GILDER TECHNOLOGY REPORT

December 2002 / Vol. VII No. 12

Moore's Law & The Value Transistor

"I thought Moore's law drove the industry, until Tredennick and Shimamoto introduced me to the value transistor. Here it is, in their words."

—George Gilder

Inside:

The big foundries, **Taiwan Semiconductor Manufacturing Corporation** (TSM) and **United Microelectronics Corporation** (UMC), have curtailed capital expenditures and are pushing out adoption dates for 300-mm (diameter) wafer production. The biggest integrated device manufacturers (IDMs), **IBM**, **Intel** (INTC), **Samsung**, and **Texas Instruments** (TXN), continue to spend on process development and are moving to 300-mm wafers. Conventional wisdom says that as the economy recovers, IDMs will be prepared for the upturn. And foundries, with lagging semiconductor processes that produce slower chips at higher costs, will lose market share.

In the early days, foundries lagged IDMs by at least a couple of process generations. Then they caught up with the IDMs—introducing large wafers and leading-edge processes right along with the major IDMs. Now, foundries seem about to fall behind. What is going on?

Moore's law is the rate of semiconductor manufacturing improvement—the number of transistors in a fixed area doubles every eighteen months. Big chips are more capable; fixed-size functions fit on smaller chips. The magic of Moore's-law progress comes from three areas. Most important is shrinking transistor size. The second contributor is increasing chip and wafer size. Third is better circuit design. Chips get faster and cheaper with manufacturing process improvements. If you find the current chips lacking, wait a generation or two and they'll have what you need. If you find the current chips capable, but they're too expensive, wait a generation or two and they'll be cheap enough. That's the way it's been for over thirty years.

But there's an interesting chart at TSMC's web site (www.tsmc.com/english/technology/t0203.htm) that isn't easy to explain. Fig. 1 is a version of TSMC's chart with its interesting features. The unit of measure for semiconductor manufacturers is "wafer starts." Chip size and transistor size vary with product and process; wafer size stays constant for years, so production capacity is measured by the number of wafers the plant processes per month.

Fig. 1 shows the percent of the foundry's wafer starts, by semiconductor process, plotted against time. Before the 1996 introduction of the 350-nm process, 100% of the foundry's wafers were at 500 nm or larger. By the beginning of 1997, more than

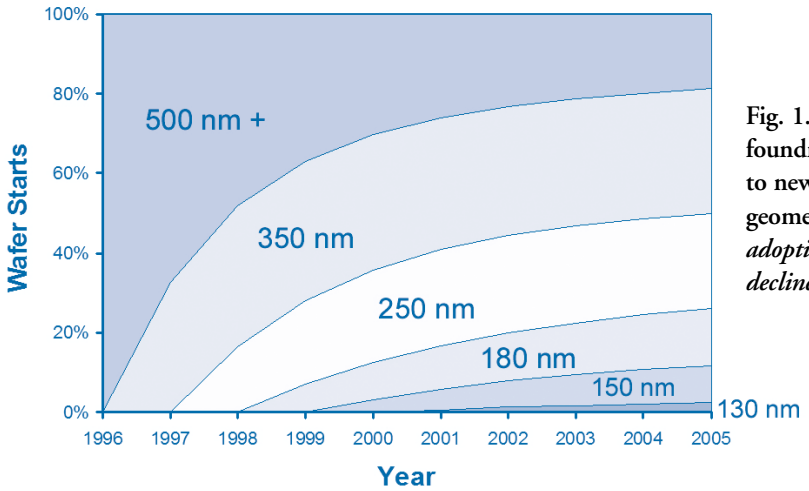


Fig. 1. As the foundry moves to new process geometries, *adoption rates decline.*

30% of wafer starts were at 350 nm. In 1997, the foundry offered a 250-nm process, but by the end of the year, fewer than 20% of its wafer starts were at 250 nm. This is an interesting trend: with each new generation of semiconductor process, the adoption rate is falling.

Moore's law says there's incentive to move to a more advanced process. Chips get smaller and faster, they use less power, and they are cheaper. Let's say you are making chips in a 500-nm process and have an opportunity to move to a 350-nm process. We've talked about the speed and power advantages, so what about cost? To a first-order estimate, the same wafer size should hold twice as many chips in a 350-nm process as it does in the 500-nm process. It costs about $500 to process a 200-mm wafer. This cost, the cost to *operate* the wafer-processing equipment, is independent of whether the wafer's patterns are at 350 nm or at 500 nm. If you sell a wafer's worth of 500-nm chips for $1,000, then $500 is profit. If you sell them at the same price, a wafer's worth of 350-nm chips will fetch $2,000, with $1,500 of profit. You might charge more for the faster, lower-power 350-nm chips, making margins even higher.

Smaller chips pack better on the wafer. The chip edges have to line up in both directions so the processed wafer can be cut into chips. While a 50% shrink in each dimension should allow 4 times as many chips, the actual number is more like 4.5 because the smaller chips pack better on the round wafer.

There's more incentive than just how many chips *fit* on the wafer. Defects reduce the wafer's yield of good chips. Yield is the percent of good chips per wafer. Fig. 2 illustrates the effect of ten random defects on the wafer's yield, for two chip sizes. For this example, moving from a half-inch chip to a quarter-inch chip *improves the yield* from 75% to 94%. Therefore, instead of 4.5 times as many chips, the same-size wafer yields almost 6 times as many *good chips*. Financial incentives for moving to smaller geometries seem compelling.

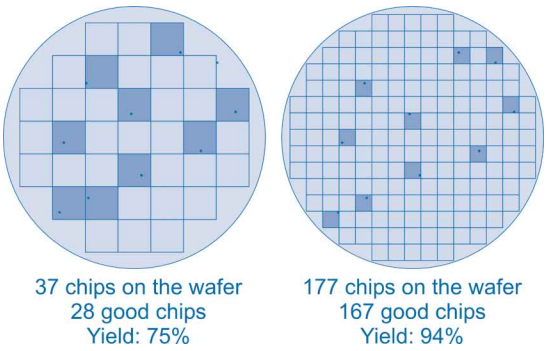


Fig. 2. The yield can be substantially higher for smaller chips. Halving the chip dimensions yields six times as many good chips.

There's an advantage when a foundry moves from 200-mm diameter wafers to 300-mm diameter wafers. Just based on area gain, we expect 2.25 times as many chips on the larger wafer. The number will be slightly higher because chips pack better on the larger wafer. Intel, for example, gets 201 of its 134-square-mm Pentium 4s on a 200-mm wafer and 482 Pentium 4s on a 300-mm wafer. It's about 2.4 times the number of chips, but the cost to process a 300-mm wafer should be only 20% higher than the cost to process a 200-mm wafer.

It seems financially compelling to move to smaller processing geometries and to move from 200-mm wafers to 300-mm wafers. But, as fig. 1 shows, at least for the foundries, the move isn't happening.

Wafer starts for old semiconductor processes don't fade to nothing. Instead, they decline and then stabilize for years as a percent of the foundry's wafer starts. Why might this be so?

The answer to "why chip designs don't move to new processes" is threefold: start-up costs, hidden costs, and physical limits.

Start-up costs

Theoretically, it costs about the same to process a 200-mm wafer whether the lines on the wafer are 180-nm wide or are 130-nm wide. This cost (wafer processing) is a variable cost. Practically, however, the foundry's 180-nm plant will be three years old, while its $2.0-billion 130-nm plant will be new. This $2.0 billion is the fixed cost of the building and the cost of the processing equipment with $500 to $600 million in process development cost (the cost to develop design rules for the 130-nm process). Wafers running the 130-nm process in the new plant have to pay their share of fixed costs. The fixed costs of the old process have been amortized over three years of production. The foundry can charge less for work in the depreciated 180-nm plant.

Hidden costs

You can't just call the foundry and say: "Move the Umptyfritz Controller production from the 350-nm plant to the 180-nm plant." Your engineering teams developed the controller for a specific 350-nm process. If you want to produce it at 180 nm, your engineering teams have work to do. They have to build the Umptyfritz Controller for the 180-nm process. For most engineering managers, the choice comes down to allocation of precious engineering talent. Do you want your engineers cost-reducing an old product or do you want them working on the next-generation product?

If your engineers cost-reduce the product, one cost will be a new mask set. Mask sets are expensive and they are getting even more expensive. As a rule of thumb, each new process generation more than doubles the cost of the mask set. For a 130-nm process, mask costs can be $600,000. At 90 nm, masks could cost $1.5 million. By 2010, a mask set could cost $10 million! If the chip doesn't work, you'll have to buy more than one mask set. Mask costs have to be amortized over the production run. If the chip you build goes into a system with expected lifetime sales of 100,000 units, then the $1 million mask cost adds $10 to the cost of each chip. Amortized engineering cost adds more. Continued production of the old chip may be cheaper.

The chip design tools the engineers used for the 350-nm

process don't work for the 180-nm process. The engineers need new tools.

Four hidden costs in moving a chip from an old process to a new one are time, engineers, masks, and design tools. Moving from an old process to a new one will get simpler as chip design descriptions get "softer," but it still costs time, talent, and money. (Soft descriptions are parameterized recipes that are independent of the manufacturing process.)

Physical limits

If the chip doesn't get smaller, there's no cost advantage in moving to a more-advanced process. Making the circuits smaller may not make the *chip* smaller. This is because the chip's connections to the outside world are through wires attached to "bonding pads" on the chip (fig. 3). Bonding pads can't shrink below the area that an automated bonding machine can hit.

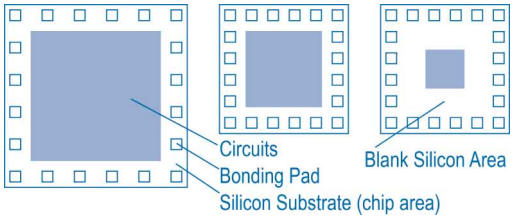


Fig. 3. Pad limited: as the semiconductor process improves, the chip and the circuits shrink, and the bonding pads get closer together (middle chip) until the bonding pads limit the size of the chip (right chip).

I've said that Moore's-law progress has left a huge wake of enabled but unexploited applications. The persistent residual percentages of wafer starts in old processes are evidence that it's true. The *total* number of wafer starts grows each year. If the 350-nm process is still 20% of wafer starts six years after its introduction, then the demand for the 350-nm process is growing at the rate that all wafer starts are growing.

The minimum unit that a foundry processes is a "boat" of twenty-five wafers. The lot size, or number of chips in a boatload, for the big chips on 200-mm wafers in fig. 2 is about 1,000. Shrinking the geometry by half brings the lot size to about 5,000. Moving to 300-mm wafers and shrinking the geometry by half again increases the lot size to about *50,000*. If you don't need chips in 50,000-unit quantities, you may not need an advanced process and 300-mm wafers.

Also, if the bill of materials for your system is $1,500, cost-reducing a two-dollar microcontroller won't be your top priority.

I can go even further in comparing the theory of Moore's law with the economic realities of semiconductor fabrication.

Most of today's semiconductor production is with 200-mm-diameter wafers, so I start with that assumption. I'll assume we're building a 25-million-transistor chip (about half the complexity of a leading-edge Pentium). How much will this chip cost? What will Moore's law do for the chip's cost as the semiconductor process advances?

More advanced semiconductor processes make smaller transistors. I'll build imaginary plants for six transistor sizes: 500 nm, 350 nm, 250 nm, 180 nm, 130 nm, and 90 nm. Today's leading-edge process is 130 nm, with leading suppliers beginning their move to 90 nm. The cost of making chips divides into fixed costs and variable costs. Fixed costs build the plant and furnish it with equipment. There are also fixed costs for developing the semiconductor process, for the software tools to build the design, and for the mask set that represents the chip. (The difference between the semiconductor equipment and the process is like the difference between the equipment and the detailed recipes-and-procedures in an industrial kitchen. Masks are like photographic negatives, defining a chip's details in a layer-by-layer buildup.) Running a wafer through the plant is a variable cost, about $500 a wafer.

I assume each plant processes 25,000 wafers a month (this is a median figure) and that it amortizes fixed costs for the plant, for the equipment, and for process development over the first three years of production. If the plant doesn't run at capacity (a frequent occurrence lately), the amortization period will likely exceed three years. Amortizing costs over four years doesn't change the qualitative values, but makes differences more difficult to see in small figures. The fixed cost for a mask set is amortized over the number of chips produced using that mask set.

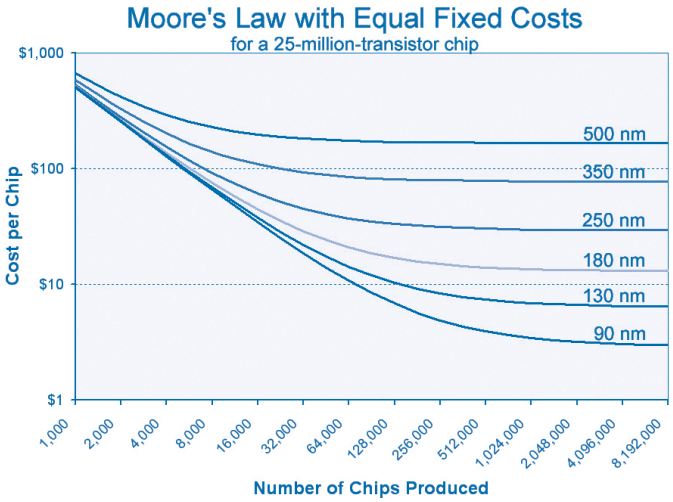


Fig. 4. If fixed costs were the same, shrinking transistors from 350 nm to 130 nm would drop the cost of a 25-million-transistor chip from $78 to $7.

Fig. 4 shows how the cost per chip varies by semiconductor process and by the number of chips built for a 25-million-transistor chip.

The cost per chip is high for building a few thousand chips because the fixed costs (plant, equipment, process development, and mask set) dominate. At low volumes, there's little cost advantage in making smaller transistors. Variable costs dominate for production runs of millions of chips.

Plants, masks, and equipment

Fig. 4 is Moore's law in theory. It assumes the *same* fixed costs across all semiconductor processes. That's not the situation in the real world. Smaller transistors are harder to

TELECOSM TECHNOLOGIES

Ciena (CIEN)

METRO WDM PLATFORMS

DECEMBER 23: 5.56 52-WEEK RANGE 2.41-17.30 MARKET CAP: 2.4B

UPSIDE SURPRISE—Ciena reported a strong quarter with revenues of $62 million versus expectations of $52 million. Cash burn came in at $100 million, $50 million less than expected, and the company exited the quarter with $1.2 billion in net cash.

Rising Requests: A number of industry analysts are noting an increase of activity in the optical-switching market, of which Ciena boasts a 50% market share. It has been reported that RFPs (request for proposal) are underway at Verizon, Sprint, British Telecom, Deutsche Telecom, Telecom Italia, France Telecom, and Telefonica.

Essex (ESEX.OB)

OPTICAL PROCESSORS

DECEMBER 23: 2.85 52-WEEK RANGE 1.50-8.25 MARKET CAP: 22M

OPTICAL ORDNANCE—Essex has been awarded two $70,000 contracts from the Department of Defense, Missile Defense Agency. One contract will focus on enhancing the performance of missile defense radars while the other will look to improve distribution and throughput of radar data.

Flowering Finances: Reinforcing our commentary from last month, Essex announced that its revenues had doubled year-over-year coming in at $4.5 million, and that it expected to double revenues again in 2003. The company also announced $500,000 of additional funding to be used toward core business expansion.

Corvis (CORV)

WDM SYSTEMS, RAMAN AMPLIFICATION, EDGE SWITCHES

DECEMBER 23: 0.71 52-WEEK RANGE 0.47-3.44 MARKET CAP: 293M

JDS Uniphase (JDSU)

ACTIVE AND PASSIVE OPTICAL COMPONENTS

DECEMBER 23: 2.54 52-WEEK RANGE 1.58-10.34 MARKET CAP: 4.7B

Avanex (AVNX)

ADAPTIVE PHOTONIC PROCESSORS

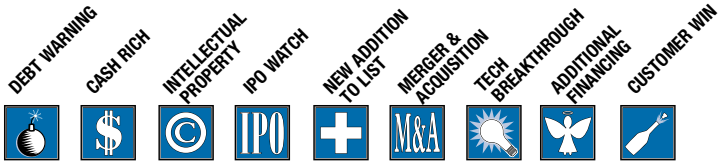
DECEMBER 23: 1.11 52-WEEK RANGE 0.63-7.20 MARKET CAP: 77M

StorageNetworks (STOR)

DATA STORAGE MANAGEMENT, SOFTWARE

DECEMBER 23: 1.04 52-WEEK RANGE 0.82-7.05 MARKET CAP: 103M

KEY



Equinix (EQIX)

SECURE INTERNET BUSINESS EXCHANGES

DECEMBER 23: 0.25 52-WEEK RANGE 0.19-3.53 MARKET CAP: 25M

The Nasdaq ruled in favor of extending Equinix's national market listing, contingent upon the company's ability to demonstrate compliance with all of the requirements for initial listing. Equinix has satisfied all such requirements except for the minimum bid price. Equinix intends to effect a reverse stock split in an amount to be determined.

Sprint PCS (PCS)

NATIONWIDE CDMA WIRELESS NETWORK

DECEMBER 23: 4.63 52-WEEK RANGE 1.75-5.20 MARKET CAP: 4.6B

PRIME PROGRESS—Sprint PCS exited its Analyst Day with the audience optimistic, after it set forth a set of realistic goals for 2003. Following 3Q02's embarrassing results, the company remains focused on improving the quality of its subscriber base. Thus far "prime" customers have made up two-thirds of 4Q02 gross additions, and this figure has risen to 70% in the last few weeks compared to 50% last quarter.

Real World Results: Actual results to date indicate that Sprint PCS's 1x network has achieved voice capacity gains of 75-85%. According to CTO Oliver Valente, delivering the same capacity with 2G infrastructure would have cost more than $100 million.

Qualcomm (QCOM)

CDMA INTEGRATED CIRCUITS, IP, SOFTWARE

DECEMBER 23: 38.77 52-WEEK RANGE 23.21-53.34 MARKET CAP: 30B

DEJA VU—Qualcomm raised its chipset target mid-period for the second consecutive quarter; the company is now forecasting F1Q03 shipments to exceed 28 million units. Even more dramatic, given historic seasonal weakness, is the strength in order rates pushing F2Q03 projections 20-35% higher into the neighborhood of 24-27 million chipsets. The Street attributes this strength largely to China, India, and Japan while warning of a disappointing North America. Electronic retailers, however, report strong interest in next-generation wireless devices, specifically Amazon.com which noted that 60% of the new phones sold through its website were color-screened devices.

Asian Update: Having surpassed the 4-million-subscriber mark on its 1x network, Japan's KDDI is confident it will meet its one-year goal of 7 million subscribers by March 31, 2003. Also, the once improbable target of 7 million subscribers is in view for China Unicom whose CDMA-subscribers now number 6.3 million. Positive news sprang forth from India as well, when the Indian Supreme Court ruled in favor of fixed-line operators Reliance and Tata Teleservices allowing these companies to proceed with their plans for building CDMA-based wireless local-loop (WLL) networks.

Altera (ALTR)

PROGRAMMABLE LOGIC DEVICES

DECEMBER 23: 12.48 52-WEEK RANGE 8.32-26.18 MARKET CAP: 4.8B

—Altera's mid-quarter update focused on the continued strength in the Stratix product rollout. Altera's 3Q conference call noted that three of the eight family members were shipping to 150 customers and that expectations were for three additional members to roll out during 4Q. The company confirmed this achievement and noted that the six members of the Stratix family are now shipping to 250 customers.

MEAD'S ANALOG REVOLUTION

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

COMPANIES TO WATCH

ANALOG DEVICES (ADI)
ATHEROS
BLUEARC
COX (COX)
ENDWAVE (ENWV)
MIRROR IMAGE
POWERWAVE (PWAV)
SAMSUNG
SCALE EIGHT
XILINX (XLNX)

Broadcom (BRCM)

BROADBAND INTEGRATED CIRCUITS

DECEMBER 23: 16.22 52-WEEK RANGE 9.52-53.35 MARKET CAP: 4.5B

MULTIMODE MOMEMTUM—Broadcom was first-to-market with an 802.11g, Wi-Fi chipset, and is believed to be aggressively pricing its solution in typical Broadcom fashion. More importantly, Broadcom is leading the charge to multimode—802.11a/g (5GHz/2.4GHz)—solutions, which should begin entering the market around the first of the year. Joining Broadcom in the multimode push are Atheros and Texas Instruments; all three appear to be leaving 802.11b leaders Intersil and Agere behind. Finally, Intel's release of its Banias chipset will solidify 802.11a in the marketplace.

Broadcom also announced approval of its baseband processor used by Sony Ericsson in a tri-band PC Card Modem. This approval should pave the way for further adoption of Broadcom's reference platform, which it acquired from Mobilink.

Terayon (TERN)

BROADBAND CABLE MODEMS, HEAD-ENDS

DECEMBER 23: 2.36 52-WEEK RANGE 0.86-9.35 MARKET CAP: 173M

CMTS CERTIFIED—The prolonged cap-ex downturn had begun to whittle away Terayon's DOCSIS 2.0 head start. However, the announced DOCSIS 2.0 certification of its CMTS (head end) and cable modem gives Terayon the industry's only complete end-to-end DOCSIS 2.0 cable data system.

J-COM Broadband, Japan's largest cable television operator, is currently running a Voice-over-IP cable telephony trial using Terayon's embedded MTA (multimedia terminal adapter).

EZchip (LNOP)

10 GIGABIT NETWORK PROCESSORS

DECEMBER 23: 6.33 52-WEEK RANGE 3.79-16.45 MARKET CAP: 46M

The entire point of a network processor is to create a platform for software so that every new network application, protocol, and upgrade does not entail new silicon. EZchip has achieved this goal through the integration of software from Level 7 Systems and can now offer the industry's first completely integrated 10-Gigabit networking hardware and software solution.

EZchip announced that it had executed a term sheet for its Series C financing totaling up to $21.5 million. LanOptics will participate with a commitment of $2.3 million with the option to invest up to another $4 million. Following the new round of financing, Lanoptics will own approximately 51% of EZchip.

Synaptics (SYNA)

TOUCH-SENSORS, FOVEON IMAGERS

DECEMBER 23: 7.47 52-WEEK RANGE 3.13-20.75 MARKET CAP: 175M

Similar to the custom-solution delivered to IBM for its ThinkPad notebooks, Synaptics unveiled an integrated dual-pointing module combining both a TouchStyk and TouchPad into a single-interface solution. This new integrated solution differs from traditional dual-pointing devices in that it joins both devices into a single module, creating a more cost-effective and easier-to-integrate solution.

National Semiconductor (NSM)

SINGLE-CHIP SYSTEMS, ANALOG EXPERTISE, FOVEON IMAGERS

DECEMBER 23: 15.95 52-WEEK RANGE 9.95-37.30 MARKET CAP: 2.9B

EMPOWERING—With Foveon as its future, National's current success in the power management space has been prodigious. Bookings for portable power management products grew more than 200% year-over-year in 2Q03. National entered the WLAN space in a huge way achieving placement of its power management solutions in Broadcom's multimode 802.11 reference platform as well as in Intel's Calexico 802.11 cards. Equally impressive is National's accruing more than ten design wins associated with Intel's Banias mobile processor, which looks to enter the market during 1H03. COO Donny Macleod noted that National had no previous power management presence in notebook PCs.

Intel (INTC)

MICROPROCESSORS, SINGLE-CHIP SYSTEMS

DECEMBER 23: 17.31 52-WEEK RANGE 12.95-36.78 MARKET CAP: 115B

HANDSET HANKERING—Intel is currently shipping prototype samples of the new Manitoba "Internet-on-a-chip." Intel's integration of the application processor, baseband and analog front-end, Flash, and SRAM-memory in a single chip will place the company in direct competition with Texas Instruments' OMAP platform and further its goal of moving the handset industry structure toward the x86 instruction set.

Wireless Wampum: Intel awarded two more private companies with an unspecified chunk of its $150 million targeted at the wireless space. Utah-based SSTN focuses on hotel and conference center connectivity while Washington-based TeleSym develops telephony software for use on mobile PCs and PDAs.

Texas Instruments (TXN)

DIGITAL, ANALOG, MIXED-SIGNAL PROCESSORS

DECEMBER 23: 15.99 52-WEEK RANGE 13.10-35.94 MARKET CAP: 28B

Texas Instruments raised its guidance for the fourth quarter based upon continued strength in sales of components into wireless handsets. Now reaching 50% of sales, TI's benefits are accretive as users transition to GPRS handsets with their higher ASPs.

Narad Networks

GIGABIT ETHERNET COAXIAL CABLE NETWORKS

PRIVATE

Disruptive technology, combined with the company's strategic alliance with IBM, won Narad placement on the Communications Engineering & Design Broadband 50.

Soma Networks

BROADBAND WIRELESS ACCESS, NETWORK SOFTWARE

PRIVATE

make. The plant and equipment to build finer geometries are more expensive. Developing processes for smaller transistors is more expensive. Mask sets for smaller transistors are more expensive. The cost of the plant and equipment approximately doubles with each new process generation. Process development escalates more slowly and mask cost escalates a little faster, but all the fixed costs rise as transistors get smaller. Fig. 5 factors in escalating costs for the plant, for the equipment, for process development, and for the mask set.

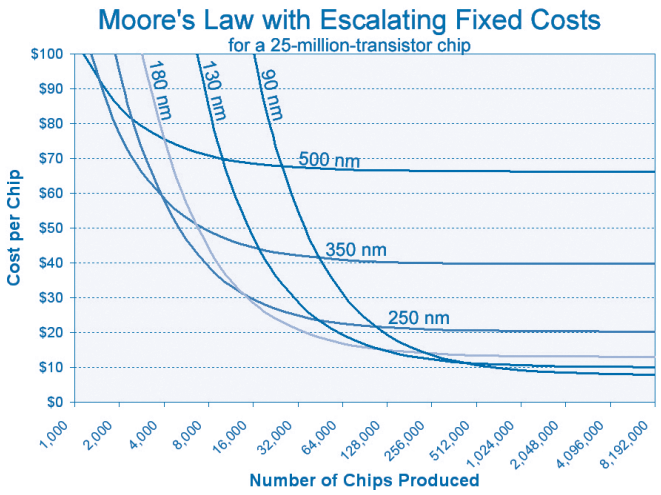


Fig. 5. Smaller transistors are cheaper than large transistors, but the advantages of shrinking shrinks with each generation.

In fig. 4, the curves don't cross; in fig. 5, they do. If fixed costs are all the same (fig. 4), it's always cheaper to build smaller transistors. If fixed costs rise as transistors get smaller (fig. 5), then for production runs of a few thousand chips (where fixed costs dominate chip cost) it's cheaper to build larger transistors. Now that we have accounted for escalating fixed costs, two important characteristics of these costs appear in fig. 5.

First, the cost-crossover point between generations is *moving to the right*. For a 25-million-transistor chip, 350-nm transistors become cheaper than 500-nm transistors at 1,500 chips, but 90-nm transistors don't become cheaper than 130-nm transistors until production runs exceed 500,000. For larger chips, crossover points move toward smaller production runs; for smaller chips, crossover points move toward larger production runs. Crossover between 90-nm transistors and 130-nm transistors for a 2-million-transistor chip requires production runs of millions of chips. Advanced processes want big chips *and* large production runs. Moving to larger wafers, from 200 mm to 300 mm, pushes crossover points toward larger production runs.

Second, the chip's *cost advantage decreases as transistors get smaller*. For large production runs, moving from 350 nm to 250 nm reduces chip cost from $40 to $20—a saving of $20 or 50%. For a production run of eight million chips, moving from 130 nm to 90 nm reduces chip cost from $10 to $8—a saving of $2 or 20%.

Escalating fixed costs push the economic crossover point to the right. If you are designing a 25-million-transistor custom chip as the brains of an espresso machine, the chip will be cheaper in a 130-nm process than in a 90-nm process unless you expect to ship more than 500,000 machines. If the custom brain is fewer than 25 million transistors, you will need to sell even more machines.

Escalating fixed costs also mean that, for a given number of transistors, the incentive to move to smaller transistors diminishes with each generation. The incentive diminishes both in absolute dollars per chip and as a percent of the cost of the chip. If you are already shipping the espresso machine with a custom chip designed in a 180-nm process, each chip is costing you $13. Your engineers could cost-reduce the design by moving the brain-chip to a 130-nm or to a 90-nm process. You would save about $2 per chip. If the bill of materials for the espresso machine totals $1,000, a $2 savings isn't much incentive. It would also cost a few engineers and several months. You would have to sell another 500,000 machines just to break even on the move to a new process.

Old processes, new processes

Fig. 6 shows what happens to chip cost with the depreciation of fixed assets. In fig. 6, I assume the three leading-edge processes (180 nm, 130 nm, and 90 nm) are being built in plants that are not fully depreciated. The cost of leading-edge chips, therefore, includes the amortization of the plant and its equipment across 25,000 wafers a month over three years of operation. The three trailing-edge processes (500 nm, 350 nm, and 250 nm) are being built in plants that are fully depreciated. The cost of trailing-edge chips, therefore, does not include dollars for plant and equipment amortization.

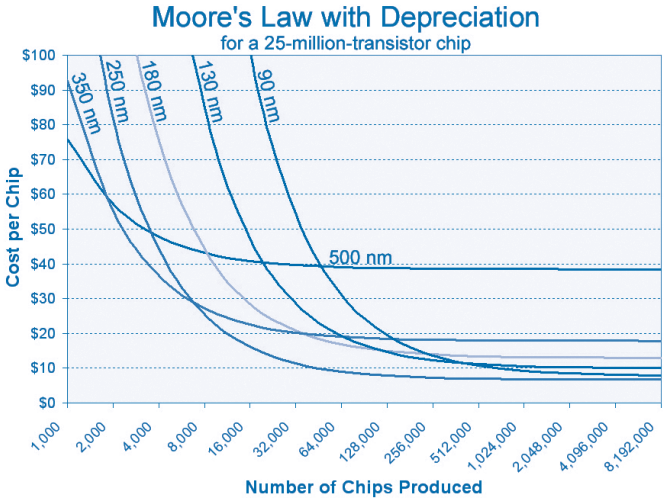


Fig. 6. Here, foundries for 180-nm, 130-nm, and 90-nm chips still amortize fixed costs into the cost of a 25-million-transistor chip. Fully depreciated foundries for 500-nm, 350-nm, and 250-nm offer lower-cost chips.

I've added another dose of reality; this time the result is even more surprising. The cost to build a 25-million-transistor chip in a newer process (180 nm, 130 nm, or 90 nm) never drops below the cost to build the same chip in a 250-

nm process! A fully depreciated 250-nm process turns out the cheapest 25-million-transistor chips. As long as cost is more important than performance, the big transistors on the old process beat the small transistors on newer processes. Once the 180-nm foundry is fully depreciated, it will turn out the cheapest 50-million-transistor chips (Moore's law).

Applications span a performance gamut from hair dryers, washing machines, and blenders to computers, video games, cell phones, and set-top boxes. The electronics market consumes billions of microprocessors each year. Most of these are four- and eight-bit microcontrollers. These applications are cost-oriented and are not performance-oriented. One or two million transistors make a very capable microcontroller. Twenty-five or fifty million transistors are more than enough for a *huge* range of applications.

Leading-edge applications pay a premium for performance and foot the bill for new processes. But with each new process generation, the range of applications available to pay premium prices shrinks.

Small chips

Fig. 6 showed cost curves for leading-edge processes and for trailing-edge processes for a 25-million-transistor chip. Fig. 7 shows the curves for a 2-million-transistor chip. The smaller chip changes the scale; small chips are less than a tenth of the cost of large chips, but the conclusions remain the same. The fully depreciated 250-nm process builds the cheapest small chip. The 2-million-transistor chip's equal-cost crossover points have moved to the right by about ten times the number of chips at the 25-million-transistor chip's crossover point. For all reasonable production volumes above 100,000, the 250-nm process makes the cheapest chips. For production volumes below 100,000 units, it would be cheaper to produce the 2-million-transistor chip

THE VALUE TRANSISTOR

What's a "value" transistor? A value transistor is the right transistor for the job.

The value transistor is a recent phenomenon. We're used to thinking that Moore's law makes the transistor better with each process generation. This is no longer true for every application. Each process generation is now creating what I call the value transistor—one that is good enough to completely satisfy a set of applications. This is analogous to what happened with the personal computer—"value PCs" now satisfy a growing segment of users. When I started thinking about how to describe the value transistor, I thought I would hit upon the perfect description to convince you that there is such a thing. I built a mathematical model and I tried ways to demonstrate its existence. It turns out that it's complicated, so there's no one description that illustrates the value transistor for all cases. One way to identify a value transistor looks at power use in chips for mobile applications.

The most common transistor—the CMOS field-effect transistor—dissipates energy in two ways: as *active power* and as *leakage power*.

Active power switches the transistor on or off. Active power depends on the size of the transistor and on its frequency of operation. Transistors burn active power while doing arithmetic or while making control decisions. Big transistors use more power (the bigger area of a large transistor requires more electrical charges). Switching the transistor faster pumps more charges from the power supply to the transistor and then to ground.

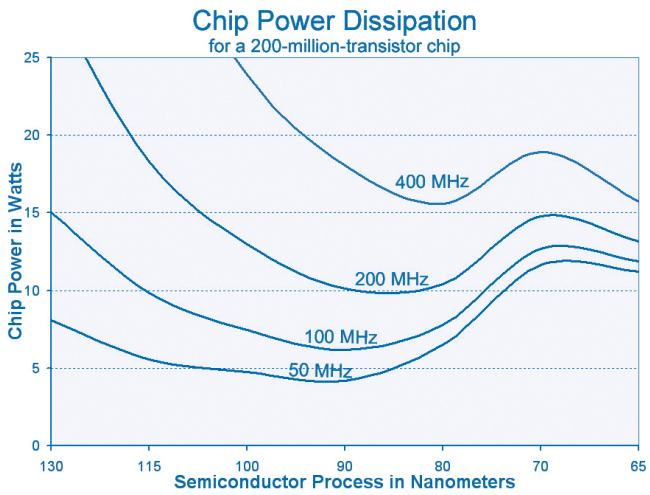


Fig. 8. Power for a 200-million-transistor chip at several frequencies. At 50 MHz, the value transistor occurs in a 90-nm process. At 400 MHz, the value transistor occurs in an 80-nm process.

Leakage power is the power dissipated as electrical current leaks when the transistor is not being switched. Big transistors don't leak much. Small transistors, with their thin conductors and thin insulation, leak more.

As transistors get smaller, the active power decreases and the leakage power increases. This means that for some applications, there is a *value transistor*: make it larger and its active power is too high; make it smaller and it leaks too much. Fig. 8 shows curves for a 200-million-transistor chip operating at frequencies from 50 MHz to 400 MHz. Note how specific the fig. 8 example is. It is for a particular number of transistors operating at a particular range of frequencies.

The percentage of transistors that are active (being switched on or off), as opposed to idle (not being switched), heavily influences the shape of the curves.

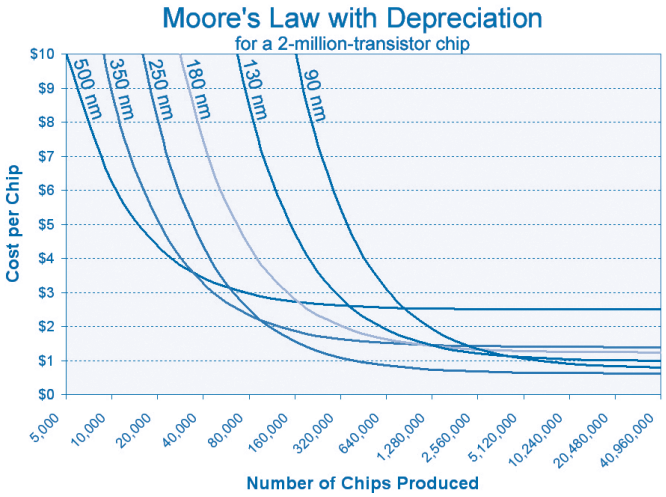


Fig. 7. Foundries for 180-nm, 130-nm, and 90-nm chips amortize fixed costs into the cost of a 2-million-transistor chip. Fully depreciated foundries for 500-nm, 350-nm, and 250-nm offer lower-cost chips.

on an even older process with bigger transistors.

Each process generation doubles the number of transistors on a pad-limited chip, so each generation takes another bite out of the application space. The bite that's taken out is for cost-oriented chips that will never need smaller transistors.

Lessons

Escalating fixed costs have two important consequences. First, a chip's equal-cost crossover point between an old process and a new one is moving toward *much higher* production runs. Unless you are building millions of systems, your chips will be cheaper in the *old* process. Second, the advantage of moving from an old process to a new process decreases in both absolute dollars and as a percentage of the cost for each new process generation. In older processes, a redesign might have saved 50%, while future processes promise only 10%.

Because its fixed costs have been amortized, a 250-nm process builds 25-million-transistor chips that are cheaper than 25-million-transistor chips built at 180 nm or at 130 nm. Once its fixed costs are amortized, the 180-nm process will build the cheapest 50-million-transistor chips (Moore's law). Any application that isn't performance-limited in 180 nm and doesn't need more than 50-million transistors (a *lot* of applications) will be cheaper in the fully amortized 180-nm process than in any process with smaller transistors. That's a huge chunk of the application space that is doubling with each process generation.

Advocates with leading-edge performance and capability requirements make the case for following Moore's law to the next smaller, faster transistor. Leading-edge applications need the performance or capability and are willing to pay a premium to get it, but these applications aren't the bulk of the market. Further, with each process generation fewer applications remain to pay the next generation's escalating costs. Intel leads the charge, betting that demand will escalate with fab cost.

The PC dominated the market for twenty years, but requirements are changing. As the PC market moves from performance to value, engineering emphasis will move to untethered applications. Untethered applications change the transistor's requirements from absolute performance to a balance of cost, power conservation, and performance. For untethered applications the right transistor for the job, the value transistor, may not come from a leading-edge process. Big transistors burn more active power; small transistors leak more.

Escalating costs have to be amortized over a fixed interval, which is causing a problem. All the while it's leaving a wake of processes that are good enough (their applications will never migrate).

Advances can change the rules, right? Maskless techniques, such as e-beam lithography, could drop mask cost to zero. Double- and triple-gate transistors can substantially reduce leakage power, which alters the position of the value transistor. Silicon-on-insulator, strained silicon, and exotic materials may change tradeoff points. The move from 200-mm wafers to 300-mm wafers affects the economics of what process is appropriate and what production runs are cost effective. So, yes, advances can change the rules. But advances take time, and the cost to change from an old process to a new process increases with time while the payoff decreases. So, for a large segment of applications, the transistors have become *good enough*.

Moore's law is an enabler, not a driver. Since the transistors are now good enough for a wide range of applications, the semiconductor industry can be healthy even if equipment makers can't sell more advanced equipment. The emergence of the value transistor invalidates "leading indicators" that assume transistors must get smaller to get better and, therefore, require the newest equipment to sustain industry growth. Up to now, the transistor wasn't good enough, and all applications shared the cost of process advances; in the future, huge segments of the market will reach their value transistor and will no longer share the cost of process advances.

—Nick Tredennick & Brion Shimamoto, December 20, 2002

GILDER TECHNOLOGY REPORT PUBLISHED BY GILDER PUBLISHING, LLC AND FORBES INC. • Copyright ©2002, by Gilder Publishing, LLC

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

EDITOR: George Gilder
EXECUTIVE EDITOR: Bret Swanson
ANALYSTS: Mary Gorski & John Hammill
MANAGING EDITOR: Marie Lavinio
ART DIRECTOR: Charles Bork
SUBSCRIPTION DIRECTOR: Rosaline Fernandes

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