arts and Sciences in 2002; he was also awarded a Foreign Membership in the Royal Academy of Engineering of the UK. Birnbaum is a Fellow of the Institute of Electrical and Electronic Engineers, of the Association for Computing Machinery, and of the California Council on Science and Technology, and is also a Sheffield Fellow of Yale University. He has received an honorary doctorate from the Technion University of Israel and was the year 2000 winner of the IEEE Weber Prize, given for career engineering leadership.

Birnbaum's board memberships include the Corporation for National Research Initiatives, the Technion University of Israel, Veridian Corporation, Hermes Softlabs, the Euphrat Museum of Art, the SETI Institute and the Monterey Bay Aquarium Research Institute. Birnbaum has served on computer science or engineering advisory councils at Cornell, Yale, Stanford, Harvard, and Carnegie-Mellon Universities as well as at the University of Southern California, the University of California at Berkeley and at Los Angeles, and the University of Illinois.

Joel S. Birnbaum
Hewlett-Packard

GRAND CHALLENGE 1: PERVASIVE COMPUTING

Once a radical idea, this rubric now defines the strategies of most large computer companies and many smaller specialized ones. It also lies at the heart of the research agenda of many university, government, and industrial research organizations, and has already been much discussed in many of the proposals leading to this gathering. It seems clear that this is the future direction of information systems, and that it will directly affect most aspects of human endeavor. What we need in order to realize its potential over the next decade or so is a clear blueprint of the most important remaining obstacles, and clear identification of areas where an architected common practice and standards need to emerge. There are so many thorny issues that it will take a sustained, open effort to make real progress toward a global information infrastructure.

It has always helped me to think of pervasive technologies as those which are not only ubiquitous, but which have also become so much a part of everyday life for most people that they are more noticeable by their absence than by their presence. I think it helps to break the grand challenge into the two fundamental quests for developing any pervasive technology:

--It must be intuitively accessible by ordinary people, and
--It must be supported by an infrastructure defined by enduring standards, perceived to be valuable enough to justify national and global investment.

The first of these involves the creation of information appliances, natural to use with minimal training because of location, user, and context awareness, and with intuitive interfaces that support a mental model of the specialized task being performed. Of course, we would also like many of the appliances to also be inexpensive and universally deployable. Information appliances need not be portable, nor personal, although many will be both; an important characteristic is that they be a means to an end, enabling users to think in terms of a particular action or solution, and not about how or where it gets done. A second key attribute is that appliances will often need to collaborate, and so must be able to communicate within families across standard interfaces.

This raises many research challenges; most, fortunately, are now being widely pursued: user interface technologies involving natural means of expression, such as voice, gesture, and handwriting; error tolerant devices, requiring a minimum of training; automatically generated, low power, embedded computer technology which can be optimized for performance, cost, and power consumption for a class of tasks; connection to distributed sensors of varied type and great number, providing information about location and context, and enabling application-driven distributed measurements; authentication of user identity through biometrics; cooperating agent technologies; better batteries and other energy sources; software agent technologies;...It is a long list, but I think that the greatest challenge will be to define a collaborative, open working environment so that clear standards for logical and physical interaction among devices and humans will emerge. From this will emerge the solutions to the great opportunities in health care, environment monitoring, personalized manufacturing, education, and the others we have all dreamed of for so long.

The second quest requires that we design and build a global supporting infrastructure that can behave like an information utility: secure, robust, always available, self-configuring and self-repairing. The role of a utility is to transform a capital investment into a competitive service. We should be able to pay for the information utility by usage, as we do for the water, gas, electric, telephone and the other utilities that defined the 20th century, and we shouldn't have to worry about where or how the service is performed.

Again, much work is being done; the dominant issue here is scale, and fundamental rethinking of many of the directions we have traveled in the first 5 decades of computing is needed. Rather

than list the myriad areas requiring serious innovation, most of which have been noted in other submissions to this meeting, I suggest that the most important one is defining the mechanisms and interfaces by which the services provided by the infrastructure can be automatically invoked and linked. Our goal should be application creation by dynamic linking of independently created service modules across heterogeneous networks; while variants of this have been demonstrated, none yet exists at the scale and with the robustness, availability, performance and maintainability required. A coordinated effort is needed to produce a lingua franca that allows the many players in the world of integrated e-services to develop their offerings with the assurance that individual components will interoperate smoothly and efficiently without requiring prior knowledge of the pairings and hierarchies.

This meeting could define the criteria and develop the roadmap for both of these critical steps in making pervasive computing a reality in this decade.

GRAND CHALLENGE 2: COMPUTER ARCHITECTURE FOR MOLECULAR ELECTRONICS

The end of Moore's Law scaling for semiconductors is a certainty at the device level due to the vanishing electrons at the switching gates. Today, thousands of electrons impinge on a gate to affect a transition; as we shrink the critical dimensions to achieve speed and density, that number will decrease by three orders of magnitude by about 2015 and the resulting statistical uncertainties will first make the devices unreliable, and then, when we reach single electron levels, unworkable. This is a matter of fundamental physics, and not engineering obstacles. Clearly, we need a new type of switching element and a cost-effective way to manufacture it, and since it is likely to be very different than the transistor it displaces we will need to rethink the architecture that defines its use.

The history of technology contains many lessons about scientific and commercial behavior during disruptive change, and one of them is that manufacturers are so loath to give up on large in-place investments that the current technology "stretches" to meet the challenge of the disruptive one, often extending its life for years and sometimes even decades. This behavior is much in evidence in the future demise of semiconductors, and we will observe enormous cleverness and invention in the years ahead to keep semiconductors competitive.

Nevertheless, many labs today are trying to replace the transistor with a smaller, faster, less expensive and far more power efficient switch depending on molecular, electronic or chemical behavior, and connected by nano-wires of varied designs. This is surely one of the grand "hardware" challenges of the new century, but there are critical systems issues beyond the device technology itself that must also be examined. In the end, these will drive the architecture of the post-Moore's Law computers of the next decade, regardless of the specific workings of the molecular switch.

My view is that the two key research questions are: How can we create a new technology which can survive the huge number of defects that will inexorably be present at the quantum scale? and, equally important, how can we overcome the economic tyranny of Moore's Second Law and manufacture the new devices at an acceptable cost?

If we start with the assumption that a large percentage of the devices will have flaws (the best biological processes make an error of one part in a couple of billion or so; man-made manufacturing processes have not yet achieved this level of accuracy, and are unlikely to), then molecular level systems will unavoidably have millions or billions of defective parts. About a decade ago, we built a machine at HP Labs that we called TeraMac, a multi-architecture computer that operated one million gates at one million cycles per second, or 10^{12} operations per second. The machine was originally designed as an emulator that could run software at useful speed for evolving architectures, but because of the prohibitive cost of implementing the FPGA-based crossbar arrays, it was decided to use many defective parts that were cheap or free, and to

program around the defects. The machine was a big success, and has achieved for a wide variety of problems at least ten times workstation performance in comparable technology in spite of having some 300,000 known defects. This is achieved by first building the machine, then experimentally detecting the defective parts, then labeling them busy and unavailable; a sophisticated compiler then creates optimized programs using this information for all subsequent uses of the machine. The Teramac was truly a junk yard computer, but it embodies the essential architectural lessons to build molecular and quantum computers. We are confident that if we design the computer first, we will be able to cope with the issue of defect tolerance regardless of which switch technology eventually prevails, but there are hard questions to answer as we scale the machine to molecular capacities.

At HP, we are pursuing both molecular and quantum computing, but we think that the time scale for the latter is at least 3 to 5 times further out for practical application, with the range of applicability also far more limited, unless there are algorithm and software breakthroughs. We are very encouraged about the practical creation of hybrid machines containing processor cores and memories built through self-assembly at molecular dimensions coexisting with semiconductors performing less repetitive tasks, providing input/output, and housekeeping functions. Today's silicon chip can be thought of as the printed circuit board for a molecular processing unit or storage, providing mechanical and electrical support, and being the conduit for signals to and from the real world. We are now working at the circuit level, and hope to be able to demonstrate working subsystems in the next 5 years or less.

It is time that architectural studies were begun in earnest, for what is already abundantly clear is that the machines cannot be wired the way today's chips are. They must be regular to permit self-assembly, and to contain the complexity at this scale. The heat dissipation at the projected speeds and densities does not encourage conventional architectures, nor will the types of problems that processors and storage of this capacity will enable. Beyond the hardware, there are fascinating problems in how application software will be written, how I/O can be specified, how signals will propagate, as well as new algorithmic approaches to search, parallelism, security, and many others.