DIGITAL ELECTRONIC DESIGN
VOLUME 2

CATT. WALTON & DAVIDSON

C.A.M. PUBLISHING

DIGITAL ELECTRONIC DESIGN

"Then there were the remarkable researches of Faraday, the prince of experimentalists, on electrostatics and electrodynamics and the induction of currents...The crowning achievement was reserved for the heaven - sent Maxwell, a man whose fame, great as it is now, has, comparatively speaking, yet to come."("Electromagnetic Theory" Vol. 1, page 8, by Oliver Heaviside.)

"We reverse this; the current in the wire is set up by the energy transmitted through the medium around it." ("Electrical Papers" Vol. 1, page 438, by Oliver Heaviside.)

DIGITAL ELECTRONIC DESIGN

Volume 2

By

Ivor Catt
David Walton
Malcolm Davidson

C.A.M. Publishing
121 Westfields
St. Albans
AL3 4JR
England

First published March 1979
Reprinted September 1979
Reprinted March 1980
Reprinted December 1980
Reprinted October 1986
Reprinted 1994

Some parts of this volume previously appeared in WIRELESS WORLD.

Diagrams drawn by Valerie Minors

121 Woodlands
St. Albans
AL3 4JB
England

ISBN: 0 906360 01 2

PREFACE

Volume 2 contains more practical advice on day to day digital design, along with some disclosures of fundamental advances in the art. Further disclosures will be found in the sister publication "Electromagnetic Theory" by the same authors.

Volume 1 was well received and is now on the shelves of about half the university and research establishment libraries. Since it was published, interest in the authors' quarterly seminar on digital design has greatly increased.

A good working relationship has developed with Wireless World, where the authors' occasional articles and the resulting correspondence are well worth reading.

-lso see "Electromagnetism 1" by Ivor Catt, pub. Westfields Press, 121 Westfields, St. Albans AL3 4JR, England, 1994.

The publishers have a helpful video driven by an Acorn Master computer which demonstrates the passage of a TEM wave.

CONTENTS

VOLTAGE SUPPLY DECOUPLING BY CAPACITORS

The conventional view is that a capacitor has stray series inductance and stray series resistance, and the model normally used for a practical capacitor is therefore as follows:

The series resistance, called "Equivalent Series Resistance", (E.S.R.), is less of an evil than the series inductance, the stray value which is thought about the most.

It is generally thought that large capacitors have large series induc-

tance. This leads people to use more than one capacitor in parallel, the smaller 10 nF capacitor having the job of dealing with the higher frequencies and the larger 10 μF capacitor dealing with lower frequencies.

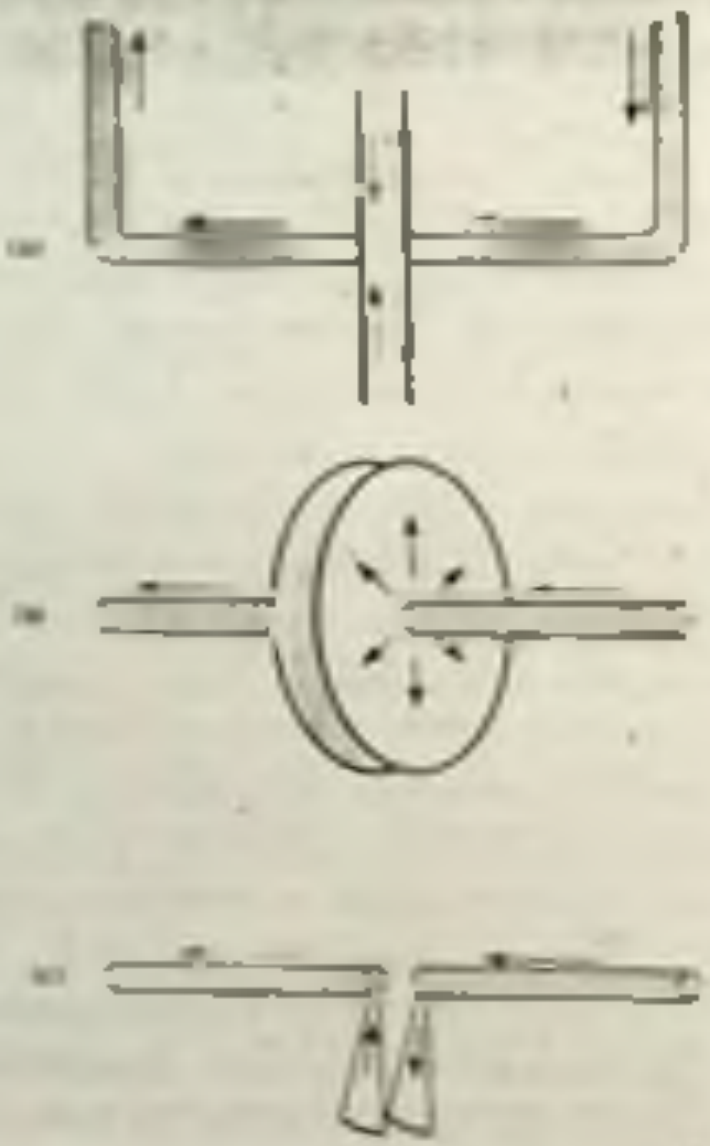
"Self Resonant Frequency" for a capacitor is the term given to the frequency ω when $1/\omega^2$ equals the product of C and the stray inductance, L. Above this frequency, it is said, the capacitor becomes inductive and is useless in a place where a capacitor and not an inductor is needed, like decoupling voltage supplies to logic.

The above arguments, although possibly reasonably harmless when designing analog circuits, become definitely counter - productive for the digital designer, who should instead follow the reasoning in the rest of this chapter

Conventional Electromagnetic Theory proposes that when an electric current flows down a wire into a capacitor, it spreads out across the plate producing an electric charge which in turn leads to an electric field between the capacitor plates. The valuable concept of continuity of electric current is then

retained by postulating, (after Maxwell) a "Displacement Current", which is a mathematical manipulation of the electric field E between the capacitor plates which has the dimensions of electric current and completes the flow of "electricity". This approach permits us to retain Kirchoff's Laws and other valuable concepts, even though superficially it appears that at the capacitor there is a break in the otherwise continuous flow of electric current.

The flaw in this model is revealed when we notice that the electric current entered the capacitor at one point only on the capacitor plate. We must explain how, then, the electric charge flowing down the wire suddenly distributed itself uniformly across the whole capacitor plate. We know that this cannot happen since charge cannot flow out across the plate at a velocity in excess of the velocity of light. This paradoxical situation is brought about by a fundamental flaw in the basic model. Work on high speed logic design has shown that the model of a lumped capacitor is faulty, and "Displacement Current" is an artefact



of this faulty model.

The true model is different. Electric current enters the capacitor through a wire and then spreads out across the plate in the same way as ripples flow out from a stone dropped into a pond. If we consider only one pie-shaped wedge of the capacitor, (see Vol.I, page 110), we can recognise it as a parallel plate transmission line whose only unusual feature is that the line width is increasing (and hence the impedance is decreasing). The capacitor is made up of a number of these pie-shaped transmission lines in parallel, so the proper model for a capacitor is a transmission line.

E.S.R. for a capacitor is the initial characteristic impedance of this transmission line at a radius equal to the radius of the input wires.

Series inductance does not exist. Pace the many documented values for series inductance in a capacitor, this confirms experience that when the so-called series inductance of a capacitor is measured, it turns out to be no more than the series inductance of the

wires connected to the capacitor. No mechanism has ever been proposed for an internal series inductance in a capacitor

Since any capacitor must now be considered to be a transmission line, it is no more necessary to postulate "Displacement Current" in a capacitor than it is necessary for a transmission line.

The excision of "Displacement Current" from Electromagnetic Theory has been based on arguments which are independent of the classic dispute over whether the electric current causes the electromagnetic field or vice versa, that is whether we adhere to Theory N or Theory H (Vol.I, page 119).

In summary, we see that when decoupling a voltage supply, a capacitor looks like a pure resistance of the order of 0.1 ohms. Its performance will not deteriorate as logic speeds increase through 1 ns and beyond.

SELF-RESONANT FREQUENCY OF A CAPACITOR

This is an extremely damaging concept, leading to the use of the least suitable capacitors (i.e. low value

capacitors) for voltage decoupling purposes.

The so-called self-resonant frequency results from the equation;

$$\omega^2 = \frac{1}{LC}$$

where L is the self inductance of the wires leading to the capacitor, and C is the nominal capacitance of the capacitor. It is easy to see from the formula that if C is increased, the so-called self-resonant frequency is reduced. This leads to the nonsensical conclusion that the best capacitor has the smallest capacitance, all other variables being kept constant. In fact, the impedance of a 1 μF capacitor, although by such a doubtful rule of thumb reaching its lowest impedance at a lower frequency, has a lower impedance at all frequencies than a so-called "R-F" 10 nF capacitor even when the latter is at its lowest impedance. The problem is that the impedance curves of the two capacitors have never been drawn on the same graph. When they are, the superiority of the 1 μF capacitor at all relevant frequencies is obvious.

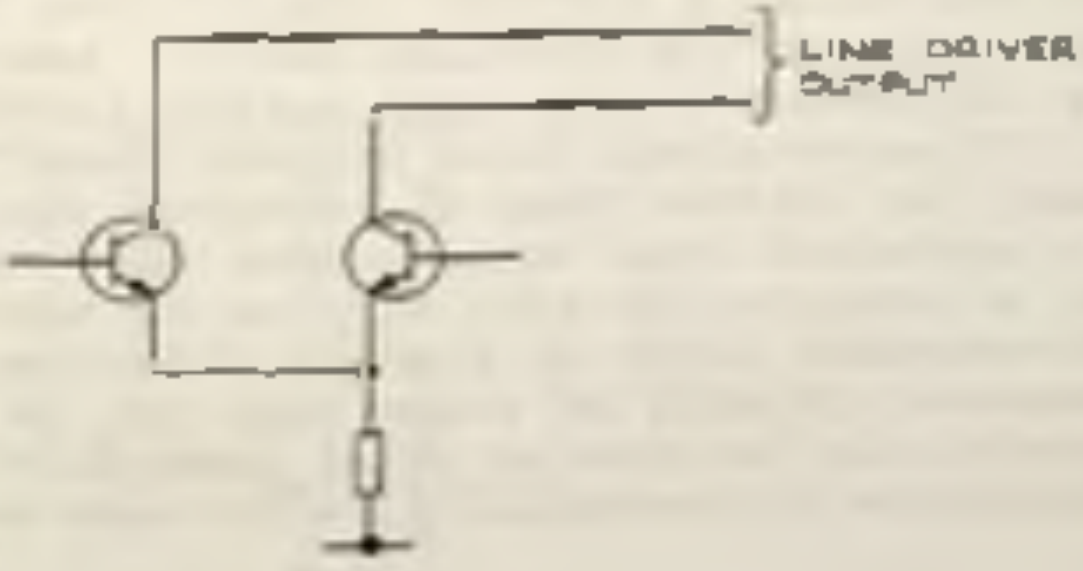
LINE DRIVERS AND RECEIVERS

In the first chapter of Volume One, entitled "Earthing in a Digital System", it is shown that the technique of sending a signal down one wire and using a common return path has severe limitations. In such a system, it appears that the best noise rejection is achieved by having the lowest possible source impedance at the driven end of the line and also the lowest possible destination impedance, so that any noise has to develop a voltage across a very low impedance, and for a given noise source the voltage developed is minimum. (It could be argued that for magnetically induced noise the source impedance should be high, but certainly the destination impedance should be as low as possible.)

When logic signals are sent over a distance, for instance from one module to another, it is prudent to send the signal in differential mode down a pair of lines. The two lines are made as similar as possible, so that any noise

introduced from elsewhere tends to be common mode (equal in amplitude on both lines), and a differential amplifier at the destination (i.e. a circuit which decides which of two inputs is higher, rather than their absolute amplitude) is unaffected by such noise.

One advantage we have when designing line drivers and receivers is that the circuit need not be as fast as a normal logic element. If maximum speed were required, the signal would surely not have been sent a long distance in the first place, with the inevitable delay that this involves (minimum 1 nsec / ft, in practice about 2 nsec / ft).

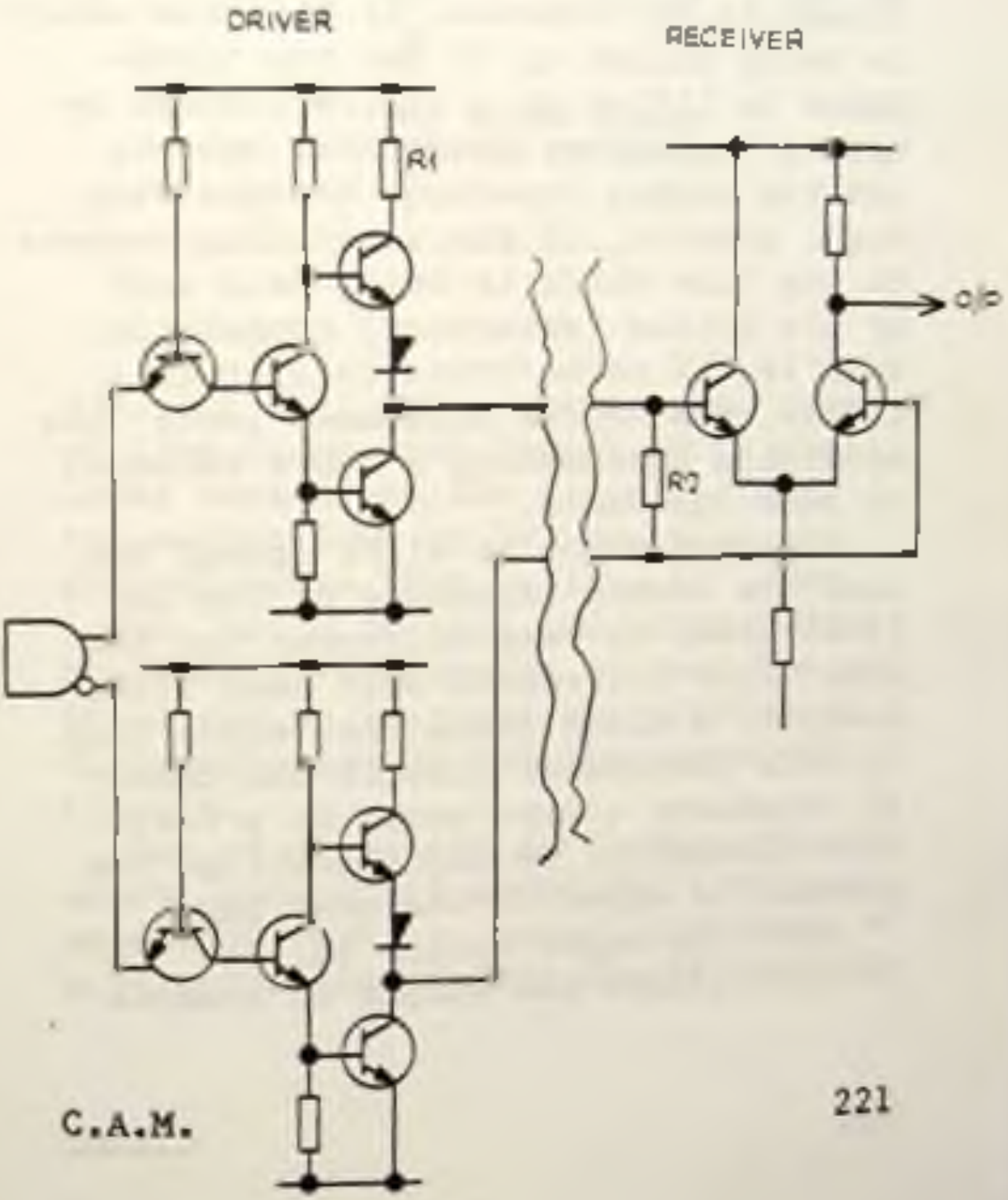


Designers of differential line drivers have given some thought to noise rejection at destination, but have consistently failed to notice that noise rejection at source, at the driver end, is also very important. As a result, they have designed line drivers with very high output impedance, which are of course very poor at soaking up noise and also poor at operating properly, that is, outputting a proper differential signal, in the presence of noise on the lines tending to lift the output pins. In particular, if the outputs come from the two collectors of a long tailed pair, there will be no signal at all on either line if the noise pulls the two collectors down below the potential of the emitters of the two transistors. This can easily happen if the output is a current source (like the one drawn on the previous page); that is, if it has a very high output impedance.

If conventional line drivers are used, it follows that the outputs should be protected from large common mode noise by a potential divider similar to that recommended later on for line receivers. However, it will be shown that it is preferable to give up using currently available differential line drivers and

replace them by 7404 outputs until something even better becomes available.

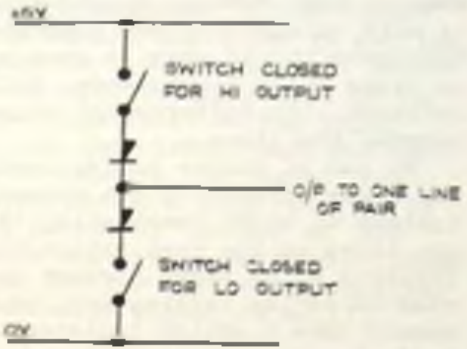
We shall end up with a circuit something like the following:



Generally, it is regarded as important to try to make the source impedance of the two lines as far as possible equal, even though they are being driven differently, one line high and the other line low. Now this is very difficult to achieve because the output drive circuit is non-linear in performance. If the line which is being pulled up by the top transistor is lifted up a little further by noise, the output transistor unhooks and the source impedance becomes very high. However, if the same thing happens to the line which is being held down by the bottom (saturated) transistor, that is, if noise tries to lift it a little, the source impedance looks like something approaching a short circuit, or zero impedance.

At first sight it might appear bad that the source impedance of the two lines looks so very different for the same noise introduced onto each line. However, a close inspection shows that in this particular circuit the change of impedance always works in a favorable direction. We might idealize the circuit as shown on the next page.

With the upper switch on, the circuit clamps the output up towards



+5v, but will not prevent noise from pulling it up even higher. With the lower switch on, the circuit clamps the output down to 0v, but will not prevent noise from pulling it even lower. So in no case can noise reduce the voltage difference between the two lines and so increase the problem of discrimination by the line receiver circuit.

Our study of the particular circuit above leads us to amend our previous assertion that the output impedance of a balanced line driver should look the

same in both states. We now see that it is more accurate to say that the output impedance should vary in such a way as to cause noise to increase, rather than decrease, the differential signal between the lines.

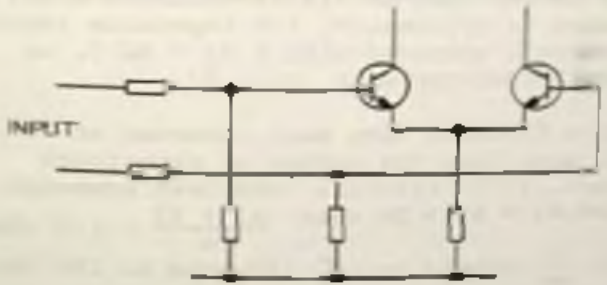
We can of course series terminate the above circuit at the driver end instead of shunt terminating between the lines at the destination with a resistor R2 as shown further on. In the case of series termination, each line should have a series resistor equal to half the characteristic impedance of the two lines. Series termination has the advantage of reducing the power dissipated, but the serious disadvantage of increasing susceptibility at the destination to "capacitively induced noise".

NOISE REJECTION AT DESTINATION

When a balanced pair of lines is used, the problem of differential noise, which affects the two lines differently, is small. However, we still have to face the possibility that common mode noise will lift both lines outside the range of voltages that the line receiver can

tolerate. For example, if the inputs to a differential amplifier go too low, all current in the differential amplifier is cut off. If the inputs go too high, the transistors at the input saturate and the circuit malfunctions.

A way to improve the common mode noise rejection of a line receiver is to have a potential divider at the destination which will attenuate common mode noise (at the cost of attenuating the differential signal as well).



A better way to do the same thing appears to be to put the potential divider between the lines, until we realize that by removing the resistor connected to ground we have increased the destination impedance seen by common mode noise to infinity, and so increased the amplitude of the noise.

An elegant solution to the problem is to common the centre points P of a number of pairs of lines and then decouple P to 0v through a very large capacitor (say 10 μF). This means that noise on the line still sees the small impedance (R1 + R2) to 0v, which is only half the characteristic impedance of the pair of lines, 2 (R1 + R2). It is important that the capacitor should be so big that at all frequencies where noise is of concern, its impedance should be small compared with (R1 + R2), or small compared with $\frac{R1 + R2}{n}$

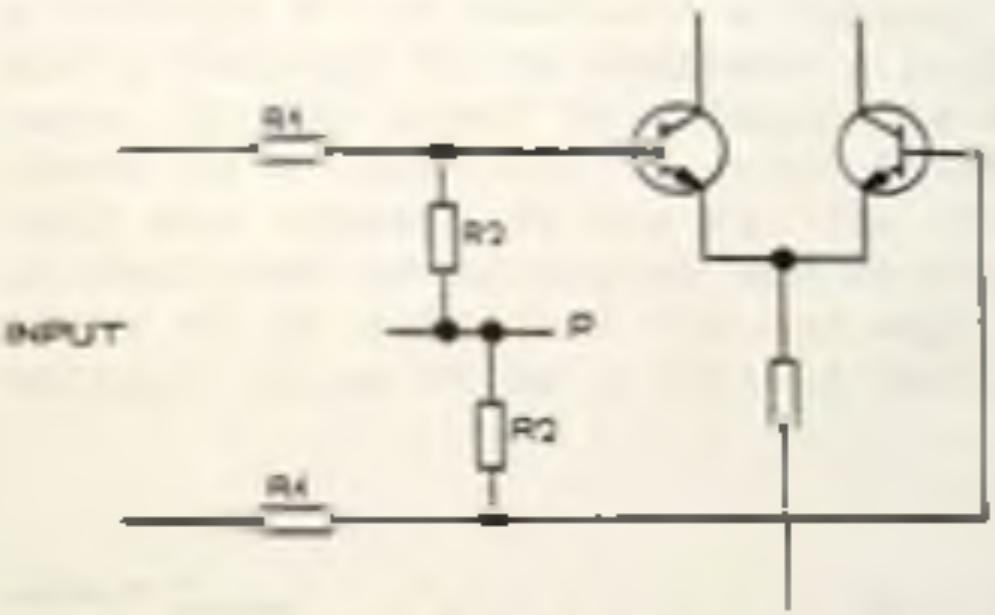
if n circuits have been commoned at P to keep down the number of capacitors used. If 32 circuits have been commoned and R1 + R2 = 50 ohms, $\frac{R1 + R2}{n}$ = 1.5 ohms.

So $\frac{1}{\omega C}$ should be $\leqslant$ 150 mohm at 100 KHz,

that is, less than one tenth of 1.5 ohms at the lowest frequency of the range of concern.

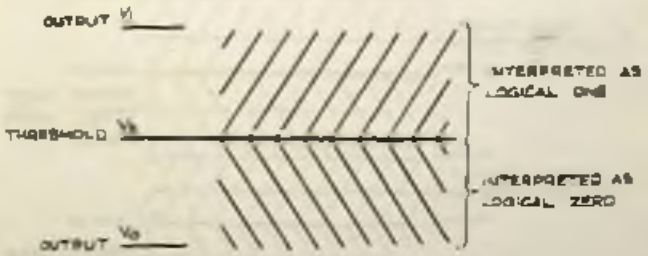
Therefore C = 10 μF approx.

The physical size of such a capacitor is acceptable in a normal system.



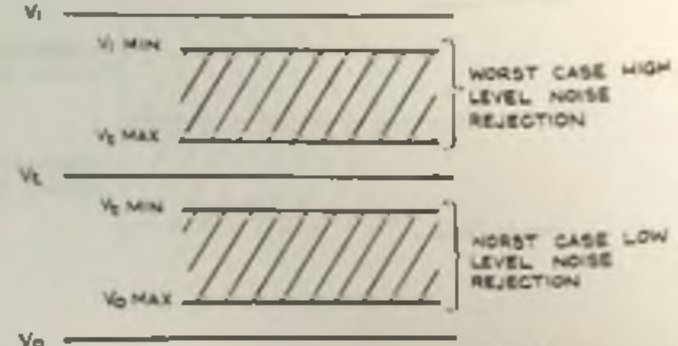
NOISE SPECIFICATIONS ON INTEGRATED CIRCUITS

A logic gate is designed to output a voltage V1 to indicate a logical One, and a voltage V0 to indicate a logical zero. At the input to a logic element there is a threshold Vt which is about half way between V0 and V1. The circuit is designed to interpret any voltage above Vt as a logical One and any voltage below Vt as a logical zero.



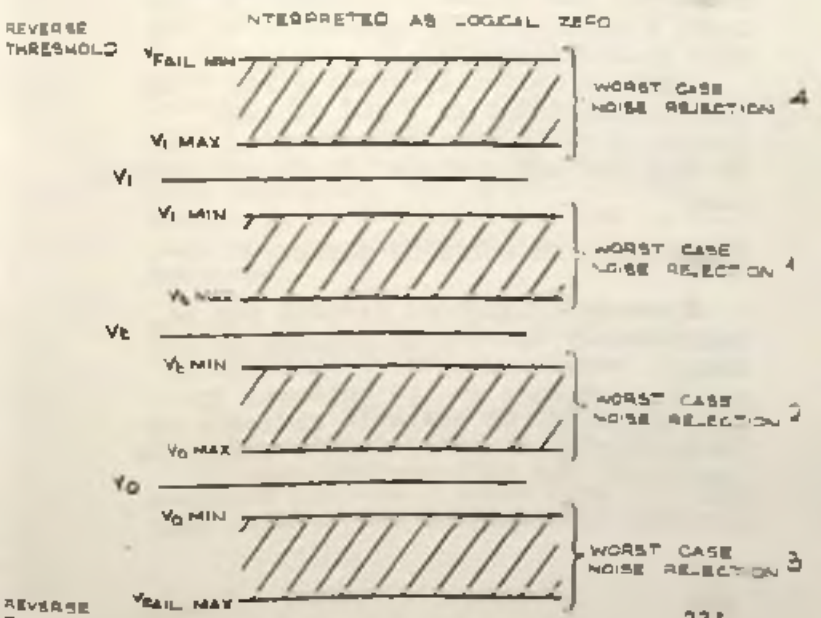
The noise rejection is commonly described as the difference between the minimum possible Vt and the maximum possible VO for circuits which are operating within specification for noise rejection at the low logic level, and the difference between the maximum possible Vt and the minimum possible Vl for noise rejection at the high logic level. Put mathematically, noise rejection is (Vtmin - VOmax) and (Vlmin - Vtmax).

The minimum (worst case) noise rejection should be specified for the case when the driving and driven logic elements are at opposite ends of the specified temperature range, and when the supply voltages are at the worst extreme within the specified range.



Even if these last provisos are kept in mind, such a specification, and therefore such a performance guarantee from the integrated circuit supplier, is quite inadequate in practical cases. This is because two other failure modes are common with present integrated circuit designs.

FAILURE DUE TO INPUTS OUTSIDE THE NOMINAL RANGE



It is common for integrated circuits to make a reversal of output when an input moves slightly above V1 or slightly below V0. A logic element will often interpret an input somewhat higher than V1 as a logical zero. It has a "reverse threshold" above V1. Alternatively, the circuit may malfunction in some other way if the input is outside the range V0 to V1. So we have to draw the more complex picture on the previous page. The worst case noise rejection for the logic elements is the narrowest of the four shaded areas.

Although obviously areas 3 and 4 are just as important as 1 and 2, such failure modes have always been ignored by supplier and customer up to the present time.

FAILURE DUE TO NOISE ON OUTPUT

A memory (bistable) element can be legitimately altered by changes at its inputs, but it can also be altered by spurious changes on its outputs. The susceptibility of bistable elements to interference on their outputs should be specified, but it never has been.

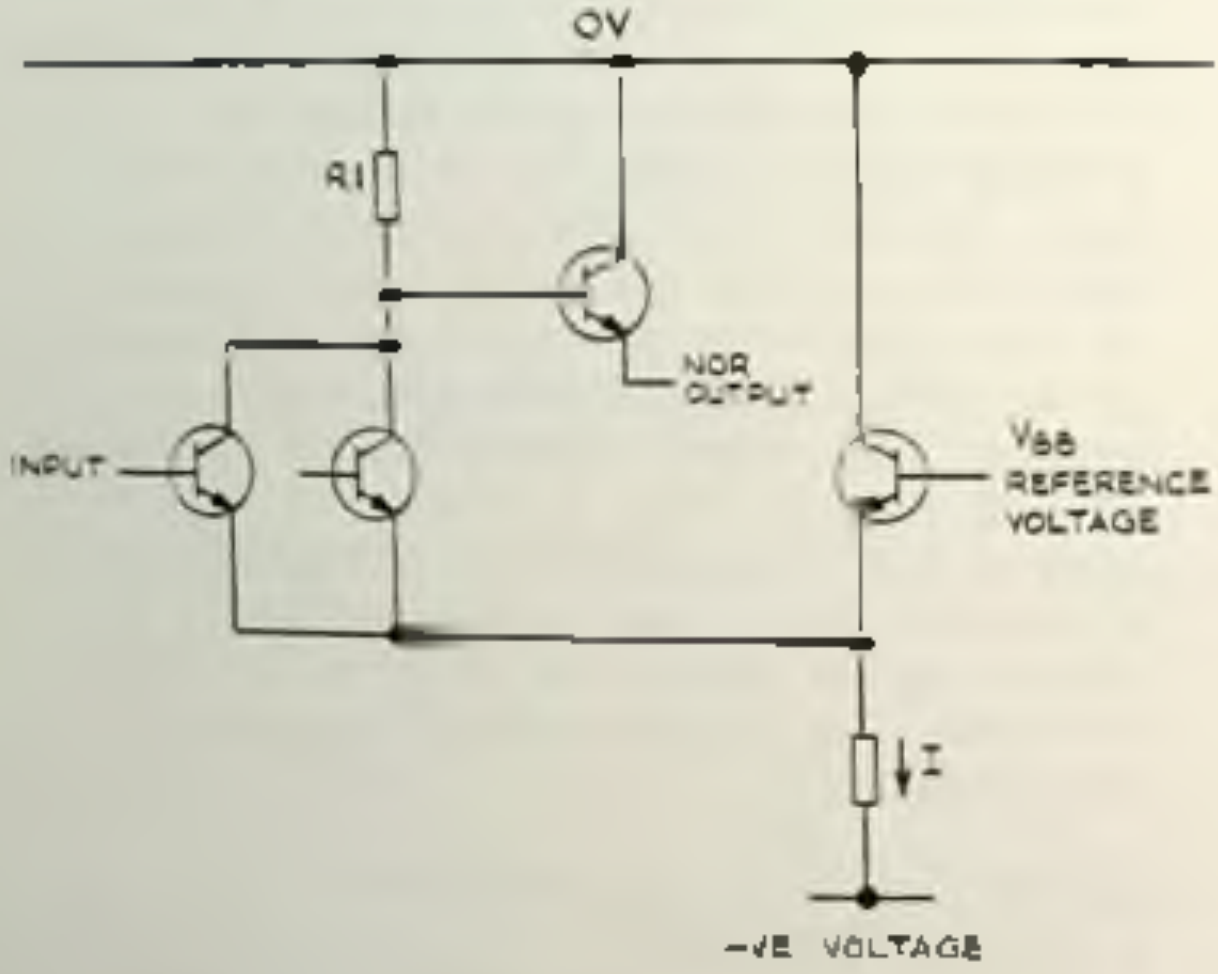
THE VALIDITY OF SPECIFYING NOISE REJECTION IN TERMS OF VOLTAGE

The foregoing was a discussion of current practices with regard to noise rejection specification in terms of voltages, and the shortcomings of the traditional methods. We now have to question whether the proper specification of noise rejection is in terms of voltage alone. Does noise appear as a pure increment of voltage on a line, irrespective of the source impedance of the line, its characteristic impedance, and the load? Obviously not, because the source impedance of the noise itself is not zero. However, if we ignore the considerations mentioned in this paragraph, but update current practices as described earlier in this section, the situation will be greatly improved.

EXAMPLE OF A LOGIC GATE WITH A REVERSE THRESHOLD

The ECL (Emitter Coupled Logic) gate has the circuit drawn on the next page.

If the input is low, the emitter resistor current I flows down the right hand side of the differential amplifier.


0V
R1
NOR
OUTPUT
INPUT
V_{BB}
REFERENCE
VOLTAGE
I
-VE VOLTAGE

No current except the small base drive current to the output transistor flows in R1, and the NOR output stands at the high level of one diode drop (Vbe) below 0V. So a low input causes a high output. As the input voltage rises above Vbb, the current I is deflected down the left hand path, and causes a drop across R. The result is that the NOR output drops to the correct low logic level.

So far so good. However, should noise raise the input even higher, above the logic 1 level, the input transistor finally saturates and its collector is driven high again via its base-collector diode, and the NOR output is also driven high again to a spurious logic 1 level. This is an example of reverse threshold.

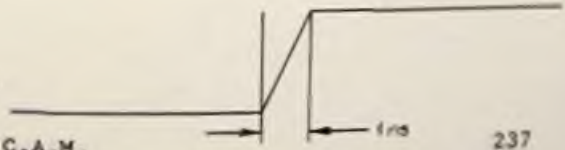
WARMING UP OF INTEGRATED CIRCUITS

The thermal time constant of integrated circuits is surprisingly high - as much as two minutes - and logic levels can vary by as much as 55 mv during this time. It is wise to allow the full two minutes of warm up time before logic is used, so that the noise margins are not significantly degraded.

COMPONENT PULSE RESPONSE

Until digital signals became widespread, particularly high speed ones, steady state sine waves were the normal state of affairs. It is well known that in VHF and Microwave regions, components cannot be thought of as lumped, and specific models must be developed. This attitude is correct, and must be applied when attempting to understand the performance of a component in a fast step situation. Firstly, we must have a clear understanding of what a step or pulse is.

The leading edge of a step is a shock wave. It is a TEM wavefront which travels at the speed of light. Now it is mathematically possible to take a single step and analyse it using Fourier Analysis.



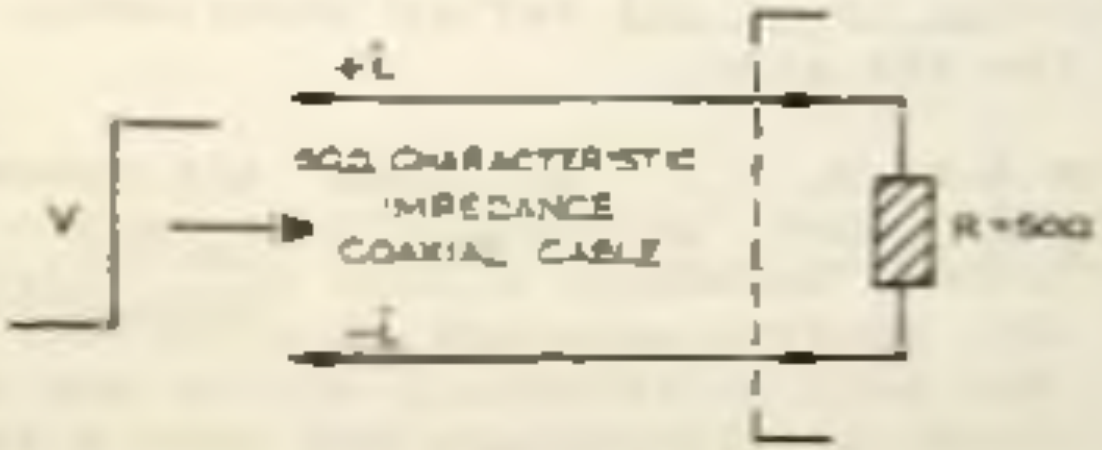
A single step has a risetime of 1 ns. To produce this from sine waves means combining an infinite number which have existed from minus infinity to plus infinity. Now this can be seen to be quite absurd and not practical at all. Instead, let us take the single event drawn on the previous page, a 1 ns. step produced from Schottky TTL or ECL. (Fairchild now do the F100K series with 700 ps. risetimes.)

It is important to realise that as this wavefront travels along a Transmission Line, everything in front of it has no knowledge of its existence whatsoever. There can be no instantaneous action at a distance. This is because the effects of any event take a finite time to propagate outwards from the source of the disturbance.

Any component, be it a resistor or a capacitor, for example, is distributed throughout space. It is therefore convenient to think of components as transmission lines, especially when considering their performance in a high speed digital environment.

THE RESISTOR

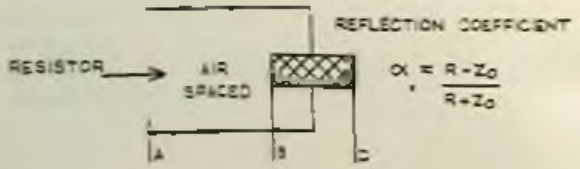
It is well known that terminating 50 ohm coaxial cable in a resistor of the same value is called 'perfect termination'. There is no reflection of the step. Therefore we can deduce that the signal cannot differentiate between a transmission line with a characteristic impedance of 50 ohms and a 50 ohm resistor.



The resistor can be seen to be a transmission line of two distinctly separate portions. The energy in the step is converted into heat and is diffused out into space. Of course, the purist would say that there is a small discontinuity, but in fact this is so short, about 12 mm, that we can treat the component as an ideal one. However, the engineer must always bear in mind that it is a very short transmission line before he makes the decision to neglect its effects. This is sometimes referred to as the series inductance of the resistor.

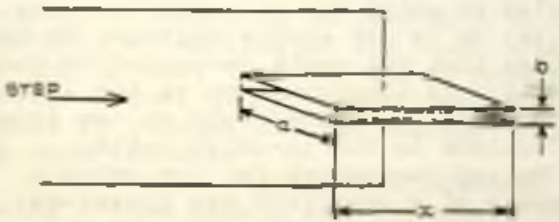
From A to B, Z_o1 = 180 ohms Air spaced

From B to C, Z_o2 = 50 ohms Carbon composite



THE CAPACITOR

Firstly, take the case of a parallel-plate capacitor.



A fast step propagates down a transmission line with a capacitor connected to the far end. The step arrives at the end of the transmission line and sees a very low characteristic impedance compared to its own Z_0 of 50 ohms, say. This is just like a short circuit, and so most of the signal is reflected and only a small portion of it enters the parallel plates of the capacitor. This small amplitude travels down the parallel plates and comes to an open circuit at the end. Voltage doubling

takes place due to the reflection coefficient of +1. At the other (left hand) end of the capacitor it sees a characteristic impedance much greater than its own (the capacitor having a Z_o of something like 0.1 ohms) and thus the voltage step is reflected once again. This process carries on until the capacitor becomes charged up to the supply voltage. We have assumed that the cable connected to the capacitor is long compared to the length of the capacitor, therefore we can ignore reflections in the co-axial cable.

The key parameters for the pulse response of a capacitor are therefore:

(1) Characteristic Impedance $Z_o = \frac{b}{a}\sqrt{\frac{\mu}{\epsilon}}$

(2) Time delay $T_d = \sqrt{\mu \epsilon}\, x$

This model of a capacitor can be applied equally well to circular plates (disc ceramics), but in that case the Z_o of the device decreases as the step travels out from the centre of the two plates.

Now, the established equation for the charging of a capacitor from a step input is

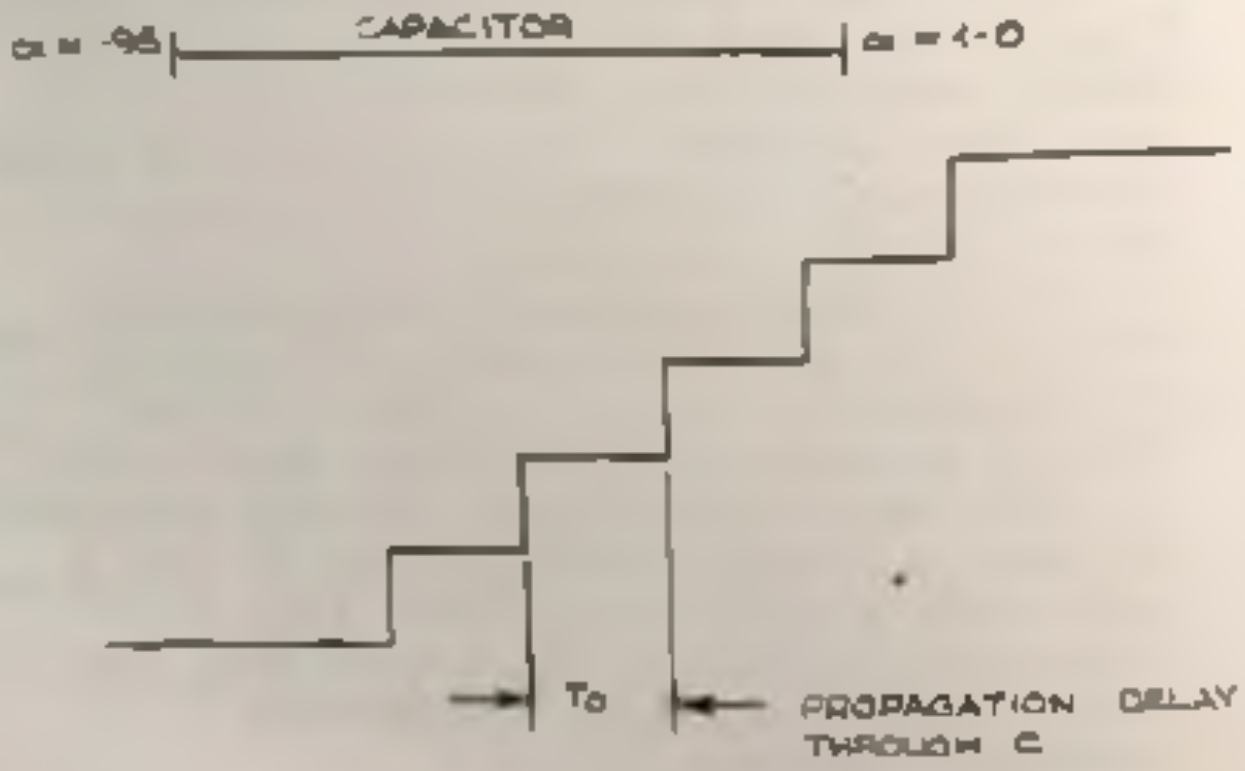
$$V = iR + \frac{1}{C}\int i\, dt$$

Looking at the second term we see that the constant C is the total capacitance of the component, and so it is not valid in a very fast step situation. This is because at the instant the step meets the capacitor the voltage developed across the plates initially cannot be defined by a value which is a function of the total size of the device. The step has no knowledge of the actual length of the plates until it reaches the far end and sees an open circuit.

Even faster edges can be produced using a time domain reflectometer (TDR). This produces a 25 psec. rise time edge which enables clear waveforms to be seen for standard components. In 50 psec, for example, a wavefront will travel about one cm in air.

The above argument shows that when components are being used in a digital environment it is of no value to use stated parameters which have been based on sine wave measurement and calculation. To have a clear understanding of how a component behaves in a fast step response situation, engineers must use the transmission line parameters developed above.

Even in a simusoidal situation this model can be used; when the wavelength of the impressed sine wave becomes comparable with the length of the leads and the plates a standing wave pattern will be present. It can be seen that the well known parameter equivalent series resistance (ESR) is in fact the characteristic impedance Z_o of the device. Also, there is no such thing as series inductance.



THE REED RELAY PULSE GENERATOR

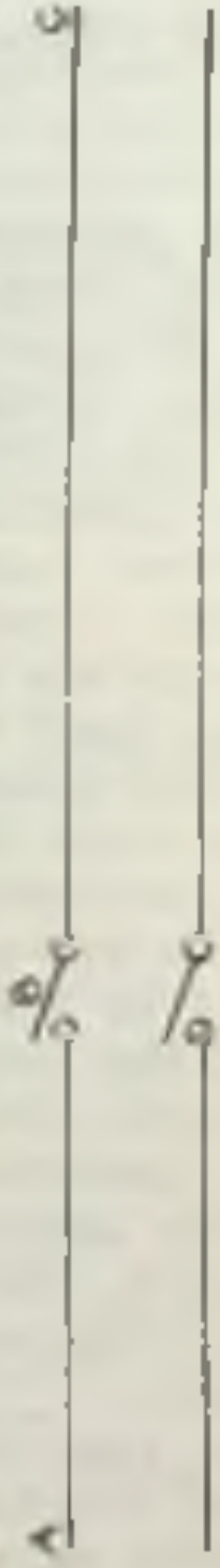
The reed relay pulse generator is described here as an aid to the understanding of the nature of power supply distribution in a digital system according to Theory H (page 119).

The reed relay pulse generator was a means of generating a fast pulse using rather primitive methods. A one metre section of 50 ohm coax AB was charged up to a steady 10 volts (say) via a one megohm resistor, and then suddenly discharged into a long piece of coax BC by the closure of two switches.

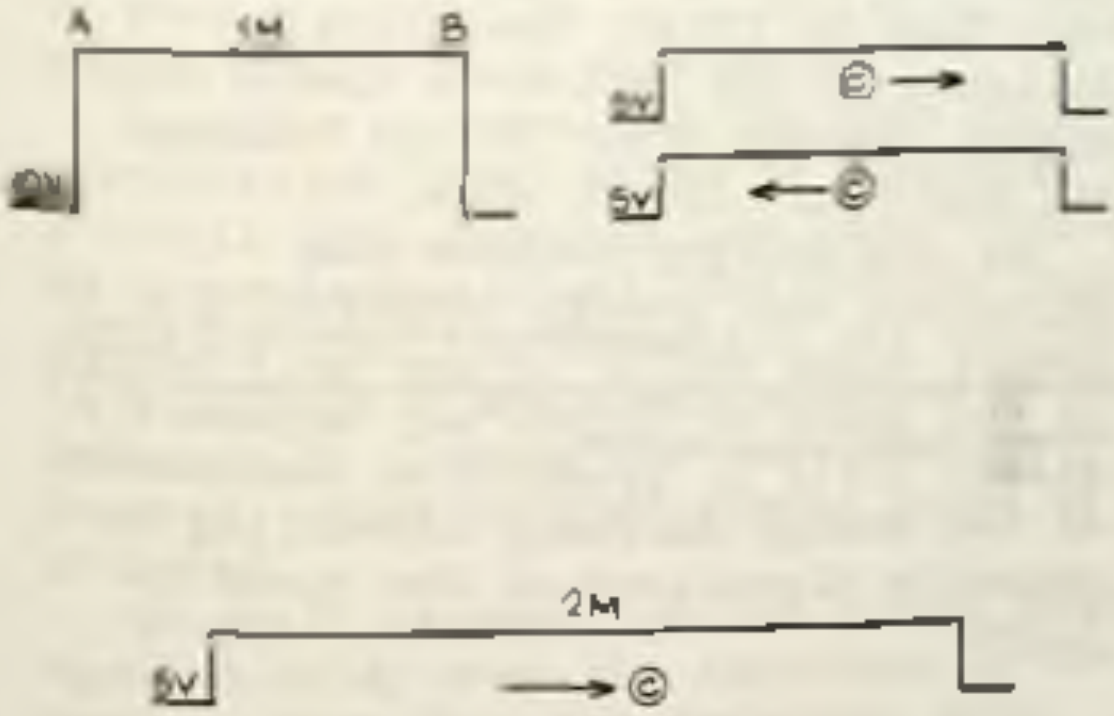
A five volt pulse two metres wide was found to travel off to the right at the speed of light for the dielectric on closure of the switches, leaving the section AB completely discharged.

(The practical device lacked the second, lower switch at B, which is added in the diagram on the next page to simplify the argument.)

The curious point is that the width of the pulse travelling off down BC is twice as much as the time delay for a



signal between A and B. Also, the voltage is half of what one would expect. It appears that after the switch was closed, some energy current must have started off to the left, away from the now closed switch; bounced off the open circuit at A, and then returned all the way back to the switch at B and beyond.



This paradox; that when the switches are closed, energy current promptly rushes away from the path suddenly made available; is understandable if one postulates that a steady charged capacitor AB is not steady at all; it contains energy current, half of it travelling to the right at the speed of light, and the other half travelling to the left at the speed of light.

Now it becomes obvious that when the switches are closed, the rightwards travelling energy current will exit down BC first, immediately followed by the leftwards travelling energy current after it has bounced off the open circuit at A.

We are driving towards the principle that Energy (Current) E x H cannot stand still; it can only travel at the speed of light. Any apparently steady field is probably a combination of two energy currents travelling in opposite directions at the speed of light.

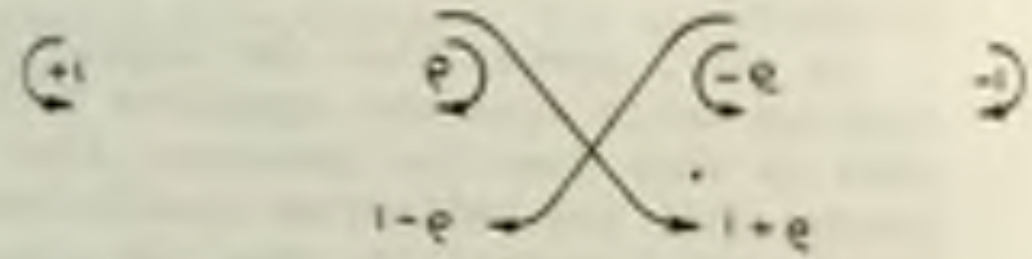
THE L - C OSCILLATOR CIRCUIT

If a charged capacitor is connected to a quiescent one-turn inductor, then energy current moves between the two components in a complex fashion. The resulting waveform is not a sine wave. Rather, it is a step waveform which approximates to a sine wave. We can see that the fundamental, idealized waveform in such a situation is a series of steps rather than a pure sine wave.

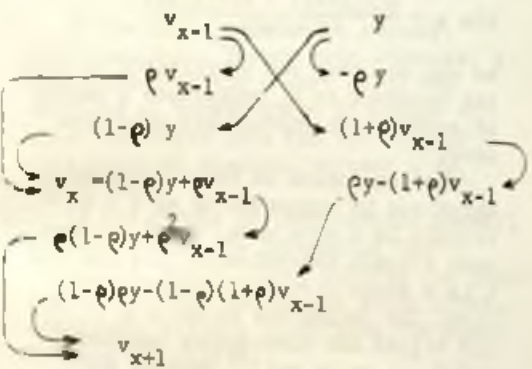
Consider a capacitor (or open circuit transmission line) which is connected to a single turn inductor (or short circuited transmission line). The initial state could be that the capacitor was charged to a voltage v and then connected to the inductor by closing the switches. However, other initial states could be proposed.

We shall assume that the signal propagation delay down the capacitor is the same as that down the inductor. Also assume that the reflection coefficient for a signal travelling down the capacitor to the right is ρ when it reaches the inductor.

In general, (see page 83), the following coefficients apply when signals reflect or pass through discontinuities in the circuit.



If at a certain time the signal in C has an amplitude v_{x-1} and in L it is y, then the following sequence occurs:



Therefore $v_x = (1-\rho)y + \rho v_{x-1}$

However,

$$v_{x+1} + v_{x-1} = v_{x-1}\left[1-(1-\rho^2)+\rho^2\right]+2\rho(1-\rho)y$$

which we can see equals $2\rho v_x$

$$\text{Therefore,} \quad v_x = \frac{v_{x+1} + v_{x-1}}{2\rho}$$

$2v_x$, $2v_{x+1}$, $2v_{x+2}$ etc. is a sequence

of amplitudes seen in the capacitor, and they obey the above formula.

Now since

$$\sin a = \frac{\sin(a+\delta a) + \sin(a-\delta a)}{2\cos\delta a}$$

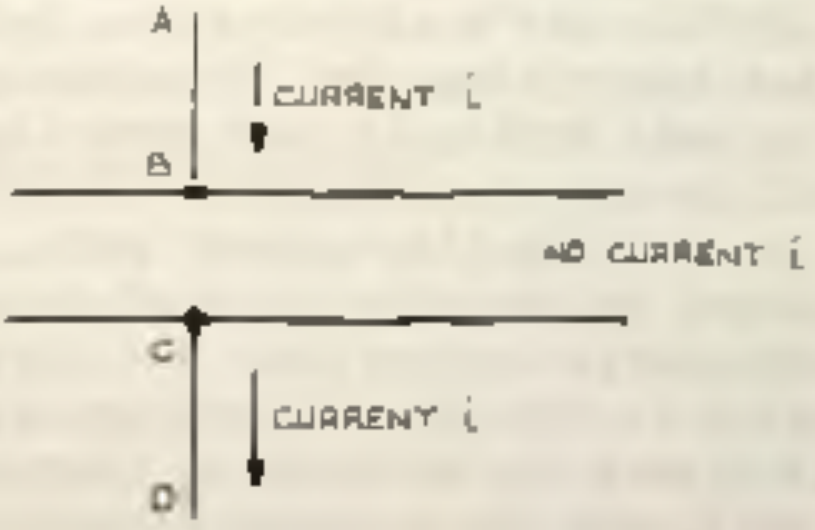
we can see that one possibility is that the sequence in v_x represents a series of steps which approximate to a sine wave.

The conclusion is that one waveform which can be supported by an L - C circuit is a sine wave, where C is an open circuit transmission line and the L is a short circuited transmission line. The larger the value of ρ (i.e. the bigger the discrepancy between Z_{oc} and Z_{oL}, or to put it another way, the more capacitive the capacitor and/or the more inductive the inductor,) the smaller is the forward flow of energy current each time across the central node between the C and the L. This means that there is more time between maxima in the energy current level in C, and a lower "resonant frequency".

THE HISTORY OF DISPLACEMENT CURRENT

In the early nineteenth century, Electromagnetic Theory made advances, a cornerstone of the theory being the doctrine of conservation of charge q, which developed into the doctrine of continuity of electric current flow, $dq/dt = i$.

In the middle of that century Maxwell struggled with the paradox of the capacitor, where charge entered one plate and then flowed out of the other plate, apparently without traversing the space between the plates.



This was an embarassment. It seemed that electric charge was being destroyed on the upper plate and being re-created when it reappeared on the lower plate. Maxwell "cut the Gordian knot" (as Oliver Heaviside put it in his Electromagnetic Theory, Volume 1, page 28 sect. 30) by postulating a new type of current, called "displacement current", as flowing across the gap BC so as to save the principle of continuity of electric current.

"Displacement current" was a result of his postulation of "electric displacement". Maxwell said that the total outward displacement across any closed surface is equal to the total charge inside the closed surface. (J C Maxwell, A TREATISE ON ELECTRICITY AND MAGNETISM, Oxford, Clarendon Press, page 253.)

This was a clever ruse, but it was not surprising that objections were raised. Notice in the next figure that if in any circuit there should be a break BC in the current path, we are bound by the principle of conservation of charge to say that the current i, that is the flow of charge, entering B from A accumulates as charge $\int i\,dt$ at B, and the current reappearing at C

"accumulates" an equal negative charge $-\int i\,dt$. By definition, electric displacement outward from B equals the total charge trapped at B; $D = \int i\,dt$ and $i = dD/dt$. It is not a coincidence that "displacement current" saves the idea of continuity of electric current; it does so by definition. With the postulation of displacement current, it would never in future be possible to devise an experiment which might refute the principle of continuity of electric current. Popper would therefore say that "displacement current" is an unscientific concept. (Karl Popper, CONJECTURES AND

REFUTATIONS, London, RKP, page 37.) Whenever charge seems to disappear at a point, displacement takes its place. Whenever electric current seems to disappear at a point, displacement current pops up to take its place.

It is important to remember that Maxwell and Heaviside believed that the current entering a capacitor plate became trapped and had nowhere to go. Writers on the subject must be glad that such a sneaky route between B and C for real current did not declare itself, since they say that the brilliant postulation of displacement current led to the postulation by Maxwell of waves in space. One text book writer, for instance, said, "Without [displacement current] there would be no electromagnetic radiation, and the greatest part of the remainder of this book would have to be omitted." (J D Jackson, Professor of Physics, University of California, Berkeley, in his CLASSICAL ELECTRODYNAMICS, pub. Wiley, page 216.)

Meanwhile, even as Maxwell was contemplating the ethereal displacement current, practical engineers were inventing and building wired telegraph

systems. The distortion of signals travelling long distances was bad, and was thought to be due to the fact that the capacitance of the telegraph wired had to be charged up through the resistance of the wires, resulting in an RC time constant which attenuated different frequencies differently. As late as 1910 virtually all electricians (including to some extent Lord Kelvin) did not accept Oliver Heaviside's claim that a telegraph wire had distributed inductance as well as capacitance, and that if only this inductance were increased by the addition of periodic loading coils, distortion-free transmission over long distances could be achieved.

It was important for Heaviside to encourage a sensible approach to the characteristic impedance of telegraph lines, because the practical payoff in telegraphy and telephony would be immense. (This misunderstanding delayed the general introduction of telephones for twenty years.) This practical payoff would be best achieved by arguing that signals travelling down (between) telegraph lines were undistorted TEM and similar to the waves in space discovered by Hertz in 1887, twenty years

before, and previously postulated by Maxwell as one implication of his proposed displacement current.

It was important for Heaviside not to rock the boat he was trying to ride, the boat of Maxwell's E-M theory. So it would have been injudicious for Heaviside to question the concept of displacement current, and he never did.

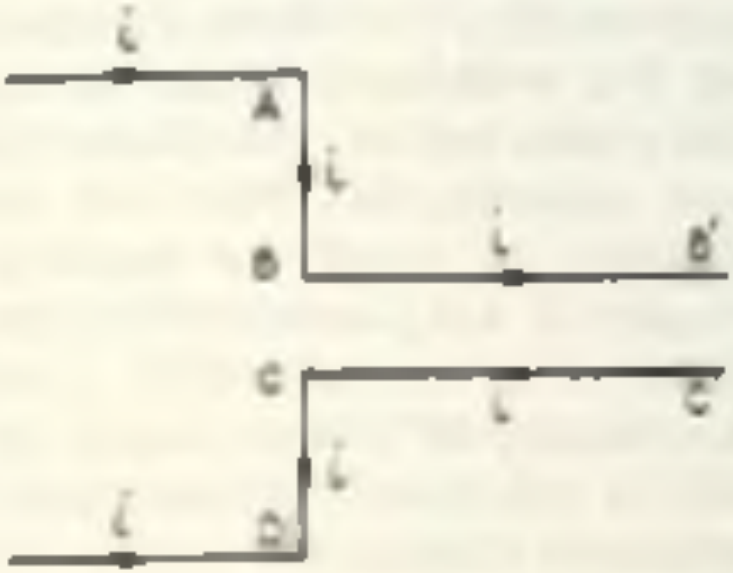
The essence of the concept of a transverse electromagnetic wave, TEM, is that nothing - field, flux or current, flows laterally across the surface of the wave front. The analogy is the Severn Bore, where we see a single step of water rushing up the Severn. Everything ahead of the step is steady, and everything behind the step is steady. There is no lateral, sideways flow. In the electromagnetic case, the idea of a lateral flow of current across the face of a TEM step is absurd, and would result in a longitudinal magnetic field; the step would "get ahead of itself". Further, since the step travels forward at the speed of light, $1/\sqrt{\mu\epsilon}$, any lateral flow would cause embarassment by travelling even faster, in the same way that when you walk across inside a

moving train, by Pythagoras' Theorem, you are travelling faster than the train.

Now although in the case of a capacitor, displacement current needed to be regarded as just like a real current, for instance causing a magnetic field; in the case of the D flux at the front of a step of TEM (E x H) energy current travelling down a telegraph line, the displacement clearly must not behave like a real current - for instance by creating a magnetic field which would reach out ahead of the wave front and ruin its TEM nature.

Maxwell and later Heaviside did not notice the discrepancy in the requirements of displacement current; that in a capacitor it must act like a real current but in a transmission line it must not; because neither of them knew that a capacitor is no more nor less than a transmission line. This is even today known by very few scientists. Maxwell, along with today's text book writers (e.g. J H Fewkes and Yarwood, ELECTRICITY AND MAGNETISM VOL 1, London; University Tutorial Press 1956, page 505; also B.I. Bleaney and Bleaney, ELECTRICITY AND MAGNETISM 2nd Edn, Oxford; Clarendon 1965, page 258)

believed that the displacement current dD/dt travelling across between the plates of a capacitor BC was uniformly distributed, and it is only very recently that it has been pointed out that the flow of current and field in a capacitor is identical with that in a transmission line; that the field moves out from the capacitor's leads as if they were links to one end of a transmission line. So the discrepancy could not become apparent.



A serious difficulty for Displacement Current arises when we realize that the two plates BB', CC' are a transmission line. We know that the current i travelling down to B from A then flows out sideways from B along the capacitor plate BB'. This route, along the capacitor plates, is the sneaky "back door" route that Maxwell failed to notice, and everyone has followed his lead.

In a transmission line, everyone agrees that the current I entering the line at B leaves B by flowing along the line BB'. No displacement current dD/dt between the lines is needed for us to retain the doctrine of conservation of charge and conservation of current. In fact, if this dD/dt were regarded as

current, far from saving the doctrine, it would destroy it, because now more current (i + dD/dt) would be leaving the first section of the plate BB' than was entering it. The last sentence is difficult to grasp; no matter, since it is easy to see, and sufficient to see, that if i enters B from A and i leaves B along BB', continuity of current is preserved without our having to postulate displacement current.

> "But surely we cannot just drop displacement current when for a century every guru has been protesting that it is the foundation of our craft? (e.g. L Solymar, LECTURES ON ELECTROMAGNETIC THEORY, OUP, page 6; also J D Jackson, CLASSICAL ELECTRODYNAMICS, Wiley, page 218;). Every expert says that 'Maxwell's leap of genius' in proposing displacement current was what got the subject going - leading to Hertz's discovery of waves in space, for instance."

The answer lies hidden in Heaviside's magnificent, regal statement, "We reverse this." We quote the full paragraph on page 120.

The discrediting of Displacement

Current merely makes Heaviside's "We reverse this" mandatory. It means that the field must be the cause and electric current an effect, rather than (as Maxwell thought) the other way round.

If we keep to "Theory H" (page 121), the theory that the field E x H, travelling along between the wires at the speed of light - what Heaviside called the "energy current", is the cause, then electric charge and electric current are merely what define the edge of an energy current. If electric current is that which defines the side of an energy current, then we may with equal justification postulate "Displacement Current" as that which defines the front face of a step of energy current. Under "Theory H", Maxwell's 'leap of genius' (when he postulated displacement current and thence waves in space) becomes tautological; "Because a wave in space if it existed would have to have a front face (displacement current), then I propose such a front face and therefore I propose waves in space."

Maxwell would have saved us from a century of confusion if he had had enough insight to say, "Since circuits containing capacitors, that is, open

circuits, work, It follows that the essence of electromagnetics cannot be electric current in closed circuits of conductors; It must be something else. What about waves in space?" Heaviside, seventy years later, missed the key point by a whisker. He failed, but but he failed gloriously. He never discovered the bad apple in the barrel, Displacement Current.

OSCILLOSCOPES FOR DIGITAL MEASUREMENT

The oscilloscope is still the most commonly used signal analysis instrument available to the digital elgineer. It is therefore relevant to consider its use carefully so that the maximum amount of information can be obtained from the display. The applications of the oscilloscope for digital signal analysis fall into two main categories:
(i) Checking the timing relationships of digital signals to verify the correct operation of the logic system.
(ii) Measurement of noise levels, checking clock edges, risetime measurements etc.

For application (i), an oscilloscope with bandwidth in the range 50 - 100 MHz (risetime 8 - 4 nsec.) will usually suffice, whereas applications in the second group usually call for a higher bandwidth and may require the use of a sampling oscilloscope (typical bandwidth 1 GHz; 350 psec risetime).

The engineer can usually assume that the oscilloscope itself is giving a reasonably accurate picture of the signal

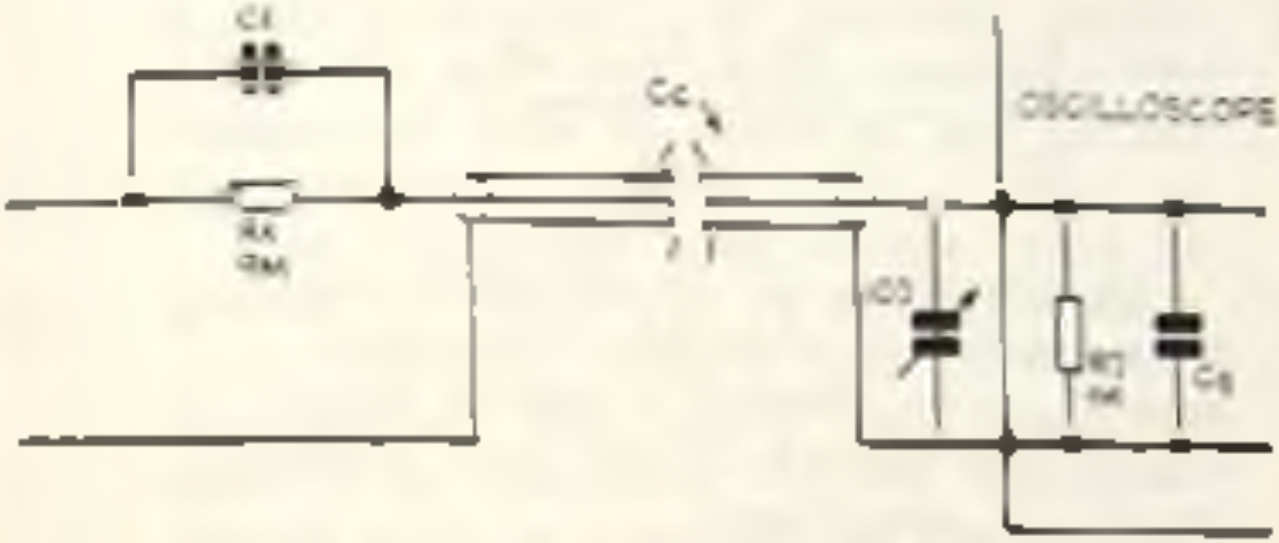
(subject to risetime limitations) which is presented to its input. In the authors' experience, the chief problem is associated with the means for conveying the signal from the circuit under test to the input of the oscilloscope. This brings us to the subject of probes.

HIGH IMPEDANCE PROBES

By far the most common type of oscilloscope probe is the high impedance 10 : 1 probe. This is designed to feed into a standard 1 Mohm oscilloscope input; the circuit is shown in the figure on the next page.

The basic principle of operation can be understood by noting that resistors R1 and R2 provide a 10 : 1 attenuator for signals at low frequencies where the effects of the various capacitances are negligible. At high frequencies, C1 together with Cc (cable capacitance), C2 (trimmer capacitor) and Cs (oscilloscope input capacitance) form a capacitive divider. Since capacitive reactance at frequency ω is given by $Z_c = 1/\omega C$ we have,

$$\frac{R1}{R2} = \frac{Z_{C1}}{Z_{(Cc + C2 + Cs)}}$$


Ci
Cc
OSCILLOSCOPE
Ri
Rm
Co
Ri
Rm
Ci

i.e. $\frac{R1}{R2} = \frac{Cc + C2 + Cs}{C1}$

In practice, C2 is adjusted until this condition obtains. The normal method is to observe a square wave signal and adjust C2 for zero undershoot and overshoot.

It will be noted that the theory of operation rests on the assumption that the cable capacitance can be treated as a lumped component. Once the transition t me of a step along the cable connecting signal to oscilloscope becomes longer than the risetime of the step, this assumption breaks down and distortions of the pulse edge will be fed to the oscilloscope.

Various 'dodges' such as the replacement of the cable core by resistive wire have been employed although, obviously, they cannot be entirely successful since the whole theory of operation of the device is based on a fiction.

In addition to the above mentioned drawback the 10 : 1 high impedance probe has an even more serious deficiency; it is extremely prone to

pickup by virtue(?) of its high impedance. The user can easily convince himself of this by means of the following simple test.

Use the ground lead of the probe to ground the probe tip so that the whole probe assembly forms a loop of 4" to 6" diameter, then place the 'loop' near to an operating digital board. Spikes will be observed on the oscilloscope. It transpires that any signal which is now injected between the probe tip and ground lead will be in series with this direct pickup; thus the display on the screen will be a composite of the signal under study plus spurious noise. It should be clear by now that this makes it impossible to tell which of the features of the displayed signal are in fact genuine, and that the usefulness of the high impedance probe to the logic engineer should be seriously questioned.

The high impedance probe originated in the days of valve technology when the high impedances of the circuitry used (up to Megohms) called for very high impedance measuring devices. In high speed logic, impedance levels are drastically lower, of necessity,

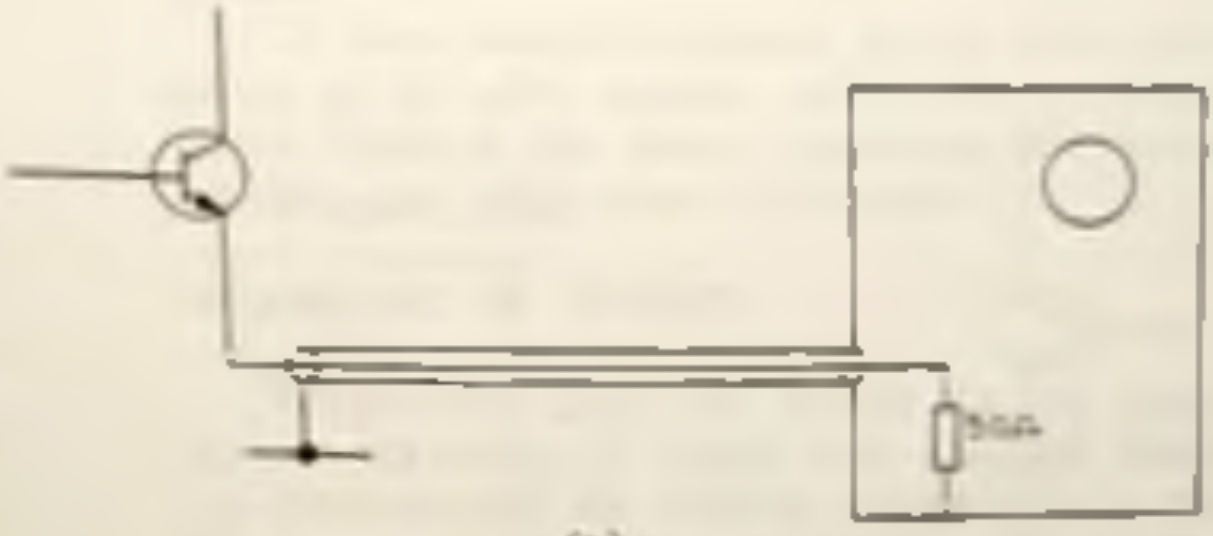
and hence the high impedance probe has ceased to be required. It should be allowed to slumber in peace, its raison d'etre having been removed.

LOW IMPEDANCE PROBES

The correct probe to be used for high speed digital signals is the low impedance probe which will be described in this section. It makes use of the following simple principle. Any resistor, R, in a circuit may be replaced by a transmission line of any length but of impedance $Z_0 = R$ terminated in resistance R. This will cause no perturbation in the circuit, and will allow the signal which would have appeared across R to be observed remotely. The only penalty to be paid is the time delay necessary for the signal to propagate along the length of transmission line. In practice this is no disadvantage; it simply means that when a number of points are being observed, equal lengths of transmission line should be employed. The diagrams on the next page explain the principle.

The emitter load resistor in the first circuit has been replaced by a 50 ohm co-ax terminated in 50 ohms and the signal

(a)

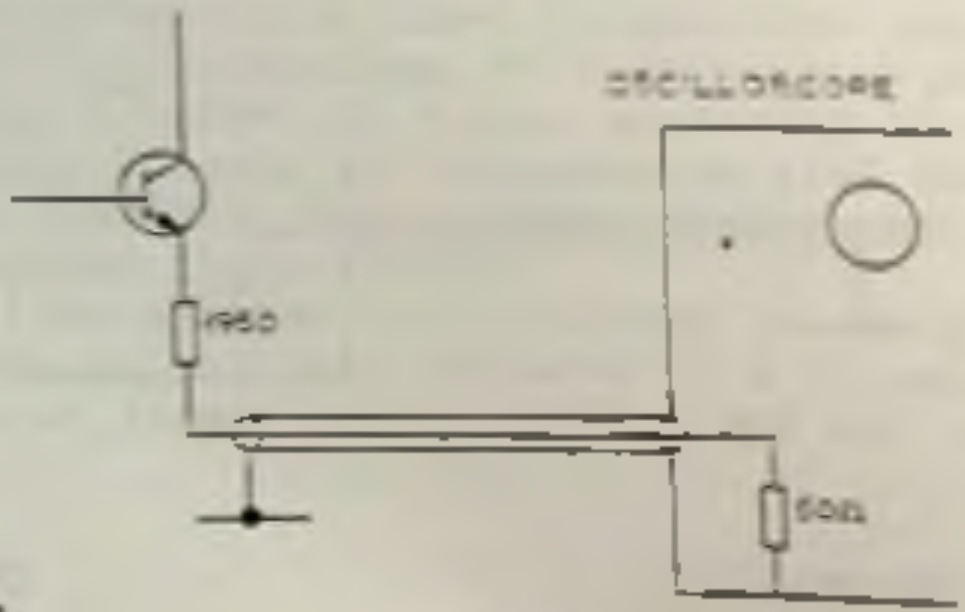


(b)

across the 50 ohm resistor is displayed at the oscilloscope.

This technique is easily extended to values of resistor higher than 50 ohms as shown in the next diagram, where the correct 2 Kohm resistor is effectively replaced by 1,950 ohms in series with 50 ohms. The signal displayed by the oscilloscope is 1/40 th of that which would appear across the 2 Kohm resistor.

It is possible to use this technique to provide built in probe points which lead to co-axial sockets which normally carry 50 ohm terminators. In order to observe any of these signals, all that is necessary is to remove the terminator and couple to the 50 ohm input of the oscilloscope with 50 ohm co-ax.



So far, we have not discussed a probe as such, but rather oscilloscope connection techniques. It is simple, however, to build a probe based on the principles just described, see diagram on the next page. In order to minimise loading of the circuit under observation, R1 should be made as high as possible considering the noise level of the oscilloscope in use.

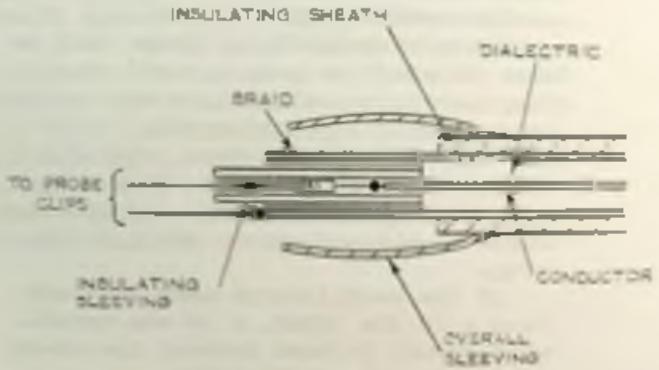
The wire ends of the probe must be kept as short as possible and ideally they should be soldered to the circuit under test. In practice this is not convenient, but miniature prodclips (azi-hook) can be fitted. These are only about ¼ inches long, and the wire leads should not exceed about 1 inch long.

If the oscilloscope used does not have a 50 ohm input, a 50 ohm terminator should be used between the co-ax cable and the oscilloscope.

GROUNDING OF PROBES

Whatever type of probe is in use, it is essential that its ground lead be connected as close as possible to the circuit point under observation. Where more than one probe is in use,

this rule should be observed for each probe.

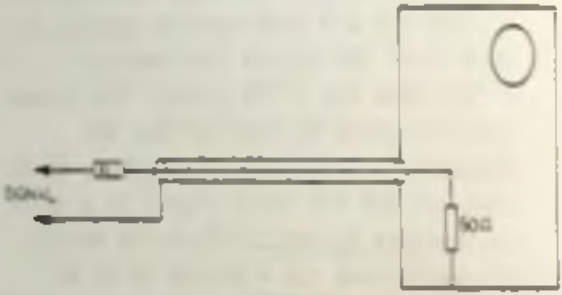


LOOKING AT VERY FAST DIGITAL SIGNALS

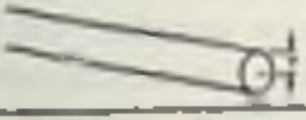
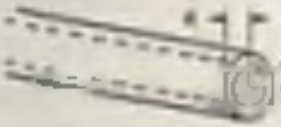
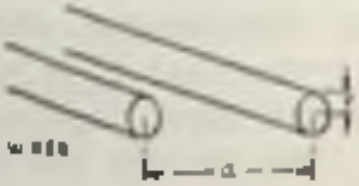
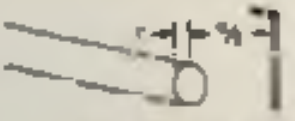
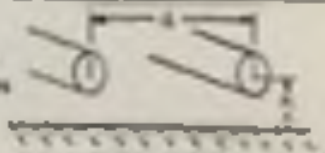
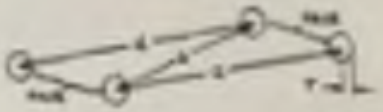
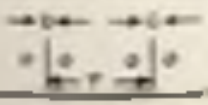
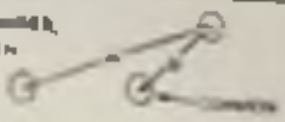
When looking at fast (1 nsec.) ECL signals, the available passive probes, although they have a satisfactory response, have either too high attenuation, (100X for a 5 Kohm passive probe), or else load the signal too heavily (500 ohms for a 10X probe). The proper passive probe to look at fast ECL signals is a 25X 1250 ohms probe, which reduces the 800 mvolt signal to a still acceptable 32 mvolts which can easily be seen above the 4 mvolts or so or noise in a typical sampling oscilloscope.

Unfortunately, no manufacturer seems willing to make a 25X passive probe. However, it is quite easy to make one by soldering a 1,200 ohm resistor onto the end of the inner conductor of a length of solid 50 ohm coax cable. For best response, a miniature 1,200 ohm resistor should be buried in the end of the coax dielectric. A better response than that of Tektronix passive probes can easily be achieved.

Traditionally, a probe has been regarded as presenting a resistive load plus a small parallel capacitance of a few picofarads. However, the home made


SIGNAL
50Ω

probe drawn opposite will look like a purely resistive load of 1,250 ohms. For Schottky TTL one can use a 10X probe containing a series resistor of 450 ohms with the coaxial cable perfectly terminated in its characteristic impedance of 50 ohms. However, as the voltage level of TTL is relatively large, a 25X probe would be more than adequate. This type of probe is by far the best when looking at, for example, noise on power rails. Conventional high impedance probes (10 megohms input impedance) tend to behave like aerials and are not suitable.

	CAPACITANCE
LONG STRAIGHT WIRE 	
CONCENTRIC CONDUCTORS 	$C = \frac{2\pi\varepsilon_r\varepsilon_0}{\ln \frac{b}{a}}$ FARAD/METRE
PARALLEL WIRES 	$C = \frac{\pi\varepsilon_r\varepsilon_0}{\ln \frac{b}{a}}$ F/m
WIRE ABOVE GROUND PLANE 	TWICE VALUE FOR PARALLEL WIRES
PARALLEL PLATES 	$C = \frac{\varepsilon_r\varepsilon_0 A}{D}$ F/m
PAIR OF WIRES GROUND RETURN 	MUTUAL CAPACITANCE ($= \frac{1}{2}$ COEFFICIENT OF INDUCTION) $C = \frac{\pi\varepsilon}{\ln \frac{2h}{a}}$
	
TWO SIGNAL WIRES COMMON RETURN 	

INDUCTANCE	CHARACTERISTIC IMPEDANCE
EXTERNAL L = [illegible] / [illegible] INTERNAL L = [illegible] / [illegible]	
HI FREQUENCY L = [illegible] / [illegible] ln [illegible] h/m LO FREQUENCY L = [illegible] [illegible] h/m	$Z_0 = \frac{1}{2\pi}\sqrt{\frac{\mu}{\epsilon}} \ln \frac{b}{a}$ $= \sqrt{\frac{L}{C}}$
SELF INDUCTANCE HI FREQUENCY $2L = \frac{\mu}{\pi} \log \frac{b}{a}$ H/m $= 0.4 \ln \frac{b}{a}$ μH/m LO FREQUENCY $2L = \frac{\mu}{2\pi}\left[1 + 4 \ln \frac{b}{a}\right]$	$Z_0 = \frac{1}{\pi}\sqrt{\frac{\mu}{\epsilon}} \log \frac{b}{a}$ $= \sqrt{\frac{L}{C}}$
HALF VALUE FOR PARALLEL WIRES	HALF VALUE FOR PARALLEL WIRES
SELF INDUCTANCE $2L = \mu \frac{b}{a}$ H/m	$Z_0 = \sqrt{\frac{\mu}{\epsilon}} \frac{b}{a}$ $= 120\pi \sqrt{\frac{\mu_r}{\epsilon_r}} \frac{b}{a}$ Ω
MUTUAL INDUCTANCE $M = \frac{\mu}{2\pi} \ln \frac{\sqrt{(a+c)^2 + d^2}}{c}$ $= \frac{\mu}{2\pi} \ln \frac{\sqrt{a^2 + d^2}}{a}$	$\mu_0 = 4\pi \times 10^{-7}$ $\epsilon_0 = \frac{1}{36\pi} \times 10^{-9} = 8.85 \times 10^{-12}$ $c = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\mu\epsilon}}$
$M = \frac{\mu}{2\pi} \log \frac{AB}{BC}$ FOR CLOSE GRIDS $M = \frac{\mu}{2\pi} \ln \frac{BC}{AC}$	$\sqrt{\frac{\mu_0}{\epsilon_0}} = 1.2 \times 10^{2}$ Ω/sec
$M = \frac{\mu}{2\pi} \ln \frac{b^2}{ad}$	C. A. M. CONSTANTS

GATING OF ASYNCHRONOUS LOGIC SIGNALS

"The Glitch" is one of the key problems which face a designer of digital systems. Unfortunately, it is completely ignored in the literature, both in books and journals, with the result that engineers by the thousand fall into the trap, designing and building systems which fail.

There is a fundamental problem, or barrier, involved when one module tries to communicate with another independent module. In this context, "independent" means that the second module is not synchronised with the first, and also that it is doing other things at the moment when the communication is attempted, and not merely waiting for the commmication. (If it were waiting, it would be a slave module rather than an independent module).

The best way to explain the problem is by way of analogy. Consider a woman sitting in the living room of a house having two doors, front and back, and likewise two doorbells. She has two pet doorbell rules which are simply

these: 1) She doesn't answer a doorbell that hasn't been rung. 2) She answers the first doorbell she hears. The solution works as long as the milkman and the postman don't come at the same time. Having foreseen this event might occur one day, she made the arbitrary decision that should they both ring at the same time, she would give priority to the front door. Content in the knowledge that she has solved her problem, she waits in the living room for the doorbell to ring. One formidable day, the postman and the milkman both decide to deliver, at exactly the same time. However, as fate may have it, the milkman's finger reaches the back doorbell just a millisecond or two before the postman's finger reaches the front doorbell. Now the poor lady of the house is fraught with indecision. It appears that two milliseconds is on the threshold of her ability to distinguish between simultaneous events and non - simultaneous events. She can not decide whether the back door rang first, or whether they both rang at the same time.

The postman and the milkman, both

being very busy men (i.e. having more than one delivery to make), cannot wait an arbitrarily long time for this lady to make up her mind. They are somewhat indifferent as to which she answers first, just as long as she makes up her mind. But there she sits, steadfastly adhering to her rule, and unable to resolve which of the two events occured; simultaneously, or non-simultaneously. Hence, this pathological event causes her system to "hang up".

Being a resourceful individual, she devises a means of making an arbitrary decision in such a circumstance. Her new rule was, when in doubt flip a coin, and abide by the outcome of the coin toss. However clever this may have seemed to her, in so doing she has inadvertently complicated, and not cured, the essential difficulty, for now she has to decide when she is in doubt and when she isn't (i.e. when to use the coin and when not to). Thus, on that pathological Monday, we find the postman and the milkman, each at his respective door with the poor frustrated lady trying to decide whether or not to use the coin.

The poor lady keeps trying to change the rules, hoping that there will be some magic formula that will not hang her up in decision, but each time her efforts are frustrated. The point of the story is that regardless of the set of rules that are made to resolve all possible cases of two independent events into one of two groups, there is always a pathological case that will "hang up" your system for an arbitrarily long time.

As another example, let us consider a car driver approaching some green traffic lights. If they change to red when he is very close to them, he will ignore them and continue on his way, if on the other hand they change to red while he is still far enough away to stop easily, he will stop. In these two cases there's no problem. However, if the lights change when he is rather near yet not very near, the driver has to make a difficult decision as to whether to stop or to continue.

The human sets a threshold; perhaps thirty yards ahead of the lights. If the lights change before he reaches this point he will stop. If they change after he has passed this point,

he will carry on. But supposing the lights change when he is only one yard past the threshold when the lights change. After a little consideration he will continue without stopping. However, imagine he is only six inches past the threshold when the lights change. In this case he will take more time to consider the matter in detail, making up his mind that he was over the threshold when the lights changed, carrying on accordingly. If he were only one inch past the threshold when the lights changed, he would be unable to decide whether to stop or not, for he would not be able to judge such a short distance.

We have arrived at a dilemma. The driver has a set threshold, a point thirty yards ahead of the lights, and has decided that he will stop if, and only if, the lights change before he reaches this point. The dilemma arises on the rare occasions when the lights change at a moment when he is so close to the threshold that he has difficulty deciding whether he has passed it or not

In an attempt to evade the problem the driver devises a third category.

Under his new system, he categorises the events as follows:

1) Lights change when he is a long distance from them.

2) Lights change when he is close to the thirty yard threshold.

3) Lights change when he is a short distance from them.

He will decide to stop in all cases which fall into category (1) and continue in all cases which fall into category (3). This leaves the difficult cases, those which fall into category (2).

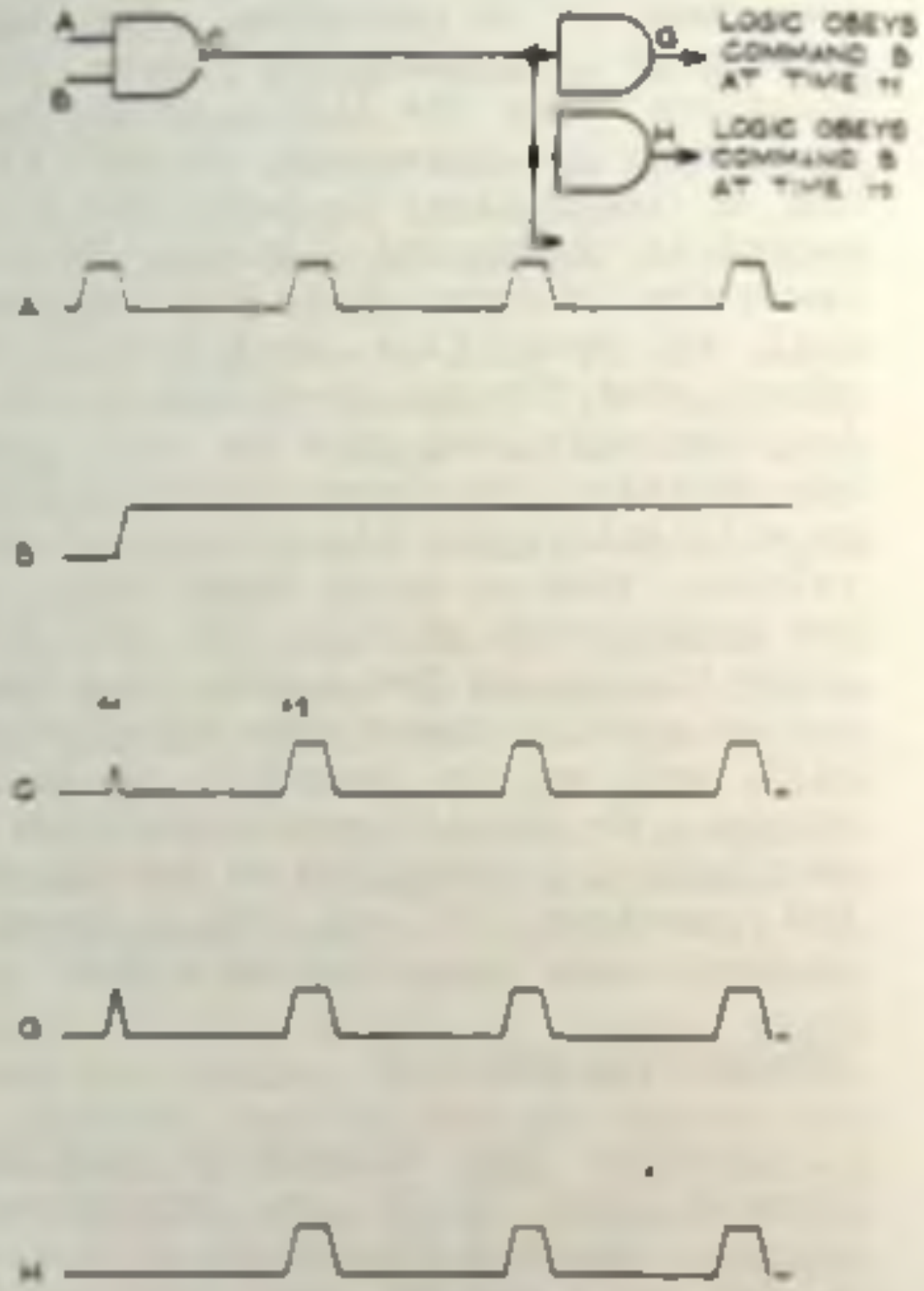
Let us suppose that he decides to stop for all cases in category (2). We see that a crucial decision now arises in the case of events which are close to the borderline between categories (2) and (3), perhaps at a distance of twenty yards from the lights. All he has done is to move the original threshold from thirty to twenty yards from the lights. The problem remains.

Another way to try to resolve the situation is to say that the driver will endeavour to judge whether or not he is past the threshold when the lights change, however small the distance may be.

This is infact the nearest he will ever come to a solution. The only difficulty is that on those rare occasions when the lights change when he is near the threshold it will take him a long time to judge the split second in timing and distance. He will leave the information in his brain and mull it over (i.e. amplify it) and after what for the co-driver can be a frighteningly long time he will make his decision. On these rare occasions, he will half apply his breaks and then release them or apply them fully. In the worse cases of all, he will half apply his breaks for a very long time, and on some of these rare occasions he will end up in hospital or in the morgue. We dismiss such casualties to statistical probability as bad drivers and continue in our own righteous, statistically more fortunate way.

THEORETICAL ANALYSIS

Consider the diagram on page 288. Suppose that B is an asynchronous signal, entering some logic with clock A, and it is brought into synchronization by ANDing (A.B). There is a sta-

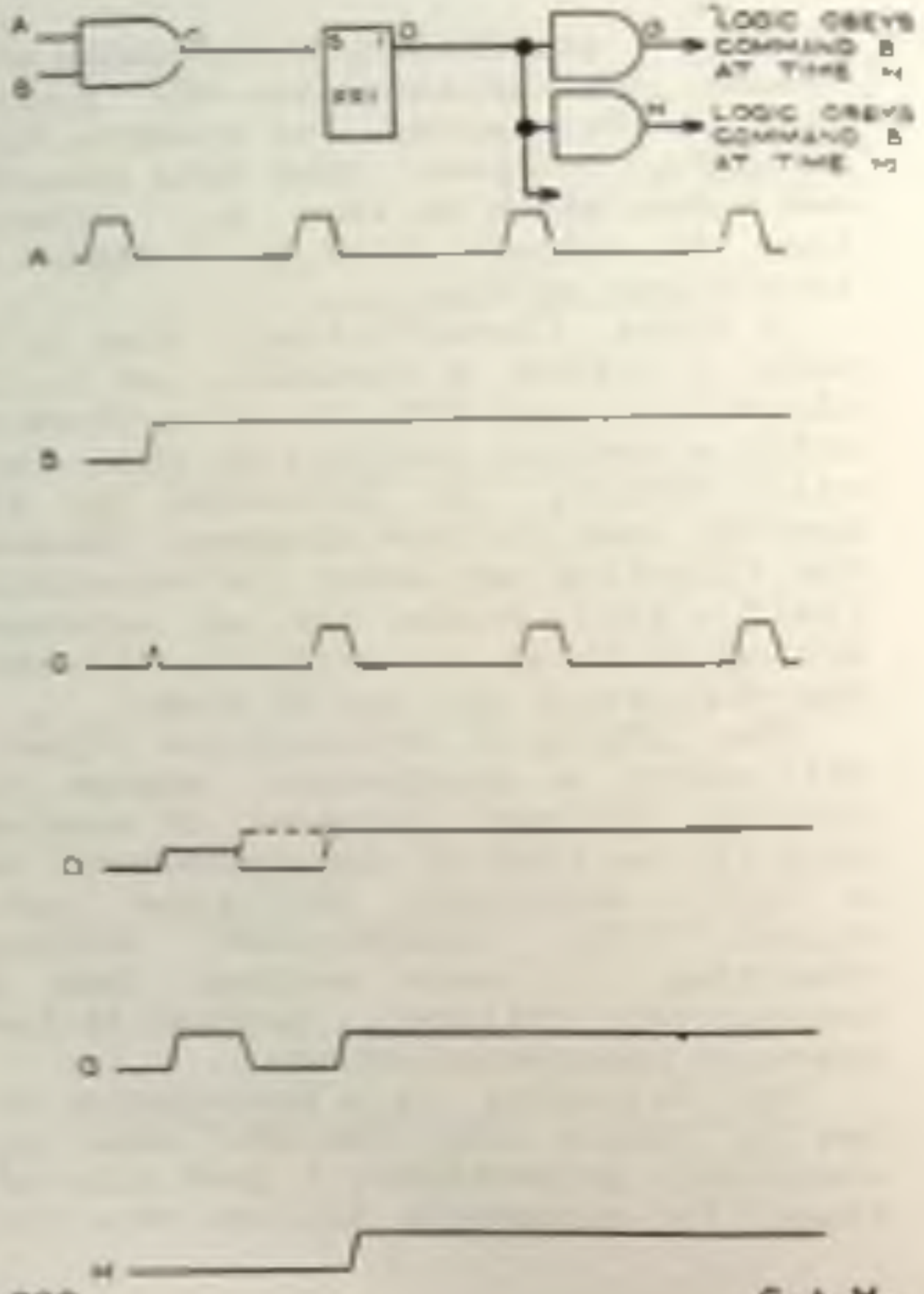

A
B
C
G
LOGIC OBEYS COMMAND B AT TIME t1
H
LOGIC OBEYS COMMAND B AT TIME t2
A
B
t0
t1
C
G
H

tistical probability that chaos will result, as indicated in the special case in the diagram. For example, line G might indicate that data transfer had taken place at time t_1, although line H only enabled the transfer to take place at time t_1.

A first (insufficient) step is to make C drive a bistable, as in the diagram on page 290. However, there is still a smaller possibility that chaos will result, as indicated by the special case in this diagram, because the flip-flop may enter its metastable (half - set) state for an extended period of time. As a result, the system will still get out of step.

The gating of asynchronous signals will carry a statistical chance of logical failure, because it carries with it the risk of the appearance of a half - amplitude, half-true logic signal. The intermittent failure resulting is more serious than a catastrophic failure, because it introduces undetected errors

The following is a description of how to reduce the failure rate to acceptable proportions. A good rule of thumb for acceptable failure rate for

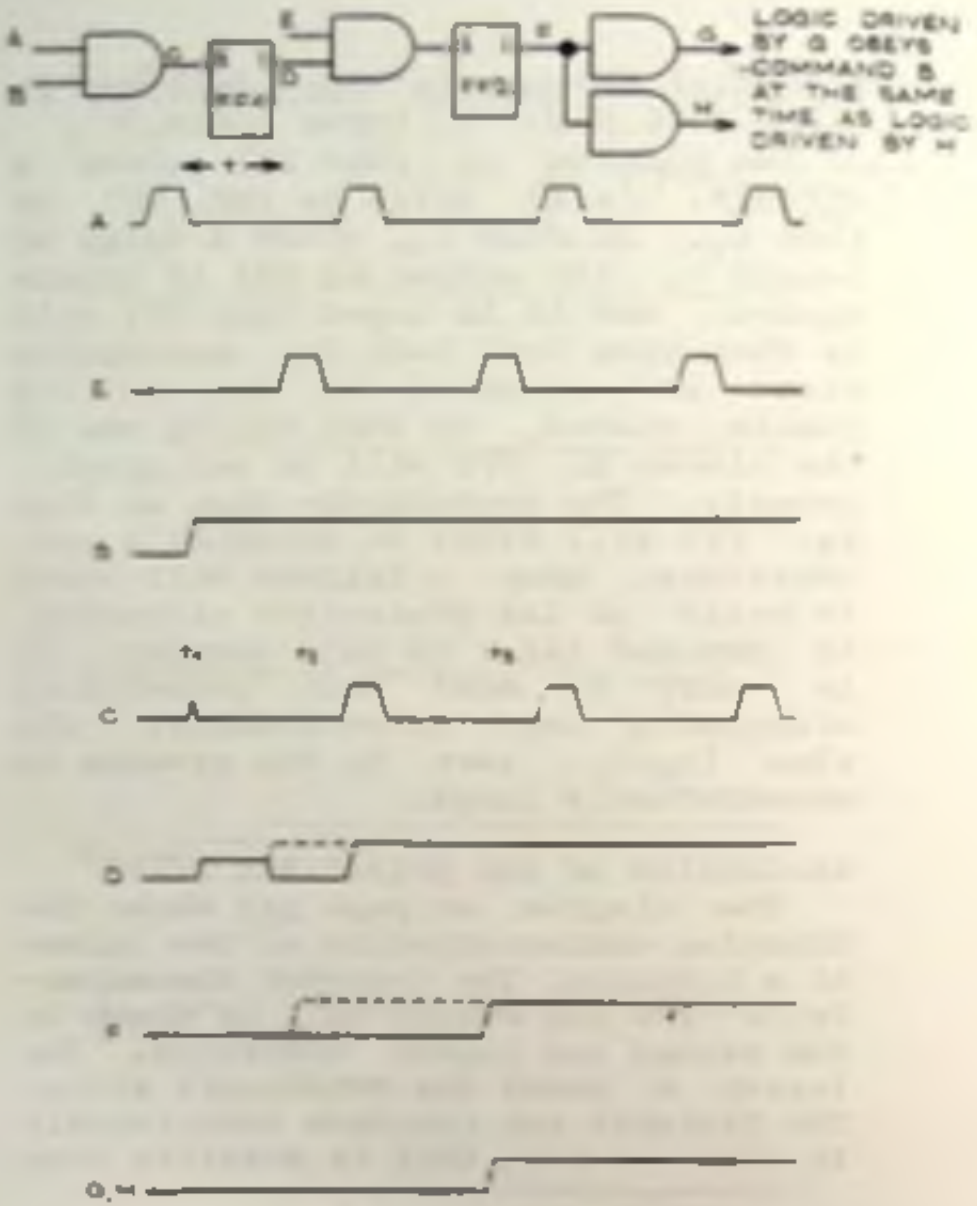

A
B
S
D
FF1
Q
LOGIC OBEYS COMMAND B AT TIME N
M
LOGIC OBEYS COMMAND B AT TIME N+1
A
B
C
D
Q
M

a logical circuit is one failure in about 1000 days, or three years.

The diagram on page 292 shows a circuit. (A.B) tries to set FF1 at time t_1. At time t_2, after a delay of length t, the output of FF1 is interrogated, and it is hoped that FF1 will by that time have left its metastable state and returned to one of its stable states, so that during one of the clocks E, FF2 will be set unambiguously. The probability that at time t_2, FF1 will still be metastable and, therefore, that a failure will occur in spite of its protective circuitry, is computed later in this chapter. It is easy to make this probability acceptably low. Unfortunately, the time (t_3-t_1) lost in the process is uncomfortably large.

DISCUSSION OF THE METASTABLE STATE

The diagram on page 293 shows the transfer characteristics of two halves of a bistable. The transfer characteristic for the second half is drawn in the second and fourth quadrants. The letter M shows the metastable state. The bistable can continue indefinitely in this state; this is possible even


A
B
C
RC4
D
E
FF2
F
G
H
LOGIC DRIVEN BY G OBEYS COMMAND B AT THE SAME TIME AS LOGIC DRIVEN BY H
t
A
E
B
C
D
F
G, H

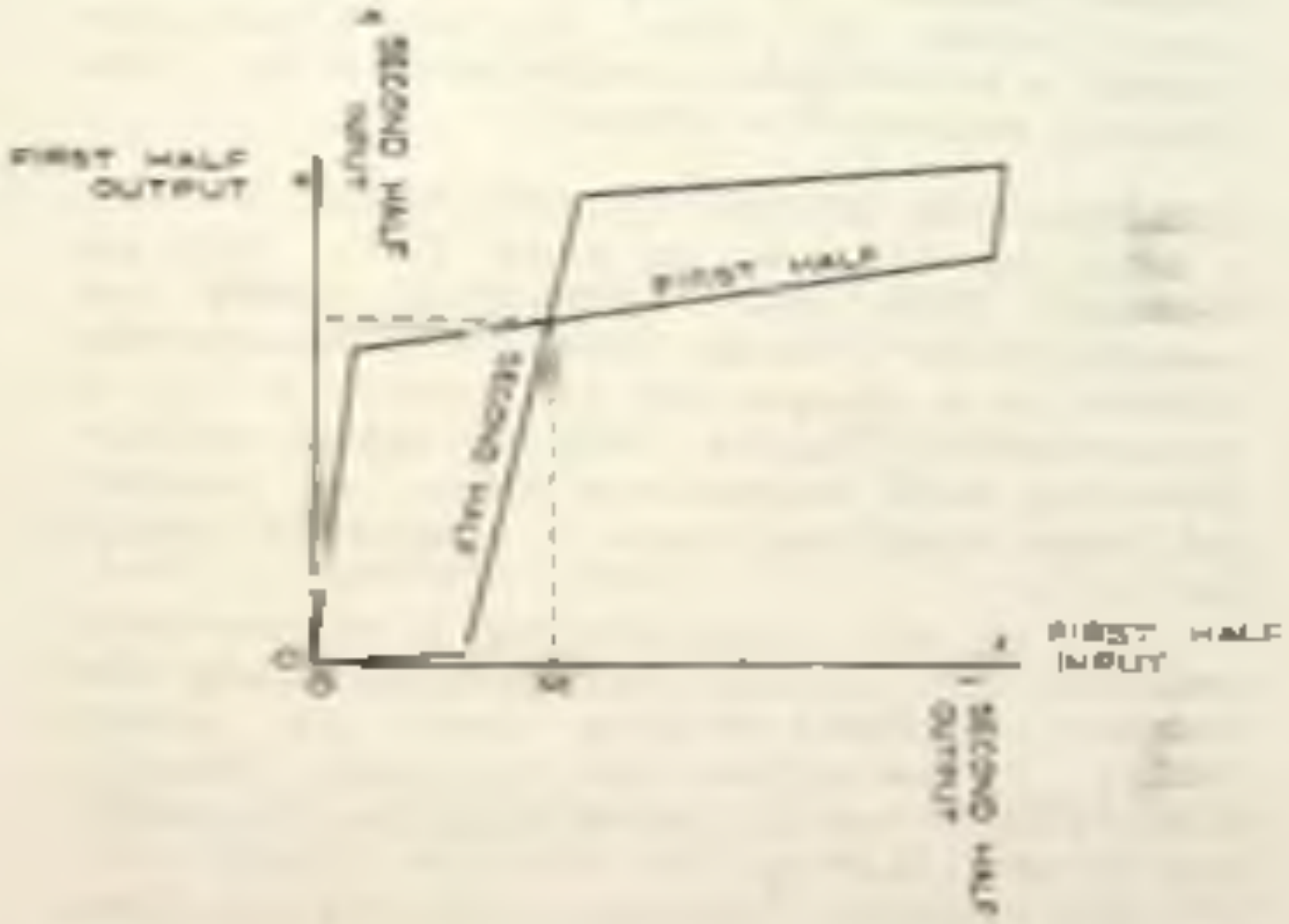
SECOND HALF OF BISTABLE
INPUT
OUTPUT
FIRST HALF OF BISTABLE
OUTPUT
START
INPUT
INPUT
OUTPUT
FIRST HALF OF BISTABLE
OUTPUT
INPUT
SECOND HALF OF BISTABLE

in a noisy environment. There is a certain pattern of random noise which will alternatively move the bistable to one side and the other of its metastable state, but not far enough for it to move to one of its two stable states.

It is possible for the bistable to continue indefinitely in its metastable state if the transfer characteristics of the two halves of the bistable differ. The diagram on page 295 shows the case where the first half of the bistable has a low threshold and the second half has a high threshold. The metastable state is where the two curves intersect.

The dotted line in the diagram on page 293 shows the case when the spike C in the previous diagram has pushed the bistable near to its metastable state - in this case, slightly further than the metastable state. We can see the events that follow by going counter-clockwise, starting with the first quadrant. In this case the bistable gets out of the transition region in a time 2d, or twice round the loop in the bistable.

The practical case of a metastable


FIRST HALF OUTPUT
SECOND HALF
FIRST HALF
FIRST HALF INPUT
SECOND HALF

state will be more complex than the steady - state picture described previously. In practice, the metastable state might be represented by an oscillation of the bistable outputs about a mean level represented by the steady metastable state.

CALCULATION OF THE FAILURE RATE

In the diagram on page 292, let us assume that the number of times per second that B goes true (and therefore there is a chance of failure) is f_B. A reasonable figure for f when multiplexing data transfers from a number of tape stations into a computer would be 10^5.

If we are considering a three-year period (about 10^8 seconds), then the number of times B goes true is about $10^8 f_B$. If we allow one failure during this time, we allow a failure probability of $1/10^8 f_B$.

The first bistable is driven into its metastable state if the width of the spike occuring at C at time t_1 is less than the delay around the loop in bistable FF1. If the transistor frequency bandwidth is F, then the delay around the bistable will be approximately 10/F seconds

If B goes positive f times per second and f<F, then the proportion of times that the trailing edge of A catches B and generates a spike is:

$$\frac{10/F}{1/f_A} = \frac{10f_A}{F}$$

where f_A is the frequency of A pulses.

Thus, the number of times in three years when a small spike occurs is:

$$10^8 f_B \frac{10f_A}{F} = \frac{10^9 f_A f_B}{F}$$

Let us assume that:

$$R = \frac{\text{Transition region amplitude}}{\text{Logic swing}}$$

Then R represents the proportion of the small spikes which will be of such amplitude as to partly succeed in setting bistable FF1, and so FF1 enters its metastable state - $(10^9 f_A f_B R)/F$ times in three years.

We now have the picture of a roughly half-sized pulse recirculating around the feedback loop of the

bistable FF1. We may assume that there is a regular distribution of amplitudes among these pulses, starting with small ones which just get into the transition region, through pulses nearly half-way through the transition region, to large pulses which are only just below the top of the transition region.

Each time this pulse passes around the loop, it is amplified twice. If the amplification in each half of the bistable is A, the amplification around the loop is A^2. In one trip around the loop, a number of pulses will drop out of the transition region. After n trips around the loop, the number of pulses still within the transition region will decrease by a factor $(1/A)^{2n}$.

Now we must wait so long that only one pulse remains in the transition region during three years. This means that:

$$1 = \frac{10^9 \cdot 6 \cdot 3^2}{r} (1/A)^{2n} .$$

Therefore, the number of times that the pulse must be allowed to circulate

round the loop is:

$$n = \tfrac{1}{2}\,\frac{\log(10^{3} f_{B} R/F)}{\log \beta}$$

The time t in the diagram on page 292 = nd where d is the delay around the loop of the bistable.

Example:

As an example, let us substitute the following values into the formula:

Number of times per second that signal B arrives - frequency of B = $f_B = 10^5$;

frequency of A - $f_A = 10^6$.

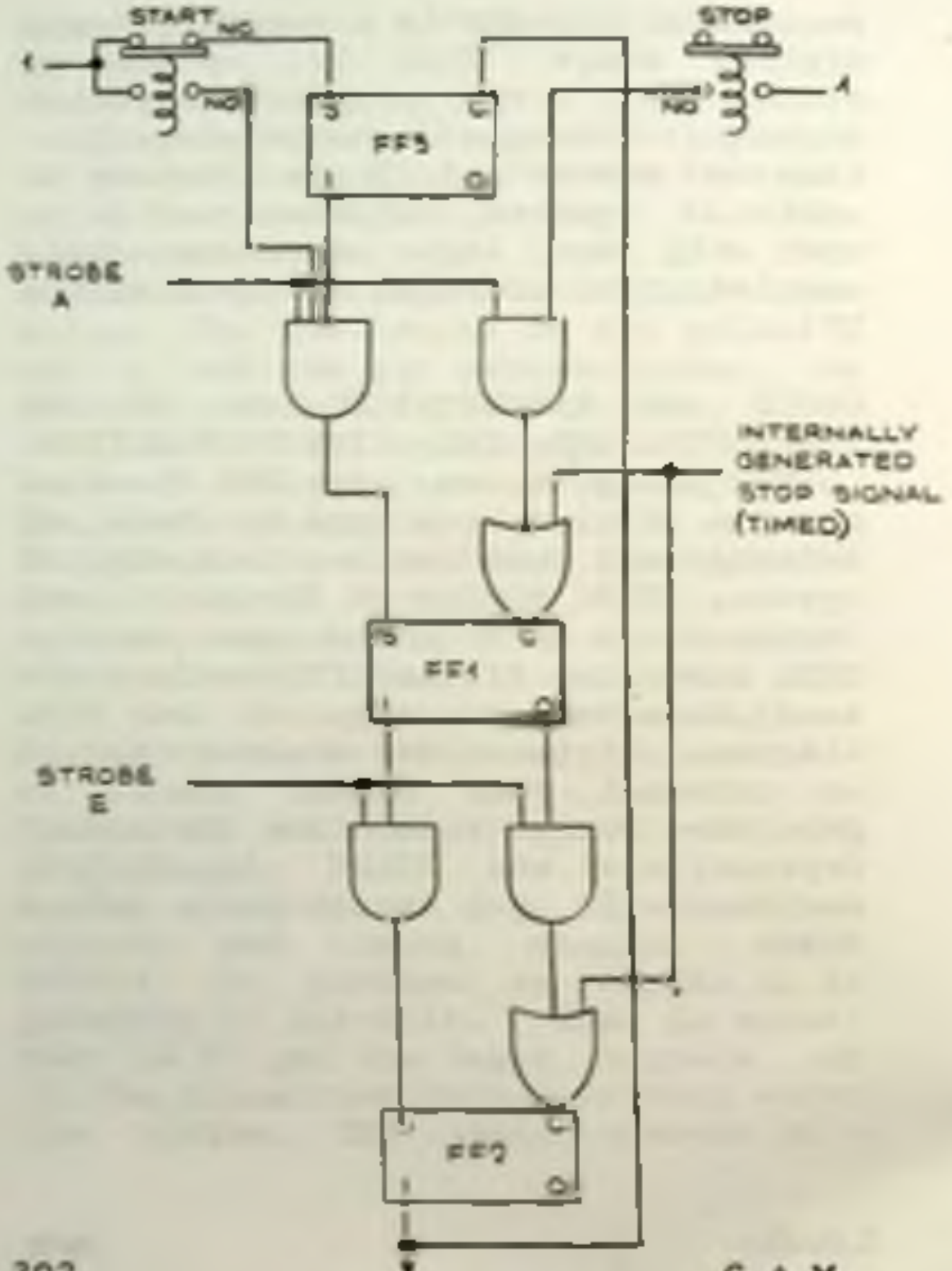
$$\frac{\text{Transition width}}{\text{Signal logic swing}} = R = 1/10;$$

Frequency bandwidth of transistors = $F = 10^9$.

Gain of one half of the bistable = $\beta = 5$.

Delay around the bistable = $d = 10/F = 10\text{ns}$. Then:

$$n = \tfrac{1}{2}\,\frac{\log\left[\frac{10^3 \times 10^{11}}{10} \times \frac{1}{10}\right]}{\log 5} = \tfrac{1}{2}\,\frac{23}{1.61} = 7.1;$$


START
NO
NO
STOP
NO
1
FF3
STROBE
A
INTERNALLY
GENERATED
STOP SIGNAL
(TIMED)
FF4
STROBE
E
FF2
MACHINE RUNNING

For more recent discussion and references, see:
L.R. Marino, "The Effect of Asynchronous Inputs on Sequential Network Reliability," IEEE Trans. Comput., vol C-26, No. 11, pp 1082-1090, Nov. 1977

See also the same journal, T J Chaney and also E G Wormald, vol C-28, No. 10, pp 802-804, Oct. 1979

The story of the postman and the milkman was written by Warren Littlefield and Thomas Chaney, Washington University, St. Louis, in 1966.

The second half of this chapter was published previously.

CHOICE OF TYPE OF LOGIC SYMBOLS

The primary aim in a logic diagram is to help someone trying to understand it, and logic diagrams should be engineered towards that end. From this point of view, it is first necessary to decide whether a block of logic is primarily

(a) a large electrical circuit made up of smaller circuits, many of which are identical, connected together according to broadly defined rules, or

(b) a large logic function which sequences through certain states in response to external stimuli, and whose outputs control other units.

It turns out that (b) is a better description of the fundamental nature of a block of logic, and that for the person trying to understand it, it is better to stress the logical function rather than the electrical nature of the circuits. It follows that whether a logic signal indicates "true" rather than "false", "active" rather than "passive", or a "one" rather than a " "zero", is generally more important

than its electrical level.

In discrete logic circuitry, it was easier (using diodes followed by an amplifying tube or transistor) to engineer AND-INVERT and OR-INVERT gates rather than straight AND and OR logic functions. From this engineering fact developed the theory that AND, OR and INVERT were the basic logic functions, and that since AND-INVERT and OR-INVERT each had two of these functions, they were regardad as more powerful than AND and OR. Thus the technical reason for using inverting gates was transmuted into a theoretical one, and we can expect our logic signals to continue to experience repeated inversion as they go through a block of logic, so that, for example, a piece of logic may generate the output signal "There is not not not not a parity failure" rather than the more somple signal "There is is is is a parity failure".

To overcome the confusions resulting from numerous cascaded inversions of logic signals, the American Military Standard (and also the inferior ASA standard) system for logic diagrams makes it possible to state the logical significance of a signal (true or false, one or zero,) regardless of its

electrical polarity, thus implying that a block of logic is primarily as described in (b) above. In contrast, the British Standard adheres to view (a), and emphasises the electrical physical reality of a block of logic at the expense of some loss of clarity in describing its logical performance.

When choosing a set of logic symbols, the first decision to be made is between (a) and (b). In general, those with experience only in (a) favour (a), and those with experience only in (b) favour (b). However, it is noticeable that those with experience in both (a) and (b) almost universally favour (b), a method of notation where the logical function is kept to the fore rather than the electrical reality.

Logic diagrams of type (b) run into immediate difficulty since the same electrical circuit can perform an AND function for positive going signals and an OR function for negative going signals. (If a gate outputs a positive level when both its inputs A AND B are positive, so acting as an AND gate, then it follows that it outputs a negative level if either its input A OR its input B is negative, so that for negative true

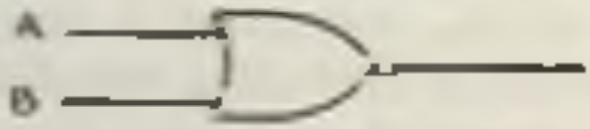
signals it acts as an OR gate. This is De Morgan's theorem.) The question arises whether such a gate should be drawn as an AND gate or an OR gate. To solve this dilemma, the logic diagrams of type (b) have two different ways of describing such a gate, one to denote AND and the other to denote OR. The idea of having two ways of describing the same thing horrifies some purists, but is well worth while in practice.

In the American Mil. Std. system of logic symbols, any logic gate can be drawn either as an OR, as follows:-

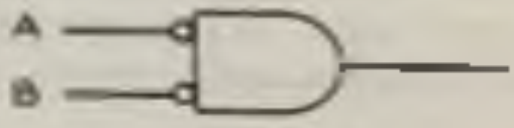
or as an AND, as follows:-

to suit the feeling of the designer as to the nature of the function performed.

The fact that a signal entering or leaving a logic gate is negative going is indicated by a circle, so that (for instance) a gate which outputs a negative level if its input A is positive OR its input B is positive can be drawn as follows:-



(This is commonly called a NOR gate.) However, since the same gate could equally well be said to have a positive output if both inputs A AND B are negative, it can also be represented as follows:-



Choice of symbol depends on whether the designer believes the logic gate is performing an AND or an OR function. Now generally, the logic designer thinks in terms of signals which are true and false, active and inactive, ones and zeros, rather than electrically positive and negative, and so it will turn out that logic interconnections in a diagram tend to have either circles at both ends or no circles at all, depending on whether the designer is thinking of the line being true, active or value one in the electrically low or in the electrically high state.

Someone studying the logic can then ignore electrical polarity and merely follow the path of a logic signal remembering whether it is true or false. It would be ideal if every line in a set of logic diagrams were one of two kinds, those with no circles and those with circles at both (or every) end. Unfortunately, this is an unachievable ideal, and some lines will remain with a circle at one end and not at the other. The number of such lines should bekept to a minimum, and can usually be reduced to only two or three in a system. (It will be found that the attempt to

EASE OF DRAWING

The U.S. Mil. Std. shapes are difficult to draw. Use of a template only solves the problem to a degree, be because the number of sizes of symbols that can be drawn is restricted.

CROSS REFERENCING

U.S. Mil. Std. 806B has a very fine system of cross referencing, and is to be strongly recommended.

While designing and drawing logic, it is invaluable to apply the discipline that whenever a signal is used as a logic input from another page, two way cross referencing to the origin of the signal is written in, even on the draft design. This cross referencing is like an invisible line drawn to link the two points; the source of the signal and its destination. Every time the designer makes a change to his design, he is well advised to update the signal cross referencing as well. The recommended system is where each logic diagram has lettered grid markings along the top of each sheet and numbered grid markings down the side. A logic signal near the

top left of sheet 3 will then have the grid reference 3B2, the letter separating the sheet number 3 from the latitude grid number 2.

DETAILS OF STANDARDS

The standard recommended here is the U.S. MIL-STD-806B, now available only from the publishers of this book. Other standards mentioned are the American ASA Y32.14 - 1962 published by the IEEE, 345 E 47th St., New York 17, N.Y., and also the British Standard BS 530.

ELECTROMAGNETIC THEORY

For half a century, electronics was largely restricted to radio telegraphy of a particular kind, where the transmitter, or sender, excited the intervening space, or ether, sinusoidally at a frequency at which this was feasible without using too much power. This excitation was called the carrier. The much lower frequency signal to be carried was added by amplitude modulation or slight frequency modulation of the carrier. The receiver would then separate out the signal from the carrier.

Since 90% of the applications for electronics were in this kind of radio telegraphy, writers on electromagnetic theory assumed that they were talking only about sinusoidal excitations, because nothing else seemed to occur in practice. In general, they slipped in the frequency term ω unannounced on about the third page.

The first significant advance beyond "steady state s.h.m." electronic

engineering probably came with the advent of radar in the forties, but this was not a major change because still a pulsed sine wave was used, albeit at very high frequency. Radar heralded the move away from frequency to time, but since it was only a small, specialised part of the industry it made little impact on fundamental theory.

Another attitude which developed from radio telegraphy and which has an unfortunate effect on digital system design was the feeling that the whole electronic art was hazardous, doubtful and somewhat magical. A sunspot, another transmitter, a storm which damaged the transmitting or receiving aerial, or a host of other things could disrupt operations. It was all a far cry from the computation of pay packets or other contemporary applications of digital computers, where complete reliability is all-important.

Two more factors drove electromagnetic theory far away from its later digital applications.

1. Fourier Series. Any time varying periodic function can be reduced to a number of sinusoidal excitations superposed. Thus it appeared that it was

only necessary to understand a circuit's response to simusoidal excitations. However, if we try to reduce a periodic series of square pulses to a family of sine waves we find that they are infinite in number and the fastest is of infinite frequency. In the practical case, the transmission of a single step (logic transition from false to true) or a single square pulse, it is doubtful if a Fourier series can be devised to replace it. The question of whether functions which are not periodic can be reduced to a combination of sine waves is always avoided. It appears that the frequencies of the necessary family of sine waves, which are infinite in number, would range from zero to infinity. So the original, simple, innocent looking step or pulse has been replaced by a dog's breakfast of sine waves, stretching from minus infinity to plus infinity in time and ranging from zero to infinity in frequency. This is surely a reductio ad absurdum of the traditional approach to electromagnetic theory. Understanding such a model turns out to be far more difficult than the original problem, that of understanding the

effect of one digital step or pulse.

2. Complex Numbers. The fact that a sine wave of certain applitude and phase can be represented by a complex number is so fascinating, and lends itself to so much enjoyable mathematical manipulation, that everyone finds it difficult to give up the complex number when they give up sine waves completely, and start a new career in digital system engineering. Everyone tries to take the baggage, the manipulative and mathematical tools and techniques which were so useful and enjoyable in radio telegraphy and thence in college examinations, with them into the field of digital system engineering, where they have little relevance and generate much confusion.

Stripped of its historical encumbrances, what electromagnetic theory remains to the digital system designer? The answer is, very little. We still sometimes have recourse to concepts grounded in sinusoidal electronics, for instance the concept of skin depth, but these should be regarded as importations needed to aid continuing weaknesses in our theoretical framework, in the same way as foreign words are

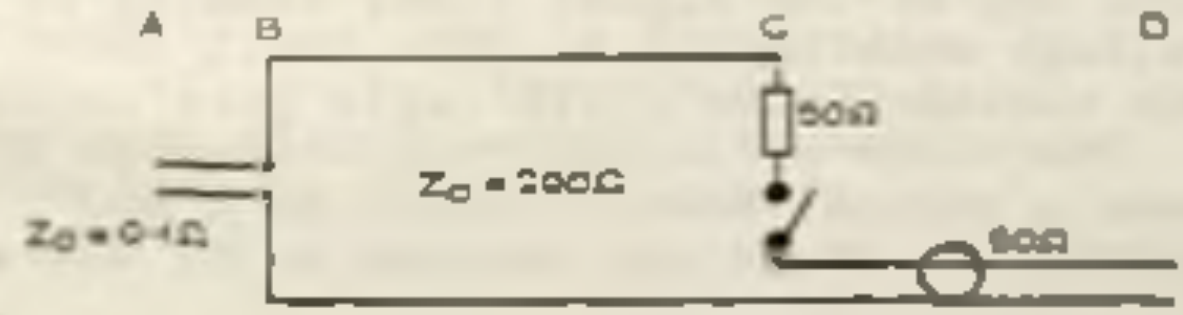
occasionally used in speech when the native language is lacking. The speaker in no way considers by this diversion that he should really be speaking the foreign language, or that the foreign language is superior. Similarly, the digital system engineer should be able to take advantage of concepts from the foreign field of telecommunication without remaining subservient to it.

The fundamental digital waveform is the step, and any waveform can be made up from a combination of small steps, each having zero rise time. If the response of a linear circuit to such a step is understood, then its response to any waveform can be deduced by superposition.

The response to a step by circuits comprising discrete components is well understood. The voltage waveform across a capacitor in an R-C circuit or across a resistor in an L-R circuit is an exponential with a rise time equal to 1/RC or L/R.

SWITCHING LOADS IN DIGITAL SYSTEMS

Consider a decoupling capacitor AB attached to the +5v line BC. Energy current is oscillating to and fro in the capacitor AB. When it reflects back at B, a very small proportion leaks out and then oscillates between B and C. At the same time, a smaller amount of energy current is oscillating to and fro between B and C. Each time it reflects at B, a small amount leaks into the capacitor BA. The line ABC appears to be sitting steadily at a voltage +5v. In fact, at any point along ABC, energy current of amplitude +2½v is travelling to the right and a second equal amount of energy current



is travelling to the left.

Because of the open switch at C, all energy current coming from the left is reflected back towards the left.

Now the logic gate at C changes. The switch (pull up transistor) closes, the intention being that the long signal line CD should rise from 0v to +5v.

Energy current travelling from the left to C along a 200 ohm line sees a termination of 100 ohms, leading to partial reflection and the remainder travelling on to the right. From page 83, equation (26), the portion reflected is given by the formula

$$5 \cdot \rho = 5 \frac{100 - 200}{100 + 200} = -5 \cdot \tfrac{1}{3} \text{ v}$$

From equation (27), $+5 \cdot \frac{2}{3}$ v goes ahead to the right, half of it into the 50 ohm resistor and the other half ($5.\frac{1}{3}$v) travelling down the signal line.

(In due course, this $5.\frac{1}{3}$v will arrive at a virtual open circuit at the right hand end of the signal line, leading to voltage doubling up to $5.\frac{2}{3}$v; well above the threshold for a TTL logic gate input.)

The $-5.\frac{1}{3}$v reflected from C towards B sees a virtual short circuit at B and inverts to $+5.\frac{1}{3}$v and returns to C, where

it causes the launching of a signal one third as large at the signal originally launched. Total energy current now flowing to the right from C is

$$5.\tfrac{1}{3}v + \tfrac{1}{3}.5.\tfrac{1}{3}v = 5.\tfrac{4}{9}v$$

With voltage doubling at the right hand end of the signal line, this rises to $5.\frac{8}{9}$ v, sufficient to switch any logic gates driven. Meanwhile, the initial signal reaching B, amplitude $-5.\frac{1}{3}v$, does not see exactly a short circuit. Actually, the reflection coefficient (equation (26)) is $\frac{1/10 - 200}{1/10 + 200}$.

A tiny signal enters the capacitor, value $v(1+\rho) = \frac{2/10}{200} . 5 = 5$ mv.

This signal travels to A where it doubles, and the capacitor now sags to 10 mv below 5v. If the only source of energy is the capacitor, then by repeated reflections the capacitor will gradually sag down to zero volts. In fact, of course, the situation is restored via yet another transmission line back to the power supply.

INTERNAL OBSTRUCTION AND SUPERFICIAL CONDUCTION

The properties of a perfect conductor are derived from those of common conductors by examining what would happen if the resistivity were continuously reduced, and ultimately became zero. In this way we find that a perfect conductor is a perfect obstructor, for one thing, which idea is singularly at variance with popular notions regarding conductors. But it is also a perfect conductor literally, though in a different sense to that commonly understood. Ohm's law has played so important a part in the development of electrical knowledge, especially on the practical side, that it is really not at all a matter of wonder that some practicians should have been so reluctant to take in the idea of a conductor as an obstructor. Scientific men who can follow the reasoning by which the functions of conductors follow from known facts have no difficulty in pursuing the consequences far beyond experimental observation. Again, younger

men, with fewer prejudices to surmount, do not find much trouble with superficial conduction and internal obstruction. But the old established practitioner with prejudices, who could not see the reason, was put into a position of some difficulty - resembling chancery. If you have got anything new, in substance or in method, and want to propagate it rapidly, you need not expect anything but hindrance from the old practitioner - even though he sat at the feet of Faraday. Beetles could do that. Besides, the old practitioner is apt to measure the value of science by the number of dollars he thinks it is likely to bring into his pocket, and if he does not see the dollars, he is very disinclined to disturb his ancient prejudices. But only give him plenty of rope, and when the new views have become fashionally current, he may find it worth his while to adopt them, though, perhaps, in a somewhat sneaking manner, not unmixed with bluster, and make believe he knew all about it when he was a little boy! He sees a prospect of dollars in the distance, that is the reason. The perfect obstruction having failed, try

the perfect conduction.

The sense in which a perfect conductor is a perfect conductor in reality is that it allows electromagnetic waves to slip along its surface in a perfectly free manner, without waste of energy. Though perfectly obstructive internally, it is perfectly conductive superficially. It merely guides the waves, and in this less technical sense of conduction the idea of a perfect conductor acquires fresh life.

ON THE TRANSMISSION OF ENERGY THROUGH WIRES BY THE ELECTRIC CURRENT

Consider the electric current, how it flows. From London to Manchester, Edinburgh, Glasgow, and hundreds of other places, day and night, are sent with great velocity, in rapid succession, backwards and forwards, electric currents, to effect mechanical motions at a distance, and thus serve the material interests of man.

By the way, is there such a thing as an electric current? Not that it is intended to cast any doubt upon the existence of a phenomenon so called; but is it a current - that is, something moving through a wire? Now, although nothing but very careful inculcation at a tender age, continued unremittingly up to maturity, of the doctrine of the materiality of electricity, and its motion from place to place, would have made me believe it, still, there is so much in electric phenomena to support the idea of electricity being a distinct entity, and the

force of habit is so great, that it is not easy to get rid of the idea when once it has been formed. In the historical development of science, static phenomena came first. In them the apparent individuality of electricity, in the form of charges upon conductors, is most distinctly indicated. The fluids may be childish notions, appropriate to the infancy of science; but still electric charges are easily imaginable to be quantities of a something, though not matter, which can be carried about from place to place. In the most natural manner possible, when dynamic electricity came under investigation, the static ideas were transferred to the electric current, which became the actual motion of electricity through a wire. This has reached its fullest development in the hands of the German philosophers, from Weber to Clausius, resulting in ingenious explanations of electric phenomena based upon forces acting at a distance between moving or fixed individual elements of electricity.

Return to our wire from London to Edinburgh with a steady current from the battery in London. The energy is poured out of the battery <u>sideways</u> into the

dielectric at a steady rate. Divide into tubes bounded by lines of energy-current. They pursue in general solenoidal paths in the dielectric, and terminate in the conductor. The amount of energy entering a given length of the conductor is the same wherever that length may be situated. The lines of energy-current are the intersections of the magnetic and electric equipotential surfaces. Most of the energy is transmitted parallel to the wire nearly, with a slight slant towards the wire in the direction of propagation; thus the lines of energy-current meet the wire very obliquely. But some of the outer tubes go out into space to an immense distance, especially those which terminate on the further end of the wire. Others pass between the wire and the earth, but none in the earth itself from London to Edinburgh, or vice versa, although there is a small amount of energy entering the earth straight downwards, especially at the earth "plates". If there is an instrument in circuit at Edinburgh, it is worked by energy that has travelled wholly through the dielectric, then finding its way into the instrument...

The last two chapters were written by Oliver Heaviside.

"Electromagnetic Theory", Vol. 1, page 337; and "Electrical Papers", Vol. 1, page 434, 1892.

Another series is now available from C.A.M. Publishing:
ELECTROMAGNETIC THEORY VOL 1 and VOL 2 by I. Catt, ISBN: 0 906340 02 0 , 0 906340 03 9.

Also "Electromagnetism 1" by Ivor Catt, pub. Westfields Press, 121 Westfields, St. Albans AL1 4JR, 1994.

10 3 6 7 9 3 4 140 24 74 99 70 19 47
5 2 3 7 2 10 6 44 28 33 78 35 80 74
8 2 12 4 3 9 6 104 28 170 40 29 118 91
6 1 5 2 2 8 7 91 3 56 28 29 75 99