

The Future of the Microprocessor Business

MICHAEL J. BASS
Hewlett-Packard Co.
&
CLAYTON M. CHRISTENSEN
Harvard Business School

Customization and speed-to-market will drive the industry from the bottom up

In a century in which technology left few aspects of life unchanged in some countries, the microprocessor may have been the most transformative of all. In three decades it has worked itself into our lives with a scope and depth that would have been impossible to imagine during its early development.

If you live in a developed country, chances are good that your household can boast of more than a hundred microprocessors scattered throughout its vehicles, appliances, entertainment systems, cameras, wireless devices, personal digital assistants, and toys. Your car alone probably has at least 40 or 50 microprocessors. And it is a good bet that your livelihood, and perhaps your leisure pursuits, require you to frequently use a PC, a product that owes as much to the microprocessor as the automobile owes to the internal combustion engine.

Throughout most of its history, the microprocessor business has followed a consistent pattern. Companies such as Intel, Motorola, Advanced Micro Devices, IBM, Sun Microsystems, and Hewlett-Packard spend billions of dollars each year and compete intensely to produce the most powerful processors, which handle data in 32- or 64-bit increments. The astounding complexity and densities of transistors on these ICs—now surpassing 200 million transistors on a 1-cm² die—confer great technical prestige on these companies. The chips are used in PCs, workstations, and other systems that, for the most part, have been lucrative, high-volume markets.

As with other ICs, microprocessors have for the past few decades been undergoing the exponential rise in perform-

ance prophesied by Moore's Law. Named for Intel Corp.'s cofounder, Gordon E. Moore, it describes how engineers every 18 months or so have managed to double the number of transistors in cutting-edge ICs without correspondingly increasing the cost of the chips. For microprocessors, this periodic doubling translates into a roughly 100 percent increase in performance, every year and a half, at no additional cost. The situation has delighted consumers and product designers, and has been the main reason why the microprocessor has been one of the greatest technologies of our time.

In coming years, however, this seemingly unshakable industry paradigm will change fundamentally. What will happen is that the performance of middle- and lower-range microprocessors will increasingly be sufficient for growing—and lucrative—categories of applications. Thus microprocessor makers that concentrate single-mindedly on keeping up with Moore's Law will risk losing market share in these fast-growing segments of their markets. In fact, we believe that some of these companies will be overtaken by firms that have optimized their design and manufacturing processes around other capabilities, notably the quick creation and delivery of customized chips to their customers.

The changes portend serious upheaval for microprocessor design, fabrication, and equipment-manufacturing firms, which have been laser-locked on Moore's Law. Executives lose sleep over whether they can keep on shrinking line widths and transistors and fabricating larger wafers. We don't blame them, given their history. Nor do we see blissfully peaceful slumber

in their near future: this is not another article forecasting the imminent demise of Moore's Law.

On the contrary, we believe that the top IC fabricators will have little choice but to invest ever more heavily so as to keep on the Moore trajectory, which we expect to go on for another 15 years, at least. We don't see these investments as sufficient for future success, however.

Will semiconductors hit a physical limit? They surely will, someday. But this is probably the right answer to the wrong question. The more important question is: as technological progress surpasses what users can use, how do the dynamics of competition begin to change?

Bottom to top

The stakes are high. The microprocessor market, which totals about US \$40 billion a year, has several main tiers. At the top are the most powerful chips, which are used in servers and workstations. Then there's the PC market, dominated by Intel microprocessors. These relatively high-end chips were a major component of a category that rang up US \$23 billion in 2001, after peaking at \$32 billion the year before, according to the Semiconductor Industry Association (San Jose, Calif.), a trade group.

Microcontrollers were another important category, with sales totaling \$10 billion in 2001. They are generally less computationally powerful than high-end microprocessors, and exert real-time control over other systems, such as automobile

engines. Finally, digital signal-processing chips, a key component of cell phones, DVD players, and other entertainment products, had sales of \$4 billion last year.

With their exponentially increasing performance, microprocessors might seem unique and unlikely to follow the broad evolutionary pattern that has played out in the past in most other technology-based industries. After all, the Moore's Law phenomenon is unprecedented in industrial history. But strong evidence shows that the same evolutionary pattern that occurred in mainframe computers, personal computers, telecommunications, banking, hospitals, and steel production is indeed occurring in the microprocessor business.

The pattern begins with a stage in which available products do not yet perform up to most customers' needs. So, not surprisingly, companies compete by making better products. In other words, competition during this stage is basically driven by performance. As engineers design each successive generation of product, they strive to fit the pieces together in ever more efficient ways to wring the most performance possible out of the technology available.

Hundreds of Athlon microprocessors, from Advanced Micro Devices, are fabricated on 200-mm wafers at a new plant in Dresden, Germany [below]. Though designed for PCs, the 1-GHz chips perform beyond the levels most PC users need.



Typically, major subsystems need to be interdependently designed—and, as a result, a competitor needs to make all the product’s critical subsystems. During this phase, the competitive advantage of vertical integration is substantial, so manufacturers do almost everything themselves. This is the way it was in the earliest days of the PC industry, for example.

The next stage of an industry’s development begins when the performance of its products has overshot the needs of customers in the less-demanding tiers of the market. These customers won’t pay premium prices for more performance, of course, but will pay extra for a product that is extraordinarily reliable, or one that has been customized to meet their specific needs—especially if they can get that ultra-reliable or customized product quickly and conveniently. Ease of use is another feature that customers typically reward with a premium.

To compete on the dimensions of customization, quick delivery, and convenience, product architectures whose pieces are strongly technologically interdependent tend to give way to modular ones, in which the interfaces among subsystems are standardized. This modularization lets designers and assemblers respond quickly to changing customer requirements by mixing and matching components. It also lets them upgrade certain subsystems without redesigning everything.

But perhaps the most important repercussion of modularization is that it usually spurs the establishment of a cadre of focused, independent companies that thrive by making only one component or subsystem of the product. Think Seagate Technology in hard-disk drives, or Delphi Delco in automotive electrical systems.

One of us (Christensen) has studied how industries that are in transition between the two stages present peril for established firms—and opportunity for upstarts. Large firms can easily become blind to shifts in the types of performance that are valued. For an established company, with its well-defined competencies and business models, the obvious opportunities for innovation are those that focus on unsatisfied customers in the higher market tiers. After all, that is where the innovations that the company has structured itself to deliver are still being rewarded by premium prices.

Inevitably, though, this high-tier, performance-hungry group of customers shrinks as performance gets better and better. At the same time, in lower market tiers there is a huge increase in customers who are willing to back off from the leading edge of performance in exchange for high reliability, customization, ease of use, or some combination of all three.

Too much of a good thing?

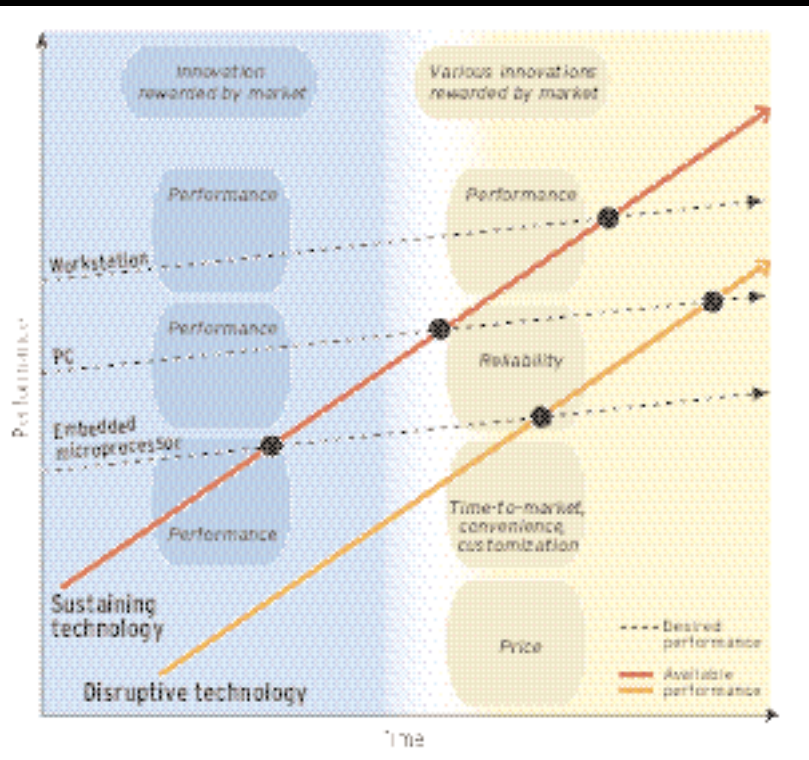
This is precisely the juncture at which the microprocessor market has now arrived. Price and performance, fueled by the industry’s collective preoccupation with Moore’s Law, are still the metrics valued in essentially all tiers of the market today. Even so, there are signs that a seismic shift is occurring. The initial, performance-dominated phase is giving way to a new era in which other factors, such as customization, matter more.

Perhaps the best evidence that this shift is under way is the fact that leading-edge microprocessors now deliver more performance than most users need. True, emerging applications like three-dimensional games, the editing of digital-video files, and speech-to-text tax the fastest available microprocessors. However, few people who regularly run such applications do so on a single-microprocessor PC. And that fact is unlikely to change. In the future, as now, many of these taxing applications will run on special-purpose or separate processors.

In any case, the users who run these applications regularly are few compared to the masses who use their PCs mainly for word processing, scheduling, e-mail, and Internet access. For this majority, high-end microprocessors—Intel’s Itanium and Pentium-4, and Advanced Micro Device’s Athlon—are clearly overkill. Running common benchmark programs, these chips

The Impact of Disruptive Technology

As the performance of conventional microprocessors improves [red line], they first meet and then exceed the requirements of most computing applications [dotted lines]. Initially, performance is key [blue area]. But eventually, other factors—reliability, time-to-market, convenience, and customization—become more important to the customer [green area]. And when a disruptive technology [orange line], such as customizable processor cores, is introduced, it inevitably takes over market share from performance leaders.





Microprocessors hundreds of times as powerful as today's should emerge from chip-making equipment using extreme ultraviolet radiation. Sandia National Laboratories unveiled such equipment a year ago [above].

can perform more than one billion floating-point operations per second (flops), and in some cases, more than 2 gigaflops. Yet, Microsoft's Windows XP, the most recent version of the ubiquitous operating system, runs fine on a Pentium III microprocessor, which is roughly half as fast as the Pentium-4.

What will the microprocessor business be like after this shift? Consider how the PC business grew from a cottage industry into a global colossus over the past 25 years.

In the early days, the 1970s, vertically integrated companies such as Apple Computer, Tandy, Texas Instruments, Commodore, and Kaypro built their computers around proprietary architectures and generally wrote their own software. Then, in 1981, IBM shook up the industry with its original PC, which had a modular architecture and subsystems built by such suppliers as Intel, Microsoft, and Seagate.

Early on, Apple's products were by consensus the best-performing and most reliable in the industry. But in time, as microprocessors, software, and other key components improved, garden-variety PCs became good enough for mainstream applications like word processing and spreadsheets. Competitive advantage shifted to the nonintegrated companies whose products made use of IBM's modular architecture. These were the clone makers—Compaq, Packard Bell, Toshiba, AT&T, and countless others. Not long after that,

dependability became the central axis of competition, and a few firms with reputations for reliability—IBM, Compaq, and Hewlett-Packard—managed to command price premiums.

By the early to mid-1990s, functionality and reliability had become more than good enough from just about everyone, and the way that computer makers needed to compete shifted again. This change set the stage for Dell Computer Corp. (Round Rock, Texas) to rocket from the lowest tiers of the market to industry dominance. Dell let consumers in their homes or offices specify a set of features and functions for their computer to meet their particular needs. Dell then delivered that computer to their door in a few days. In effect, by coming up with a business model that emphasized customization, speed of delivery, and convenience, Dell pushed the industry into its next stage of development.

Although our example is PCs, other technology-based industries have evolved similarly, from automobiles to mainframes. In every case, the primary dimension of competition migrated from an initial focus on performance to reliability, convenience, and customization. When performance began exceeding what customers in a tier of the market could use, competition redefined the types of improvement for which those customers would pay extra. Further, the types of features that let suppliers demand premium prices shifted predictably

from performance to reliability, convenience, and so on. As this happened, moreover, competitive advantage moved from companies that were highly integrated to ones that were not.

Why don't microprocessor makers simply start producing lower-performance chips? A few years ago, Intel Corp. (Santa Clara, Calif.) began doing just that with its Celeron microprocessor. The problem is that Celeron is a one-size-fits-all proposition. Its architecture isn't more modular than that of the Pentium products it is displacing, and it cannot be customized nearly as much as emerging alternatives.

Interestingly, while the latest microprocessors offer higher processing rates than most users need, semiconductor fabrication facilities now offer circuit design teams more transistors than they need. Put another way, the rate at which engineers are capable of using transistors in new chip designs lags behind the rate at which manufacturing processes are making transistors available for use.

This so-called design gap has been widening for some time. In fact, the National Technology Roadmap for Semiconductors noted it five years ago, observing that while the number of transistors that could be put on a die was increasing at a rate of about 60 percent a year, the number of transistors that circuit designers could design into new interdependent circuits was going up at only 20 percent a year.

The fact that microprocessor designers are now "wasting" transistors is one indication that the industry is about to reenact what happened in other technology-based industries, namely, the rise of customization. Keep in mind that in order to develop a modular product architecture with standardized interfaces among subsystems, it is necessary to waste some of the functionality that is theoretically possible. Modular designs by definition force performance compromises and a backing away from the bleeding edge.

Core customization

A form of customization has already taken hold in the lower tiers of the microprocessor industry. System-on-a-chip (SoC) products are modular designs constructed from reusable intel-

lectual property (IP) blocks that perform specific functions [see "Crossroads for Mixed-Signal Chips," *IEEE Spectrum*, March 2002, pp. 38-43]. IP blocks vary in size and complexity, ranging from simple functions such as an RS-232 serial port interface or a DRAM memory controller, to a complex subsystem, such as an entire 32- or 64-bit microprocessor. These IP blocks can be used within multiple designs within a company, or used in designs at different firms.

In the lower market tiers, several firms are bundling IP blocks into both soft cores (software-like descriptions of the IP blocks that can be synthesized into hardware designs) and hard cores, that is, pre-verified hardware designs. Such cores range from hundreds of thousands to a few million transistors, and their availability in the marketplace enables firms to focus

Engineers can specify a microprocessor and in some cases completely design one in weeks, rather than months

less on new design and more on system integration. They can select and integrate microprocessor and other types of cores into SoC designs that are then manufactured as special-purpose components for a specific product.

Recently, a few companies have been pushing this trend toward component selection and integration even further, into microprocessor cores themselves. Using special design tools, engineers can specify such a microprocessor, and in some cases completely design one, in weeks rather than months.

Leading companies in this so-called customizable core movement include Tensilica, ARC Cores, Hewlett-Packard, and STMicroelectronics. They have a similar philosophy, but target different markets and application needs. The Tensilica Xtensa processor, for instance, offers customization within the framework of a simple microprocessor core. Customers can specify their own instruction set extension by accessing a Web site and using a high-level language, such as Verilog.

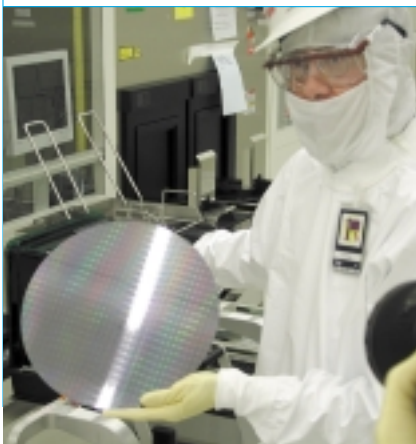
ARC Cores' ARCTangent family targets the digital signal processor markets. Like Tensilica, it allows users to customize

Custom Chips and Future Fabs

As custom-configured chips carve out a larger portion of the microprocessor market, major changes are in store for the way chip fabrication facilities operate. To understand why, consider the underlying trends. One is the ever-increasing abundance of transistors on a chip, which portends a day when the number of chips per product will approach one. Also every chip will, to some degree, be custom-

tailored to its application, and product life cycles will be short. Put those trends together and the inescapable conclusion is that, in the foreseeable future, chips will have to be made in a few days, not weeks, as they are now.

To achieve that kind of speed, chip fab plants must operate differently, starting with the way wafers flow through them. The dies are fabricated in batches on round semicon-



both processor features (bus widths, cache sizes, and so on) and instructions. Hewlett-Packard's HP/ST Lx family is aimed at scalable multimedia acceleration using Very Long Instruction Word (VLIW) techniques. It lets customers choose the amount of instruction-level parallelism—in other words, how many functional units to include, and how many operations can be performed in parallel.

Intel's dilemma

So what will happen in the microprocessor market as a whole? It will be a repeat of what has already happened in other technical industries. The trends of customization and speed-to-market will continue to take hold in the lower tiers—in digital signal processing (DSP) and in processors that are embedded within such products as MP3 players, digital cameras, and set-top boxes. Increasingly, sales of stand-alone digital signal-processing and embedded chips will give way to SoCs that incorporate DSP or other functions.

Gradually, over a period of years, the trend will creep upward into higher tiers of the market, including PCs [see graph on p. 36]. In fact, companies such as SuperH, MIPS Technologies, and ARM already produce reduced-instruction-set microprocessor cores that can be combined on a single die with other functional units, and that are easily powerful enough to serve as a PC's central processing unit.

For high-end companies such as Intel, the dilemma will be that their best and most profitable customers will continue to need exactly the sort of general-purpose, leading-edge processors that Intel is so good at designing and delivering. Nevertheless, broad trends in electronics suggest that growth will increasingly stem from applications in the lower tiers.

For example, sales of PCs, the largest market for high-end microprocessors, declined in 2001 for the first time since 1985, according to market research firm Gartner Dataquest (Stamford, Conn.). Meanwhile, although the market for DSPs shrank even more than the rest of the microprocessor category last year, on the whole they have been one of the fastest-growing segments of that category. The growth has been fueled by such up-and-coming sectors as wireless communications, handhelds, and game, music, video, and other entertainment systems.

Implications, implications

If the microprocessor market does become dominated by multitudes of targeted chips produced in relatively small numbers, several intriguing ramifications could develop. First, for chip-makers, time to market will matter much more [see "Custom Chips and Future Fabs," below]. Second, as firms target smaller, more specific markets, they will differentiate their products more by discovering specific needs that general-purpose products do not address.

This discovery process works better when products are introduced faster and more frequently, in response to feedback from customers. Tensilica, for one, now boasts that new microprocessor cores used within a system-on-a-chip design can be created and their performance tested through simulation in two weeks. The manufacturing latency of most chip fabrication facilities, on the other hand, remains as high as 10 weeks.

With product life cycles approaching a year or less, some microprocessors currently spend a good part of their lives being manufactured. In the future this situation will be competitively intolerable.

In coming years, success in the microprocessor business will increasingly demand that companies:

- Adeptly use modular designs that reuse and recombine silicon intellectual property.
 - Include multiple circuit types and possibly process technologies on a single die, creating custom-tailored systems-on-a-chip.
 - Shrink the design cycle for microprocessors and systems-on-a-chip and dramatically reduce manufacturing latency times (time spent in the plant).
 - Tolerate, both technically and economically, a manufacturing mix composed of a multitude of low-volume runs of narrowly targeted products with short market windows.
- The evolution toward this future will not be driven by or grounded in the choices of managers in today's industry-leading companies. Competition in the relevant tiers of the market will force these new trajectories of improvement to become critical. The only question is which companies will have developed the capabilities and organizational structures required to thrive in these markets. ●

Glenn Zorpette, *Editor*



ductor wafers about the size of a dinner plate [photo, left].

In today's fabs, those wafers are also produced in batches. Therein lies the problem: as these groups of wafers move from station to station in the fab plant, they have to wait until an entire batch is complete before going on to the next station. At any given time, large amounts of valuable product are simply sitting around the plant.

In the future, wafers will need to move rapidly, one by one, through these plants. The difficulty will be finding a way to imple-

ment this single-wafer processing in a variety of fabs, all of which have their own proprietary architecture. To our knowledge, no firm as yet has figured out how to do this, though Taiwan Semiconductor Manufacturing Co. (Hsinchu) and United Microelectronics Corp. (Taipei) have made significant progress in this direction.

Single-wafer processing will also probably compromise the economic viability of investments in 300-mm wafer-processing equipment. The reason is that, in the past, when a factory converted to a low-inventory

process flow (such as single-wafer processing), the effective capacity of that plant typically doubled.

What this means for fab plants is that it makes more economic sense to convert a 200-mm-wafer fab plant to single-wafer processing, rather than build a 300-mm-wafer fab plant. The 200-mm, single-wafer-processing fab plant could be expected to process twice the number of wafers per month and therefore as many die as the 300-mm fab—and to do so at much lower levels of cost and complexity.

—M.J.B. & C.M.C.