LECTURE 2

FOUNDATIONS OF THE DIGITAL (DISCRETE) REVOLUTION

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 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 Lectures, 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 that 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 that we have seen how it can arise. It should be evident that 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 <u>repeaters, not amplifiers</u>. 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, Lectures 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, Lectures 14-17. We should note here that <u>transmission through space</u> (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 that 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 that, 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 WWII most workers were in factories - now less than half are there. In (1993) there were more people in government, (excluding the military), than there ~~are~~ were in manufacturing! What will the situation be in 2020? As a guess I would say that 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 things 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 that 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 that we produced by hand.

Indeed, one of the major items in the conversion from hand to machine production is the imaginative redesign of an <u>equivalent product</u>. 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 this or that current version - if you want a significant success in the long run.

I need to stress this point; mechanization requires that you produce an equivalent product, not identically the same one. Furthermore, in any design it is now essential that field maintenance be considered 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 had a critical mass or you did 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, N.M..

From that one experience, on thinking it over carefully and what it meant, I realized that 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 50's in an address to the President and V.P.s 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 that real observations were the key to experiments and that I was just a wild theoretician from the mathematics department, but you all realize that 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 that can not be done in any lab, that it is inevitable that 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 that started the modern scientific revolution - look at Nature not in books! But what was I saying above? That we are now looking more and more in books and less and less at Nature! There is clearly a risk that we will go too far occasionally - and I expect that this will happen frequently in the future. We must not forget, in all the enthusiasm for computer simulations, that 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 that <u>computers are now often an essential component of a good design</u>.

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 that will determine good design - a topic that needs your serious attention at all times.

7. The effects on society are also large. The most obvious illustration is that 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 that some big corporation is decentralizing, but when you follow it for several years you see that they merely intended to do so, but in fact did not.

Among other evils of micromanagement is the fact that 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 ~~much~~ 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, ideas that arise 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 their own 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 inter-trade 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 that 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 that 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 that 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 - and that 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 reservations systems), word processing which is spreading everywhere, symbol manipulation as is done by many programs such as those that 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), Lectures 6-8.

9. In the military it is easy to observe, (in the Gulf War for example), the central role of information, and that 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 as military people to try to foresee the situation in the year 2020 when you are at the peak of your careers. I believe that computers will be almost everywhere since I once saw a sign that read, "The battle field is no place for the human being." The many advantages of machines over humans were listed near the end of the last Lecture 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 doctrines you were taught 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 to the military 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 that you must rethink <u>everything</u> 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 - <u>you must assume the responsibility for what you believe.</u>

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 is that 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

$$dy/dt = ky$$

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 that 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

$$dy/dt = ky(L - y)$$

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

$$dz/dx = z(1 - z)$$

as the <u>reduced form</u> for the growth problem, where the saturation level is now 1. Separation of variables plus partial fractions yields

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

$$z/(1 - z) = Ae^{x}$$

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

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)

$$dz/dx = z^a(1 - z)^b \qquad (a, b > 0)$$

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

$$z = a/(a + b)$$

we have the maximum slope

$$a^a b^b/(a + b)^{a + b}$$

A direction field sketch Figure 2-1 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 <u>there is really only one shape</u>, and the initial conditions determine where you look, not what you look at. When the differential equation has coefficients that do not depend on the independent variable then you have this kind of effect.

In the special case of a = b we have

$$\text{maximum slope} = 1/2^{2a}$$

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

$$z = \sin^2(x/2 + C) \qquad (-C \le x/2 \le \pi - C)$$

Here we see that 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 that 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 Lecture). 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) that the world is made out of molecules, and (2) using the evidence that 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 that 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 Lecture.

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 that takes off from around the saturation level of the old one, Figure 2-2. 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 that 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 manner to get what you want, and (3) writing the corresponding programs. Awareness of the chips, and circuit boards that 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, that 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.

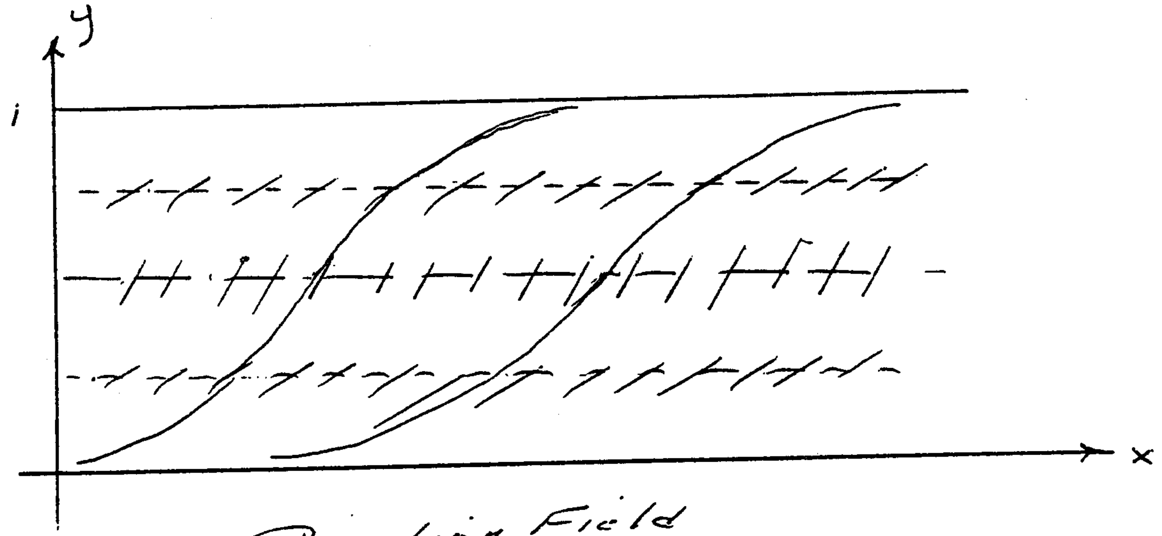
3. Other users, including the manufacturer, will suggest upgrades of the chip, hence you can expect a steady stream of improved chips with little or no effort on your part.

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 that 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 that you can adapt by mere program changes _if_ you have used general purpose chips and good progamming methods rather than your special purpose chipwhich will almost certainly tie you down to your first design .

Hence beware of special purpose chips!

though many times they are essential.



Direction Field

Figure 2-1

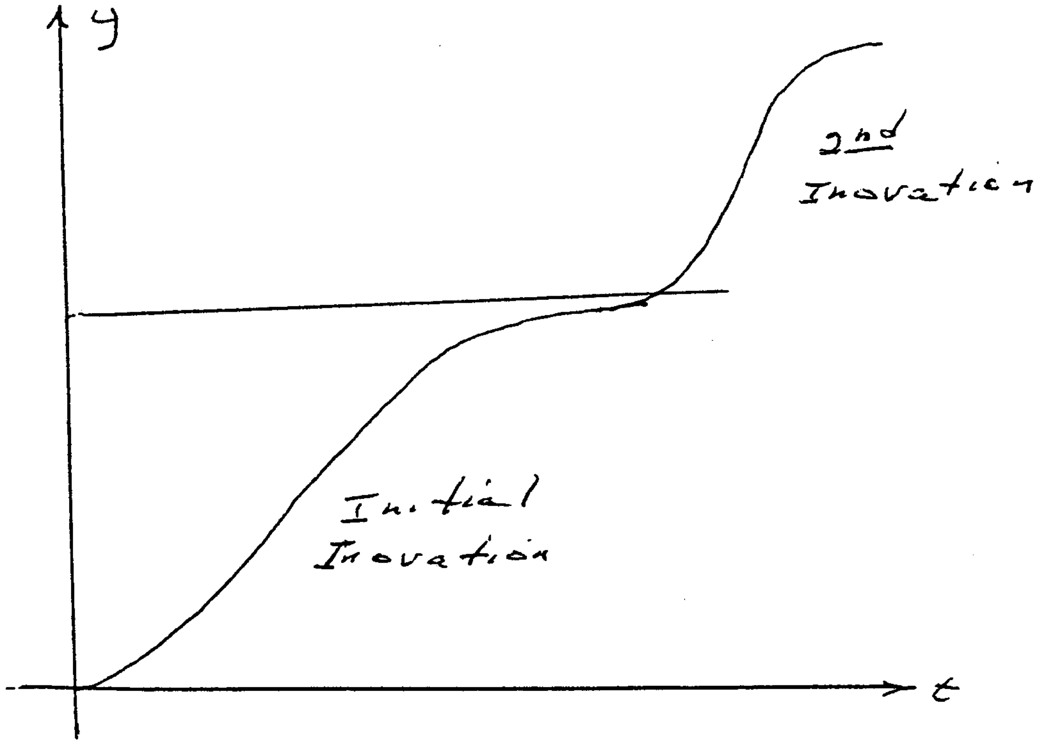


Figure 2-2