

The supercomputer in your pocket

Over the past decade, the computing landscape has shifted from beige boxes under desks to a mix of laptops, smartphones, tablets, and hybrid devices. This explosion of mobile CPUs is a profound shift for the semiconductor industry and will have a dramatic impact on its competitive intensity and structure.

Harald Bauer, Yetloong Goh, Joseph Park, Sebastian Schink, and Christopher Thomas The shift toward mobile computing, at the expense of tethered CPUs, is a major change that has raised the competitive metabolism of the semiconductor sector. Mobile, now the central battleground of the technology industry, is having an intense impact on the larger semiconductor landscape. Mobile-computing processor requirements now drive the industry, setting design requirements for transistor structures, process generations, and CPU architecture. It's the must-win arena for all the semiconductor sector's top companies. Over the next half decade, leading players will spend hundreds of billions of dollars on R&D.

Two major contests will play out for semiconductor vendors competing in the mobile arena: the clash between vertical-integration and horizontal-specialization business models, and the clash between the two dominant architectures and ecosystems, ARM and x86. Each of these battles will be explored in detail below.

More and more smartphones are as capable as the computers of yesteryear. Laptops have displaced desktops as the most popular form factor for PCs, and, thanks to the success of Apple's iPad, tablets have stormed into the



marketplace. PC original equipment manufacturers (OEMs) aren't waiting to lose consumer share of wallet to tablets. Instead, they are generating a wide range of new form factors. For example, hybrid computers—those with attributes of laptops and tablets in one device—are on store shelves across Asia. Combination machines are coming to market. Intel is pushing "ultrabooks," which combine aspects of both tablets and traditional PCs.

Over the next five years, we expect mobile phones, tablets, portable computers, and new hybrid devices to dramatically exceed overall industry growth. Adjacent technologies that feed these mobile CPUs are growing fast, too. Cisco reports that global 3G network rollouts have helped increase mobile traffic by a factor of 2.6 from 2010 to 2011, the third straight year of such rapid growth rates. Changes in the types of data transmitted over these networks, as well as an increase in wireless speeds, will drive additional demand for mobile processing. In 2011, mobilevideo traffic accounted for more than 50 percent of all data traffic. The use of video, which is more processing intensive than static text or pictures, dramatically increases the level of sophistication required in mobile-device CPUs. What's next? Operators are now launching 4G technologies; Cisco says that the average early adopter of a 4G phone downloads 28 times more data than a 3G phone user.

Based on analysis of current trends, we expect fivefold growth in smartphone unit sales by the beginning of 2016. Tablets, another booming category, are expected to grow threefold, and the connected PC segments should see unit sales double in that time frame. If Moore's law, which assumes a doubling of processing power every 18 months, holds the same path for the next few years as it did for the last 40, the global, mobile CPU processing power of the installed base could grow 40- to 60-fold between 2011 and 2016 (Exhibit 1).

This massive growth provides a compelling opportunity for the semiconductor industry—but also leaves it facing two inflection points. Executives should take note because these types of transitions have again and again upset the competitive order, leading to new winners and losers.

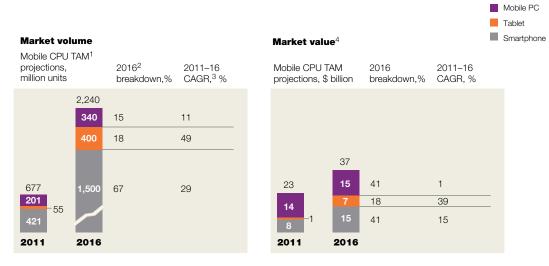
The first of these transitions is the serious challenge that Apple, Samsung, and others pose to the merchant silicon business model. An industry that has long accepted a horizontal business model is now revisiting vertical integration through the internal chip development occurring at some of its largest players.

At the same time, the two formerly separate worlds of the ARM and x86 architectures have grown into direct competition. Previously, these two CPU ecosystems largely moved in parallel, holding dominant positions in separate product lines with different customers. Further complexity is added as mobility changes the rules of traditional CPU technology competition. In small devices with limited battery density, mobile CPUs cannot fully benefit from the performance improvements of Moore's law; the power drain would be prohibitive.

Given this state of affairs, there are two issues that each semiconductor company should consider. The first is the competition between

Exhibit 1

The growth rates for mobile devices with advanced CPU processing requirements are strong.



¹Total addressable market.

Source: ARM; DB Market Research; IC Insights; iSuppli; McKinsey analysis

two different models for industry structure, vertical or horizontal. Next, there is the battle between competing technical architectures, which pits x86 against ARM. The resolution to each of these debates will go a long way toward separating the winners from the losers in the years ahead.

Competing industry structures

When smartphones first appeared in the market, each handset vendor had its own software stack and distinct sets of services, and players looked to differentiate themselves through hardware features. Different OEMs created distinct products, and proprietary technology created a barrier to entry for low-cost players.

Then Android emerged. As an open operating system, Android leveled the playing field, introducing both a standardized user experience and hardware specification. This allowed new low-cost players such as Chinese handset makers to develop products with a look, feel, and capabilities similar to those of the established players. The rules of the phone game changed-making it much more similar to the PC arena, where Intel and Microsoft drive the technology cadence and OEM differentiation is limited. The popularity and mass adoption of Android has required smartphone vendors to find new sources of differentiation. If they can't crack the code, these OEMs may be perceived as little more than product assemblers, with corresponding low margins.

²Estimated.

³Compound annual growth rate.

⁴Includes value for application processor/CPU; wireless baseband (discrete or integrated) is excluded.

Apple and Samsung, who collectively held over 50 percent share of the smartphone and tablet markets in 2012, have been winning with a different approach. They used their internal control of key CPU technology to provide differentiated customer experience to their device users. Apple assembled a silicon design team and began designing ARM-based custom CPUs for smartphones, with manufacturing performed by external foundries. Samsung, the world's number-two semiconductor vendor, expanded its design and manufacturing capabilities to smartphone and tablet CPUs. In today's system-on-a-chip (SOC) world, these two OEMs realized that the CPU is the system and thus they needed their own CPUs to truly differentiate their products.

This approach also gives the companies greater negotiating leverage with the leading merchant silicon vendors such as Intel, NVIDIA, Broadcom, MediaTek, and Qualcomm. Following this lead, others are increasingly looking to bring chip design back in-house. Both Microsoft and Google launched branded tablets in the second half of 2012. While both are powered by merchant silicon, these "software" companies are taking strict control of hardware specification and directly engaging CPU vendors for specific features.

This is a new and significant challenge to traditional merchant silicon vendors. First, it reduces their revenue opportunity. Second, and more important, it removes them from their traditional position of defining and driving the leading edge of product design; leading OEMs are pushing the envelope just as fast as the CPU vendors are.

Despite this shift toward internal design, the merchant markets are still robustly competitive. They maintain the greatest share of the chip market and the greatest concentration of technology capability. Merchant players are still providing complementary technologies beyond the CPU, such as baseband chips or analog silicon, that appear in top products from Samsung and Apple. In addition, merchant chip vendors still hold traditional advantages. First, their higher collective volume gives them scale advantages. They can develop technologies that extend into product lines and technologies far more diverse than any one OEM could support. They can push their technology cadence as fast as possible and provide that technology to multiple competing OEMs that iterate it and deploy it collectively across many devices. Vertical players, on the other hand, have stand-alone OEMs that must tackle all the technical and market hurdles individually. Merchant silicon vendors can also focus solely on making great silicon, rather than designing and manufacturing a full consumer end product. Finally, many mobile devices require CPUs to be integrated with other components such as digital-baseband chips—technologies that OEMs such as Apple, Samsung, ZTE, and others do not currently possess.

OEMs that attempt to control all aspects of their silicon may swim against the tide of history. All high-growth computing markets with rapid technology cadences have eventually adopted the horizontal silicon model; it has simply been too hard for integrated players to keep up. Apple, Samsung, and others will have expensive and technologically difficult challenges in aiming to win consistently against the merchant silicon players over the longer term.



Competing architectures

All the growth in mobile computing over the last two decades has been driven by the x86 and ARM architectures. For more than 30 years, the x86 microprocessor instruction set architecture (ISA) has been the technical foundation of the personal-computing industry, and for the last 15 years, it has powered the Internet and server ecosystem. Intel, with about 80 percent market share of x86 CPU shipments, and Advanced Micro Devices (AMD) have been the drivers behind this architecture.

Recently, the ARM ISA, developed by ARM Holdings, has grown to match and, in some ways, exceed the scale and scope of the x86-based computing industry. Using a very different business model, the ARM ecosystem has shipped more than 15 billion ARM-based CPU and microcontroller chips in the last five years, and its sales are growing at a 25 percent annual rate. Traditionally, these two ISAs have not competed against each other directly as they served different end markets: x86 targeted personal computers, servers, and high-end embedded-computing applications, while ARM offered power-efficient chips for mobile phones, tablets, and microcontrollers.

Could ARM, a midsize company headquartered in the United Kingdom with fewer than 2,000 employees and a bit more than \$700 million in annual revenue, actually thwart the ambitions of the company that has been the number-one semiconductor manufacturer for over 20 years? Or will the x86 architecture, the mainstay of personal computing since the Intel 8086 processor launched in 1978, conquer one more end market?

Converging road maps

This architectural battle has been the subject of speculation and controversy for years in the semiconductor ecosystem. Yet 2012 may be a tipping point, the year a theoretical discussion becomes a real one (Exhibit 2).

In the past, economies of scale and learning effects—the fact that semiconductor design and manufacturing knowledge accumulates through previous generations and becomes a core asset required to make the next generation of designs—have tended to create winner-take-all dynamics for hardware architectures. In each market, one architecture has won, by far, the largest market and profit share.

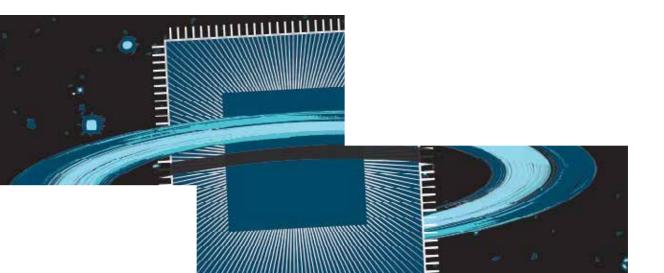
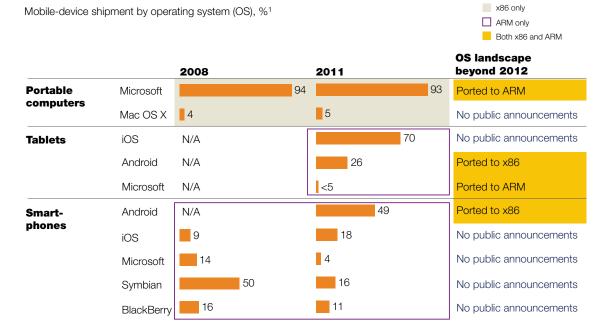


Exhibit 2 In 2012 and beyond, the leading PC and smartphone operating systems will work on both x86 and ARM architectures.



¹Figures do not sum to 100% because "other" category has been excluded. Source: Gartner; McKinsey analysis

Clashing ecosystems

We believe the robustness and success of each architecture's ecosystem—the OEMs, original design manufacturers, and software vendors that build the device and services around a chip—will determine the outcome, rather than technical superiority. The robustness of the ecosystem hinges largely on the different business models of the two architectures.

There are three key criteria to assess the success of each business model. First, which architecture's technical ecosystem will have the greatest amount of engineering resources to drive the technology forward? Second, which will attract the most capital to fund increasingly expensive

and difficult technology development? Third, which will be most successful in encouraging OEMs and software vendors to build innovative devices with the architecture?

It is ironic that in the current competition, ARM is now playing the role that the x86 architecture played in previous battles with proprietary mainframe and server CPUs. The x86 architecture built a foundation for standard hardware platforms. Many different OEMs and software providers then built industry-standard products around that ecosystem. This model has been successful, for the most part, displacing any vertically integrated system that challenged it.

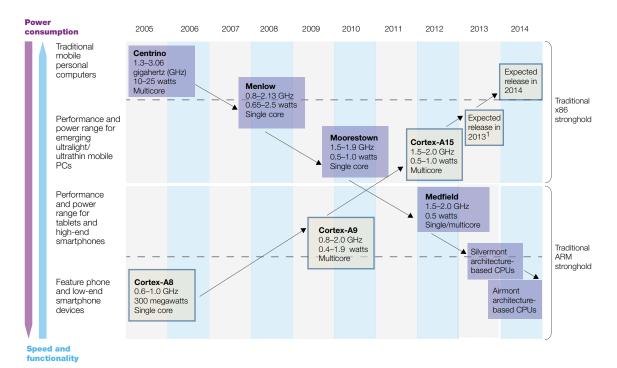
However, as system architectures move from ones with discrete CPUs to SOCs, the CPU itself becomes the system (Exhibit 3). In a SOC model where a CPU core is surrounded by integrated peripherals such as switch fabrics, graphics processing, embedded flash memory, and multimedia processing, all on the same silicon die, the ARM CPU takes the role of a standard platform. Multiple chip-set vendors build SOCs on top of ARM cores, integrating their own and other vendors' intellectual-property (IP) "blocks" to make final products. Multiple wafer-fabrication facilities can then produce the final chip.

In x86, Intel and AMD are dependent primarily on their own engineers to develop these additional IP blocks. They must also integrate them into SOCs and develop the process technology to manufacture the chips containing those IP blocks. And they must depend solely on their own sales volume in mobile CPUs to fund process-technology development and manufacturing capacity.

Clashing business models

The cost to develop the most advanced microprocessors has risen dramatically—from as much

Exhibit 3 Intel and ARM are moving into each other's strongholds.



¹ARM Holdings has not publicly announced its 2013 and 2014 product road map; Intel has announced it 2012–14 Atom product line microarchitecture road map, but no detailed product specifications.

Source: "A guide to mobile processors," The Linley Group, August 2012; Gartner ARM investor reports; product specifications

as \$4 billion to build one scale-capacity manufacturing facility with associated process development at 65 nanometers (2008–09 technology) to as much as \$10 billion at 22 nanometers, the current leading edge. This figure will continue to grow through future technology generations. The increasing performance requirements of mobile devices will continue to demand the latest semiconductor technology. The technical and funding challenge is so large that, in 2012, the largest semiconductor players, including Intel and TSMC, began sharing some of the load via joint investments in the leading lithography tool vendor, ASML.

Despite a few instances of some joint cost sharing, under the current x86 business model, Intel and, to a lesser extent, AMD are required to fund the majority of these technologies internally. These two companies will be required to develop the process technology, build the factories, improve the microprocessor microarchitecture, develop complementary silicon IP to support the core CPU function, and figure out how to integrate all these elements. Intel certainly has the financial and technical capability to do this.

However, ARM will use a different model to approach this investment requirement. There are 15 ARM foundry licensees, and each can work to develop the best process technology to manufacture ARM-based CPUs. There are more than 275 ARM core licensees, and the design team for each can use diverse methods to solve technical problems. These licensees collectively have between 60,000 and 100,000 engineers driving forward ARM semiconductor technology,

which is most likely more than the total number of engineers working at AMD and Intel directly on x86 technologies.

ARM licenses its technology using two primary mechanisms: the first is a set of microprocessor core/ISA licenses that enables chip designers to build application processors. The second is a physical IP license that enables third-party foundries to manufacture these cores and related IP blocks. The broad number of licensees enables ARM to be customized and sold into many different markets, whether large or niche, and it ensures vibrant competition among ARM chip vendors. That competition also enlarges the technology road map, as multiple design teams across the ARM ecosystem will try different implementations to solve end-customer problems.

License flexibility allows fabless semiconductor players to customize their own business model. Marvell and Qualcomm, for example, invested in architectural licenses by purchasing the ARM v7 ISA. This license has in turn enabled them to invest in custom ARM CPU cores. In one case, Qualcomm's Snapdragon product uses a customized and proprietary Krait CPU core, which the company claims has better performance than standard ARM Cortex-A8 CPUs. Other mobile CPU vendors make circuit-level modifications to the ARM standard core to improve performance without having to build a proprietary stand-alone core.

Each licensee's business-development and sales organization can experiment with a variety of business models and search for new customers. At the same time, the open model drives

competition, advancing ARM technology and creating challenging dynamics for CPU vendors. There are often five to ten highly capable ARM-based CPU vendors competing for the same "socket" in a new OEM phone design.

ARM-based CPU designers have a wider range of CPU core IP from which to choose because ARM does not carry any manufacturing overhead when offering its legacy cores. ARM offers the higher-performance, higher-power dissipation A15 core as well as the much lower-power, smaller A7 companion core specifically for phones. (It should be noted that even older, generalpurpose, lower-power cores can be used in these designs.) As a result, chip designers have a range of options for partitioning computing tasks. For instance, in an integrated CPU baseband chip, a legacy, low-power core can manage telephony (making and receiving calls), while a new, higher-performance core handles Internet access and video.

Intel's model may have less diversity, but the company believes this simplicity and focus on industry-standard platforms gives it unmatched agility and speed in technical development. Intel's chairman, Andy Bryant, has spoken often about how the company's integrated business model, in which "everybody works for the same owner," helps it move faster than the multiheaded ARM ecosystem and allows it to more thoroughly transfer knowledge while keeping the technology proprietary, thereby producing the unique benefit to those using Intel's products. The company can use its unmatched scale to transfer this technology leadership broadly and rapidly to the whole industry in a way a single, smaller ARM CPU vendor cannot. This scale

allows technology innovation to diffuse much faster into the OEM ecosystem, and it creates a "level playing field" for OEMs with regard to raw hardware features, enabling them to compete on other factors such as branding, supply chain management, software, and device user experience (Exhibit 4).

Clashing technologies: Convergence and competition

Beyond the rather stark differences in business models, each architecture brings different technical strengths to bear. Technology still matters, and CPUs are highly complex products—among the most difficult products in the world to design and manufacture. However, the technology competition between ARM and x86 will not only be about which architecture is technically "more efficient" or "better" for mobile computing. It will primarily involve what happens beyond the actual CPU architecture. We believe there are four key success factors required to build a better CPU chip; we can examine how each architecture currently stacks up in these dimensions.

CPU microarchitecture. Fundamental differences between x86 and ARM CPU cores remain, but those will lessen over time as each architecture works to adopt the best technical features of the other. ARM has always possessed less power than x86 because of its reduced ISA, which resulted in much lower performance. As ARM increases its processing capability to match the requirement of new mobile devices, it is adopting x86 core technologies such as multicore chips and deeper processing pipelines. Historically, ARM chips were much smaller than x86 CPUs. However, with the emergence of the low-end Atom architecture

Exhibit 4 The x86 mobile-device value chain is more integrated than that for ARM.

	x86 ecosystem	ARM ecosystem	
Instruction set and CPU core		ARM	
Additional processing blocks and CPU core modifications	Intel Advanced Micro Devices (AMD)	Qualcomm Synopsys ARM CEVA Texas Instruments Imagination Tec Mentor Graphics Marvell	chnologies
System-on-a-chip integration	-	Qualcomm ST-Ericsson MediaTek Broadcom Texas Instruments Samsung NVIDIA Apple	
Chip fabrication ¹	Intel AMD GLOBALFOUNDRIES	Taiwan Semiconductor Manufacturing Company United Microelectronics Corporation SMIC Samsung GLOBALFOUNDRIES	
Operating systems	Microsoft Android	Android Microsoft iOS Symbian BlackBerry	
Other equipment manufacturers	Hewlett-Packard Acer Dell ASUSTeK Computer	Nokia Huawei HTC Sony Ericsson ZTE Samsung	

that th x86 RM nced ts)



¹Other ARM-licensed foundries include Chartered Semiconductor Manufacturing, Dongbu HiTek, IBM Microelectronics, TowerJazz, and Vanguard.
²Original equipment manufacturers.

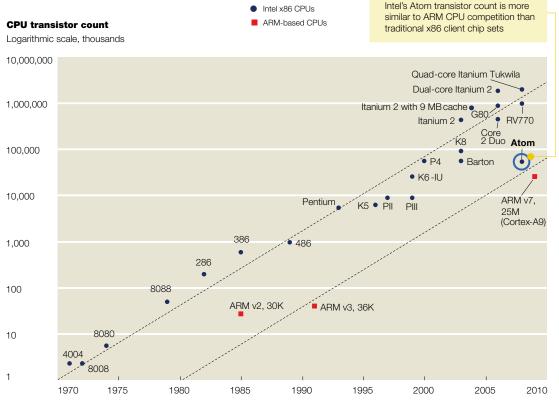
from Intel and the increasing transistor count in ARM chips, the physical sizes of cores are converging, from the low end of x86 to the high end of ARM (Exhibit 5).

Process technology. By moving from one semiconductor manufacturing node to the next with smaller transistors, CPU products can gain 20 to 40 percent in performance, translating into lower power or lower cost than previous generations—or a combination of both. Intel has been the process leader for decades, traditionally

holding a 12- to 18-month lead over major competitors in moving to a new generation of process technologies. It also ramps new technologies much faster than foundries, which maintain a significant amount of volume at lagging-edge lithographies. However, because there are more foundries pursuing the ARM business, they can offer a greater variety of process recipes (such as low-power processes focused specifically on maximizing battery life) than can Intel alone. The technology transition to smaller and smaller technologies will not be easy,

Exhibit 5

The transistor count in ARM processors has increased, growing closer to that for traditional x86 CPUs.



Source: ARM; Intel; Sanford Bernstein; McKinsey analysis

and it poses several hurdles. Intel began shipping 22-nanometer process products based on 3-D transistors in the second quarter of 2012, while the ARM ecosystem was only at 28 nanometers. If Intel can maintain or even extend this technology lead, its x86 products will possess greater processing or power-saving capabilities than ARM-based competition.

CPU and wireless-baseband integration capability. As mobile devices get smaller and price competition increases, there are size, power, and cost advantages to reducing the number of semiconductor chips. In basic and feature phones, all CPUs are integrated into wireless-baseband semiconductors that control the phone's communications. Personal computers and tablets have discrete semiconductors for CPUs and wireless basebands. Smartphones have both integrated and discrete configurations. ARM-based CPU vendors with discrete and integrated configurations can offer greater breadth in CPU products to OEMs competing in different categories across all mobile devices. The ARM-based CPU vendors can simplify the software-integration task that the device seller will have to undertake to make a radio work with a CPU. To that end, Intel purchased Infineon's baseband business in 2010 and has the option of integrating x86 CPUs and basebands. However, ARM-based vendors Qualcomm (which had 43 percent market share through the fourth quarter of 2011), MediaTek, Spreadtrum Communications, Broadcom, and Intel's Infineon division—the top five baseband suppliers already sell a complete portfolio of integrated ARM CPU products. Intel's x86 product line will need to master the technically difficult task of logic and wireless integration to catch up to

ARM. Qualcomm's 60 percent or more of 3G baseband share makes this an uphill climb. As such, Intel's integrated products have to take share from an ARM-based competitor with a highly defensible installed base.

Application compatibility. Consumers purchase a device because of the software and applications it can run. Application programmers write to an application environment or operating system—not an ISA. In the past, x86 and ARM have supported different operating environments for different applications. However, that distinction is fading, as both architectures are working to fully support all programming environments across all mobile devices. In fact, 2012 is proving to be the year in which cross-architecture operating-system compatibility becomes reality. By the end of 2012, all the major operating systems (for example, Android, Windows, and Windows Phone) and application environments (such as Flash, HTML5, and Webkit) will work on both architectures. Much like in the vertical-versus-horizontal debate, Apple could be a swing vote; it has been reported that Apple is also porting its iOS and Mac OS operating systems between the two architectures, but the company has made no public announcements on the matter.

Operating systems supporting both architectures use a "middleware" framework that abstracts the hardware from the software, making the end applications run on both architectures with minimal difference in performance. Android's system-development kit (SDK), for instance, allows developers to write applications that work on any Android system, regardless of the CPU. Certain high-performance applications for gaming or multimedia processing use native Android software code—not the SDK—to leverage

application-programming-interface command sets to maximize processing performance. While we expect Android and other operating systems to provide supplementary middleware allowing developers to write one set of native code that will work on both architectures, software translation always involves processing overhead and reduces performance. If one architecture is able to gain the lion's share of natively developed applications, the other will be able to maintain application compatibility, but at a performance cost.

Looking to the future

Where does this leave us? One view of the future sees Apple and Samsung extending their lead and deepening their investment in semiconductor capabilities that were once solely owned by merchant silicon vendors. Samsung could enter the wireless-baseband market through acquisition or internal development, or Apple could partner with a semiconductor foundry to develop proprietary access to new process technologies. Other OEMs may well follow these paths and the semiconductor industry could become primarily a vertically integrated, OEM-driven market. For this model to be successful, vertically integrated players would need to keep their market shares high to justify the technology investment, while driving the CPU technology as fast as the best merchant silicon vendors would.

Alternatively, the Chinese smartphone vendors, the smaller global OEMs and the global PC manufacturers could break this global smartphone-tablet duopoly with the strong support of Intel, Qualcomm, and other merchant silicon players, leaving the horizontal model supreme.

The x86-versus-ARM architectural battle is a multiround game. Both the ARM and x86 ecosystems have the financial model, the annuity cash flows, and the technology base to compete in the mobile-computing space for the next five years. Intel and AMD's revenue exceeds \$50 billion, with about \$15 billion in operating cash flow. ARM's annual revenue is modest, at about \$700 million, but its partners generate \$30 billion to \$50 billion in silicon revenue—enough to drive multiple generations of process technology and new designs. Both architectures can fund investments to advance their design and process technologies for several years without prevailing over the other architecture in the battle to dominate the mobile-computing landscape. With such rapidly expanding consumer demand, even the "losing" architecture could still see revenue growth. ARM's best chance for success will not be achieved through displacing x86 from its traditional home, PCs, but rather through the expectation that ARM-based tablets and smartphones will cannibalize PCs. The x86 camp's best chance would be if Intel builds a sustainable lead in process technology to create products for its growth markets of smartphones and tablets with unmatched performance and power dissipation. These products would need to be so good that leading OEMs had to adopt them, despite the challenges of adopting a new architecture.

However, even if one architecture gains the upper hand, every new CPU product launch, every new version of Android or Windows, and every device-level transition is an opportunity for an OEM to choose a new architecture. For one architecture to truly triumph over the other, either ARM or x86 would need to string together an unbroken set of "transition wins" over many

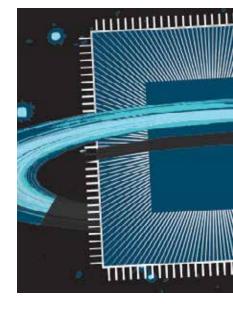
years to develop a permanent lead that convinces the other architecture and its ecosystem to accept a secondary role.

This increased competitive intensity in the semiconductor industry will drive an increase in the competitive metabolism of the device industry. The heightened competition will also drive consolidation along the ARM value chain, either by vendors exiting the market or shifting their design focus from mobile devices to other promising markets, such as Texas Instruments' recent announcement that it will focus its ARM-based CPU business on home networking and machine-to-machine communications.

participants face the most expensive and hardest-fought battle in their history. Semiconductor industry players need unmatched market insight, aggressive technology road maps, world-class business development, and operational excellence just to punch their entry ticket to compete in this arena, and the winners will need to combine all four elements (and perhaps a little luck) to emerge victorious. •

• • •

Technology transitions have always created winners and losers. The mobile device industry is experiencing several changes right now, as the conversion from feature phones to smartphones reaches its apex, tablets move from a niche to a must have, and semiconductor industry



Harald Bauer (Harald_H_Bauer@McKinsey.com) is a principal in McKinsey's Frankfurt office, **Yetloong Goh** (Yetloong_Goh@McKinsey.com) is a consultant in the Singapore office, **Joseph Park** (Joseph_Park@McKinsey.com) is an associate principal in the Seoul office, **Sebastian Schink** (Sebastian_Schink@McKinsey.com) is a consultant in the Munich office, and **Christopher Thomas** (Christopher_Thomas@McKinsey.com) is an associate principal in the Beijing office. Copyright © 2012 McKinsey & Company. All rights reserved.