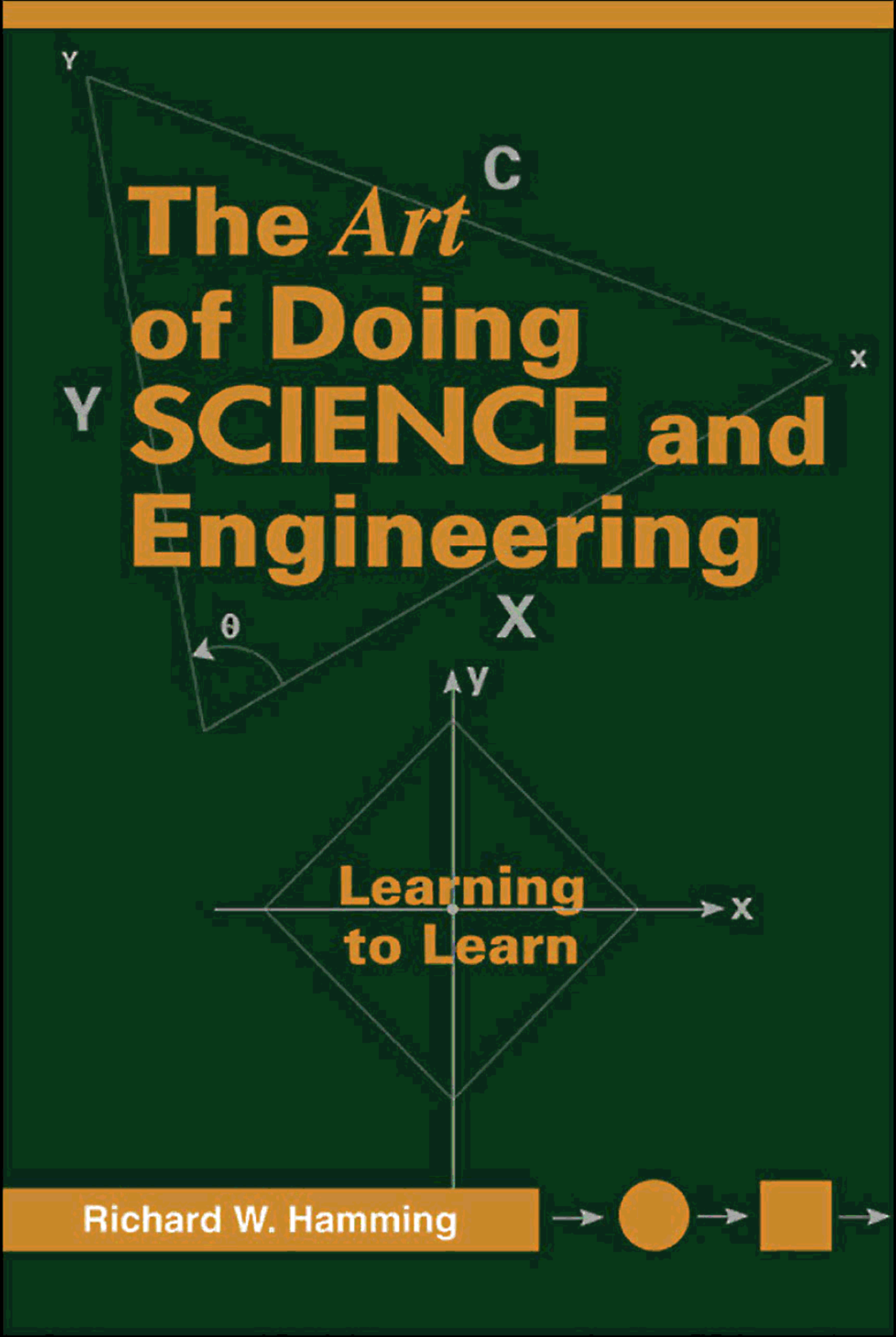

The *Art* of Doing SCIENCE and Engineering

Learning to Learn

Richard W. Hamming

The Art of Doing Science and Engineering

Learning to Learn

Richard W.Hamming
U.S. Naval Postgraduate School
Monterey, California

GORDON AND BREACH SCIENCE PUBLISHERS
Australia • Canada • China • France • Germany • India •
Japan • Luxembourg • Malaysia • The Netherlands •
Russia • Singapore • Switzerland • Thailand •
United Kingdom

This edition published in the Taylor & Francis e-Library, 2005.

Printed in India.

Amsteldijk 166
1st Floor
1079 LH Amsterdam
The Netherlands

British Library Cataloguing in Publication Data

Hamming, R.W. (Richard Wesley), 1915–
The art of doing science and engineering: learning to learn
1. Science 2. Engineering
I. Title
500

ISBN 0-203-45071-X Master e-book ISBN

ISBN 0-203-45913-X (Adobe eReader Format)
ISBN 90-5699-501-4 (Print Edition)

CONTENTS

PREFACE

After many years of pressure and encouragement from friends, I decided to write up the graduate course in engineering I teach at the U.S.Naval Postgraduate School in Monterey, California. At first I concentrated on all the details I thought should be tightened up, rather than leave the material as a series of somewhat disconnected lectures. In class the lectures often followed the interests of the students, and many of the later lectures were suggested topics in which they expressed an interest Also, the lectures changed from year to year as various areas developed. Since engineering depends so heavily these days on the corresponding sciences, I often use the terms interchangeably.

After more thought I decided that since I was trying to teach "style" of thinking in science and engineering, and "style" is an art, I should therefore copy the methods of teaching used for the other arts—once the fundamentals have been learned. How to be a great painter cannot be taught in words; one learns by trying many different approaches that seem to surround the subject. Art teachers usually let the advanced student paint, and then make suggestions on how they would have done it, or what might also be tried, more or less as the points arise in the student's head—which is where the learning is supposed to occur! In this series of lectures I try to communicate to students what cannot be said in words—the essence of style in science and engineering. I have adopted a loose organization with some repetition since this often occurs in the lectures. There are, therefore, digressions and stories—with some told in two different places—all in the somewhat rambling, informal style typical of lectures.

I have used the "story" approach, often emphasizing the initial part of the discovery, because I firmly believe in Pasteur's remark, "Luck favors the prepared mind." In this way I can illustrate how the individual's preparation before encountering the problem can often lead to recognition, formulation, and solution. Great results in science and engineering are "bunched" in the same person too often for success to be a matter of random luck.

Teachers should prepare the student for the student's future, not for the teacher's past. Most teachers rarely discuss the important topic of the future of their field, and when this is pointed out they usually reply: "No one can know the future." It seems to me the difficulty of knowing the future does not absolve the teacher from seriously trying to help the student to be ready for it when it comes. It is obvious the experience of an individual is not necessarily that of a class of individuals; therefore, any one person's projection into the future is apt to be somewhat personal and will not be universally accepted. This does not justify reverting to impersonal surveys and losing the impact of the personal story.

Since my classes are almost all carefully selected navy, marine, army, air force, and coast guard students with very few civilians, and, interestingly enough, about 15% very highly selected foreign military, the students face a highly technical future—hence the importance of preparing them for *their* future and not just *our* past.

The year 2020 seems a convenient date to center the preparation for their future—a sort of 20/20 foresight, as it were. As graduate students working toward a master's degree, they have the basics well in hand. That leaves me the task of adding "style" to their education, which in practice is usually the difference between an average person and a great one. The school has allowed me great latitude in trying to teach a completely non-technical course; this course "complements" the more technical ones. As a result, my opening words, occasionally repeated, are: "There is really no technical content in the course, though I will, of course, refer to a great deal of it, and hopefully it will generally be a good review of the fundamentals of what you have learned. Do not think it is the *content* of the course—it is only illustrative material. *Style of thinking* is the center of the course."

The subtitle of this book, *Learning to Learn,* is the main solution I offer to help students cope with the rapid changes they will have to endure in their fields. The course centers around how to look at and think about knowledge, and it supplies some historical perspectives that might be useful.

This course is mainly personal experiences I have had and digested, at least to some extent. Naturally one tends to remember one's successes and forget lesser events, but I recount a number of my spectacular failures as clear examples of what to avoid. I have found that the *personal* story is far, far more effective than the *impersonal* one; hence there is necessarily an aura of "bragging" in the book that is unavoidable.

Let me repeat what I earlier indicated. Apparently an "art"— which almost by definition cannot be put into words—is probably best communicated by approaching it from many sides and doing so repeatedly, hoping thereby students will finally master enough of the art, or if you wish, style, to significantly increase their future contributions to society. A totally different description of the course is: it covers all kinds of things that could not find their proper place in the standard curriculum.

The casual reader should not be put off by the mathematics; it is only "window dressing" used to illustrate and connect up with earlier learned material. Usually the underlying ideas can be grasped from the words alone.

It is customary to thank various people and institutions for help in producing a book. Thanks obviously go to AT&T Bell Laboratories, Murray Hill, New Jersey, and to the U.S.Naval Postgraduate School, especially the Department of Electrical and Computer Engineering, for making this book possible.

INTRODUCTION

This book is concerned more with the future and less with the past of science and engineering. Of course future predictions are uncertain and usually based on the past; but the past is also much more uncertain—or even falsely reported—than is usually recognized. Thus we are forced to *imagine* what the future will probably be. This course has been called"Hamming on Hamming" since it draws heavily on my own past experiences, observations, and wide reading.

There is a great deal of mathematics in the early part because almost surely the future of science and engineering will be more mathematical than the past, and also I need to establish the nature of the foundations of our beliefs and their uncertainties. Only then can I show the weaknesses of our current beliefs and indicate future directions to be considered.

If you find the mathematics difficult, skip those early parts. Later sections will be understandable provided you are willing to forgo the deep insights mathematics gives into the weaknesses of our current beliefs. General results are always stated in words, so the content will still be there but in a slightly diluted form.

1
Orientation

The purpose of this course is to prepare you for your technical future. There is really no technical content in the course, though I will, of course, refer to a great deal of it, and hopefully it will generally be a good review of the fundamentals you have learned. Do not think the technical content is the course—it is only illustrative material. Style of thinking is the center of the course. I am concerned with educating and not training you.

I will examine, criticize, and display styles of thinking. To illustrate the points of style I will often use technical knowledge most of you know, but, again, it will be, I hope, in the form of a useful review which concentrates on the fundamentals. You should regard this as a course which complements the many technical courses you have learned. Many of the things I will talk about are things which I believe you ought to know but which simply do not fit into courses in the standard curriculum. The course exists because the department of Electrical and Computer Engineering of the Naval Postgraduate School recognizes the need for both a general education and the specialized technical training your future demands.

The course is concerned with "style", and almost by definition style cannot be taught in the normal manner by using words. I can only approach the topic through particular examples, which I hope are well within your grasp, though the examples come mainly from my 30 years in the mathematics department of the Research Division of Bell Telephone Laboratories (before it was broken up). It also comes from years of study of the work of others.

The belief anything can be "talked about" in words was certainly held by the early Greek philosophers, Socrates (469–399), Plato (427–347), and Aristotle (384–322). This attitude ignored the current *mystery cults* of the time who asserted you had to "experience" some things which could not be communicated in words. Examples might be the gods, truth, justice, the arts, beauty, and love. Your scientific training has emphasized the role of words, along with a strong belief in *reductionism,* hence to emphasize the possible limitations of language I shall take up the topic in several places in the book. I have already said "style" is such a topic.

I have found to be effective in this course, I must use mainly first hand knowledge, which implies I break a standard taboo and talk about myself in the first person, instead of the traditional impersonal way of science. You must forgive me in this matter, as there seems to be no other approach which will be as effective. If I do not use direct experience then the material will probably sound to you like merely pious words and have little impact on your minds, and it is your minds I must change if I am to be effective.

This talking about first person experiences will give a flavor of "bragging", though I include a number of my serious errors to partially balance things. Vicarious learning from the experiences of others saves making errors yourself, but I regard the study of successes as being basically more important than the study of failures. As I will several times say, there are so many ways of being wrong and so few of being right, studying successes is more efficient, and furthermore when your turn comes you will know how to succeed rather than how to fail!

I am, as it were, only a coach. I cannot run the mile for you; at best I can discuss styles and criticize yours. You know you must run the mile if the athletics course is to be of benefit to you—hence *you* must think carefully about what you hear or read in this book if it is to be effective in changing you—which must obviously be the purpose of any course. Again, you will get out of this course only as much as you put in, and if you put in little effort beyond sitting in the class or reading the book, then it is simply a waste of your time. *You* must also mull things over, compare what I say with your own experiences, talk with others, and make some of the points part of your way of doing things.

Since the subject matter is "style", I will use the comparison with teaching painting. Having learned the fundamentals of painting, you then study under a master you accept as being a great painter; but you know you must forge your own style out of the elements of various earlier painters plus your native abilities. You must also adapt your style to fit the future, since merely copying the past will not be enough if you aspire to future greatness—a matter I assume, and will talk about often in the book. I will show you my style as best I can, but, again, you must take those elements of it which seem to fit you, and you must finally create your own style. Either you will be a leader, or a follower, and my goal is for you to be a leader. You cannot adopt every trait I discuss in what I have observed in myself and others; you must select and adapt, and make them your own if the course is to be effective.

Even more difficult than what to select is that what is a successful style in one age may not be appropriate to the next age! My predecessors at Bell Telephone Laboratories used one style; four of us who came in all at about the same time, and had about the same chronological age, found our own styles and as a result we rather completely transformed the overall style of the Mathematics Department, as well as many parts of the whole Laboratories. We privately called ourselves "The four young Turks", and many years later I found top management had called us the same!

I return to the topic of education. You all recognize there is a significant difference between *education* and *training*.

> Education is what, when, and why to do things, Training is how to do it.

Either one without the other is not of much use. You need to know both what to do and how to do it. I have already compared mental and physical training and said to a great extent in both you get out of it what you put into it—all the coach can do is suggest styles and criticize a bit now and then. Because of the usual size of these classes, or because you are reading the book, there can be little direct criticism of your thinking by me, and you simply have to do it internally and between yourselves in conversations, and apply the things I say to your own experiences. You might think education should precede training, but the kind of educating I am trying to do must be based on your past experiences and technical knowledge. Hence this inversion of what might seem to be reasonable. In a real sense I am engaged in "meta-education", the topic of the course is education itself and hence our discussions must rise above it—"meta-education", just as metaphysics was supposed to be above physics in Aristotle's time (actually "follow", "transcend" is the translation of "meta").

This book is aimed at your future, and we must examine what is likely to be the state of technology (Science and Engineering) at the time of your greatest contributions. It is well known that since about Isaac Newton's time (1642–1727) knowledge of the type we are concerned with has about doubled every 17 years. First, this may be measured by the books published (a classic observation is libraries must double their holdings every 17 years if they are to maintain their relative position). Second, when I went to Bell

Telephone Laboratories in 1946 they were trying to decrease the size of the staff from WW-II size down to about 5500. Yet during the 30 years I was there I observed a fairly steady doubling of the number of employees every 17 years, regardless of the administration having hiring freezes now and then, and such things. Third, the growth of the number of scientists generally has similarly been exponential, and it is said currently almost 90% of the scientists who ever lived are now alive! It is hard to believe in your future there will be a dramatic decrease in these expected rates of growth, hence you face, even more than I did, the constant need to learn new things.

Here I make a digression to illustrate what is often called "back of the envelop calculations". I have frequently observed great scientists and engineers do this much more often than "the run of the mill" people, hence it requires illustration. I will take the above two statements, knowledge doubles every 17 years, and 90% of the scientists who ever lived are now alive, and ask to what extent they are compatible. The model of the growth of knowledge and the growth of scientists assumed are both exponential, with the growth of knowledge being proportional to the number of scientists alive. We begin by assuming the number scientists at any time t is

$$y(t) = a\exp\{bt\}$$

and the amount of knowledge produced annually has a constant k of proportionality to the number of scientists alive. Assuming we begin at minus infinity in time (the error is small and you can adjust it to Newton's time if you wish), we have the formula

$$\frac{1}{2} = \frac{\int_{-\infty}^{t-17} kae^{bt}\,\mathrm{d}t}{\int_{-\infty}^{t} kae^{bt}\,\mathrm{d}t}$$

$$= \frac{(ka/b)\mathrm{e}^{b(t-17)}}{(ka/b)\mathrm{e}^{bt}} = \mathrm{e}^{-17b} = \frac{1}{2}$$

hence we know b. Now to the other statement. If we allow the lifetime of a scientist to be 55 years (it seems likely that the statement meant living and not practicing, but excluding childhood) then we have

$$\frac{\int_{t-55}^{t} ae^{bt}\,\mathrm{d}t}{\int_{-\infty}^{t} ae^{bt}\,\mathrm{d}t} = \frac{\mathrm{e}^{bt} - \mathrm{e}^{\{bt-b(55)\}}}{\mathrm{e}^{bt}} = 1 - \mathrm{e}^{-55b}$$

$$= 1 - (\tfrac{1}{2})^{55/17} = 1 - 0.106\ldots = 0.894\ldots$$

which is very close to 90%.

Typically the first back of the envelop calculations use, as we did, definite numbers where one has a feel for things, and then we repeat the calculations with parameters so you can adjust things to fit the data better and understand the general case. Let the doubling period be D, and the lifetime of a scientist be L. The first equation now becomes

$$\tfrac{1}{2} = \mathrm{e}^{bD},$$

and the second becomes:

$$\tfrac{9}{10} = 1 - \mathrm{e}^{bL} = 1 - (\tfrac{1}{2})^{L/D},$$

$$(\tfrac{1}{2})^{L/D} = \tfrac{1}{10},$$

$$\frac{L}{D} = \frac{\log 10}{\log 2} = \frac{1}{0.30103} = 3.3219\ldots$$

With D=17 years we have 17×3.3219=56.47…years for the lifetime of a scientist, which is close to the 55 we assumed. We can play with ratio of L/D until we find a slightly closer fit to the data (which was approximate, though I believe more in the 17 years for doubling than I do in the 90%). Back of the envelop computing indicates the two remarks are reasonably compatible. Notice the relationship applies for all time so long as the assumed simple relationships hold.

The reason back of the envelop calculations are widely used by great scientists is clearly revealed—you get a good feeling for the truth or falsity of what was claimed, as well as realize which factors you were inclined not to think about, such as exactly what was meant by the lifetime of a scientist. Having done the calculation you are much more likely to retain the results in your mind. Furthermore, such calculations keep the ability to model situations fresh and ready for more important applications as they arise. Thus I recommend when you hear quantitative remarks such as the above you turn to a quick modeling to see if you believe what is being said, especially when given in the public media like the press and TV. Very often you find what is being said is nonsense, either no definite statement is made which you can model, or if you can set up the model then the results of the model do not agree with what was said. I found it very valuable at the physics table I used to eat with; I sometimes cleared up misconceptions at the time they were being formed, thus advancing matters significantly.

Added to the problem of the growth of new knowledge is the obsolescence of old knowledge. It is claimed by many the half-life of the technical knowledge you just learned in school is about 15 years—in 15 years half of it will be obsolete (either we have gone in other directions or have replaced it with new material). For example, having taught myself a bit about vacuum tubes (because at Bell Telephone Laboratories they were at that time obviously important) I soon found myself helping, in the form of computing, the development of transistors—which obsoleted my just learned knowledge!

To bring the meaning of this doubling down to your own life, suppose you have a child when you are x years old. That child will face, when it is in college, about y times the amount you faced.

y factor of increase	x years
2	17
3	27
4	34
5	39
6	44
7	48
8	51

This doubling is not just in theorems of mathematics and technical results, but in musical recordings of Beethoven's Ninth, of where to go skiing, of TV programs to watch or not to watch. If you were at times awed by the mass of knowledge you faced when you went to college, or even now, think of your children's troubles when they are there! The technical knowledge involved in your life will quadruple in 34 years, and many of you will then be near the high point of your career. Pick your estimated years to retirement and then look in the left-hand column for the probable factor of increase over the present current knowledge when you finally quit!

What is my answer to this dilemma? One answer is you must concentrate on fundamentals, at least what *you think* at the time are fundamentals, and also develop the ability to learn new fields of knowledge when they arise so you will not be left behind, as so many good engineers are in the long run. In the position I found myself in at the Laboratories, where I was the only one locally who seemed (at least to me) to have a firm grasp on computing, I was forced to learn numerical analysis, computers, pretty much all of the physical sciences at least enough to cope with the many different computing problems which arose and whose solution could benefit the Labs, as well as a lot of the social and some the biological sciences. Thus I am a veteran of learning enough to get along without at the same time devoting all my effort to learning new topics and thereby not contributing my share to the total effort of the organization. The early days of learning had to be done while I was developing and running a computing center. You will face similar problems in your career as it progresses, and, at times, face problems which seem to overwhelm you.

How are you to recognize “fundamentals”? One test is they have lasted a long time. Another test is from the fundamentals all the rest of the field can be derived by using the standard methods in the field.

I need to discuss science vs. engineering. Put glibly:

> In science if you know what you are doing you should not be doing it.
> In engineering if you do not know what you are doing you should not be doing it.

Of course, you seldom, if ever, see either pure state. All of engineering involves some creativity to cover the parts not known, and almost all of science includes some practical engineering to translate the abstractions into practice. Much of present science rests on engineering tools, and as time goes on, engineering seems to involve more and more of the science part. Many of the large scientific projects involve very serious engineering problems—the two fields are growing together! Among other reasons for this situation is almost surely we are going forward at an accelerated pace, and now there is not time to allow us the leisure which comes from separating the two fields. Furthermore, both the science and the engineering you will need for your future will more and more often be created after you left school. Sorry! But you will simply have to actively master *on your own* the many new emerging fields as they arise, without having the luxury of being passively taught.

It should be noted that engineering is not just applied science, which is a distinct third field (though it is not often recognized as such) which lies between science and engineering.

I read somewhere there are 76 different methods of predicting the future—but very number suggests there is no reliable method which is widely accepted. The most trivial method is to predict tomorrow will be exactly the same as today—which at times is a good bet. The next level of sophistication is to use the current rates of change and to suppose they will stay the same—linear prediction in the variable used. Which variable you use can, of course, strongly affect the prediction made! Both methods are not much good for long-term predictions, however.

History is often used as a long-term guide; some people believe history repeats itself and others believe exactly the opposite! It is obvious:

> The past was once the future and the future will become the past.

In any case I will often use history as a background for the extrapolations I make. I believe the best predictions are based on understanding the fundamental forces involved, and this is what I depend on mainly. Often it is not physical limitations which control but rather it is human made laws, habits, and organizational rules, regulations, personal egos, and inertia, which dominate the evolution to the future. You have not been trained along these lines as much as I believe you should have been, and hence I must be careful to include them whenever the topics arise.

There is a saying,"Short term predictions are always optimistic and long term predictions are always pessimistic". The reason, so it is claimed, the second part is true is for most people the geometric growth due to the compounding of knowledge is hard to grasp. For example for money a mere 6% annual growth doubles the money in about 12 years! In 48 years the growth is a factor of 16. An example of the truth of this claim that most long-term predictions are low is the growth of the computer field in speed, in density of components, in drop in price, etc. as well as the spread of computers into the many corners of life. But the field of Artificial Intelligence (AI) provides a very good counter example. Almost all the leaders in the field made long-term predictions which have almost never come true, and are not likely to do so within your lifetime, though many will in the fullness of time.

I shall use history as a guide many times in spite of Henry Ford, Sr. saying, "History is Bunk". Probably Ford's points were:

1. History is seldom reported at all accurately, and I have found no two reports of what happened at Los Alamos during WW-II seems to agree.
2. Due to the pace of progress the future is rather disconnected from the past; the presence of the modern computer is an example of the great differences which have arisen.

Reading some historians you get the impression the past was determined by big trends, but you also have the feeling the future has great possibilities. You can handle this apparent contradiction in at least four ways:

1. You can simply ignore it.
2. You can admit it.
3. You can decide the past was a lot less determined than historians usually indicate and individual choices can make large differences at times. Alexander the Great, Napoleon, and Hitler had great effects on the physical side of life, while Pythagoras, Plato, Aristotle, Newton, Maxwell, and Einstein are examples on the mental side.
4. You can decide the future is less open ended than you would like to believe, and there is really less choice than there appears to be.

It is probable the future will be more limited by the slow evolution of the human animal and the corresponding human laws, social institution, and organizations than it will be by the rapid evolution of technology.

In spite of the difficulty of predicting the future and that:

> Unforeseen technological inventions can completely upset the most careful predictions,

you must try to foresee the future you will face. To illustrate the importance of this point of trying to foresee the future I often use a standard story.

It is well known the drunken sailor who staggers to the left or right with n independent random steps will, on the average, end up about $\sqrt{n}$ steps from the origin. But if there is a pretty girl in one direction, then his steps will tend to go in that direction and he will go a distance proportional to n. In a lifetime of many, many independent choices, small and large, a career with a vision will get you a distance proportional to n, while no vision will get you only the distance $\sqrt{n}$. In a sense, the main difference between those who go far and those who do not is some people have a vision and the others do not and therefore can only react to the current events as they happen.

One of the main tasks of this course is to start you on the path of creating in some detail *your vision of your future*. If I fail in this I fail in the whole course. You will probably object that if you try to get a vision now it is likely to be wrong—and my reply is from observation I have seen the accuracy of the vision matters less than you might suppose, getting anywhere is better than drifting, there are potentially many paths to greatness for you, and just which path you go on, *so long as it takes you to greatness,* is none of my business. You must, as in the case of forging your personal style, find your vision of your future career, and then follow it as best you can.

No vision, not much of a future.

To what extent history does or does not repeat itself is a moot question. But it is one of the few guides you have, hence history will often play a large role in my discussions—I am trying to provide you with some perspective as a possible guide to create your vision of your future. The other main tool I have used is an active imagination in trying to see what will happen. For many years I devoted about 10% of my time (Friday afternoons) to trying to understand what would happen in the future of computing, both as a scientific tool and as shaper of the social world of work and play. In forming your plan for your future you need to distinguish three different questions:

What is possible?
What is likely to happen?
What is desirable to have happen?

In a sense the first is Science—what is possible. The second in Engineering—what are the human factors which chose the one future that does happen from the ensemble of all possible futures. The third, is ethics, morals, or what ever other word you wish to apply to value judgments. It is important to examine all three questions, and in so far as the second differs from the third, you will probably have an idea of how to alter things to make the more desirable future occur, rather than let the inevitable happen and suffer the consequences. Again, you can see why having a vision is what tends to separate the leaders from the followers.

The standard process of organizing knowledge by departments, and subdepartments, and further breaking it up into separate courses, tends to conceal the homogeneity of knowledge, and at the same time to omit much which falls between the courses. The optimization of the individual courses in turn means a lot of important things in Engineering practice are skipped since they do not appear to be essential to any one course. One of the functions of this book is to mention and illustrate many of these missed topics which are important in the practice of Science and Engineering. Another goal of the course is to show the essential unity of all knowledge rather than the fragments which appear as the individual topics are taught. In your future anything and everything you know might be useful, but if you believe the problem is in one area you are not apt to use information that is relevant but which occurred in another course.

The course will center around computers. It is not merely because I spent much of my career in Computer Science and Engineering, rather it seems to me computers will dominate your technical lives. I will repeat a number of times in the book the following facts: Computers when compared to Humans have the advantages:

Economics	—far cheaper, and getting more so
Speed	—far, far faster

Accuracy	—far more accurate (precise)
Reliability	—far ahead (many have error correction built into them)
Rapidity of control	—many current airplanes are unstable and require rapid computer control to make them practical
Freedom from boredom	—an overwhelming advantage
Bandwidth in and out	—again overwhelming
Ease of retraining	—change programs, not unlearn and then learn the new thing consuming hours and hours of human time and effort
Hostile environments	—outer space, underwater, high radiation fields, warfare, manufacturing situations that are unhealthful, etc.
Personnel problems	—they tend to dominate management of humans but not of machines; with machines there are no pensions, personal squabbles, unions, personal leave, egos, deaths of relatives, recreation, etc.

I need not list the advantages of humans over computers—almost every one of you has already objected to this list and has in your mind started to cite the advantages on the other side.

Lastly, in a sense, this is a religious course—I am preaching the message that, with apparently only one life to live on this earth, you ought to try to make significant contributions to humanity rather than just get along through life comfortably—that the life of trying to achieve excellence in some area is in itself a worthy goal for your life. It has often been observed the true gain is in the struggle and not in the achievement—a life without a struggle on your part to make yourself excellent is hardly a life worth living. This, it must be observed, is an opinion and not a fact, but it is based on observing many people's lives and speculating on their total happiness rather than the moment to moment pleasures they enjoyed. Again, this opinion of their happiness must be my own interpretation as no one can know another's life. Many reports by people who have written about the "good life" agree with the above opinion. Notice I leave it to you to pick your goals of excellence, but claim only a life without such a goal is not really living but it is merely existing—in my opinion. In ancient Greece Socrates (469–399) said:

> The unexamined life is not worth living.

2
Foundations of the Digital (Discrete)

We are approaching the end of the revolution of going from signaling with continuous signals to signaling with discrete pulses, and we are now probably moving from using pulses to using solitons as the basis for our discrete signaling. Many signals occur in Nature in a continuous form (if you disregard the apparent discrete structure of things built out of molecules and electrons). Telephone voice transmission, musical sounds, heights and weights of people, distance covered, velocities, densities, etc. are examples of continuous signals. At present we usually convert the continuous signal almost immediately to a sampled discrete signal; the sampling being usually at equally spaced intervals in time and the amount of the signal being quantized to a comparatively few levels. Quantization is a topic we will ignore in these chapters, though it is important in some situations, especially in large scale computations with numbers.

Why has this revolution happened?

1. In continuous signaling (transmission) you often have to amplify the signal to compensate for natural losses along the way. Any error made at one stage, before or during amplification, is naturally amplified by the next stage. For example, the telephone company in sending a voice across the continent might have a total amplification factor of 10^{120}. At first 10^{120} seems to be very large so we do a quick back of the envelop modeling to see if it is reasonable. Consider the system in more detail. Suppose each amplifier has a gain of 100, and they are spaced every 50 miles. The actual path of the signal may well be over 3000 miles, hence some 60 amplifiers, hence the above factor does seem reasonable now we have seen how it can arise. It should be evident such amplifiers had to be built with exquisite accuracy if the system was to be suitable for human use.

Compare this to discrete signaling. At each stage we do not amplify the signal, but rather we use the incoming pulse to gate, or not, a standard source of pulses; we actually use *repeaters, not amplifiers*. Noise introduced at one spot, if not too much to make the pulse detection wrong at the next repeater, is automatically removed. Thus with remarkable fidelity we can transmit a voice signal if we use digital signaling, and furthermore the equipment need not be built extremely accurately. We can use, if necessary, error detecting and error correcting codes to further defeat the noise. We will examine these codes later, Chapters 10–12. Along with this we have developed the area of digital filters which are often much more versatile, compact, and cheaper than are analog filters, Chapters 14–17. We should note here *transmission through space* (typically signaling) is the same as *transmission through time* (storage).

Digital computers can take advantage of these features and carry out very deep and accurate computations which are beyond the reach of analog computation. Analog computers have probably passed their peak of importance, but should not be dismissed lightly. They have some features which, so long as great accuracy or deep computations are not required, make them ideal in some situations.

2. The invention and development of transistors and the integrated circuits, ICs, has greatly helped the digital revolution. Before ICs the problem of soldered joints dominated the building of a large computer,

and ICs did away with most of this problem, though soldered joints are still troublesome. Furthermore, the high density of components in an IC means lower cost and higher speeds of computing (the parts must be close to each other since otherwise the time of transmission of signals will significantly slow down the speed of computation). The steady decrease of both the voltage and current levels has contributed to the partial solving of heat dissipation.

It was estimated in 1992 that interconnection costs were approximately:

Interconnection on the chip	$\$10^{-5}=0.001$ cent
Interchip	$\$10^{-2}=1$ cent
Interboard	$\$10^{-1}=10$ cents
Interframe	$\$10^{0}=100$ cents

3. Society is steadily moving from a material goods society to an information service society. At the time of the American Revolution, say 1780 or so, over 90% of the people were essentially farmers—now farmers are a very small percent of workers. Similarly, before WW-II most workers were in factories—now less than half are there. In 1993, there were more people in Government (excluding the military), than there were in manufacturing! What will the situation be in 2020? As a guess I would say less than 25% of the people in the civilian work force will be handling things, the rest will be handling information in some form or other. In making a movie or a TV program you are making not so much a thing, though of course it does have a material form, as you are organizing information. Information is, of course, stored in a material form, say a book (the essence of a book is information), but information is not a material good to be consumed like food, a house, clothes, an automobile, or an airplane ride for transportation.

The information revolution arises from the above three items plus their synergistic interaction, though the following items also contribute.

4. The computers make it possible for robots to do many things, including much of the present manufacturing. Evidently computers will play a dominant role in robot operation, though one must be careful not to claim the standard von Neumann type of computer will be the sole control mechanism, rather probably the current neural net computers, fuzzy set logic, and variations will do much of the control. Setting aside the child's view of a robot as a machine resembling a human, but rather thinking of it as a device for handling and controlling things in the material world, robots used in manufacturing do the following:

A. Produce a better product under tighter control limits.
B. Produce usually a cheaper product.
C. Produce a different product.

This last point needs careful emphasis.

When we first passed from hand accounting to machine accounting we found it necessary, for economical reasons if no other, to somewhat alter the accounting system. Similarly, when we passed from strict hand fabrication to machine fabrication we passed from mainly screws and bolts to rivets and welding.

> It has rarely proved practical to produce exactly the same product by machines as we produced by hand.

Indeed, one of the major items in the conversion from hand to machine production is the imaginative redesign of an *equivalent product*. Thus in thinking of mechanizing a large organization, it won't work if you try to keep things in detail exactly the same, rather there must be a larger give-and-take if there is to be a significant success. You must get the essentials of the job in mind and then design the mechanization to do that job rather than trying to mechanize the current version—if you want a significant success in the long run.

I need to stress this point; mechanization requires you produce an equivalent product, not identically the same one. Furthermore, in any design it is now essential to consider field maintenance since in the long run it often dominates all other costs. The more complex the designed system the more field maintenance must be central to the final design. Only when field maintenance is part of the original design can it be safely controlled; it is not wise to try to graft it on later. This applies to both mechanical things and to human organizations.

5. The effects of computers on Science have been very large, and will probably continue as time goes on. My first experience in large scale computing was in the design of the original atomic bomb at Los Alamos. There was no possibility of a small scale experiment either you have a critical mass or you do not—and hence computing seemed at that time to be the only practical approach. We simulated, on primitive IBM accounting machines, various proposed designs, and they gradually came down to a design to test in the desert at Alamagordo, NM.

From that one experience, on thinking it over carefully and what it meant, I realized computers would allow the simulation of many different kinds of experiments. I put that vision into practice at Bell Telephone Laboratories for many years. Somewhere in the mid-tolate 1950s in an address to the President and V.Ps of Bell Telephone Laboratories I said, "At present we are doing 1 out of 10 experiments on the computers and 9 in the labs, but before I leave it will be 9 out of 10 on the machines". They did not believe me then, as they were sure real observations were the key to experiments and I was just a wild theoretician from the mathematics department, but you all realize by now we do somewhere between 90 % to 99 % of our experiments on the machines and the rest in the labs. And this trend will go on! It is so much cheaper to do simulations than real experiments, so much more flexible in testing, and we can even do things which cannot be done in any lab, that it is inevitable the trend will continue for some time. Again, the product was changed!

But you were all taught about the evils of the Middle Age scholasticism—people deciding what would happen by reading in the books of Aristotle (384–322) rather than looking at Nature. This was Galileo's (1564–1642) great point which started the modern scientific revolution—look at Nature not in books! But what was I saying above? We are now looking more and more in books and less and less at Nature! There is clearly a risk we will go too far occasionally—and I expect this will happen frequently in the future. We must not forget, in all the enthusiasm for computer simulations, occasionally we must look at Nature as She is.

6. Computers have also greatly affected Engineering. Not only can we design and build far more complex things than we could by hand, we can explore many more alternate designs. We also now use computers to control situations such as on the modern high speed airplane where we build unstable designs and then use high speed detection and computers to stabilize them since the unaided pilot simply cannot fly them directly. Similarly, we can now do unstable experiments in the laboratories using a fast computer to control the instability. The result will be that the experiment will measure something very accurately right on the edge of stability.

As noted above, Engineering is coming closer to Science, and hence the role of simulation in unexplored situations is rapidly increasing in Engineering as well as Science. It is also true *computers are now often an essential component of a good design*.

In the past Engineering has been dominated to a great extent by “what can we do”, but now “what do we want to do” looms greater since we now have the power to design almost anything we want. More than ever before, Engineering is a matter of choice and balance rather than just doing what can be done. And more and more it is the human factors which will determine good design—a topic which needs your serious attention at all times.

7. The effects on society are also large. The most obvious illustration is computers have given top management the power to *micromanage* their organization, and top management has shown little or no ability to resist using this power. You can regularly read in the papers some big corporation is decentralizing, but when you follow it for several years you see they merely intended to do so, but did not.

Among other evils of micromanagement is lower management does not get the chance to make responsible decisions and learn from their mistakes, but rather because the older people finally retire then lower management finds itself as top management —without having had many real experiences in management!

Furthermore, central planning has been repeatedly shown to give poor results (consider the Russian experiment for example or our own bureaucracy). The persons on the spot usually have better knowledge than can those at the top and hence can often (not always) make better decisions if things are not micromanaged. The people at the bottom do not have the larger, global view, but at the top they do not have the local view of all the details, many of which can often be very important, so *either extreme gets poor results*.

Next, an idea which arises in the field, based on the direct experience of the people doing the job, cannot get going in a centrally controlled system since the managers did not think of it themselves. The *not invented here* (NIH) syndrome is one of the major curses of our society, and computers with their ability to encourage micromanagement are a significant factor.

There is slowly coming, but apparently definitely, a counter trend to micromanagement. Loose connections between small, somewhat independent organizations, are gradually arising. Thus in the brokerage business one company has set itself up to sell its services to other small subscribers, for example, computer and legal services. This leaves the brokerage decisions of their customers to the local management people who are close to the front line of activity. Similarly, in the pharmaceutical area some loosely related companies carry out their work and intertrade among themselves as they see fit. I believe you can expect to see much more of this loose association between small organizations as a defense against micromanagement from the top which occurs so often in big organizations. There has always been some independence of subdivisions in organizations, but the power to micromanage from the top has apparently destroyed the conventional lines and autonomy of decision making—and I doubt the ability of most top managements to resist for long the power to micromanage. I also doubt many large companies will be able to give up micromanagement; most will probably be replaced in the long run by smaller organizations without the cost (overhead) and errors of top management. Thus computers are affecting the very structure of how Society does its business, and for the moment apparently for the worse in this area.

8. Computers have already invaded the entertainment field. An informal survey indicates the average American spends far more time watching TV than in eating-again an information field is taking precedence over the vital material field of eating! Many commercials and some programs are now either partially or completely computer produced.

How far machines will go in changing society is a matter of speculation—which opens doors to topics that would cause trouble if discussed openly! Hence I must leave it to your imaginations as to what, using computers on chips, can be done in such areas as sex, marriage, sports, games, “travel in the comforts of home via virtual realities”, and other human activities.

Computers began mainly in the number crunching field but passed rapidly on to information retrieval (say airline reservation systems), word processing which is spreading everywhere, symbol manipulation as is done by many programs such as those which can do analytic integration in the calculus far better and cheaper than can the students, and in logical and decision areas where many companies use such programs to control their operations from moment to moment. The future computer invasion of traditional fields remains to be seen and will be discussed later under the heading of artificial intelligence (AI), Chapters 6–8.

9. In the military it is easy to observe (in the Gulf War for example), the central role of information, and the failure to use the information about one's own situation killed many of our own people! Clearly that war was one of information above all else, and it is probably one indicator of the future. I need not tell you such things since you are all aware, or should be, of this trend. It is up to you to try to foresee the situation in the year 2020 when you are at the peak of your careers. I believe computers will be almost everywhere since I once saw a sign which read, "The battle field is no place for the human being". Similarly for situations requiring constant decision making. The many advantages of machines over humans were listed near the end of the last chapter and it is hard to get around these advantages, though they are certainly not everything. Clearly the role of humans will be quite different from what it has traditionally been, but many of you will insist on old theories you were taught long ago as if they would be automatically true in the long future. It will be the same in business, much of what is now taught is based on the past, and has ignored the computer revolution and our responses to some of the evils the revolution has brought; the gains are generally clear to management, the evils are less so.

How much the trends, predicted in part 6 above, toward and away from micromanagement will apply widely and is again a topic best left to you—but you will be a fool if you do not give it your deep and constant attention. I suggest you must rethink *everything* you ever learned on the subject, question every successful doctrine from the past, and finally decide for yourself its future applicability. The Buddha told his disciples, "Believe nothing, no matter where you read it, or who said it, no matter if I have said it, unless it agrees with your own reason and your own common sense". I say the same to you—*you must assume the responsibility for what you believe*.

I now pass on to a topic that is often neglected, the rate of evolution of some special field which I will treat an another example of "back of the envelop computation". The growth of most, but by no means all, fields follow an "S" shaped curve. Things begin slowly, then rise rapidly, and later flatten off as they hit some natural limits.

The simplest model of growth assumes the rate of growth is proportional to the current size, something like compound interest, unrestrained bacterial and human population growth, as well as many other examples. The corresponding differential equation is

$$\frac{dy}{dt} = ky$$

whose solution is, of course,

$$y(t) = Ae^{kt}.$$

But this growth is unlimited and all things must have limits, even knowledge itself since it must be recorded in some form and we are (currently) told the universe is finite! Hence we must include a limiting factor in the differential equation. Let L be the upper limit. Then the next simplest growth equation seems to be

$$\frac{dy}{dt} = ky(L - y).$$

At this point we, of course, reduce it to a standard form that eliminates the constants. Set $y=Lz$, and $x=t/kL^2$, then we have

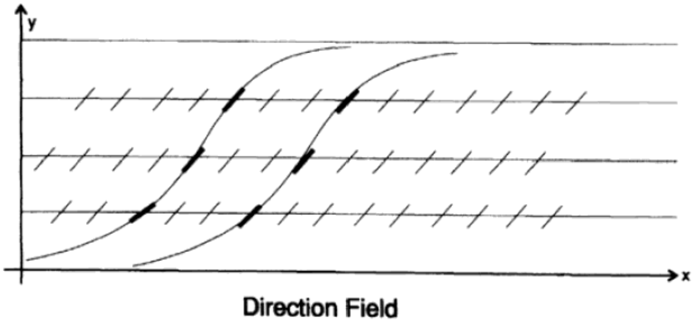

Figure 2.I

$$\frac{dz}{dx} = z(1 - z)$$

as the *reduced form* for the growth problem, where the saturation level is now 1. Separation of variables plus partial fractions yields:

$$\ln z - \ln(1 - z) = x + C,$$

$$\frac{z}{1 - z} = Ae^{x},$$

$$z = \frac{1}{[1 + (1/A)e^{-x}]}.$$

A is, of course, determined by the initial conditions, where you put *t* (or *x*)=0. You see immediately the "S" shape of the curve; at $t = -\infty$, z=0; at t=0, z=$A/(A+1)$; and at $t = +\infty$, z=1.

A more flexible model for the growth is (in the reduced variables)

$$\frac{dz}{dx} = z^{a}(1 - z)^{b}, \quad (a, b > 0).$$

This is again a variables separable equation, and also yields to numerical integration if you wish. We can analytically find the steepest slope by differentiating the right hand side and equating to 0. We get

$$a(1 - z) - bz = 0.$$

Hence at the place

$$z = \frac{a}{(a + b)},$$

we have the maximum slope

$$\frac{a^{a}b^{b}}{(a + b)^{a+b}}.$$

A direction field sketch Figure 2.I will often indicate the nature of the solution and is particularly easy to do as the slope depends

only on *y* and not on *x*—the isoclines are horizontal lines so the solution can be slid along the *x*-axis without changing the "shape" of the solution. For a given *a* and *b there is really only one shape,* and the

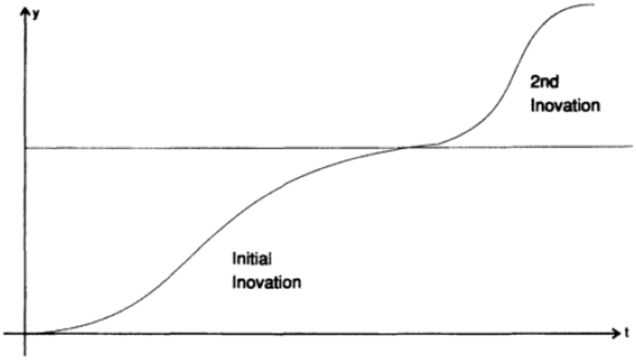

Figure 2.II

initial conditions determine where you look, not what you look at. When the differential equation has coefficients which do not depend on the independent variable then you have this kind of effect.

In the special case of $a=b$ we have

maximum slope$=1/2^{2a}$.

The curve will in this case be odd symmetric about the point where $z=1/2$.

In the further special case of $a=b=1/2$ we get the solution

$$z=\sin^2(x/2+C), \quad (-C\leq x/2\leq \pi - C).$$

Here we see the solution curve has a finite range. For larger exponents a and b we have clearly an infinite range.

As an application of the above consider the rate of increase in computer operations per second has been fairly constant for many years—thus we are clearly on the almost straight line part of the "S" curve. (More on this in the next chapter.) In this case we can more or less know the saturation point for the von Neumann, single processor, type of computer since we believe: (1) the world is made out of molecules, and (2) using the evidence from the two relativity theories, special and general, gives a maximum speed of useful signaling, then there are definite limits to what can be done with a single processor. The trend to highly parallel processors is the indication we are feeling the upper saturation limit of the "S" curve for single processor computers. There is also the nasty problem of heat dissipation to be considered. We will discuss this matter in more detail in the next chapter.

Again we see how a simple model, while not very exact in detail, suggests the nature of the situation. Whether parallel processing fits into this picture, or is an independent curve is not clear at this moment. Often a new innovation will set the growth of a field onto a new "S" curve which takes off from around the saturation level of the old one, Figure 2.II. You may want to explore models which do not have a hard upper saturation limit but rather finally grow logarithmically; they are sometimes more appropriate.

It is evident Electrical Engineering in the future is going to be, to a large extent, a matter of: (1) selecting chips off the shelf or from a catalog, (2) putting the chips together in a suitable manner to get what you want, and (3) writing the corresponding programs. Awareness of the chips, and circuit boards which are currently available will be an essential part of Engineering, much as the *Vacuum Tube Catalog* was in the old days.

As a last observation in this area let me talk about special purpose IC chips. It is immensely ego gratifying to have special purpose chips for your special job, but there are very high costs associated with them. First, of course, is the design cost. Then there is the "trouble shooting" of the chip. Instead, if you will find a general purpose chip, which may possibly cost a bit more, then you gain the following advantages:

1. Other users of the chip will help find the errors, or other weaknesses, if there are any.
2. Other users will help write the manuals needed to use it.
3. Other users, including the manufacturer, will suggest upgrades of the chip, hence you can expect a steady stream of improved chips with little or no effort on your part.
4. Inventory will not be a serious problem.
5. Since, as I have been repeatedly said, technical progress is going on at an increasing rate, it follows technological obsolescence will be much more rapid in the future than it is now. You will hardly get a system installed and working before there are significant improvements which you can adapt by mere program changes *If* you have used general purpose chips and good programming methods rather than your special purpose chip which will almost certainly tie you down to your first design.

Hence beware of special purpose chips!

though many times they are essential.

3

History of Computers— Harware

The history of computing probably began with primitive man using pebbles to compute the sum of two amounts. Marshack (of Harvard) found what had been believed to be mere scratches on old bones from cave man days were in fact carefully scribed lines apparently connected with the moon's phases. The famous Stonehenge on the Salisbury plain in England had three building stages, 1900–1700, 1700–1500, and 1500–1400 B.C., and were apparently closely connected with astronomical observations, indicating considerable astronomical sophistication. Work in archeoastronomy has revealed many primitive peoples had considerable knowledge about astronomical events. China, India, and Mexico were prominent in this matter, and we still have their structures which we call observatories, though we have too little understanding of how they were used. Our western plains have many traces of astronomical observatories which were used by the Indians.

The *sand pan* and the *abacus* are instruments more closely connected with computing, and the arrival of the Arabic numerals from India meant a great step forward in the area of pure computing. Great resistance to the adoption of the Arabic numerals (not in their original Arabic form) was encountered from officialdom, even to the extent of making them illegal, but in time (the 1400s) the practicalities and economic advantages triumphed over the more clumsy Roman (and earlier Greek) use of letters of the alphabet as symbols for the numbers.

The invention of logarithms by Napier (1550–1617) was the next great step. From it came the slide rule, which has the numbers on the parts as lengths proportional to the logs of the numbers, hence adding two lengths means multiplying the two numbers. This analog device, the slide rule, was another significant step forward, but in the area of analog not digital computers. I once used a very elaborate slide rule in the form of a (6–8”) diameter cylinder and about two feet long, with many, many suitable scales on both the outer and inner cylinders, and equipped with a magnifying glass to make the reading of the scales more accurate.

Slide rules in the 1930s and 1940s were standard equipment of the engineer, usually carried in a leather case fastened to the belt as a badge of one's group on the campus. The standard engineer's slide rule was a “10 inch loglog decitrig slide rule” meaning the scales were 10” long, included loglog scales, square and cubing scales, as well as numerous trigonometric scales in decimal parts of the degree. They are no longer manufactured!

Continuing along the analog path, the next important step was the differential analyzer, which at first had mechanical integrators of the analog form. The earliest successful ones were made around 1930 by Vannevar Bush of MIT. The later RDA #2, while still analog and basically mechanical, had a great deal of electronic interconnections. I used it for some time (1947–1948) in computing Nike guided missile trajectories in the earliest design stages.

During WW-II the electronic analog computers came into the military field use. They used condensers as integrators in place of the earlier mechanical wheels and balls (hence they could only integrate with respect to time). They meant a large, practical step forward, and I used one such machine at Bell Telephone Laboratories for many years. It was constructed from parts of some old M9 gun directors. Indeed, we used parts of some later condemned M9s to build a second computer to be used either independently or with the first one to expand its capacity to do larger problems.

Returning to digital computing Napier also designed “Napier’s bones” which were typically ivory rods with numbers which enabled one to multiply numbers easily; these are digital and not to be confused with the analog slide rule.

From the Napier bones probably came the more modern desk calculators. Schickert wrote (Dec. 20, 1623) to Kepler (of astronomical fame) that a fire in his lab burned up the machine he was building for Kepler. An examination of his records and sketches indicates it would do the four basic operations of arithmetic — provided you have some charity as to just what multiplication and division are in such a machine. Pascal (1623–1662) who was born that same year is often credited with the invention of the desk computer, but his would only add and subtract—only those operations were needed to aid his tax assessing father. Leibnitz (of calculus fame) also tinkered with computers and included multiplication and division, though his machines were not reliable.

Babbage (1791–1871) is the next great name in the digital field, and he is often considered to be the father of modern computing. His first design was the *difference engine,* based on the simple idea that a polynomial can be evaluated at successive, equally spaced, values by using only a sequence of additions and subtractions, and since locally most functions can be represented by a suitable polynomial this could provide “machine made tables” (Babbage insisted the printing be done by the machine to prevent any human errors creeping in). The English Government gave him financial support, but he never completed one. A Norwegian father and son (Scheutz) did make several which worked and Babbage congratulated them on their success. One of their machines was sold to the Albany observatory, New York, and was used to make some astronomical tables.

As has happened so often in the field of computing, Babbage had not finished with the difference engine before he conceived of the much more powerful *analytical engine,* which is not far from the current von Neumann design of a computer. He never got it to work; a group in England constructed (1992) a machine from his working drawings and successfully operated it as he had designed it to work!

The next major practical stage was the Comptometer which was merely an adding device, but by repeated additions, along with shifting, this is equivalent to multiplication, and was very widely used for many, many years.

From this came a sequence of more modern desk calculators, the Millionaire, then the Marchant, the Friden, and the Monroe. At first they were hand controlled and hand powered, but gradually some of the control was built in, mainly by mechanical levers. Beginning around 1937 they gradually acquired electric motors to do much of the power part of the computing. Before 1944 at least one had the operation of square root incorporated into the machine (still mechanical levers intricately organized). Such hand machines were the basis of computing groups of people running them to provide computing power. For example, when I came to the Bell Telephone Laboratories in 1946 there were four such groups in the Labs, typically about six to ten girls in a group; a small group in the Mathematics department, a larger one in network department, one in switching, and one in quality control.

Punched card computing began because one far seeing person saw the Federal census, that by law must be done every 10 years, was taking so much time the next one (1890) would not be done before the following one started *unless* they turned to machine methods. Hollerith, took on the job and constructed the first punched card machines, and with succeeding censuses he built more powerful machines to keep up with both the increased population and the increased number of questions asked on the census. In 1928 IBM began to use cards with rectangular holes so electric brushes could easily detect the presence or absence of a hole on a card at a given place. Powers, who also left the census group, kept the card form with round holes which were designed to be detected by mechanical rods as “fingers”.

Around 1935 the IBM built the 601 mechanical punch which did multiplications, and could include two additions to the product at the same time. It became one of the mainstays of computingthere were about 1500 of them on rental and they averaged perhaps a multiplication per 2 or 3 seconds. These, along with some special triple product and division machines, were used at Los Alamos to compute the designs for the first atomic bombs.

In the mechanical, meaning relay, area George Stibitz built (1939) the complex number computer and exhibited it at Dartmouth (1940) when the main frame was in New York, thus an early remote terminal machine, and since it normally had three input stations in different locations in the Labs it was, if you are kind, a “time shared machine”.

Konrad Zuse in Germany, and Howard Aitken at Harvard, like Stibitz, each produced a series of relay computers of increasing complexity. Stibitz’s Model 5 had two computers in the same machine and could share a job when necessary, a multiprocessor machine if you wish. Of the three men probably Zuse was the greatest, considering both the difficulties he had to contend with and his later contributions to the software side of computing.

It is usually claimed the electronic computer age began with the ENIAC built for the U.S. Army and delivered in 1946. It had about 18,000 vacuum tubes, was physically huge, and as originally designed it was wired much like the IBM plug boards, but its interconnections to describe any particular problem ran around the entire machine room! So long as it was used, as it was originally intended, to compute ballistic trajectories, this defect was not serious. Ultimately, like the later IBM CPC, it was cleverly rearranged by the users to act as if it were programmed from instructions (numbers on the ballistic tables) rather than from wiring the interconnections.

Mauchly and Eckert, who built the ENIAC, found, just as Babbage had, before the completion of their first machine they already envisioned a larger, internally programmed, machine, the EDVAC. Von Neumann, as a consultant to the project, wrote up the report, and as a consequence internal programming is often credited to him, though so far as I know he never either claimed or denied that attribution. In the summer of 1946, Mauchly and Eckert gave a course, *open to all,* on how to design and build electronic computers, and as a result many of the attendees went off to build their own; Wilkes, of Cambridge, England, being the first to get one running usefully, the EDSAC.

At first each machine was a one-of-a-kind, though many were copied from (but often completed before) the Institute for Advanced Studies machine under von Neumann’s direction, because the engineering of that machine was apparently held up. As a result, many of the so-called copies, like the Maniac-I (1952) (which was named to get rid of the idiotic naming of machines), and built under the direction of N.C. Metropolis, was finished before the Institute machine. It, and the Maniac-II (1955), were built at Los Alamos, while the Maniac-III (1959) was built at the University of Chicago. The Federal government, especially through the military, supported most of the early machines, and great credit is due to them for helping start the Computer Revolution.

The first commercial production of electronic computers was under Mauchly and Eckert again, and since the company they formed was merged with another, their machines were finally called UNIVACS. Especially noted was the one for the Census Bureau. IBM came in a bit late with 18 (20 if you count secret cryptographic users) IBM 701s. I well recall a group of us, after a session on the IBM 701 at a meeting where they talked about the proposed 18 machines, all believed this would saturate the market for many years! Our error was simply we thought only of the kinds of things we were currently doing, and did not think in