1

INTRODUCTION: THE IMPORTANCE OF SIGNAL INTEGRITY

1.1 COMPUTING POWER: PAST AND FUTURE

It is estimated that sometime between the years 2025 and 2050, commonplace personal computers will exceed the calculation power of a human brain. Further extrapolation based on historical trends indicates that a single commonplace computer could exceed the computational power of the human race sometime between 2060 and 2100. Are such vast increases in computational power possible in less than 100 years? We cannot say for certain because it is impossible to predict the future. However, hindsight is always 20/20, and if we subscribe to the notion that history tends to repeat itself, we can look at the progress of computational capabilities over the last century to see if historical data support a rate sufficient to achieve such performance. Hans Moravec, a researcher from the Robotics Institute at Carnegie Melon University, estimated that a computer would require 100 million megainstructions per second (MIPS) to mimic sufficiently closely the behavior of a human brain [Moravec, 1998]. Based on the

Advanced Signal Integrity for High-Speed Digital Design, By Stephen H. Hall and Howard L. Heck

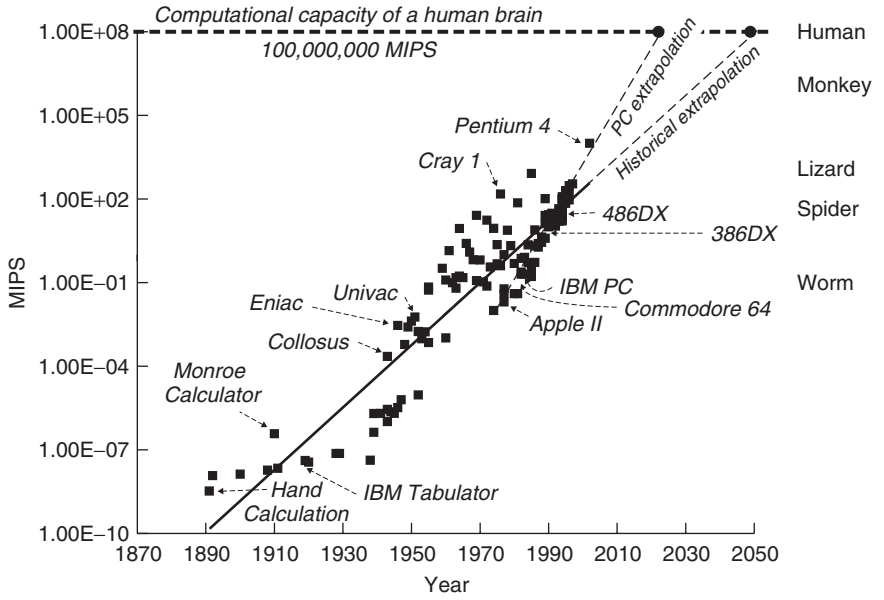

Figure 1-1 Historical computational power and extrapolation into the future. (Adapted from Moravec [1998].)

number of neurons, he was also able to compare the current state of computer technology to the estimated computational power of animal brains. These data points outline a particularly interesting way to examine the history of computational power while level-setting computer performance against the brainpower of common animals.

Figure 1-1 plots the computational power for mechanical and electrical computers used over the last 100 years. Some of the more interesting data points are labeled on the plot, ranging from hand calculation (ca. 1/100,000,000 MIPS) to the Pentium 4 processor of 2002 (10,000 MIPS), which is only two orders of magnitude away from the estimated brain power of a monkey (1,000,000 MIPS). As the plot indicates, computers comparable to the human brain could appear as early as the 2020s based on the extrapolation of personal computer (PC) performance over the last three decades. If the historical data for the entire twentieth century are used, the time frame is extended to 2050. The predictions get even more outrageous if we extend the extrapolation to the estimated computational power of all humans presently on Earth (ca. 6 billion), which would require 6×10^{17} MIPS. Such a computer could exist by 2060, as shown in Figure 1-2. The question is: Can the historical pace of development be sustained? Observation of the data indicates that the historical trend shows no sign of slowing. In fact, the rate appears to be increasing.

However, one often encounters articles by knowledgeable people in the computer industry who believe that the trend cannot be sustained and that the decades of exponential growth must stop. In 1998 it was estimated in such an article that

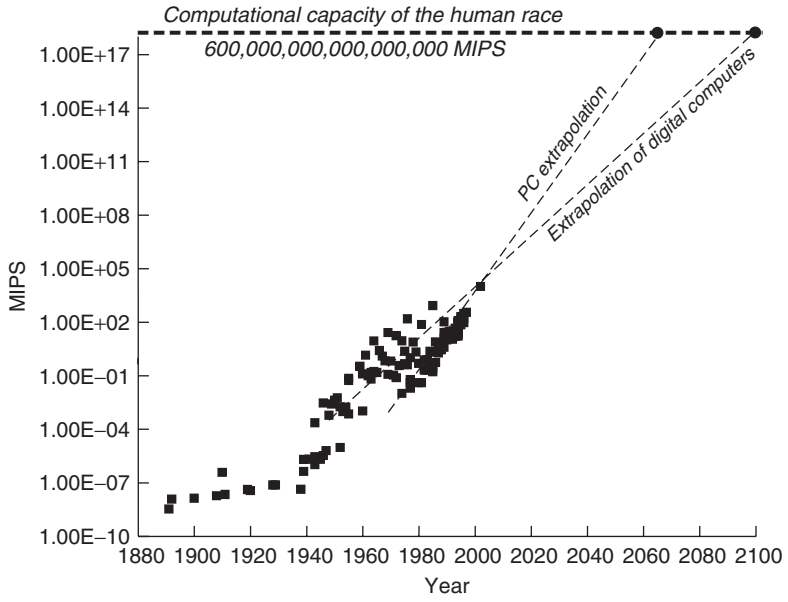

Figure 1-2 Historical computational power and extrapolation into the future.

commonplace printed circuit boards (PCBs) built on an FR4 dielectric could not support bus speeds faster than 300 MHz [Porter, 1998]. Current designs using FR4 substrates exceed that bus speed in commonplace personal computers by almost 10-fold (PCI Express Gen 2 buses run at 5 giga-transfers per second, which has a fundamental frequency of 2.5 GHz). History is filled with "experts" who mispredicted the future:

> Heavier-than-air flying machines are impossible.
> —Lord Kelvin, British mathematician and physicist, president of the British Royal Society, 1895

> Fooling around with alternating current is just a waste of time. Nobody will use it, ever.
> —Thomas Edison, American inventor, 1889

> Rail travel at high speed is not possible because passengers, unable to breathe, would die of asphyxia.
> —Dr Dionysys Larder (1793–1859), professor of natural philosophy and astronomy, University College London

In the mid-1970s, integrated circuits held approximately 10,000 components, which was enough to construct an entire computer with devices as small as 3 μm. Experienced engineers worried that semiconductor technology had reached its pinnacle. Three micrometers was barely larger than the wavelength of the light

used to sculpt the chip. Interactions between ever-closer wires were about to ruin the signals. Chips would soon generate so much heat that they would be impossible to cool without a refrigeration unit. The list goes on [Moravec, 1998].

A look at the computer growth graph shows that the industry found solutions to all those problems. Chip progress not only continued—*it sped up*. Technology companies, motivated by the potential of high profits, dedicated tremendous resources to making the "impossible" possible: developing more efficient transistor designs, better heat sinks, new manufacturing processes, and more advanced analysis techniques. History indicates that the rate of performance will continue to grow at exponential rates.

Historically, the mechanism for advancing computation has been to miniaturize components, allowing more devices to fit in and operate in a smaller space, thus producing more performance per unit volume. First, the gears in mechanical calculators shrunk, which allowed them to spin faster. Then the relays in electromechanical machines became smaller, which allowed them to switch faster. Next, the switches in digital machines evolved from shotgun shell–sized vacuum tubes, to pea-sized transistors, to tiny integrated-circuit chips [Moravec, 1998]. Each of these technological advancements came with a price: *New problems that were never before considered arose that needed to be solved*.

How does this relate to signal integrity? The field of signal integrity arose directly from the exponential growth of computing power. A computer system is comprised of many integral components in addition to the processor, such as the memory, cache, and chip set. The interconnections between these parts within a computer system are known collectively as system *buses*. Essentially, a bus is an integrated set of interconnections used to transfer data between different parts of a digital system. Accordingly, to capitalize on the benefits of increased processor power, system buses must also operate at higher data transfer rates. For example, if the memory bus fails to transmit data at a sufficiently fast rate, the processor simply sits idle until data are available. This bottleneck would negate much of the performance gained from a more powerful processor. Subsequently, it is vital that the bus performance scale correspondingly with processor performance.

1.2 THE PROBLEM

The two mechanisms used historically to scale bus design to feed the growing performance of computer processors have been speed and width. *Speed* facilitates higher information transfer rates by sending more bits in a given amount of time. *Width* facilitates more information transfer by sending more bits in parallel. From now on, the rate of information transfer on a bus will be referred to as the *bus bandwidth*.

Increasing the bus speed to overcome bandwidth limitations becomes problematic for many reasons. As bus frequencies increase in speed, the pathways that comprise the bus, called *interconnects*, begin to exhibit high-frequency behavior, which thoroughly puzzles many conventional digital designers. What is

required is complete comprehension of the relevant analog techniques and theory commonly used in microwave system and radio designs applied carefully to the digital realm. As the operating frequencies of digital systems increase, these analog effects become more prevalent and severely impede the overall performance if not resolved properly. Furthermore, increased bus speed usually requires more power, which is a precious commodity, especially in mobile designs such as laptops, which rely on battery power.

Increasing the bus width to overcome bandwidth roadblocks is self-limiting. Practical mechanical limitations arise quickly due to increased pin counts on packages, sockets, connectors, and the shortcomings of PCB technology. Furthermore, interactions between closely spaced interconnects lower the signal-to-noise level, making clean data transmission more difficult. Since Moore's law results in computer performance doubling every 18 months, and the bus bandwidth must scale in proportion, doubling the number of signals in the bus to facilitate the required bandwidth provides a solution to the problem that lasts less than two years. Increasing the width of the bus is simply a short-term "band-aid." At some point, faster bus speeds will be required.

The problem is that as bus designs become both wider and faster and form factors shrink to provide more computational power per unit volume, the assumptions used for past designs become outdated and new techniques must be developed. As a result, the field of signal integrity is evolving continuously to encompass new effects that were not relevant to earlier designs. Modern bus designs have become so fast that the designer must calculate the voltage and timing numbers to a resolution as small as a few picoseconds and a few millivolts. This degree of resolution was unheard of in computer designs just a few years ago. Just to put this problem in perspective, the light that is reflected off your nose takes a little over 85 ps to travel to the surface of your eye, which is well over 10 times the required timing resolution of some modern bus designs. This dramatic decrease in bus timing requirements leads to several problems. First, the number of effects that must be accounted for in the design stage increases. This is because effects that were either second order, or ignored completely in previous designs, begin to dominate the performance. Consequently, the total number of variables that must be accounted for increases, which makes the problem more difficult. Furthermore, the new variables are often very difficult or impossible to model using conventional methods. So we have not only many more variables to worry about, but most of the new variables are very difficult to model correctly. Finally, to top it off, the laboratory equipment currently available is often insufficient to resolve these small timing variations, making it difficult or impossible to verify the completed design or to correlate the models to reality.

1.3 THE BASICS

As the reader undoubtedly knows, the basic idea in digital design is to communicate information with signals representing 1's or 0's. Typically, this involves

sending and receiving series of trapezoidal voltage signals in which a high voltage is a 1 and a low voltage is a 0. The conductive paths carrying the digital signals are known as *interconnects*. The interconnect includes the entire electrical pathway from the chip sending a signal to the chip receiving the signal. This includes the chip packages, connectors, sockets, transmission lines, and vias. A group of interconnects is referred to as a *bus*. The region of voltage where a digital receiver distinguishes between a high and a low voltage is known as the *threshold region*. Within this region, the receiver will either switch high or switch low. On the silicon, the actual switching voltages vary with temperature, supply voltage, silicon process, and other variables. From the system designer's point of view, there are usually high- and low-voltage thresholds, known as V_{ih} and V_{il}, associated with the receiving silicon, above and below which a high or low value is guaranteed to be received under all conditions. Thus, the designer must guarantee that the system can, under all conditions, deliver high voltages that do not, even briefly, fall below V_{ih}, and low voltages that remain below V_{il}, to ensure the integrity of the data.

To maximize the speed of operation of a digital system, the timing uncertainty of a transition through the threshold region must be minimized. This means that the rise or fall time of the digital signal must be as fast as possible. Ideally, an infinitely fast edge rate would be used, although there are many practical problems that prevent this. Realistically, edge rates as fast as 35 ps are encountered in real systems. The reader can use Fourier analysis to verify that the quicker the edge rate, the higher the frequencies that are found in the spectrum of the signal. Herein lies a clue to the difficulty. Every conductor has a frequency-dependent capacitance, inductance, conductance, and resistance. At a high-enough frequency, none of these things are negligible. Thus, a wire is no longer a wire but a distributed, frequency-dependent parasitic element that has delay and a transient impedance profile that can cause distortions and glitches to manifest themselves on the waveform propagating from the driving chip to the receiving chip. The wire is now an element that is coupled to everything around it, including power and ground structures, heat sinks, other traces, and even the wireless network. The signal is not contained in the conductor itself but is, instead, carried in the local electric and magnetic fields around the conductor. The signals on one interconnect will affect, and be effected by, the signals on another. The inductance, capacitance, and resistance of all the structures in the vicinity of the interconnect have vital roles in the simple task of guaranteeing proper signaling transitions with appropriate timing at the receiver.

One of the most difficult aspects of high-speed design is the fact that there are many codependent variables that affect the outcome of a digital design. Some of the variables are controllable, and others force the designer to live with the random variation. One of the difficulties in high-speed design is how to handle the many variables, whether they are controllable or uncontrollable. Often, simplifications can be made by neglecting or assuming values for variables, but this can lead to unknown failures down the road for which it will not be

possible after the fact to locate the root cause. As timing becomes more constrained, the simplifications of the past are rapidly dwindling in utility to the modern designer. In this book we also show how to incorporate a large number of variables that would otherwise make the problem intractable. Without a methodology for handling the large number of variables, a design ultimately incorporates some guesswork, no matter how much the designer understands the system physically. The final step in handling all the variables is often the most difficult part and the one most readily ignored by designers. A designer crippled by the inability to handle large numbers of variables will ultimately resort to proving a few "point solutions" instead and hope that they plausibly represent all known conditions. Although such methods are sometimes unavoidable, this can be a dangerous guessing game. Of course, a certain amount of guesswork is always present in design, but the goal of the system designer should be to minimize uncertainty.

1.4 A NEW REALM OF BUS DESIGN

Technology has progressed to a point where digital design has entered a new realm, where new design techniques and concepts are required that baffle even the most seasoned digital system designers. The fact is that present and future state-of-the-art digital systems, such as personal computers, cannot be designed without a thorough understanding of the principles outlined in this book. Why hasn't this been a problem before? The answer is that digital designers didn't need to understand these things. But digital circuits are reaching speeds where design will not be possible without an understanding of this subject. Seasoned engineers face the threat of becoming a legacy if they do not adapt to the new design space. This book will help practicing engineers adapt.

From the Monroe calculator to the Pentium, from punch cards to flash memory, from vacuum tubes to integrated circuits, computer performance is increasing at an exponential rate. In this book we address the needs of the contemporary digital designer as he or she encounters the numerous new challenges with modern and future high-speed digital systems and is forced to learn material previously not needed. As the conventional digital designer transitions to faster designs, he or she will indeed experience a completely different view of logic signals at high speeds. This book will help to make sense of the ugly, distorted, and smeared waveforms produced in a high-speed digital system.

1.5 SCOPE OF THE BOOK

This book was written to be an advanced study in signal integrity. Although some basic material is covered, it is assumed that the reader is well acquainted with basic electromagnetic theory, vector calculus, differential equations, statistics, and transmission-line analysis. The book builds on the traditional knowledge base and covers topics required to design present and future digital systems.

1.6 SUMMARY

All of this leads to the present situation: There are new problems to solve. Engineers who can solve these problems will define the future. This book will equip readers with the necessary practical understanding to contend with contemporary problems of modern high-speed digital design with enough theory to see beyond this book and solve problems that the authors have not yet encountered.

ERRATA

Inevitably, some errors will slip past the layers of review. Although they will be remedied in subsequent printings, it is useful to summarize the corrections in one place. The errors, along with the corrections, will be posted at ftp://ftp.wiley.com/public/sci_tech_med/high_speed_design.

REFERENCES

Moravec, Hans, 1998, When will computer hardware match the human brain?, *Journal of Evolution and Technology*, vol. 1.

Porter, Chris, 1998, High chip speeds spell end for FR4, *Electronic Times*, Mar. 30.