GILDER
TECHNOLOGY
REPORT

November 2005 / Vol. X No. 11

The March of Moore

Altera and Xilinx own about 85% of the rapidly growing FPGA market. FPGAs will win by serving the bulk of the market with chips that are good enough.

Overflowing the Resort at Squaw Creek in Tahoe in late September into hotels in the nearby ski village, this was the Telecosm of the "Singularity." A singularity designates a point in the future beyond which the "event horizon" darkens, as the horizons of the past darken beyond the reaches of the Big Bang. In between, we are to believe, is the known universe. But still in the dark remains the question of when and whether **Broadwing** (BWNG) will break out into profits using MPLS (multi-protocol label switching) on the intelligent edge and fast all-optical switching at the core of its still industry-leading network, while Cisco (CSCO) struggles to keep the smarts in the core. Even the present is enigmatic. We have little assurance whether bandwidth prices are stabilizing, as Jay Adelson of **Equinix** (EQIX) reported in a fascinating speech, or whether they are continuing on a downward plunge as confidently testified **Cogent** (COI) CEO Dave Shaeffer. Nor, even after all of the earnest explanations of CEO Eli Fruchter and CTO Amir Eyal, do we know when **EZchip** (LNOP) will begin an explosive ascent of revenues for 10 gigabit Ethernet and line-card processors. And despite the presence of representatives of both **Foveon** and **Synaptics** (SYNA), we still don't know when these two kindred companies will burst into the huge markets for teleputer sensors and imagers (though news from Foveon has been picking up since the conference, including the special "Progress Medal" from the Royal Photographic Society in London).

Introducing a dazzling new best seller, *The Singularity is Near*, and generously giving a copy to each of the attendees, Ray Kurzweil acknowledged that macrofuturism, projecting Moore's law in all directions, is much easier than micro, predicting what will happen to specific companies and technologies. Nonetheless, on stage the first night of Telecosm, Ray faced a skeptical micro question from yours truly on the dismal failure of several teams of robotic engineers last year to create a device that could negotiate a DARPA course through the Mojave desert without plunging off the road into a ditch or an infinite loop. In response, Ray confidently asserted that teams from Carnegie Mellon and Stanford would succeed in this task in October. Sure enough, he was on the button with this prophecy. So far, so good.

At the heart of his larger prophecies is the continued exponential progress of all the arts and sciences of information technology on beyond machines into a biological Singularity. Ray's intriguing argument is that today's exponential curves merely follow in the train of the original evolutionary curve, which also reveals an ever accelerating pace of advance—some 13 billion years from the exquisitely calibrated bang to the biosphere, with DNA processing in the eukaryotic (nucleated) cell, then the Cambrian explosion of life forms some 3 billion years ago, and then the rushed ascent of punctuated equilibrium to the emergence of man and Ray and the Telecosm list, after which things really start popping.

Discerned in all this heroic ascent is scant intelligence at all until the arrival of human technology, though the information processing underway in the some 300 trillions of cells in your body, each with some 6 billion base pairs of DNA programming, excels the output of all the world's supercomputers with all their intricate software and firmware. As Ray points out (p. 209), the ribosomes that translate DNA into amino acids accomplish 250 million billion "read" operations every second just in manufacturing the hemoglobin that carries oxygen from the lungs to body tissues. While the genes are digital, much of the biocomputing is inscrutably analog. But in another four decades, so Kurzweil calculates, digital machine intelligence will exceed human intelligence, precipitating the Singularity.

Humans, he predicts, will use the machines massively to extend our lifespans and to project the reach of our learning both into our own brains, mastering the mysteries of consciousness, and out into space, with an imperial march of human intelligence incarnate in our machines and in our newly bionic bodies. It is a grand and triumphant trajectory of thought on which Kurzweil is launched, and his argument is finely mounted and gracefully written, with much self-deprecating humor in artfully shaped "dialogs" at the end of each chapter. But as some attendees groused, it would be nice if by the time of the Singularity, or even before, **Microsoft** (MSFT) could get Windows to boot in less than four seconds and could avoid the darkened event horizons of its chronic blue screens. And after many projects at Caltech attempting to use neuromorphic models as the basis of electronic simulations of brain functions, Carver Mead observed that we still have no idea of the workings of the brain and nervous systems of a common housefly. As I describe in *The Silicon Eye*, it goes about its business, eluding the swatter and garnering chemical sustenance in the air, all on microwatts of power using means that remain beyond the grasp of our most sophisticated neuroscience.

Oh, well, observed Nick Tredennick, author of the following masterly ruminations, all these exponential curves look flat to the engineer attempting to solve the immediate problems he faces. So back to work, folks. ***– George Gilder***

The Cost of Culture

As a pilot in the air force, I learned the value of rote procedure and of attention to detail. When an engine fails or a wing is on fire, you must do the right things immediately—in the right order—or people might die. There's no time for rumination, discussion, or negotiation. Do the right things, now, or die. These rote procedures, extensively rehearsed, work. That they are so successful in specialized situations such as aircraft in-flight emergencies has encouraged their spread to the organization at large.

Fellow pilot Ken McClure summed the sometimes unfortunate consequences of this broader application, in response to a high-level decision that was insane in its local implementation: "Don't let judgment interfere with procedure."

The corporate rebuttal is that something that has always worked in the past needs no justification, even if the circumstances have changed.

The circumstances have changed.

But the semiconductor industry still offers these rote solutions as the solutions to anything: "shrink the transistors" … "use a microprocessor" … and "use PC memory." Each of these can be summarized as "ignore power consumption" … "ignore power consumption" … and "ignore power consumption."

With the new measure of goodness for electronic devices being cost-performance-per-watt, you see why I think major semiconductor companies are headed for a hard time.

Shrink the transistors

Shrinking transistors grows the market in two ways. More low-end applications are affordable and more high-end applications are doable. Suppose the largest-selling chip in the previous semiconductor process was a one-million-transistor chip. In the new process, that same one-million-transistor chip is much cheaper because its transistors fit in half the space occupied by the previous generation's transistors. Cost varies exponentially with chip area because yields rise and because more chips fit on the wafer. The lower cost per transistor justifies applications that were too expensive before. And the new process offers a two-million-transistor chip for the cost of the previous generation's one-million-transistor chip, enabling new applications at the high end.

The magic of increasing the market by shrinking transistors was built into the business models of chip companies beginning in the late 1960s and is still the core business strategy of today's integrated device manufacturers (IDMs) such as **Advanced Micro Devices** (AMD), **Freescale** (FSL), **Infineon** (IFX), **Intel** (INTC), **STMicroelectronics** (STM), and **Texas Instruments** (TXN). The strategy has been working for 40 years—it made Intel the world's largest and most profitable semiconductor company—so who's to say that it won't be successful for the next 10 to 15 years? These IDMs have built enormous momentum. (Momentum is what you *build* when you keep doing what you're *doing;* inertia is what you feel when you try to change course.) But, in the course of growing their businesses and of developing their corporate cultures, these IDMs may have transformed beneficial momentum into crippling inertia.

How so? Costs are rising and demand for leading-edge transistors is slowing. Chip cost divides into fixed cost and variable cost. Fixed cost is the cost to operate the production line, usually calculated on a per-wafer basis. Variable cost includes building the plant and furnishing the equipment, process development, and masks. For the first three decades of shrinking transistors, fixed costs dominated and variable costs were essentially inconsequential. But, while fixed costs have risen only slightly, variable costs have been doubling with each process generation. Variable cost is now as significant as fixed cost and it is on its way to dominating total cost.

The consequence of the escalating variable cost is that the

investment to shrink transistors grows rapidly from generation to generation. Eventually, the investment to reach the next smaller transistor exceeds the investment for other circuit innovations that supply equivalent advantages. I believe that the industry is years *past* the point when it should have invested in less-expensive alternative innovations.

It's an opportunity for an innovator such as startup **Tezzaron Semiconductor**. Tezzaron has developed a process for wafer stacking. The 3-D wiring of a chip stack is significantly shorter than the 2-D wiring of conventional chips, increasing speed and decreasing power (short-distance drivers are smaller). Tezzaron's 3-D DRAMs (dynamic random-access memory) run five to ten times faster than normal DDR-2 DRAMs. With stacking, it's possible to stack logic, memory, analog, and even to mix process generations. Fault tolerance can be improved through redundancy. Other companies, such as **Cubic Wafer** (once Xanoptix), **Matrix Semiconductor**, **Ziptronix**, and **ZyCube** are working on innovations that could be cost-effective alternatives to shrinking transistors.

When shrinking transistors was cheap, only process developers (IDMs and foundries) participated in semiconductor progress. With the increasing cost of shrinking transistors, progress in semiconductors is now open to a broader range of participants. Semiconductor development will accelerate.

Do we really need smaller, faster transistors? Ninety-nanometer semiconductor processes are good enough for most applications. I expect the foundries to add capacity in 130-nm and in 90-nm processes in lieu of extending to 65 nm and below. Partly this is due to emerging markets in China, India, and other countries, where demand will be for microwave ovens, hair dryers, blenders, and other consumer appliances that don't need leading-edge transistors. Among volume consumer products only mixed systems-on-a-chip for cell phones, games, and video players entail leading edge processes. But these devices may require analog, media processing, and radio-frequency circuit advances that are not part of the standard CMOS Moore's law roadmap.

For the bulk of the market, value transistors, the cheapest transistor that's good enough, will do. Demand will grow rapidly in applications that don't need leading-edge processes, which will be reflected in the growth of foundries that add capacity to meet demand rather than in the IDMs that have shrinking transistors built into their business models.

Once nimble enough radically to change its strategy, Intel gave up its primary memory-chip business to focus on shrinking transistors for microprocessors in personal computers, building huge momentum that will become inertia as it tries to adjust its business to the changing environment. To give two examples, I believe that Intel should sell foundry services and that it should license a soft-core x86 if it wishes to meld smoothly into the future of the semiconductor industry. Either of these suggestions would be heresy inside Intel; that's the effect of culture.

Similarly, suppliers of semiconductor processing equipment have been following their customers for 40 years, building their businesses and corporate cultures on leading-edge equipment for shrinking transistors. If transistors are good enough (value transistors), will **Applied Materials** (AMAT), **KLA-Tencor** (KLAC), **Lam Research** (LRCX), and **Novellus** (NVLS) shift from building leading-edge equipment to building equipment for value fabs? It seems unlikely.

Mask costs approximately double with each process generation. Plant cost and process development cost also double with each generation, but here I'll consider only the effect of the doubling mask cost. In the early days, a mask set was less than $20,000. Because the mask set is unique for each design, its cost is amortized across the chip's market. If the application is a medical instrument that sells 20,000 units, the contribution of the mask cost to the integrated circuit is one dollar. Double the mask cost for the next-generation product and the instrument will need a market of 40,000 units to maintain the one-dollar contribution to the total cost. If the size of the end market is fixed, the masks contribute two dollars to cost. Today's million-dollar mask set needs a million-unit market to maintain its cost contribution or it adds $50 to the unit cost if the market size is fixed at 20,000 units.

As mask costs rise, fewer applications are cost effective.

Ascendant are foundries, such as **Taiwan Semiconductor** (TSM), **United Microelectronics** (UMC), **Semiconductor Manufacturing International** (SMI), **Grace**, and even the eternally frustrating, on and off the list, but now impressively reviving **Chartered** (CHRT) of Singapore with the shared knowledge base of advanced fabs at IBM, Infineon, and Samsung. Pushing upstream against the trend are the long dominant integrated device manufacturers, who fuse design with fabrication, such as Intel, AMD, **IBM** (IBM), Texas Instruments, and **Analog Devices** (ADI). The ascendant foundries are demand driven; they build a mix of production capacities for leading-edge and for value transistors to match customer orders. This change in dominance of the industry may go even further. Just as the advent of "value steel" encouraged the development and ascendance of minimills, the advent of the value transistor may encourage the development and ascendance of mini-fabs.

Vendors of leading-edge semiconductor processing equipment will either change their strategy to meet the growing market demand for cost-effective equipment or they will lose market share. They will have to change from building custom equipment for leading-edge processes to building modular equipment with standard interfaces for high-throughput, high-efficiency semiconductor processing. Second-tier equipment manufacturers, such as **Semitool** (SMTL) and **Ultratech** (UTRA) may thrive, as may businesses such as **Tower** (TSEM) of Israel that can use refurbished trailing-edge equipment.

Use a microprocessor

A microprocessor is a chip that simulates the chip you actually want by throwing transistors and software at the problem. Until now, that has been good enough.

TELECOSM TECHNOLOGIES

Advanced Micro Devices (AMD)
Agilent (A)
Altera (ALTR)
Analog Devices (ADI)
Broadcom (BRCM)
Broadwing (BWNG)
Cepheid (CPHD)
Corning (GLW)
Equinix (EQIX)
Essex (KEYW)
EZchip (LNOP)
Flextronics (FLEX)
Intel (INTC)
JDS Uniphase (JDSU)
Microvision (MVIS)
National Semiconductor (NSM)
NetLogic (NETL)
Power-One (PWER)
Qualcomm (QCOM)
Semiconductor Manufacturing International (SMI)
SK Telecom (SKM)
Sprint Nextel (S)
Synaptics (SYNA)
Taiwan Semiconductor (TSM)
Texas Instruments (TXN)
Wind River Systems (WIND)
Xilinx (XLNX)
Zoran (ZRAN)

Note: The Telecosm Technologies list featured in the Gilder Technology Report is not a model portfolio. It is a list of technologies that lead in their respective application. Companies appear on this list based on technical leadership, without consideration of current share price or investment timing. The presence of a company on the list is not a recommendation to buy shares at the current price. George Gilder and Gilder Technology Report staff may hold positions in some or all of the stocks listed.

Advanced Micro Devices (AMD)

PARADIGM PLAY: INTERNET COMPATIBLE PROCESSORS

OCTOBER 17: 21.37; 52-WEEK RANGE: 13.66 – 26.07; MARKET CAP: 8.51B

AMD used to stand for "Hey, over here, we Also Make Desktop chips!" Today, the acronym has earned the title the company has always claimed for it, as the chipmaker increasingly reaps the rewards of the far-sighted technology bets it placed over the past several years in multicore, low-power processors and in the x86 architecture for 64-bit chips. AMD continues to gain on Intel in microprocessors for servers, desktops, and notebooks. In the June quarter, 8% of all servers shipped had AMD chips inside, almost double the 4.8% a year earlier and sales of dual-core Opteron chips were crucial to AMD's 89% sequential revenue increase in server products. Overall, microprocessor revenue grew 2% to a record $767m. Also helping was a strong ramp of Turion64 notebook processors, having been designed into 60 laptop models.

Then came September. Processor revenue surged 26% to $969m, with the charge this time led by desktop chips, followed by notebooks. Meanwhile, growth in the server market continues, where AMD claims customers such as IBM, Sun, and HP (Intel's partner on the ill-fated Itanium but now the largest seller of Opteron-based systems).

So, on cue, the stock price fell 8% the day after last weeks' earnings call, as the worrywarts came out of the woodwork, finding nits in all-time high gross margins, 10% sequential growth projections, and a possible bulge in capacity from the opening of Fab 36 in Germany, not to mention burdensome depreciation and operating costs associated with new facilities. Whoa, folks. That's how ascendant companies grow. That's how AMD could introduce 20 new 64-bit processors in the quarter. And it's how AMD will take 30% of the server market by the end of next year, as some inside the company believe they can.

AMD currently trades at an enterprise value of $11b or 17.8x the annualized operating profit of its microprocessor business. It's a buy. *– CB*

Amedia Networks (AANI.OB)

COMPANY TO WATCH

OCTOBER 17: 0.90; 52-WEEK RANGE: 0.75 – 2.00; MARKET CAP: 18.86M

Only 15 months after opening its doors, Amedia Networks (see March GTR) was selected in June by Tai Long Communications to enable 100 Mbps symmetrical access over fiber for up to 50,000 subscribers in China. During the next two years, Tai Long will spend some $9 million deploying Amedia's Ethernet equipment in its central offices and customer homes and businesses. The contract gives Amedia an enterprise value of 3.5x sales at the current share price of 90 cents. On the surface, that's not a bargain, considering that the company is burning about $1.7m per quarter with net cash down $3.25m as of June. But Amedia is young and has the potential to earn large returns for investors since small increases in sales dollars translate to large percentage increases in revenue; just two additional modest wins similar to Tai Long's would triple revenue and double the stock price at today's enterprise-value-to-sales multiple.

However, fraught with the risks of early-stage companies and technologies, we keep Amedia on our "Companies to Watch" list. Financially, Amedia has a chunk of options, warrants, and converts outstanding that could dilute shares by up to 70 percent while raising another $23 million of much needed cash. Marketwise, Amedia competes with scores of companies vying to supply last-mile optics or to obviate them with copper enhancements such as the 200 Mbps DSL links described by DSL inventor John Cioffi at Telecosm. Since none of the major global telcos are currently considering fully-active last-mile architectures, Amedia probably will have to settle for wins at niche networks, while innovating wireless capabilities of the products.

Longer term, CEO Frank Galuppo awaits a repentant Verizon, as it attempts to migrate from its current ATM-based BPON (broadband PON) technology to IP-based GPON (gigabit PON) next year. An apparent kludge, GPON uses ATM for voice, Ethernet for data, and proprietary encapsulation for video. Even now, according to Amedia, Verizon and SBC are buying large, layer-2 terabit Ethernet switches from Cisco and Force10 Networks—and not just for data. One of the keys to Amedia's win in China was the interoperability of its equipment with other vendors' layer 2/3 switches. A follow-on win with an RBOC could make Amedia a home run, but it will be a very hard feat to pull off. *– CB*

Equinix (EQIX)

PARADIGM PLAY: STOREWIDTH STAR. WHERE STORAGE & BANDWIDTH CONVERGE

OCTOBER 17: 37.14; 52-WEEK RANGE: 31.39 – 46.39; MARKET CAP: 886.31M

Equinix shares are off more than 20% from their recent summer highs of 46. Most of the drop came in the last week since the company filed a shelf registration of 10.2 million shares on October 6. STT Communications, an affiliate of Singapore Telemedia, plans to offer the shares, bringing its total equity stake to around 35%, depending on your accounting method. But total shares outstanding are just around 30 million, not an unmanageable float and the stock price has already taken its hit on this transaction.

Like many capital intensive businesses, Equinix has lost money as it built out its infrastructure. But with few variable costs and recurring revenue representing some 95% of its business, Equinix can leverage its financial model and strategic position at the heart of

MEAD'S ANALOG REVOLUTION

NATIONAL SEMICONDUCTOR (NSM)
SYNAPTICS (SYNA)
SONIC INNOVATIONS (SNCI)
FOVEON
IMPINJ
AUDIENCE INC.
DIGITALPERSONA

COMPANIES TO WATCH

ADAPTIX
AMEDIA (AANI.OB)
ATHEROS
ATI TECHNOLGOIES (ATYT)
BLUEARC
COX (COX)
ENDWAVE (ENWV)
FIBERXON
LINEAR (LLTC)
LUMERA (LMRA)
ISILON
LENOVO
MEMORYLOGIX
NOVELLUS (NVLS)
POWERWAVE (PWAV)
SAMSUNG
SEMITOOL (SMTL)
SIRF
SOMA NETWORKS
STRETCH INC.
SYNOPSYS (SNPS)
TEKNOVUS
TENSILICA
VIA TECHNOLOGIES
XAN3D

the Net into a very profitable business. Continued sequential growth of 7%, for example, would yield 2006 sales of $285 million, compared to total 2005 expenses of around $231 million.

With Internet traffic continuing to double each year, Equinix has been acquiring all available data center facilities at discount prices from other companies now exiting the business, thus preempting possible competitors in its targeted markets. With some racks of blade servers now running at almost 2 kilowatts, each Equinix IBX now consumes some 25 megawatts. Sucking this much power into a building and then pulling the heat out is a major challenge, and a further barrier to entry for rivals.

Some Equinix customers, like Google, have now gotten so large they have begun building their own data center buildings right next door to the Equinix facilities. Confirming the value of the IBX model, they run short optical links to the networks housed at Equinix. *– BTS*

EZchip (LNOP)

PARADIGM PLAY: A GENERATION AHEAD IN NETWORK PROCESSORS

OCTOBER 17: 6.53; 52-WEEK RANGE: 5.83 – 15.17; MARKET CAP: 69.39M

EZchip added at least 7 new design wins for its NP-2 network processor in August and September. Presenting at our 9th annual Telecosm conference at the end of September, CEO Eli Fruchter said sales of the company's first generation chip, NP-1c, were still slow and unpredictable. But NP-2, he said, now enjoys 30 total active designs, including wins at "three of the top five networking companies," defined as Cisco, Juniper, Alcatel, Huawei, and Nortel. EZ is actively courting the other two networking giants. After four-plus months of testing in-house and at top customers, NP-2 samples also have not encountered any significant problems, meaning volume shipments of the chip can begin on schedule at the beginning of 2006.

With 64 total active designs in the pipeline, Fruchter says his NPUs are beginning to show up in telecom linecards, such as new ATCA standard devices, much as microprocessors and other parts are used to build industry standard PC motherboards. Only highly programmable general purpose NPUs can be used in such a modular way.

Fruchter revealed that the company's next product, NP-3, will be produced using 90-nanometer technology and will integrate Ethernet PHYs (physical layer circuitry normally performed on stand-alone chips) and also the switch fabric interface. A "line-card-on-a-chip," NP-3 will thus contain everything but the optics.

At 130 nanometers and below, NPUs become "pad limited," meaning the chip's size is dependent on the I/O pins, not the logic circuitry. Apparent across the industry, this I/O bottleneck, which brings Moore's law to a halt, gives impetus to the 3D chip stacking technologies of Cubic Wafer (once Xanoptix), Zycube, Matrix, and others. Stacking memory on top of logic would not only increase memory bandwidth but would also eliminate the crunch created by proliferating I/O pins. Chips could thus keep getting smaller and cheaper, and Moore's law could resume its march. *– BTS*

MICROPROCESSOR WARS

Mastering the complexity of hundreds of millions to billions of transistors, microprocessor developers have recently abandoned the quest for clock speed and have begun to integrate multiple microprocessor cores on a single chip. All that can be done for uni-processors—pipelines, out-of-order execution, threads, higher clock speeds, caches—has been done. With multiprocessors, the problem is being thrown over the wall from hardware to software. Instead of struggling to turbocharge the performance of a single processor, designers now struggle with multiprocessor coordination, memory, and connectivity issues, such as how does the chip deal with the flood of input/output needed by the chip's many processors?

Experimenting with a wide range of applications are such companies as **Broadcom** (BCM 1840, 4 cores), **Cavium** (Octeon, 16 cores), **IBM** (Cell, 9 cores), **RMI** (XLR, 8 microprocessor cores), and Sun Microsystems (Niagara, 8 cores).

Meanwhile, at the bottom end we've been suggesting for years that the x86 will invade embedded systems and that it will eventually displace ARM in cell phones. Here are some x86 advantages.

1. PCs based on the x86 provide the brick and mortar for the structure of the Internet, with hundreds of millions of more-or-less permanently connected x86 computers that run tens of thousands of applications. Mobile devices connect individually to this aggregation; it is a one-to-many connection no matter how many mobile devices there are. This means that x86 instructions provide a benefit to cellphones and teleputers linking to the net, but there is no aggregation benefit derived by the ARM processors in the 600 million individually connected cell phones.

2. Beyond the users's familiarity with the Windows graphical user interface, the x86-based PC is the universal development platform. Even ARM applications are built and tested on x86-based PCs. Portability of applications and data is easier and, therefore, cheaper, if the underlying microprocessor is the same.

With the potential for hundreds of millions of transistors on a chip, the complexity of implementing the instruction set is not important. The x86 instruction set isn't enough worse than anything else to make a difference. X86 microprocessors can be designed for low power and for cost-effective operation. MIPS and PowerPC began life as performance-oriented workstation designs. Only after being pushed out of the market by x86-based PCs coming up market did they become embedded microprocessors. ARM started life as the CPU in the Acorn RISC Machines' computers and was pushed into embedded applications by the all-consuming x86. All that is required is a redesign from performance (in the case of workstations) or from cost-performance (in the PC) to a cost-performance-per-watt orientation. If ARM, MIPS, PowerPC, Sparc, and others can be redesigned for embedded applications, so can x86. X86 can compete with ARM on monetary cost and energy cost, and **Via Technologies** and **Advanced Micro** (AMD) and **MemoryLogix** and perhaps even **Intel** (INTC) will do so. (Already 60 percent of Via's chips go into embedded applications.)

For builders of a consumer system, it might be a good idea to begin with a small-form-factor PC motherboard. All of the I/O and drivers for any imaginable peripheral will already be available at low cost. The engineers can do development and testing on their desktop or laptop computers with no need to port to a new instruction set or operating system—avoiding major duplication of test and verification.

The simplest way to achieve compatibility with the development system is to be compatible with the development system. The simplest way to achieve compatibility with x86 applications is to be an x86. We expect cell phones to add an x86 to the ARM and DSP (digital signal processor) that are already there in order to facilitate Windows compatibility. Once the x86 is there, it's there to stay and the ARM and DSP are candidates to be displaced. It will take some time because no one yet offers an x86 core for licensing.

ARM is displacing older microprocessors and it will continue to do so, but that's no guarantee that it cannot be displaced. If we just took a snapshot of who owns the market at a particular time, we would have predicted that IBM's 360/370 would dominate, then DEC's PDP-11, and then Sun's SPARC or **Silicon Graphics**' (SGI) MIPS microprocessor.

As for other processors that "can do Windows" or can run Linux, the libraries and drivers available for anything except x86 are skeletal, barely enough to show that it can be done. And there's always a cost to supporting alien architectures that gives x86 developers a cost advantage.

Intel still dominates in dollar volume of embedded microprocessors, but is focused elsewhere. With its priorities oriented toward x86 and an impressive array of new offerings, AMD is well positioned with its HyperTransport processor, Athlon, and its Geode line of embedded microprocessors. Via is also positioned well for emerging markets in Asia and for its embedded products.

– Nick Tredennick and Brion Shimamoto

In 35 years, integrated circuits have grown from a few hundred transistors to a few billion. In early designs, active power (the energy to switch transistors) mattered because transistors were big and they were often busy. Leakage currents (and, therefore, leakage power) could be ignored because the active power of a transistor was a billion times that lost to leakage. Active power decreases as transistors shrink, but the leakage current increases. In two process generations, the leakage current increases tenfold—four times as many transistors, all leaking, fit on the chip. If the chip's active power remains constant, then the leakage power makes up its nine-orders-of-magnitude deficit in fewer than a dozen process generations.

That's where we are today; 5 microns to 90 nm has been twelve process generations. Leakage power is as important as active power, and business as usual—programming microprocessors to simulate custom hardware—isn't efficient enough. The new metric of cost-performance-per-watt implies the need for more efficient design. Transistor designs and efficient transistors go from being the abundant resource to being precious.

The problem with microprocessor-based designs is their dismal efficiency. Low efficiency was OK for cost-performance-based systems because these systems could burn more power to reach higher performance. Most of these products satisfied consumer markets that demanded low cost. Because the microprocessor was cheaper than custom logic both in design cost and in component cost, it displaced custom logic where its performance was adequate. Most of these systems got power from wall sockets, so energy use wasn't a design constraint. But the emergence of the value PC and the industry's transition to mobile systems is changing the design goal from cost performance to cost-performance-per-watt. Microprocessor-based implementations aren't efficient enough for these systems. The quest for more efficient implementations will bring back innovation in chip design from companies such as **Altera** (ALTR), **ARC Cores**, **Celoxica**, **CriticalBlue**, **Stretch Inc.**, **Tensilica**, and **Xilinx** (XLNX).

Use PC memory

Just as the industry focused on shrinking transistors and on speeding up microprocessors, it also only cared about memory that worked in PCs.

The personal computer was introduced with read-only memory (ROM) and with DRAM. The ROM held the computer's initialization software, the DRAM was the working memory between the microprocessor and the disk's bulk storage. Flash memory displaced ROM because it enabled updating of the initialization software. As the PC improved and as its market grew, the designers of its microprocessor optimized performance while the designers of its DRAM chips optimized storage capacity. This divergence of design goals between the microprocessor and memory led to a performance gap. Microprocessors grew faster and DRAM grew larger. On-chip and off-chip caches, built with static random-access memory (SRAM), bridged the performance gap between fast microprocessors and slow DRAMs.

As the PC's memory components came down the learning curve. High-volume production led to low cost for SRAM, DRAM, and flash memory. Each of the PC's memory components has advantages and disadvantages. Flash memory keeps its content through power cycling, but is slow for reading, very slow for writing, and it eventually wears out. DRAM offers high storage density, but is slow for reading and writing and it loses its content on power cycling. SRAM is fast, but it lacks density, it hogs power, and it loses its content on power cycling. The PC benefits from the good features of these memory types and it masks their flaws. Thus, there has been little incentive to develop new memory types.

At its introduction, the PC wasn't good enough to meet the requirements of any of its users. After 25 years of development, however, PC performance satisfies most users, who now buy value PCs with good-enough performance and a low absolute price. Engineering resources that were once dedicated to improving the PC are being reallocated to higher-margin markets, such as mobile applications. That changes the design goal from the PC's cost-performance orientation to mobile systems' cost-performance-per-watt orientation, which requires a memory component with the non-volatility of flash memory, the density of DRAM, and the speed of SRAM. None of the PC's memory types is adequate, either alone or in combination, so there is now enormous and growing market incentive to create a satisfactory memory type.

Within the next couple of years, we will see the emergence of *non-volatile* memory that is suitable for cell phones and other mobile devices and which will displace DRAM, SRAM, and flash memory in these applications. Non-volatile memory will be used to great advantage in programmable logic devices and these businesses, which already have a healthy growth rate, will expand even more rapidly.

Pent-up innovation

We are near the end of the IDMs' dominance in semiconductor process development. For cost-performance-per-watt-oriented systems the microprocessor isn't efficient enough and current memory chips have debilitating flaws. As a result, innovation will accelerate in chip design.

There are barriers to innovation: chips are complex, development software is expensive, fabs are costly, mask costs are high, and leading-edge processes are expensive or are unavailable. There's too much regulation, too much litigation, and too much bureaucracy. There's lack of infrastructure and lack of standards. There are dominant and controlling companies, risk-averse venture capitalists, concerns about intellectual property protection, and a weak patent system. Offsetting these barriers are today's innovation enablers, such as globalization, specialization, and industry transformation.

The ASIC (application-specific integrated circuit) market is about $30 billion. Rising costs of masks, design, and verification have encouraged encroachment of the ASIC market by makers of programmable logic, application-specific standard products, structured ASICs, and micro-

processors. These usurpers are enabled by continuing semiconductor improvements. Startups competing against established companies follow two broad approaches: top-down and bottom-up design methods.

Top-down competitors, including the microprocessor and microcontroller vendors and most of the startup companies, raise the level of abstraction as a means of lowering design cost. But doing so changes the design method, which harms acceptance by the hardware engineers in charge of the process. **AsyncArray Devices**, **Rapport**, Stretch, and Tensilica are examples of top-down competitors using microprocessor-based designs. Though changing the design method and raising the level of abstraction is a hard sell in the near term, the top-down competitors will be the long-term winners because raising the level of abstraction conserves the critical resource—the designer's time. For now, these companies are generally in the microprocessor market and are not in the ASIC market.

Configurable computing will finally make some headway. Stretch Inc. will be the first successful reconfigurable-processor company (though the company never uses the term reconfigurable).

Bottom-up competitors retain the design method and they lower costs by other means, with the advantages that their products appeal to known customers of current products and that their customers retain familiar design methods. Bottom-up competitors include companies offering ASSPs (application-specific standard product), structured ASICs, and FPGAs (field programmable gate arrays).

An ASIC typically uses a CPU (central processing unit) core and one or more DSPs (digital signal processors) surrounded by custom peripherals for special functions, as compute accelerators, and for communication. The performance of the ASIC may depend heavily on how well it can run software and on the performance of the on-chip CPU and DSP implementations.

ASSPs have the same high costs as ASICs because they are application specific. Their advantage is that costs are amortized across vendors using the same chip; their disadvantage is that vendors have limited means of differentiating products based on ASSPs. ASSPs will successfully encroach on ASIC markets. Structured ASICs reduce mask costs, but do little to reduce design and verification costs.

Structured ASICs will likely have only limited success because they do not substantially reduce the major costs (design and verification).

Sitting somewhere in between the two stools of bottom-up configurability and top-down programmability are coarse-grain structures such as FPGA-like arrays that use arithmetic logical units, CPUs, DSPs, multipliers, and other special purpose modules instead of the fine-grain programmable look-up tables (LUTs) used by ordinary FPGAs. The advantage of coarse-grain structures is that they raise the level of abstraction; their disadvantages are that they change the design method and that they are less universally applicable. Changing the design method requires new development software and it requires reeducating the design community. The range of cost-effective applications is narrowed by their coarse-grain structure, containing on-board microprocessor cores, for example.

FPGA makers

The FPGA makers are bottom-up competitors and the main threat to ASIC and ASSP makers. The principal advantage for FPGAs is that they are generic in manufacture and are customized in the field. (ASICs, structured ASICs, and ASSPs are customized at manufacture.) FPGA makers use fine-grain structures (look-up tables and multiplexers) along with personalization memory and block memory. They also use standard ASIC design methods and offer cheap development software. FPGAs directly verify the design on the chip that will be in the end product.

Altera and Xilinx together own about 85 percent of the rapidly growing market for programmable logic. These companies began by building logic-consolidation chips for engineers designing board-based systems. One general-purpose chip could be programmed to subsume the functions of miscellaneous chips that were a thorny characteristic of board designs. The miscellaneous chips matched interfaces and logic functions between the board's subsystems. Engineers at Altera and at Xilinx are hardware designers. The market for programmable logic, which is about $3 billion, is encroaching on the $30-billion ASIC market. Altera and Xilinx know the ASIC market because it and the programmable logic market employ the same engineers, design methods, and development software. They also know how to compete for ASIC business, so the programmable logic business is growing at the expense of the ASIC suppliers.

But microprocessors are also growing into the ASIC market—and from a much larger base of $40 billion. Because microprocessors are accessible to programmers, they have the advantage over programmable logic, which is primarily

LOGIC-DESIGN PRODUCT HIERARCHY

ASIC: The ultimate in density and performance; custom logic for a single customer. Representative companies are: **LSI Logic** (LSI), **Fujitsu**, and **Toshiba**, together with such in-house producers as **Sony** (SNE) and **Cisco** (CSCO).

ASSP: Excellent density and performance; custom logic for one application, often in modules or chipsets sold to several customers. Representative companies are **Texas Instruments** (TXN), **Freescale** (FSL), **Analog Devices** (ADI), **Via**, and **Infineon** (IFX). Even **Intel** (INTC) fits with its Centrino modules.

Structured ASIC: OK density and performance; fits in the large cost and performance gap between ASICs and FPGAs. Representative companies are **Actel** (ACTL), **eASIC**, **Chip Express**, and **QuickLogic** (QUIK) with its ultra low power anti-fuse process.

FPGA: Lowest in performance and density; best in convenience. Representative are **Altera** (ALTR), **Xilinx** (XLNX), and following behind, **Atmel** (ATML) and **Lattice Semiconductor** (LSCC).

accessible only to hardware designers. Programming is a higher-level abstraction than logic design, which makes the microprocessor's engineers more efficient. Further, there are ten times as many programmers as hardware designers, giving the microprocessor a double advantage in its competition with programmable logic for ASIC applications. Programmable logic companies need to transform themselves into microprocessor companies. With their soft- and hard-core microprocessors, these companies seem positioned for such a change. But it will be difficult for businesses dominated by circuit designers and logic designers to make the transition to a microprocessor orientation. Their business culture is wrong and their customer base is wrong.

The cost of culture: industry transition

I don't know how the semiconductor industry's battles will turn out, but I can give you the general rule from my experience: Battles are fought at the leading edge, but the war goes to what is good enough. The problem for corporations is that their momentum takes them *past* the point where their solution is good enough. Inertia means they have to have the will to change when their solution is *still* good enough.

During the long-running war between assembly programming and high-level languages, conference battles were fought on performance and on program space. Assembly's advocates won all the battles. They pointed to examples of dismal performance and of bloated object code produced by compilers of high-level-language programs. Assembly-language programmers could always do better by a wide margin. However, the critical resource turned out not to be performance or memory space, but programmers. High-level languages won the war by raising programmers' productivity. Moore's-law progress grew performance and memory size. What mattered was the cheapest route to a solution that was good enough.

Similarly, in the 20-year war between RISC (reduced instruction-set computer) and CISC (complex instruction-set computer), RISC won all the battles on reported performance, research support, and press attention; CISC won the war on unit volumes and on revenue. At the dawn of the personal computers, workstations, built for performance, won conference battles. Workstation makers offered high-performance microprocessors, expensive memory systems, and high-end graphics. They could always demonstrate better absolute performance for their high-end systems. PCs, built for volume and riding Moore's law, won the war by squeezing workstations into ever-diminishing market corners. Best performance at any price didn't matter as much as best performance below the sweet-spot cost.

In today's war between ASICs and FPGAs, ASICs win the conference battles on logic density, on performance, and on cost by wide margins. Some examples show ASICs to be 10 times faster, 10 times smaller, and 10 times cheaper than their FPGA counterparts. That matters for those few applications at demand's leading edge, but it doesn't matter for the bulk of the market. FPGAs will win the war by serving the bulk of the market with chips that are good enough. The critical resource isn't logic density, performance, or chip cost; it's design cost and designer productivity.

Looking 20 years into the future shouldn't be attempted in semiconductors. Five years is probably too far. This is particularly true now that we are within a few years of a major transition from business as usual in shrinking transistors. That said, here are a few suggestions:

Fabs will transition from batch processing to single-wafer manufacturing. Also, 3-D wafer stacking will emerge, making it possible to mix processes easily.

Memories will transition to something that is dense, fast, and is *non-volatile*, and something that does not resort to conventional use of electric charge or magnetic domains. That explains the enthusiasm of premier venture capitalists, such as Vinod Khosla of Kleiner Perkins and Steve Jurvetson of Draper Fisher Jurvetson for a non volatile porphyrin molecule technology from **ZettaCore** of Denver and the persistent interest in exotic memory spins at **Ovonix**, **Nantero**, and **NVEC** (NVEC).

Logic modules will connect via serial interfaces. There won't be a need for defining, designing, and verifying gigantic logic systems; functions will plug together simply through self-identifying serial interfaces. Many logic functions will carry their own sensors, their own data reduction and analysis circuitry, and their own wireless communication systems.

Engineers will adopt and adapt Nature's solutions. We will harness naturally occurring systems, using viruses, for example, to grow systems for our use. Biological systems and electronic systems will merge.

The cost of culture is the years-long wait for the industry to develop approaches that are inherently cost-performing, low-power solutions.

— Nick Tredennick and Brion Shimamoto
October 18, 2005

GILDER TECHNOLOGY REPORT PUBLISHED BY GILDER PUBLISHING, LLC AND FORBES INC. • Copyright ©2005, by Gilder Publishing, LLC

291A MAIN STREET, GREAT BARRINGTON, MA 01230, TEL: (413) 644-2100, FAX: (413) 644-2123 • EMAIL: INFO@GILDERTECH.COM

EDITOR	EDITORS	EXECUTIVE EDITOR	TECHNOLOGY ANALYST	MANAGING EDITOR	ART DIRECTION	SUBSCRIPTION DIRECTOR	CUSTOMER SERVICE
George Gilder	Nick Tredennick Brion Shimamoto	Bret Swanson	Charlie Burger	Mary Collins	design-O-matic	Rosaline Fernandes	Sandy Fleischmann

For subscription information telephone toll free: (800) 292-4380 • WWW.GILDERTECH.COM