Moore’s Law:  
A 50th Anniversary Assessment

“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.”

The article, “Cramming More Components Onto Integrated Circuits,” appeared in the 35th anniversary issue of *Electronics* in April 1965. “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. A few years later, he and several members of the team would leave to found Intel.

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 U.S. economic, cultural, and military power.

“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."

Moore’s 10-year projection was not only correct but 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. Mead dubbed the phenomenon Moore’s law, and exponential chip scaling has continued, more or less, for 50 years.

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. 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. They think post-War America was a zenith of innovation and income growth and that info tech 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 pauper prospects well into the future.

Others argue just the opposite — that information technology is so dangerously powerful that robots will steal most jobs. If, that is, artificial intelligence doesn’t wipe out the human race first.

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 won’t 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.

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.

*Processing Power* — Around the year 2000, a now-famous chart began showing up at technology conferences everywhere. Moore’s law was going strong.\(^{10}\) It 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.\(^{11}\) This temperature trend was clearly unsustainable. Engineers knew for at least the next decade they could keep packing more transistors in smaller areas. But they couldn’t continue to boost clock frequencies, the number of times per second a chip’s transistors turn on and off, which had been a key source of performance improvements. Because power density is a function of clock speed and voltage, increasing clock speeds requires reduced 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, clock speeds leveled off at 2-3 gigahertz (GHz), and around this time, some said Moore’s law was ending. By some measures, it appeared true. The National Academy of Sciences in its comprehensive 2011 analysis even referred to the situation as “the crisis in computing performance.”\(^ {12}\) NAS concluded that “After many decades of dramatic exponential growth, single-processor performance is increasing at a much lower rate, and this situation is not expected to improve in the foreseeable future.”\(^ {13}\) Yet transistors kept shrink-
Today’s leading edge process is 14 nanometers (nm), and the International Roadmap says we are headed toward 7 nm. And something else happened: overall chip performance kept getting better.

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 the SPEC CPU2000 and CPU2006 benchmarks, economist Stephen Oliner and colleagues show that total CPU performance 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 1971-1990 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, Oliner 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 total computing power of 6.14 teraFLOPS, which alone would have qualified it as the world’s fastest supercomputer until the year 2000. GPUs are increasingly used in cluster- and super-computers and are being applied to a range of more general purpose applications in the form of GP-GPUs.

Memory and Storage — Moore’s law also revolutionized memory and data storage. The nearby photo shows a 1956 IBM 305 RAMAC storage system with a capacity of 5 megabytes (MB) being loaded with a forklift onto an airplane. Today, a 128 gigabyte (GB) flash drive the size of your fingertip stores 25,600 times more information and costs around $40. New 3D memory technologies from Intel and Micron will soon yield a 3.5 TB flash drive and a tiny SDD 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 non-leading edge end of the spectrum and across a range of attributes. It provides for extremely inexpensive embedded chips — chips we never see and rarely think about — in automobiles, for example, which now contain more than 100
microprocessors and microcontrollers, in running shoes, toothbrushes, thermostats, and all kinds of consumer and industrial devices. Moore’s law also provides the tools to build extremely low-power chips and components (like MEMS accelerometers) for untethered devices like mobile phones and for RFID tags, which are manufacturable at just a few pennies a piece.

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 nanotech innovations in materials, manufacturing, and design. Increasingly, the specialized sensor designs are biologically inspired and highly parallel — what Mead in the early 1980s called neuro-morphic technology.21 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 (most of its devices contain a front and rear camera). That was twice the total of all stand-alone cameras sold worldwide in the record year of 2012. With nearly every mobile phone now containing a camera, if not two, and with increasing deployment of webcams and security systems, 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, or the emerging Internet of Things, is not nearly as dependent on continued growth of top line computer speed, as was PC performance in the 1980s and 90s. But these sensory, embedded, and low-power applications nevertheless vindicate Moore’s original goal — unimaginably huge volumes at very low cost.

The Economic Effects

Seventy-three private firms, according to The Wall Street Journal, are now members of “the billion-dollar startup club.” Fifty of these start-ups are American, and a number of them have recently achieved valuations of $10, $20, even $40 billion. The total value of the 50 US club members is $223.9 billion and does not include the 10 club members that went public or were acquired in 2014. 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 $2.25 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.22

Yet measuring the economic impact of information is difficult. Indeed, it may be the central question in all of economics right now. We know the advance and spread of information technology is powerful. By some intuitive measures, it seems almost miraculous. By many orthodox measures, however, it seems tepid. We encountered this paradox in the 1980s when Robert Solow famously noted that “You can see the computer age everywhere but in the productivity statistics.”23 Computers finally flowed through to the 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 economist Tyler Cowen ignited the debate with his book The Great Stagnation, which argued that the Internet was a great source of “cheap fun” but not jobs or incomes.24 Paul Krugman in 1998 had famously said by 2005 we would view the Internet as no more important than the fax machine.25 With “The Demise of U.S. Economic Growth” in 2012, Northwestern’s Robert Gordon drew an even darker picture. The First and Second Industrial Revolutions, Gordon argued, were far more potent than the Third (the Information Revolution) and, more importantly, that the Information Revolution may already have ended. The first two Industrial Revolutions account for the sharpest leap of living standards in history, and so it will be difficult for any economic era ever to match their import. But Gordon ar-
gues 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 its effects across the economy.

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 in 2014) but is also the basis for increasing portions of 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 telecommunications. Software is “eating the world,” in the famous phrase of venture capitalist Marc Andreessen.26 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.

Some have pointed to a slow-down in the chip price declines we had become accustomed to and which helped fuel the productivity surge in the 1990s. Microprocessor prices from 2000 to 2013, according to the producer price index (PPI), fell at an annual rate of 28 percent, a fast pace for almost any other product but slower than the historical rate. After 2008, prices fell by just 8 percent per year, a dramatic and seemingly ominous slow down. Oliner, however, found something curious. Intel, which makes up nearly half the U.S. 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).27 This superficial change appears to mask continued rapid price drops. After adjusting prices for real chip performance and to avoid Intel’s list price anomaly, Oliner and his colleagues estimate that the actual annual price drop for 2000-13 was 44 percent.28 This meant that according to the official government measure, $100 worth of computing power in 2000 could be purchased for $1.40 in 2013, which sounds impressive. Oliner, however, shows that the actual cost in 2013 may be just 5 cents ($0.05). According to this measure, consumers can purchase 28 times more computing power per dollar than the official data suggests.

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 that:

• 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).29 Today, more than two billion people own smartphones, what some have called “supercomputers in your pocket.”

• In 1990, Internet traffic totaled one terabyte (TB) per month, about as much data as fits on a PC desktop hard drive.30 In 2015, Internet traffic totals around 75 million TB per month.31

• The smartphone-cloud combination created a whole new software industry — mobile apps. After the 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 is just $5,000 and is rapidly headed toward $1,000.32 (We may not yet have seen much resulting innovation, 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 to 2006.33 Over this period, 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 vast bulk of progress coming after World War II. Extrapolating from Nordhaus’s figures, which end in 2006, we 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 Jorgensen 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 non-farm 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 post-War boom, but it’s no fault of IT.

The notion that IT has not produced gains for the middle class is mistaken for at least two big reasons. First, the base argument of middle class stagnation is 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. 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.

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 U.S. incomes between 1967 and 2014 shows a near catastrophic rise of just 4.2%. 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%. 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 information technology. How much, for instance, is Wikipedia worth? In how many ways do time-saving information tools improve our non-work 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.

It is true median incomes have stagnated since the 2008 financial crisis, and employment trends have deteriorated, but these serious problems likely reflect demographics and an anti-growth policy environment, not a failure of IT.

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 upon an environment of freedom where firms and entrepreneurs could both iterate rapidly and also take big risks on unproven technologies and business models. For example:

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 multi-billion 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 upon the same virtuous circle among broadband, app, device, cloud, and content firms.

Robert Graboyes has offered a particularly vivid description of the innovation gap between a highly regulated sector like health care and a mostly unregulated sector like IT:

In the same years that IT exploded, changing how billions of people live their daily lives, health care was painfully slow to innovate.

Step into a time machine and travel back to 1989. Gather a group of people and tell them of the advances that medical science has made in 25 years — statins, new vaccines, face transplants, and so forth. The audience will be pleased and gratified by the news, but there is little that will shock them.

Now tell them the following story:

“While camping high in the Rockies, Efram signed and deposited his paycheck in his bank account. Then he purchased The Complete Works of Shakespeare and read Macbeth. A bit later, on YouTube, he watched the Beatles sing ‘Yellow Submarine.’ Using Google Translate, he converted the lyrics into Hindi and then Skyped his friend Arjun, who is working at McMurdo Station, Antarctica. Efram sang his translation to Arjun, who grimaced, but then commented on the beauty of the towering mountain behind Efram. After hanging up, Arjun emailed a restaurant in Denver (a city he has never visited), and an hour later a drone delivered Indian food to Efram’s campsite—all paid for with bitcoins. While eating his tikka masala, Efram toured McMurdo Station via Street View and asked Siri for the current temperature there. ‘Brrrr. It’s 10 degrees below zero Fahrenheit, Efram,’ she answered. Then he accessed Netflix and watched Seven Samurai before dozing off to a selection of Malian jazz, courtesy of iTunes Radio. The entire cost of this sequence of events was $34.77 — $0.99 for the Kindle edition of Shakespeare, $2.00 for the film, $26.78 for the food, and $5.00 for the drone delivery service. And the whole set of interactions required only
Efram’s iPad and Arjun’s cell phone — the two devices together costing less than $1,000.”

Now, your audience will assume you are lying or delusional. And yet to our 2014 eyes, every step of this story is mundane and familiar — except for the drone delivering dinner to the mountain. And drone deliveries are perfectly feasible; it’s just that drones are the only part of the story that remain [regulated].

Information technologies enabled the globalization of trade and capital markets, which modernized many developing economies. Most directly, the Internet and smartphones in the hands of individuals provided 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.

As free enterprise policies helped Moore’s law prosper, Moore’s law also spread freedom around the world. Although a number of factors were in play, it is hard to ignore the fact that during Moore’s law’s run, global poverty dropped from over 60 percent in 1965 to just 16 percent in 2011. Information technologies helped achieve this expansion of the world’s middle classes in a number of direct and indirect ways. First, the U.S. lead in computers helped it win the Cold War, unleashing wave of political freedom. China, anticipating these changes, chose in the late 1970s to free its economy and focus, like nearby Taiwan, on the electronics industries.

Today, the single 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.”

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,” 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 nano-electronics, including carbon nanotubes (CNTs), graphene, and even diamond.

New ways of storing, reading, and manipulating information itself 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 don’t 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 processor improvements, they 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.”

Carver 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.48

The vision was compelling. But the complexity of 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. It 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 and storage. Amazon Web Services, however, is the largest provider of cloud services and now employs some two million computers in its cloud offerings alone.49

In keeping with Moore’s original vision of radically low cost computing, the pricing of these cloud services is often measured in pennies.50 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.”51 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: the massive, centralized air conditioning operations that cool servers in our data centers in effect remove heat from our mobile devices.52

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 at 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.53

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.

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.54

The automobile turned out to be an engine of the “American Century.” At the outset, however, it wasn’t 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 are also fearful of 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 overdone. The two pessimistic views — that technology is either impotent (Gordon’s demise of growth) or dangerously dystopian (runaway robots and AI) — are relatively unlikely. Yes, some technologies won’t 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, we believe information technology has only begun its vast impact on every existing and new industry. Far from ending, we are only at the end of the beginning of information technology. The case for “rational optimism” is strong.55

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, until recently medicine 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.”56 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. An information based medicine will also provide for diagnostic “sniffers” — molecular sleuths meandering through our bodies and smartphone apps gauging 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, and the environment. Pooling and then analyzing data from millions of subjects will yield new insights. For information-based medicine and health care to impart its most powerful economic benefits, however, will require a rethink of 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.57

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 of 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’s a 60 percent differential. If health care could escape its cage of hyper-regulation and truly exploit information science, this bloat industry could turn into a productivity growth industry.58 At one-sixth of the economy, such a transformation could substantially improve the prospects for overall economic growth.

Gordon and other economists correctly note that unfavorable demographics will limit one important source of economic growth over the next several decades. But another factor, which most view as an additional suppressant of growth, doesn’t have to be.

Economists are pessimistic about the almost certain slow down in educational attainment.59 They say that formal years of education has already leveled off and that it 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 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 due to 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 — has 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 don’t require bleeding edge processing speeds.

New computer architectures, which emerged in the last decade, will continue 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, the combination 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 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 encourage the spread of this policy to other sectors of the economy.61

“It’s extremely important,” Carver Mead summed up, “that there’s a living example of people’s belief in the future, bringing it to pass.” Moore’s law made the future “tangible.” “Because that’s really down deep what Moore’s law is about. If you don’t have that belief . . . it won’t happen . . . . Semiconductors are not the only arena where learning curves of that sort work. And so I think it’s not amiss to take that belief system into other parts of our industry.” EE
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; and George Gilder and Carver Mead for foundational insight and many helpful conversations over the years.

Selected References


White House report PCAST NITRD https://m.whitehouse.gov/sites/default/files/microsites/ostp/pcast-nitrd-report-2010.pdf


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.”


10 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 is the number of components that minimizes the cost?) By the 1980s, better designs had mostly eliminated wasted space, and by the 1990s chip sizes weren’t 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.

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


13 Ibid. p. 10.

14 For one measure of transistor scaling, see the Stanford CPU database (CPUDB), from which Figure [X] 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.

We have elsewhere 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, etc.). See “Into the Exacloud.” Entropy Economics. November 21, 2011. http://entropyeconomics.com/wp-content/uploads/2011/11/Into-the-Exacloud-21-Nov-2011.pdf


Again, Moore’s projection was for the number of components per unit area, 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.


Paul Krugman. “Why most economists’ predictions are wrong.” Red Herring. “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.” June 1998. Access at http://web.archive.org/web/19980610100009/www.redherring.com/mag/issue55/economics.html


See Oliner 2015 at p. 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.”


Data for 1990 comes from Andrew Odlyzko of the University of Minnesota (MINTS — Minnesota Internet Traffic Studies), and can be found in Bret Swanson. “Into the Exacloud.”
31 See Cisco Visual Networking Index.


34 See Tables 6 and 7 in Nordhaus and the data appendix. We 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, which approximates Moore’s paper in 1965.


40 http://www.pbs.org/newshour/making-sense/need-liberate-americas-health-care/


42 Moore and Mead interview. Minute 55:00.


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

51 I first heard it described this way by George Gilder.
Again, Mead told us 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 Cisco Cloud Index.


See Adam Thierer. “Permissionless Innovation.”