

Kurzweil, Moore, and Accelerating Change

Ilkka Tuomi

Visiting Scientist

Joint Research Centre, Institute for Prospective Technological Studies

Working paper, 27 August 2003

Abstract

Ray Kurzweil has advanced an influential hypothesis that computing technology has developed at accelerating speeds during the last century, and that the increasing rate of technical change will lead to profound changes in the society and human life. The paper focuses on a review of empirical evidence for this claim, focusing on Moore's law, which in its simplest form says that the number of transistors on semiconductor chips has been growing exponentially. The paper also reviews evidence about accelerating returns in the semiconductor industry, and discusses the conceptual framework that underlies Kurzweil's argument that technological development can be understood as an evolutionary process. We find that the available empirical data do not support Kurzweil's hypothesis. The underlying conceptual assumptions have, however, broad relevance for students of technological change.

Introduction

Although the overwhelming speed of change has been commented at least since the Medieval Age, it has become a central part and a defining characteristic of the modern society. Science and technology have frequently advanced at dizzying speeds, creating both enthusiastic responses such as the futurist movements of the early 20th century and more pessimistic interpretations of the imminent destruction of values and the foundations of society. An influential analysis of technical progress was crystallized in the Communist Manifesto, which claimed that “all that is solid melts into air.” Marx and Engels argued in 1848 that:

“The bourgeoisie cannot exist without constantly revolutionizing the instruments of production, and thereby the relations of production, and with them the whole relations of society... Constant revolutionising of production, uninterrupted disturbance of all social conditions, everlasting uncertainty and agitation distinguish the bourgeois epoch from all earlier ones. All fixed, fast-frozen relations, with their train of ancient and venerable prejudices and opinions, are swept away, all new-formed ones become antiquated before they can ossify. All that is solid melts into air... The bourgeoisie, by the rapid improvement of all instruments of production, by the immensely facilitated means of communication, draws all nations, even the most barbarian, into civilization.”

Such poetic expressions, of course, could not have gained credibility without some psychological and empirical support. Quantitative studies on change have provided additional support in recent decades. In particular, the idea that science and technology expand exponentially originates with the studies by Derek J. de Solla Price, at the end of the 1940s. According to Price (1986), the number of scientific journals had doubled about every 15 years since 1750, the number of “important discoveries” had doubled every 20 years, and the number of U.S. engineers about every 10 years.

One reason for the increasing number of engineers and scientific articles is the fact that during the last millenniums the world population has grown with accelerating exponential speeds. Estimates for the world population in 6500 B.C range from 5 to 10 million people. At 1 A.D., the population had grown to 300 million. It reached 1 billion in 1804, and 2 billions 123 years later. The next billion was added in 33 years, and 5 billion was reached 13 years later, in 1987 (UN, 1999).

Historical data shows that global production has grown systematically faster than global population. This means that humans have been able to increase their effectiveness and efficiency in producing goods and services. The speed of growth accelerated after the 15th century, increased impressively during the 19th century, and skyrocketed in the 20th century. Knowledge and technology have clearly been important factors in this process.

During the last decade or two, probably the most influential statement of rapid technical progress has been Moore's Law. Moore's Law combines two fundamental dimensions of the story of long-term economic growth. Semiconductors show impressive technical progress in quantitative terms. Semiconductors, however, are also essential components in information processing and communication systems. If economic growth is to a large extent determined by knowledge creation, the rapid changes in semiconductor technology perhaps not only reflect technical progress but may also become drivers of it. In other words, semiconductors could well be a source of a new mode of economic growth.

It is therefore important to understand the nature of change in semiconductors and information processing. Furthermore, we need to understand their relationships to economic and social transformation. Kurzweil's hypothesis provides a useful starting point here.

Kurzweil's hypothesis

Kurzweil (1999; 2001) has proposed that technical progress can be characterized as accelerating exponential development. According to Kurzweil's "law of accelerating returns," technical change is generated in an evolutionary process where the outputs

of the process are used as inputs in the next phase of the development. This leads to exponential growth. Kurzweil maintains that the rate of exponential growth itself increases. When a particular evolutionary process becomes more effective than its alternatives, greater resources are deployed for the further progress of the effective process. As a result, the rate of exponential growth itself grows exponentially. Evolution, and technology—evolution continued by other means—therefore is a process that leads to accelerating change.

Kurzweil also argues that when a specific paradigm that provides exponential growth exhausts its potential, a paradigm shift occurs. The overall progress, therefore, consists of a sequence of hyper-exponential patterns of progress.

Kurzweil provides a selection of examples to support his argument, ranging from Moore's Law to the speed of DNA sequencing and the number of hosts on the Internet. Kurzweil points out that biological evolution is a process with accelerating returns, and that technological evolution can be understood as a continuation of biological evolution.

Using his observations as a basis, Kurzweil argues that technical development will rapidly lead to computational devices that are more intelligent than humans. The increasing capabilities of computers will further accelerate the speed of technological progress and, at least from our limited perspective, lead to an apparently infinite speed of change. Kurzweil calls this approaching point of extremely rapid change "Singularity." The approaching Singularity will transform all aspects of our lives. According to Kurzweil, the exponential growth of computing is a marvelous quantitative example of the exponentially growing returns from an evolutionary process.

Kurzweil's hypothesis of approaching "Singularity" therefore consists of four major claims. One is the empirical claim that technical developments in computing show exponential growth and that the growth speed has increased over time. Another claim is that this process is generated by a "law of accelerated returns" which feeds back resources to those areas of technology development that grow rapidly. The third claim is that technology creation can be understood as an evolutionary process. The fourth

claim is that the speed of change will soon reach levels where the human intelligence, at least in its current forms, perceive infinite pace of change and that this will lead to major changes in the society and in individual lives. I will discuss the first two of these claims below, starting with the claim that technical developments in computing demonstrate increasing exponential growth rates.

Moore's Law

In its simplest form, Moore's Law is a relatively straightforward claim, which says that the number of components on a lowest-cost semiconductor chip grows exponentially in time. Moore originally noted that the number of components roughly doubled every year during the first seven years of integrated semiconductor technology. His main argument in his famous 1965 paper was that there were no obvious technical bottlenecks that would slow down the increase in component count before 1975. In 1975, Moore revisited his analysis. He noted that the speed of technical change was slowing down, and that during the next decade the growth rate would probably be close to 24 months. The basis for this prediction was Moore's belief that the possibilities for innovation were becoming exhausted (Moore, 1975).

I have shown elsewhere (Tuomi, 2002) that the different versions of Moore's Law never very accurately described developments in the semiconductor industry. One has to be quite creative to find support for them. This is sometimes difficult to see, as a large number of authoritative sources claim that Moore's Law has been valid for several decades, that it will remain so for at least for the next decade, and that Moore himself has said so.¹

¹ The most common claim is that Moore's Law says that the number of transistors on a chip doubles every 18 months. Moore has never made such a claim. He has recently pointed out that microprocessor MIPS ratings have been growing at about 20 months doubling time (in a keynote of the 2003 International Solid-State Circuits Conference). The graph that Moore has used to illustrate this (ftp://download.intel.com/research/silicon/Gordon_Moore_ISSCC_021003.pdf) is, however, incompatible with processor data provided by Intel (for example, the 8086 MIPS rating seems to have been only about one third of what Moore uses in his graph and the 80386 MIPS rating is about 3 years off its actual value). In fact, if we use 8080 (0.64MIPS) and Pentium 120 (203 MIPS), we get a 317-fold increase in 21 years, implying a 30.3-month doubling time.

The validity of such claims depends on the exact version of Moore's Law used. Relatively small changes in the formulation of "Moore's Law" easily lead to large errors across decades. A closer look on the history of semiconductors reveals that claims related to Moore's Law typically have had problems with both historical facts and empirical data.

Apparently, Moore's Law has been widely accepted because it is intuitively right. On average, the number of transistors on semiconductor chips has grown rapidly. The speed of growth, however, has varied in time. Although it is always possible to draw an exponential curve between two points, such simple exercises hide important aspects of technical development in semiconductors. Furthermore, such exercises do not typically support any known versions of Moore's Law.

Different types of semiconductors have very different patterns of complexity growth. It is, however, possible to argue that a fundamental bottleneck in this process is technical and engineering innovation and that an exponential trend defines a technical trajectory that is independent of any external factors. Moore's Law, in its original form, is basically such a claim. Its exponential form implies that—at least for the time being—developments in integrated circuits are effectively independent of economic, organizational, social, or any other forces. Whatever they are, we do not need to care. We may, of course, test this hypothesis also directly.

As Moore noted, on average integrated circuit component counts grew rapidly during the 1960s, almost with a one-year doubling time. During the first decade of microprocessors, the transistor counts grew at about 22 months doubling time, when measured using a least squares trend, which slowed to about 33 months during the following decade. During the 1990s, transistor counts grew at varying speeds. In the Intel's Pentium chips, the transistor counts grew with around 54 month doubling time. After that the transistor counts grew very rapidly, partly because large amounts of memory were added onto microprocessor chips (Tuomi, 2002b).

Alternative growth rate estimates can be based simply on transistor counts on representative microprocessors introduced at two points of time. If we use the first

microprocessor, Intel 4004, as a starting point, the exponential growth time for the 1971-82 period is 21 months, for the 1971-91 period 26 months, and for the 1971-98 period 27 months.² This calculation indicates a slowing down in the component growth rate. If we separately calculate the growth rate for the 1982-91 period, it is about 35 months, and for the 1991-98 period somewhat over 30 months. It therefore appears that during the first decade of microprocessors component counts increased much more rapidly than during the 80s. In the 1990s the growth rate was faster, 30 months for both the first half and the whole 1990-1998 period, but considerably slower than during the first decade of the microprocessor history.

These two-point calculations are the simplest possible estimates for growth rates and probably resemble some of the calculations that have led to the popularity of Moore's Law. One should note, however, that they start from the assumption that the growth has been exponential. A simple calculation of the growth rates in different time periods shows that the exponential assumption is not historically valid.

The International Technology Roadmap for Semiconductors (ITRS, 2001: table 1d) uses a 36 month doubling time for microprocessor transistor counts for the next years and notes that there was a period of faster growth in 1995-2001.³ A more detailed study shows that microprocessor families have developed in complex ways, which are difficult to describe by using only technical parameters.

Many commentators of Moore's Law have used it to describe technical progress and therefore they have focused on chips that contain the largest numbers of transistors. As memory chips are not in any obvious way connected with progress in computing power, the most advanced microprocessors have commonly been used to support Moore's Law. It is not exactly clear, however, what criterion we should use to

² Data for these calculations come from Intel's Microprocessor Quick Reference Guide, <http://www.intel.com/pressroom/kits/quickrefyr.htm>.

³ One should note a fundamental difference between Moore's original observation and ITRS forecasts: Moore's Law was explicitly based on the observation that there were no known technical obstacles for the predicted development. ITRS, in contrast, specifically points out that there are no known solutions for achieving the roadmap milestones. Indeed, ITRS talks about a "red brick wall" which the industry will hit if it is not able to come up with radically new ideas.

measure advances in microprocessors. There are many trade-offs in microprocessor technology, such as those between clock speed and power consumption, that are valued differently in different applications. Technical characteristics such as transistor count may therefore be evaluated quite differently in different product contexts. For example, manufacturers of portable machines are interested in power consumption and may therefore prefer chips with lower clock speeds and transistor counts. Indeed, in such applications, fewer transistors could sometimes mean progress.

In particular, semiconductors in consumer electronics and telecommunications devices, cars, measurement instruments, and other similar uses, rarely have the largest transistor counts or state-of-the-art technical performance parameters. The main reason is that in such applications costs are counted with care. Many application specific semiconductor chips and microprocessor chips have an order of magnitude lower component complexities and technical parameters than state-of-the-art chips. Most chips are also not manufactured with the latest manufacturing technologies. This can be seen from Figure 1, which shows the global volume of metal-oxide semiconductor (MOS) production in the first quarter of 2003. The newest manufacturing technologies used a 130-100 nm feature size (0.13-0.10 μm).

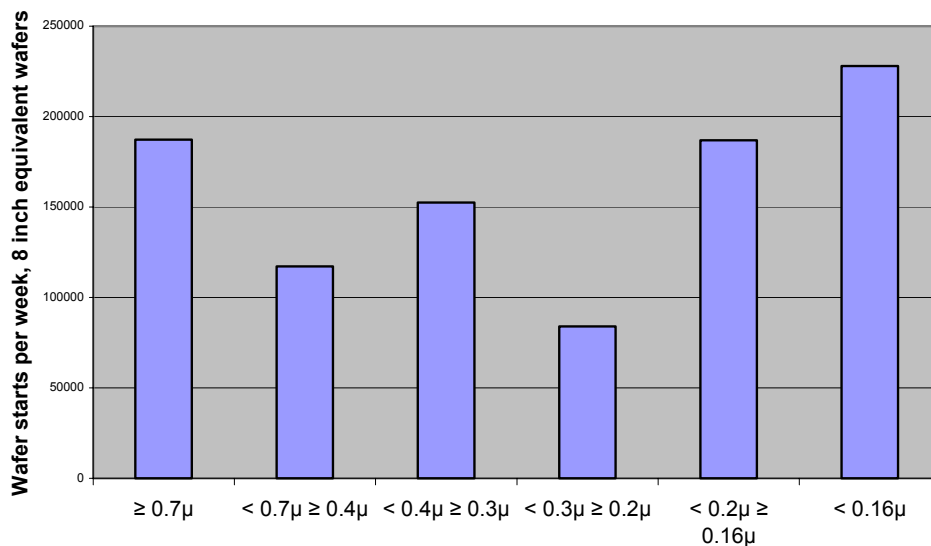


Figure 1. Global metal-oxide semiconductor production by manufacturing technology, 1Q/2003. (source: Semiconductor International Capacity Statistics).

Computer markets are an exception here. In particular, personal computers are aggressively marketed by using technical characteristics such as processor clock-speed, memory size, and novelty value of their processors. This has led to extremely rapid rates of obsolescence and short expected lifetimes of computer investments. In such a hyper-deflation environment, economic decision-making easily breaks down. An important role is played here by the common belief that semiconductor technology progresses at exponential speed, popularized as Moore's Law.

Accelerating exponential growth?

Transistor counts and the number of instructions per second are purely technical characteristics. Kurzweil, however, has argued that cost per computation estimates show accelerating returns and hyper-exponential declines. According to Kurzweil, Moore's Law represents the fifth paradigm of computing technology, and the rate of progress in computation has increased since at least the first decade of the 20th century. The number of computations per unit cost doubled every three years between 1910 and 1950, doubled every two years between 1950 and 1966 and after that it has doubled yearly (Kurzweil, 1999).

Most existing references on Moore's Law point to his 1965 paper, where Moore noted that the number of components double roughly every 12 months. In 1975 Moore discussed the reasons why the speed of development is slowing down. Indeed, Moore argued that the speed of exponential growth was about to drop to a half of what it had been during the first two decades of semiconductor development. This would apparently contradict Kurzweil's claim of accelerating speed of change.

Computational power has often been described using the number of instructions that a computer can handle per unit of time. A common unit has been MIPS (millions of instructions per second), which in historical computers was directly related to the processor clock speed. The MIPS data provided by Intel actually shows an accelerating trend when it is complemented with benchmark MIPS ratings for the newest processors. During the first decade of microprocessors (from Intel 4004 to 80286) the maximum MIPS ratings grew 15 fold, during the next decade 45 fold, and during the last decade 100 fold. The last number would in fact correspond to the

commonly mentioned 18 month doubling period. One should, however, note that the MIPS ratings for the last decade are theoretical and do not correspond to the number of instructions that a processor can handle. Although the name MIPS has remained the same, its meaning has changed. The MIPS ratings have become essentially a marketing device, and this change became particularly important in the 1990s.

It is difficult to verify Kurzweil's calculations for the full period he is describing, as reliable data is difficult to find for historical computers.⁴ For recent periods, this is somewhat easier. In current dollars (not adjusted for inflation), the median price of desktop computers sold in the US in the years 1976, 1979, 1980, 1994, 1995 and 1997 was almost exactly the same \$2000 (Berndt, Dulberger, & Rappaport, 2000: table 1). If we use the MIPS ratings for Intel 80286 (0.9 MIPS; released in February 1982) and Pentium 133 (219 MIPS; released in June 1995), the MIPS compound annual growth rate was about 51 per cent. This would correspond to a doubling rate of 20 months. If we subtract the effect of general annual inflation of some 3.5 percent, the cost per MIPS would seem to have halved in about 19 months in the 1982-95 period.⁵

⁴ For example, Ceruzzi (1998:71,74) gives \$1.6 to about \$2 million as the price of a full IBM 7090 installation. Kurzweil uses \$3 million. Kurzweil has also moved Babbage's Analytical Engine about half a century in time, with the explanation that it probably could have been built in 1900. Other authors have argued the machine could have been built using available manufacturing capabilities, but the project did not succeed as Babbage kept changing his designs. Historical data also reveals that the early computers rarely were working at their theoretical speeds. For example, the mean time between failures of IBM 701, introduced in 1952, was about 20 minutes (Ceruzzi, 1998:35). In modern computing, on the other hand, idle times of more than 90 percent are not uncommon.

⁵ To avoid confusion, it is useful to note that commentators of Moore's Law frequently describe geometric growth ($x_t = x_{t-1} * r$) as exponential growth. Annual compound growth rate (ACGR) of about 69 percent is sufficient to produce doubling in a year, whereas geometric growth with yearly compounding obviously requires a 100 percent annual rate. The growth rate of 51 percent is calculated as annual compound rate in the geometric sense, i.e., with yearly cumulation. Geometric growth equals exponential growth only when the time period of geometric growth steps becomes infinitely small. The doubling and halving times are also not symmetrical in geometric growth or decline. Annual decline of 29.5 percent leads to halving in two years ($0.705 * 0.705 = 0.497$). Doubling in two years requires 41.5 percent growth ($1.414 * 1.414 = 2$). When we talk about growth, we are climbing up the relatively slowly upward curving slope of an exponential curve; when we talk about decline, we are falling down the steep slope of the same curve.

The above calculation, of course, is a very rough estimate as it simply uses median desktop computer prices, general inflation and two processors to estimate the MIPS growth rate. Broader measures for computing costs that adjust for quality improvements typically have estimated prices to drop in the 18-32 percent annual range (Landefeld & Grimm, 2000). This would imply a 50 to 30 month doubling time for “computing” in constant prices. Quality adjusted price indexes used in the US National Income and Product Accounts show average annual price index declines of 24 percent for computers and peripheral equipment in the 1995-1999 period, when prices dropped especially rapidly (Landefeld & Grimm, 2000). This would correspond to a doubling time of 39 months. For the longer 1985-1994 period, Intel 80x86 microprocessors show an average annual decline of 27 percent (Grimm, 1998), corresponding to a 35 month doubling time. For Motorola’s 680x0 processors the average decline is 23 percent. This last number corresponds to a 40 month doubling time.

The last half of the 1990s was a unique period in the semiconductor and computer industry. In 1995 and 1996 quality-adjusted Intel 80x86 microprocessor prices dropped over 60 percent per year (Grimm, 1998). Also DRAM prices fell fast. On average, prices per unit of memory have declined 32 percent per year during the 1978-2000 period (CBO, 2002:18). This corresponds to a 30 month doubling time. The DRAM prices were almost unchanged during the first half of the 1990s, but dropped then like a stone. An important factor was the Asian currency crisis, which created competitive pressures on the market. PC makers responded by bundling their machines with large amounts of DRAM (CBO, 2002:19). Competitive pressures led also to accelerated technology development cycles and product introductions. This became recorded in the price indexes as quality increase. The end results was that the quality adjusted PC prices also seemed to drop extremely rapidly. This price drop did at least partially come from declining profit margins in the computer and semiconductor industry.⁶ In fact, Intel’s margins in its microprocessor business

⁶ The profits of US electronic and other electric equipment industry dropped about \$19 billion between 1997 and 1999, and about \$6.5 billion between 1999 and 2001, according to BEA data. Earnings growth in the semiconductor industry started to decline at the end of 1995, rebounded slightly in 1997, and fell then again until the end of 1998 (Svoren & Khan, 2002). Since then, price developments have

declined from about 90 percent in 1993 to about 73 percent in 1999. In 1998, when the Asian economic crisis had its greatest impact, about 30 percent of the price declines in Intel's microprocessors can be attributed to falling margins (Aizcorbe, 2002).

The full cost of computing consists of computing equipment and other complementary costs. In modern computing environments, these obviously include software costs, but also system maintenance, cooling and heating, physical security, training, debugging, and other costs. When the evolution of computing costs is studied across decades, it is important to account for all the relevant costs. For example, accounting only for the cost of a single transistor or a semiconductor chip would clearly be an inadequate measure of the cost of computing.

The cost of computation is therefore only partially associated with microprocessor costs. About one third of semiconductor costs in computers are microprocessor costs. About another third is memory, about 10 percent are microperipheral chips, about 9 percent other logic chips, and about 2.6 percent microcontroller chips. During the 1990s, the quality-adjusted prices of these different types of chips have declined at different rates. According to Aizcorbe et al. (2002: table 1) microprocessor prices declined at about 52 percent annual compound rate. For memory the rate was about 31 percent, for microperipherals about 14 percent, for other logic chips about 13 percent, and for microcontroller units about 7.5 percent. The associated doubling times range from 16 months to 111 months.

A microprocessor chip with a given MIPS rating does not do any calculations even after it is connected with other semiconductors. It has to be connected to a power supply and other components before it becomes functional. In fact, according to the official US estimates, semiconductors account for only 14 percent of the value of computer shipments.⁷ If one studies the share of inputs that are used computer

certainly been influenced by market conditions. The projected losses of the five leading Japanese semiconductor manufacturers were about \$11.4 billion in 2002. The ITRS 2002 update noted that the industry struggled "through the worst recession in its history."

⁷ This Census Bureau estimate is probably too low. It includes only those semiconductors that are reported as specific components. Alternative estimates put the semiconductor content to 30 or 45 percent of the value of computer systems (CBO, 2002: 20-23). According to the World Semiconductor

manufacturing in the US, one can see that the biggest input cost is associated with wholesale trade (about 14 percent of total output) and semiconductor devices (also about 14 percent). This is followed by payments for other electronic component manufacturing, software publishers (about 9.5 percent), computer storage device manufacturing, computer peripheral manufacturing, internal sales in the computer industry, and management consulting services. A more detailed study reveals that computer manufacturing requires such inputs as air transportation, computer terminals, sheet metal work, and food services and drinking places.

The average price changes in the 1990s are greatly influenced by the extremely rapid drop of prices in the second half of the decade. For example, the semiconductor input prices in the computer industry dropped over 40 percent annually during 1995-99. In the first half of the decade, however, they declined only 11 percent annually (Aizcorbe, Flamm, & Khurshid, 2002: table 2). If we assume that factors such as the Asian currency crisis, declining profit margins and the Internet stock market boom led to unsustainable growth rates during the latter half of the 90s, the 11 percent growth rate might actually be closer to sustainable rate in the near future.⁸ If the semiconductor costs would drive the declining cost of computing, this would imply a doubling rate of computations per constant dollar of about six and half years. In the communications industry, the growth rate in the first half of the 1990s was much slower, about 3.33 percent. This would lead to doubling in 22 years.

These estimates do not support Kurzweil's claim that the number of computations per constant dollar would have recently been doubling yearly. The data also show that the rate of price change has not been stable. Quality-adjusted computer and microprocessor prices dropped extraordinarily rapidly during the last years of the 1990s. This rapid decline seems to have resulted from factors such as the Asian currency crisis, decline in industry margins, scale advantages from the Internet boom,

Trade Statistics, semiconductors represented about 22 percent of computer value in 1998, and 26 percent in 2000 (c.f. Aizcorbe, Flamm, Khurshid, 2002: table A7).

⁸ Quality-adjusted price indexes take into account improvements in MIPS ratings but also other chip performance characteristics, such as on-chip memory sizes, multitasking functionality, and external bus widths. The price indexes, therefore, decline faster than indexes that would be based simply on MIPS counts.

and architectural changes in microprocessors during this time. A reasonable estimate for the average annual decline in quality-adjusted computer prices is probably about 18-30 percent during the last couple of decades, which corresponds to 2.6 to 4.2 year “doubling times.” If the semiconductor price change rates of the early 1990s are closer to currently sustainable rates, the annual price declines might be somewhere between 11 percent and 30 percent, leading to constant dollar “computing” price declines that would be slower than the ones Kurzweil has given for the early 20th century. It therefore also seems that Kurzweil’s claim about the accelerating speed of progress in computing would not be supported.

As I have pointed out elsewhere (Tuomi, 2002b), the economic extensions of Moore’s Law are conceptually independent of its original formulation. Although economic considerations were built into the original formulation of Moore’s Law, they appeared only as a criterion for lowest current cost of manufacturing. This is essentially an optimization criterion, which can be handled without reference to time or the evolution of prices. Extensions of Moore’s Law that make claims concerning the evolution of prices or costs have the additional challenge of defining what is the value of money at different points of time. This is a theoretically difficult task. It means, for example, that improvements in semiconductor technology have to be discounted when we compare the value of money across time. This implies that the value of money becomes dependent on exactly those technical developments that we try to value. It is therefore not at all clear what we mean by “constant dollars” in such contexts.

Positive returns?

Kurzweil’s hypothesis posits that the speed of progress accelerates because resources are increasing deployed to improve the most effective technologies. In the case of the semiconductor industry we saw above that the speed of progress radically slowed down in the mid-1970s, and that it has varied across time. But to what extent new resources have swarmed to semiconductor and computer development?

In the US, the investment in computers and peripherals has grown considerably during the last four decades. In 1965, roughly 0.1 percent of the GDP was invested in them. The share of investments in computers and peripherals peaked in 2000 at close to 1

percent, dropping then back to about 0.7 percent of GDP in 2002.⁹ Their share from total nonfarm business output was about 0.98 percent from 1973 to 1995, 1.06 percent during the 1991-1995 period and increased to 1.22 percent during the 1996-2000 boom. The worldwide growth of information and communication technology markets reached its peak in 2000 when the size of the market grew almost 13 percent. In 2001 the market grew about 2.8 percent and in 2002 about 1.2 percent (EITO, 2003). About 74 percent of the market in Europe was software and services.

The US private fixed investment in computers and peripheral equipment has grown relatively slowly, although in the 1994-2000 period computer investments grew rapidly. During the 1984-91 period, total annual investments in computers and peripherals did not grow much. Since the 1980, software investments have grown systematically faster than the hardware investments, and the absolute volume of software investments exceeded in current dollars hardware investments in 1990. In 2001, software investments were about \$180 billion in the US, whereas investments in computer hardware and peripherals were about \$75 billion.¹⁰

Adding up software and hardware investments, we can see that their share of fixed private investments has grown about 0.3 percent annually in the 1960-2002 period in the US. There is, however, a clear increase in the speed of growth around 1979. During the 1960s, the share of computer investments grew at about 0.3 percent annually, during the 1970s their share stabilized to around 3.3 percent, and then started to grow about 0.5 percent per year. This development of computer and software investment share is shown in Figure 2. The share of computer and software

⁹ These numbers come from Gordon, 2003, p. 20. CBO, 2002, p. 11 shows shares ranging from 1.08 percent to 1.33 in 2000 of nonfarm business sector output. Gordon, 2000, p. 51 shows a relatively constant share of about 1.3 percent of nonfarms nonhousing business GDP. The exact number depends on what is included in the measure of the total economy.

¹⁰ These current dollar values do not adjust for improvements in computer technology. Such adjustments would be needed if we were to estimate the productive value of current and accumulated computer investments. Here we do not make quality adjustments as we are interested in the amount of money that flows to computer technology producers. It is also worth noting that different methods of collecting data about software lead to different results as firms don't always account their software investments as investments (cf. Grimm, Moulton, & Wasshausen, 2002:5).

investments has grown approximately linearly as a percentage of total investments in the US.

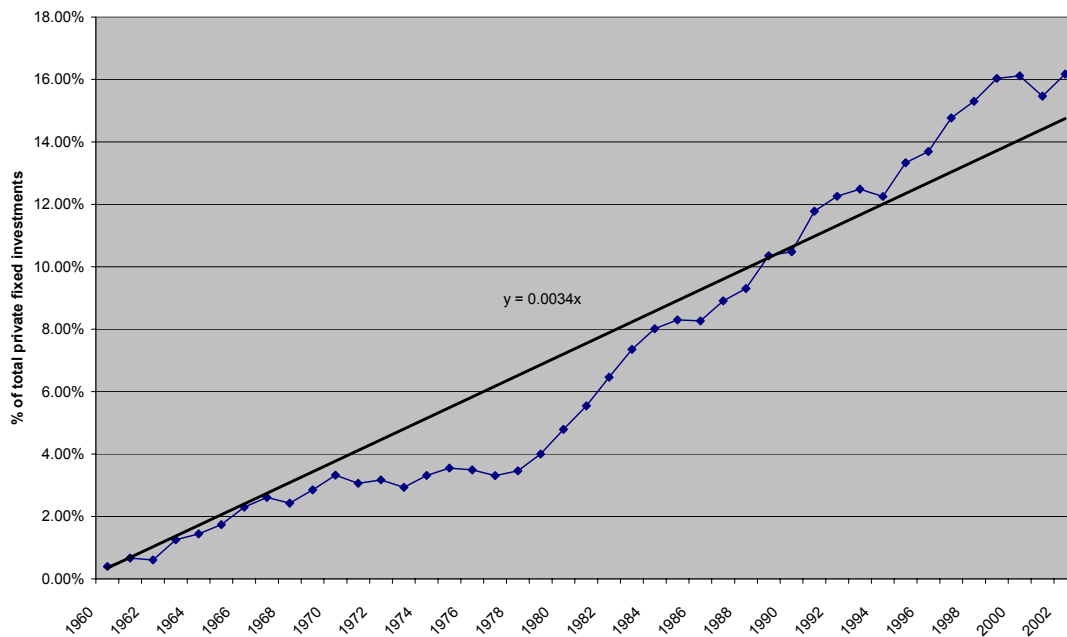


Figure 2. Computer and software investments in the US, 1960-2002 (calculated from US National Income and Product Accounts).¹¹

Technological developments in the semiconductor industry are generally viewed as the drivers of progress in computing. According to Kurzweil’s hypothesis, one would expect semiconductor industry to enjoy increasing positive returns that would speed up technical developments in the industry. Indeed, in Kurzweil’s model, the rapid technical developments would be caused by the increase in resources available for developers. The semiconductor industry, however, has not seen huge or accelerating increases in its resources. The rapid price declines in this industry have meant that although production volumes have grown rapidly, total revenues have been growing only at about 10 percent annually since 1958. The semiconductor industry is a very cyclical industry that has repeatedly been in a crisis (Langlois & Steinmueller, 1999), and there have also been several years of negative growth. For example, during the 1995-98 period the total value of semiconductor shipments decreased from \$144 billion to \$125 billion in current dollars, according to the numbers provided by the

¹¹ <http://www.bea.doc.gov/bea/dn/nipaweb>

Semiconductor Industry Association. The global semiconductor market growth in the 1982-2002 period is shown in Figure 3.

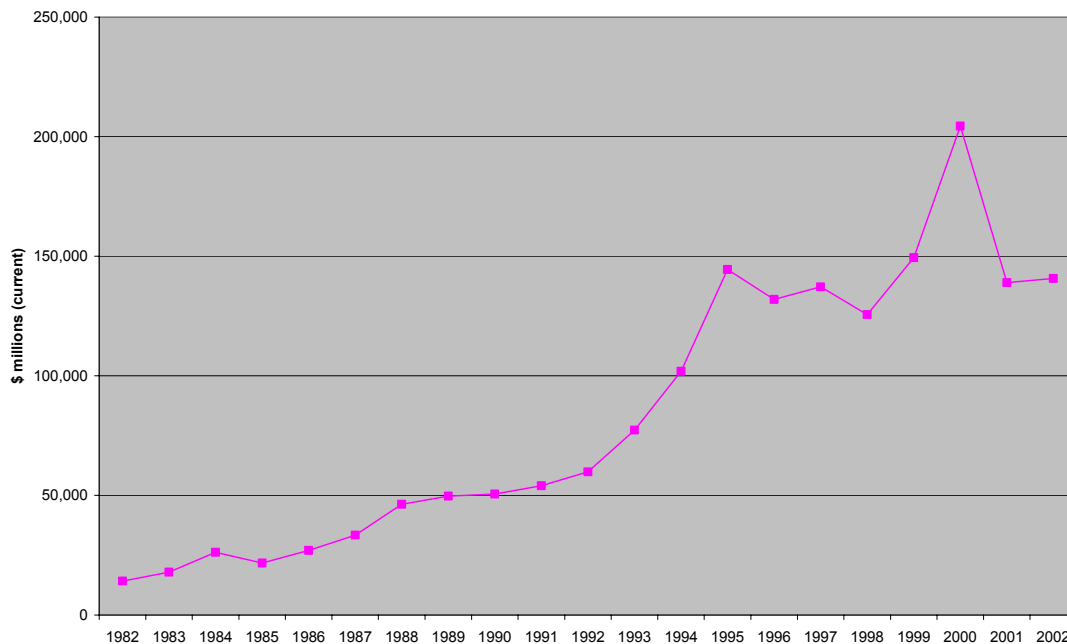


Figure 3. Worldwide semiconductor sales, 1982-2002 (source: World Semiconductor Trade Statistics).

International currency fluctuations, different inflation rates in different countries, and international transfers of intermediate products make it difficult to accurately describe the real growth of the global semiconductor markets. One way to extract the growth trend of semiconductor industry is to use annual market growth rates and try to see how much faster the industry grows than the rest of the economy. If we use estimates from the World Semiconductor Trade Statistics (WSTS), the average year-to-year change in semiconductor shipment value during the 1958-2002 period is 18 percent. This number is based on current dollars and does not take into account inflation or the overall growth of total output. If we simply adjust the growth rates by subtracting the annual US GDP growth rates from the global semiconductor market growth rates, the average semiconductor market growth in 1958-2002 becomes 10.8 percent. In this approximation, we assume that the global economy would have been expanding at the same rate as the US economy.

Calculated this way, the trend of semiconductor market growth has been decreasing

slowly during the 1958-2002 period. At the beginning of the period, semiconductor markets grew at about 20 percent annually, whereas towards the end of the period the nominal growth rates were close to zero. The annual nominal growth rates are shown in Figure 4.

The Semiconductor Industry Association currently forecasts global growth at about 10 percent rate during the next four years with rapid expansion for 2004, followed by slowing growth rates for the remaining forecast years. In 2002, the global semiconductor production was \$140 billion according to WSTS, slightly down from its 1995 value of \$144 billion in current dollars.

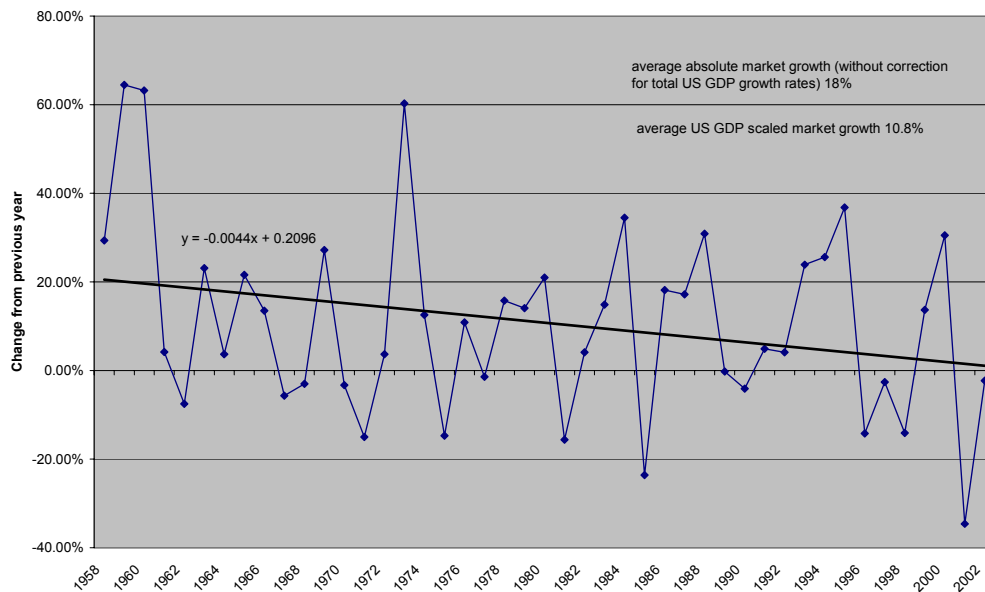


Figure 4. Annual growth of worldwide semiconductor shipments, 1958-2002, adjusted by US GDP growth (calculated from WSTS and BEA data).

It therefore does not seem that the semiconductor industry would have attracted resources at an accelerating pace. On the contrary, ITRS and the business press describe current semiconductor markets as probably the worst recession in the history. The slowdown is related to the general economic slowdown, and partly cyclical. The downturn in the economic cycle, however, is also directly related to the information and communication technology boom. From the point of view of Kurzweil's model of accelerating returns, the current post-boom cycle is part of the internal dynamic of technology development and not an external accident. The decline in resources in the computer industry also cannot be reduced to the effects of the recent downturn. As

Aizcorbe (2002) has noted, the margins of Intel's microprocessor business declined during the 1990s from about 90 percent to about 73 percent. The recent golden era of Intel was preceded by difficult times. The Japanese DRAM manufacturers grew their market share from a few percent in 1974 to about 98 percent in 1990, and in 1985 Intel's losses were more than its book value (Langlois & Steinmueller, 1999).

Progress and change?

We have discussed above some empirical aspects of Kurzweil's hypothesis. When we analyze the historical developments in computing, it is relatively easy to assess whether specific technical characteristics have followed exponential or hyper-exponential trajectories. Such an analysis becomes more difficult when we combine economic and technical arguments. Kurzweil, however, goes beyond that. His main claim is that technological evolution itself follows the logic of accelerating returns. This leads to additional and interesting challenges. What, indeed, do we mean by a generic pace of technological change? What has change to do with progress? Are there fundamental similarities between technical and biological evolution?

These questions are conceptual, but they are also critical for understanding technical evolution, development, and change. Here it is possible only to briefly comment these questions.

Kurzweil's hypothesis is about the pace of change. The world, indeed, is changing rapidly. Change, however, is a difficult concept. As Heraclitus noted over two thousand years ago, we cannot step into the same river twice. The only constant in nature is change. What, then, could we mean by the rate of change?

There is a lot of change around us, and normally we don't count it. Clouds are continuously changing their form. The ripples on a stormy sea encode huge amounts of information. Any argument about speed of change therefore has to neglect most sources of change. The selection is obviously made by relevance. When we say that the evolution is progressing at an accelerating pace, we have to abstract away all change that doesn't matter.

Which sources of change are left out of the equation depends on our present interests. If we talk about semiconductors, it really does not matter if a cloud on a sky looks like America in one moment and the African continent a minute later. A cloud is still a cloud and it has nothing to do with semiconductors.

Generic claims of the rate of change are therefore possible only if we already have made important choices about the types of change that interest us. Each qualitative claim about change, furthermore, has to define a unit of measurement and an initial value. For example, when Kurzweil describes the rate of change in computing technology, he seems to count only algorithmic calculators. In practice, other types of calculators have been widely used in the human history. At least until 1947, abacus, invented sometime between 3000 B.C. and 1000 B.C., computed faster than the most advanced current electronic computers.¹²

Many mathematical problems that require an infinite number of algorithmic computations can be solved by intelligent humans and by non-algorithmic calculating machines. A classical technical method of doing this has been to use analog computers. Indeed, in many classes of mathematical problems the computational power of an analog computer is infinitely greater than the computation power of conventional digital computers.¹³

To be able to observe progress we have to define a set of criteria that allows us to distinguish progress from regress. These criteria or their range of validity are usually not explicitly defined in discussions about progress, as they are core elements in our system of value, and therefore taken for granted. For example, more transistors may

¹² “In 1947 Kiyoshi Matsuzake of the Japanese Ministry of Communications used a soroban (the Japanese version of the abacus) to best Private Tom Wood of the United States Army of Occupation, who used the most modern electrically driven mechanical calculating machine, in a contest of speed and accuracy in calculation.” (Williams, 1990:7). Slide rulers were extensively used during the Apollo 13 flight, in 1970.

¹³ The reasons include what I have earlier called “the starting problem of Turing machines” (Heinämaa & Tuomi, 1989; Tuomi, 1988). As Wiener (1975) noted, the main advantage of digital computers is that they can iterate computations without noise. Strictly speaking, Wiener’s argument however presumes that there are no rounding errors, or software or hardware bugs in the computing system.

indicate progress in specific applications of computing. In other applications, less transistors may be more. We can solve mathematical equations by turning them into sequences of elementary computations. Sometimes more intelligence is demonstrated by solving the equation with pencil and paper.

This implies that all generic claims about speed of change necessarily fail. There is no “generic” change. To speak about speed of change, we have to specify what kind of change we believe to be relevant for the present discussion. This selection can always in principle be contested and there are no a priori guarantees that the specific type of change that we now find relevant will remain relevant also for the future.

A conceptual problem with Kurzweil’s generic claim about accelerating change therefore is that its core concepts—progress and change—are not well defined. More specifically, the links between transistor count, the speed of computation, and evolutionary efficiency remain unclear.

Conclusion

The paradoxical character of conventional economic and technical discussions on technological change is that they can talk about history, but they are conceptually blind to essential novelty and change. Engineers try to improve existing functionalities and the value and quality systems of engineering communities set the criteria for improvements in relation to existing technologies (Constant, 1987; Tuomi, 2002a). In economic theory, in turn, the quantitative approaches that provide its cornerstones require that qualitative change be reduced to quantitative change. Essentially the same problem underlies the Darwinistic models of technological evolution. This problem was carefully analyzed by Bergson more than a century ago, when he tried to understand the creative dimension of evolutionary processes (e.g., Bergson, 1983). This analysis led Bergson to the view that time and cognition are closely linked, in ways that still have clear relevance for theorists of intelligence and students of technological change (c.f. Tuomi, 1999). As technological trajectories fundamentally talk only about relevant change, we therefore need theories of technical development that link change with relevance. As I have suggested elsewhere (Tuomi,

2002a), this leads to theories of innovation that are grounded on social and cognitive theories of technology.

Paradoxically, therefore, the apparent radical nature of Kurzweil's claims—for example the claim that computers are soon becoming more intelligent than humans and that technological system will continue evolution—has its roots in making coherent conclusions from assumptions that are generally taken for granted in discussions about technological change. Kurzweil draws conclusions that appear logical, as long as we share the assumption that his axioms are true. Yet, the conclusions are provocative because they are against our intuitive beliefs. In a stark contrast with the projections that predicted limits to growth in the 1970s, Kurzweil uses a model where infinite growth follows from the evolutionary nature of technical development. His choice is to keep the underlying axioms uncontested and extrapolate their outcomes into future, where most other conventional and obvious assumptions have to be rejected. Although Kurzweil's hypothesis is about change, it is therefore radically conservative in its core.

References

- Aizcorbe, A. (2002). Why are semiconductor prices falling so fast: Industry estimates and implications for productivity measurement. Federal Reserve Board. FEDS DP 20-2002.
- Aizcorbe, A., Flamm, K., & Khurshid, A. (2002). The role of semiconductor inputs in IT hardware price decline: computers vs. communications.
<http://www.federalreserve.gov/pubs/feds/2002/200237/200237pap.pdf>.
- Bergson, H. (1983). *Creative Evolution (first edition 1907)*. Lanham, MD: University Press of America.
- Berndt, E.R., Dulberger, E.R., & Rappaport, N.J. (2000). Price and quality of desktop and mobile personal computers: a quarter century of history. 17 July 2000,
<http://www.nber.org/~confer/2000/si2000/berndt.pdf>.
- CBO. (2002). The Role of Computer Technology in the Growth of Productivity. The Congress of the United States, Congressional Budget Office. A CBO Paper. <http://www.cbo.gov>.
- Ceruzzi, P.E. (1998). *A History of Modern Computing*. Cambridge, MA: The MIT Press.

- Constant, E.W. (1987). The social locus of technological practice: community, system, or organization? In W.E. Bijker, T.P. Hughes, & T.J. Pinch (Eds.), *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. (pp. 223-242). Cambridge, MA: The MIT Press.
- EITO. (2003). *European Information Technology Observatory 2003*.
- Gordon, R.J. (2000). Does the "New Economy" measure up to the great inventions of the past? *Journal of Economic Perspectives*, **14** (4), pp.49-74.
- Gordon, R.J. (2003). Hi-tech innovation and productivity growth: Does supply create its own demand? National Bureau of Economic Research. NBER Working Paper No. w9437. <http://www.nber.org/papers/w9437>.
- Grimm, B.T. (1998). Price indexes for selected semiconductors, 1974-96. *Survey of Current Business*, (February 1998), pp.8-24.
- Grimm, B.T., Moulton, B.R., & Wasshausen, D.B. (2002). Information processing equipment and software in national accounts. Bureau of Economic Analysis. <http://>.
- Heinämaa, S., & I. Tuomi. (1989). *Ajatuksia synnyttävät koneet: tekoölyn unia ja painajaisia (Thought Provoking Machines: Dreams and Nightmares of Artificial Intelligence; in Finnish)*. Porvoo: Werner Söderström Osakeyhtiö.
- ITRS. (2001). International Technology Roadmap for Semiconductors, 2001 Edition, Executive Summary. <http://public.itrs.net>.
- Kurzweil, R. (1999). *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. New York: Viking.
- Kurzweil, R. (2001). The Law of Accelerating Returns. <http://www.kurzweilai.net/articles/art0134.html>.
- Landefeld, J.S., & Grimm, B.T. (2000). A note on the impact of hedonics and computers on real GDP. *Survey of Current Business*, (December 2000), pp.17-22.
- Langlois, R.N., & Steinmueller, W.E. (1999). The evolution of competitive advantage in the worldwide semiconductor industry, 1947-1996. In D.C. Mowery & R.R. Nelson (Eds.), *Sources of Industrial Leadership: Studies of Seven Industries*. (pp. 19-78). Cambridge: Cambridge University Press.
- Moore, G.E. (1975). Progress in digital integrated electronics. *IEEE, International Electron Devices Meeting, IEDM Tech.Digest - 1975*, pp.11-13.

- Price, D.J.d.S. (1986). *Little Science, Big Science ...and Beyond*. New York: Columbia University Press.
- Svoren, J., & Khan, T. (2002). Semiconductor & Semiconductor Equipment Industry Earnings Outlook - Q4:02. WR Hambrecht + Co. Earnings Outlook, December 2002.
- Tuomi, I. (1988). Neural networks as dynamical systems: some theoretical reasons for non-algorithmic information processing. In M. Mäkelä, S. Linnainmaa, & S. Ukkonen (Eds.), *Proceedings of the Finnish Artificial Intelligence Symposium, STeP-88, Vol 2*. (pp. 593-601). Helsinki: Finnish Artificial Intelligence Society.
- Tuomi, I. (1999). *Corporate Knowledge: Theory and Practice of Intelligent Organizations*. Helsinki: Metaxis.
- Tuomi, I. (2002a). *Networks of Innovation: Change and Meaning in the Age of Internet*. Oxford: Oxford University Press.
- Tuomi, I. (2002b). The lives and death of Moore's Law. *First Monday*, 7 (11)
http://firstmonday.org/issues/issue7_11/tuomi
- UN. (1999). The World at Six Billion. United Nations Population Division, Department of Economic and Social Affairs. <http://www.un.org/esa/population/publications/sixbillion/sixbillion.htm>.
- Wiener, N. (1975). *Cybernetics: or Control and Communication in the Animal and the Machine*. Cambridge, MA: The MIT Press.
- Williams, M.R. (1990). Early calculation. In W. Aspray (Ed.), *Computing Before Computers*. Iowa: Iowa University Press.