Moore’s Law at 50

The Performance and Prospects of the Exponential Economy

BRET SWANSON

NOVEMBER 2015
Executive Summary

Fifty years ago, semiconductor pioneer Gordon Moore offered a theory of exponential improvement for electronics. Integrated circuits, he estimated, could get smaller, faster, and cheaper at an astounding rate—doubling in cost performance every one to two years. “Moore’s Law” has proven more or less true ever since. Today, however, many think Moore’s Law has already run out of steam. Others argue that information technology has become even more efficient and thus dangerously powerful.

The bulk of the evidence suggests information technology has delivered both technically and economically: it has achieved the promise of Moore’s Law in both its narrowest sense of transistor scaling and its broadest effect of widespread economic uplift. The effects of Moore’s Law are especially evident in information technology (IT), which has provided nearly all the productivity growth in the US economy over the last 40 years. IT has also powered a wave of globalization that has brought two billion people around the world out of material poverty. In short, Moore’s Law has been both a chief driver and an exemplar of the American experiment.

Although Moore’s Law may not continue to scale using the conventional metrics, such as transistor counts, a variety of innovations in materials, devices, state variables, and parallel architectures will likely combine to deliver continued exponential growth in computation, storage, and communications. Far from reaching an end point, information technology will help transform a number of lagging industries, including health care and education.

This paper examines the technical achievements and economic impact of Moore’s Law; the public policies that encouraged rapid innovation in information technology and the way IT, in turn, helped foster greater freedom around the world; and the technical future of Moore’s Law and the economic potential of IT.
Moore’s Law at 50:  
The Performance and Prospects of the Exponential Economy

It was a tangible thing about belief in the future.
—Carver Mead

In 1971, Intel’s first microprocessor, the 4004, contained 2,300 transistors, the tiny digital switches that are the chief building blocks of the Information Age. Between 1971 and 1986, Intel sold around one million 4004 chips. Today, an Apple A8 chip, the brains of the iPhone 6 and iPad Air, contains two billion transistors. In just the last three months of 2014, Apple sold 96 million iPhones and iPads. In all, in the fourth quarter of last year, Apple alone put around 30 quintillion (30,000,000,000,000,000,000) transistors into the hands of people around the world.\(^1\)

This rapid scaling of information technology—often referred to as Moore’s Law—is the foundation of the digital economy, the Internet, and a revolution in distributed knowledge sweeping the globe. It is also a central factor of US economic, cultural, and military power.

Moore’s Law began rather humbly, however. “I was concerned that people thought integrated circuits were very expensive,” Gordon Moore recalled, thinking back to 1965. “They were up until that time. The military was the only one that could afford them. There were all kinds of arguments about why this was an expensive technology. But in the laboratory, things were really starting to change. And I wanted to get the idea across that this was going to be the way to make inexpensive electronics. . . . So that was the motivation behind the article.”\(^2\)

The article, “Cramming More Components onto Integrated Circuits,” appeared in the 35th anniversary issue of *Electronics* in April 1965.\(^3\) “Integrated circuits,” Moore wrote, “will lead to such wonders as home computers—or at least terminals connected to a central computer—automatic controls for automobiles, and personal portable communications equipment.” Moore was a physical chemist working with a brilliant team of young scientists and engineers at Fairchild
Semiconductor, one of Silicon Valley’s first semiconductor firms.\(^4\) A few years later, he and several members of the team would leave to found Intel.

“I just looked at the data we had for the first few years,” remembered Moore. “Starting with one transistor in ’59. Then in ’61, the first integrated circuits. I could see we’d been about doubling every year. The next one coming [later in 1965] had about 60 components on it. So I just took that doubling every year and extrapolated from 60 to 60,000 components for the next 10 years . . . I never had an idea that it was going to be at all precise.” (See figure 1.)

Moore’s projection that the trend would continue for 10 years was, in some ways, conservative. In 1970, an applied physicist and Moore collaborator named Carver Mead figured out that transistors could continue to get smaller—and thus faster, cooler, more reliable, and cheaper.\(^5\) He found that transistor densities—the number of digital switches per chip—would likely continue to double every 18 months to two years for decades to come.\(^6\) Mead dubbed the phenomenon Moore’s Law, and exponential chip scaling has continued, more or less, for 50 years. (See figure 2.)

Today, however, many think Moore’s Law has already run out of steam, that the pace of chip performance improvements may have substantially slowed beginning around 2005.\(^7\) Information technology more generally is at the center of a number of economic and cultural debates. The information revolution, some economists believe, has not lived up to the hype. It was, they say, not as powerful as the Agricultural and Industrial Revolutions, or even the age of plumbing, electricity, and air conditioning.\(^8\) They think post–World War II America was a zenith of innovation and income

---

**Figure 2. Transistors Continue on Moore’s Path . . . for Now**

![Graph showing transistor density doubling every 18 months to two years from 1960 to 2020.](Image of graph)

Note: Red dot is the author’s estimate for feature size of 14 nm in 2014.
Source: Stanford CPU Database.
growth and that information technology has not delivered the goods for the past generation. In any case, they say, the Information Age has peaked and is now waning. They forecast a new normal of paltry prospects well into the future.

Others argue just the opposite—that information technology is so dangerously powerful that robots will steal most jobs.9 This is if artificial intelligence does not wipe out the human race first.10

Yes, shrinking silicon transistors is getting more difficult. And of course technology is rendering some jobs obsolete. The bulk of the evidence, however, suggests information technology has delivered both technically and economically: it has achieved the promise of Moore’s Law in both its narrowest sense of transistor scaling and its broadest effect of widespread economic uplift. Although future information processing technologies will not mirror the last five decades of silicon scaling in their technical particularities, it is unlikely we have reached the end of the road. Instead, Moore’s Law has built a foundation of information tools that will make nearly every sector of the economy an information industry and provide the capacity for new discoveries and enterprises. Multiple Moore’s Law paths of exponential technology and economic growth are open to us—if, like Moore himself, we commit to building the future.

In the remainder of this paper, I examine Moore’s Law’s technical achievements and economic impact to date. I then consider the public policies that encouraged rapid innovation in information technologies (IT) and the way IT, in turn, helped foster greater freedom around the world, and finally, examine the technical future of Moore’s Law and the economic potential of information technology.

The Technical Track Record

In every act of the Moore’s Law story, the technical and economic hurdles on the horizon were daunting. And yet scientists and engineers—and the business leaders who funded their research—pushed forward and kept extending the horizon. The narrow definition of Moore’s Law was about the economies of component scaling. But the early silicon pioneers also knew this would translate directly into similar exponential trends in computer processing power, abundant data storage and memory, and large volumes of tiny chips of ever-greater variety, such as sensors, low-power electronics, and embedded devices. The technical track record should thus be judged not only by transistor sizes and counts but also by the resulting practical performance measures.

Processing Power. Around the year 2000, a now-famous chart began showing up at technology conferences everywhere.11 Moore’s Law was going strong.12 Advances in chip speeds had actually outperformed the trend line in the 1990s. But the engineers saw a problem on the horizon—heat. Computers already required fans to cool the innards of tower PCs and laptops, and designers were even contemplating liquid-cooled systems similar to radiators for automobile engines. The chart, created by Intel engineers, showed that by 2008 leading-edge chips would generate the heat of a rocket nozzle. By 2010, chip “power densities” might approach the heat found on the surface of the sun. This temperature trend was clearly unsustainable.

The Heat Equation

On a microchip, the power density (roughly equal to temperature) is a function of voltage and clock speed. In simplified form, power is equal to the voltage squared multiplied by the frequency (how many times per second the transistors turn on and off): \[ P = V^2 f \], where \( P \) is power, \( V \) is voltage, and \( f \) is frequency.

Faster clock speeds allow chips to perform more calculations per unit of time. But as clock speed (\( f \)) rises, so does temperature (\( P \)). For many years, engineers were able to reduce voltage (\( V \)) enough so that even as \( f \) rose, \( P \) did not explode. In the mid-2000s, however, further reductions of voltage became far more difficult. Clock speeds thus had to level off to avoid catastrophically high temperatures.
Until the mid-2000s, faster clock speeds, the number of times per second a chip’s transistors turn on and off, had been a key source of performance improvements. But because heat is a function of clock speed and voltage, increasing clock speeds requires reducing voltages. Below a threshold voltage level, however, chips stop working. Clock speeds and voltages would thus have to level off. But without faster clock speeds, where would performance improvements come from?

In 2004–05, just as predicted, maximum clock speeds leveled off at 2–3 gigahertz (GHz), and around this time, some said Moore’s Law was ending. By some measures, this appeared true. The National Academy of Sciences (NAS) in its comprehensive 2011 analysis even referred to the situation as “the crisis in computing performance.”

Yet transistors kept shrinking in size. Today’s leading-edge silicon manufacturing process is 14 nanometers (nm), and the International Technical Roadmap for Semiconductors, a key industry authority, says we are headed toward 7 nm. And something else happened: overall chip performance kept getting better (figure 3).

The solution was found in parallelism. The rate of improvement in single-processor performance had indeed slowed a bit. But as clock speeds and voltages leveled off, firms began putting two, then four, then more processors on each chip, and the results were encouraging. This “multicore” strategy was new to microprocessors, or CPUs, but it was already familiar in other types of chips that specialize in real-time processing of high-speed data, such as graphics processors (GPUs), network processors (NPUs), and digital signal...
processors (DSPs). New materials, such as “strained” silicon and high-k dielectrics, and new device technologies, known as FinFET or tri-gate transistors, also helped sustain performance increases.

Using large data sets of Intel microprocessors and applying standard industry benchmarks known as SPEC CPU2000 and CPU2006, economists David Byrne, Stephen Oliner, and Daniel Sichel have shown that total CPU performance improvements decelerated only marginally in recent years. Between 2000 and 2013, performance grew at a 32.4 percent compound annual rate, which was close to the 1971–90 rate of 36.4 percent. The 1990s were the real outlier at 59.7 percent annual growth. For the entire period, stretching nearly back to Moore’s paper, Byrne, Oliner, and Sichel found microprocessor performance increased at an astounding rate of 40.3 percent per year. This works out to almost exactly a doubling (2.00629) every two years, which was Moore’s revised prediction in 1975. A microprocessor in 2013 was thus 1.5 million times more powerful than in 1971.

Leading-edge microprocessors found in desktops, laptops, and servers, however, were not the only beneficiaries, nor the sole measures of success, of Moore’s Law. At the high end, it also enabled a revolution in graphics processors. Nvidia’s most advanced GPU, the GTX Titan X, for example, now contains eight billion transistors and 3,072 cores, provides 336 gigabytes per second of memory bandwidth, and has total computing power of 6.14 teraFLOPS, which would have qualified this single chip as the world’s fastest supercomputer until the year 2000. GPUs are increasingly used in cluster and cloud computers and are being applied to a range of more general-purpose applications in the form of general-purpose GPUs.

Memory and Storage. Moore’s Law also revolutionized memory and data storage. Figure 4 shows a 1956 IBM 305 RAMAC storage system with a capacity of 3.75 megabytes (MB) being loaded with a forklift onto an airplane. (For reference, 3.75 MB is about the size of one high-resolution digital photograph or a single five-minute digital song.) IBM leased the machine to customers for around $3,000 per month. Today, a 128 gigabyte (GB) flash drive the size of your fingertip stores 34,000 times more information and costs around $40. Flash memory and dynamic random access memory (DRAM) are as important to smartphones and, increasingly, to cloud computing as are microprocessors. New 3D memory technologies from Intel and Micron will soon yield a 3.5 terabyte (TB) flash drive and a tiny SSD (solid state drive) card holding 10 TB—two million times more than the two-ton IBM machine.

Embedded, Low-Power, and Sensing Chips. The Moore’s Law phenomenon, however, is just as important at the nonleading-edge end of the performance spectrum and across a range of attributes. It provides for extremely inexpensive embedded chips—chips we never see and rarely think about—in automobiles (which now contain more than 100 microprocessors and microcontrollers), running shoes, toothbrushes, thermostats, and other kinds of consumer and industrial devices. Moore’s Law also provides the tools to
build extremely low-power chips and components, like MEMS (microelectrical mechanical systems) accelerometers, for untethered devices like mobile phones and for radio frequency identification (RFID) tags, which cost just a few pennies to manufacture.

Moore’s Law has also enabled the digital camera revolution. Imaging sensors do not depend on the smallest transistors available but, like many embedded and low-power applications, do depend on nanotechnology innovations in materials, manufacturing, and design. Increasingly, the specialized sensor designs are biologically inspired and highly parallel — what Mead in the early 1980s called neuromorphic technology.\(^{22}\)

Apple’s sales of 101 million iPhones, iPads, and Macs in the fourth quarter of 2014 meant that in just three months it sold close to 200 million cameras, since most of its devices contain a front and rear camera. That was twice the total of all standalone cameras sold worldwide in the record year of 2012. With nearly every mobile phone now containing one or two cameras, and with increasing deployment of webcams and security systems, worldwide camera sales now total around three billion per year. In addition, nearly all these cameras now double as video recorders.

This world of ubiquitous sensors, part of the emerging Internet of Things, does not depend on continued growth of top-line computer speed nearly as much as did PC performance in the 1980s and ’90s. But these sensory, embedded, and low-power applications nevertheless vindicate Moore’s original goal — unimaginably huge chip volumes at very low cost.

The Economic Effects

One hundred private firms, according to the Wall Street Journal, are now members of “the billion-dollar start-up club.”\(^{23}\) Sixty-six of these start-ups are American, and a number of them have recently achieved valuations of $10 billion, $20 billion, or even $40 billion. The total value of the 66 US club members is $256.7 billion, not including the 11 club members that went public or were acquired in 2014 and the first half of 2015. Most of these are information technology firms. Many public US technology firms are, likewise, booming. The market values of just seven tech leaders — Apple, Google, Microsoft, Facebook, Oracle, Intel, and Amazon — total nearly $2.38 trillion, more than the entire value of the German or Australian stock markets. Apple alone is twice as valuable as any firm in the world, including Exxon Mobil and Google.\(^{24}\)

Measuring the economic impact of information is difficult. Indeed, it may be the central question in all of economics right now. We know the advancement and spread of information technology is powerful. By some intuitive measures, it seems almost miraculous. By many orthodox measures, however, its recent economic impact seems tepid. We encountered this paradox in the 1980s when Robert Solow famously noted, “You can see the computer age everywhere but in the productivity statistics.”\(^{25}\) Computers finally showed up in productivity data during the brief productivity boom in the late 1990s and early 2000s. The conventional view that median incomes have stagnated for the last four decades, however, runs counter to a presumption that Moore’s Law should have substantially improved living standards.

In 2011, George Mason University economist Tyler Cowen ignited this debate with his book The Great Stagnation, which argued that the Internet was a great source of “cheap fun” but not of jobs or incomes.\(^{26}\) Paul Krugman in 1998 famously said that by 2005 we would view the Internet as no more important than the fax machine.\(^{27}\) With “The Demise of US Economic Growth” in 2012, Northwestern University’s Robert Gordon drew an even darker picture. Gordon argued that the First and Second Industrial Revolutions were far more potent than the third (the Information Revolution) and, more important, that the Information Revolution may already have ended.\(^{28}\)

The first two Industrial Revolutions account for the sharpest leap in living standards in history, and it will be difficult for any economic era ever to match their import. But Gordon argues that because Amazon is more than 20 years old, Google more than 15, and Facebook more than a decade old, and because ATMs and barcodes stopped adding to productivity years ago, IT is a spent force. This betrays a remarkably narrow view of the industry’s past and prospects and of its effects across the economy.
The notion that IT has not produced gains for the middle class is mistaken—first of all, because the base argument of middle-class stagnation is itself overstated and, in many cases, plain wrong. Thomas Piketty, for example, claims that real incomes for the bottom 90 percent of Americans have not risen since 1968. This is very far from the truth. Real consumption per person over the period, economist Alan Reynolds notes, has tripled.29

Among other problems, the Piketty figures do not include realized capital gains or a dramatic rise in private and public benefits, from health insurance to food stamps. They do not account for a plunge in the middle-class tax burden. They do not include some $20 trillion of retirement savings in IRAs and 401(k)s. Related assertions of income stagnation often do not account for important shifts in household size. It should not surprise us that a household of one may have an income lower than a household of three. The incomes of some of today’s households may thus appear to have shrunk, yet shrinking household sizes might reflect growing prosperity, not stagnation.30

Second, the measuring rod for our standard of living may have become less accurate over time. Using the conventional consumer price index (CPI), for example, another typical measure of real median US incomes between 1967 and 2014 shows a near catastrophic rise of just 4.2 percent.31 Using the only slightly more advanced personal consumption expenditures (PCE) deflator, however, real median incomes (not including those important adjustments for benefits, taxes, and savings) rose 33.0 percent. Real income measures are thus highly sensitive to the chosen inflation measure, and the price deflators themselves may not fully account for the true benefits bestowed by IT and other technologies.

Some have pointed to an apparent slowdown in the chip price declines we had become accustomed to and that helped fuel the productivity surge in the 1990s. Microprocessor prices from 2000 to 2013, according to the Producer Price Index, fell at an annual rate of 28 percent, a fast pace for almost any other product but slower than the historical rate for microprocessors. After 2008, prices fell by just 8 percent per year on average, a dramatic and seemingly ominous slowdown. Byrne, Oliner, and Sichel, however, found something curious. Intel, which makes up nearly half the US semiconductor market, “dramatically” changed its pricing practices in the middle of last decade (as it happens, about the time clock speeds leveled off and around the time cloud computing began).32 This superficial change appears to mask continued rapid price drops. After adjusting prices for differences in chip performance and using only the prices of new models to avoid the likely problems with list prices for older chips, Byrne, Oliner, and Sichel estimate that the actual annual average price drop for 2000–13 was 44 percent.33 This meant that according to the official government measure, $100 worth of computing power in 2000 could be purchased for $1.31 in 2013, which sounds impressive. The authors, however, show that the actual cost in 2013 may have been just 4.8 cents ($0.048). According to this measure, consumers can purchase 27 times more computing power per dollar than the official data suggest. (See figure 5.)

“[W]ith millions of new and rapidly changing products and services,” Harvard economist Martin Feldstein notes, government statisticians “are supposed to assess whether $1,000 spent on the goods and services available today provides more ‘value’ or ‘satisfaction’ to American consumers than $1,000 spent a year ago.”34 “These tasks are virtually impossible,” he writes. How, for example, do we measure an inflation rate of a lifesaving pharmaceutical or surgery that did not exist several decades ago? How much is Wikipedia worth? In how many ways do time-saving information tools improve our nonwork lives? How does the variety and quality of entertainment options factor in our standard of living? Although it is difficult to measure, IT provides hundreds of billions of dollars in consumer surplus every year—goods and services consumers would be willing to pay for but essentially get for free. “Today’s pessimists,” Feldstein concludes, “are wrong because the official statistics underestimate the

The notion that IT has not produced gains for the middle class is mistaken.
growth of real GDP, of productivity, and of real household incomes.”

Median incomes appear to have stagnated since the 2008 financial crisis and employment trends have deteriorated, but these serious problems likely reflect demographics and an antigrowth policy environment, not a failure of IT.

Where would the economy be without IT? The entire software industry, for example, is an outgrowth of Moore’s Law. It is a huge industry in its own right (around $500 billion the US in 2014) but is also increasingly the foundation for firms in every industry. It accounts for most of the value of firms like Google and Apple but also is replicating, in bits, all kinds of services from the old economy, from taxis to hotels to money itself. Software is “eating the world,” in the famous words of venture capitalist Marc Andreessen. Human creativity, delivered in software apps, services, and platforms, is not nearly at an end. But like much of the information economy, it is difficult to describe using traditional economic measures.

This dramatic gap between the official government data and reality is just one discrete example of a likely trend across the range of price, output, and productivity measurements for the information economy. Consider:

- In 1990, building a device with the computer, storage, and communications power of a single iPhone 6 would have cost more than $5 million (perhaps $10 million, adjusted for inflation). Today, more than two billion people own smartphones, what some have called “supercomputers in your pocket.”

Figure 5. Microprocessor Prices Continue Rapid Decline

According to the producer price index (PPI), microprocessor price declines appeared to level off around 2005. By the PPI measure, $100 of computing in 2000 could be purchased for $1.31 in 2013.

In fact, the “real price” of computation kept falling and was just 4.8 cents in 2013. Consumers could thus purchase 27 times more computer power than the official data suggests.

Source: David M. Byrne, Stephen D. Oliner, and Daniel E. Sichel, AEI, 2015.
In 1990, worldwide Internet traffic totaled one terabyte (TB) per month, about as much data as fits on a PC desktop hard drive. In 2015, Internet traffic totals around 75 million TB per month.

The smartphone-cloud combination created a whole new software industry—mobile apps. After Apple’s launch of the App Store in 2008, it took just four years to go from zero to 60 billion app downloads. But the value of the productivity, creativity, and variety of software is difficult to measure and is often buried deep inside other products.

In 2001, the cost to sequence one genome was $100 million, but today the cost, which is mostly a function of computer speed and better algorithms, is just $5,000 and is rapidly headed toward $1,000. (We may not yet have seen much resulting innovation from genomics, but we soon will.)

William Nordhaus, in a reprise of his famous study of the history of lighting technologies, looked at the price of computation over the last two centuries until 2006. Over this period, he estimated that computation per dollar grew by a factor of seven trillion, and the labor cost of computation—the number of work hours needed to purchase a unit of computation—dropped by a factor of 73 trillion, with the bulk of progress coming after World War II. Extrapolating from Nordhaus’s figures, which end in 2006, I estimate that in the five decades of Moore’s Law, the labor cost of computation has dropped by a factor of around one trillion. It is very difficult for
conventional economic measures to capture all of the value in such exponential trends.

For 20 years, economist Dale Jorgenson has been measuring information technology’s impact on the economy, and his latest estimates take advantage of new data from the Bureau of Economic Analysis. Despite the likely underestimation of the impact of information technology, Jorgensen finds that it still accounts for nearly all gains in total factor productivity—or “innovation”—over the last 40 years. And it probably accounts for between 50 percent and 70 percent of all productivity gains over this span. Using his own methods, Oliner estimates around half of all nonfarm productivity growth since 1974 is due to information technology (possibly an understatement given his view that the official data underestimate semiconductor productivity). Official productivity over the last 50 years may not have risen as fast as in the postwar boom, but IT is not the culprit.

Technologies of Freedom

“The world of bits was unregulated, the world of atoms was highly regulated. We’ve had a lot of innovation in bits, not much in atoms.”

—Peter Thiel

The digital world, from microchips to the Internet, has flourished in a policy environment almost entirely free of top-down regulatory constraints. It is no coincidence that these industries have also been the most explosively innovative.

The key to Moore’s Law was relentless spirals of innovation, lower prices, surprising consumer demand, more capacity, and further innovation. These learning curves depended on an environment of freedom where firms and entrepreneurs could iterate rapidly and also take big risks on unproven technologies and business models. As I argued in 2009:

Microsoft and Intel built upon each other in a virtuous interplay. Intel’s microprocessor and memory inventions set the stage for software innovation. Bill Gates exploited Intel’s newly abundant transistors by creating radically new software that empowered average businesspeople and consumers to engage with computers. The vast new PC market, in turn, dramatically expanded Intel’s markets and volumes and thus allowed it to invest in new designs and multibillion dollar chip factories across the globe, driving Moore’s law and with it the digital revolution in all its manifestations.

Software and hardware. Bits and bandwidth. Content and conduit. These things are complementary. And yes, like yin and yang, often in tension and flux, but ultimately interdependent.

Just like the previous generation of PCs and software, today’s digital economy depends on the same virtuous circle among broadband, app, device, cloud, and content firms. But as Peter Thiel explains, not every industry has enjoyed the highly competitive, complementary, and generally freewheeling environment of IT. There is a huge divergence in innovation, and thus incomes, between industries depending on their regulatory status. In a vivid description of the differences between industries and the regulation thereof, Robert Graboyes concludes, “In the same years that IT exploded, changing how billions of people live their daily lives, health care was painfully slow to innovate.”

As free enterprise policies helped Moore’s Law prosper, Moore’s Law also spread freedom around the world. Although a number of factors were at play, it is hard to ignore the fact that during Moore’s Law’s run, global poverty dropped from more than 60 percent in 1965 to just 9.6 percent in 2015.

Information technologies helped achieve this expansion of the world’s middle classes in a number of direct and indirect ways. First, the US lead in computers helped it win the Cold War, unleashing waves of political freedom. China, anticipating these changes, chose in the late 1970s to free its economy and focus, like nearby Taiwan and Singapore, on the emerging electronics industries. Information technologies enabled the globalization of trade and capital markets, which modernized many developing economies. Most directly, the Internet and smartphones provided individuals access to all the world’s knowledge and communication...
with anyone across the globe. History is complicated, but Moore’s Law is an important facet in this epochal transformation.

Two policy challenges threaten to slow, or in some cases even reverse, the spread of these technologies of freedom. First, nations such as China, Iran, and Russia are undermining the Internet’s openness and universality, two of its primary attributes. These regimes censor content and communications and repress those who challenge the central authorities’ edicts. These nations are also attempting to use international bodies to assert more top-down bureaucratic control over Internet governance, an effort that could lead to fragmentation, or “balkanization,” of the Internet and thus the worldwide digital political economy. Widespread hacking and other online crimes also threaten the security and usefulness of the Internet and the trust on which any economy—real or virtual—relies.

In the US, the biggest threat to the virtuous cycle of innovation in the information economy is the effort by the Federal Communications Commission to regulate the Internet. The Internet is the central nervous system of the computer, device, mobile, and software industries and, increasingly, of every other industry and of the culture. A reversal of the unambiguously successful policy of regulatory humility is inexplicable and perhaps the only thing that can interrupt IT’s innovative spirit and poverty-fighting power.

As silicon scientist Chris Mack has written, “Nothing about Moore’s Law was inevitable. Instead, it’s a testament to hard work, human ingenuity, and the incentives of a free market.”

---

Figure 7. Global Poverty Plunged over the Last Half-Century

Source: Max Roser, “Our World in Data.”
The Technical Future

“There always seems to be a barrier two to three generations away.”

—Gordon Moore

Can chips continue to scale at even a rough Moore’s Law pace? For the next five years, the answer is probably yes—and in relatively conventional ways.

Micro Innovations. Silicon-based (CMOS) transistors will shrink to 10 nm and then 7 nm. After that, we will finally reach the atomic limits that skeptics thought would end Moore’s Law long ago. But already we are seeing innovations that can push Moore’s Law further than we believed possible two or three generations ago. Three-dimensional memory cells, stacked 32 or 48 layers thick, are coming soon. Fundamentally new memory technologies, such as “spintronics,” which use not just the presence or absence of electrons but also the direction of their “spins,” will allow us to store more data per transistor and cell. True 3D chip stacking may also provide a boost for logic circuits.

Scientists are also optimistic that a range of new materials will pick up where silicon leaves off. We have used “exotic” materials like gallium arsenide (GaAs) for decades to make high performance communications devices. But now GaAs (and variants) will be used more extensively and possibly make its way into many mainstream products. Gallium nitride (GaN) will be used to create a new generation of “power chips,” which already control power management in the electromechanical world but could now enable new applications like wireless charging of mobile devices and LiDAR (think radar) for autonomous cars and virtual reality systems. Germanium (Ge), III-V materials, and ultra-high-k dielectrics will be used more extensively in conjunction with traditional CMOS.

An even more radical transformation may come from carbon nanoelectronics, including carbon nanotubes, graphene, and even diamond. IBM, for example, just announced a breakthrough in carbon nanotube transistors that it says “could overcome contact resistance challenges all the way to the 1.8 nanometer node—four technology generations away.”

New ways of storing, reading, and manipulating information may also extend performance increases. Today, information is represented mostly through electron charge. But other “state variables,” such as spin, phase, polarity, and even molecular configuration could be used to boost the “information density” of our materials and device structures.

We do not today know which of these device, material, chip design, and state variable technologies will yield high-performance, cost-effective, manufacturable gains. But with so many good options, it is likely some of them will bear fruit. We will likely use a number of these material, device, chip, and state variable innovations in combination. And although a key goal will be to maintain microprocessor improvements, the technologies will be used to produce an increasingly diverse array of chips for numerous applications—sensing, power conversion, electromechanical, low-power—that may not require bleeding-edge speed.

Macro Innovations. A decade ago, as chips started moving to parallel architectures (multiple computing cores) at the micro scale, engineers also began employing parallel computing architectures at the macro scale, linking processors, memory, storage, and software stacks across the data center—and then across numerous data centers—to perform integrated computing functions. Supercomputers had used some of these cluster-computing techniques for decades, but they were few in number and used for niche applications. The rarity of supercomputers and a lack of adequate communications bandwidth meant the advances in parallel hardware architectures, and the programming of these parallel computers, came slow.

Fiber optics, gigabit Ethernet, and broadband access networks, however, would fundamentally transform our computing architectures. “When the network becomes as fast as the processor,” Eric Schmidt said in the early 1990s, nearly a decade before joining Google, “the computer hollows out and spreads across the network.” Or, as Schmidt’s then-employer Sun Microsystems stated in its motto, “The Network Is the Computer.”

Mead also presaged this development, in more technical terms:
In any physical computing system, the logical entropy treated by classical complexity theory is only part of the story. There is also a spatial entropy associated with computation. Spatial entropy may be thought of as a measure of data being in the wrong place, just as logical entropy is a measure of data being in the wrong form. Data communications are used to remove spatial entropy, just as logical operations are used to remove logical entropy. Communications is thus as fundamental a part of the computation process as are logical operations.\textsuperscript{56}

The vision was compelling. But coordinating thousands of servers and data dispersed across the globe was a huge challenge. Google achieved major advances with its Google File System and MapReduce paradigms, which empowered disparate computing resources across the cloud to talk to one another and act as a single large computer when needed, or to split up varied smaller tasks based on the needs of end users on their PCs or smartphones. The company has continued to innovate in parallel architectures and programming for its own web services, like search and Gmail, and for its third-party cloud services, such as Compute Engine and Google Storage. Amazon Web Services, however, is the largest provider of cloud services and now employs some two million computers in its cloud offerings alone.\textsuperscript{57}

In keeping with Moore’s original vision of radically low-cost computing, the pricing of these cloud services is often measured in pennies.\textsuperscript{58} We are extending the economics of Moore’s Law from the microcosmic chip to the macrocosmic data center. When we link computing and memory resources with light-speed optics, we are, in a rough way, building “a chip on a planet.”\textsuperscript{59} The move to multicore chips was a way to relieve chips of the growing heat burden at the micro scale. Warehouse scale computing extends this function to the macro level: the massive, centralized air conditioning operations that cool servers in our data centers in effect remove heat from our smartphones and mobile devices.\textsuperscript{60} This—the physics of computation—is the only fundamental limit on the information economy.

In 1965, the only computers were large, centralized mainframes at universities and large corporations that were still fed by punch cards. A small number of lucky workers, professors, and students had to wait in line to access scarce computer time. Today, the centralized computers of our era—in “the cloud”—generate 4.7 zettabytes (ZB, $10^{21}$) of data per year.\textsuperscript{61} The smartphone-cloud partnership gives us a glimpse of the new paradigm. But we are only a decade into this wave, which will yield even more impressive gains for decades to come. The computers, devices, networks, software platforms, services, and apps we will build with this new infrastructure is limited mostly by self-imposed constraints and our imaginations.

\textbf{During Moore’s Law’s run, global poverty dropped from more than 60 percent in 1965 to just 9.6 percent in 2015.}

\textbf{The Economic Future}

On October 9, 1910, the New York Times published an analysis of a perplexing new question in transportation technology. The crucial inputs were oil, gasoline, hay, and oats. The output was passenger miles per dollar. The title of the article was “Auto vs. Horse.” The verdict? “Six-Day Test Shows Motor Car Cheaper and More Efficient Than Animal.” At 1.57 cents per passenger mile versus 1.84 cents, the auto beat the equine.\textsuperscript{62}

The automobile turned out to be an engine of the “American Century.” At the outset, however, it was not even obvious the automobile outclassed animal as a vessel of mobile horsepower, let alone that it would become an industrial and cultural juggernaut. We tend to underestimate the power of technology over the medium to long term.

But we also fear technology. Consider the reaction a century ago if citizens were told that by the late 1960s auto accidents would annually kill 50,000 Americans. We may have banned further development of this new technology. Experience suggests today’s worries are overestimated. The two pessimistic views—that technology is
either impotent (Gordon’s demise of growth) or dangerously dystopian (runaway robots and AI)—are both relatively unlikely. Yes, some technologies will not live up to their billing. And yes, technology always brings its share of economic and cultural dislocations, and even safety hazards. But in terms of both wealth and health, technology has always delivered massive net gains.

It is far more likely that technology will confer substantial and surprising benefits than that it will fail or catastrophically succeed. Because information is so fundamental in nature, technology, and the economy, information technology has likely only begun its vast impact on every existing and new industry. Far from ending, we are at only the end of the beginning of information technology. The case for “rational optimism” is strong.

Medicine and health care, to take perhaps the two most important examples, are in the midst of profound transformations into information industries. Vaccines and antibiotics brought us out of the dark ages and dramatically boosted human longevity. And although researchers and the Food and Drug Administration over the last century used systematic information in crude ways, medicine until recently was mostly a trial-and-error, hit-and-hope world. Understanding the codes of the genome, proteome, epigenome, and metabolome, however, will unleash molecular medicine. “The vital core of medicine,” writes Peter Huber in *The Cure in the Code*, “is now on the same plummeting-cost trajectory as chips and software.” Just as the macrocosm of vacuum tubes gave way to the microcosm of silicon chips, we are moving from the goopy world of petri dishes to the biocosm of DNA and protein codes, the information networks of molecular metabolics.

The new knowledge and tools will yield therapies customized not to symptoms or broad disease categories but to the individual person. Information-based medicine will also provide for diagnostic “sniffers”—molecular sleuths meandering through our bodies and working with smartphone apps to gauge chemicals in our breath and ominous signals in our retinas.

Smartphones and other devices will vacuum up huge amounts of data on our health and our responses to therapies, nutrition, exercise, and the environment. Pooling and then analyzing data from millions of subjects will yield new insights and expand the possibilities for experimentation. For information-based medicine and health care to impart the most powerful economic benefits, however, we must rethink policy, research, and the hospital-based delivery system. As Robert Graboyes writes, we need to replace the “fortress” mentality that governs today’s industry and regulatory apparatus with a “frontier” ethos of entrepreneurial business models and scientific discovery.

Total productivity growth is substantially weakened by the poor performance of a few sectors like health care and education—the famous sufferers of Baumol’s cost disease. Annual productivity growth outside health care between 1990 and 2012 was around 2.0 percent. But health care lost ground at a rate of 0.6 percent per year. Over 20 years, that is a 60 percent differential. If health care could escape its cage of hyperregulation and truly exploit information science, this bloat industry could turn into a productivity growth industry.

At one-sixth of the economy, such a transformation could substantially improve the prospects for overall economic growth.

Unsatisfactory educational attainment, likewise, is often billed as a suppressant to growth. Failing K–12 schools and expensive college degrees of questionable quality have indeed weighed on our economic possibilities. Precisely because of this, however, improvements in education could yield a surprising burst of growth. Economists are pessimistic about the almost-certain slowdown in educational attainment. They say that average formal years of education has already leveled off and that formal education will stop adding to productivity. It is of course true that most people cannot continue dedicating ever more years to education—at some point, most people have to go to work.

But what about informal education? The digital world makes instantly available most of the world’s basic knowledge, which people of every background and occupation can use throughout their careers and lives. New digital educational services will allow people to continue learning in a variety of structured and unstructured ways. The new educational tools may also exert pressure on, and thus help transform, many of our inadequate and expensive traditional educational
institutions. A focus on “years of education” may thus be misleading.

In fact, it is possible that a revolution in all forms of education—from preschool through doctoral programs, from vocational training to professional schools to lifelong learning—will dramatically improve the quality and quantity of education even if the orthodox “years of education” measure appears to have stalled. Higher educational attainment around the world should also continue adding to the stock of ideas and technology.

Health care and education are just two of the largest and most obvious possible beneficiaries of radical transformations based on information technologies.

Conclusion

“Fifty years of mind-numbingly fast progress,” writes Chris Mack, “have turned a hundred dollar chip with a few dozen transistors into a 10 dollar chip with a few billion transistors. Much of what we enjoy about modern life is a direct result of this progress.”

By a few measures—clock speeds and the performance of single-processor cores—Moore’s Law appeared to slow over the last decade. By other important measures, however, it continues apace. Transistor doubling per unit area has been sustained. Multicore architectures and new transistor designs, meanwhile, have offset most of the performance deceleration due to clock speed leveling. Advances in programming of parallel systems—on the chip and among linked computers—have likewise yielded performance advances beyond the traditional Moore’s Law measures. A wider variety of chips for an expanding array of applications take advantage of the Moore’s Law advances even if they do not require bleeding-edge processing speeds.

New computer architectures that emerged in the last decade will continue this rapid evolution. The combination of powerful mobile devices and the even-more-powerful cloud—linked by fiber optic and wireless broadband—gives always-on supercomputer capabilities to billions of people and to developers of nearly unlimited software services and apps.

Despite a number of very serious challenges with traditional silicon materials and approaching atomic limits, the semiconductor industry is successfully experimenting with a wide range of new materials and device designs. Although Moore’s Law may not continue to scale using the conventional metrics, such as transistor counts, a variety of innovations in materials, devices, state variables, and parallel architectures will likely combine to deliver continued exponential growth in computation, storage, and communications.

Information technology, powered by Moore’s Law, provided nearly all the productivity growth of the last 40 years and promises to transform industries, such as health care and education, which desperately need creative disruption. It has powered a wave of globalization that helped bring two billion people out of material poverty and has delivered the gift of knowledge to billions more.

Information technology, powered by Moore’s Law, provided nearly all the productivity growth of the last 40 years and promises to transform industries, such as health care and education, which desperately need creative disruption.

Information is fundamental. Information technology is not a spent force. It will continue to advance in its own right and power other industries—provided we stay on the frontiers of discovery and entrepreneurship. This requires us to maintain the successful environment of “permissionless innovation” that has characterized the information economy—and to encourage the spread of this policy to other sectors of the economy.

“It’s extremely important,” Mead summed up, “that there’s a living example of people’s belief in the future, bringing it to pass.” In making the future tangible, Moore’s Law was, and is, a chief driver and exemplar of the American experiment.
Notes


5. Mead was building off of Richard Feynman’s famous notion that “there’s more room at the bottom.” In the quantum world, embodied most tangibly in silicon electronics, the rule would be: “the less the space, the more the room.”

6. In 1975, at the 10-year anniversary of his original paper, Moore projected transistor densities would double roughly every 18 months to two years, instead of every year.


11. See a reference to the ubiquitous Intel “heat” chart in, for example, this UC Berkeley presentation on silicon transistor technology, 2011, www-inst.eecs.berkeley.edu/~ee130/sp13/lectures/Lecture28.pdf.

12. In the beginning, Moore said three factors would help generate more components per chip: shrinking feature size (making transistors smaller), increasing chip area, and reducing wasted space between components. (In the original paper, Moore also emphasized the economy of boosting component counts: what number of components minimizes the cost?) By the 1980s, better designs had mostly eliminated wasted space, and by the 1990s chip sizes were not growing as fast. For a while clock speed, or frequency, was linked to Moore’s law, but frequencies leveled off in 2004–05. Over the last three decades, Moore’s Law (as originally understood) thus increasingly came to be defined by one factor: feature size reduction.


15. For one measure of transistor scaling, see the Stanford CPU database (CPU DB), from which figure 2 is taken. I have added the red dot in the lower right corner. It represents the data point of the 14 nm (0.014 µm) node in 2014.

17. Elsewhere I have referred to this concerted move toward parallel architectures across a range of information technologies and platforms as the “Paralleladigm.” In addition to CPUs, GPUs, and NPUs, parallel architectures are increasingly important in neuromorphic analog electronics, fiber optics (wavelength division multiplexing, WDM), wireless radio air interfaces (OFDM, MIMO), wireless networks (small cells), and large-scale computing platforms (cloud computing, warehouse-scale computing, GPU-based supercomputers, and so forth). See Bret Swanson, “Into the Exacloud,” *Entropy Economics*, November 21, 2011, http://entropyeconomics.com/wp-content/uploads/2011/11/Into-the-Exacloud-21-Nov-2011.pdf.


19. Again, Moore's projection was for the number of components per chip, not necessarily broad cost-performance measures, to double every two years. Yet the two metrics were roughly equivalent, both in a technical sense and in the popular understanding.


27. Paul Krugman, “Why Most Economists’ Predictions Are Wrong,” *Red Herring* (June 1998), accessed at http://web.archive.org/web/19980610100009/www.redherring.com/mag/issue55/economics.html: “By 2005 or so, it will become clear that the Internet’s impact on the economy has been no greater than the fax machine’s.”


30. See, for example, the work of Scott Winship exploring the household size topic and its relationship to real income growth, including “Debunking Disagreement over Cost of Living Adjustment,” *Forbes*, June 15, 2015, www.forbes.com/sites/scottwinship/2015/06/15/debunking-disagreement-over-cost-of-living-adjustment/; and “Whether and How to Adjust Income Trends for


32. See Byrne, Oliner, and Sichel, “How Fast Are Semiconductor Prices Falling?” 9: “Between 2003 and 2006, the properties of Intel’s posted prices for MPU chips changed dramatically. Prior to 2003, the price of a specific Intel MPU model tended to drop fairly rapidly in the year or two following its introduction, especially once a new, higher performance model became available. By 2006, this pattern had completely changed; the posted price of a specific model tended to remain constant, even after a new, higher performance model became available at a similar price.”

33. Ibid.


37. Data for 1990 come from Andrew Odlyzko of the University of Minnesota (MINTS—Minnesota Internet Traffic Studies) and can be found in Swanson, “Into the Exacloud.”


41. See tables 6 and 7 in Nordhaus, “Two Centuries of Productivity Growth in Computing,” and the data appendix. I estimate the 2015 figure for labor cost of computation to be roughly 3.00E-12, which is around one-trillionth the value for the average of the 1960s, approximating Moore’s prediction in 1965.


45. Peter Thiel (with Blake Masters), Zero to One (New York: Crown Business, 2014.)


58. See, for example, Google Cloud pricing charts at https://cloud.google.com/pricing/.

59. I first heard it described this way by George Gilder.

60. Again, Mead told us back in 1980 that the physics of computation would lead in this direction—to the Internet and cloud computing: “The communication of information over space and time, the storage and logical manipulation of information by change of state at energy storage sites, and the transport of energy into and heat out of systems, depend not only on abstract mathematical principles but also on physical laws. The synthesis and the functioning of very large scale systems, whether artificial or natural, proceed under and indeed are directed by the constraints imposed by the laws of physics” (emphasis added). See Mead and Conway, *Introduction to VLSI Systems*.


63. Gordon and other economists correctly note that unfavorable demographics will limit one important source of economic growth over the next several decades.

64. For a thorough argument on the reasonableness of optimism, see Matt Ridley, *The Rational Optimist* (New York: Harper-Collins, 2010).


About the Author

Bret Swanson is a visiting fellow in AEI’s Center for Internet, Communications, and Technology Policy and president of Entropy Economics LLC, a strategic research firm specializing in technology, innovation, and the global economy.

Acknowledgments

I thank Steve Oliner for his expert insight and for the courtesy of sharing his data, Ari Rabkin for helping me understand the latest parallel programming strategies, Nick Tredennick for thoughts on both the past and future of Moore’s Law, George Gilder and Carver Mead for foundational insight and many helpful conversations over the years, Evelyn Smith and Matt Au for research support, Guro Ekrann for expert editorial assistance, and Jeff Eisenach for helpful suggestions.