MAGNETIC RECORDING CIRCUITS

Ian H. Graham

Mr. Graham is presently the Senior Staff Scientist to the Recording Technology Center of the MEMOREX CORPORATION, Santa Clara, CA. 95052

TABLE OF CONTENTS

HIGH DENSITY DISK DRIVE TECHNOLOGY

2/77

Introduction

Rotating storage devices have traditionally occupied a niche to themselves by providing low cost storage of large amounts of data. Slow access times always characterize this area of storage. This is in contrast to the core and semiconductor memories which feature fast access but at high cost. With disk or drum memories, large amounts of data can be made readily available to the computer as "on line" storage.

History

During the past twenty years of disk drive development, the cost per stored bit has gone down considerably while the amount of stored information per machine has greatly increased. The earliest disk drives used 24 inch fixed disk arrays with hydraulic accessing mechanisms. These were usually for large size computers. Their physical size usually precluded their use with small office computers.

With the invention of the removable 14 inch disk and disk assemblies, a new market was opened up providing disk drives to the small computer user. These disk packs could be removed and stored at will. Programs were written to call for a certain pack or packs to be installed to complete the job at hand. The concept of a resident computer program further increased the use of disk files.

The capacity of disk files increased with each new technology step. In order to permit these technology steps, improvments needed to be made to the disk surface finish, the magnetic coating materials, the air bearing or air lubricated head construction, the read/write head positioning mechanism and associated electronics, including the logic family, used to control each machine function and many

other areas. Each new improvement required finer tolerancing of most parts associated with the disk drive mechanisms. Higher storage densities are usually achieved by increasing the radial track density and the circumferential bit density. Increasing the track density has been a problem largely controlled by the tolerance build up of the mechanical parts associated with the disk drive spindle bearing system and the access mechanism. With the invention and successful implementation of a track following servo system, further increases in track density were possible until the tolerance build up associated with pack interchange forced the designers back to the concept of fixed disk storage again. By now, the amount of storage per disk drive and the present requirement to have all data on line to the computer at all times has reduced the need for pack interchange thus making possible still further increases in track density. Track densities have been increased from about 20 tracks per centimeter to a present value of 189 (378) tracks per centimeter. Developments presently underway in track following techniques involve the individual addressed head with staggered servo data and read-write data. This may eliminate these last barriers and permit removable packs on very high density machines.

Circumferential bit density increases usually require reductions in magnetic head to disk surface spacings. These changes have not come easily as the finish or flatness of the disk surface must be improved with each decrease. The magnetic oxide coating materials must change along with the size and shape of each magnetic oxide

particle. On early disk drives the air bearing formed between the magnetic head and the disk surface was controlled by forcing compressed air between the two surfaces. An inventive application of air lubrication principles provided the present self lubricated head air bearing. Typical spacings started out at around 12 microns. Today the head to disk spacing is around a half micron. The gap between the magnetic pole pieces of the head have also been reduced to permit closer bit spacing. Values presently used are around 1 micron. The materials used to make the head pole pieces have changed from permalloy to ferrites because of the increased frequencies involved in record and read back functions.

Requirements for increased logic speed have brought their own family of improvements. These range from the vacuum tube or valve, through transistors to the present specialized integrated circuits. Typical data speeds have gone from 1 bit per 100 microseconds to a present 1 bit in one tenth of a microsecond. Storage capacities have changed from one million bytes per machine to over 300 million. These rapid improvements and increases in capacity will continue for at least another decade. There are designs on the drawing boards of several manufacturers that will permit a four to eight fold increase in capacity within the next two years with no real end in sight. For each limiting factor new technologies have been invented. For example, as track densities increase, the width of a track decreases. The head materials presently used have a grain size equal to the track width of the next generation disk drives. Already many firms are working on thin film heads. These heads are made by depositing thin films of magnetic metals or alloys to

dimensions far smaller than the grain size of the best ferrites. The disk coating materials, which presently consist of tiny particles of ferric oxide bonded in an epoxy resin layer of about one micron thickness, will be replaced with thin films vacuum deposited on the disk at thicknesses approaching 50 thousandths of a micron or 5×10^{-8} meters.

Tremendous improvements have been made in the codes used to transmit the data. Error detection and error correction codes permit accurate data even with disk defects encompassing more than a whole byte of data in a record. Concepts have now been developed which permit a disk surface defect to be skipped during the write process. Further improvements in addressing will permit many such defects to be transparent to the user.

Competition

The extension of disk drives as low cost, high density storage devices is expected to continue for many years to come. Magnetic recording requires low energy per bit to write and takes a short time to write. There is a lower limit to the time needed to write a bit. It is controlled by the domain switching time of the disk coating material. For ferric oxide films this is about 50 nanoseconds or about half the present bit spacing time. Transmission speed is therefore limited to 20 million bits per second. The actual density of the recording for both track density and circumferential density is limited only by the magnetic domain size. This limit will not be reached for many years.

Competing technologies are electron beam, holographic, semiconductor RAM, charge coupled devices, and bubble memories. Of these, holographic and thermal electron beam memories are slow writers. Certain dyes permit write, read and rewrite capability

for holographic memories but most are read only devices. The same limitation to read only after an initial write is true of thermal electron beam memories. Their usefulness is limited to large library storage such as legal cases or court histories where the data does not need to change over many years. Bubble memories, charge coupled devices, semiconductor-electron beam and semiconductor MOS and bipolar RAM will compete and replace core memories or fixed head per track machines within the next few years but they cannot replace the large capacity disk drive without a more than tenfold decrease in cost and a more than doubling of the world's semiconductor capacity. Such is not likely within ten years.

Future

I suppose this is the hardest part to summarize. Since there is easily an eight to tenfold increase in capacity presently available within the current technology, one might suppose further technological changes might produce another decade increase in capacity. The amount of data available in a single disk drive could well become 30 thousand million bytes by 1990. Thru put is limited to 20 million bits per second or 2.5 million bytes per second because of the magnetic domain switching time limitation. This may well equal the best channel acceptance times of the next generation of computers. Byte size may be increased which will reduce the cycle time per byte. Interleaved by byte records can double circumferential density without increasing channel speeds. Staging devices may be employed to buffer the disk data and the channel.

Presently a storage control unit is required for a group of drives.

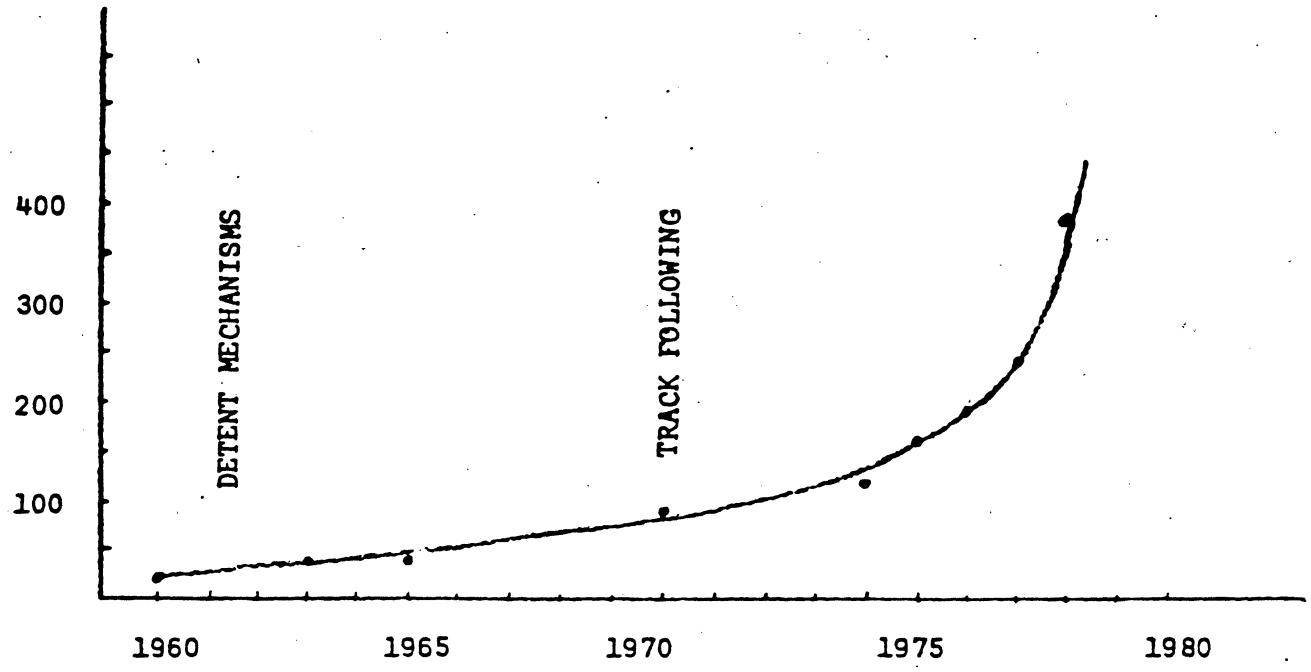
This storage control unit has its own microprocessor constructed of discrete logic blocks. It is controlled by a resident microprogram that performs all the housekeeping functions for a large number of drives. Future drives may each have their own microprocessor. Each drive may then be tailored to a specific storage function by means of its own microprogram. Many tasks presently performed by the controller or even the main computer can now be delegated to an integrated drive. Processing of data for storage is an easy task for such a drive. Processing the data prior to transmittal to the main computer is an easy step, particularly if we have individualized disk drives that are tailored by a particular microprogram. Combinations of disk drive and mass tape systems are currently available. Their future usage may well place a company or government in real time control of its resources or records.

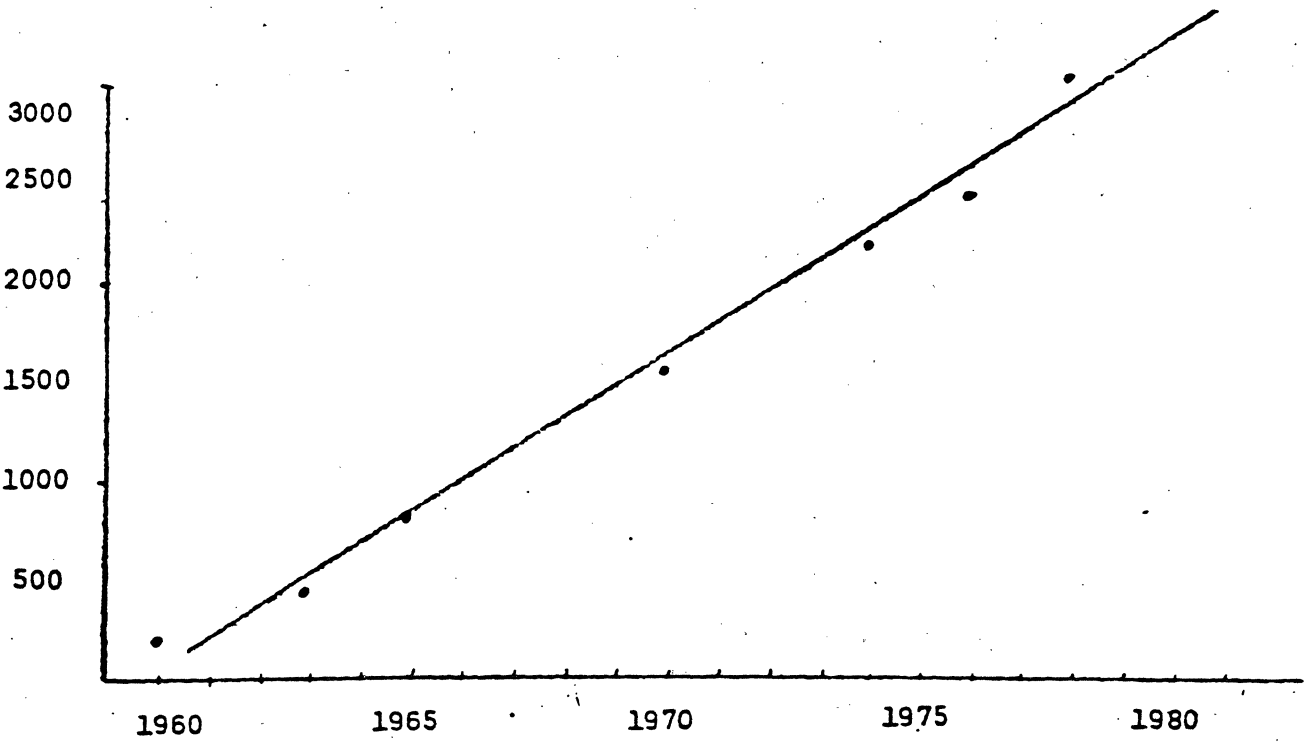
Conclusions

Disk drives offer large, non volitile data storage that is accessible in miliseconds. It has an advantage of not requiring periodic replacement such as tapes. Destruction of data due to catastrophic malfunctions such as head crashes have been minimized by the use of low mass, light load magnetic heads in sealed environments.

Data storage and retrieval has made possible the present growth in computer technology. As the storage capacity of a computer installation is increased so is its capacity to handle complex programs. Presently there are a few programs developed or being developed that require very large data bases. These are mainly in the field of simulation, modeling, and pattern analysis. As these fields progress in their complexity and capability larger

data storage devices will be required. The technology presently available can provide storage capacities that challenge our ability to manage them. Considerable work is needed in data management and programming to provide the type of environment needed to handle large data base systems for tomorrows research and development. As our data base expands, so do the risks to the freedom of individuals caught up in such a network of data storage. Responsible governments will, therefore, need to guard against such encroachments.


400
300
200
100
DETENT MECHANISMS
TRACK FOLLOWING
1960
1965
1970
1975
1980


3000
2500
2000
1500
1000
500
1960
1965
1970
1975
1980

Summary

High Density Disk Drive Technology

The development of the digital computer has required a parallel development of storage devices. Of the many available technologies for data storage, magnetic disk drives lend themselves to the best solution by providing low cost, easily accessible, non volitile storage. The history of their development extends over 20 years first with the drum memories and then disk memories. During this period storage capacities have increased over a hundredfold while costs have tumbled making todays cost per bit the lowest in history. The paper presents some of the history of the development of the disk drive by outlining the major improvements in technology that have taken place. A comparison is provided that compares the various other technologies used for data storage and lists some of their advantages and disadvantages.

The trend towards larger data based systems and the storage devices needed to handle the storage requirements of the future is discussed. Some concepts of future usage couple the now popular micro processor with the disk drive which can provide a compact intelligent storage device. This power is only hinted at in the drive and controller combination which is presently in use. The controller portion can be expanded to process much of the data before it is passed on to the computer instead of just doing housekeeping and sequencing.

iker

?

Ian Graham was born in Rexburg, Idaho U.S.A. in February, 1930. He was educated in Australia and returned to the United States of America in 1951. He served in the U.S. Navy for 4 years as an aviation electronic technician. Upon discharge he attended The University of Utah and received his Batchelor of Science degree in Electrical Engineering in 1960. Following graduation he worked for IBM Corporation for 9 years working on the development of magnetic disk drives. In 1969 he joined Memorex Corporation. He is presently the manager of Recording Technology and has the responsibility of supervising the development of all disk read/write functions within Memorex.

RECORDING ELECTRONICS

THE HEAD STRUCTURE

The magnetic head is a modified torroid of magnetically permeable material. It is provided with an air gap and a suitably dimensioned window around which a coil is wound. The shape may vary with intended use but every attempt is made to keep the structure magnetically efficient.

The terminology is illustrated in Fig. 1.1. The core has some thickness and width. The width defines the track width recorded on some media. The throat height is the thickness of the core at the air gap, and the length of the gap is referred to as the gap length.

The coil is usually referred to by the number of turns and whether it is centertapped or not.

The ring structure is the one most used in the literature, particularly in writing the equations describing its action or interaction. No attempt will be made here to go into this aspect, but it is well described in the literature, Hoagland and Karlquist being the earliest authors.

For our purposes we will be satisfied by looking at the field lines and their behavior, as affected by the various mechanical dimensions. The magnetic fields produced by current in the windings is mostly developed across the higher reluctance of the air gap. The field lines leave the higher permeability core surface normal to that surface and seek the opposite side, terminating normal to that surface.

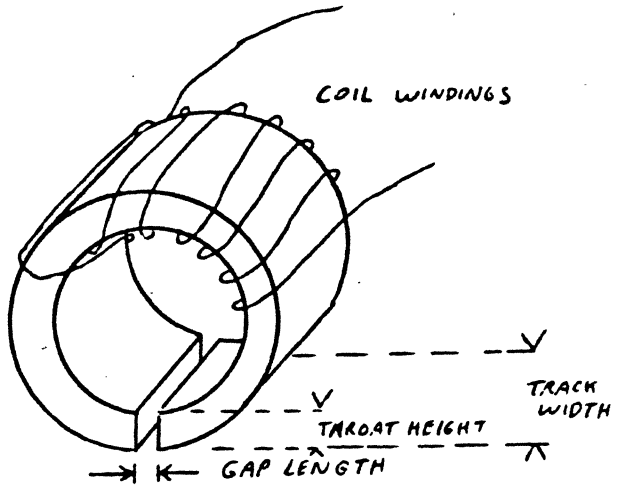


FIG 1·1 RING HEAD

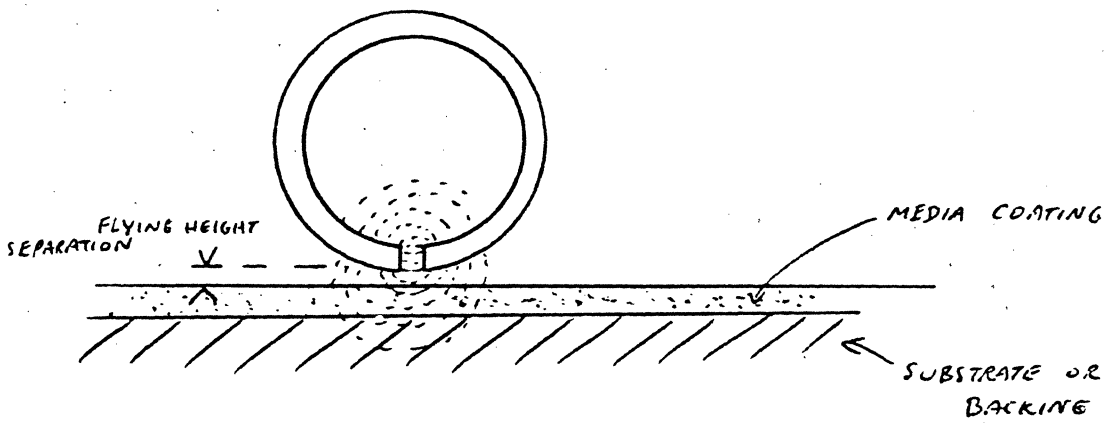


FIG 1·2 FIELDS

The field intensity is greatest within the gap and diminishes with increasing distance following the inverse square law. As can be seen in Fig. 1.2, the field expands out from the gap. It is this portion of the field that is used for writing, the remainder is wasted. Obviously, the closer to the gap the media is kept, the more efficient the Write process. This separation then becomes a fundamental parameter in the recording and reproduction process.

In hard disc drives it is referred to as flying height and in tape drives as separation. In tape applications where the tape is expected to be kept in contact, any separation of the head and media is deterimental. In disc drives it is deliberate and is part of the design. This is necessary in order to minimize head-media wear expected at the higher velocities used.

Other structures that have been used to date include those shown in Fig. 1.3. The windings may be either around the core itself or around the back bar. This structure has been implemented in ferrite in the IBM 2314, and 3330 machines. There are two back gaps that are shorter and larger in area than the main gap. Here the reluctance is minimized to increase efficiency. The CI structure was used in all the earlier disc machines from the IBM Ramac 350 to the 2311. The pole pieces were made of laminated Permalloy in order to reduce both core and hysteresis losses. Notice the poor back gap contact in Fig. 1.4. This was due to the slight angle necessary to produce the front gap using lapped parts. This head structure was later abandoned due to the poor frequency response of the thick laminations of the Permalloy. Ferrite afforded improved permeability at higher frequencies

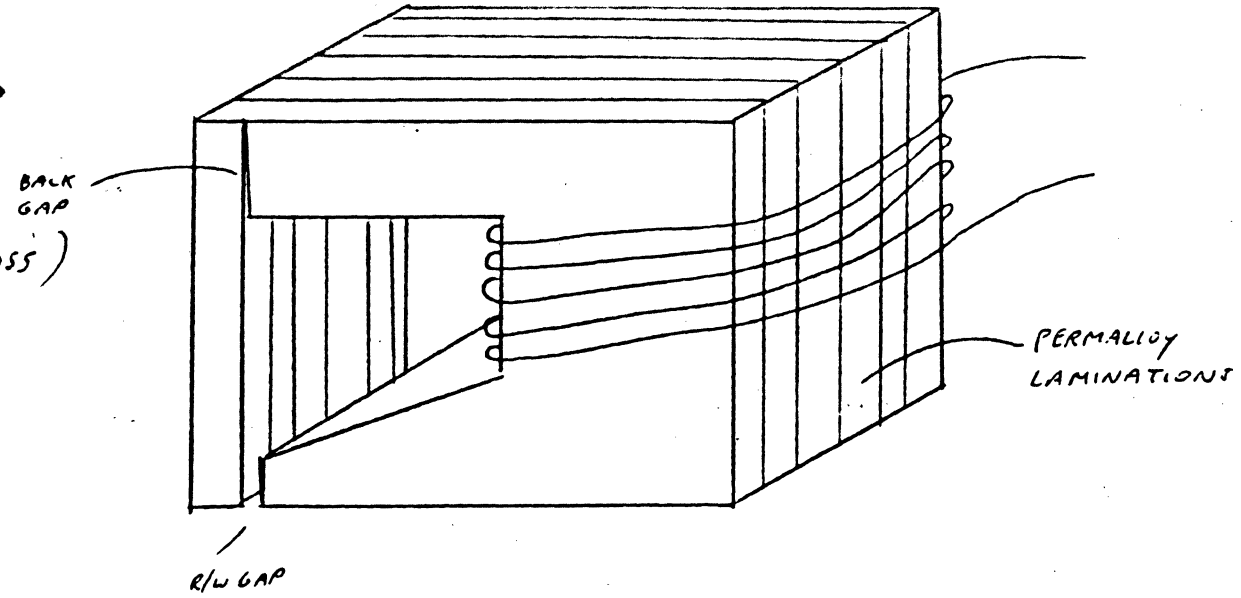


FIG 1.4 C I STRUCTURE

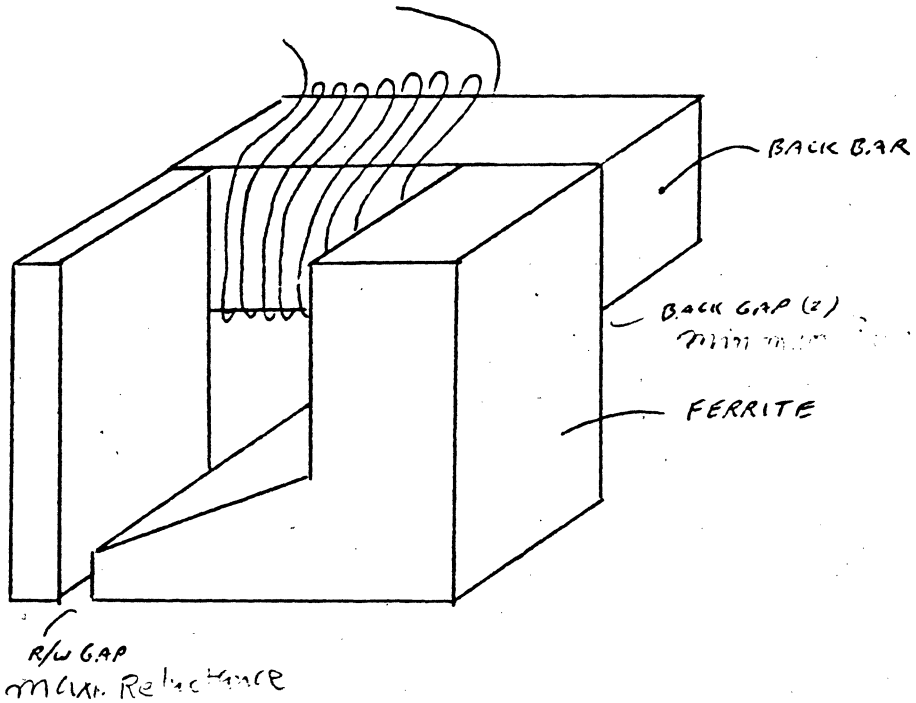


FIG 1.3 FERRITE STRUCTURE

IBM 2314, 3330

and was therefore used extensively for the next twenty years until grain size became comparable to trackwidths.

IBM announced the Winchester head in its 3340 product in 1974. Its structure permitted lower flying heights with less mass and therefore a lower loading force with less energy content on contact. Its structure is shown in Fig. 1.5. A small C structure is bonded to the face to provide the gap and the coil winding window. The two outside rails, A, B, constitute the air bearing surface, replacing the large ceramic or barium titanate sliders used in the earlier high mass heads. The center rail carries the head C core and is machined to the width of the track. There are variations of this slider form using only two rails that carry two head C cores or two thin film heads. Sometimes this structure has a machined cavity that produces a low pressure area. This low pressure area is balanced against the high pressure area under the rails to make a self loading slider that does not require an external load force. The earliest heads required an air supply to establish the air bearing required to maintain head-disc separation. The development of a self lubricated slider removed the requirement for a pressurized air supply. These heads were loaded onto a spinning disc through a cam arrangement with a load of 350 grams. The heads required removal before the disc was stopped. With the introduction of the Winchester head, the head load and mass were low enough to permit contact start and stop, thus permitting a sealed environment. A comparison of the two types of air bearing is shown in Fig. 1.6. The dimensions are exaggerated in order to show the principle.

Head-1 Head-2

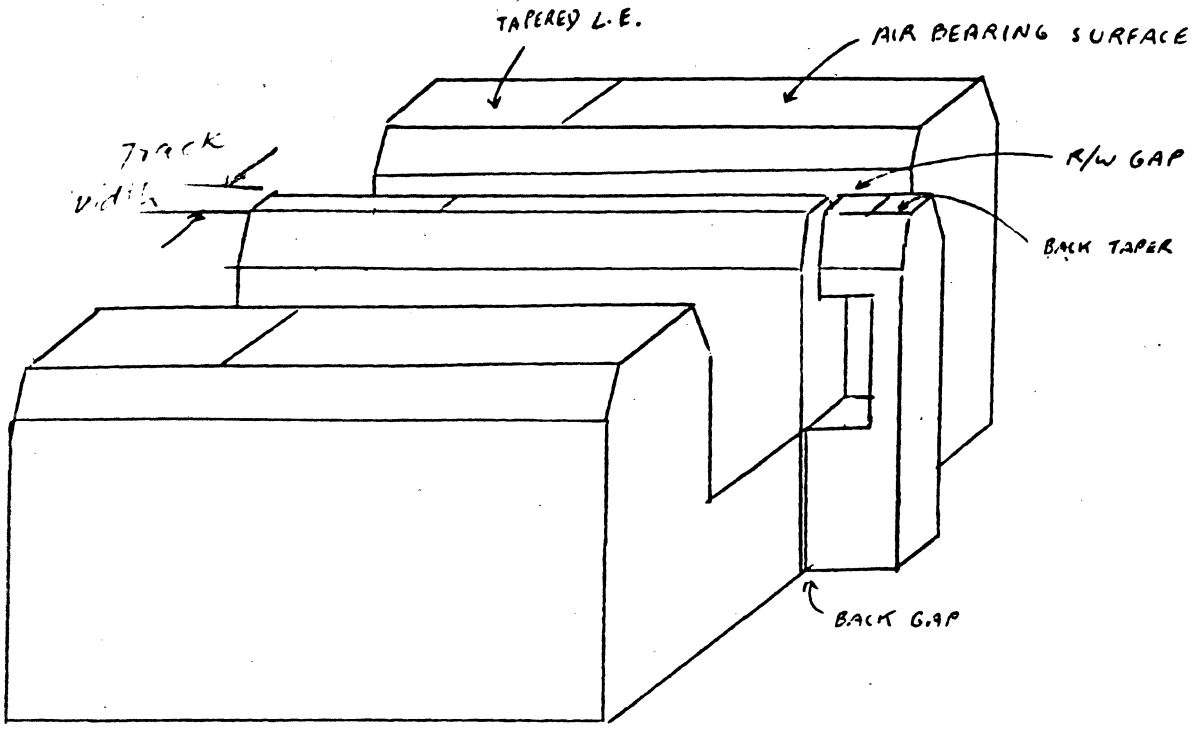


FIG 1.5 WINCHESTER HEAD (UPSIDE DOWN)

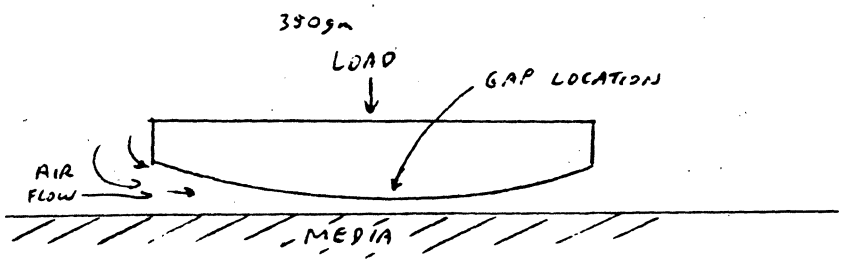


FIG 1.6 A

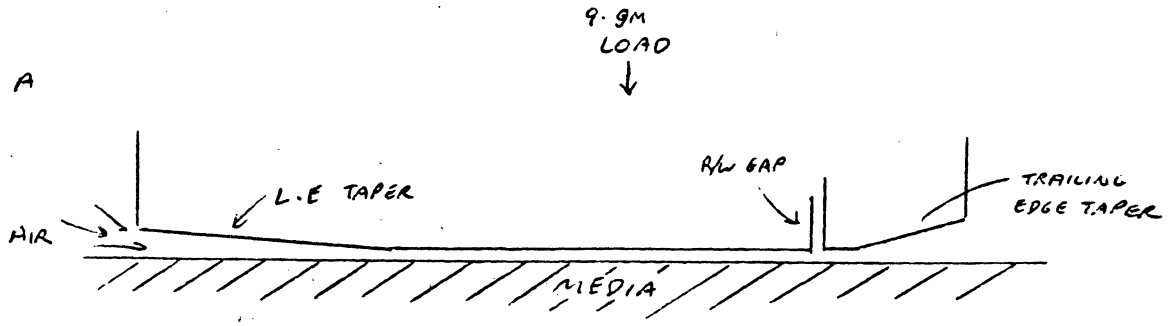


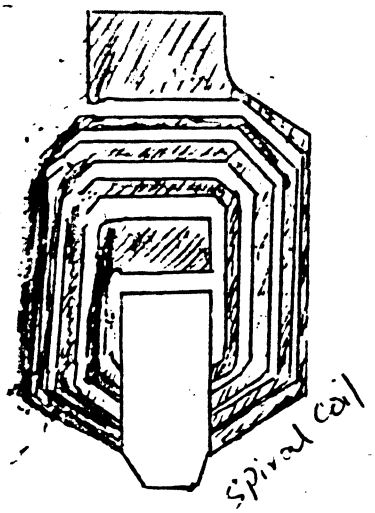
FIG 1.6 B

The next head type used is the thin film head, so named because it is manufactured using thin film techniques. Here the various parts of a head are deposited as films of magnetically permeable materials such as Permalloy, conductors such as copper or aluminum, and various insulators. The precision of photomasking techniques permit precise trackwidth control to dimensions down to the sub micron level. The structure of the thin film head is shown in Fig. 1.7 A and B. The actual shape of the various etched deposits varies with design.

The return to a Permalloy core structure is permitted because the core losses are greatly reduced. The very thin films permissible by the technique reduce these losses significantly.

In tape drives the core material remained Permalloy for a long time. This was due to the relatively low tape velocity compared to discs. Recently they have moved to ferrite to improve frequency response and head wear. Their structure is not unlike that previously given, except that multiple heads are sandwiched together to provide the required number of parallel tracks simultaneously used. The tape is held in contact by the use of pressure pads and guides. Some heads have two cores per track: One specifically for writing which has a wide trackwidth and a wide gap length; The second head follows the the Write head in tape direction and is constructed with a narrower trackwidth and a narrower gap length. This is done to reduce off track positioning errors and to improve the Read frequency response.

Disc heads must, of necessity, be a compromise in gap length, as they are used for both reading and writing. Fig. 1.9 shows why only one head is used.



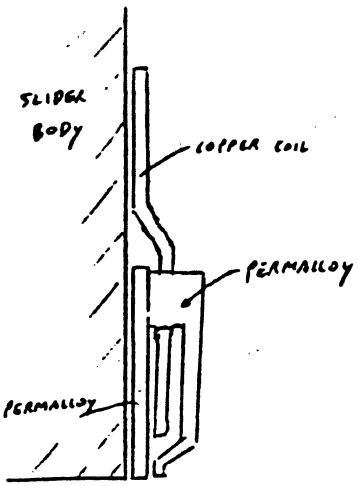


FIG 1.7A FIG 1.7B

THIN FILM HEAD

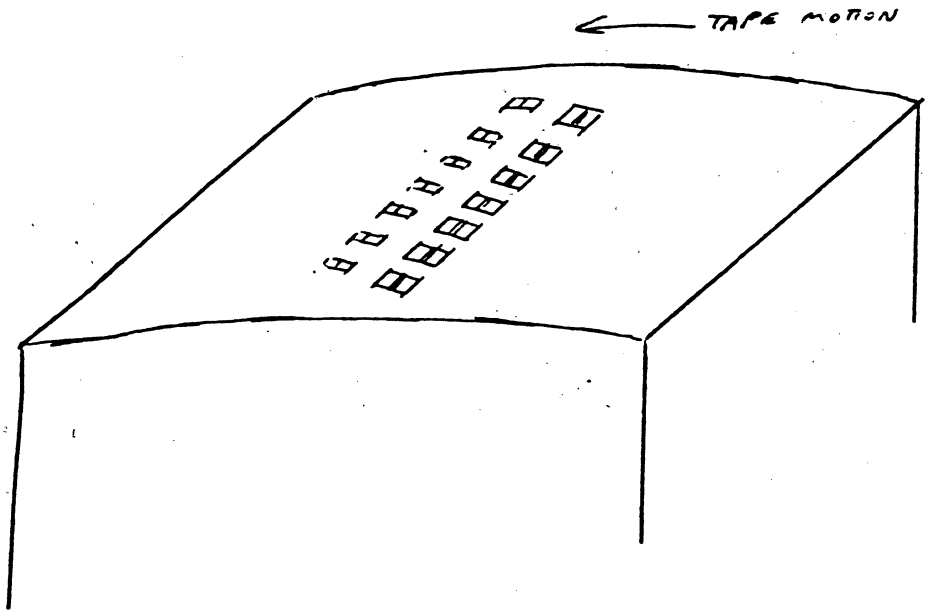


FIG 1.8 TAPE HEAD MULTI TRACK

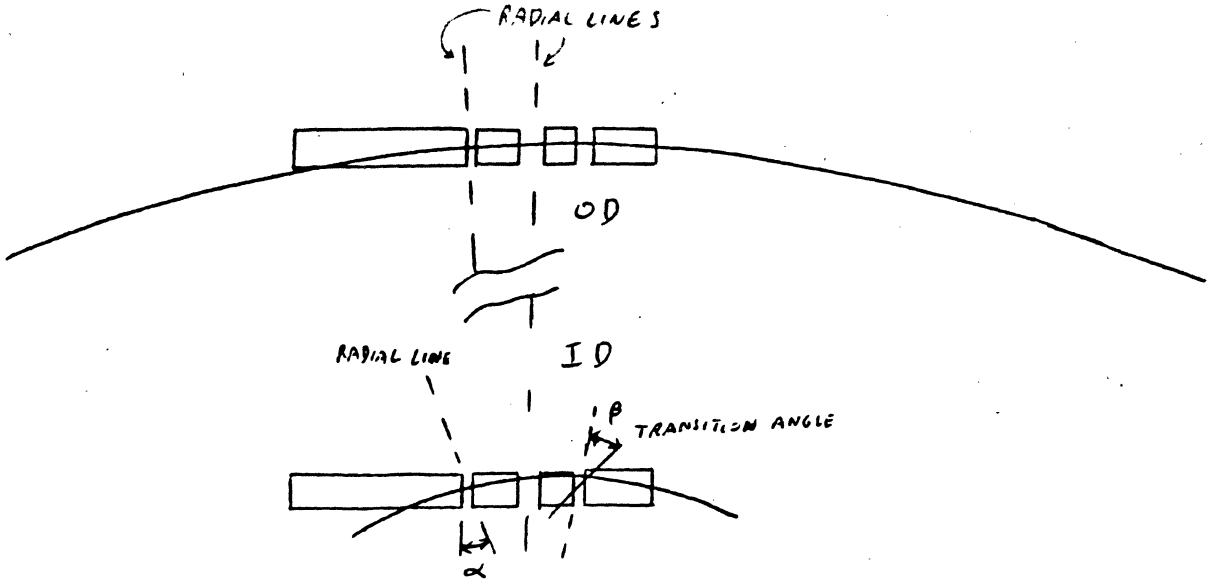


FIG 1.9A DISK HEAD

EXAGURATED DIMENSIONS SHOWING ERRORS IN HEAD POSITIONING AS A FUNCTION OF RADIUS AND TRANSITION ANGLE BETWEEN READ-WRITE.

RECORDING ELECTRONICS

DISC/TAPE STRUCTURE

Magnetic tapes have long been manufactured using a backing material usually of plastic, but earlier tapes used paper. Today Mylar is extensively used, as it stretches or deforms less. The magnetic material is gamma ferric oxide imbedded in a binder and coated onto the surface of the backing uniformly. Calendaring, a Memorex invention, was later used to reduce the surface roughness and hence head wear.

Discs are made from an aluminum alloy blank stamped from sheet stock of high purity. The blank is then polished such that its flatness is controlled within microinches. A mirror finish to within a light band is the result. This disc is coated with a slurry of gamma ferric oxide and a suitable binder.

Down through the years this coating has become thinner and thinner, going from about 1 mil to 35μ" in twenty years. Changes in formulation have occurred to improve coating hardness, uniformity, coercivity, particle size, particle dispersion, and adhesion.

Gamma ferric oxide has been used extensively due to its fairly square hysteresis curve. This curve relates the B and H fields as functions of the intensity (see Fig. 2.1). It is noted that the permeability of the oxide changes both from field intensity and from past history. What makes the particle so useful is the ease of saturation and the retained B field, Br. This is the chacteristic that permits recording. In saturation recording the coating is saturated first in one direction and then in the other as a

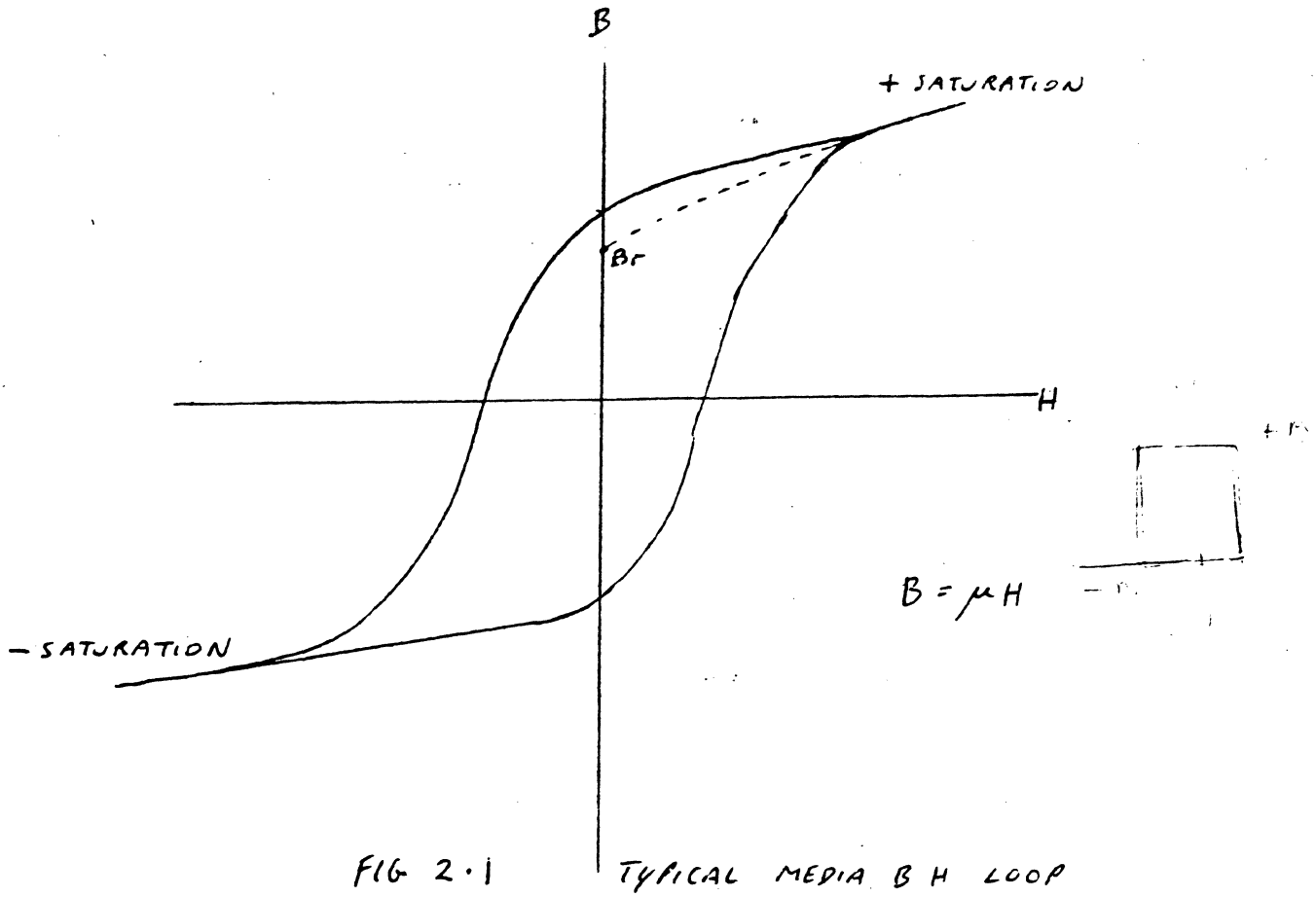


FIG 2.1 TYPICAL MEDIA B H LOOP

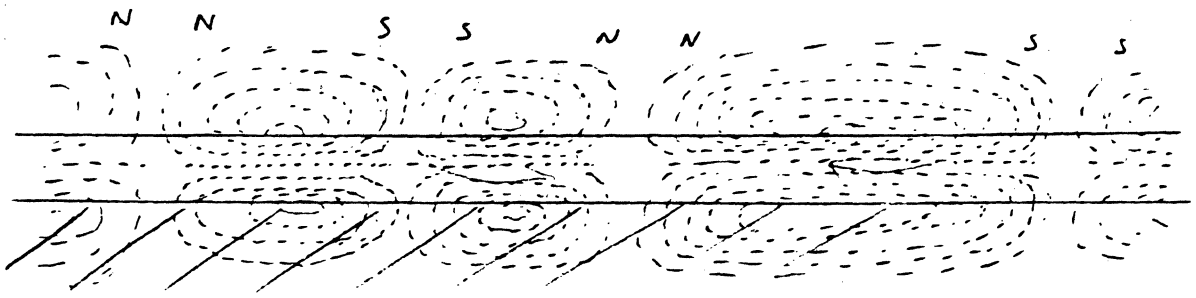


FIG 2.2 TAPE - DISC MAGNETIZATION FIELDS

function of the data to be recorded. The spacing between flux reversals determines linear data density. The linear distance is divided into cells to which is assigned a bit of known value, thus on play-back (read) each bit is reproduced in its correct cell and the data is recovered. A typical disc or tape magnetization pattern is illustrated in Fig. 2.2. The cross section is taken longitudinally along the track. The location of the N,-N juxtaposition or S,-S juxtaposition is referred to as a transition, the center line being the exact location of the transition. Since this is the moment at which a moving head sees a maximum time rate of change in flux, a voltage is developed in the coils surrounding the core as the core gathers the flux, due to its higher permeability than the surrounding air. This simplistic explanation will suffice for here. More precise development of the theory is given in the literature.

The surface of discs is polished to the desired flatness in order to minimize the head-disc spacing variations. Any variation in flatness is seen by the head as an up and down motion as the disc rotates, which excites the mass-spring mechanics of the head, causing further head-disc separation and possible contact on the negative excursion. Contact has been a problem with the high load, high mass head, as the disc is damaged extensively due to the energy of contact. Particles are removed which further contaminates the air stream under the head, which causes further disturbances and further contact. The final effect is called a crash. Crashes have essentially been eliminated with the Winchester style slider. Some disc manufacturers deliberately add alumina particles to the coating slurry in order to force a contacting head

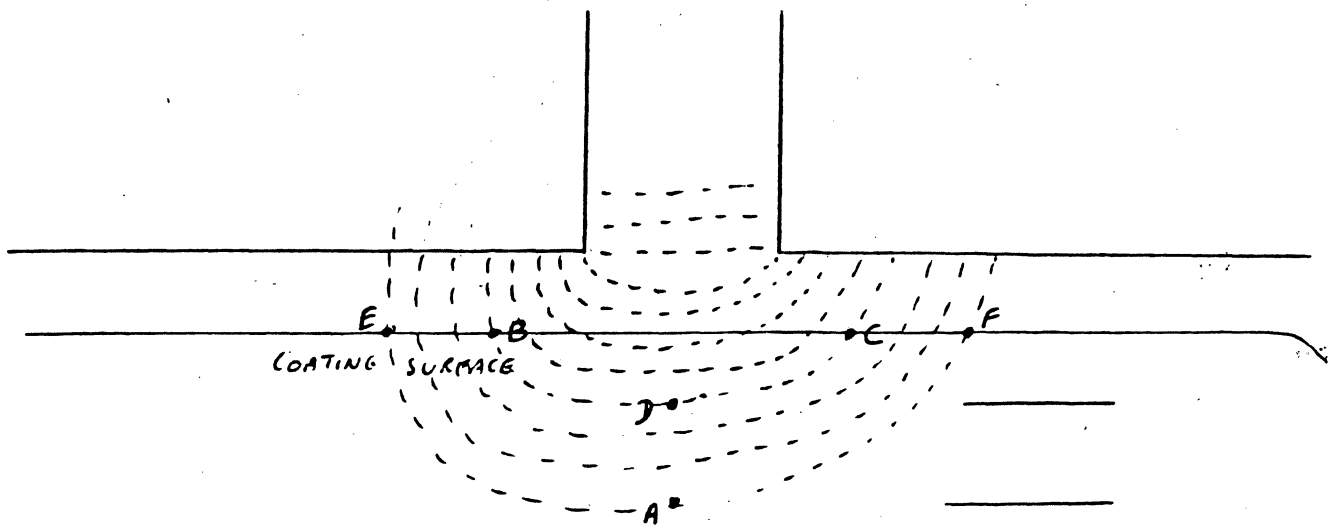


FIG 2.3 FIELD INTENSITY AND DEPTH

to rebound from the hard particle. A problem with the alumina is that it is non-magnetic and therefore represents a magnetic discontinuity which is read by the head as a noise voltage. Size control is required in order to keep the top of the particle below the expected position of the head. Contact of the Winchester head is deliberate during start-stop operations. The disc coating is given a thin coating of a fluorocarbon in order to improve its wearability without causing stiction.

The coating thickness influences the spacing between transitions. Hence as the data density has increased, so the coating thickness has reduced. This effect is easily seen when one considers that the field required for saturation must emanate from the head gap which reduces as the inverse square of the distance from the gap. The further the field must penetrate, the larger the initial field; therefore the wider the field lines. If point D on Fig. 2.3 is 300 Oe or saturation value, then the particle at A is not saturated. But if A is 300 Oe, then D is much higher and its influence extends to E and F, thus widening the field or reducing the obtainable density. The height of the head is above the media and has the same effect of reducing the potential transition density.

RECORDING ELECTRONICS

HEAD CIRCUIT

In order to write a transition, current must be passed through the coils of the head windings first in one direction and then in the other in time with the imaginary cells assigned to each bit recorded on the moving disc or tape. To see the effects of such current reversal, we need to develop an equivalent circuit for the head. We expect it to include resistance due to the conductivity of the wire used. It must have inductance due to the turns and the core structure materials. we would also expect interwinding and wiring capacitance. See Fig. 3.1. This then becomes a simple RLC circuit, as illustrated in Fig. 3.2A. The equations for a step current in LaPlace form concern the voltage developed across the head windings as well as the current through the head windings.

$$V_{(s)} = \frac{I_{(s)}}{S} Z_{h_{(s)}} \tag{3·1}$$

$$= \frac{I_{(s)}}{S} \left(\frac{\frac{1}{CS}(LS+R)}{\frac{1}{CS} + LS+R} \right) \tag{3·2}$$

$$= I_{(s)} \left(\frac{S + \frac{R}{L}}{SC(S^2+\frac{R}{L}S + \frac{1}{LC})} \right) \tag{3·3}$$

This can be rewritten as (3·4) which is the standard form:

$$I_{(s)} \left(\frac{S + 2\zeta W_n}{SC(S^2+2\zeta W_n S + W_n{}^2)} \right) \tag{3·4}$$

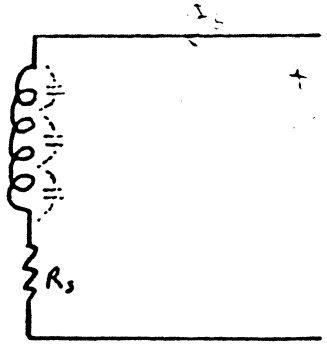


FIG 3.1

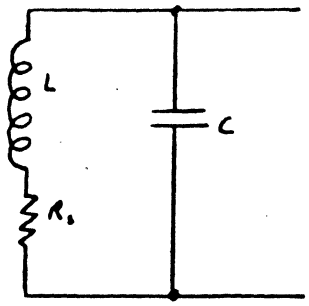


FIG 3.2A

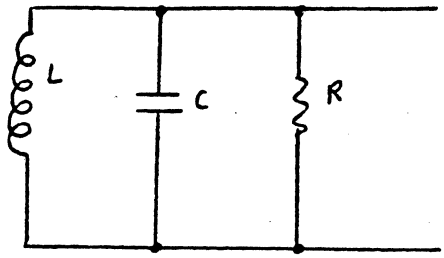


FIG 3.2B

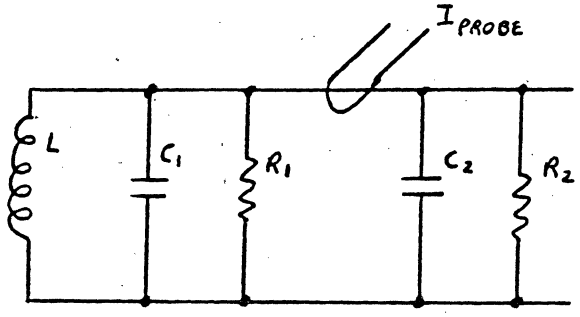


FIG 3.2C

HEAD CIRCUIT

The current through the head winding which produces the flux is simply the voltage divided by the R and L of the head. This is not exactly true, since the interwinding capacitance plays a role in the true current, but it is sufficiently accurate for our purposes.

$$I_{(s)Head} = \frac{V_{(s)}}{R+LS} = \frac{I_{(s)}}{S}\left(\frac{\frac{1}{CS}(LS+R)}{(\frac{1}{CS}+LS+R)(LS+R)}\right) \quad (3\cdot5)$$

$$= \frac{I_{(s)}}{S}\left(\frac{\frac{1}{CS}}{\frac{1}{CS}+LS+R}\right) \quad (3\cdot6)$$

$$= \frac{I_{(s)}}{S}\left(\frac{\frac{1}{LC}}{S^2+\frac{R}{L}S+\frac{1}{LC}}\right) \quad (3\cdot7)$$

which, when written in the standard form, becomes Eq.(3·8):

$$\frac{I_{(s)}}{S}\left(\frac{W_n^2}{S^2+2\zeta W_n S+W_n^2}\right) \quad (3\cdot8)$$

As we examine these equations we see the terms $2\zeta W_n$ and W_n^2 are identical for both the voltage and the current. The R value is small, being typically only a few ohms for low winding heads. The damping factor, ζ, calculated from the algebraic equation

$$\frac{R}{L} = 2\zeta W_n \quad \text{or} \quad \zeta = \frac{R}{2LW_n} \quad (3\cdot9)$$

would be small, indicating that both the voltage and current will be an exponentially damped sinusoid instead of a modified square wave, as we would wish. In order to do this, we must add a resistance either in series or in parallel.

HEAD CIRCUIT

The equation becomes (from Fig. 3.2B):

$$V = \frac{I_s}{S} Z_{(s)} = \frac{I_{(s)}}{S} \left[\frac{\frac{R(\frac{1}{CS})}{R + \frac{1}{CS}} \quad LS}{\frac{R(\frac{1}{CS})}{R + \frac{1}{CS}} + LS} \right] \tag{3.10}$$

which reduces to Equation 3.11:

$$V_{(s)} = \frac{I_{(s)}}{C(S^2 + \frac{S}{RC} + \frac{1}{LC})} \tag{3.11}$$

when written in the standard form it becomes Equation 3.12:

$$V_{(s)} = \frac{I_{(s)}}{C(S^2 + 2\zeta W_n S + W_n^2)} \tag{3.12}$$

Similarly we develop the current equations as before:

$$I_{(s)} = \frac{V_{(s)}}{LS} = \frac{I_{(s)}}{LCS(S^2 + 2\zeta W_n + W_n^2)} \tag{3.13}$$

$$= \frac{I_{(s)} \; W_n^2}{S(S^2 + 2\zeta W_n S + W_n^2)} \tag{3.14}$$

We need both equations 3.12 and 3.14 as they describe the voltage swing across the head during a a write and the current wave form. With R properly chosen to make ζ = 1.0 for no overshoot we obtain the case of no ringing in either voltage or current. Practice shows that a ζ of .95 is best as it improves

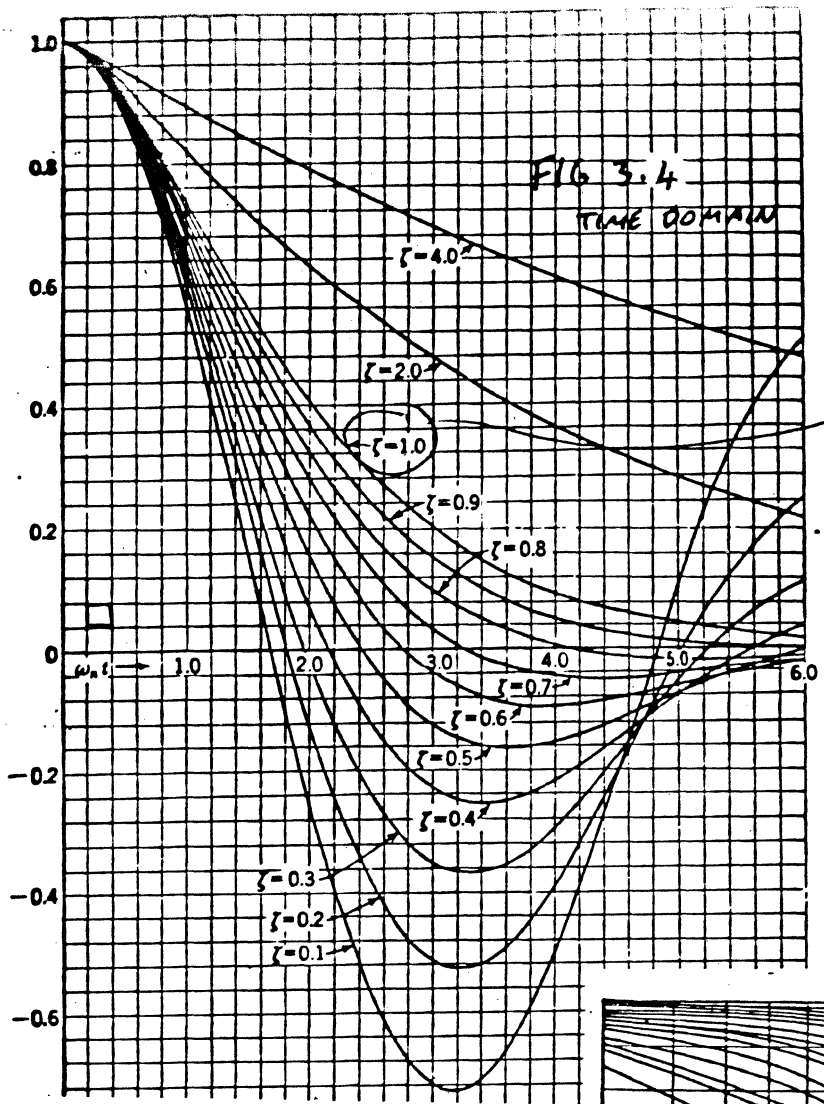


FIG 3.4
TIME DOMAIN

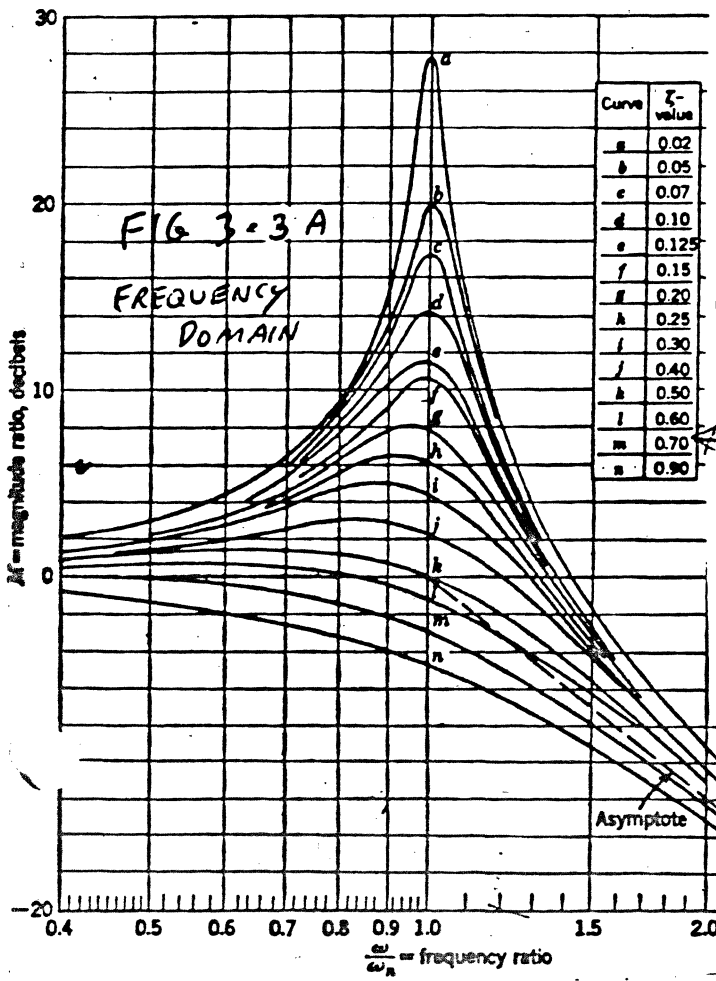


FIG 3.3 A
FREQUENCY DOMAIN

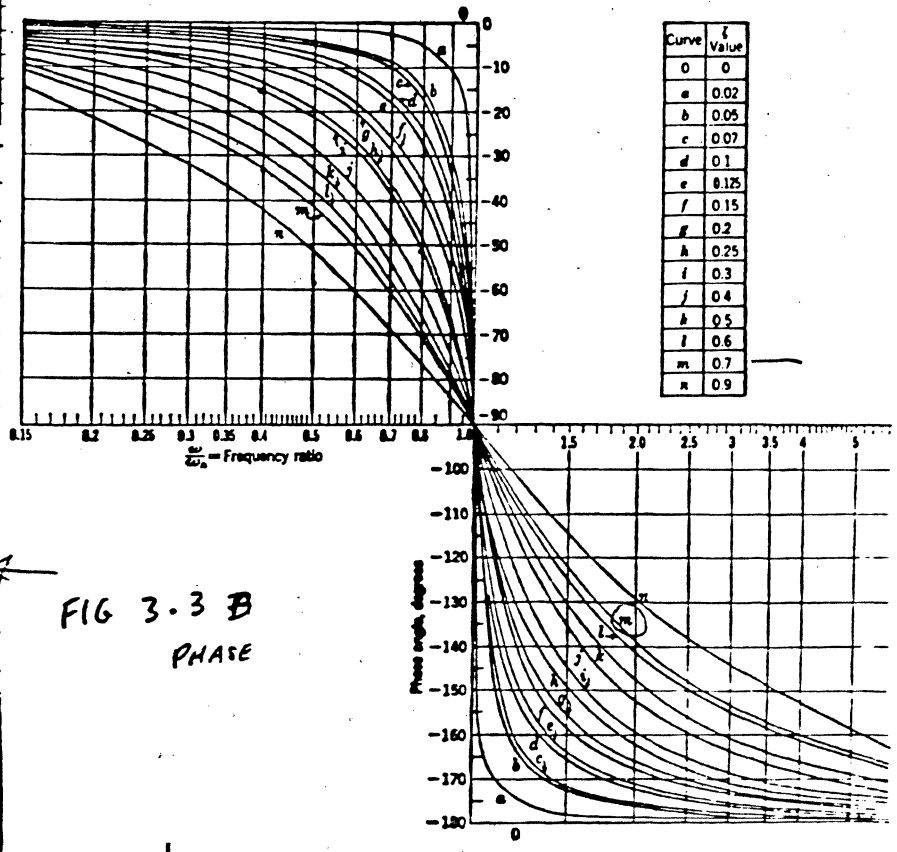


FIG 3.3 B
PHASE

the rise time with minimum ringing. The actual overshoot is about 3-5% which is acceptable. When the current is measured using a current probe, the ideal waveform is not seen. To see why, we must relook at the equivalent circuit. FIG 3.2C We see a capacitor and a resistor on both sides of the current probe. The equation can be modified and does reflect the true waveform.

$$I_{Probe(s)} = \frac{I_s}{S}\left(\frac{R_p^2}{R_1^2}\right)\left(\frac{S^2 + 2\zeta_p W_{np} S + W_{np}^2}{S^2 + 2\zeta_1 W_{n1} S + W_{n1}^2}\right)\left(\frac{W_{n1}^2}{W_{np}^2}\right)$$

Where R_p, W_{np}, and ζ_p are the parallel equivalents and the terms with the subscript 1 are those on the head side of the probe.

For reading the damping must be adjusted for a $\zeta = 0.7$. The reason for this is that for current we are talking about a time domain response and for reading we are talking about frequency domain response. See Figures 3.3 and 3.4.

Refering to the current's time domain response and the hysterisis curves for the media as shown in Figures 3.6 and 3.5 respectively, we can see the magnetic effect of ringing of the write current. The overshoot A, causes the media to be pushed further into saturation while the undershoot B, brings the media back out of saturation. This is undesirable.

Since the write current cannot change instantaneously, there is a period of time during which the media sees less than a saturating field. If the rise time of the write current is short compared to the time a media particle travels from one edge of the head gap to the other, then that particle is assured of leaving the influence of the gap saturated in the new direction. From this we can see that the trailing edge of the head gap exerts the final influence on the media.

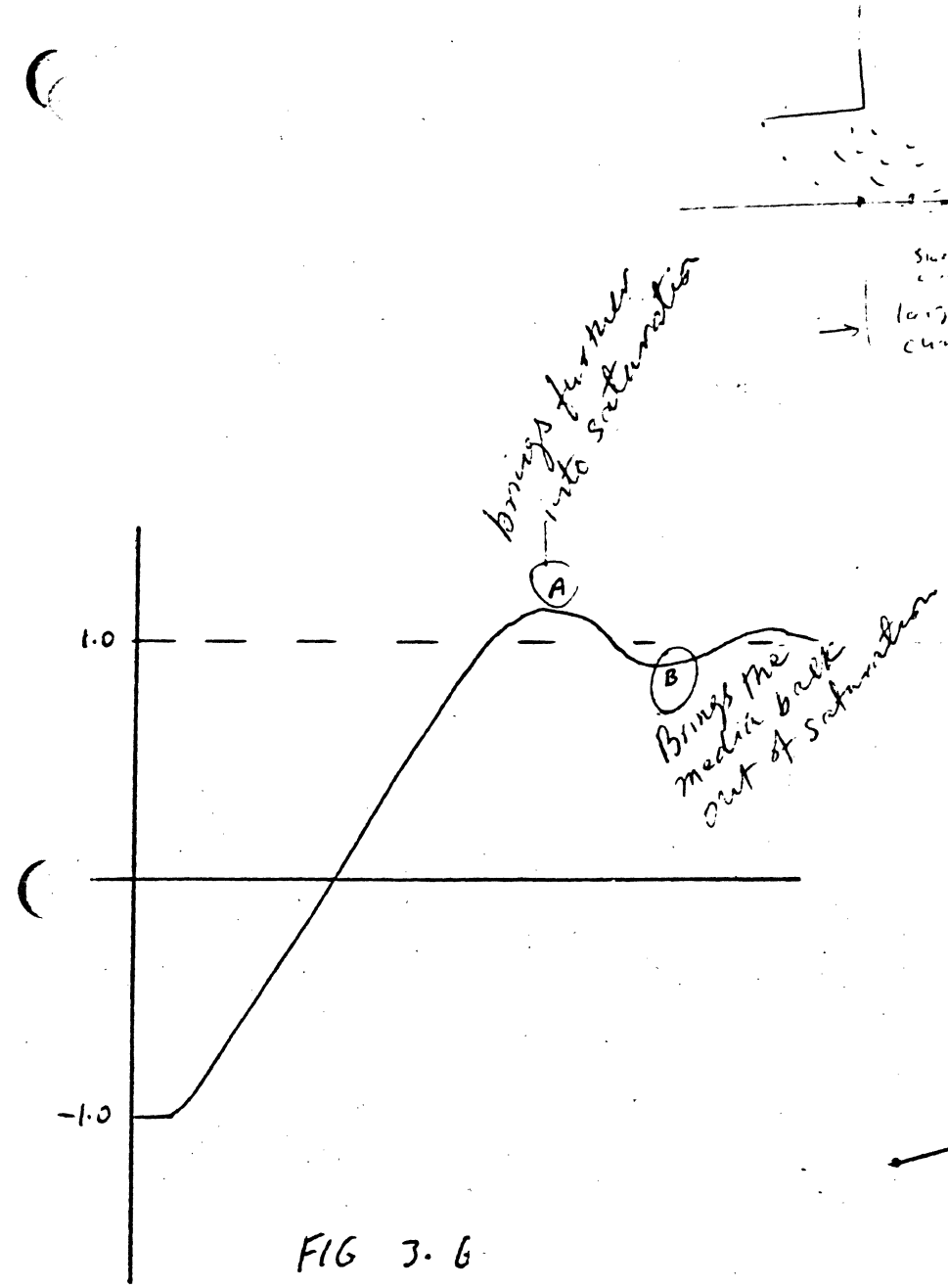


FIG 3.6

HEAD CURRENT WAVEFORM

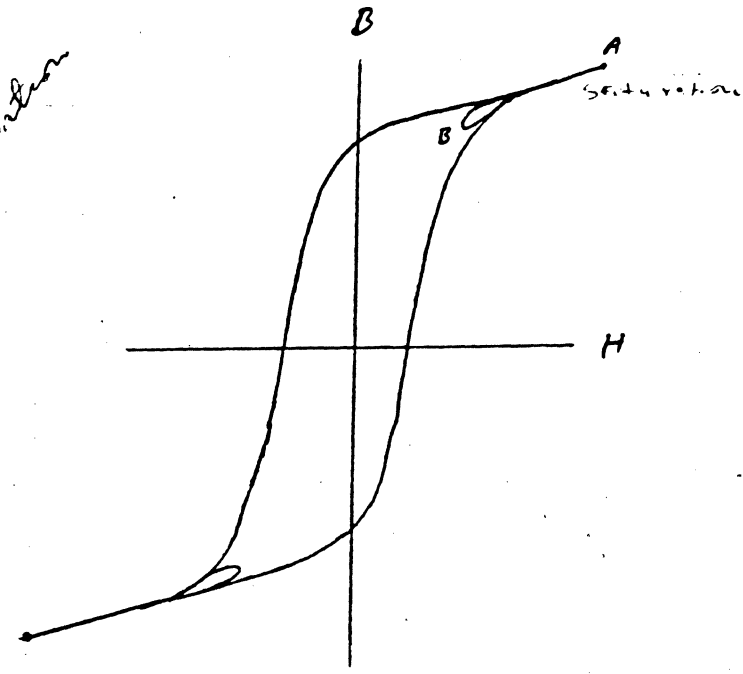


FIG 3.5

MEDIA HYSTERISIS CURVE

HEAD CIRCUIT

Figure 3.7 shows the positional relationship of a particle of oxide as it travels within the gap and the field strength it sees at each position. Clearly current curve 1 takes the particle from -M to +M within the distance of the gap travel; whereas, with current curve 2 the particle is well outside the gap before +M level is reached at the trailing edge. This indicates that the particle will not be saturated and will therefore retain old information. The saturation is not quite this bad as the write current is usually greater than required for saturation. Similarly a particle at the gap center at the start of the transaction remains saturated at -M as the field when crossing the trailing edge is nearly zero.

Magnetically the head circuit can be described by a reluctance diagram. (Fig. 3.8). In the construction of the head these reluctances must be considered. The core leg and back gap reluctances total must be small compared to the front or working gap.

When writing the front gap should be wide in order to assure complete saturation during current rise time. Its reluctance will therefore be greatest as desired. However, in its construction the core area is considerably reduced at the throat in order to maximize the external field as shown in cross section in Fig. 3.9. The reluctance which is a function of cross section will be increased, hence the field strength in the area is increased thereby creating the possibility of pole tip saturation. Pole tip saturation effectively widens the gap as that portion saturated has a μ of 1 like air.

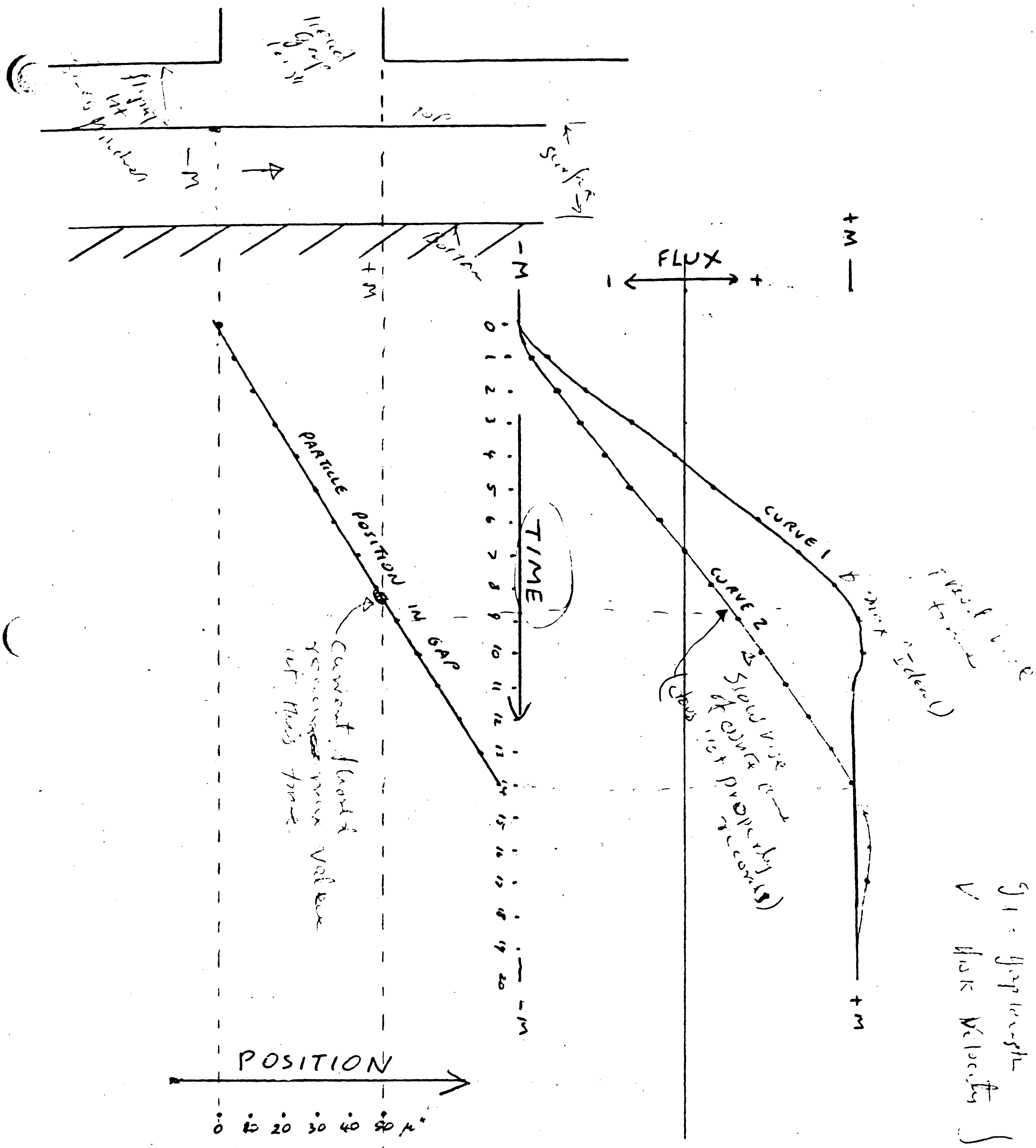

FLUX
TIME
PARTICLE POSITION IN GAP
CURVE 1
CURVE 2
POSITION
-M
+M
0 10 20 30 40 50 μ

FIG 3.7

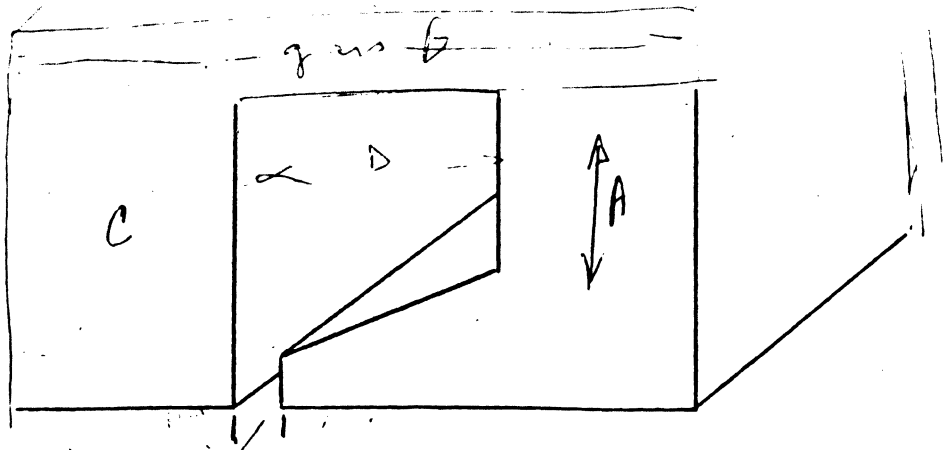
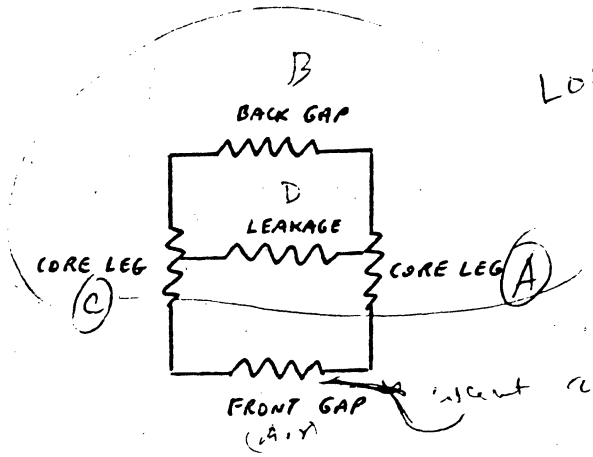


FIG 3.9 CORE AREA REDUCTION AT GAP

front



LOSSES

FIG 3.8 Reluctance Circuit

In summary, the field strength seen by the media depends on the current value, the ratio of reluctances, the flying height or spacing, and the coating thickness. The design of the head must therefore accommodate all these when attempting to maximise the lineal transition density. Also the head inductance increases with the square of the turns, whereas the output voltage only increases as a direct function of turns. Trying to compromise output and rise time becomes difficult because of the inductance.

RECORDING ELECTRONICS

HEAD PERFORMANCE

In inductive heads which are the only ones considered so far, the read back performance of the head is directly releated to the velocity of the recorded transition, the number of turns on the core, and the efficiency of the flux gathering paths. The instantaneous read back voltage is then proportional to $KN \frac{d\Phi}{dt}$. The flux resulting from a transition is complex having field lines changing in slope from some positive value to some negative value or visa versa over some distance. The work of Karlquist and Hoagland's studies have provided the basis for these interactions with considerable work done by others following. It is not the purpose here to detail the derivations, but we will use their results.

KARLQUIST

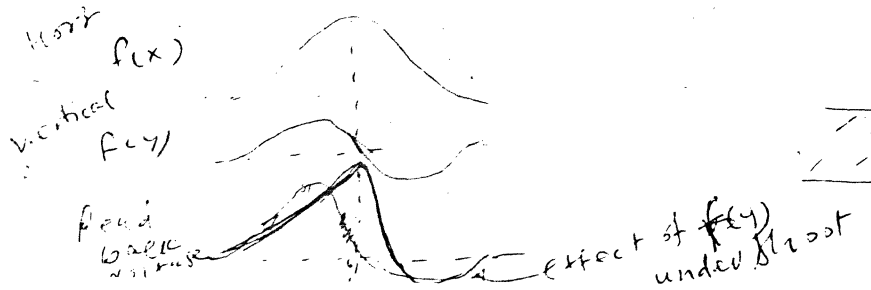
$$Hx\ (x,y) = \frac{1}{\pi g}\left[\tan^{-1}\left(\frac{g/2 + x}{y}\right) + \tan^{-1}\left(\frac{g/2 - x}{y}\right)\right] \qquad (4.1)$$

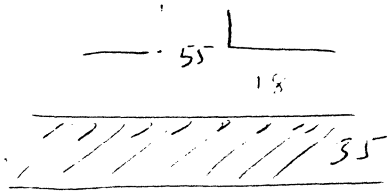
$$Hy\ (x,y) = \frac{-1}{2\pi g}\ln\left(\frac{(g/2 + x)^2 + y^2}{(g/2 - x)^2 + y^2}\right) \qquad (4.2)$$

Many thought this was true for small particle/gap but not true

These two equations show that there is both a horizontal x component as well as a y component of flux. Where g is the gap length, x and y are the component vectors.

Most authors have neglected the y component for simplification by assuming a thin media; however, there are features of the read back pulse that can only be predicted by using the y component.





HEAD PERFORMANCE

The idealized thin media pulse is given by:

$$e(\bar{x}) \tilde{=} K^{1}\int M^{1}_{x}(x - \bar{x})\, Hx(x)dx = (M^{1}x * Hx)\,\bar{x}$$ where * is the convolution.

There are several other derivations that should be looked at besides the arctangent equations. Others have used the Gausian, Lorentzian and modified Lorentzian versions. We will use the results of their work here, but will not go into the magnetics nor derive the equations. Our purposes will be filled as we understand the effect of the various parameters of the head on the read back and writing process.

As expected the center of the transition is the point of the maximum time rate of change of the recorded flux; therefore, the read back voltage will be a maximum trailing off on either side. We will use the Gausian or bell shaped curve for understanding as shown in Figure 4.1. We refer to this pulse as an isolated pulse. Hoagland and others have shown that linear superposition holds for this pulse. Therefore as we record positive and negative transitions alternately on the disc the resulting waveform will be a train of positive and negative pulses of the general shape shown in Figure 4.1. As these pulses are crowded together we can use superposition in order to predict the resulting waveform or interaction.

In Figure 4.2A the peaks of the two pulses do not interface, but there is interference between them. The resultant waveform remains nearly the same in peak-to-peak amplitude, but does not return to the base line between them. In Figure 4.2B the spacing is closer. Here the pulses interact strongly, influencing both amplitude and peak position. Note the reduction in amplitude of the resultant peaks and also the shift in position of the peaks compared to the original. Since a train of data is time dependent as to its value in a data stream, this shift becomes significant. We refer to the shift as bit shift or peak shift

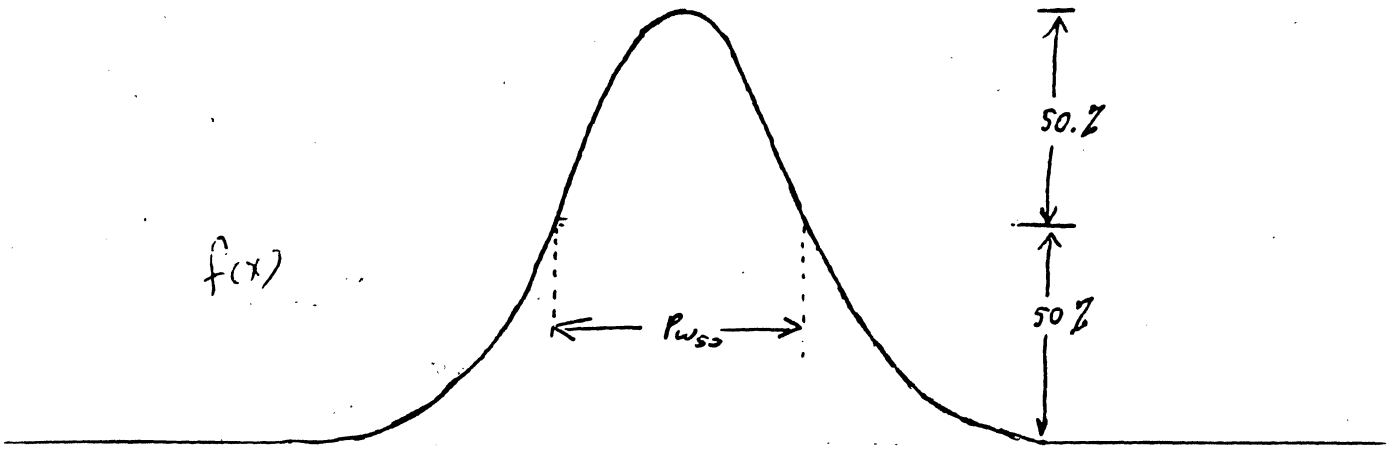


FIG 4.1 ISOLATED IDEAL PULSE.

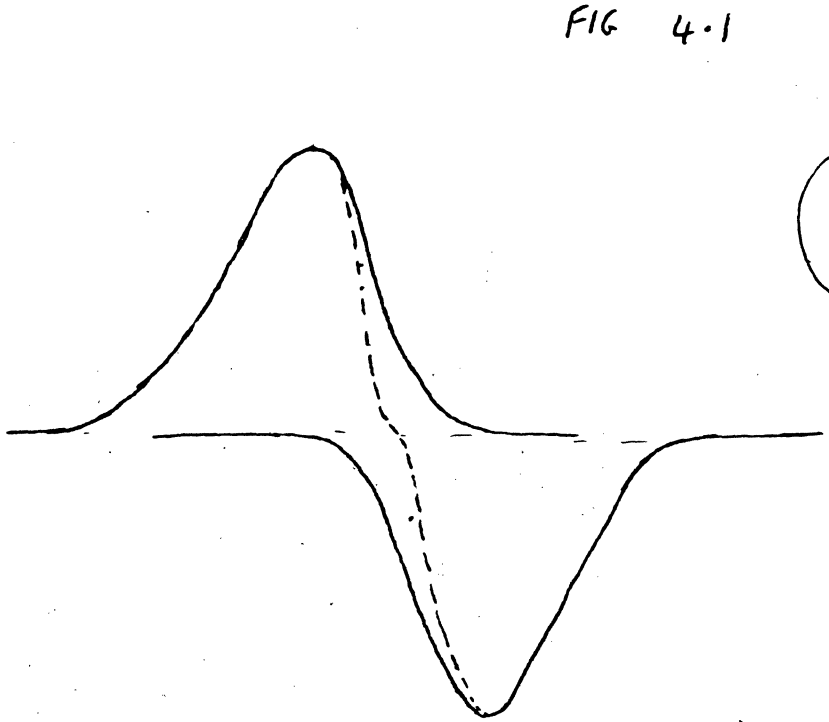


FIG 4.2 A

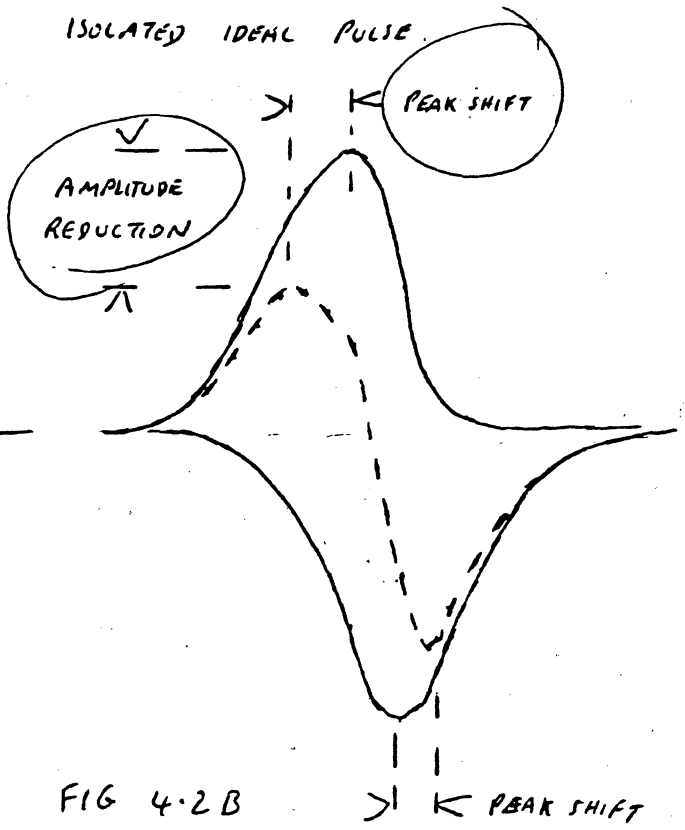


FIG 4.2 B

LINEAR SUPERPOSITION OR PULSE INTERACTION IN BOTH PULSE AMPLITUDE AND PULSE POSITION.

and it results strictly from pulse interaction.

If we were to test the peak amplitude of the read back waveform as a function of transition spacing or transition density, then we get what is called a transition or bit density curve. This is shown in Figure 4.3.

Each head and disc combination has its own curve depending on their many parameters. Bit shift or transition shift can also be similarly plotted on the same graph coordinates. The extension of the amplitude curve relates to the wavelength of the transition spacing and the gap length. If the gap field includes two transitions the net flux is zero, hence a maximum at B in Figure 4.3. The head disc parameters are gap length, throat height, flying height or spacing, media coating thickness, media coercitivity and remenance, and head core reluctance. Amplitude is affected by throat height, head spacing, coating thickness and remenance particularly in the flat or non-interacting portion of the curve. The point at which the roll off occurs is affected by gap length, flying height, coating thickness and media coercitivity.

From the above it can be seen that some parameters affect both amplitude and roll off. Generally speaking, if we want to increase transition density we need to fly closer, use thinner media of higher coercive force and use a narrow gap head. All this shows up in the equations for Pw_{50} or the ½ voltage pulse width of the isolated pulse as shown back in Figure 4.1.

There is an equation that has been derived to express the Pw_{50} in terms of distance.

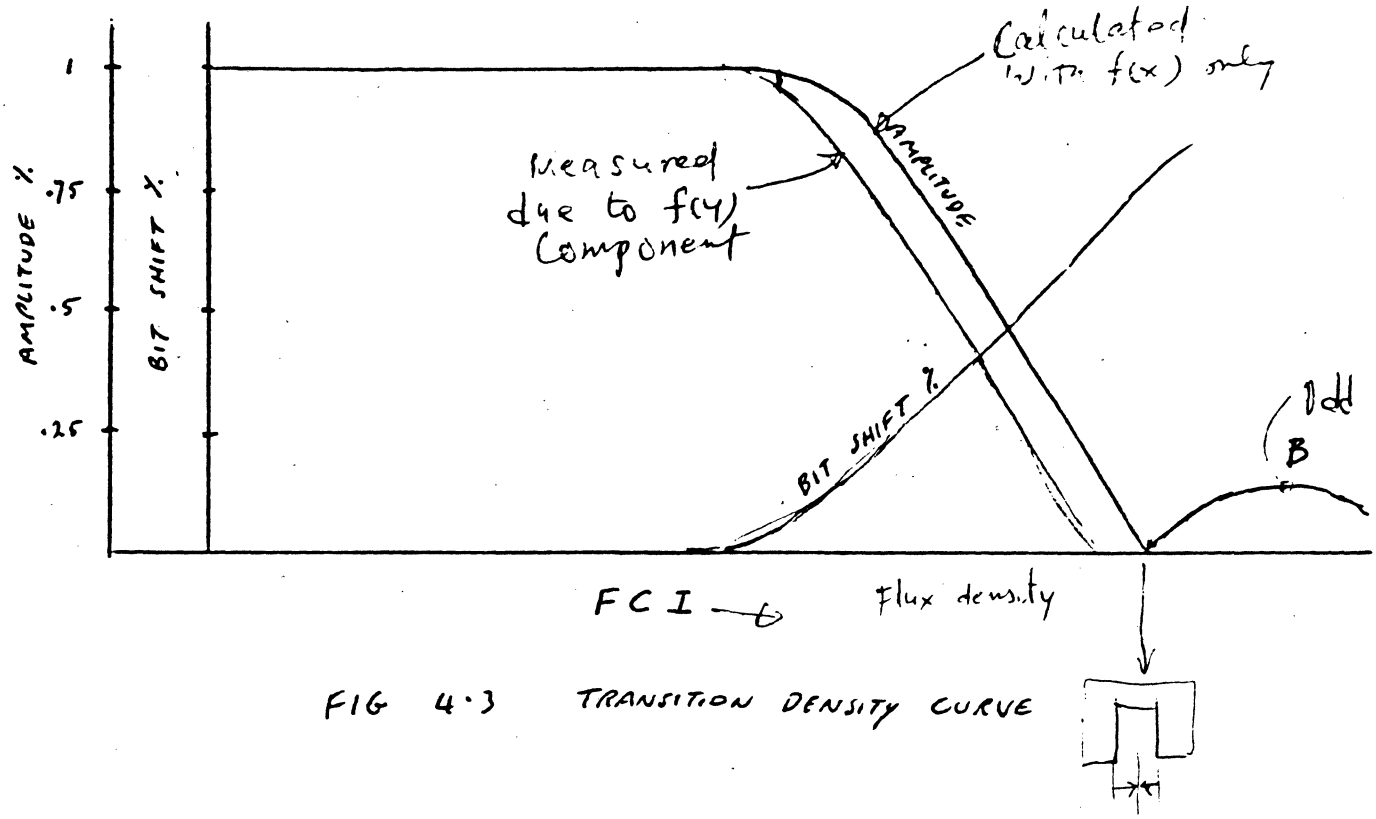


FIG 4.3 TRANSITION DENSITY CURVE

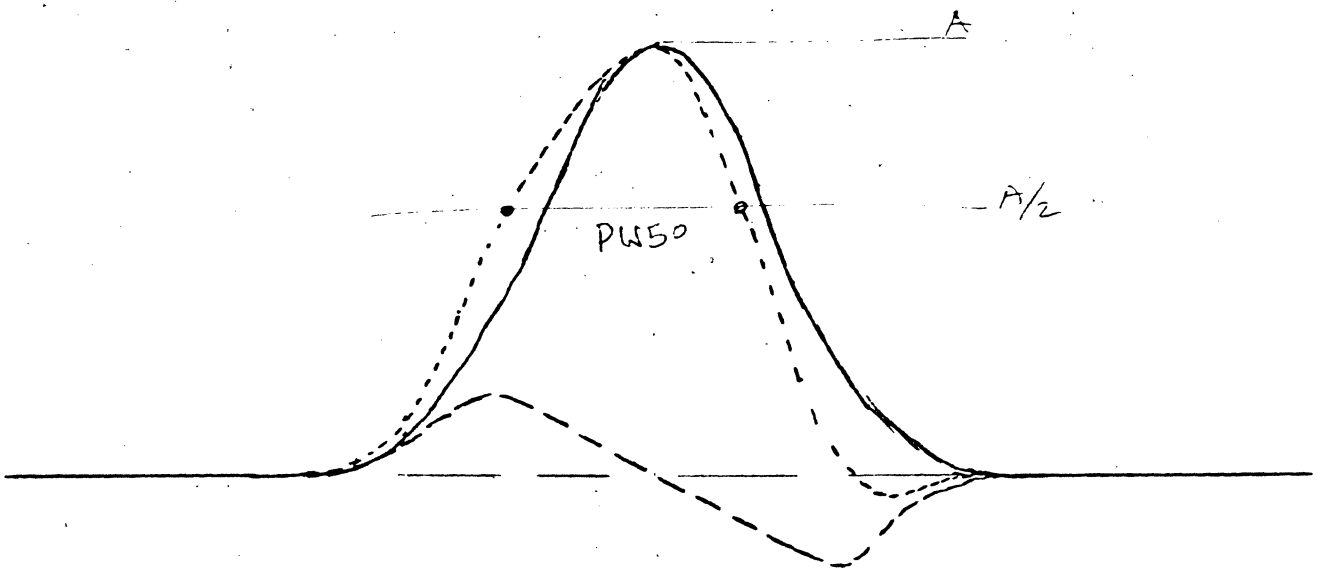
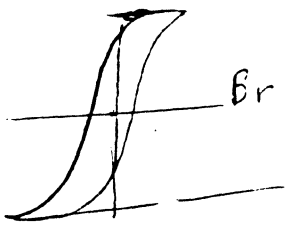


FIG 4.4 HEAD X AND Y COMPONENTS

HEAD PERFORMANCE

$$PW_{50} = \sqrt{g^2 + 4\,(d + a + \delta)(d + a)} \qquad (4.3)$$

Where g = gap length in μ"
d = head media separation in μ"
δ = media thickness in μ"
a = transition length in μ"



The transition length has been expressed as 'a' for N_1Z_n ferrite heads.

$$a = \frac{\delta}{2}\left(\frac{Br}{Hc \bullet Kd} - 1\right) \qquad (4.4)$$

No mention is made of the field spreading effects of finite rise time nor of the core reluctance and permeativity. Kd is an empirical number equal to 0.75 for particulate media and about 0.9 for thin metal films. The equation does not hold too well for MnZn ferrite heads. A possible explanation is that N1Zn heads usually have a magnetic dead layer therefore flying height is incorrect as is possibly the gap length. If it were perfectly annealed, the equation for 'a' might be in error due to Kd not counting the effect of finite rise time.

If we observed an isolated pulse on an oscilloscope, we would see a slight asymmetry and a trailing undershoot. Going back to the earlier Karlquist equation, we can see that there is predicted a y component. It is this y component that causes the asymmetry as illustrated in Figure 4.4.

This distortion must be considered when predicting bit shift and amplitude using superposition. It is presently done by entering points on the curve into a computer and having the computer do the work to generate the transition density curve. A general density curve can be drawn relating amplitude to Transition Spacing/PW50.

HEAD PERFORMANCE

SATURATION CURVE

If the amplitude of the read back signal were plotted as a function of the write current amplitude or given transition density, we get a new curve called a saturation curve. As the value of current is increased, we would expect the read back amplitude to increase as it would in a linear system. However, as we approach saturation in the media the amplitude levels off and remains steady for increasing amplitude. If the media is thick, the saturation curve rolls off instead of remaining flat with increasing current.

To understand why this is so, consider Hoagland's terminology of near field and far field. The near field is defined as the field within one gap length from the gap center as shown as point A in Figure 4.5. Point B is in what is called the far field.

It can be shown that for a head disc interface where the combination of flying height and coating thickness is equal to or less than the gap length the saturation curve remains essentially flat for increasing current provided the pole tip is not saturated. If the furthest particle of the media is further away from the gap than one gap length then the total effect is to broaden the transition width which reduces the amplitude the same as if the PW50 were increased which is exactly what happens. This was explained in Figure 2.3. The resultant saturation curve looks like that of Figure 4.6.

As expected from the transition density curve earlier discussed, the amplitude for higher transition densities is reduced by superposition. A saturation curve may be drawn for each density; therefore a typical saturation curve is a multiple curve showing at least the minimum and maximum density curves for the prepared recording system. Note that as in Figure 4.7

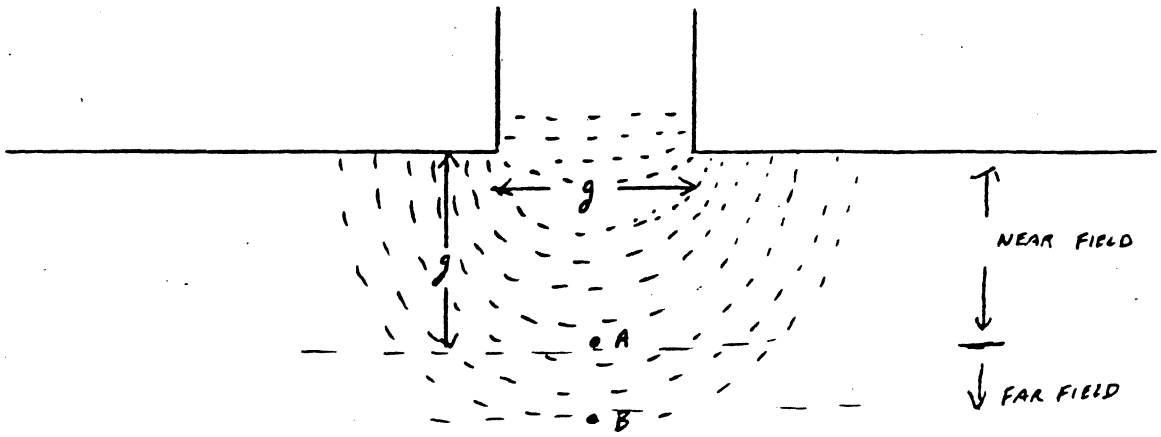


FIG 4.5 NEAR FIELD AND FAR FIELD

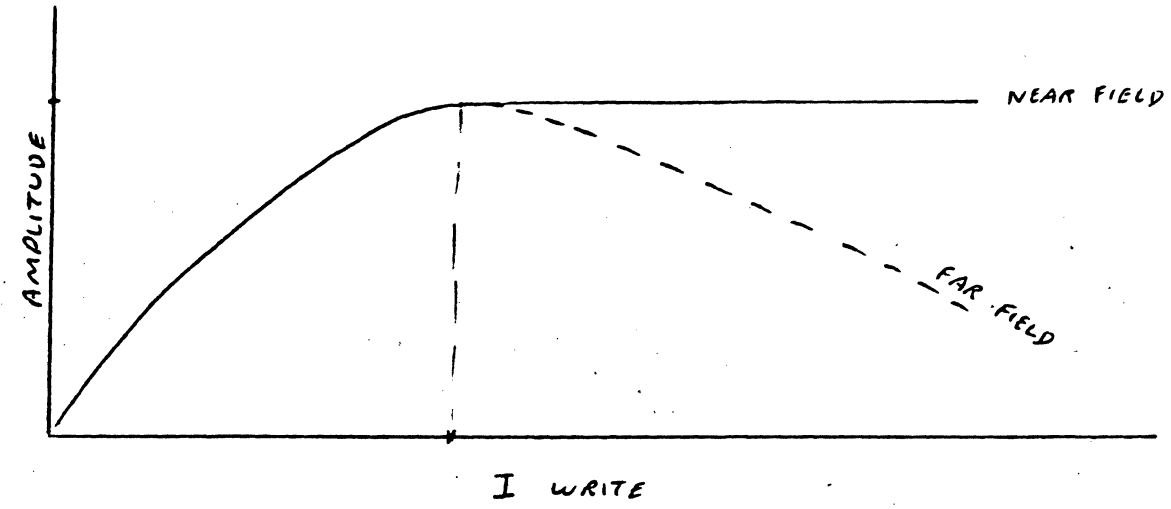


FIG 4.6 SINGLE DENSITY SATURATION CURVE

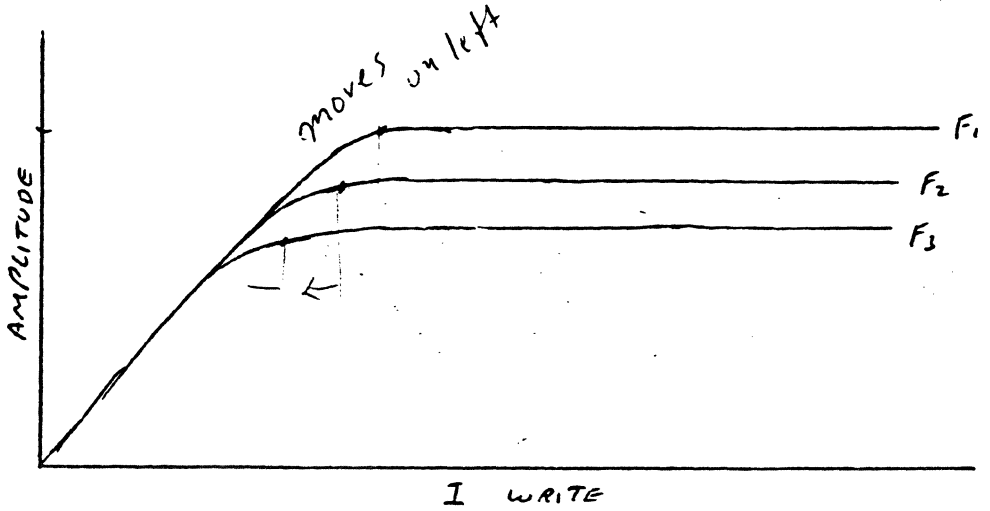


FIG 4.7 MULTIPLE DENSITY SATURATION CURVE

the current of saturation for each density is different indicating that saturation is also a function of transition density. The usual transition density curve can be taken at a single current value or it can be plotted using the minimum saturation current level for each transition density. To optimize a system it is profitable to choose the current value that best overwrites old information. It should also be noted that if the recording involves the far field, the slope of roll off increases with increasing density. This is shown in Figure 4.8. This roll off can be expected from the field spreading effect of the particle in the far field vs. the recorded wavelength. The correct write current must always be chosen to the right of the maximum for the lowest density to be recorded.

Since we noted that the so called saturation peaks occur at lower write current values for increasing transition density, we might expect the ability of writing higher transitions to erase a lower transition signal previously recorded to be diminished. Such is the case and results in a new curve called the write over curve. It is usually drawn on the same graph as the saturation curve, Figure 4.9. The curve data is taken by first writing the lower density signal and measuring its amplitude. This amplitude is called 0.db and becomes the reference. The higher density is then written over the lower density using the same value of write current. The residual low density signal amplitude is measured. This is done by using a high Q filter turned to the low density frequency in both cases. The high density signal is thus eliminated from the measurement. The ratio is taken as a -db level and is plotted on the graph. The resulting curve then indicates the degree of erasure and the quality of the recorded signal. As could be expected, any degradation of a signal affects the ability to read a transition and then assign it to its correct time slot.

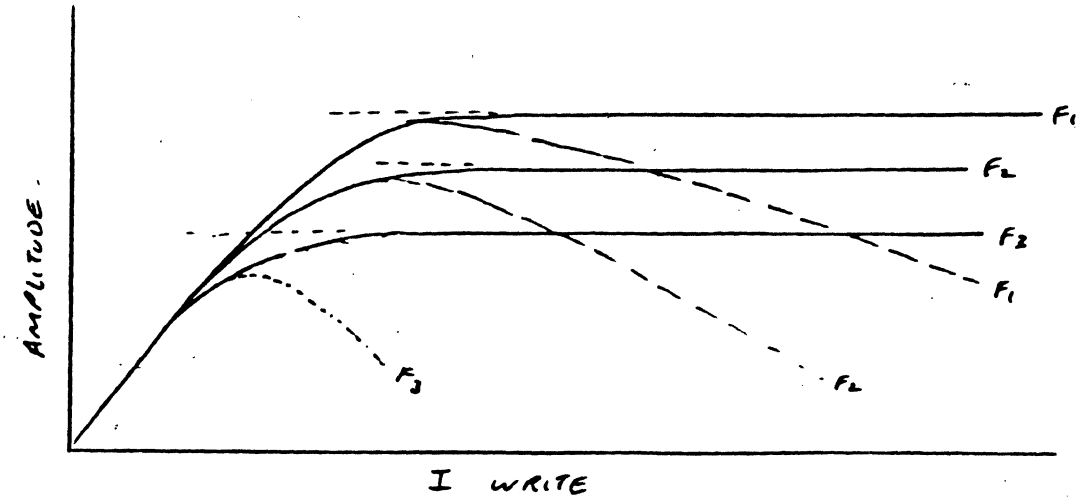


FIG 4.8 MULTIPLE DENSITY WITH FAR FIELD EFFECTS
SOLID LINE - NEAR FIELD. DASHED LINE - FAR FIELD.

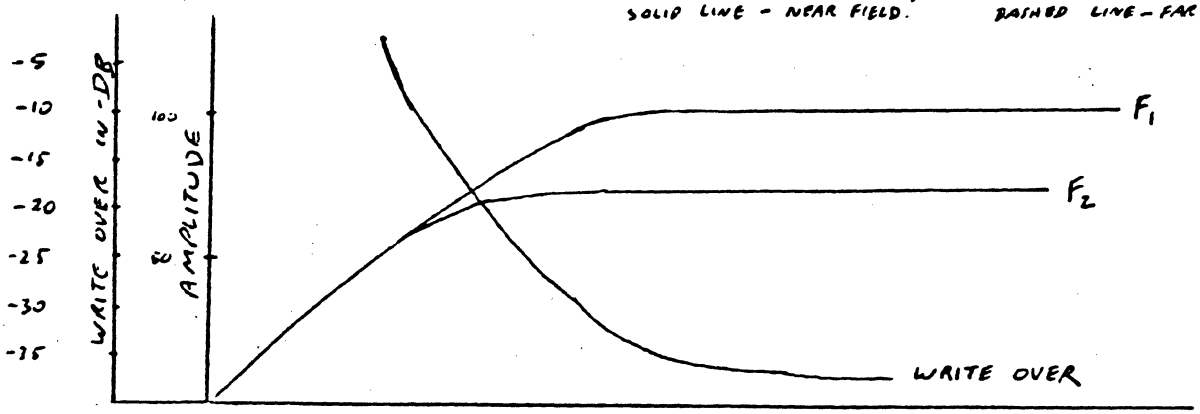


FIG 4.9 SATURATION AND WRITE OVER CURVES

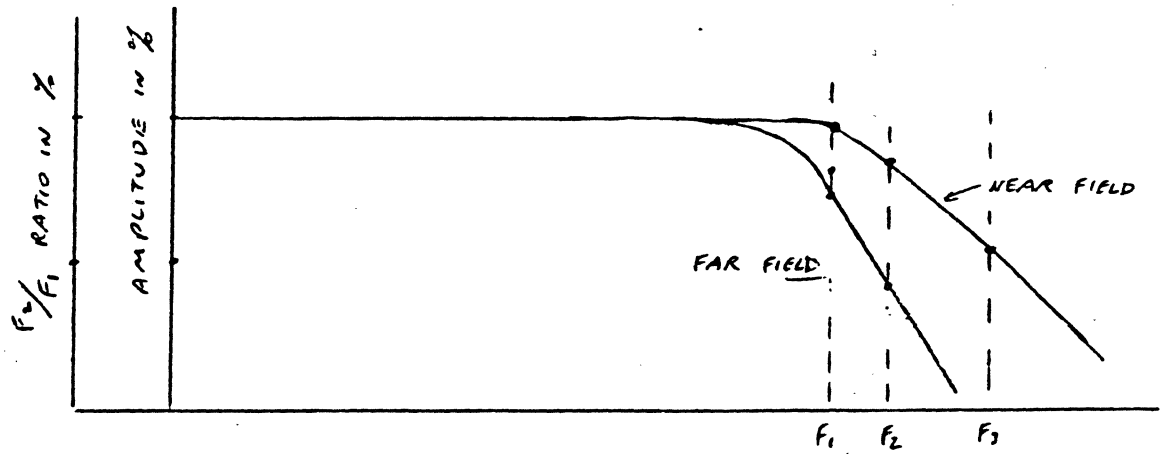


FIG 4.10 FAR FIELD EFFECTS ON RESOLUTION

HEAD PERFORMANCE

A second valuable measurement is the ratio of the amplitudes of the ~~of the~~ highest to lowest densities recorded. This is usually expressed as a percent. The lower the percentage the further the two points are apart on the bit density curve, or if the two points are a fixed density ratio apart then it indicates the points are further to the right on the bit density curve. This is particularly true if the recording involves the far field. Figure 4.10 illustrates this effect.

In the near field case, the ratio of ~~F2/F1~~ V_{f_2}/V_{f_1} is about 0.8, whereas the far field ratio is .4/.75 or .53. Back to the near field case, to get the same .53 ratio the transition density separation is F1 to F3.

Because of the write over requirements the write current must be kept high, but if the far field effect are involved, both the amplitude and resolution, hence bit shift, suffer. A compromise must then be made between the two. It is then obvious that far field recording is undesirable. Write over values above -26 db are unacceptable. Usually we require at least -30 db to keep from degrading the amplitude and resolution or bit shift. The current value is always to the right of the saturation point regardless of the write over value. This is necessary to ensure erasure of old information.

The last important measurement is the signal to noise ratio. Noise consists of five general components. First is the electronic noise assoicated with the amplifier first stage, the amplifier input current noise times the head impedance plus the amplifier voltage noise referred to the input. These two add as the square root of the sum of the squares. Barkhausen noise in the head core is also similarly added. The second noise is the media noise associated with the particle size, particle distribution and dispersion.

For particulate media this noise is considerable particularly as the track width diminishes. This noise increases as an inverse power function of track width. The third major noise source is the write over noise already discussed. The fourth noise is side fringing noise as read by the head from the adjacent track. The fifth noise source is the minor bit noise. These and electronic noise will be considered in a later chapter. The media noise will be worse for particulate media and best for thin film media such as metal films. This can be seen by considering the particles as separate magnets, each surrounded by a non-magnetic binder. Thus each particle contributes to the overall field, but as the view of the head decreases either in gap length or in track width, then the individual fields dominate which thus modulate the head signal.

If we record a single frequency signal (single density) and we were to read it back noiselessly the resultant spectrum would be a single line equal to the bandwidth of the measuring equipment. As we allow noise to enter the system the spectrum broadens into the typical bell shaped distribution for white noise, or if colored, as by media noise, a different shape. We could plot the peaks of all pulses in the presence of this noise and we would get a similar curve. Since we are most interested in these peaks as they represent the true position of the reproduced bit, we need to concern ourselves with the amount and sources of the noise. Similarly, as we move further to the right on the bit density curve, we must add the time shift caused by pulse superposition or interaction when we write bits of at least two different spacings randomly. The result is three curves or more each centered on the predicted peak shift 'd' for the indicated bit spacing and each containing the probability of peak position due to noise. This is illustrated in Figures 4.11 a, b, and c. The work was first described by D. E. Katz and is the subject of a paper by him and Dr. Campbell published later.

The ordinate may be changed to that of the time deviation from the expected time position of a recorded transition in a data stream. When this is done, Figure 4.11C becomes a plot of the probability of a transition being detected as a function of the expected transition of a noiseless non interacting system. If the time window allowed for each transition to be assigned to its correct time slot in a data stream were to be drawn on the curves of Figure 4.11C, we would notice that a portion of the transitions on either side of 'w' would be misplaced or be in error. We will discuss this further at a later time as there are many other effects that contribute to the number of transitions detected outside of its assigned window.

SIDE FRINGING

As mentioned earlier a significant noise source is side fringing. This signal has two components. Consider the head gap. It is three dimensional. So far we have only considered the field directly under the head core but the field emanates from the side of the gap just as much as below it. The field intensity limits for saturation are just as far as the depth of recording and worse as the field of non saturation extends even further. The head can read this field every bit as well as that under the head. Also it is as if the media were infinitely thick (to the side). Thus we would expect the field to behave as if it were a thick media or "far field" recording. This results in low density signals to be read at a higher amplitude than high density signals. Now we measure write over as a ratio of two low density amplitudes before and after a high density overwrite. It can be easily seen that the write over value is degraded by the side fringing signal since non saturated information is available to influence the head. The side fringing signal pick up is greater for low density signals. If two tracks were

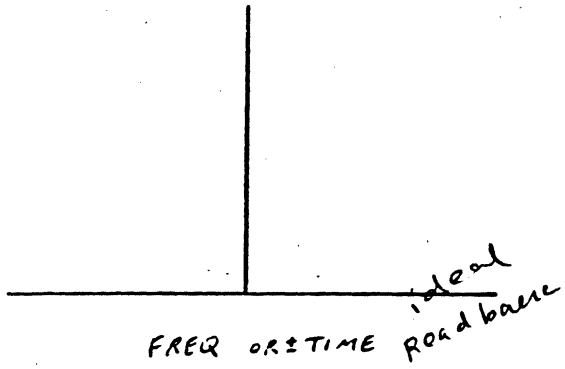


FIG 4·11 A

NOISELESS NON INTERACTING

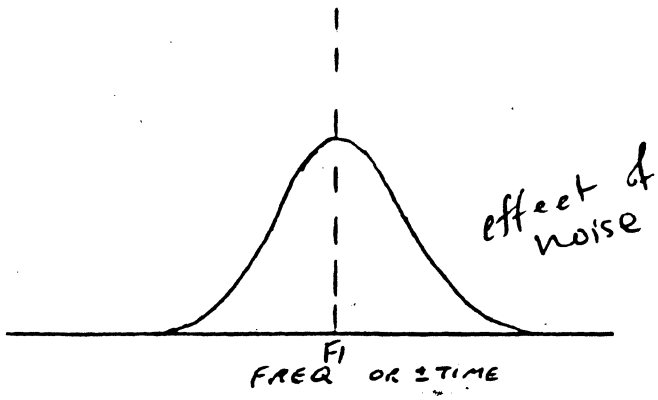
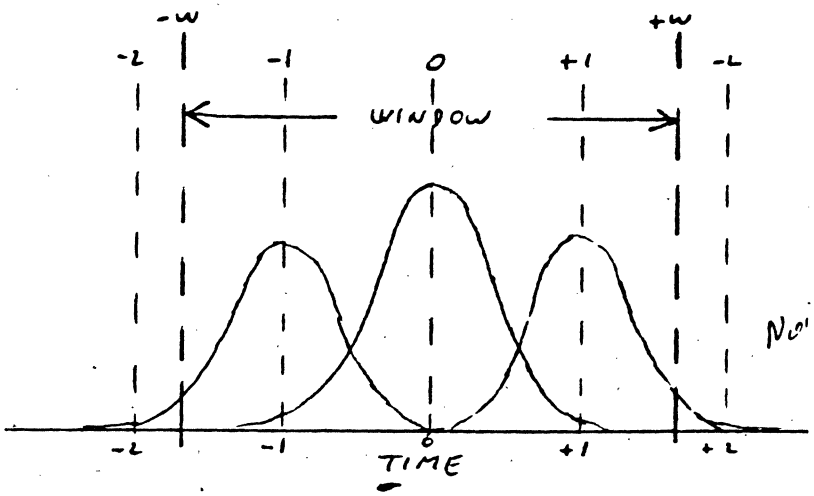


FIG 4·11 B

NOISY, NON INTERACTING



$$\text{Noise} = \frac{\text{peak-peak}}{\text{RMS}}$$

NOISY INTERACTING

FIG 4·11 C

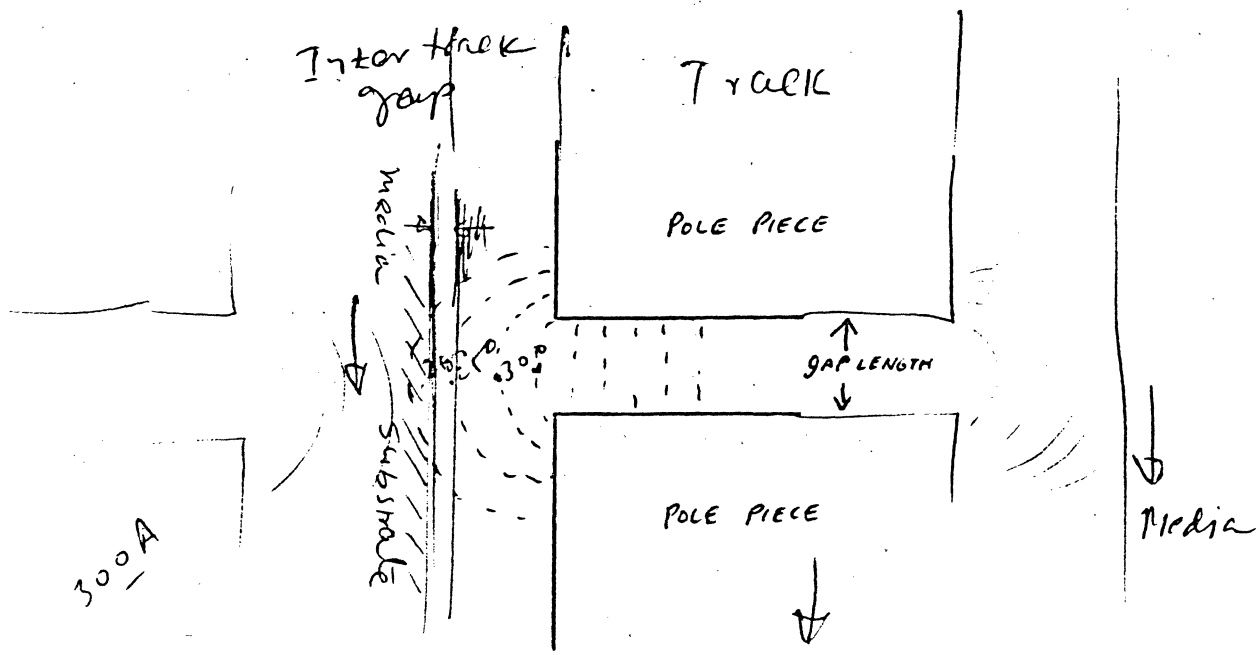


FIG 4·12 SIDE FRINGING

ONE SIDE ONLY SHOWN

immediately adjacent, the adjacent track recorded with a low density signal and the true track recorded with a high density signal then when reading the true track the read back signal would contain the low density signal as read to the side. If we were to plot the value of the fringing signal as a function of the low density frequency, we would observe an increasing fringing pickup with decreasing low density or decreasing frequency. All this means that the signal to noise ratio is further degraded from both on track low density signals previously recorded as well as adjacent track low density signals Figure 4.13 and 4.14.

MINOR BIT

Another noise source is the effect of the edges of the core away from the gap. These also represent a discontinuity in permeability and thus will appear as a partial gap. The gap length being infinity. On closer inspection, infinity is not correct as some field lines prefer to travel around the core and exit the side of the core thus generating a voltage in the coils.

This is illustrated in Figure 4.15 A, B. The resultant pulse is very broad and of low amplitude but contains significant energy. An experiment can be set up in which a low density signal is recorded and read back as isolated pulses. The amplitude and position of both the isolated pulse and the minor pulse are plotted as a function of the low density bit spacing. At a certain spacing which coincides with an exact multiple of bit spacings equal to the core length the isolated pulse is dramatically affected by the minor bit as it adds, Figure 4.16, or subtracts its energy to the isolated pulse height by the few percent amplitude of the minor bit, but such is not

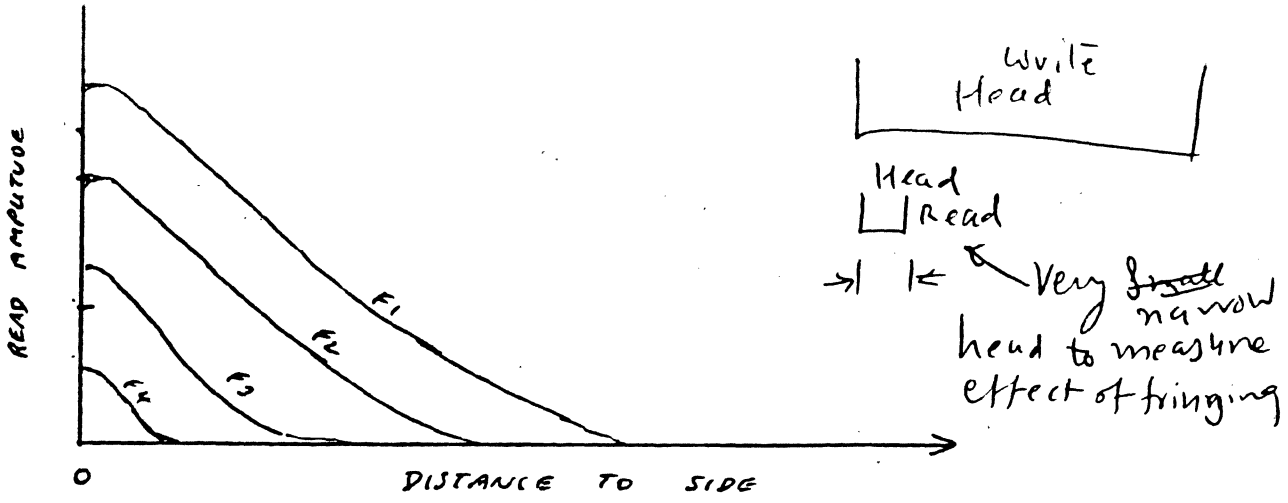


FIG 4.13 ON TRACK FRINGING CONTRIBUTION

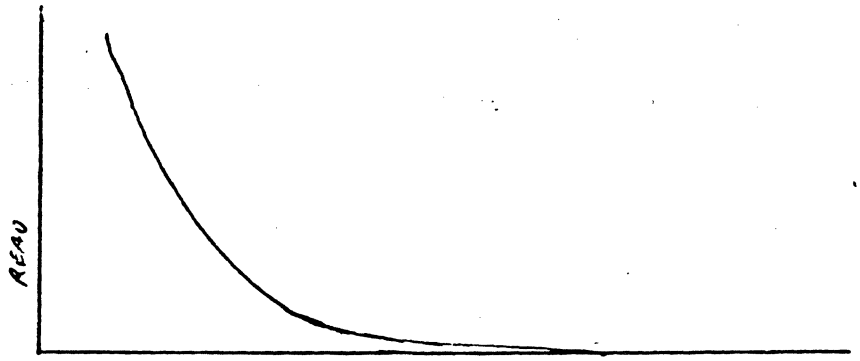


FIG 14.14

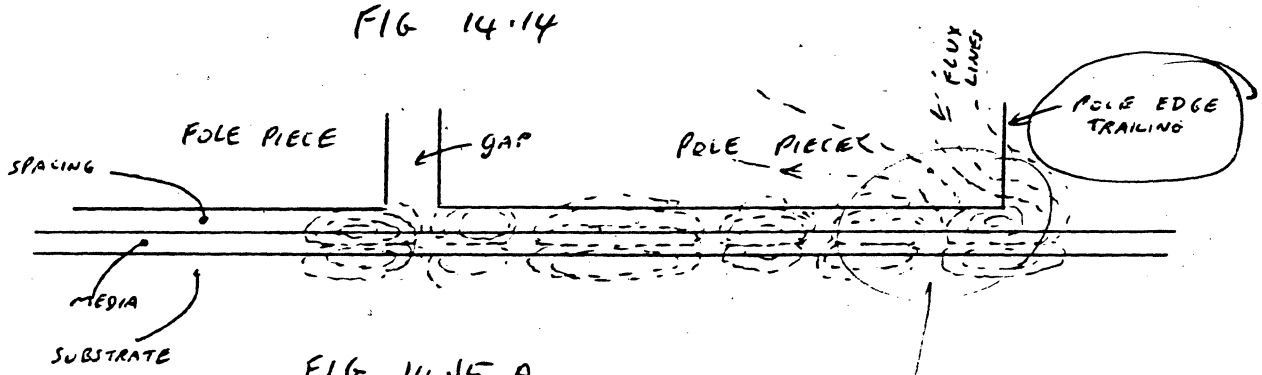


FIG 14.15 A

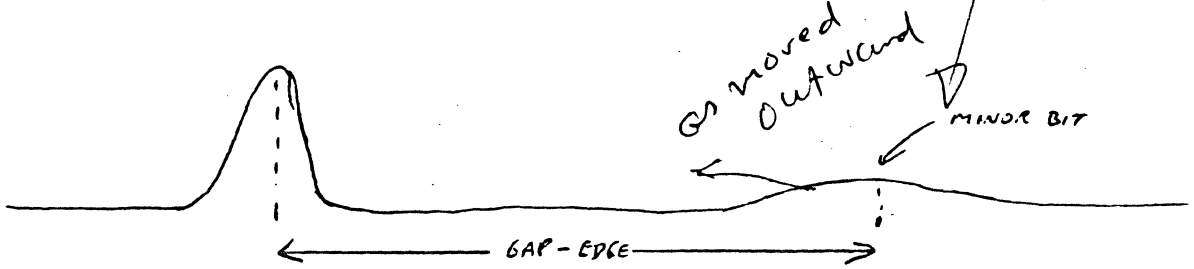


FIG 4.15 B

SIGNAL WAVEFORM SHOWING TRUE SIGNAL (ISOLATED PULSE) AND A MINOR BIT DUE TO THE TRAILING EDGE LATER

the case. Amplitude increases of ~100% have been observed indicating that energy is involved, Figure 4.16. The reason this is not observed more often is that normal recording is of higher density which masks some of the effect. As this noise does affect the recording performance the head is modified to reduce the pick up.

Head manufactures usually degrade the leading and trailing core edges either by increasing the flying height at these edges or by machining the edges so that it is not parallel to the recorded transition or by crumbling the corner so that it does not present a uniform edge equal to the track width. This phenomenon is only a reading phenomena. The write field strength at the trailing core edge is not sufficient to move the media remenant field, Br, enough to influence the read back process.

This can best be seen when recording on a disc on the inner diameters where the pole edges 'B' are not over the track 'A' made by the regular gap. Then moving the head to have the gap over the 'B' track. No evidence is seen of the signal recorded while writing 'A' even when using a spectrum analyzer as the measuring device. The thin film heads have significantly shorter core pieces, therefore the minor bit is substantial. It shows itself as an undershoot on both sides of the isolated pulse. A second effect in disc recording is an amplitude modulation as a function of radius for constant frequency record. These two effects are shown in Figures 4.17 and 4.18 respectively.

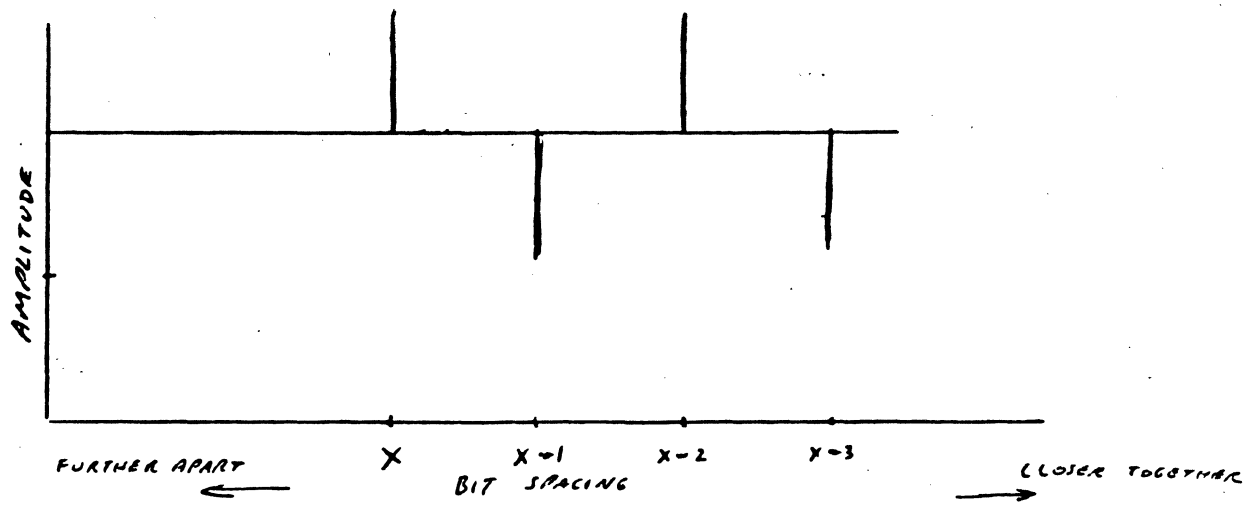


FIG 4.16 A AMPLITUDE VS BIT SPACING, LOW DENSITY

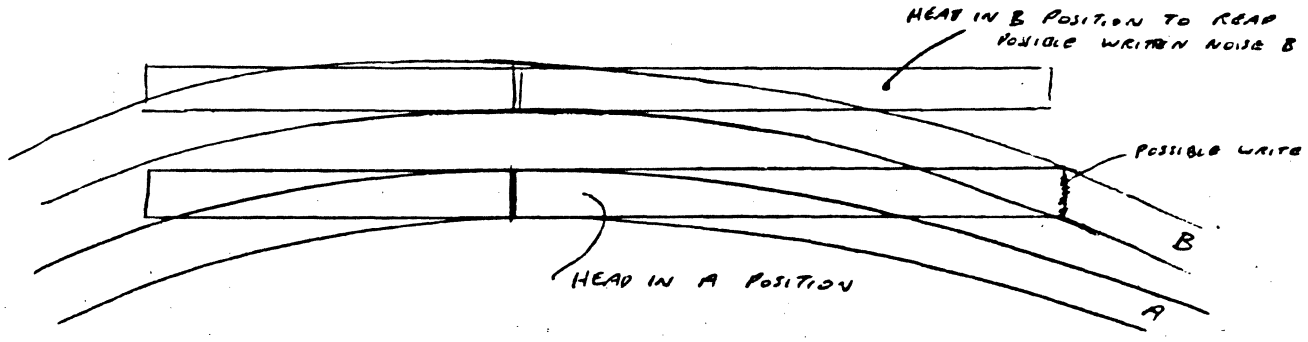


FIG 4.16 B

POLE EDGE PHENOMENA.

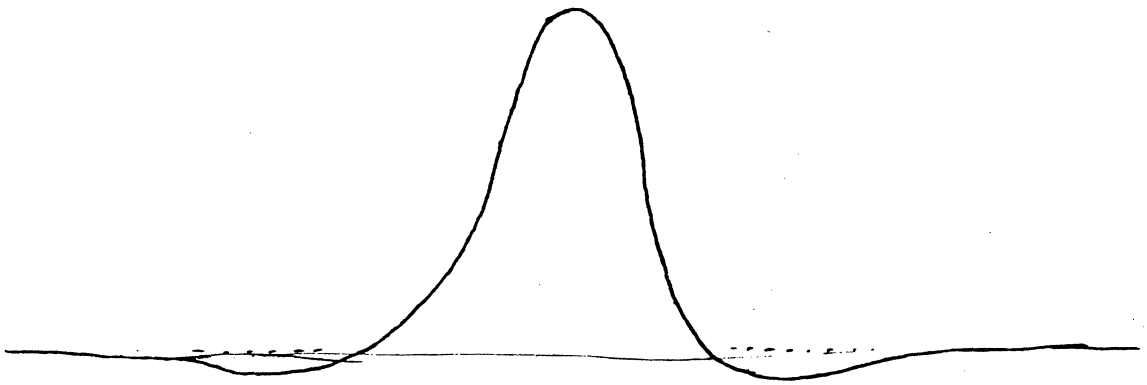


FIG 4.17 THIN FILM HEAD OUTPUT

HEAD PERFORMANCE

The number of undulations being determined by the ratio of the pole thickness and the diameter change from ID to the OD. FIG 4.18 Similarly, we would expect a modulation if we wrote varying bit density signals on a constant track as shown in Figure 4.19 which is the standard density curve. At very low densities the transition spacing exceeds the pole tip length, therefore, no modulation occurs. The above assumes equal pole tip lengths.

The isolated pulse shape is the same for all low density signals below the pole tip length. When the transition spacing nears the pole tip length, the shape of the isolated pulse changes until it affects the amplitude. Thereafter the density curve is modulated for all higher density signals.

During this chapter we have focused on three fundamental curves that describe the performance of the heads and discs together. We can summarize by drawing several curves that relate the various mechanical dimensions of the head and disc. The unlabeled dimensions are considered unchanging.

The five mechanical parameters that affect head-media performance significantly are the head gap length, head spacing, head coil turns, media thickness and media coercitivity. The actual shapes of the above curves are only to show trends not actual ratios. Of these curves the head gap length, head spacing and media coercitivity control the transition density performance as long as the signal to noise ratio remains the same. Generally we can say that as head gap length decreases, as long as the combination of flying height and media thickness is kept within the near field definition, transition density can increase.

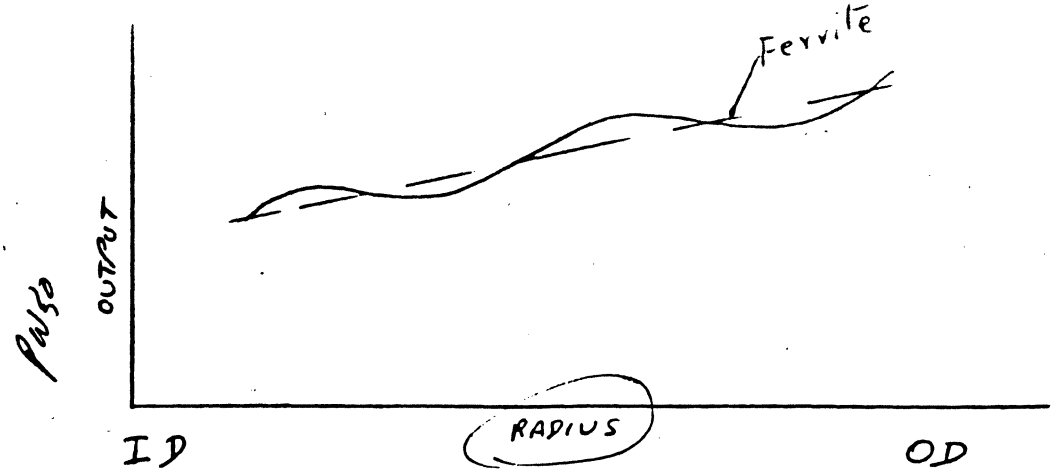


FIG 4-18 AMPLITUDE VARIATIONS WITH RADIUS of DISK

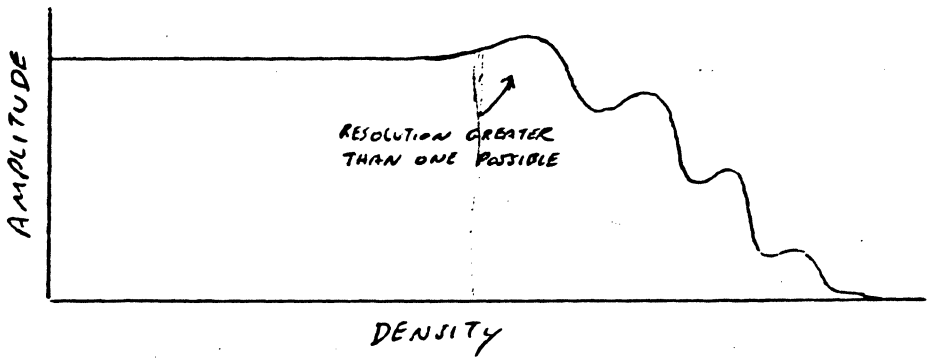


FIG 4-19, AMPLITUDE VARIATIONS WITH WAVELENGTH

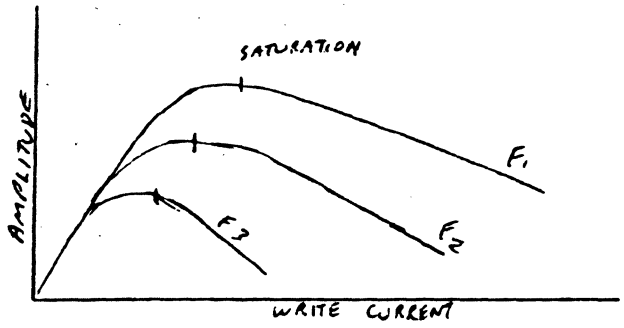


FIG 4-20

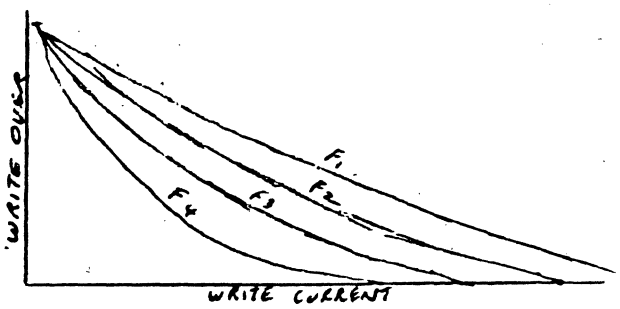


FIG 4-21

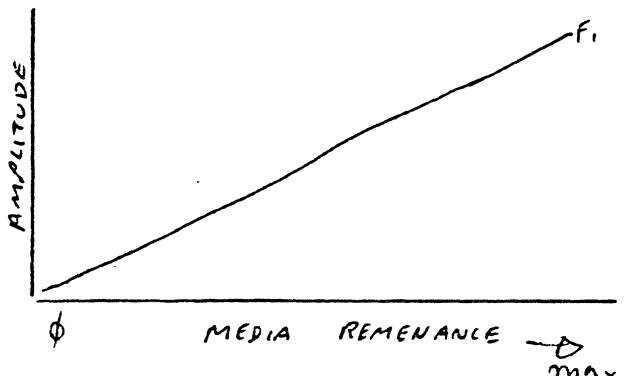


FIG 4-22

magnetic material has high Remenance

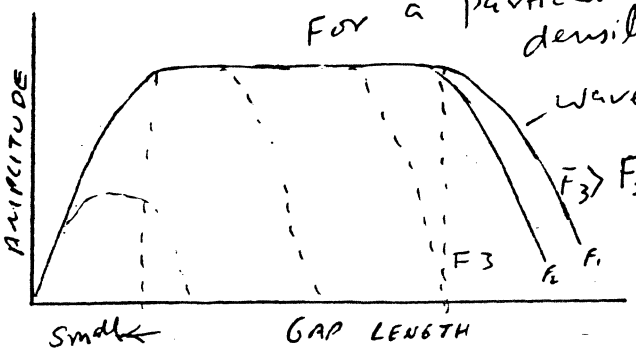


FIG 4-23

OFF TRACK CONSIDERATIONS

Both tape and disc machines exhibit problems with registration of the written track and the reading head. In tape machines this occurs in two areas. First the skew of the head centerline from the centerline passing thru the center of all parallel transitions. The angle produces two problems. The angle produces a cosine error in the track width which lowers the signal amplitude and a cosine function that broadens the transition as seen by the head gap thus lowering the amplitude and effectively increasing the Pw50 which reduces frequency response. The other is tape registration which is a problem relating to the guides and the slitting process of the tape itself.

In disc drives part comes in the form of disc runout which is similar to the tape guide-slit edge problem wherein the disc does not always rotate around the same point. This is due to bearing problems. Earlier disc drives have a cantilever bearing system which accentuates the problem. Also pack mounting repeatability is a problem. These together cause the disc line of rotation to precess which moves the track from its expected position as a cosine error. With a disc stack of more than one disc this makes the error subject to vertical location.

Another area of concern is the carriage and ways. These are the moving parts that hold the head arms and allows movement into and out of the pack, a radial change in position. Any tilt of this assembly either due to machining or due to debris on the bearing surfaces will again cause a cosine error which worsens the further the head position is from the bearing surface. The manufacturing repeatability of the head arm and its alignment introduce either direct off track position error due to misalignment or cosine and cosine error from gap skew as previously discussed. The latter group of

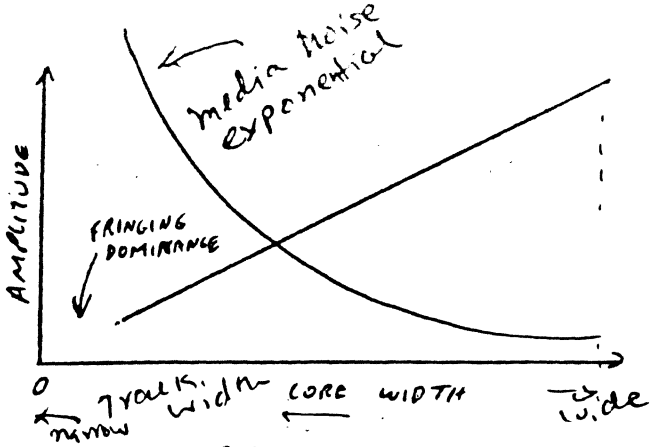


FIG 4-24 TRENDS

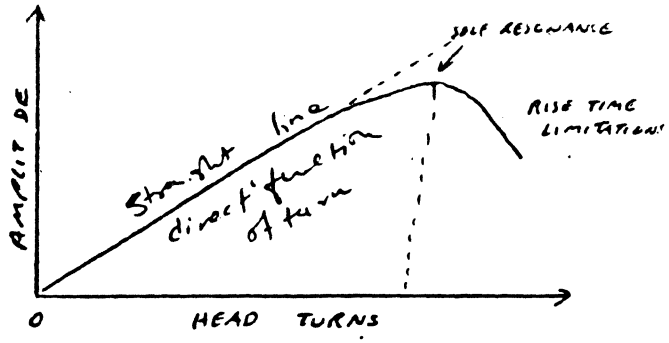


FIG 4-25 TRENDS

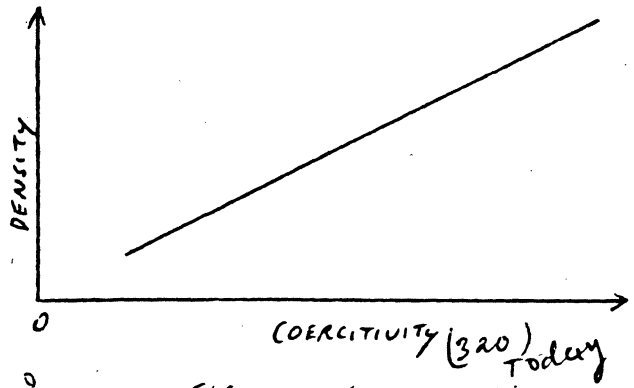


FIG 4-26 TRENDS

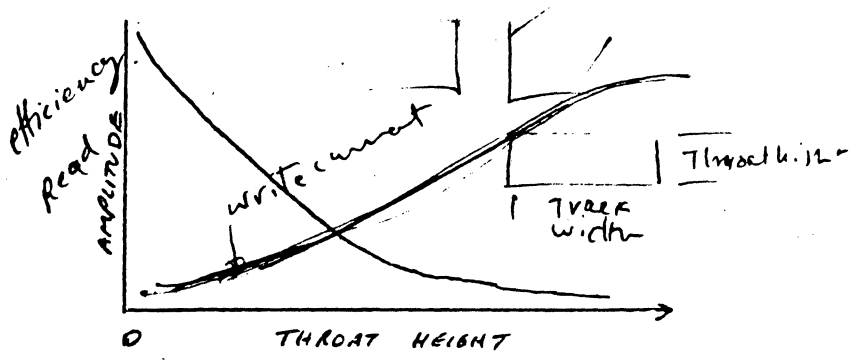


FIG 4-27 TRENDS

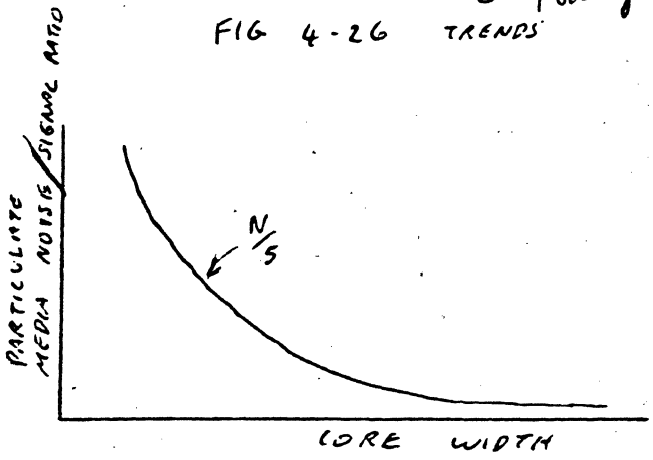


FIG 4-28 TRENDS

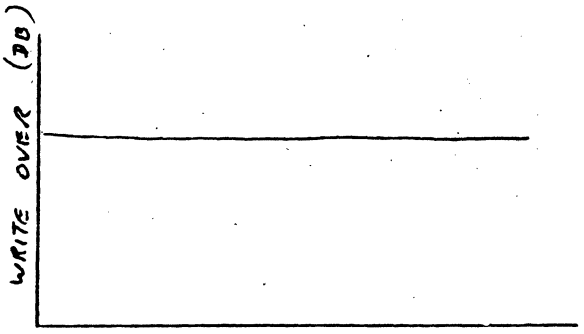


FIG 4-29 TRENDS

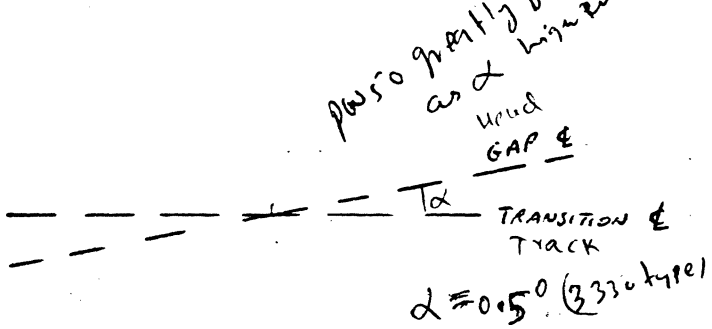


FIG 4-30

GAP SKEW AS IT AFFECTS PW50 AND AMPLITUDE

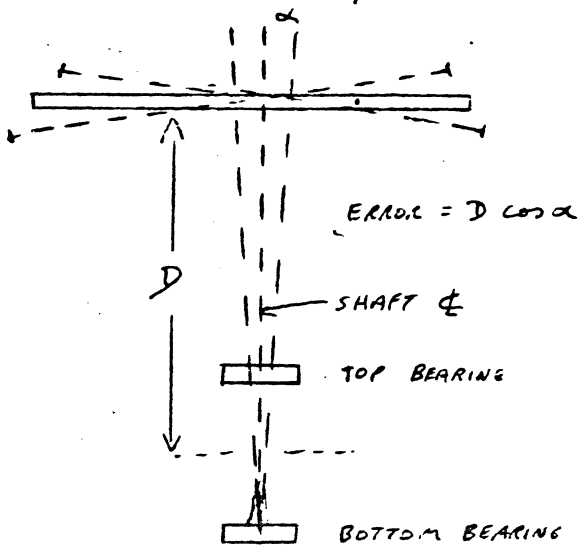


FIG 4-31 DISC TRACK ERROR (RADIAL)

errors have been eliminated in the fixed pack concept which was introduced in 1974 by IBM in the 3340 machine wherein the heads, carriage and way are included with the spindle in a separate package or module. The remaining tolerances remain until they can be reduced by changing the location of the bearings to either side of the spindle and carriage.

Early disc drives used a detent arrangement to locate the position of each track. These tolerances were enormous compared to the track spacings. For example in the 2314 the track spacing is 10 mills. A total of 3 mills was allowed for all the above tolerances, or 30% of the track width. Later machines achieved better registration by utilizing a closed loop positioning servo to locate each track. Here a single head, the servo head, is made to follow a pre-recorded track containing positioning information. This cut the carriage tilt error to about half and similarly the precession errors. Added though is the ability of the servo system to follow the track.

The total savings were positive thus permitting a present 960 tracks per inch or about 1.04 mill track spacing for the Memorex 3652 machine. Any mispositioning of a head in relationship to its recorded track results in increased noise in the form of adjacent track signals during read. A misplaced written track similarly creates problems for both the track of interest as well as the adjacent track and finally a reduction in signal amplitude due to the mispositioning. As can be seen the closer together the tracks, the less movement can be accepted before the signal is degraded. Typical ratios of head width to track separation remain fairly consistent for all machines

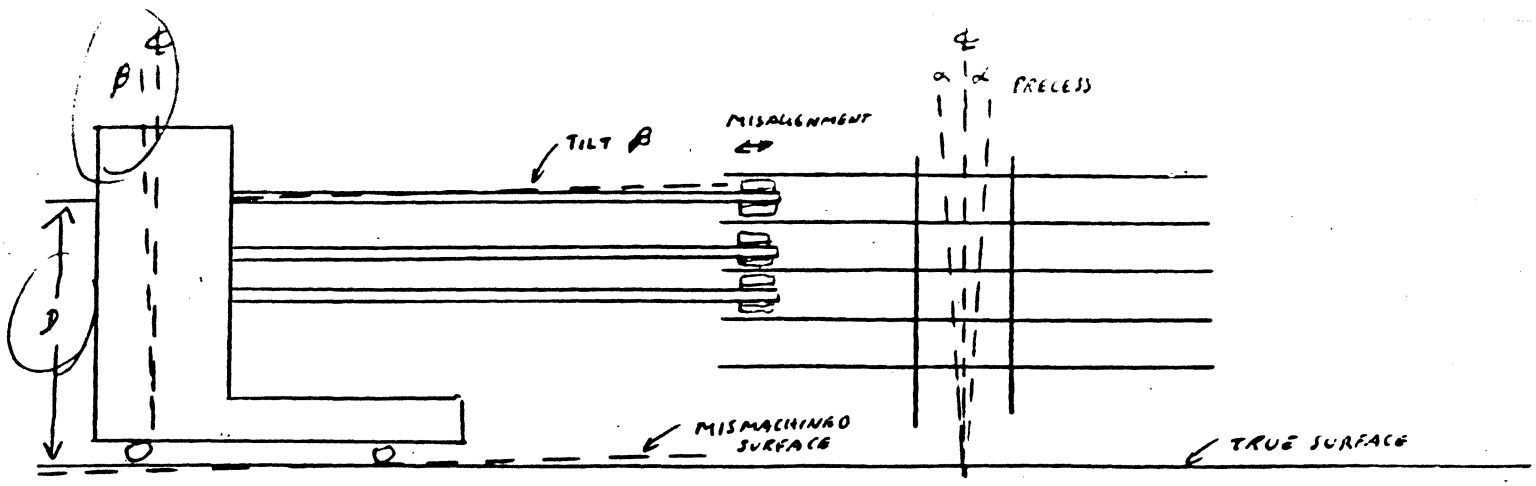


FIG 4·32 MISALIGNMENT, $\pm D\cos\beta \pm D\cos\alpha$

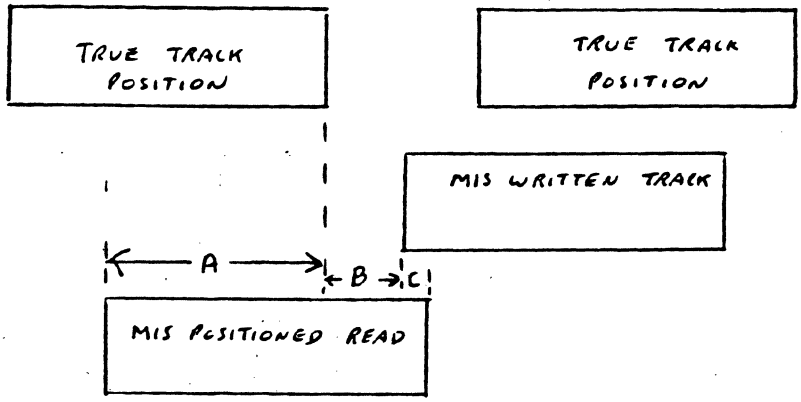
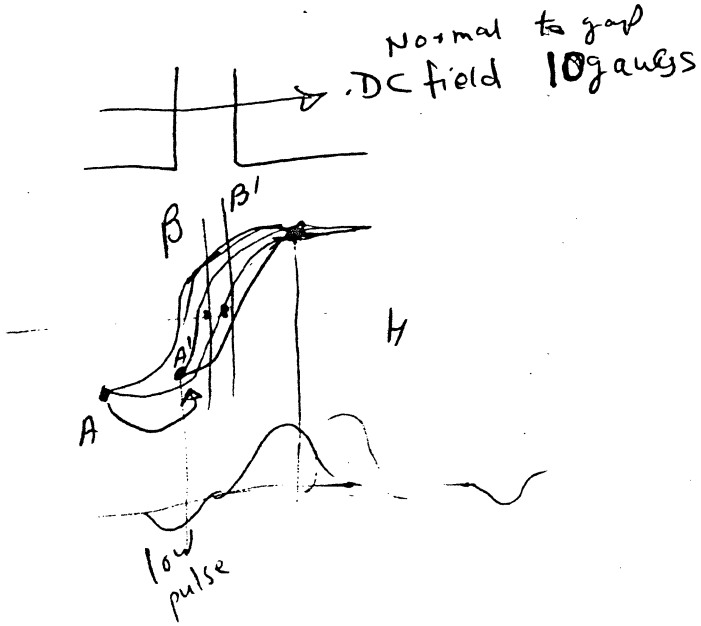


FIG 4·33
TRACK MISALIGNMENT R-W-R.



However, if we align the gaps to the radial line at some mid position then we would get $\pm(\alpha=\beta)$ for both the I.D. and the O.D. The true alignment point would be at the radius where $\beta=0.5^{o}$ sec Fig. 3.35.

Now the difference between α and β is large and normal skew misalignment is usually kept to within ± 30 or $\pm 0.5^{o}$ to minimize amplitude loss. This then restricts the total travel of the head. For example:

$$R = 6.5" \qquad D = 0.02"$$

$$\alpha = \sin^{-1}\left(\frac{\frac{0.02}{2}}{6.5"}\right) = .08814736"$$

$$R_2 = \frac{.01}{\sin 1.08814736} = \frac{.01}{.01899061} = .5265" \text{ radius}$$

if $\frac{d}{2} = .1"$ $\quad R_0 = 6.5"$ $\quad R_1 = 4.0"$ $\quad \alpha = .44074106^{o}$, $\quad \beta = .71621585^{o}$

Δb .120 mills. This says that $\Delta \angle$ is $2(\beta - \alpha) = (.275475^{o})2$ This is close to the spec for $\pm 0.5^{o}$ skew.

$$b_0 = 6.5(1 - \cos(\sin^{-1}\tfrac{.05}{6.5})) = 1.4234 \times 10^{-4}$$

$$b_1 = 4.0(1 - \cos(\sin^{-1}\tfrac{.05}{4})) = 3.125 \times 10^{-4}$$

This says that there is no positioning reason for not having two heads, one for write and one for read with a radial head movement.

The problem is the isolation required.

$$Vw = 7.V_{BP} \qquad V_R = 1.0 \times 10^{-3} \text{ V @ -30 db S/N}$$

$$Vwn = \frac{.001}{(31.622777)}$$

$$= 3.162 \times 10^{-5} V$$

for 106.9 ~~10590~~ db isolation for a noise contribution of -30 db.

at around 70% which is just the same as for the old detent machines of the 1960's. The difference being that some tolerances have been reduced permitting an increase in track density up to the next limitation.

An example of the signal degradation due to mispositioning is illustrated in Figure 4.33. The signal read will be A, the intertrack gap is B and the adjacent track signal is C. Fringing is also a factor.

$$\text{Sig} = T(\text{on track amplitude})\ \frac{A}{T_1} + F_1$$

$$\text{Noise} = T(\text{on track amplitude})\ \frac{C}{T_2} + F_2 + T\ (\text{surface noise})$$

SKEW

At the ID the gaps are separated by $2\left[\left(\frac{d}{2}\right)(BPI)\right]$ bits

At the OD the gaps are separated by $\frac{2d}{2}\left[(BPI)\left(\frac{R_1}{R_0}\right)\right]$ bits

If the gaps are symetrical around the radial line, then ÷ at OD the length

$b = R(1 - \cos \alpha) = R_0(1 - \cos(\sin^{-1}(\frac{d/2}{R})))$

At ID the length of $b = R(1 - \cos\ ) = R\ (1-\cos(\sin^{-1}(\frac{d/2}{R_0})))$

$\therefore$ the difference is $R_0(1 - \cos^{-1}\frac{d/2}{R}) - R_1(1 - \cos^{-1}\frac{d/2}{R_1})$

This could be compensated for by the servo track spacing as far as track centerline is concerned as well as the intertrack spacing.

The skew will be twice the difference between α and β if we align the gaps to the radial line at the O.D. I.D. = $2(\alpha - \beta)$.

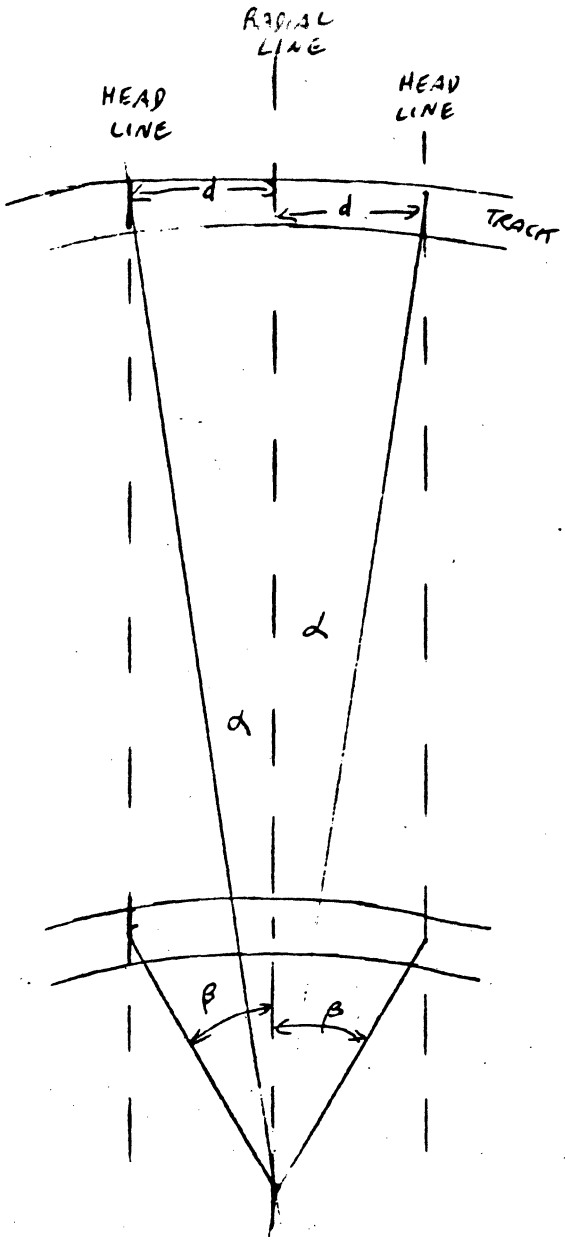


FIG 4-34

THE PROBLEM OF AN OFFSET HEAD READING RADIALLY WRITTEN DATA

HEAD PERFORMANCE

$$b_0 = 6.5\left(1 - \cos\left(\sin^{-1}\frac{.01}{65}\right)\right) = .00000768$$

$$b_1 = 4.0\left(1 - \cos\left(\sin^{-1}\frac{.01}{4}\right)\right) = .00000312$$

$$\Delta b = .000,004,56 \quad \text{or } 4.56\ \mu"$$

that is for a gap spacing of 10 mills

Try gap spacing of .10"

$$b_0 = 6.5\left(1 - \cos\left(\sin^{-1}\frac{.05}{6.5}\right)\right) = (.00002959)6.5 = .00019234$$

$$b_1 = 4.0\left(1 - \cos\left(\sin^{-1}\frac{.05}{4}\right)\right) = (.00007313)4 = .00031251$$

$$\Delta b = .120 \text{ mills}$$

RECORDING ELECTRONICS

R/W BLOCK DIAGRAM

Figure 5.1 is a general block diagram that may be used to define the circuits required. All R/W channels have this form. Its complexity may be increased depending on the sophistication of the recorded signal or it may be decreased for very simple signals. As most disc drives have multiple heads, it is obvious that some means be provided to isolate the individual heads from each other while allowing one head to function. This is the function of the block marked Matrix. It is fed from an address register that contains the head number selected. For reasons to be discussed later, these two blocks may be repeated. The blocks marked Read and Write perform these basic services. A means must be provided to select either. That is the function of the blocks marked Read Select or Write Select. Part of the write chain includes the Write Pre Driver, the Trigger and any encoding functions. The Read Chain includes the Linear Amplifier and Filter, the Detector, and a decoding or declocking scheme. Some subfunctions include Address Mark Detection and synchronization. A necessary set of functions include the Safety Circuits. These are provided in order to protect the recorded data either from simultaneous commands or from failed components or circuits. These circuits do not respond to legitimate though unintended commands. Tape drives generally perform these functions multiply in groups of 5, 7, or more depnding on the machine type. It is the p upose of the remainder of this book to address each of these blocks in turn. We will discuss the various interactions and requirements particularly those related to the head-disc interface.

WRITE CIRCUITS

The write circuit used depends on the head winding structure whether it is single-ended to reference, single ended floating differential or centertapped

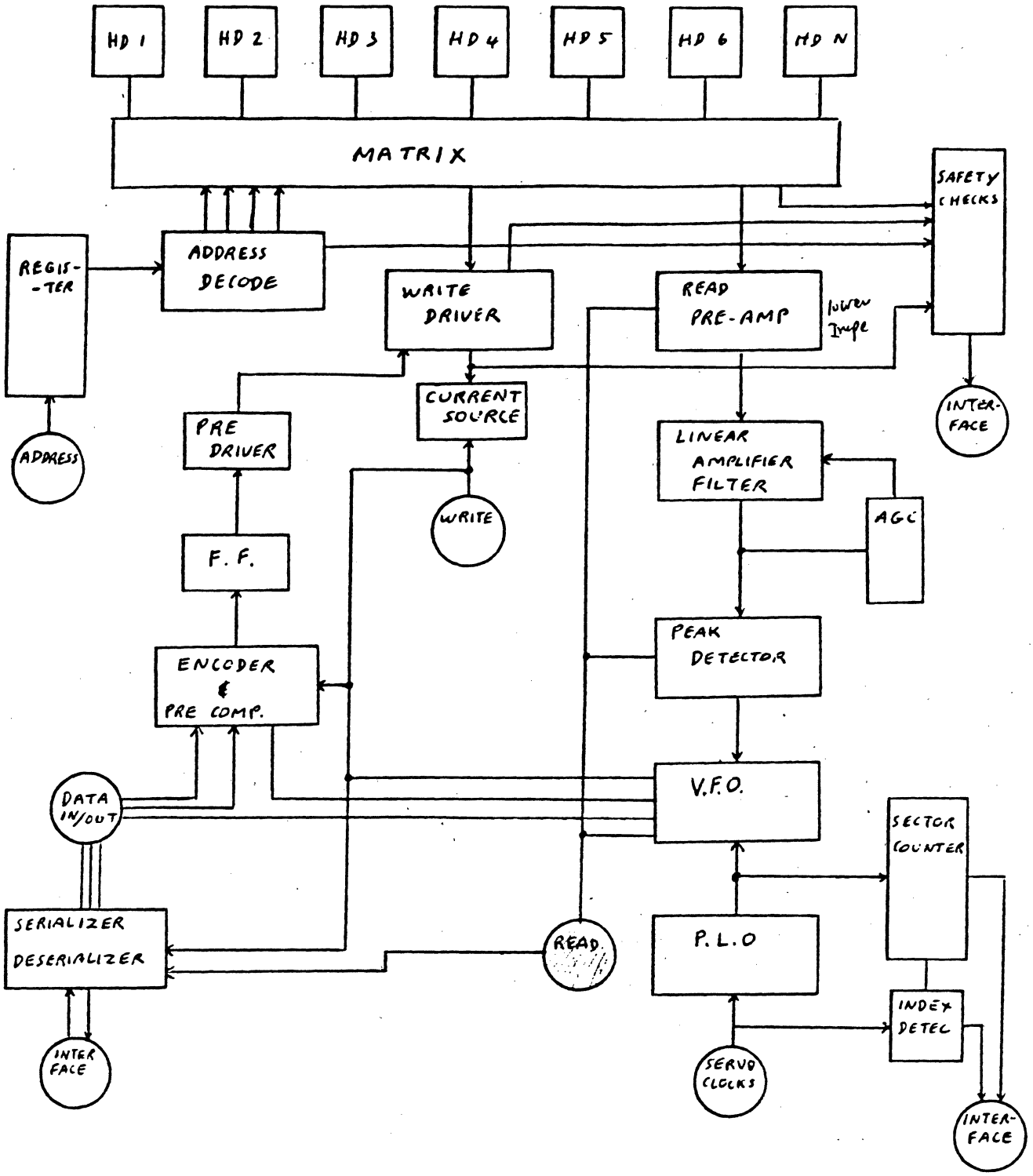


FIG 5.1 TYPICAL R/W BLOCK DIAGRAM

differential. The circuit also depends on the time between transitions. In the single ended version the write current required for saturation is alternately reversed in the windings producing the alternating flux reversals required for writing. This can be accomplished by the circuit of Figure 5.2A. Here the complimentary emitter follower drives is driven by a square wave that is carried above and below ground. The current flow is then determined by the voltage level out of the driver and the value of resistance in series. With large input voltage swings the value of the series resistor can be made large which minimizes the L/R time constant and thus reduces the time of the recorded transition. Power dissipation is large both in the Driver transistors, the input driving circuit and the series resistor.

The current is determined from EQ 5.1 and 5.2

$$DC\ I_{+_1} = \frac{Vin+ - Vbe^1}{R + Rh} \qquad (5.1)$$

$$DC\ I- = \frac{Vin- - Vbe^2}{R + Rh} \qquad (5.2)$$

If the circuit is balanced to ground then these two currents are equal except for the slight differences in Vbe and the input voltage swings. The circuit is worse cased by considering input swing variations, the Vbe variations and the two resistors variations, one a fixed and the other the winding resistance.

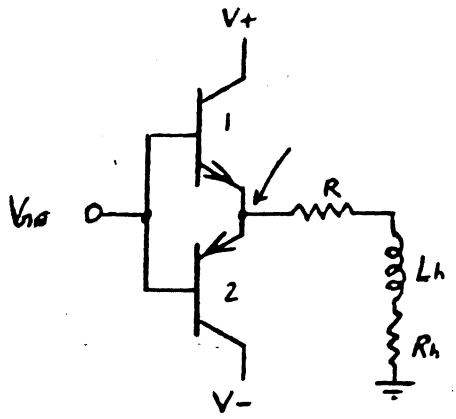


FIG 5.2 A

COMPLEMENTARY WRITE DRIVER FOR TWO TERMINAL HEAD

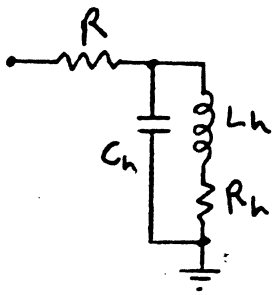


FIG 5.2 B

TWO TERMINAL HEAD WITH STRAY CAPACITANCE

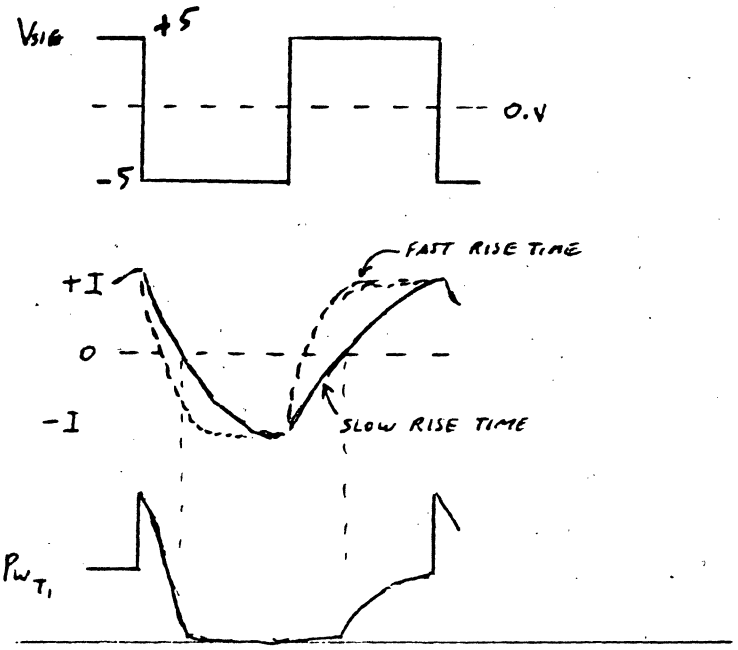
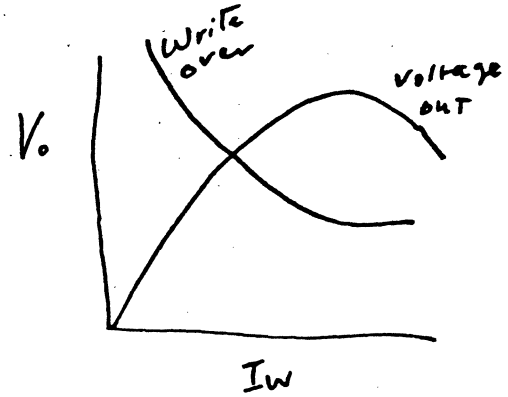


FIG 5.3

VOLTAGE, CURRENT AND POWER WAVEFORMS FOR CIRCUIT OF FIG 5.2



Power dissipation for the transistors is simply calculated, again worse case conditions must be assumed.

$$P_{T_1} = \frac{(\text{Vsupply max} - \text{Vsig min}_+ + \text{Vbe max})(V_{sig\ min} - V_{be\ max})}{\text{R min} + \text{Rh min}} \qquad \text{EQ 5.3}$$

It will be noted that the current is a function of time; therefore, the actual transistor power dissipation is less than EQ 5.3 would indicate during the time of the transition. Also the true maximum may not occur at Vsig min but at some other value. At the time after switching, the current thru the inductor cannot reverse instantaneously, therefore, the transistor power dissipation is increased in the same transistor until the current falls off to zero on its way to the opposite maximum. The base voltage changes to the opposite polarity but the current remains the same. The power peak is given by EQ 5.4.

$$P_{T_{Peak}} = \underline{(\text{Vsupply max} - (-\text{Vsig min}) + \text{Vbe max})\ \text{I Max}} \qquad \text{(EQ 5.4)}$$

Where I max is the current determined by equation 5.1 (or 5.2).
This transient power dissipation must be considered, particularly when secondary breakdown can occur. The choice of transistor then not only depends on the voltage and current, but unfortunately both at the same time. Figure 5.3 shows the relationships.

R/W BLOCK DIAGRAM

The head circuit cannot really neglect the capacitance; therefore, the actual head current is determined by EQ 5.4_2 for a step function, using the circuit of Figure 5.2B.

$$I_h = \frac{V\ sig\ (s)\left(\dfrac{(LS + Rh)\dfrac{1}{CS}}{LS + Rh + \dfrac{1}{CS}}\right)}{S\left(R + \dfrac{(LS + Rh)\dfrac{1}{CS}}{LS + Rh + \dfrac{1}{CS}}\right)(LS + Rh)} \qquad (5.4)_2$$

This breaks down to a third order step:

$$\frac{V\ sig\ (s)}{S\left[RLCS^2 + (RR_hC + L)S + R_h + R\right]} \qquad (5.5)$$

6/14/82

All this slows down the rise time, widens the transition width, which in turn widens the PW50.

Another circuit that could be used is shown in Figure 5.4. Here the write current is determined by the series combination of R, the head circuit, and the saturation resistance of the transistor.

$$I_{h(DC)} = \frac{V - Vsat}{R + R_h} \qquad (EQ\ 5.6)$$

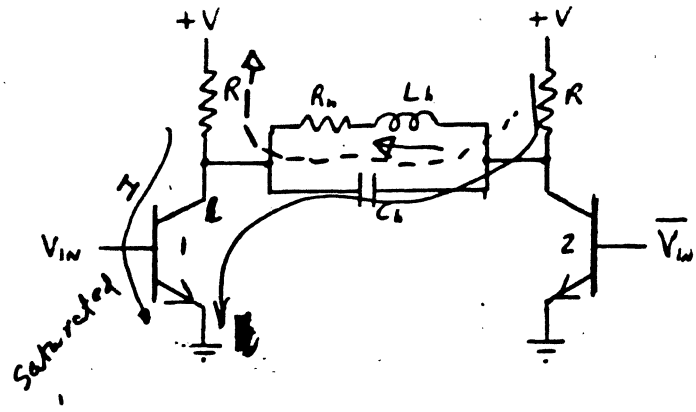


FIG 5.4

SATURATED DRIVER FOR TWO TERMINAL HEAD

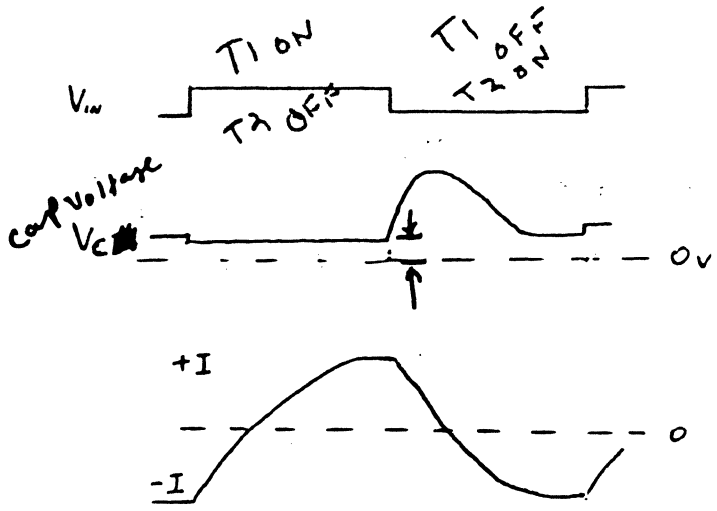
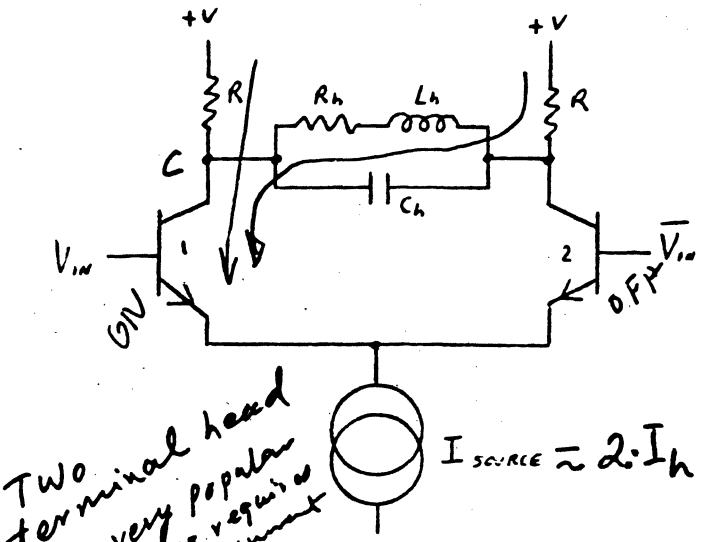


FIG 5.5

INPUT, COLLECTOR AND CURRENT WAVEFORMS



- Two terminal head not very popular circuit since requires two times current

FIG 5.6

CURRENT SOURCE DRIVEN LINEAR DRIVER FOR A TWO TERMINAL HEAD

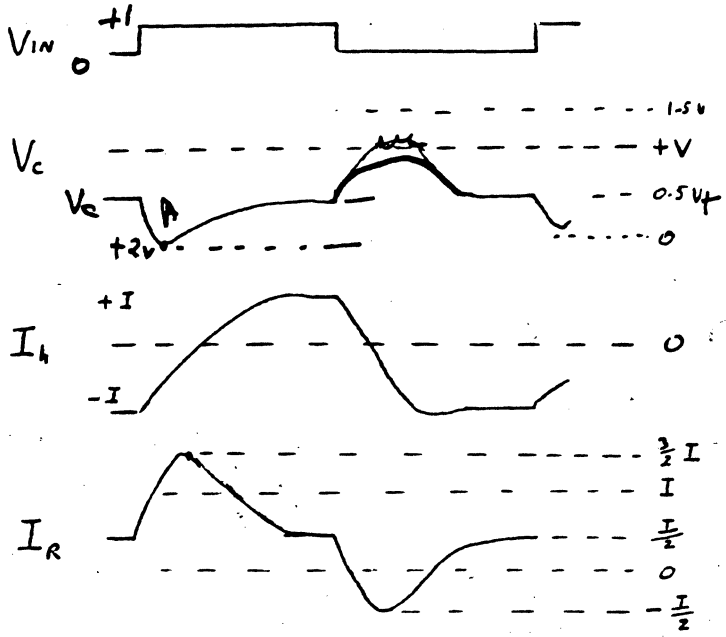


FIG 5.7

INPUT, COLLECTOR, CURRENT WAVEFORMS

Point A should never go below 1 volt ($\simeq V_{in}$)

$$V_C = V_+ - \left(I_T \cdot \frac{R}{2}\right)$$

$$\text{Steadystate } I_h = \frac{I_S \left(\frac{R(R+R_h)}{R+R+R_h} \right)}{R+R_h} \simeq \frac{1}{2} \cdot I_S.$$

The transient behavior is the same as EQ 5.5 only V sig is replaced by +V.

It should be noted that the rise time is affected by the storage time of the transistor. If the storage time is very small compared to the transition time then it might be a useful circuit. Note that the transistor current is nearly double being

$$I_{T_x} = \frac{V - Vsat}{R + Rh} + \frac{V - Vsat}{R} \qquad \text{(EQ 5.7)}$$

The power dissipation in the resistors are very nearly constant. The voltage breakdown requirements for the transistor include the voltage developed across the head at turn off time due to the inductance. This can be nearly the same as +V meaning the transistor will see 2V during the transient.

The damping of the head for a zeta of .95 can be accomplished by the collector resistors or by the addition of a third resistor in parallel with the head.

Tolerances on the Resistor, the Vsat, the supply, and the head winding resistance determine the range of write current expected in a manufacturing run.

A third circuit is shown in Figure 5.6. Here the transistor storage time is eliminated, but the current source must supply nearly twice the head

current as is also required in Figure 5.4. The commutating diodes are eliminated by making +V equal to twice that required which leaves a bias of +V on the collectors. This accommodates the negative V swing of the head without saturating the transistor. A penalty is that the transistor power dissipation is high. The average Pw being for the transistor,

$$Pw\ ave = (+V - I_S\frac{R}{2} + Vbe - Vb)I_S \qquad \text{(EQ 5.8)}$$

The time domain transient equation is the same as equation 5.9 and 5.10

$$I_h = \frac{I(s)\ \frac{1}{LC}}{S(S^2 + \frac{1}{2RC}S + \frac{1}{LC})} \qquad \text{(EQ 5.9)}$$

$$= \frac{I(s)\ Wn^2}{S(S^2 + 2\zeta WnS + Wn^2)} \qquad \text{(EQ 5.10)}$$

The transistor voltage breakdown requirement is 1.0 V due to the voltage rise resulting from inductive current. Again the damping is achieved via 2R or a third resistor in parallel with the head. It will be noted that the current thru a resistor at switching time goes from I/2 to 3/2I during the transient and back to I/2 again for one half of the cycle. On the second half cycle it goes from +I/2 to -I/2 and then back thru zero to + I/2 again. The degree of achieving these excursions is controlled by both zeta and the head capacitance.

The peak power dissipation for the resistor is threfore approximately

$$P_{R\ Peak} \leq \left(\frac{3I}{2}\right)(V) \qquad \text{(EQ 5.11)}$$

occurring at time A on Figure 5.7.

A fourth circuit and its variations can be used which reduces the power dissipation by requiring a current source of only I instead of the 2I as used in the previous two circuits. The basic circuit is that of a current controlled bridge. In this circuit the current path is controlled by a pair of emitter followers in the upper half of the bridge. The base voltage swing Vb1 - Vb2 must be large. The negative going portion must be greater than the voltage developed across the head during switching. Fig 5.8

The average power dissipation of the upper transistors is half the DC value if the signal on Vb1 - Vb2 exceed the transient head voltage.

$$P_{T(1\ or\ 2)} = \frac{(V - Vb_{1+} + Vbe)\ I}{2} \qquad \text{(EQ 5.12)}$$

The head current equation is the same as in EQ 5.10. If the input Vb1 and Vb2 is less than the transient voltage then current must flow thru T1 or T2 during a portion of the transient; therefore, the power dissipation is increased by that current flowing times the V-Vb difference.

$$P_{T_1} = (V - (-Vb_1) + Vbe)\ I_t \qquad \text{(EQ 5.13)}$$

Where I_t is that portion of I_h supplied thru the transistor.

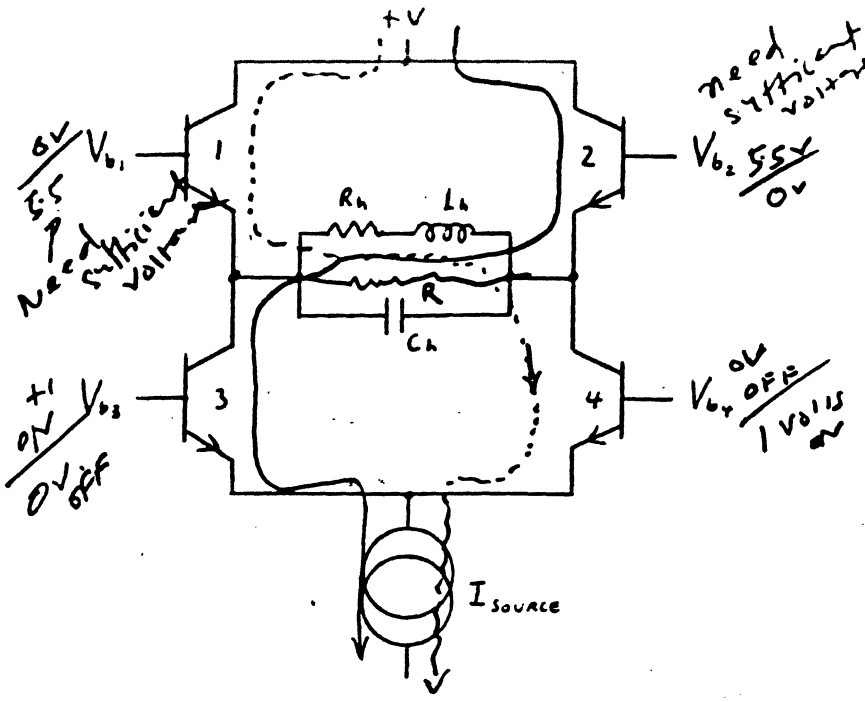


FIG 5.8

COMPLIMENTARY BRIDGE DRIVER FOR A TWO TERMINAL HEAD

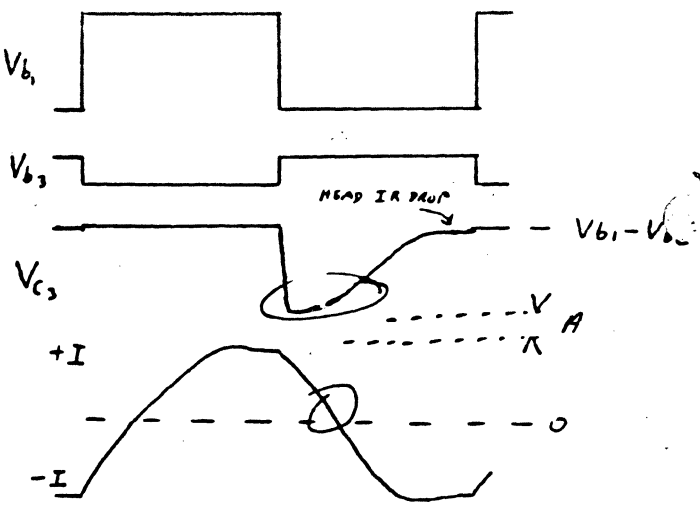


FIG 5.9

6/16/82

INPUT, COLLECTOR, CURRENT WAVEFORMS

worst case

$$V_{b3} < V_{m2_{min}} - V_{bez\,max} - V_{peak}$$

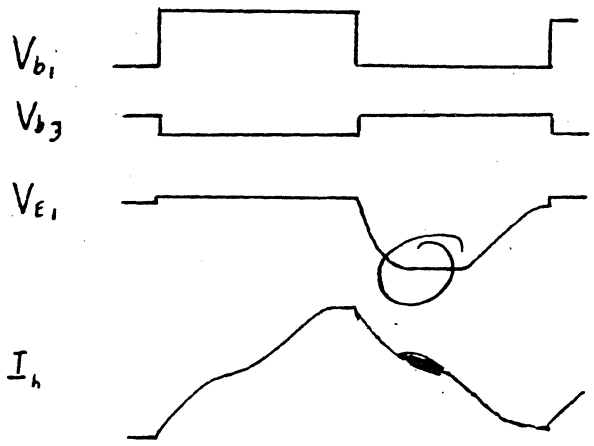


FIG 5.10

CURRENT DISTORTION DUE TO INSUFFICIENT INPUT TO V_{b_1} OR V_{b_2}

The modification of the head current is due to a portion of I source being supplied thru the non off upper bridge transistors. One disadvantage of this bridge circuit is the circuits that are required to drive the bridge. These circuits also have power dissipation particularly the circuits driving the upper half of the bridge due to the large swings required, and if fast speed is required, low impedance, high current.

There are several circuits that may be used. Note the phasing required. Because of the various propagation delays and turn on - turn off times, the bridge may exhibit current spiking where both ~~transitions~~ TRANSISTORS may be on momentarily at the same time providing a path directly from +v to the current source. Fortunately, the current source prevents the larger currents that occur in saturated bridges.

With these drivers the current sources determine the swing available. The tolerance of the various resistors and the tolerances on the current source must ensure adequate swings on Vb1 and Vb2 to maintain an unaltered current waveform. Care should also be exercised to minimize this margin as the power dissipation of the bridge depends on these voltages and the current. If too large a margin is provided, A in Figure 5.9, then the lower half of the bridge has a higher than necessary dissipation. If not enough margin is provided, then the bridge saturates and rise time is degraded. Further if the swing on Vb1 - Vb2 is small then the upper half of the bridge experiences a higher dissipation. Normally, the upper half of the bridge only sees the difference +V and Vb1 or Vb2 times the current source value. By using the circuit of Figure 5.11B this is minimized. One nice thing about the combination of

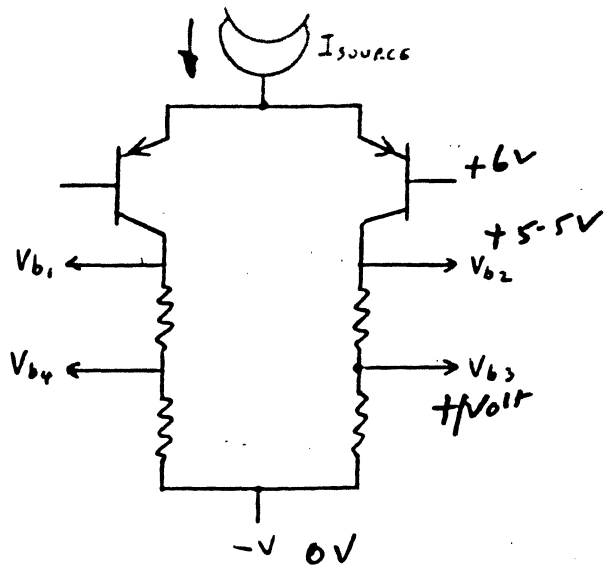


FIG 5.11 A

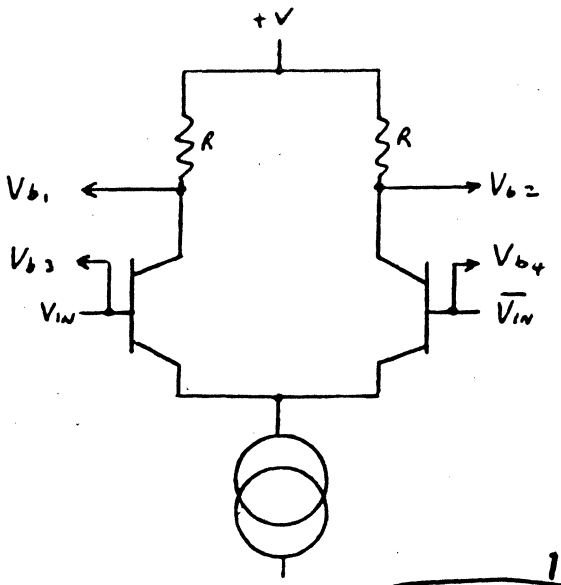


FIG 5.11 B

BRIDGE DRIVER C

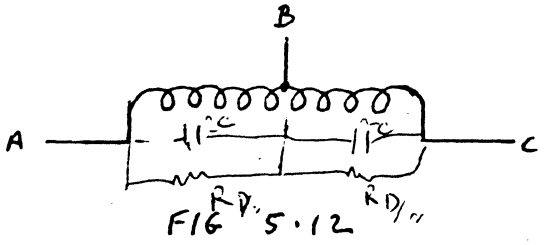
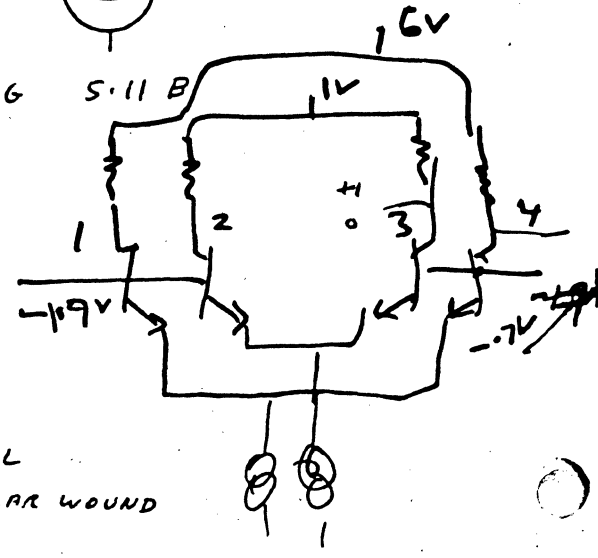


FIG 5.12

CENTER-TAPPED OR DIFFERENTIAL HEAD WINDING. USUALY BIFILAR WOUND TO IMPROVE FIELD SYMETRY



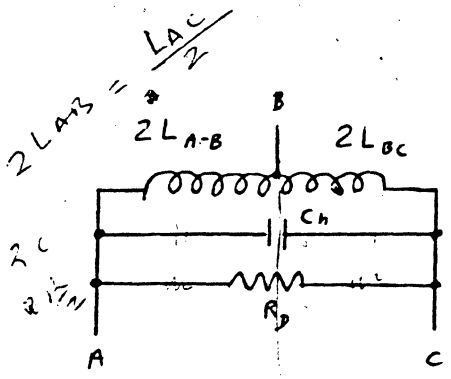


FIG 5.13 A

CENTER TAPPED HEAD EQUIVILENT CIRCUIT

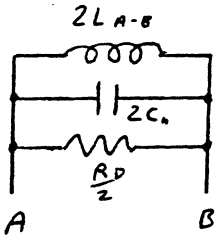


FIG 5.13 B

ONE HALF OF THE CENTER TAPPED HEAD CIRCUIT

Figure 5.8 and 5.11B is that it is easily integrated. Integrated circuits cannot tolerate PNP switches at either high currents or high speeds, therefore, theyare avoided. This last combination is very effective for two terminal thin film heads where the voltage transient is below the base - emitter ~~zones~~ ZENER voltage. Those heads that have large voltage transients must necessarily use a different circuit such as Figure 5.6. The head field for all two terminal heads is proportional to NI. The read back voltage is also proportional to $N\ d\phi/dt$.

There is another class of head circuit that is very popular for reasons to be discussed later under multiple heads. These heads feature a centertap. They are therefore a three terminal device as Figure 5.12.

The circuits used to drive this head are necessarily different. One principle is immediately obvious and that is that the write current flow is into either terminal A or terminal C and out terminal B depending on the direction of the writing flux desired. The head inductance is proportional to N^2; therefore, the number of head turns required for the same N I as previously discussed needs to be double, therefore the inductance is multiplied by four. One advantage is that the read back voltage is twice the previous value. Bandwidth restrictions force the use of a total of N turns therefore the readback voltage is the same but the write current is double to keep the same NI.

The head circuit can be either the full differential, or it can be half where the inductance is equal to $L_{A-C/2}$ (EQ 5.14) as can be seen by

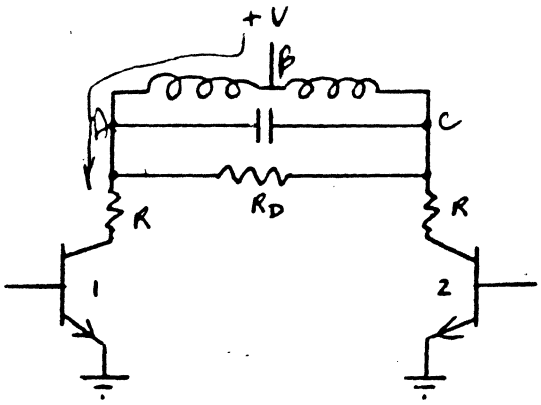
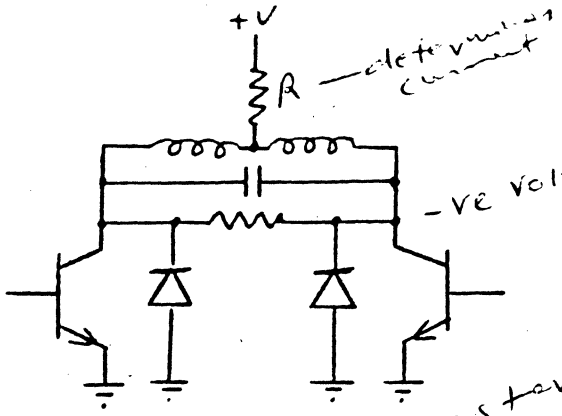


FIG 5.14 A



Disaster
Not Recommended

FIG 5.14 B

ALTERNATE FORMS OF A SATURATED SWITCH DRIVER FOR A CENTERTAPPED HEAD

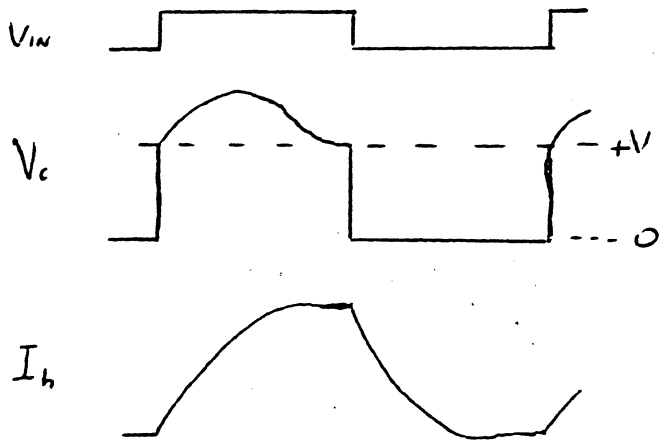


FIG 5.15 A

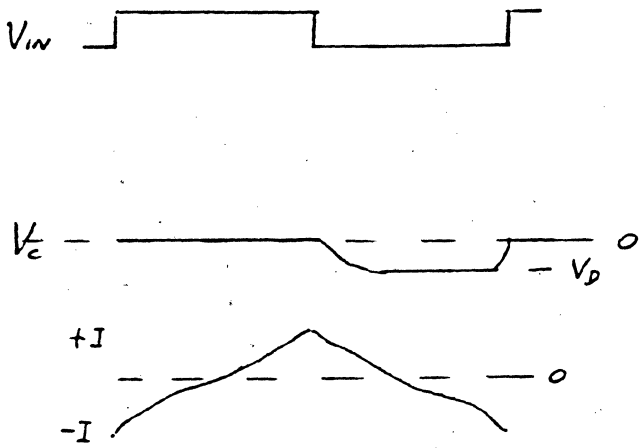


FIG 5.15 B

$$I_h = \frac{V_o - V_{CE\,sat}}{R_h + R}$$

(I_h = steady state)

$I_{AB} = 2 \cdot I_{AC}$ for same flux ($NI = \phi$)

$L_{A-B} + M_{B-C}$ (EQ 5.15) where M_{B-C} is the mutual inductance of the section B - C reflected into A - B. The capacitance for the half equivalent is twice the value of the differential capacitance. The damping resistor is half. All this is shown in Figure 5.13 A and B.

$$L\ Total = L_{A-C} = L_{A-B} + M_{B-C} + L_{B-C} + M_{A-B} \quad \text{(EQ 5.16)}$$

$+L_{A-B}$

Either circuit will yield the correct results when used in equation 5.10.

The circuits that are used are discussed below.

The first circuit is the saturated switch version as shown in Figure 5.14A and B. In Figure 5.14A the DC current is established from EQ 5.17.

$$I_{h_{DC}} = \frac{V - Vsat -}{R + \frac{R_h}{2}} \quad \text{(EQ 5.17)}$$

Worse case values can be assigned that give the range of currents over production runs. Note that the current I is only passing thru half of the head windings when calculating the current for the field required. This circuit is only useful where the storage time is acceptable.

The damping resistor R_D is not affected by the series current determining resistor R in contrast to that of the two terminal head circuits of Figure 5.4.

The voltage excursions on the collector are the same due to twice the current. No commutating diodes are required as the voltage on the collectors never go below ground.

For this circuit, though, the collector - emitter voltage breakdown must be greater than twice the +V supply.

The circuit of Figure 5.14B is different. It also suffers from storage time in the switching transistors, but the voltage waveform is different even though the DC value is identical to EQ 5.17.

It will be noticed that the collector voltage goes below ground, while the second one goes to ground. This requires commutating diodes, also a second look at the equivalent circuit. The commutating diode places both ends of the head at near ground forcing a head equivalent circuit of just a series R_h and the inductance L_T for the duration of the conduction of the diodes. The time for rise during this period is essentially $^{L}A\text{-}C/R_h$ which can be very long. When the transient voltage reduces as the change in current drops, then the circuit reverts to the standard parallel RLC of Figure 5.13A or B. Obviously this is not a desirable circuit.

The most popular circuit is shown in Figure 5.16. Here the full speed can be achieved but at the cost of transistor power dissipation. The voltage V is chosen to keep the negative transient voltage at the collectors above the input Vin +. The damping resistor is chosen to satisfy EQ 5.17 for a zeta of 0.95.

$$\zeta = 0.95 = \frac{1}{2W_hRC} \qquad \text{(EQ 5.17)}$$

The waveforms are shown in Figure 5.17. Notice the collector voltage relationship to the base voltage marked as 'margin' also the peak voltage to the $\overline{Vw}$ level that must be within the V_{CEO} breakdown voltage, (collector to emitter).

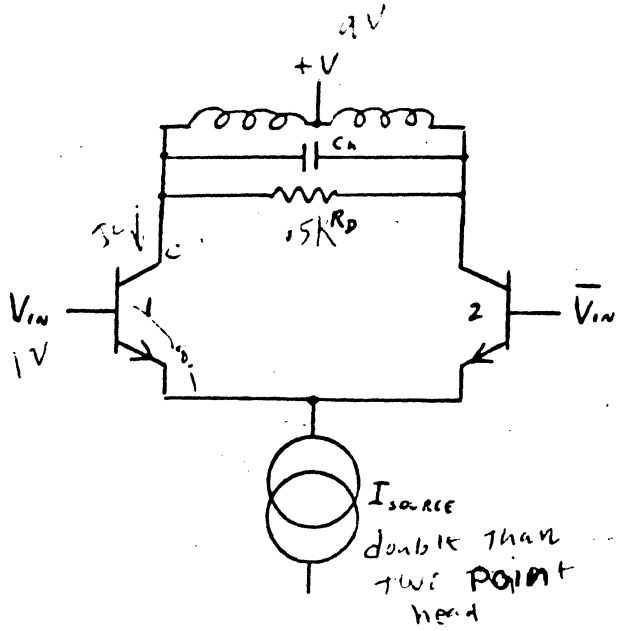


FIG 5.16

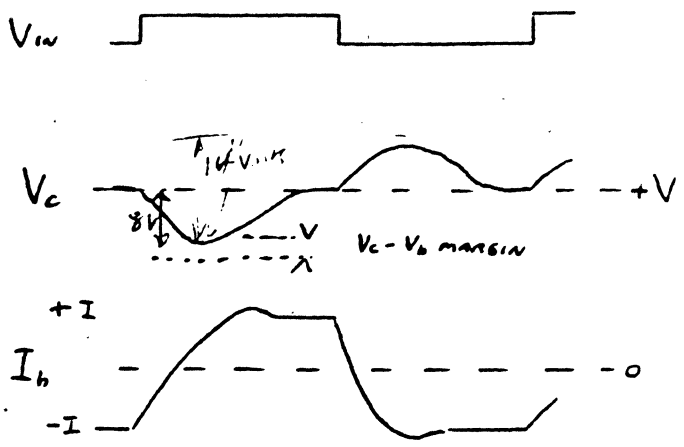


FIG 5-17

PREFERED DIFFERENTIAL DRIVER FOR A CENTERTAPPED HEAD

INPUT, COLLECTOR, CURRENT WAVEFORMS

$$(V_+ - V_D)\cdot I_C + V_D I_b = P_{W_{Trans}} \quad (1)$$

$$5V \times 80mA = 640\ mW$$

$$\frac{2}{3}\cdot 640\ mW$$ — for MFM

$$I_h = I_S(\cdot\alpha) = I_S\cdot\frac{\beta}{1+\beta} \qquad I_{RD} = \frac{16\ Volts}{.5K} =$$

Three terminal (single) allows head switching for one wire driver and preamplifier

2) NI = constant

$L \propto N^2$

$$\text{Voltage} = V_c = \frac{I\cdot L\cdot\frac{1}{\sqrt{LC}}}{\sqrt{1-\zeta^2}} = \frac{I\sqrt{L/C}}{K} = \frac{NI}{K}\left(\frac{N}{\sqrt{C}}\right)$$

For a P maximum we use EQ 5.18.

$$P_{MAX_{TX}} = \frac{(Vcc\ MAX - V_{b,\ Min_+} + Vbe\ MAX)\ I\ source\ MAX}{} \quad \text{(EQ 5.18)}$$

We may divide power by two only if the switching signal has no dc component. If the DC average of the input waveform is not zero, then some other factor must be used. Its value will lie between 1 and 2 depending on the asymmetry. Another consideration is the length of time one transistor is conducting. This is due to the thermal lag of the transistor structure. For slow waveforms the power dissipation must be considered as the full value even if there is no dc component of the input signal. Localized heating of the junction may exceed the allowable junction temperature.

The junction temperature for all circuits can be calculated using the transistor thermal resistivity value published for that device.

$$T_J = (R_{JC} + R_{CA})(^oC/W)(Pw\ Max)(Watts) + T_A\ Max \quad \text{(EQ 5.19)}$$

Where R_{JC} is the thermal resistance in ^oC/Watt from junction to case,

RCA is the thermal resistance in ^oC/watt from case to ambient air

PW_{Max} is the power dissipation in watts, and T_A is the ambient maximum temperature in ^oC.

For best reliability the junction temperature, T_J, should not exceed 100^oC even though a device may be rated to 125^oC or even 150^oC. The temperature rise is the first half of the equation. It may be modified by adding a heat sink which alters the parameter R_{CA}. Nothing can be done for R_{JC} though.

Air flow also enters into R_{CA} value and is usually published as a family of curves. For writing circuits the power dissipation is fairly high in comparison to standard circuits particularly as large currents are required in high inductance circuits. The requirement to keep the collectors out of saturation forces higner collector voltages.

$$I_h = I_{Source}\left(\frac{\beta}{1+\beta}\right) \qquad \text{(EQ 5.20)}$$

BASE DRIVE

A further consideration is the base drive. The impedance of the base driving circuit needs to be kept low in order to reduce the Miller effect feedback. If the input impedance is high the head voltage transient will be capacitively coupled to the base circuit possibly forcing the transistor back out of conduction and the opposite transistor back into conduction. Figure 5.18 illustrates this effect where the dotted line represents the feedback thru Miller capacitance. The transistor β also requires consideration when designing the base driver circuits. It also affects the current thru the head and the current source. All the circuits previously mentioned that are driven from current sources will have these limitations. Those that are saturated switches will have only the Miller effect to contend with. Equations 5.21 and 5.22 describe these effects.

$$I_h = \text{I source} - \frac{Ic}{\beta} = I_{source}\left(1 - \frac{1}{1+\beta}\right) \qquad \text{(EQ 5.21)}$$

$$\text{V base} = \text{Vin} - \left(\frac{\text{Vh Transient}}{\frac{1}{CS} + \text{Rin}}\right) \text{Rin} \qquad \text{(EQ 5.22)}$$

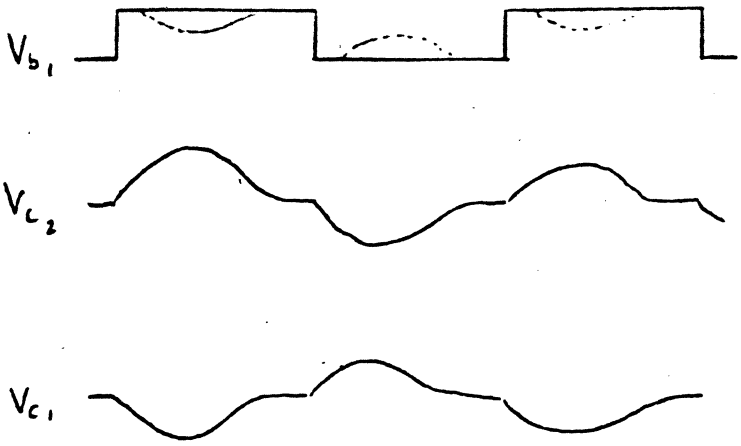


FIG 5·18 MILLER FEEDBACK FROM HEAD TRANSIENT

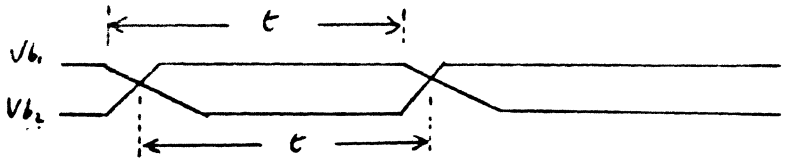


FIG 5·19

BASE RISE-FALL UNSYMETRY EFFECTS SHOWING FINAL SYMETRY

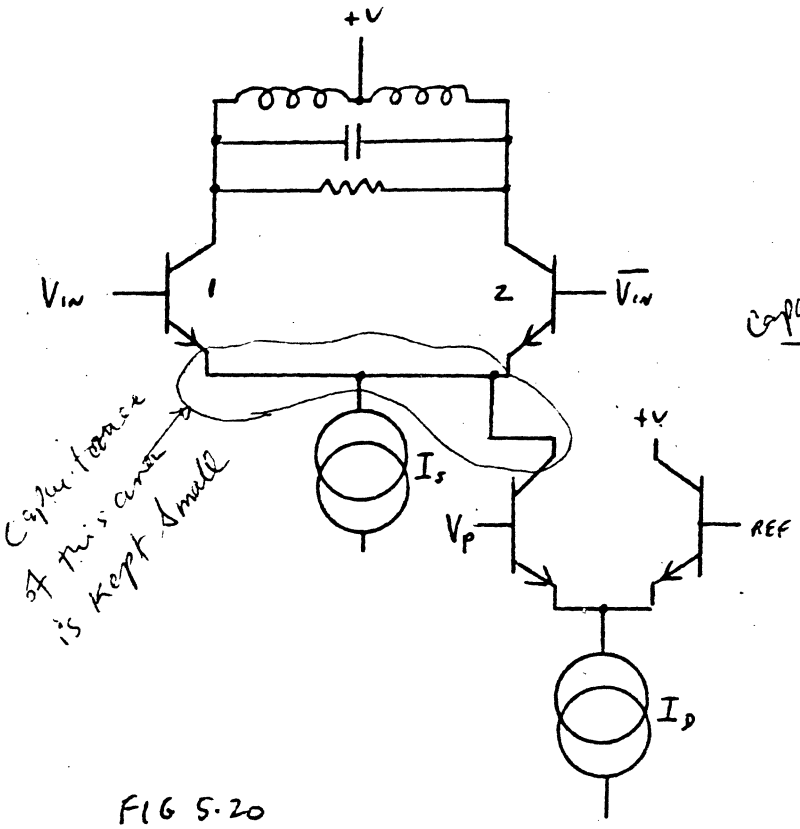


FIG 5·20

CURRENT PULSING

A METHOD TO IMPROVE RISE TIME USING HIGH INDUCTANCE HEADS

Used but not often

allows to go higher density for same head

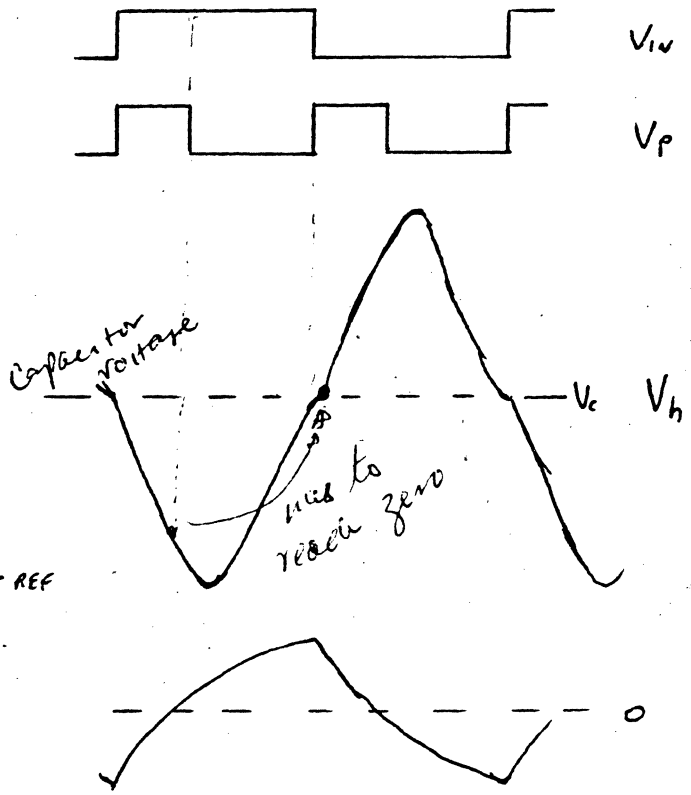


FIG 5·21

WAVE FORMS DUE TO CURRENT PULSING

TIME DOMAIN SOLUTION

The time domain solution for the head voltage and the head current as described in EQ 3.11 and EQ 3.14 are given in EQ 5.23 and 5.24 respectively.

from chapter-3

$$\mathcal{L}^{-1} V_{(s)} = \mathcal{L}^{-1} \frac{I(s)\, L\, Wn^2}{(S^2 + 2\zeta WnS + Wn^2)} = \frac{IL\ Wn}{\sqrt{1-\zeta^2}}\, e^{-\zeta Wnt} \sin(\sqrt{1-\zeta^2}\ Wnt) \quad \text{(EQ 5.23)}$$

$$\mathcal{L}^{-1} I_{(s)} = \mathcal{L}^{-1} \frac{I(s)\ Wn^2}{S(S^2 + 2\zeta WnS + Wn^2)}$$

$$= I\left[1 - \frac{1}{\sqrt{1-\zeta^2}}\, e^{-\zeta Wnt} \sin\left(Wn\sqrt{1-\zeta^2}\,t - \tan^{-1}\frac{\sqrt{1-\zeta^2}}{-\zeta}\right)\right] \quad \text{(EQ 5.24)}$$

It may be noticed that most write driver circuits bases are driven differentially. This type of input is forgiving of any slight unsymmetry in the input waveform as long as the unsymmetry is repeated on each input. This is illustrated in Figure 5.19 where the crossovers are not occurring at the centerline due to slope unsymmetry. Such unsymmetry may be caused by variations in rise and fall times. A typical switching input swing requirement for differential unsaturated switches is about 1.0V. This value guarantees total cut off of the opposite transistor. We assume that 0.4 volts Vbe are required to bring a transistor into a slightly conductive condition and by 0.7 to 1.0 volts the transistor is completely on. When using transistors with larger Vbe sat voltages, they need to be provided larger input swings in order to correctly switch them.

WRITE VOLTAGE CONSIDERATIONS

Since the write voltage transient forces the collector voltage to be high to accommodate the swing, we might profitably look at what we can do to limit the total swing. Restating EQ 5.23 again, we can ignore the time varying terms and just look at the magnitude portion as shown in EQ 5.25.

$$V_h \simeq \frac{I\ L\ Wn}{\sqrt{1 - \zeta^2}} \qquad \text{(EQ 5.25)}$$

by substituting $KN^2 = L$ and ignoring the damping term in the denominator as ζ is a constant for all write system = 0.95, we get:

$$V \simeq \frac{I\ KN^2}{\sqrt{KN^2C}} = \frac{NI\sqrt{K}}{\sqrt{C}} \qquad \text{(EQ 5.26)}$$

Now we see that NI is proportional to the flux required to saturate the media. For a given head - media interface, NI is a constant. If we change the flying height and/or the gap length in order to reduce the current then we can reduce the transient voltage, but just reducing I forces N to be increased to keep the same saturating flux which accomplishes nothing. The only other alternative is to either improve the head efficiency by reducing the throat height provided we can do so without saturating the core pole tip or increasing the capacitance. This latter will lower Wn which increases the rise time which may be excessively detrimental.

WRITE PULSE SHAPING

One way to improve the rise time in a head that requires a large number of turns, such that the Wn is lower than desired, is to pulse the current source in time with each switching edge. The effect is to force the head current to

rise towards the higher value and then just before the required current value is reached to drop the current source value to its normal value. A penalty for doing this is that there is a voltage across the head capacitance remaining that needs to be removed before the head current can settle to its final value. The width of the pulse will require careful control in order to orchestrate the desired result. A circuit for doing this is shown in Figure 5.20, along with the waveforms in Figure 5.21.

As can be seen the voltage transient is very large. The rise is fast during the pulse then it reverts to a negative slope until the transient is over. The equation takes the form of two parts where the

$$V_h = \left(\frac{Im}{S_{A-B}} - \frac{I_D}{S_{B-C}} \right) Z_h \qquad \text{(EQ 5.27)}$$

notation is for two step functions at differing times and $Im = I + I_D$ (EQ 5.28).

B. PRE DRIVER CIRCUITS

The circuits used to drive the Write Drivers can range all the way from a direct connection to the Flip-Flop to a intermediate amplifier or switch that is used to establish the bias levels required and/or the base current requirements.

For the saturated versions the driving circuit need only provide the base current required and a voltage output swing capable of turning the driver transistors on and off. Standard T^2L logic blocks are usually sufficient. If higher base current is required an open collector output device can be used efficiently. An example of both is shown in Figures 5.22 A, B, and C.

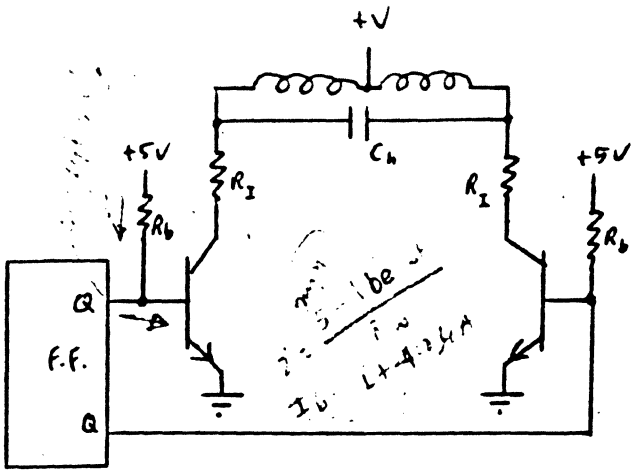


FIG 5.22 A

T^2L, SATURATED DRIVERS

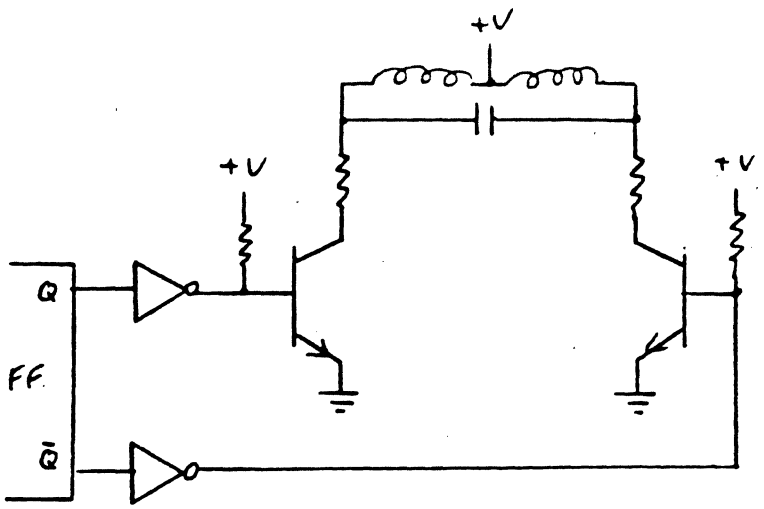


FIG 5.22 B

T^2L BUFFERED . SATURATED DRIVERS

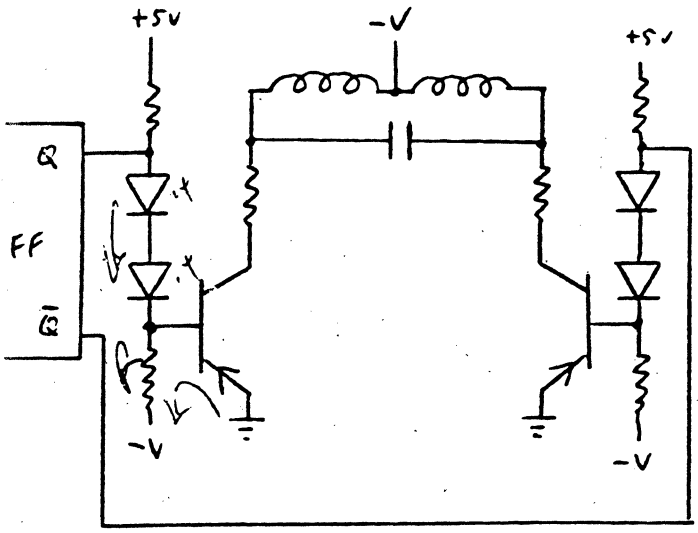


FIG 5.22 C

PNP SATURATED DRIVERS
GROUNDED EMITTERS

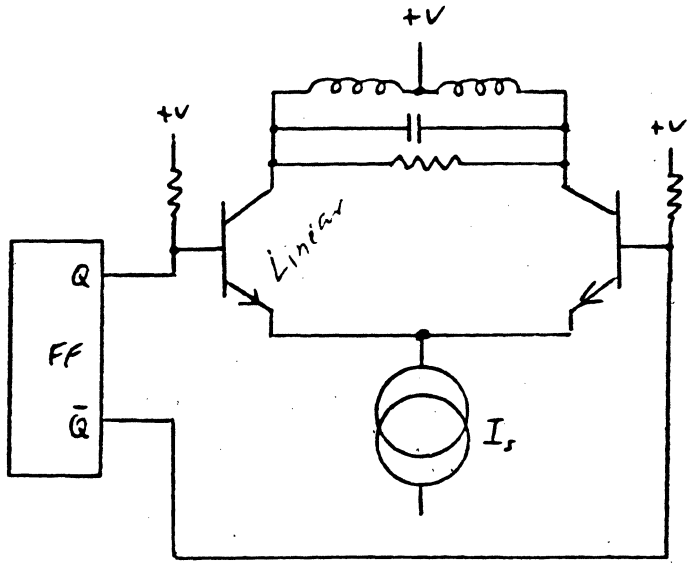


FIG 5.23 A

NON SATURATING DRIVERS WITHOUT BUFFERS NPN

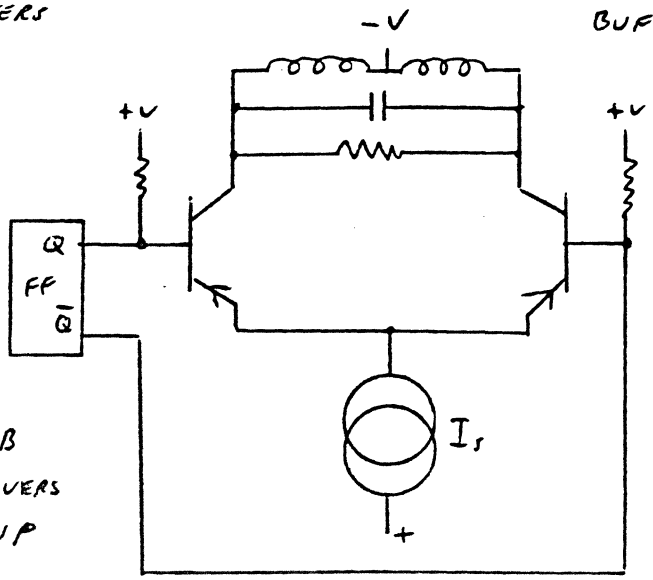


FIG 5.23 B

NON SATURATED DRIVERS WITHOUT BUFFERS PNP

- Increase voltage
- Lower Z
- minimize noise effect

B. PRE DRIVER CIRCUITS

For the non saturating switches either a T^2L, ECL or a voltage translating switch can be used. If T^2L logic blocks are to be used, care must be taken to minimize the Miller feedback transients during the up level by maintaining low impedances or by using pull up resistors as required in the saturated version. ECL logic has the advantage of low impedance and a voltage swing sufficient to switch the driver transistors.

If the write drivers are PNP and the head is tied to a negative voltage, then the type requires no base translation as shown in Figure 5.22C but may be connected directly if sufficient base drive is supplied. If the head centertap is grounded, then the bases of the write drivers need to be driven from a potential sufficient to keep the driver transistors out of saturation. This function is best performed by a current switch unless the storage time of saturated switches and their voltage swing can be tolerated.

With the current switch Pre Driver both the impedance and the voltage swing requirements can be designed in. Figure 5.24 shows an NPN driver with a PNP Pre Driver. The -V ref is chosen to keep the Write Driver collectors (3,4) out of saturation during the head transient. The bases of the Pre Driver can be driven directly from either T^2L or ECL logic blocks. This kind of circuit lends itself to large separations between the Pre Driver and the Write Driver wherein the impedance can be that of an interconnecting cable for termination purposes. The current in the Pre Driver needs to be large enough to produce the Write Driver base drive voltage swing required. When this circuit is worse cased both the Write Driver turn on and turn off requirements must be met but also the Miller feedback from the head transient must be allowed for. Lastly, the Write Driver base breakdown voltage Vber

Always use differential signals

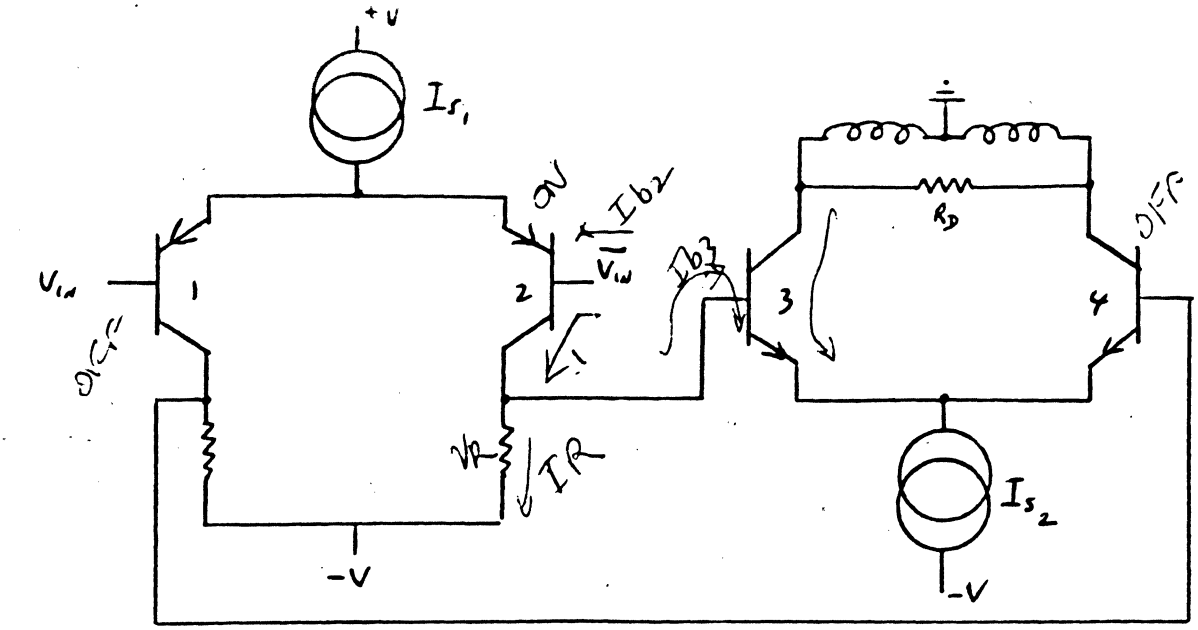


FIG 5.24

NON SATURATED DRIVER WITH NON SATURATED PRE DRIVER SWITCH

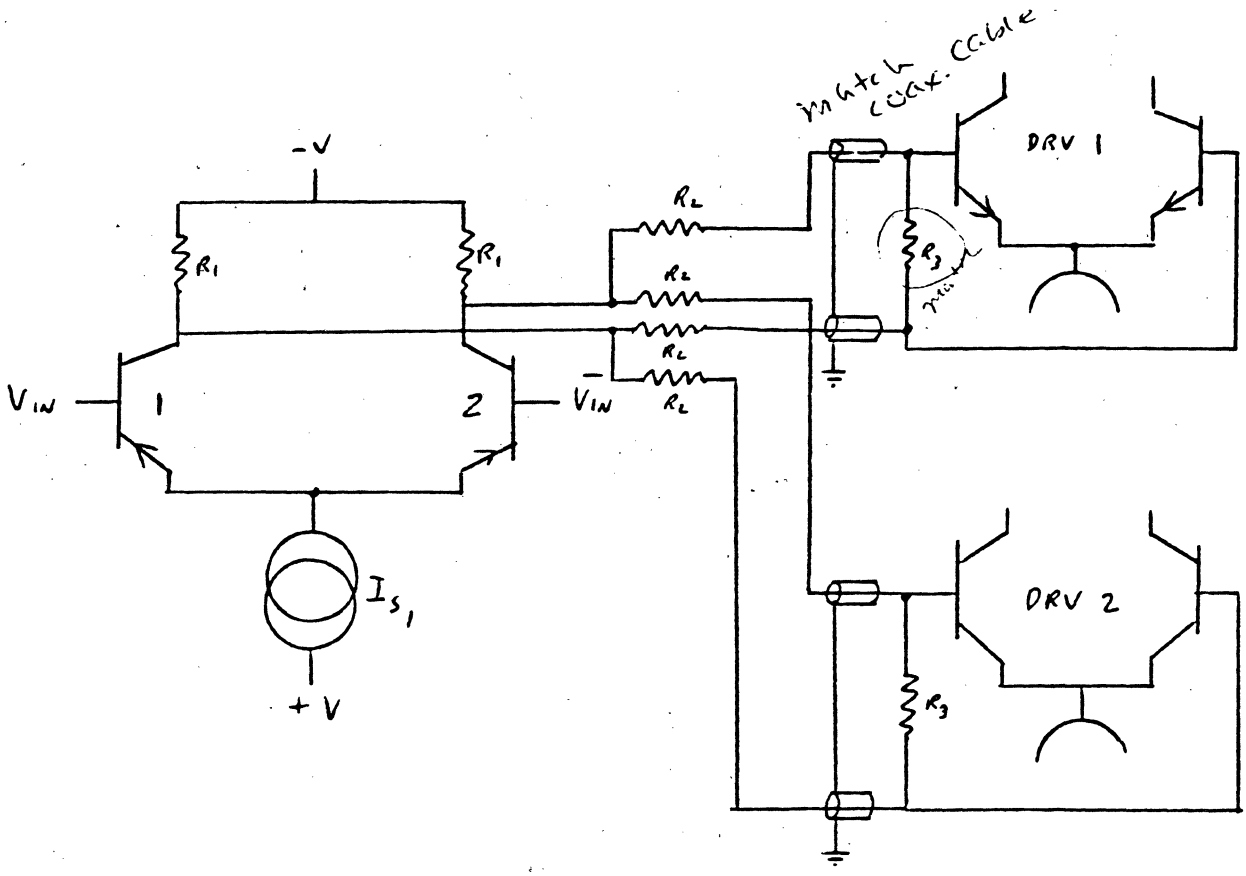


FIG 5.25

A SINGLE PRE DRIVER DRIVING TWO MATCHED CABLE PAIRS TO TWO DRIVERS.

B. PRE DRIVER CIRCUITS

must not be exceeded. These equations are complicated by the base current requirements of the Write Driver. A set of equations follows.

$$\Delta V_{bW.D._{min}^{4}} = \left[Isource_{min}^{1}\left(1 - \frac{1}{\beta_1 min+1}\right) - Isource_2 \frac{1}{\beta_4 min+1} \right] R_{min} \quad \text{(EQ 5.29)}$$

$$\Delta V_{bW.D_{max}^{4}} = \left[Isource_{max}^{1}\left(1 - \frac{1}{\beta_1 max+1}\right) - Isource_{max\,2} \frac{1}{\beta_4 max+1} \right] R_{max} \quad \text{(EQ 5.30)}$$

$$P_{TX_2}max = I_{SOURCE_1}\left(\frac{\left(V_{REF_{max}} - \Delta V_{bWD_{4_{MAX}}} + V_{in} + V_{be_{max}}\right)\beta_{2_{max}}}{\beta_{2_{max}}+1} + \frac{V_{be}}{1+\beta_{2_{max}}} \right) \text{(EQ 6-31)}$$

$$\left| \Delta V_{bWD_4 max} \right| < Vber \quad \text{(EQ 5.32)}$$

If more than one Write Driver is desired to be connected to a common Pre Driver, then due consideration needs to be paid to capacitance as associated with the RC of the Pre Driver load. One problem when driving long cables between the Pre Driver and the Write Driver is that both ends must be terminated in the characteristic impedance of the cable in order to absorb the transients associated with both the Pre Driver output and the Miller feedback of the Write Drive. This will ensure quiet operation with no reflections. A network can be designed to drive multiple cables with their characteristic impedance at both ends. A circuit for doing this is shown in Figure 5.25

B. PRE DRIVER CIRCUITS

Symmetry shows that half the impedance of a twinaxial cable or the impedance of a coaxial cable must be used for Z_0.

$$Z_0 = \frac{R_3}{2} = R_2 + \frac{R_1(R_2 + \underline{Z_0})}{R_1 + R_2 + \underline{Z_0}} \qquad \text{(EQ 5.33)}$$

The voltage swing at the bases of the Write Driver will be a function of the two current sources as before (EQ 5.29, -30) but now R needs to be modified to include the effects of the network. This is best illustrated by considering Figure 5.26 when only one Write Driver is activated and the second is idle.

$$V_A = I_{source_1}\left(\frac{\beta_1}{1+\beta_1}\right)\left[\frac{(R_2 + \frac{R_3}{2} + R_1)R_1}{R_2 + \frac{R_3}{2} + 2R_1}\right] \qquad \text{(EQ 5.34)}$$

$$\Delta V_{3-4_A} = \frac{V_A\left(\frac{R_3}{2}\right)}{R_1 + R_2 + \frac{R_3}{2}} \qquad \text{(EQ 5.35)}$$

This is the base$_3$ to base$_4$ voltage with no base current effects from the Write Driver.

C. CURRENT SOURCES

The current sources considered are those used to generate the write current. Several design requirements must be met. First the current source must be stable with temperature and supply voltages. Second the manufacturing tolerances must be minimized. Two circuits are considered here. The first is the zener controlled emitter degenerative circuit of Figure 5.27. This is shown as a negative current source.

For this circuit to function correctly the voltage on the collector of Q1 must always be more positive than its base. This prevents saturation. When the collector is connected to the Write Driver this means that the most positive base of the Write Driver must be at least two Vbe drops above the base of Q1. Notice that the Diode D1 is added to compensate for the Vbe of Q1 over temperature. This is only true if the diode characteristics of both Q1 and D1 are the same and the currents are the same. Doing this is rather wasteful so a compromise is made allowing a degree of temperature compensation. The zener D2 is chosen for a sharp knee or at least a fairly flat zener potential around the maximum and minimum currents expected thru R1. If the diode drop V_{D1} is the same as the Vbe at the operating current then the current source is essentially:

$$\frac{Vz}{R_2}\left(1 - \frac{1}{B_1+1}\right) \qquad \text{(EQ 5.36)}$$

Since this is fairly ideal we need to consider the whole circuit. The circuit includes the T^2L interface and Q2.

C. CURRENT SOURCES

First we will saturate Q2 for a maximum of 25 ma. This will ensure that the zener will be operating well past its knee.

$$\frac{25.ma}{\beta_2} \geq I_{b_{2_{min}}} = \frac{+Vmin - Vbe_2max - Vsat_4}{R_3max} - \frac{Vbe_2max}{R_4min} \qquad \text{(EQ 5.37)}$$

With Q2 saturated we can proceed to the input of Q1.

$$I_{Z_{min}} = \frac{(+Vmin - (-V\ min) - V_{D_1}Max - V_Z\ Max - V_{CE_{1\ SAT_{MAX}}} - I_{Q_1}\left(\frac{1}{\beta_1\ Min}\right)}{R_1\ Max} \qquad \text{(EQ 5.38)}$$

The voltage at the base of Q_1 will be, realtive to the minus supply, as follows if we ignore the fact that the first term Vz_{min} is contrary to Vz_{max} used to calculate Iz_{min} as given in EQ 5.38.

$$V_{b_1} = Vz_{min} + R_{Z\ min}(I_{Z\ min}) + V_{D_1} + R_{D_1 min}(I_Z)_{min} \qquad \text{(EQ 5.39)}$$

Therefore the current source will be:

$$I_{source\ min} = \frac{Vb_{1min} - Vbe_{1max}}{R_{2max}}\left(\frac{\beta_{1min}}{1+\beta_{1\ min}}\right) = I_{Q_1} \qquad \text{(EQ 5.40)}$$

If this current source were to feed the Write Driver of Figure 5.16, then the actual head current would be reduced by the base current drawn by the Write Driver as indicated in EQ 5.20.

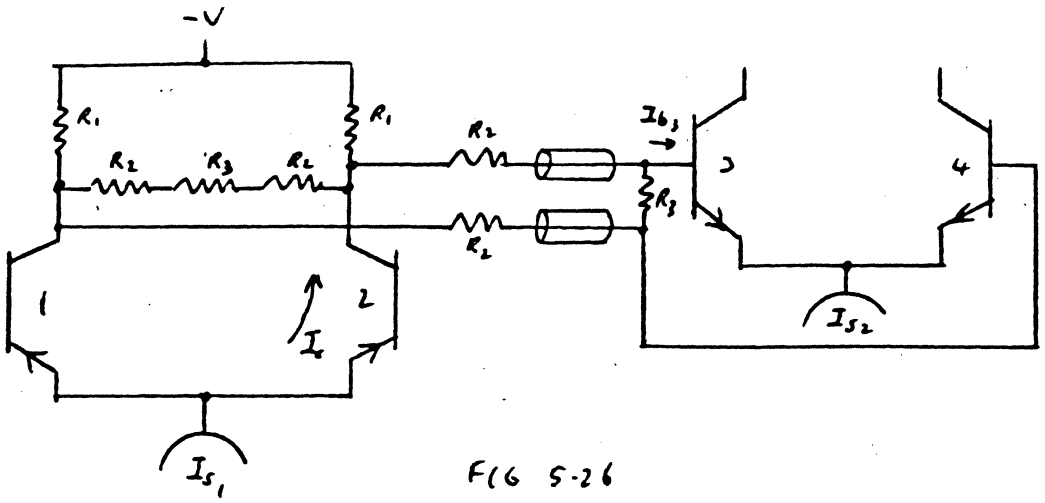


FIG 5.26

SIMPLIFIED CIRCUIT WHEN 2nd DRIVER IS OFF

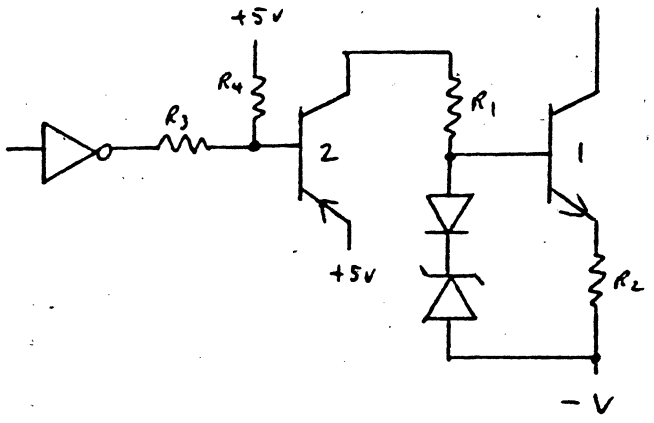


FIG 5.27 CURRENT SOURCE NEGATIVE

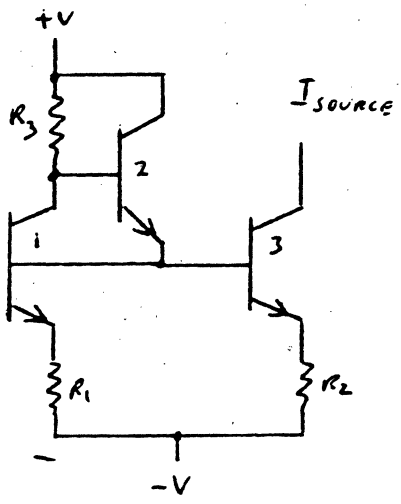


FIG 5.28 A MIRROR NEGATIVE CURRENT SOURCE

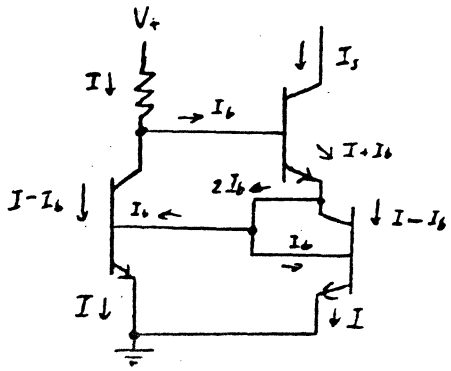


Fig 5.28 B

WILSON MIRROR

C. CURRENT SOURCES

The maximum write current can be found as follows:

$$I_{z_{max}} = \frac{+V_{max} - (V_{max}) - V_{D_1 min} - V_{z\ min} - V_{CE SAT_2 min}}{R_{1min}} - I_{Q_1}\left(\frac{1}{\beta_1\ max}\right) \quad \text{(EQ 5.41)}$$

$$V_{b_{1max}} = Vz_{max} + Rz_{max}\ (I_{z\ max}) + V_{D_1 max} + R_{D1max}\ (I_{z\ max}) \quad \text{(EQ 5.42)}$$

$$I_{source_{max}} = \frac{Vb_{1max} - Vbe_{1max}}{R_2 min}\left(\frac{\beta_{1\ max}}{1+\beta_1\ max}\right) \quad \text{(EQ 5.43)}$$

The manufacturing tolerance (more than worse case) is then:

$$\Delta I_{source} = I_{source\ max} - I_{source\ min.} \quad \text{(EQ 5.44)}$$

It should be noted that several factors can be controlled by choosing both the zener voltage large compared to Vbe_1 and V_{D1} and using a temperature compensated zener with 1% or better resistors for R_2. Also closer tolerances on the zener voltage Vz and the zener impedance Rz.

Going back to the saturation curves of Figure 4.8, we can see reasons for a small delta I source when we are forced to use thick media where the saturation curve rolls off. If we are using *thin* media where the saturation curve is flat above saturation then we can use cheaper wider tolerance parts for the current source.

C. CURRENT SOURCES

We can go thru a similar procedure if we choose a positive current source.

The second type of current source is the current mirror. This circuit finds favor if the whole is to be integrated on single chip. The circuit of Figure 5.28 is a simple Wilson current mirror. The requirements for stable current are the value of R_3 and the matching of R_1, R_2, Q_1 and Q_2. Often the current thru Q_1 is multiplied by the junction area ratios of Q_1 and Q_3 with due consideration for the periphery of the emitters. The function of Q_2 is to supply base current to Q_1 and Q_3 bases at the cost of the error I_{b_2} .

$$I_{b_2} = \text{I Error} = \frac{(I_{b_1} + I_{b_3})}{\beta_2} \qquad \text{(EQ 5.45)}$$

Current multiplication can also be achieved by varying the relative value of R_1 and R_2. Since the resistors in integrated circuits typically have a tolerance of 25%, this means that some other resistor type must be used for R_3 or it can be laser trimmed as one manufacturer has done.

Power dissipation for both types need to be calculated to ensure the junction temperature is not exceeded nor the devise forced into second breakdown. The output voltage is simply the conducting base voltage of the write driver less one V_{be} or V_C max.

$$P_{source_{max}} = \frac{\beta(Vc\ max - Ve_1\ min)I_s\ max}{\beta_{max} + 1} + \frac{I_s\ Vbe\ min}{\beta_{max} + 1} \qquad \text{(EQ 5.46)}$$

D. DATA

In most recording applications the Write Data is received on multiple lines which must be converted to serial form before writing on the media. This is easily handled by a parallel to serial converter under the control of the write clock. The output of the shift register, or serial data is then changed to pulses if the data is true, or no pulses if the data is false. These operations are shown in Figure 5.29 which includes a means of providing alternations of the input lines to the Write Pre Driver if used and/or the Write Driver. The alternations in input level provide the current switching which in turn provides the flux changes of the recording.

The function of the 'and' block A can be modified to suit the code used for recording by the use of an encoder. These circuits will be covered later when we discuss codes. There is one other function that can be included in the Block A and that has to do with Pre Compensation. Consider for a moment the transition density curve Figure 4.3 and the interaction between transitions that cause the reduction in amplitude and pulse shift. When writing a data pattern there is not a constant density but discreet changes in density depending on the data content and the code used. The plot for bit shift or pulse shift included in the density curve was achieved by measuring the peak spacing between two adjacent transitions separated by long areas of no transitions. This type pattern can also occur in a data stream for some codes. If we were to write the transition in such a way that a pulse that is shifted early in time compared to its true position could be compensated for by writing the transition late. Similarly a pulse that is shifted late can be corrected by writing it early. Thus when this signal is read back the pulses are very nearly back to their true position. This is know as Pre Compensation. When a head - media choice

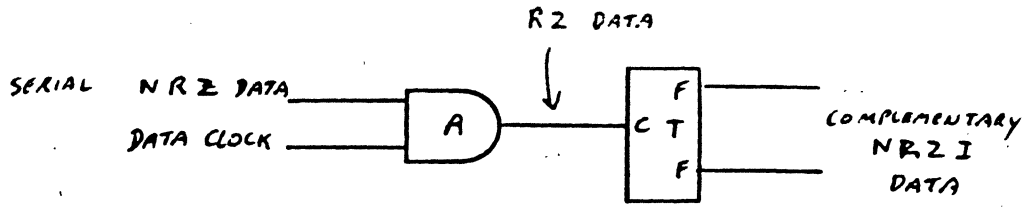


FIG 5.29 A

NRZ TO NRZI CONVERTOR

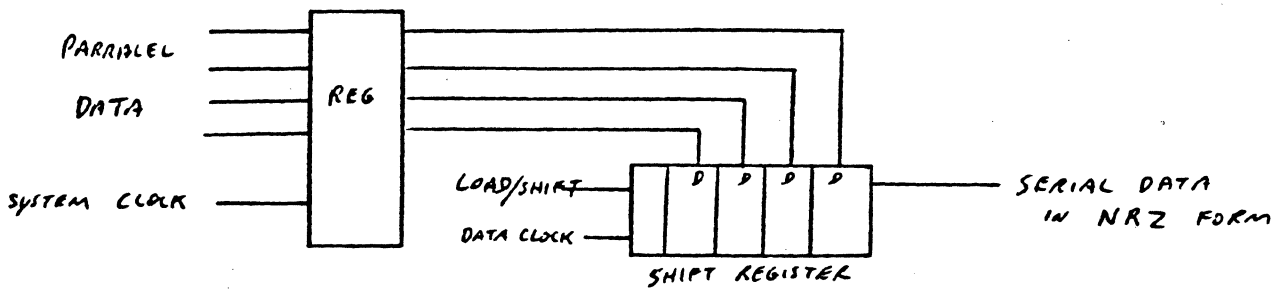


FIG 5.29 B

PARRALEL TO SERIAL CONVERTOR

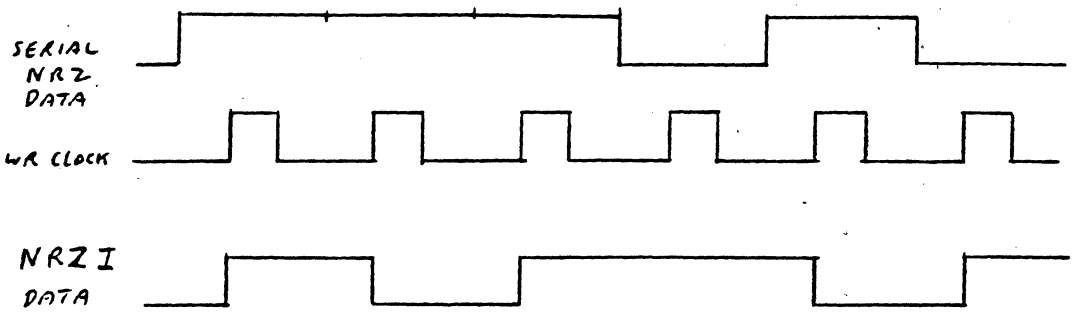


FIG 5.29 C

DATA WAVEFORMS NRZ TO NRZI

is made for a particular machine design, compromises can be made that can increase the density beyond that safely obtainable by using Pre Compensation. Generally speaking for the FM codes Pre Compensation is advisable below a resolution of 0.7 and definitely required below 0.6. The subject of codes is discussed later. The circuits chosen to implement Pre Compensation must consider any parallel delays in the logic paths as any unsymmetry there will write bit shift. This can best be achieved by using logic gates from the same clip for all parallel functions. As we begin the design we need to determine the number of discreet shifts required. These depend on the code used and the transition density chosen. For example, one code might exhibit two levels of bit shift, $\pm$ 5 and $\pm$ 9ns. These are sufficiently far apart that it would be expedient to design a system that implemented the shifts. A truth table then needs to be generated that describes the pattern and the expected shifts. We will leave this function to the chapter on codes as the implementation of the code is done simultaneously. This will suffice for the present.

We have now completed the blocks used for writing with a single head. There were many blocks described for each function. How they are put together and which block is chosen depends on the power supply, biasing, bit timing vs. circuit delays such as saturated transistors, intended cost goal, and the head - media interface magnetically and electronically.

READ CIRCUITS

Referring to the block diagram Figure 5.1, locate the Read Pre Amplifier. This particular block determines the basic signal to noise ratio of the machine. It also provides the functions of signal amplification and impedance change. When reading a head signal which is the result of the $\frac{d\phi}{dt}$ of the recorded transitions the windings of the head are connected to the Pre Amplifier. The amplifier also has some input capacitance and some input resistance, Zin. Since we are concerned with a maximum voltage at the Pre Amplifier input for voltage amplifiers, the concept of impedance matching is incorrect. We must, however, properly damp the RLC network as previously discussed such that we have a zeta of 0.7 for a maximally flat bandpass. Now R and C of the head adds appropriately to the Zin and Cin of the amplifier and must be included in the calculations.

Single ended amplifiers, which most engineers are familiar with, have poor common mode rejection meaning that for any ground shift voltage, power supply voltage noise, or magnetic and electric field noise coupled into the signal leads the amplifier will treat them as if they were signal. This is disastrous for high speed magnetic recording. For this reason all wide bandwidth read amplifiers use the differential connection as illustrated in Figure 6.1. Differential amplifiers have excellent common mode signal rejection and common mode power supply noise rejection.

The differential connection itself needs some basic understanding. Head signals are usually referred to in volts, peak to peak, Differential. This means that the voltage across the two inputs or outputs is measured between the two inputs or outputs as a Peak to Peak value. An oscilloscope is the usual measuring instrument. The usual oscilloscope set up is A - B for the two inputs.

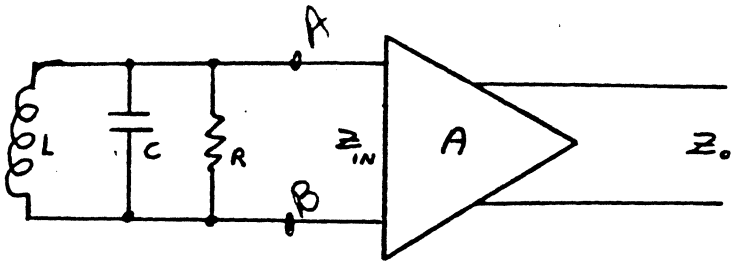


FIG 6.1

READ PRE AMP CONFIGURATION

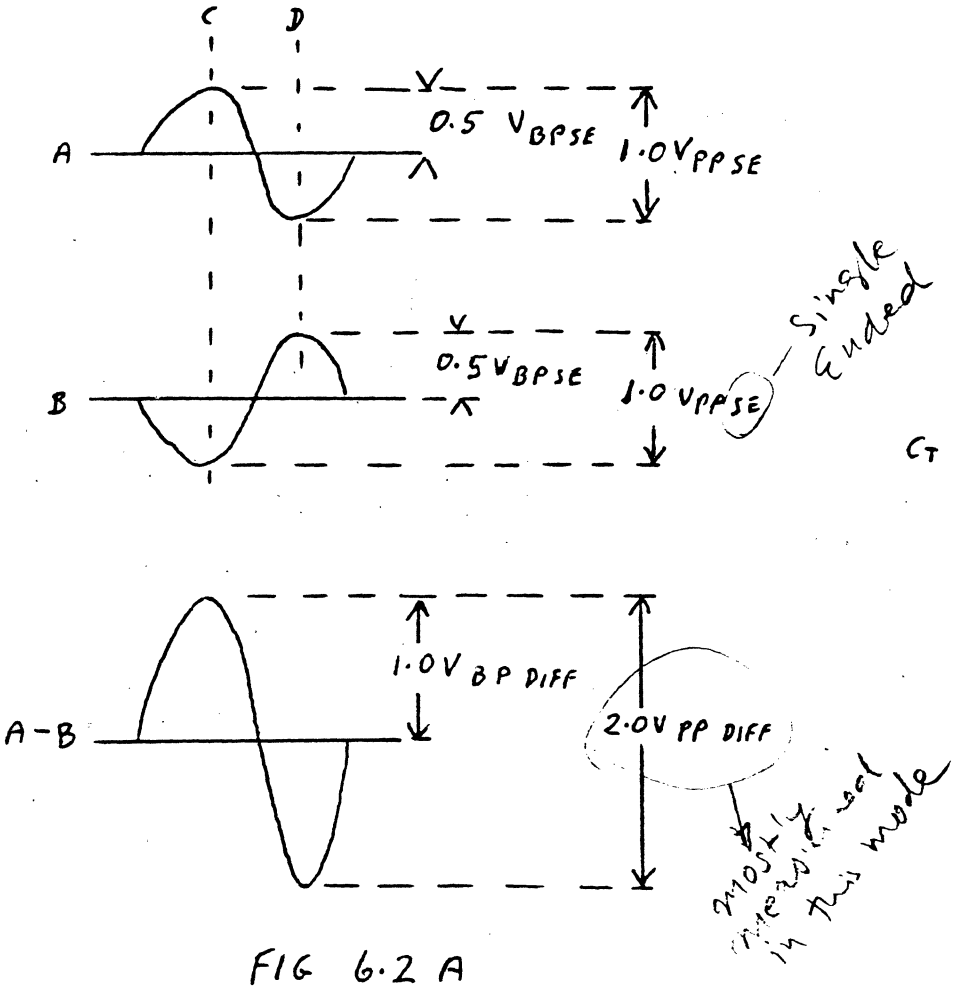


FIG 6.2 A

SINGLE ENDED AND DIFFENTIAL SIGNAL LEVELS

So many ways to measure voltages

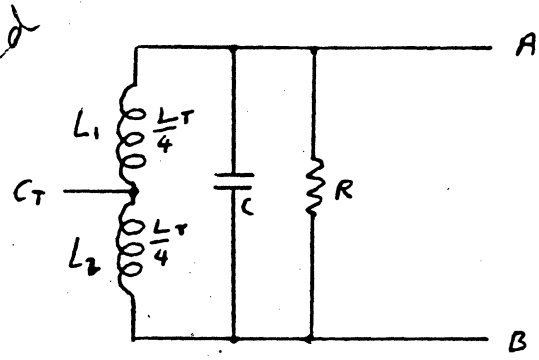


FIG 6.2 B

CENTERTAPPED HEAD

$L_T = L_1 + M_1 + L_2 + M_2$

If we measured 2 mV PP differential signal between points A and B, we would then expect to measure 1.0 mV PP_{diff} between point A and ground also from point B and ground. This is referred to as 1.0 mV PP single ended. (S.E.) The term "differential" means the difference in voltage between terminals A and B. In Figure 6.2A we can see that the voltage difference between terminal A and B at point C is +0.5mV - (-0.5mV) = +1.0mV (EQ 6.1). Similarly, at point D we measure -0.5mV - (+0.5mV)=-1.0mV (EQ 6.2).

The resultant waveform would be a voltage with an amplitude of 1.0mV - (-1.0mV = 2.0mVpp differential (EQ 6.3). We could look at the following relationships.

$$2mV\ PP\ diff = 1mV\ PP\ SE = 0.5mV_{BP\ SE} \qquad (EQ\ 6.4)$$

where S.E. is single ended, and B.P. is base to peak.

We could add to the complexity and say that this signal is 0.707mV RMS Differential or we could say it is 0.3535mV RMS single ended.

With the above background we can now talk about the amplifier itself. The parameters we are most concerned with are high gain, wide bandwidth, low noise, low output impedance, and high input impedance with a differential connection and high common mode signal and power supply rejection.

The input signals are typically in the low millivolt to microvolt range. This immediately requires that the amplifier noise referred to the input must be considerably lower than these levels. For example, we require an amplifier that has a Signal to Noise ratio of +30 db, meaning

$$20 \log \frac{S}{N} = 30\ db \qquad (EQ\ 6.5)$$

READ CIRCUITS

For an expected 1.0 mVPP signal, S, we need to first convert this to .3535 mV RMS differential. The noise limit can then be calculated from

$$\text{Antilog}\ \frac{30}{20} = 31.622 = \frac{S}{N} = \frac{0.3535\ \text{mV RMS}}{N} \qquad \text{(EQ 6.6)}$$

$$N = \frac{.3535\ \text{mV RMS}}{31.622} = 11.17\ \text{micro volts RMS Diff.} \qquad \text{(EQ 6.7)}$$

If the amplifier gain were 100 then we would expect to measure 1.117 mV RMS of noise at the amplifier output. The amplifier input impedance and the source impedance play a dominant role. There are two sources of noise to consider, first the voltage and shot noise, meaning with the inputs shorted together we would measure an output noise equal to this internal noise voltage source times the amplifier gain. The second noise source is a noise current. To develope a voltage we simply multiply this noise times the input circuit impedance. In our case this is an RLC circuit; therefore, we would expect it to vary with frequency. There is a third noise source called $\frac{1}{F}$ noise, but as this is below a few cycles and most magnetic recording occurs at much higher frequencies, we can effectively ignore this noise.

If the head were purely resistive then we could add the two noise sources as the root mean square:

$$K\sqrt{Vn^2 + \bar{I}n^2R^2} = K(\text{effective noise}) \qquad \text{(EQ 6.8)}$$

where K is the gain of the amplifier and R is the resistive head.

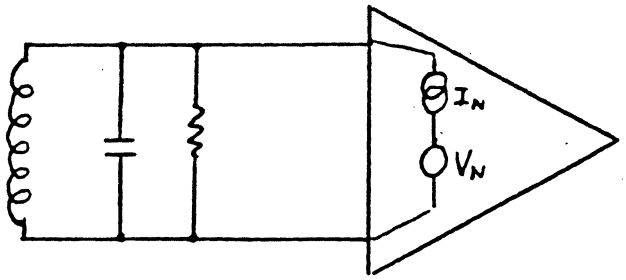


FIG 6.3

AMPLIFIER NOISE SOURCES

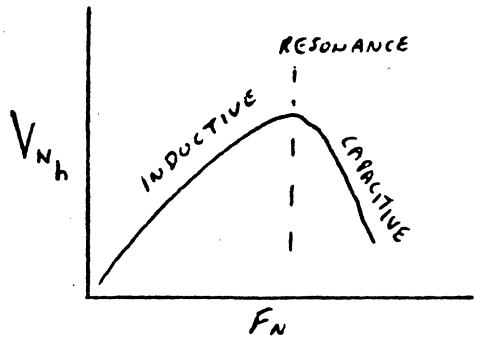


FIG 6.4

V_{N_h} = (HEAD IMPEDANCE)(I_N)

This becomes complicated as we use the true head impedance. The noise is no longer white noise, but is coloured by the reactive head, Figure 6.4. Generally we connect the head and measure the noise as a total noise instead of trying to separate the various types of noise.

We may choose a commercially available Pre Amplifier or we may design our own. The Fairchild μα733 is one that has desirable characteristics. Flexible gain, reasonable input impedance, fairly low output impedance, very good Common Mode Rejection Ratio and about 12 μv of noise measured in 10MHZ bandwidth. The amplifier bandwidth is around 70.MHZ. A variation of the μα733 design is the Signetics SE592. The basic difference is in the use of a pair of current sources instead of a single source supplying the first stage. The basic connection is a common emitter differential pair driving a common emitter second stage with shunt feedback. The output stage is common collector. These two commercial devices will suffice as long as the head signal is several mV minimum, and the head impedance is low. When lower level head signals are involved, then a better amplifier is needed. There is another connection that might be better and that is the cascode stage. Here the input impedance is about the same, but Miller feedback is considerably reduced. The shunt feedback connection does reduce the Miller feedback from that of a straight gain stage using a common emitter circuit. Compare these circuits in Figure 6.5 thru 6.7.

The low noise is achieved by the use of transistors that have very low base resistance, r_{ib}. A selection can be made based on r_{ib}, breakdown voltage and F_t. If desired, the amplifier could be designed and integrated as

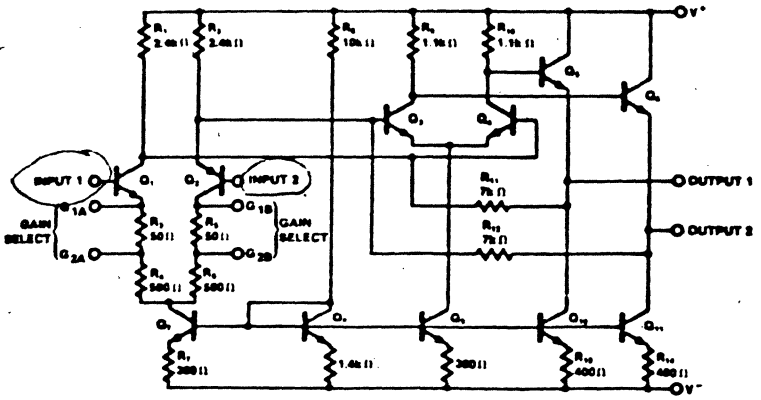
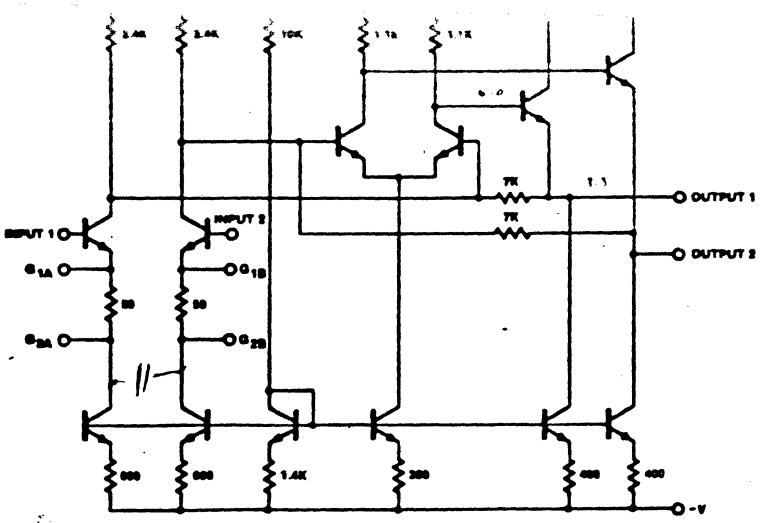
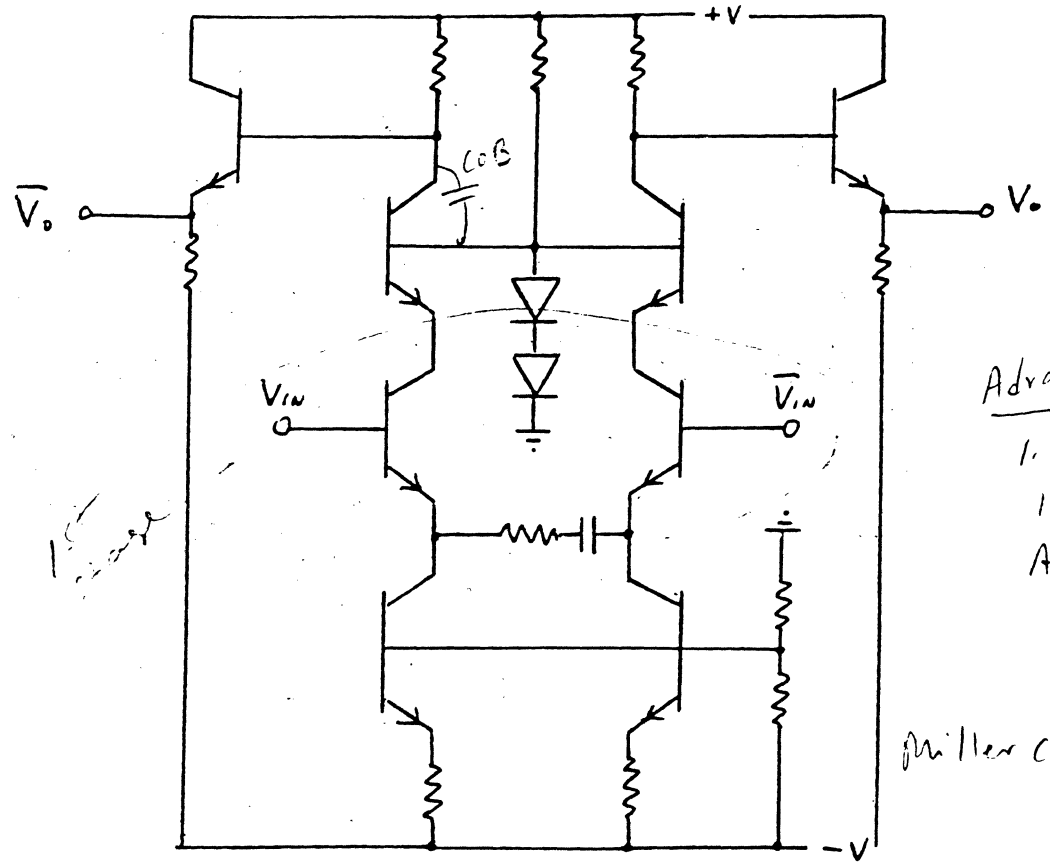


FIG 6.5 μA 733
SCHEMATIC (FAIRCHILD)



Two current source

FIG 6.6 NE 592
SCHEMATIC (SIGNETICS)



Advantages:

1. gain of 1st part is less than one

$$A = \frac{R_L}{R_E} = \frac{V_e}{V_e - R_E \frac{V_e}{R_E}} = \frac{V_e}{V_e} \approx 1$$

Miller cap = $(1+A)\,C_{OB}$
$= 2\,C_{OB}$

FIG 6.7 CASCODE CONNECTION

an IC using the design rules for the pertinent parameters.

We will design several Pre Amplifiers here in order to show the method, considerations and procedures.

The basic amplifier will be done first; see Figure 6.8. Simply, the input impedance differentially is equal to $2(r_e + Rm)\,\beta_{CE}$ (EQ 6.9). There are other considerations involving the collector, but we will ignore those. The output impedance differently is $2(\frac{R_L}{\beta_{EF_1}} + r_{e_2} + Rm_2)$ (EQ 6.10)

The gain differentially is $\frac{2\,R_L}{2(re_1 + Rm_1)}$ or $\frac{R_L}{re_1 + Rm_1}$ (EQ 6.11)

where r_e is the emitter resistance, Rm is the bonding resistance internal to the transistor.

These simple equations suffice as they will give us the true value within a few percent. If we have chosen a transistor with sufficient Ft, then the bandwidth will be determined by the Miller effect and any stray capacitance.

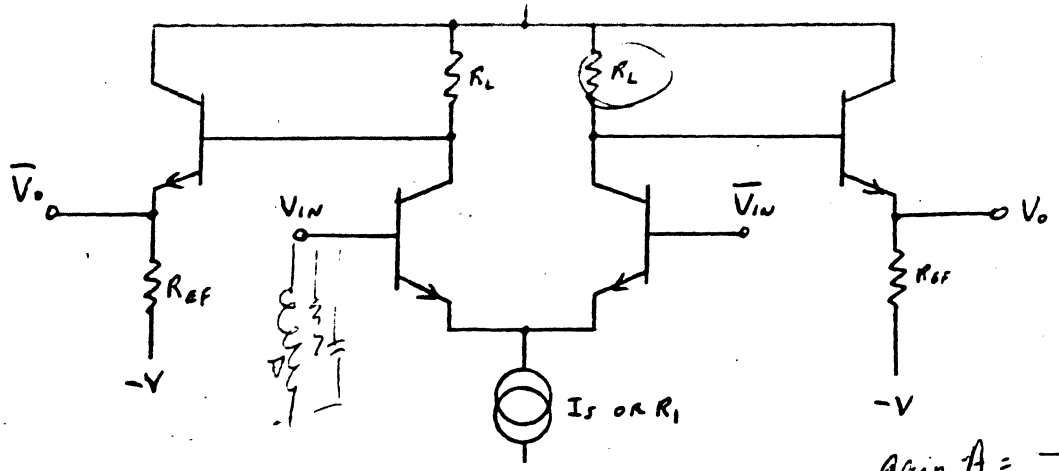
The Miller effect is worse if the input source resistance is large and less if it is low.

FROM FIG 6.9

$$V_1 = V_{in} + \frac{V_o R_s}{Rs + \frac{1}{C\,S}} \qquad \text{(EQ 6.12)}$$

$$V_o = V_1\left(\frac{-R_L}{r_e + Rm}\right)\left(\frac{R_s + \frac{1}{CS}}{R_L + Rs + \frac{1}{CS}}\right) \qquad \text{(EQ 6.13)}$$

needs correction



half circuit

gain $A = \dfrac{R_L}{r_e + R_L}$

FIG 6·8 SIMPLE DIFFERENTIAL PRE AMPLIFIER

$Z_{in} = r_e \cdot \beta$

$Z_o = \dfrac{R_L}{\beta}$

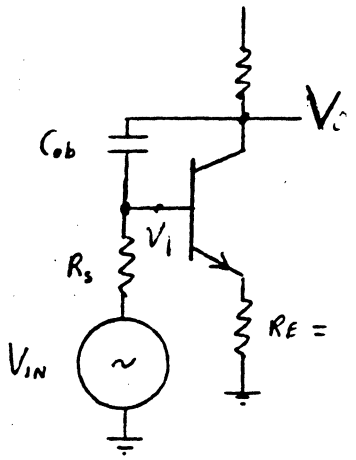


FIG 6·9
MILLER FEEDBACK CONSIDERATIONS

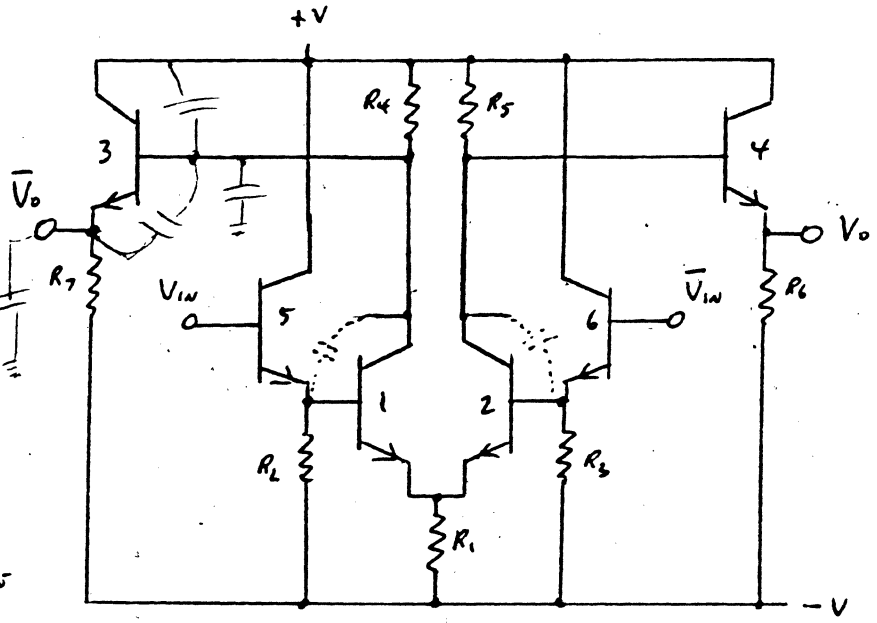


FIG 6·10
MILLER FEEDBACK REDUCTION

Choose Low Cob device

$I_R = V_{in\,peak} \times C_o \times W$

all wrong

substituting and rearranging we get:

$$A = \frac{V_o}{V_{in}} = \frac{R_L\left(R_S + \frac{1}{CS}\right)}{(re + Rm)\left(R_L + Rs + \frac{1}{CS}\right) + R_L} \qquad \text{(EQ 6.14)}$$

If we allow Rs → 0 then we have the case of zero input resistance which is close to the case of being driven by an emitter follower.

$$A_{R_S \to 0} = \frac{\frac{1}{R_L(CS)}}{(re + Rm)\left(R_L + \frac{1}{CS}\right) + R_L} \qquad \text{(EQ 6.15)}$$

If the frequency is raised so that $\left|\frac{1}{CS}\right| = R_L$ in magnitude, then the equation reduces to:

$$A = \frac{R_L^2}{(re + Rm)(2R_L) + R_L} = \frac{R_L}{2(re + Rm) + 1} \qquad \text{(EQ 6.16)}$$

which indicates that the true -3db point for the zero Rs case is slightly lower than where $|XC| = R_L$.

The whole object is to show that as long as we use the circuit of Figure 6.8, we will not get good bandwidth even if we drive the inputs with emitter followers in order to reduce Rs (Figure 6.10).

Notice also that the bandwidth reduces quickly if R_L is large. This may be acceptable, though, so we will finish the design. The current source and the dynamic range needs to be considered next. The power supply +V can be

determined from the current source

$$V_{o_{dc}} = +V - \left(\frac{I_{source}}{2}\right)\left(R_L\right)\left(\frac{\beta}{1+\beta}\right) \qquad \text{(EQ 6.17)}$$

In order to get sufficient reverse bias on the collector junction, we can refer to the transistor plots of constant bandwidth as a function of V_{CE} and I_C. Choosing the V_{CE} for the best bandwidth, we only need to assure ourselves that the negative output signal swing which is the input signal times the gain cannot saturate the collector junction.

$$V_{in\,\$i\,max} + V_{in\ max\ PP\ SE} << (V_{cc} -)(I_{source})\ \frac{\beta}{1+\beta}\ RL \qquad \text{(EQ 6.18)}$$

$$\text{and} \quad V_{in} - V_{be} << (I_{source})\left(\frac{\beta}{1+\beta}\right) RL \qquad \text{(EQ 6.19)}$$

A'

If these three equations are satisfied in the worse case, we have established the +V level. For example, using the parameters below determine the values required using the circuit of Figure 6.10.

β	= 70 Min
F_t @ 2.ma	= 400 MHZ
C_{ob}	= 5.PF
Vin max	= 10.$MV_{pp\ diff}$
F sig max	= 5.MHZ

READ CIRCUITS

@ 10 Times signal freq

First we will design for a bandwidth of at least 50 MHZ so that we have control of the phase over a manufacturing run.

With an emitter follower input we can assure that Miller effect is small therefore the roll off is approximately when $|X_C| = R$. Notice that we have several capacitors in parallel, C_{ob} of the amplifier, C_{ob} of the emmiter follower and some C_{be} of the emitter follower plus stray capacitance.

$$C_t = 2\,C_{ob} + C_{be} + C_{st} = 10 + 3 + 5 = 18._{pf} \qquad \text{(EQ 6.20)}$$

assume 20_{pf}

$$X_C = \frac{1}{2\pi FC} = \frac{1}{(2\pi)(5\times10^{7})(2\times10^{-11})} = 1.59\times10^{2}\ \Omega \qquad \text{(EQ 6.21)}$$

Therefore R_L cannot be greater than 150 Ω

At a current of 2.0 ma per transistor we need a current source of 4.0 ma.

The gain of the 2nd stage is approximately

$$A_2 = \frac{R_L}{re + Rm} = \frac{150}{\frac{26}{2ma} + 5} = 8.333 \qquad \text{Eq 6.22}$$

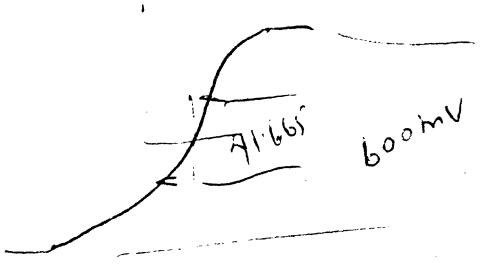
The V swing across the R_L is

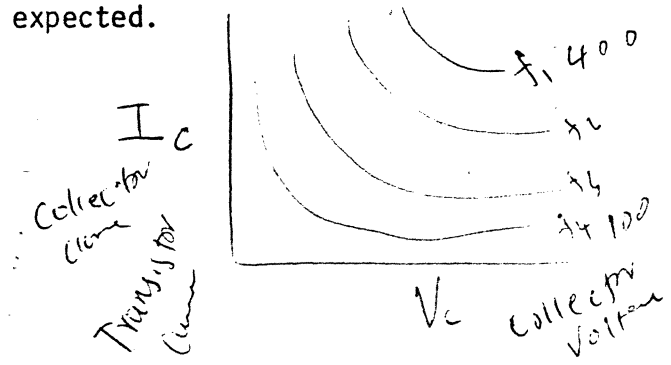
$$V_{RL_{pp}} = \frac{Vin_{pp}}{2} A_2 = \frac{10mv}{2}\ 8.333 = 41.665 \text{ mv pp se} \qquad \text{(EQ 6.23)}$$

The max DC capability of the output V swing is

$$(I_S)(R_L) = (4.ma)(150\ ) = 600.mv \text{ pp se} \qquad \text{(EQ 6.24)}$$

which is well above the 41.665mV expected.





READ CIRCUITS

We need about 5 volts reverse bias on the collector junction as the Vcc needs to be greater than

$$V_{cc} \quad -V_{b_1} - V_{RL} \gtrsim 5V \qquad \text{(EQ 6.25)}$$

$$V_{cc} \quad -(-0.75V) - (2.0ma)(150\Omega) \gtrsim 5.V$$

$$\therefore \; Vcc \gtrsim +4.55V$$

to allow for worse case conditons let us choose 6.0V for Vcc.

The output quiescent voltage is then (nominal)

$$V_{c_2} - V_{bc_3} = 6.0_V - (2.0ma)(150\Omega) - 0.75v = 4.95v \qquad \text{(EQ 6.26)}$$

If we choose the negative supply as -6.0v then the current source, if a resistor, should be (nominal)

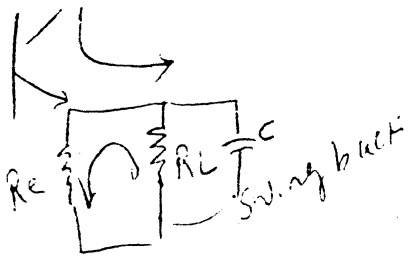
$$R_1 = \frac{|2V_{be} - 6.0V|}{2I_C} = \frac{|1.5v - 6.0v|}{4.0\ ma} = 1.125K\Omega \qquad \text{(EQ 6.27)}$$

Similarly we can calculate the input emitter follower resistor for a 2.0ma current as (nominal)

$$R_2 = \frac{|Vbe - 6.0V|}{2.0ma} = \frac{|0.75 - 6.0|}{2.0ma} = 2.625K\Omega \qquad \text{(EQ 6.28)}$$

The output emitter follower can only be calculated if we know the impedance we will be driving. Let us assume we will drive a 300Ω load. Our output swing is 41.66 mV_{pp}. This requires that we be able to pull down the emitter voltage such that it can follow.

$$\frac{V}{T} = \frac{I}{C}$$



READ CIRCUITS

$$\frac{V_o\ pp}{R_o} = \frac{41.66\ mvpp}{300\ \Omega} = 0.1388\ ma\ required. \qquad (EQ\ 6.29)$$

Output capacitance will increase this value.

We can provide this current easily with our 2.0 ma sources

$$R_6 = \frac{V_o - (-6V)}{2.0\ ma} = \frac{4.9v + 6.0v}{2.0\ ma} = 5.45K,\ nominal \qquad (EQ\ 6.30)$$

The true gain is not the 8.33 of EQ 6.22, but is modified by the two emitter followers.

The gain is approximately

$$\frac{R_2}{r_{c_1} + R_2}\ (8.333)\ \frac{\frac{R_0\ R_6}{R_0 + R_6}}{r_{e_3} + \frac{R_0 R_6}{R_0 + R_6}} = \qquad (EQ\ 6.31)$$

$$\frac{2.625K}{\frac{26}{2} + 2.625}\ (8.333)\ \frac{\frac{(5.45K)\ (0.5K)}{5.45K + 0.5K}}{\frac{26}{2} + \frac{(5.45K)(0.5K)}{5.45K + 0.5K}} =$$

$$(.995)(8.333)(.972) = 8.059 \qquad (EQ\ 6.32)$$

The above was done to show the attenuation of the other stages and in practice you will measure very close to this.

FIG 6.11

ATTENUATION THRU EMITTER FOLLOWERS

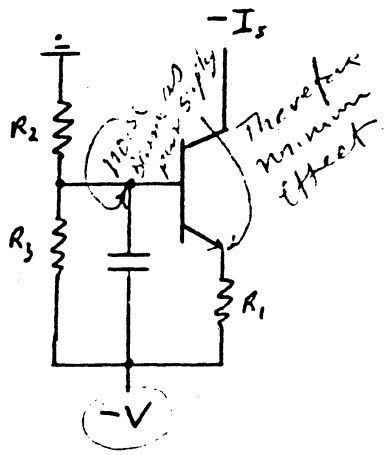


FIG 6.12 A

CURRENT SOURCE FROM DESCRETE COMPONENTS

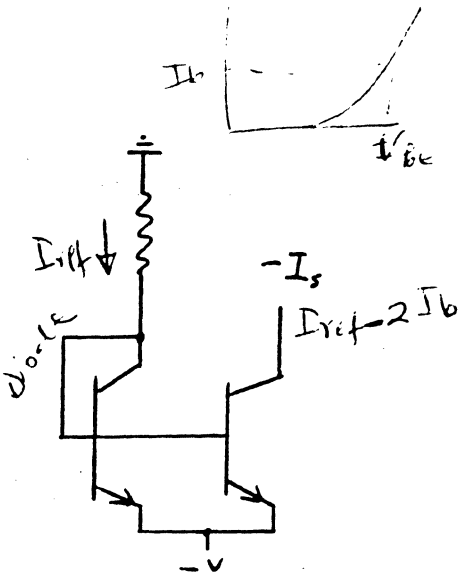


FIG 6.12 B

CURRENT SOURCE MATCHED V_{be} TRANSISTOR PAIR OR INTERGRATED

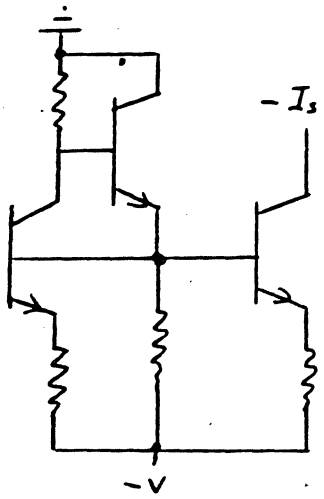


FIG 6.12 C

STABALIZED WILSON CURRENT SOURCE INTERGRATED VERSION

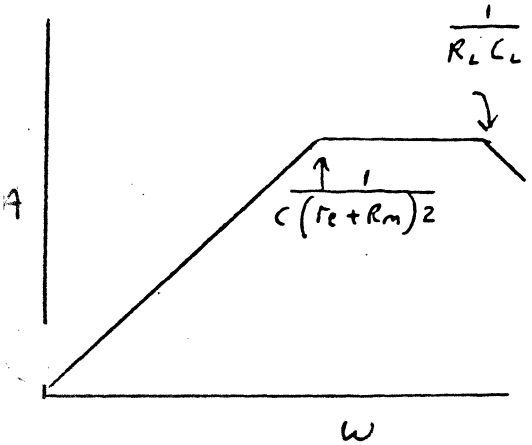


FIG 6.14

BODE PLOT OF AMPLIFIER OF FIG 6.13

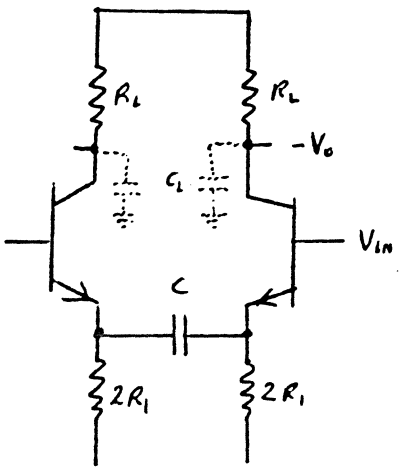
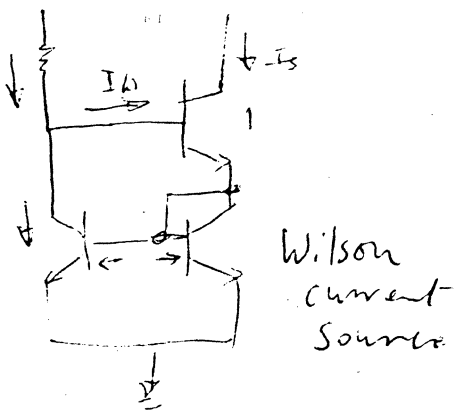


FIG 6.13



DESCRETE OR INTERGRATED AMPLIFIER WITH FORCED CURRENT - V_{be} BALANCE

Let us look at the effect of the current source resistor. If there is a 1.0 Volt$_{pp}$ noise signal on the input, then we would expect the output quiescent voltage to vary.

It would be

$$\Delta V_o = R_L \Delta I = \frac{1.0V}{1.125K\Omega} (150\Omega) = 0.1333 \text{ volts} \qquad \text{(EQ 6.33)}$$

Depending on the balance of the circuit we would expect some of this change to appear in the output as a differential signal. Assume the balance was 2% off then the output would contain 2% (0.1333v) or 2.666 mV noise.

If our minimum input signal were 1.0 mV then the output would be (8.059)(1.0 mv) = 8.059 mv with a 2.666 mv noise for a $^{S}/_{N}$ of $\frac{8.059}{2.666}$ = 3.02:1 (EQ 6.34) which is disastrous.

Well, what can be done? There are several. One is to provide the best balance in both transistor parameters and resistor values, and second to make R_1 a current source. Now the current will not vary with noise and the current balance is optimized. A current source is shown in Figure 6.12.

The current source is calculated as follows:

$$I_S = \frac{\frac{|V| R_3}{R_2 + R_3} - V_{be}}{R_1} \qquad \text{(EQ 6.35)}$$

Notice the placement of the capacitor C. This is positioned in order to reduce noise across R_1 which determines the actual current. It's value is high enough so that the lowest frequency component of noise is sufficiently attenuated.

The second current source type is shown in Figure 6.12B. It is basically a Wilson source.

Now we have got a very good idea of what the nominal case should be. We have no idea of what will happen worse case. Let us pursue this as it is a very important consideration. Worse case is always figured to use the various parameters in the direction that emphasizes the calculation in the direction desired.

The minimum value of stage current (non current source version) is obtained by modifying EQ 6.27. (use 5% values)

$$I_{min} = \frac{|-2\ V_{be}\ max - V_-\ min|}{R_1\ max} = \frac{|-(2)(0.80v) - (-5.7v)|}{1181\ \Omega} \qquad \text{(EQ 6.36)}$$

$$= \underline{3.471\ ma} \quad \text{instead of our desired 4.ma}$$

This includes the temperature effects on V_{be}.

Similarly, we can calculate the maximum current

$$I_{max} = \frac{|-2\ V_{be}\ min - V_-\ max|}{R_1\ min} = \frac{|-(2)(0.7v) - (-6.34v)|}{1068\ \Omega}$$

$$= \underline{4.588\ ma} \qquad \text{(EQ 6.37)}$$

The output voltage variations are complex. We will take the straight forward case first.

$$V_{o\,min} = V_{cc\,min} - \left[\left(\frac{I_{s\,max}}{2}\right)\left(\frac{\beta_{2\,max}}{1+\beta_{2\,max}}\right) - \frac{I_{3\,max}}{1+\beta_{3\,min}}\right] R_{L\,max} - V_{be_3\,max}$$

$$= 5.70v - \left[\left(\frac{4.588\ ma}{2}\right)\left(\frac{300}{1+300}\right) - \frac{I_{s\,max}}{1+300}\right] 157\Omega - 0.8V \qquad \text{(EQ 6.38)}$$

Notice that I_3 max really depends on V_0 min, therefore, the current is not the true maximum at all but less. We can best calculate a usable value by assuming a straight 2 ma for I_3 and ignoring the fact that it is worse than worse case, but this is acceptable.

$$\therefore\ V_{0\,min} = 5.70V - (2.286\ ma - .006\ ma)157\Omega - 0.8V \qquad \text{(EQ 6.39)}$$

$$= 4.460\ V$$

Similarly we can obtain $V_{O\,max}$

$$V_{o\,max} = V_{cc\,max} - \left[\left(\frac{I_{s\ min}}{2}\right)\left(\frac{\beta_{2\ min}}{1+\beta_{2\,min}}\right) - \frac{I_{3\,min}}{1+\beta_{3\,max}}\right] R_{L\ min} - V_{be_3\ min}$$

$$= 6.30v - \left[\left(\frac{3.471\ ma}{2}\right)\left(\frac{70}{1+70}\right) - \frac{2.0\ ma}{1+70}\right] 143\,\Omega - 0.70v$$

$$= 6.30v - (1.711\ ma - .028)143 - 0.70v$$

$$= 5.36v \qquad \text{(EQ 6.40)}$$

There is a 0.9v difference between the two worse cases meaning that in manufacturing we will see this spread.

In actuality it is worse than this because we cannot buy discreet transistors with Vbe's so well matched. Let us look at the Ic current again.

For this we need to refer to the Vbe vs. Ic curves as well as the spread between devices. This spread can be as great as a tenth of a volt at these currents. The unbalance then becomes,

$$\Delta I \doteq \frac{0.1v}{\frac{2(26)}{\frac{Ima}{2}}} = 4.41 \text{ ma} \qquad \text{(EQ 6.41)}$$

This means that one transistor is drawing almost all the current and the second is nearly cut off - drawing only

$$4.588 - 4.41 = 0.178 \text{ ma}$$

The amplifier is useless to us if built out of discreet transistors. Now do you see the advantage of doing a worse case analysis. We can modify the circuit to force current balance by employing two current sources of half the value and adding a capacitor of a suitable value between the emitter as shown in Figure 6.13. Now balance is restored, but at the cost of a zero and Pole in the gain equation (6.42) for low frequencies.

$$A = \frac{R_L}{r_e + Rm + \frac{1}{cs}}$$

$$= \frac{R_L \; CS}{C(r_e + Rm)\, S + 1} \qquad \text{(EQ 6.42)}$$

The gain curves are shown in Figure 6.14.

Usually for a low voltage low current stage we do not need to worry about the power dissipation of the transistors, but we will calculate those values anyway. This combination cannot occur but it will assure us that we are safe.

$$P_{w\,max} = (I_{c\,max})\,V_{ce\,max}$$

$$= \left(\frac{4.588\text{ ma}}{2}\right)\left(\frac{\beta_2\text{ max}}{1+\beta_2\text{ max}}\right)\left[V_{cc\,max}^{-2}V_{be_{max}} - \left(\left(\frac{4.588\text{ ma}}{2}\right)\left(\frac{\beta_2\text{ max}}{1+\beta_{2_{max}}}\right) - \frac{I_3{}^1}{1+\beta_{3\,max}}\right)R_{L_{min}}\right]$$

$$= \left(\frac{4.588_{ma}}{2}\right)\left(\frac{300}{301}\right)\left[6.3v + 1.6v - \left(\left(\frac{4.588_{ma}}{2}\right)\left(\frac{300}{301}\right) - \frac{2ma}{301}\right)\ 143\right]$$

$$= (2.286 \times 10^{-3})(6.3 + 1.6 - 3.26 \times 10^{-1}) = 17.314\text{ mW} \qquad \text{(EQ 6.44)}$$

For a transistor that has a derating factor of 1.7 mw/^{0}C this amounts to a

$$\frac{17.314\text{ mw}}{1.7\text{ mw}/^0\text{C}} = 10.18^0\text{ C rise} \qquad \text{(EQ 6.45)}$$

The two worst resistors are R_1 and R_6. These are respectively

$$(I_{1\,max})(V_{R\,max}) = \frac{(|-2\,V_{be\,min} - V_{\_\,max}|)^2}{R_{1\,min}} = 22.48\text{ mw} \qquad \text{(EQ 6.46)}$$

and

$$(I_{6\,max})(V_{R\,max}) = \frac{(V_{0\,max} + V_{e\,max})^2}{R_{6\,min}} =$$

$$\frac{(5.36V + 6.3V)^2}{5.177K} = 26.26\text{ mw} \qquad \text{(EQ 6.47)}$$

READ CIRCUITS

This completes the design except for the noise. This circuit is quite noisy for several reasons. First the effective noise voltage source resistance r_{ib} is twice due to the emitter follower input r_{ib} used to increase the bandwidth and the regular gain transistor input source resistance r_{ib}. Second, the gain is only 8.03 which is not enough to ensure adequate Signal to Noise ratio into the following stages. Third, the common mode power supply rejection and common mode input signal rejection is very poor. All this adds up to a poor choice. Some degree of immunity can be achieved by using transistors that have lower Cob and using current sources and an emitter capacitor. These 3 changes improve the design and permit higher gain. The capacitor could be eliminated if the circuit were integrated where Vbe matching is typically better than 5.mv.

A much better circuit is the cascode amplifier. We will discuss this next. It is easily integrated.

Transistors 5 and 6 are the current scources. The current is fixed by R_2 and R_3 with R_1. Transistors 3 and 4 is the first stage. Its emitter feedback is thru C and its load is r_e of transistors 1 and 2. Then transistors 1 and 2 provide the gain where their load is R_4 and R_5. The output stage is transistor7and 8.

The advantage of this circuit is that the gain of the first stage is one, therefore, Miller capacitance is only 2(Cob). This devise can be made large in order to ensure r_{ib} is small therefore low noise. Bandwidth is determined by R_4 and Cob of transistors 1 or 2 and these can be made small in order to reduce Cob.

Let us proceed as we did before and start with the value of R_4 and R_5 . From 2N918 transistor data,

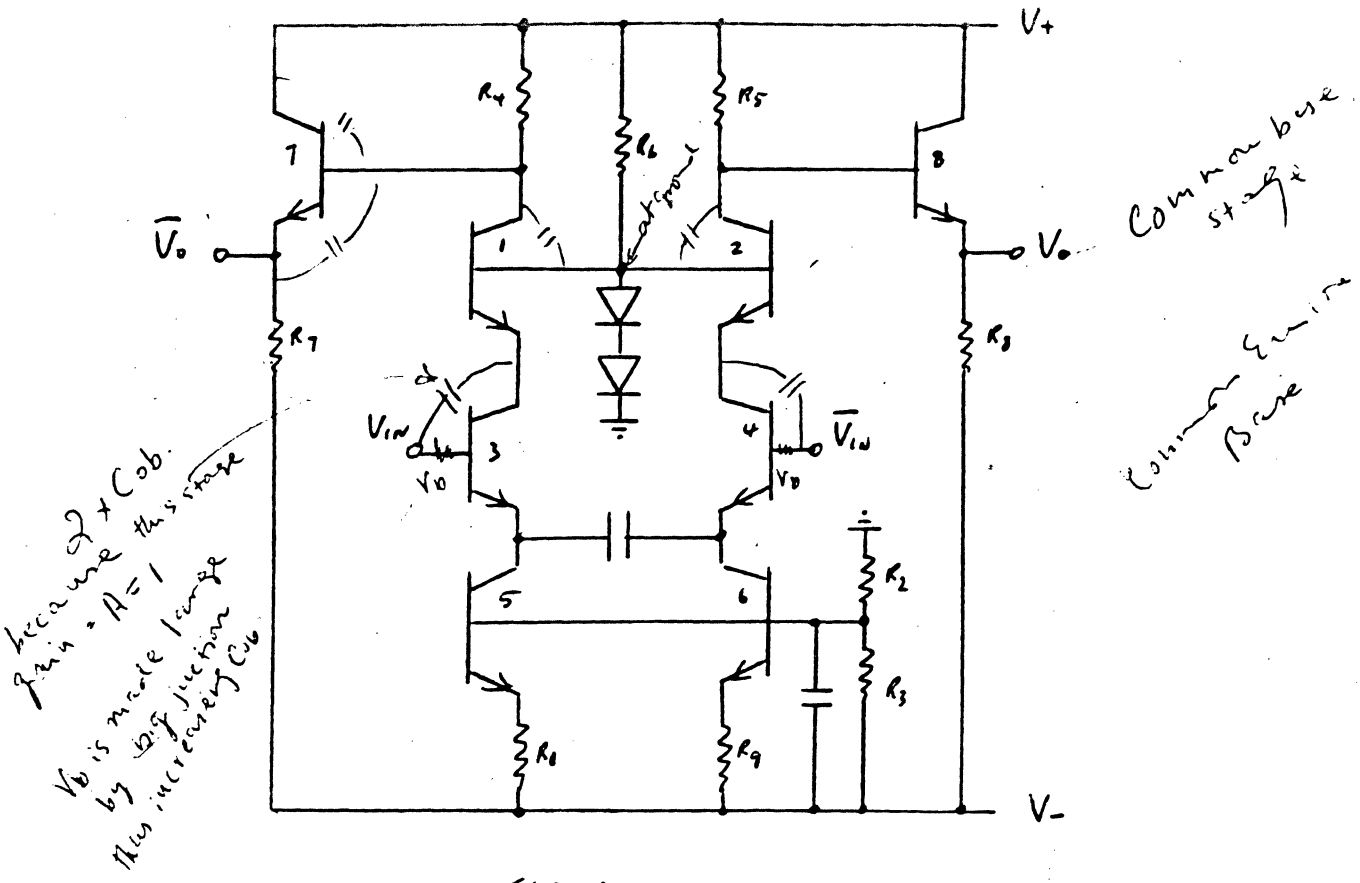


FIG 6.15

CASCODE PRE AMPLIFIER
INTERGRATED OR DESCRETE

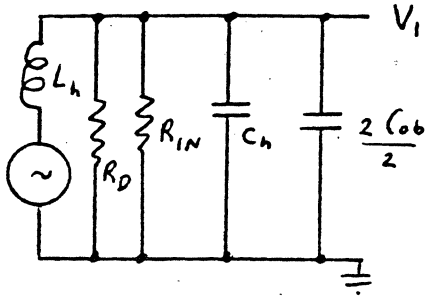


FIG 6.16

INPUT ATTENUATION DUE TO
HEAD, AMPLIFIER IMPEDANCES

$$T_{1,2}\ Cob = 1.9_{pf} \quad @\ 5.v$$

$$F_t = 1.2\ GHZ$$

$$X_c = \frac{1}{2\pi\ F_{bw}(Cob_1 + Cob_7 + C_5)} \qquad \text{(EQ 6.48)}$$

$$X_c = \frac{1}{50.MHZ(2\pi)(1.9 + 1.9 + 3)_{pf}} = \frac{1}{(5x10^7)(2\pi)(6.8x10^{-12})} = 4.68x10^2\Omega$$

use 450 R_L nominal.

$$\text{Gain } A_1 = \frac{R_L}{r_{e_1} + R_{m_1}} = \frac{450}{\frac{26}{2ma} + 5\,\Omega} = 25 \text{ nominal} \qquad \text{(EQ 6.49)}$$

$$\text{Gain } A_3 = \frac{r_{e1} + R_{m1}}{r_{e2} + Rm_2} = \frac{\frac{26}{2} + 5}{\frac{26}{2} + 5} = 1.000 \qquad \text{(EQ 6.50)}$$

Therefore the total gain is (25)(1) for the same bandwidth.

Next we will calculate the current sources for 2.0 ma each using our ± 6V power supplies letting R_2 and R_3= 500Ω each.

$$R_1 = \frac{\frac{Vs\ R_3}{R_2 + R_3} - \frac{2(2ma)}{\beta + 1}\ \frac{R_2\ R_3}{R_2 + R_3} - V_{be}}{2.0\ ma} \qquad \text{(EQ 6.51)}$$

$$= \frac{\frac{(6.0)(500)}{1000} - \left(\frac{4x10^{-3}}{151}\right)\left(250\Omega\right) - 0.75V}{2x10^{-3}}$$

$$= \frac{3 - 6.622x10^{-3} - 0.75}{2x10^{-3}} = 1.12K\Omega$$

Notice the effect of the R_2 R_3 network as a result of base current; it reduces the effective base voltage. This is why we used 500Ω each. If we chose a value to save current then the loss could be substantial in the worse case analysis.

The output voltage, T_{X_7} becomes an interesting function of all the series bases and the output base.

$$V_0 = V_{CC} - R_4(I_{R_1})(1 - \frac{(3 - 1)}{\beta + 1}) - V_{be_7}$$

$$= 6V - 450(2.0ma)(1 - \frac{2}{151} - 0.75v = 4.362 \text{ Volts} \qquad \text{(EQ 6.52)}$$

The value of R_{7-8} for the same 2 ma of current simply is,

$$R_7 = \frac{V_0 + V_-}{2.ma} = \frac{4.36 + 6.0v}{2.0ma} = 5.18K\Omega \qquad \text{(EQ 6.53)}$$

And lastly, the value of R_6 should be such that the variations in base current of $T_{X\ 1}$ and $_2$ do not disturb the voltage.

Choose I_d of 6 ma then

$$R_6 = \frac{V_{CC} - 2\ V_d}{6.ma} = \frac{6.0v - 1.6v}{6.ma} = 733\Omega \qquad \text{(EQ 6.54)}$$

Now we could look at the bandwidth of the first stage collector. Since $A_3 = 1$ the C effective is =

$$(1 + A)Cob = (1 + 1)Cob = 2\ Cob = 2(5_{pf}) = 10.pf \qquad \text{(EQ 6.55)}$$

$$BW = \frac{1}{2\pi\ RC} = \frac{1}{2\pi\left(\frac{26}{2ma}\right)\left(10_{pf}\right)} = 1.22\ GH$$

or not worth bothering about except for the effect on the imput impedance Z_s. For a 5 MHZ signal into our head circuit, we get the following due to the differential connection.

If we choose a typical head with L_h = 10 μh, C_h = 10.pf and Cob = 5pf, we can calculate R.

$$\frac{1}{C_t R} = 2\zeta W_n \quad \therefore \quad R = \frac{1}{C_t 2\zeta W_n} \quad \text{and} \quad W_n = \frac{1}{\sqrt{L\ C_t}}$$

$$Wn = \frac{1}{\sqrt{(10^{-5}H)(1.5\times10^{-11}F)}} = 8.165 \times 10^7 \text{ rad} \qquad \text{(EQ 6.56)}$$

$$R = \frac{1}{(1.5\times10^{-11})(2)(.707)(8.165\times10^7)} = 577.\ \Omega \qquad \text{(EQ 6.57)}$$

At 5.0 MHZ this then becomes an attenuator α of the input circuit.

$$\alpha = \frac{\frac{(R)(-jXc)}{R - jXc}}{X_L + \frac{R(-jKc)}{R-jKc}} = \frac{\frac{(577)(-j2.122K)}{577 - j2.122K}}{314 + \frac{(577)(-j2.122K)}{577 - j\ 2.122K}} = 0.644\ \angle -5.48^0 \qquad \text{(EQ 6.58)}$$

↑ very large

where $X_C = \frac{1}{2\pi(5\text{ MHZ})(C_h + 2\ \frac{Cob}{2})} = 2.122K\Omega$

$$X_L = 2\pi\ (5\text{ MHZ})(10.\mu h) = 3.14\times10^2\Omega$$

This value of attenuation is better than the case where Miller effect is large which in turn both lowers the resistor value to keep ζ the same, but also increases the capacitance which worsens the attenuation.

We could complete the design by doing a worse case analysis for gain, V_0, Power dissipation, and dynamic range, but we have already done that. The use of the cascode stage only adds a slight complication yet permits a low noise design.

The amplifier noise contribution can be calculated from the following equation:

$$V_{n_A^2 \text{ diff}} = 2\left[(4kT\ B_W)(r_{ib} + \frac{1}{2\ gm}) + 4\ Z_S^2 q\ I_C B_W\ (\frac{1}{\beta_0} + \frac{1}{\beta^2(F)})\right] \quad \text{(EQ 6.59)}$$

where K is Boltzman's constant, T is the temperature in ^{O}Kelvin, Bw is the Bandwidth of interest, r_{ib} is the base resistance (base thermal noise), $gm = \frac{1}{r_e}$ is the collector transconductance ($\frac{1}{2gm}$ is the collector shot noise), Z_S is the Head impedance, q is the charge on the electron, I_C the collector current, $\frac{1}{\beta_0}$ the Base current shot noise, and $\frac{1}{\beta^2(F)}$ the collector current noise. The function of frequency is that obtained from the usual noise - frequency curves. If the amplifier bandwidth is much higher than the frequency of interest, we can use the value of β^2 unmodified. We will address this again.

Lastly, we should consider the two commercially available amplifiers. In using these amplifiers great care should be exercised in adhering to the specifications. For example, to rely on the typical specifications is to invite trouble during a manufacturing run. As is done in worse case analysis we use the parameter in the direction that accentuates the result. When using the μA733C, the gain at the 400 setting can be anywhere between 250 and 600. To calculate the worse case for a minimum input signal we would use the gain of 250 and when doing the maximum input case we use the maximum gain of 600. Now we know what our true output variations will be. These then should be

considered against S/N ratios for the low gain low input case and against the linearity specifications for the high gain high input case. Assume our minimum input signal is 0.5 mV_{pp} diff, and our maximum signal is 5.5 mV_{pp} diff. The S/N ratio is calculated from the input noise data. But the manual only gives a typical value. We could guess that this value might vary ± 6db and use that in our equations.

First we must convert the 0.5 mV_{pp} diff to RMS diff by dividing by $2\sqrt{2}$ to give 1.767×10^{-4}V RMS diff

peak to peak

$$\text{the S/N} = \frac{1.767 \times 10^{-4} V_{\text{rms diff}}}{(2)(1.2 \times 10^{-5} V_{\text{rms diff}})} = 7.365 \quad \text{(EQ 6.60)}$$

Always rms

$$\frac{0.5}{2\times(1.2\times10^{-5})\ (\text{rms})} = 20.83 \quad 2 = 3.06$$

in db it is 20 log 7.365 = 17.344 db (EQ 6.61)

This value is very low, therefore, another devise is indicated that has a lower noise or a narrower bandwidth of interest. Similarly for the linearity case.

$$(V_{sig\ max})(A_{max}) = (5.5\ mV)(600) = 3.3\ V_{pp}\ diff. \quad \text{(EQ 6.62)}$$

This value exceeds the minimum output voltage swing into a differential load of 2.KΩ by 0.3V. Again the devise is not suitable. Now these two examples were only given to emphasize the parameters of interest that we should concern ourselves with. As long as we use these parts within their specifications we are assured of good performance.

What is the value of input signal that guarantees 30db S/N at 10 MH Bandwidth?

$$Vin_{min_{pp}} = (2)(\text{antilog}\ \frac{30}{20})(1.2\times10^{-5}\ V)(2)(\sqrt{2}) = 2.146\ mV_{pp}\ diff \quad \text{(EQ 6.63)}$$

READ CIRCUITS

If we restricted our Bandwidth of interest to 5 MHZ then we could use signals the $\sqrt{2}$ lower or 1.517 mV_{pp}.

$$\sqrt{\frac{\text{Bw of interest}}{\text{Bw given}}} = \sqrt{\frac{5.\text{MHZ}}{10.\text{MHZ}}} = 0.707 = \frac{1}{\sqrt{2}} \qquad \text{(EQ 6.64)}$$

Another parameter that requires attention is the offset. At the 400 gain setting, the maximum output offset is 1.5 V which must be subtracted from the dynamic range as published as output voltage swing. Going back to our example, what is the maximum input signal that we can accept and still use the devise at this gain.

$$Vin_{pp}\ \text{max} = \frac{(Vo_{pp}\ \text{min} - V_{offset\ max})}{A\ \text{max}} = \frac{(3.0V_{pp} - 1.5V_{pp})}{600} = 2.50mV_{pp} \quad \text{(EQ 6.65)}$$

From what we have discussed then for a 30 db S/N we need 2.14 mVpp min to input, from the maximum input we are limited to only 2.50 mV or Too close for the restrictions we have placed on the circuit.

Lets look at this again for a gain of 100. Again the noise of EQ 6.63 holds. The Vin pp max needs to be calculated at the new gain using EQ 6.65.

$$\text{Vin pp max} = \frac{3.0\ V_{pp} - 1.5V_{pp}}{110} = \frac{\cancel{13.6}}{12.5}\ mV_{pp} \qquad \text{(EQ 6.66)}$$

which is much more sensible. Now we have at our disposal a dynamic range of 2.14 mV to $\frac{\cancel{13.6}}{12.5}$ mV that is guaranteed to meet our specifications.

When using the SE 592 there is some improvement in the specification for offset. This is due to the dual current sources used in the input stage. This can be taken advantage of to increase the input dynamic range from the last example to:

$$Vin_{pp}\ max = \left(\frac{3.0 - 0.75}{110}\right) = 20.45\ mV_{pp}diff \qquad \text{(EQ 6.67)}$$

We will need this feature in a later chapter when we are dealing with considerable offsets at the input.

Some improvement in the output swing at lower than 2K output loads can be achieved on both devices by providing more pulldown current at the output emitter followers. This must only be done within the limitations of the output emitter followers current handling capability, the alteration of the quiescent operating point due to the increased base current requirement and the power dissipation increase. This output pulldown current can be supplied either from a pair of resistors connected to the negative supply pin or from a pair of current sources.

These two devices then when used within their specifications can perform quite well as preamplifiers.

One added feature of the SE 592 is in its use of external feedback elements that can perform network filter functions.

A second requirement for the pre amplifier function is to interface with the following functions. If the pre amplifier, Read, and following amplifiers are very close then the emitter followers provided in all the examples given so far will suffice to isolate the collector load from any following capacitance which is its purpose.

In most cases however the pre amplifier is mounted close to the head and any head moving mechanism such as actuators, linear motors, etc. In these cases the output emitter follower is not adequate to drive any intervening cable.

For example, if we expect a 3.0 V less offset pp differential signal maximum, and we want to drive this into a 92Ω or 50Ω coaxial cable pair, each cable must carry a $1.5V_{pp}$ SE less offset/2 signal. At 92Ω Z_0 terminated at one end we need a current drive capability of over ± 8 ma. For best linearity an emitter follower should have over 10.ma current load. This can be provided by a second emitter follower capacitively coupled to the cable. A transistor should be chosen to handle the voltage, current and power dissipation. The function can also be provided by a common emitter amplifier with the collector loads equal to the cable impedance. Both circuits are shown in Figure 6.17A and B.

In Figure 6.17A we need to provide a 92Ω coaxial cable with a maximum of 1.5V pp SE signal. (Use 5% tolerances) (EQ 6.68)

$$R_{E\ max} = \frac{V_{\omega dc\ min} - V_{be\ max} + |V-|min}{\frac{1.5\ V_{pp\ SE\ max}}{Z_0\ min}} = \frac{2.0V - 0.80V + 6.7V}{\frac{1.5\ V}{87.4\Omega}} = 402\,\Omega\ max$$

tc allow some margin for linearity we need about 10% less or about 360.Ω max. The value of C needs to be large enough to handle the lowest frequency F_L of interest without attenuation

$$C_{min} = \frac{1}{(2\pi)(F_L)(Z_0)} \qquad \text{(EQ 6.69)}$$

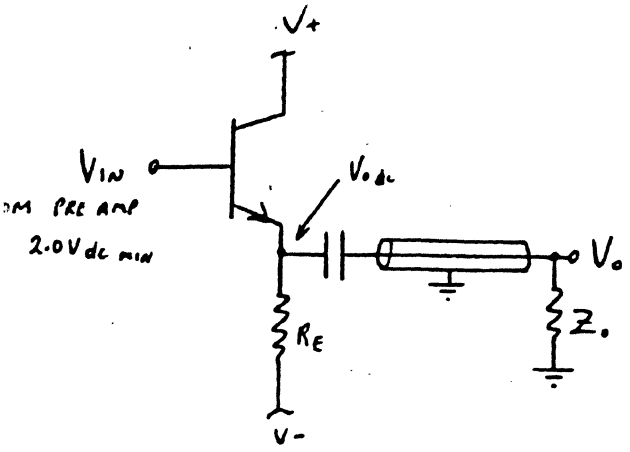


FIG 6.17 A

EMITTE FOLLOWER CABLE DRIVER (HALF SHOWN)

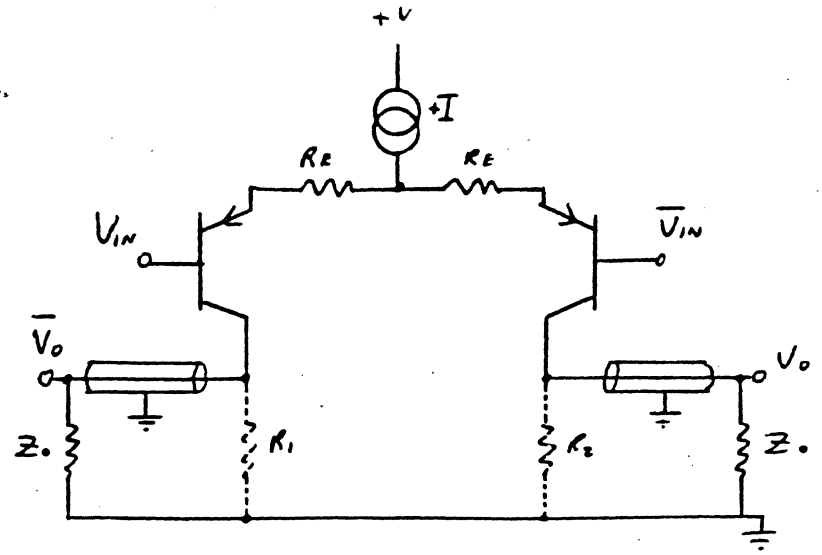


FIG 6.17 B

COMMON EMITTER CABLE DRIVER

R_1 AND R_L MAY BE ADDED IF TERMINATION IS REQUIRED ON BOTH ENDS OF CABLE

IF USED GAIN MUST USE $Z_0/2$ INSTEAD OF Z_0

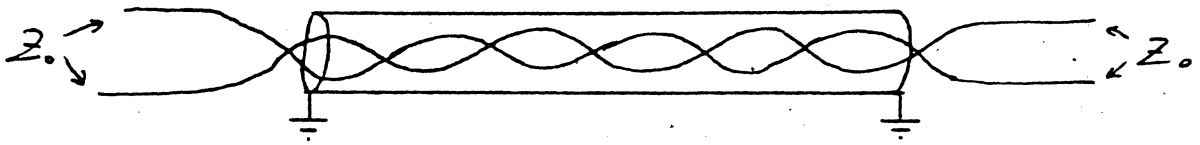


FIG 6.18 A

TWINAXIAL CABLE

FIG 6.18 B

CABLE IMPEDANCE Z_0 AS SEEN BY DRIVING CIRCUIT FROM EITHER END

READ CIRCUITS

The maximum power dissipation in the transistor is

$$P_{w\ max} = \left[\left(1 + \frac{1}{\beta_{max}}\right)V_{be_{max}} + V_{+\ max} - Vin\ min\right] (I_E)\left(\frac{\beta\ max}{1 + \beta max}\right) \quad \text{(EQ 6.70)}$$

$$\text{and } I_E = \frac{Vin\ min - V_{be}\ max + |V_-|max}{R_{E\ min}} \quad \text{(EQ 6.71)}$$

substituting we get

$$P_{w\ max} = \left[\left(1 + \frac{1}{300}\right) 0.80V + 6.3V - 2.0V\right]\left[\frac{2.0V - 0.8V + 6.3V}{324\Omega}\right]\left(\frac{300}{1 + 300}\right)$$

$$= 117.7\ mW \quad \text{(EQ 6.72)}$$

with a TO18 can transistor with 1.7 mW/ °C thermal transconductance we would get a $\frac{117.7\ mv}{1.7\ mw/\ ^oC} = 69.2^o$ C rise at the junction. (EQ 6.73)

The second circuit, Figure 6.17B is designed as follows:

The single current source required needs to supply, (~~this includes 10% margin for linearity~~).

$$I_{s\ min} = 1.1\left(\frac{V_{in\ pp\ max}}{Z_{o}\ \Omega min}\right) = \left(\frac{1.5V}{87.4\Omega}\right) 1.1 = 18.8\ ma\ min \quad \text{(EQ 6.74)}$$

If this current source were a resistor to + 6v and Vin were +3.0 V_{max} then that resistor would be

$$R_{5max} = \frac{V+\ min - Vin\ max + V_{be}\ max}{18.8\ ma} = \frac{5.7v - 3.0v + 0.8v}{18.8\ ma}$$

$$= 186.1\ \Omega \quad \text{(EQ 6.75)}$$

The true resistor will be 5% less or 176.8 Ω

We had best use a current source and we would get better results for CMR performance.

We next need to calculate the resistor R_E. If we design for a gain for the stage of 1_{NOM} then

$$R_m + r_e + R_E = Z_0 \qquad R_{E\ nom} = 92\Omega - \frac{26}{20.ma} - 5\Omega = 85.7\Omega \qquad \text{(EQ 6.76)}$$

Notice that we could ignore r_e + Rm as they are small compared to R_E. If we did we would only be off a few percent.

The same stage could be designed with two current sources of half value with a single resistor of 2 R_E between them and still get the same DC and AC results. But if the current were supplied by resistors, then the gain equation is modified and the CMRR would be considerably degraded.

We next need to verify that the transistors will not be saturated. If we had a 3.0V min input DC and a 1.5V max AC $_{pp}$ signal $_{SE}$ then the base will be

$$V_{DC_{min}} - \frac{V_{AC\ SE\ max}}{2} = 3.0V - \frac{1.5V}{2} = 2.25V\ min \qquad \text{(EQ 6.77)}$$

The collector swing is the maximum current times the Z_0 max or

$$(20.0\ ma)(96.6\Omega) = 1.93V \qquad \text{(EQ 6.78)}$$

this leaves us 2.25V - 1.93V = 0.32V of margin worse case.

If we were to use the μa 733 or SE 592 the Vin_{DC} would range from 2.4 to 3.4 volts so some restrictions would need to be placed on the maximum output V swing to avoid collector saturation worse case.

If we used 50Ω coax cable all the currents would need to be increased accordingly. ALSO IF WE INTEND TERMINATING THE CABLE AT BOTH ENDS THE CURRENTS AND GAIN RESISTORS WILL NEED CORRECTION. (SEE FIG 6.17B R1 ADDED)

A better cable is a twinaxial cable. It consists of a shielded twisted pair with good control of Z_0. The impedance is listed as ohms differential. For our 92Ω coax case, they could be replaced with a 184Ω twinax cable with all the equations for current ect remaining the same as 184Ω differential equals 92Ω single ended. Normal twisted pair is around 125Ω requiring increased driving currents for the same signal swing. The Z_0 of the calculations is ½ the Z_0 of twinax cable. (FIG 6.18A-B)

We will leave this exercise up to the student to worse case the design. The main benefit of twinaxial cable is its inherent balance. This is required for phase balance as well as amplitude balance which maintains the Common Mode Rejection of the system while reducing noise pick up.

ADDENDUM FOR CHAPTER 6

A separate series of preamplifiers are used in the tape drive industry. None are presently used in the disc drive industry, although some thought has been given to their use. The preamplifiers considered are the common base type. This type can be made true differential by driving the centertap with a single current source or by capacitive coupling and a pair of current sources. The interposition of the diode matrix forces this type of coupling due to the large offset voltages.

Figure 6.18 A thru C show variations of the same basic type. The gain of these stages is not much different from the gain equations previously given. We will develop this equation.

XR_D is the total number of series diodes that would be used if a matrix were required. It is for this reason that this type of Preamplifier is not used in the disc drive industry. Also the noise contribution of each diode should be taken into account. The gain is a function of the head impedance. Whereas the attenuation of the head circuit was previously considered, it now shows up in the total gain of the amplifier.

Deriving the gain equation as EQ 6.79, we can see the total effect. We will not include a damping resistor as we do not need it because the amplifier load is high ($2r_e$). From Figure 6.19 we get:

$$i_{IN(s)} = \frac{\frac{\frac{1}{CS} R_T}{\frac{1}{CS} + R_T} V_{sig(s)}}{R_T \left(LS + \frac{\frac{1}{CS} R_T}{\frac{1}{CS} + R_T} \right)} \qquad \text{where } R_T = 2XR_D + 2r_e \qquad \text{(EQ 6.79)}$$

$$i_{IN(s)} = \frac{V_{sig(s)}}{R_T LCS^2 + LS + R_T} = \frac{V_{sig(s)}}{R_T LC \left(S^2 + \frac{S}{R_T C} + \frac{1}{LC} \right)} \qquad \text{(EQ 6.80)}$$

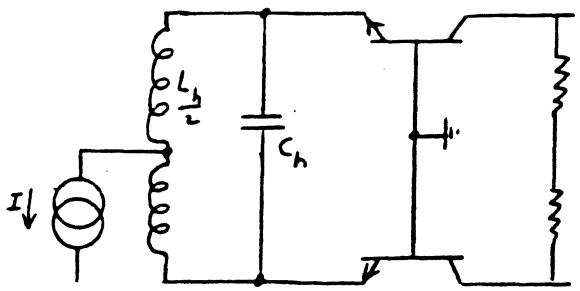


FIG 6.19A
SINGLE HEAD COMMON BASE AMPLIFIER

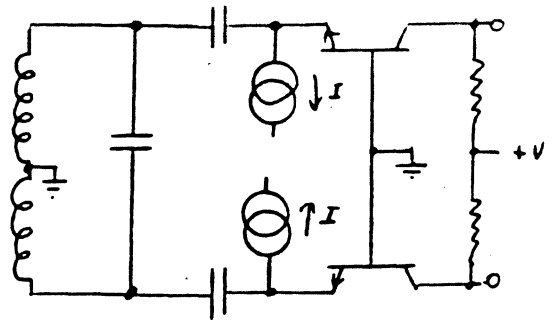


FIG 6.19B
CAPACITAVLY COUPLED HEAD TO A COMMON BASE AMPLIFIER

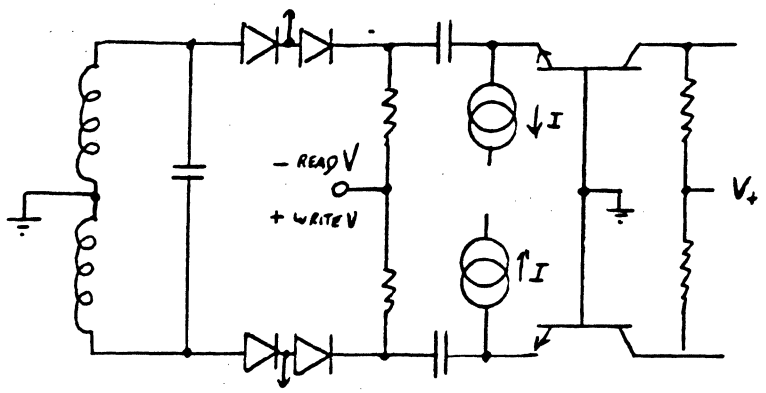


FIG 6.19C
MATRIX CONNECTED COMMON BASE AMPLIFIER WITH CAPACITIVE COUPLING TO ELIMINATE DIODE OFFSETS.

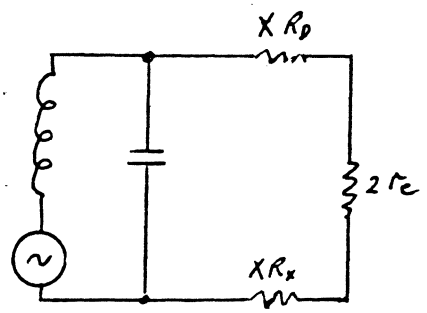


FIG 6.20A
EQUIVILENT CIRCUIT

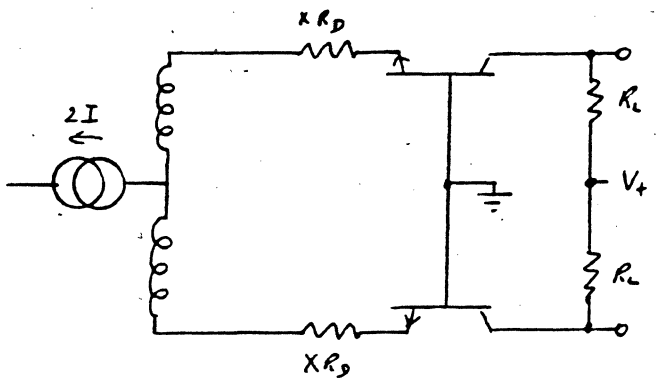


FIG 6.20B
GAIN STAGE

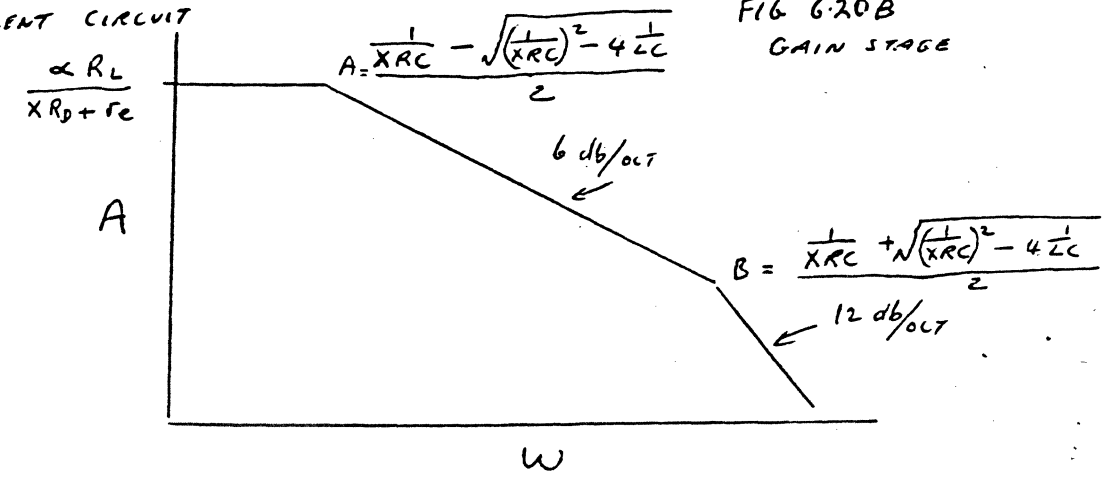


FIG 6.21

ADDENDUM FOR CHAPTER 6

which becomes in the standard form:

$$i_{in(s)} \qquad \frac{V_{sig(s)}\ \omega_n^2}{R_T\left(s^2 + 2\zeta\omega_n s + \omega_n^2\right)} \tag{EQ 6.81}$$

$$\text{Now} \quad V_{o(s)} = \quad 2R_L\, i_{in(s)}\, \alpha \tag{EQ 6.82}$$

$$\therefore\ A_{(s)} = \frac{V_{o(s)}}{V_{sig(s)}} = \quad \frac{2R_L\omega_n^2\alpha}{R_T\left(s^2 + 2\zeta\omega_n s + \omega_n^2\right)} \tag{EQ 6.83}$$

$$\text{where} \quad R_T = 2XR_D + 2r_e = 2(XR_D + r_e)$$

We could cancel out the two's then to get

$$A_{(s)} = \quad \frac{R_L\omega_n^2\alpha}{(XR_D + r_e)\left(s^2 + 2\zeta\omega_n s + \omega_n^2\right)} \tag{EQ 6.84}$$

which would be the single ended gain which is the same as the differential gain.

Since $2\zeta W_n = \frac{1}{R_T C}$ we can see the effect of the series diodes and the input resistance $2r_e$ of the amplifier on the gain. The gain is inversely proportional to the matrix diodes added.

If we were to damp the circuit for a zeta of .707 for maximally flat current input and/or gain as a function of frequency we would need

$$R_{SERIES_{TOTAL}} = \frac{1}{2\zeta\omega_n C} = \frac{1}{(1.414)(\omega_n C)} \tag{EQ 6.85}$$

If we had a 10μh head with 25.PF capacitive load, we would require

$$R_T = \frac{\sqrt{10^{-5} \cdot 2.5\times10^{-11}}}{1.414\left(2.5\times10^{-11}\right)} = 4.47\times10^{2}\,\Omega \tag{EQ 6.86}$$

Obviously this would cut down the gain available considerably.

ADDENDUM FOR CHAPTER 6

If we restricted ourselves to $\frac{R_T}{2} = XR_D + r_e = 13\Omega + 13\Omega$ then the Bode plot as shown in Figure 6.20 is obtained.

$$A_{corner} = \frac{\frac{1}{XRC} - \sqrt{\left(\frac{1}{XRC}\right)^2 - 4\frac{1}{LC}}}{2} = 5.247\times10^6 \text{ RAD or } 8.35\times10^5 \text{ HZ} \quad \text{(EQ 6.87)}$$

$$B_{corner} = \frac{\frac{1}{XRC} + \sqrt{\left(\frac{1}{XRC}\right)^2 - 4\frac{1}{LC}}}{2} = 1.536\times10^9 \text{ RAD} \approx 2.44\times10^8 \text{ HZ} \quad \text{(EQ 6.88)}$$

The significance of all this is that for signals of interest between the corners we have a 6 db/octave gain reduction. It might be of interest to compare the gain equation if there were no capacitor. Fig 6.21 A, B

$$A_i = 2R_L\alpha i_{in} = \frac{2R_L\alpha}{LS + 2XR_T + 2r_e} = \frac{2R_L\alpha}{L\left(S + \frac{2(XR_D + r_e)}{L}\right)} \quad \text{(EQ 6.89)}$$

The Bode plot as shown in Figure 6.22 has only one pole located at L/R_T. For our 10 μh head with 2(26)Ω series resistance it would be located at $5.2\text{X}10^6$ radians or $8.27\text{X}10^5$ Hz. The addition of the capacitance then makes a considerable difference in the plot.

The advantages of such an amplifier, of course, is the reduction of noise. This is achieved from two sources -- first the reduction in bandwidth, and second the very low input impedance lowers the electrostatic noise field interference, but it does worsen the electromagnetic noise interference. Shielding and twisted cable will then help. We will discuss the bandwidth effects at a later time.

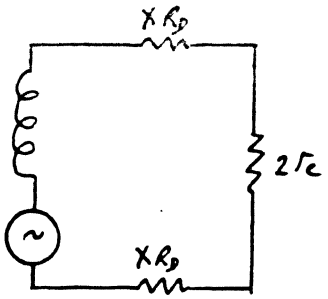


FIG 6.22 A

NO CAPACITIVE LOADING EQUIVILENT CIRCUIT

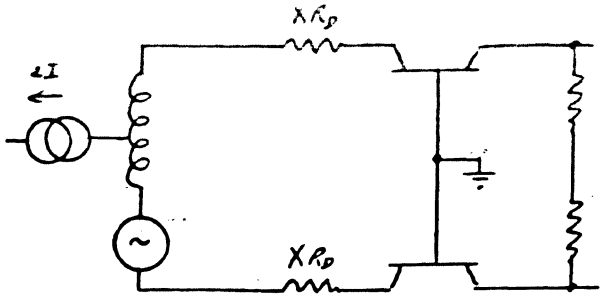


FIG 6.22B

GAIN STAGE

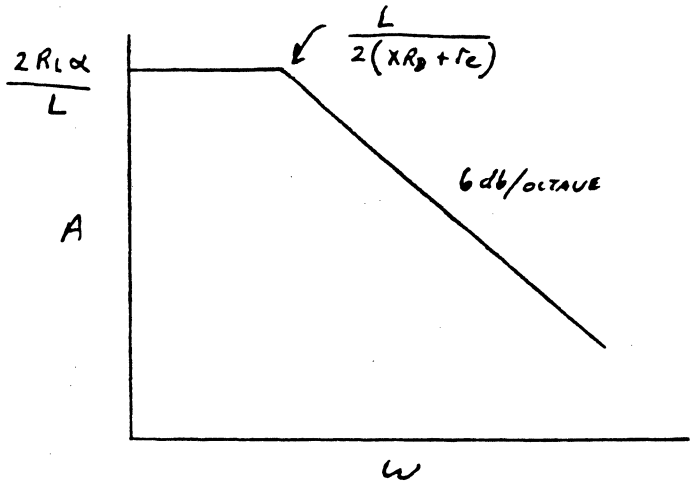


FIG 6.23

BODE PLOT FOR NO CAPACITIVE LOADING

MATRIX CIRCUITS

As shown in the Block Diagram, Figure 5.1, when more than one head is to be used alternately, or non simultaneously, then some means must be provided to electrically separate the heads both during writing and during reading. Early machines accomplished this function by the use of centertapped heads and a diode or diode transistor network.

First let us discuss the write separation function.

In Figure 7.1A two heads are to be electrically separated by the use of diodes. Current flow is from the PNP transistor T_1 emitter-collector to the head centertap, thru 1/2 the head winding, the series diode, the conducting Write Driver transistor and out the current source, I source. The reversed voltages and diode polarities are used for the PNP version of Figure 7.1B. Again, as discussed in Chapter 6 on Write Drivers, collector saturation is avoided by providing a sufficient voltage on Vin or $\overline{Vin}$ to allow for the transient, but now also the transistor T, V_{CE} sat and the diode drop. The diode direction is in the direction of current flow when writing so unless some means of reverse biasing them is provided the capacitance of the parallel heads is still connected during the transient. Resistor R are added to reverse bias these diodes when the centertap transistor is cut off. The voltage chosen must exceed the transient. When calculating transistor current this extra Resistor current must be added. The reverse bias voltage must be greater than the transient voltage peak value. Figure 7.2 A and B show the relationships.

Figure 7.2 B is a good way to visualize the various bias drops required to maintain linearity. The transients in both the negative and positive direction are compatible with the diode polarity; therefore, there need only be one damping resistor for all the heads, provided they are all the same, of course, See Figure 7.2C. The capacitance of the head is increased

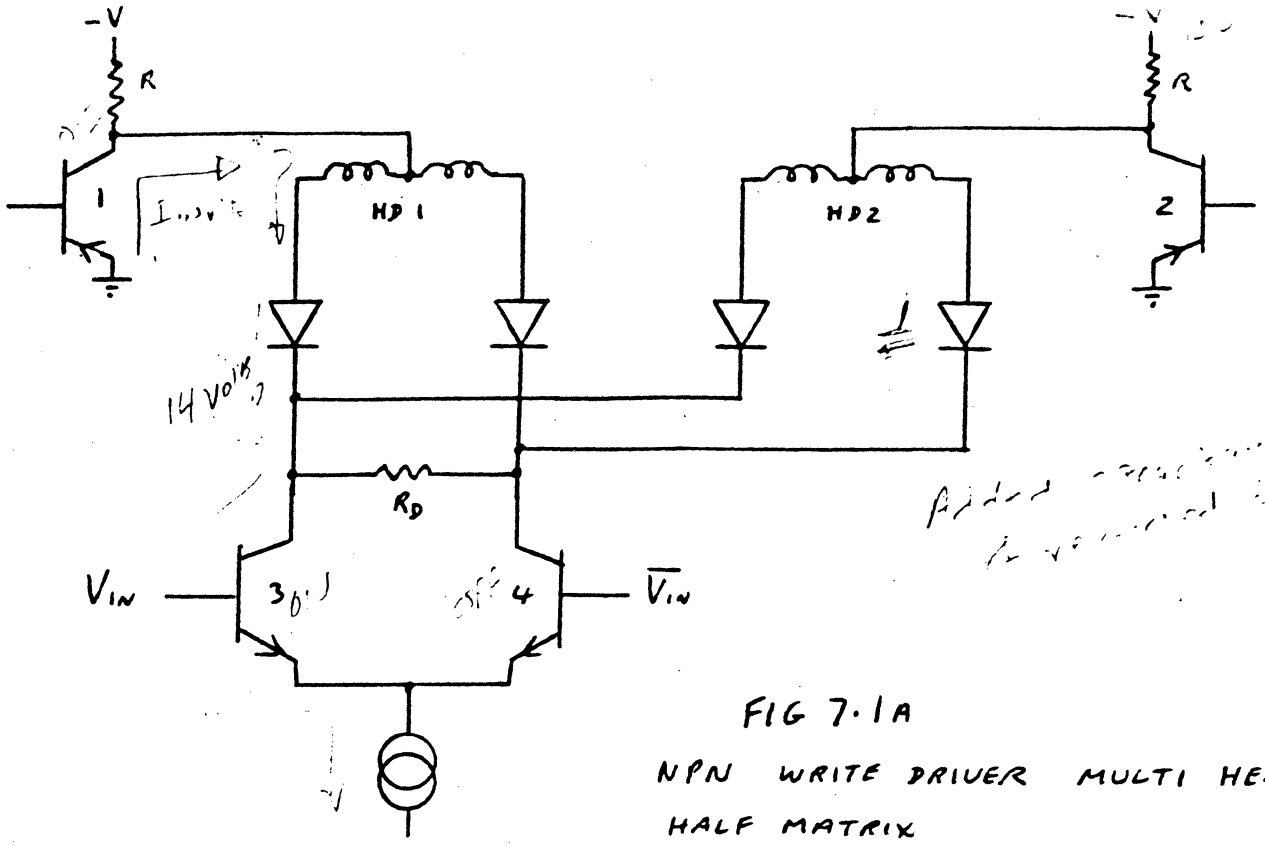


FIG 7.1A

NPN WRITE DRIVER MULTI HEAD HALF MATRIX

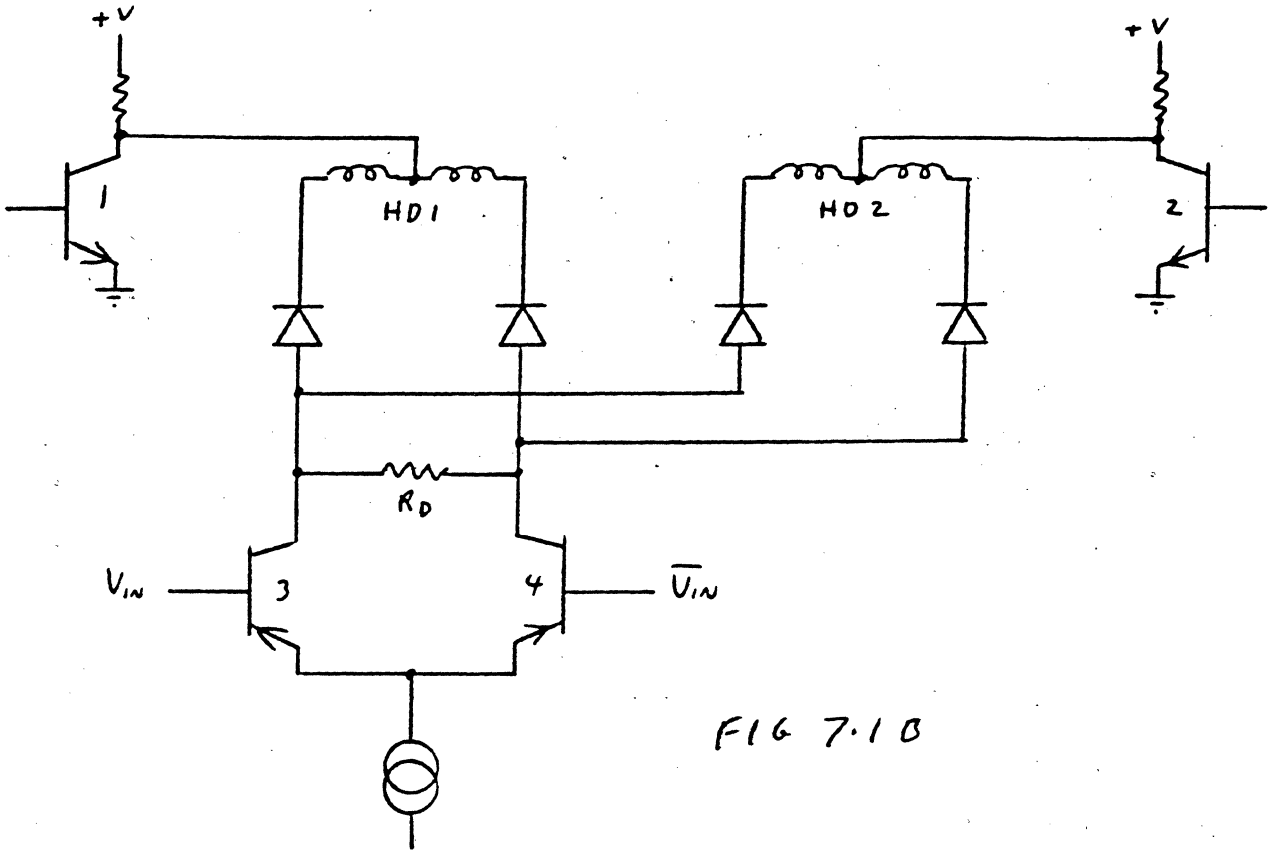


FIG 7.1B

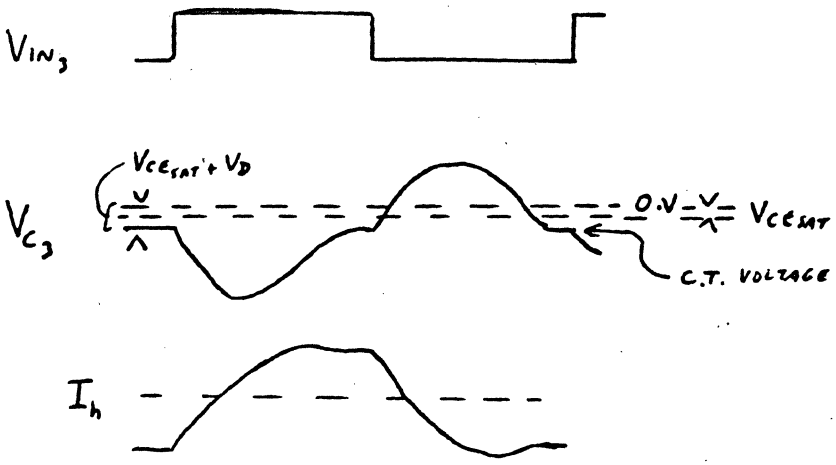


FIG 7.2 A

WAVEFORMS FOR CIRCUIT OF FIG 7.1A

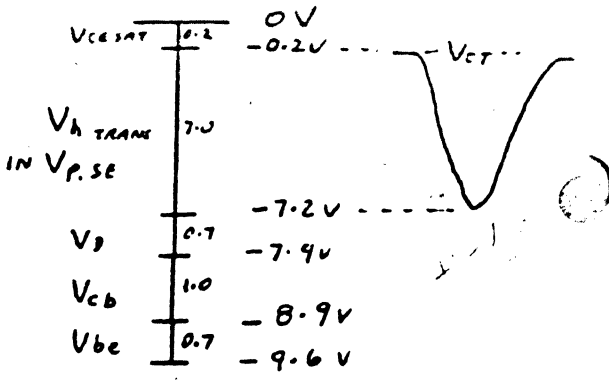


FIG 7.2 B

BIAS DIAGRAM FOR CIRCUIT OF FIG 7.1A

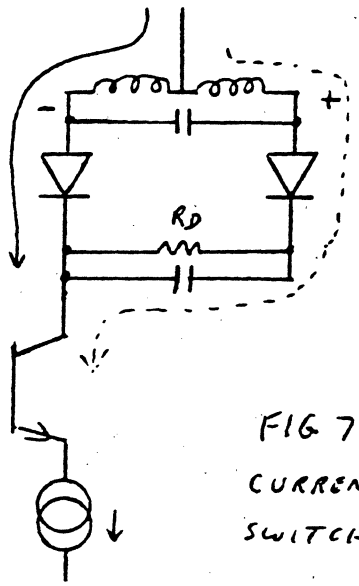


FIG 7.2 C

CURRENT FLOW DURING SWITCHING TRANSITION

as expected as a function of the capacitance of the reversed biased diodes and any parallel capacitance. This is shown in Figure 7.3. Looking at only one side for convenience, we see that the 2 (head capacitance) becomes.

$$2C_T = 2C_h + C_{w1} + C_{w3} + \frac{\frac{(C_D)(C_{w2} + 2C_h)(C_{w4} + C_{Tr1})}{(C_D + C_{w2} + 2C_h)}}{\frac{(C_D)(C_{w2} + 2C_h)}{C_D + C_{w2} + 2C_h} + C_{w4} + C_{Tr1}} \qquad \text{(EQ. 7.1)}$$

As we look at the matrix and lump together some of the wiring capacitances Cw_{1-4} then we can write a new equation.

$$2C_T = 2C_h + \frac{(C_D)(2C_h)}{C_D + 2C_h} \qquad \text{(EQ. 7.2)}$$

This is handy because we can now address the case where there are more than two heads that are separated by the matrix circuits. It is obvious that just adding more and more heads in parallel will just increase the capacitance and thus lower Wn which slows down the rise time. If we can make the matrix two level, meaning that we group the heads into subgroups and then connect them to the write driver thru a second diode, we can take advantage of this series parallel network to reduce the capacitance. The general equation becomes EQ 7.3 if there are B branches of X sub branches making a total of X•B = N heads.

$$2C_T = 2C_h + \frac{(X-1)(2C_h)(C_D)}{2C_h + C_D} + \frac{\left[\frac{X(2C_h)(C_D)}{2C_h + C_D}\right] C_D(B-1)}{\frac{X(2C_h)(C_D)}{2C_h + C_D} + C_D} \qquad \text{(EQ. 7.3)}$$

It is easy to see that this equation can be minimized as a function of X and B if we substitute C_E as the equivalent of 2 C_h and C_D in series.

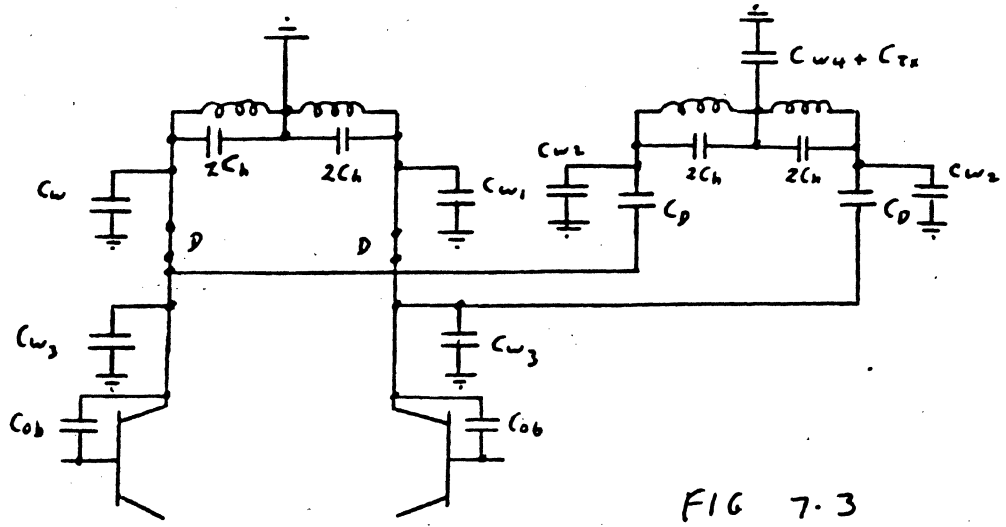


FIG 7.3

CAPACITANCE OF MULTIPLY CONNECTED HEADS

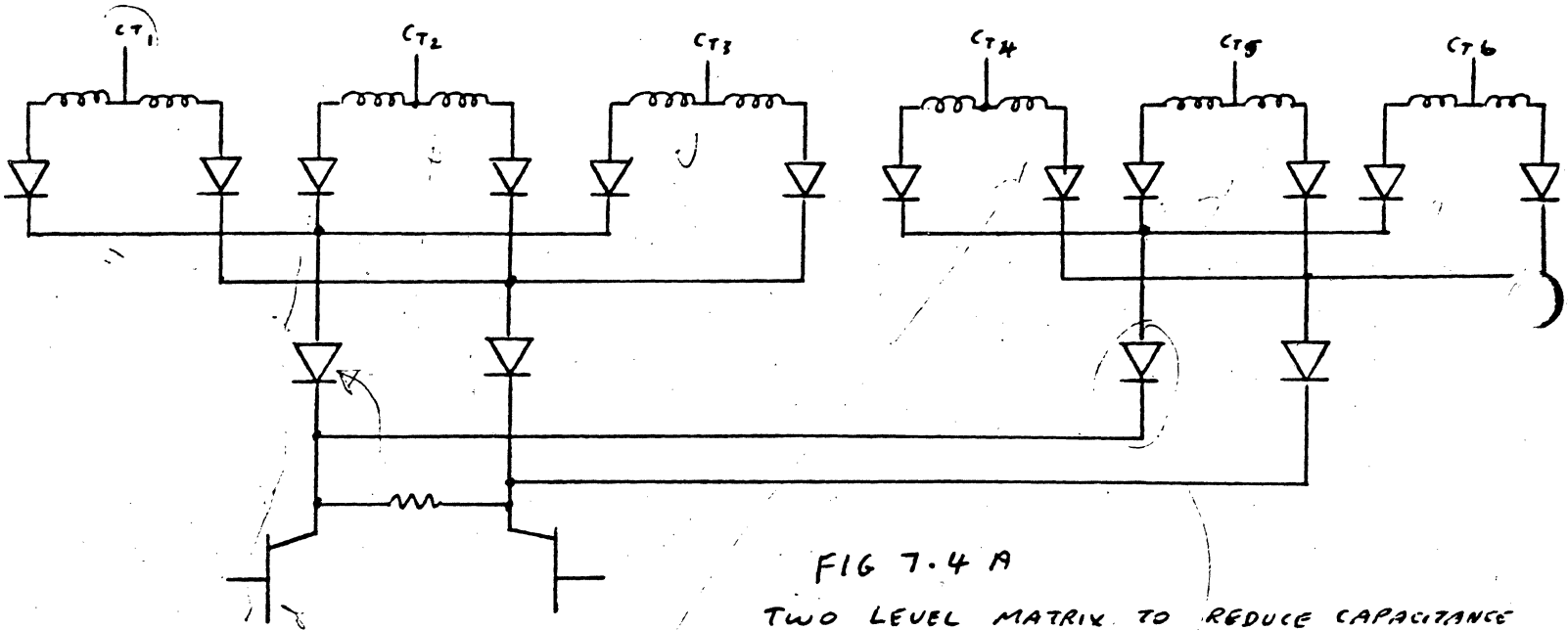


FIG 7.4 A

TWO LEVEL MATRIX TO REDUCE CAPACITANCE

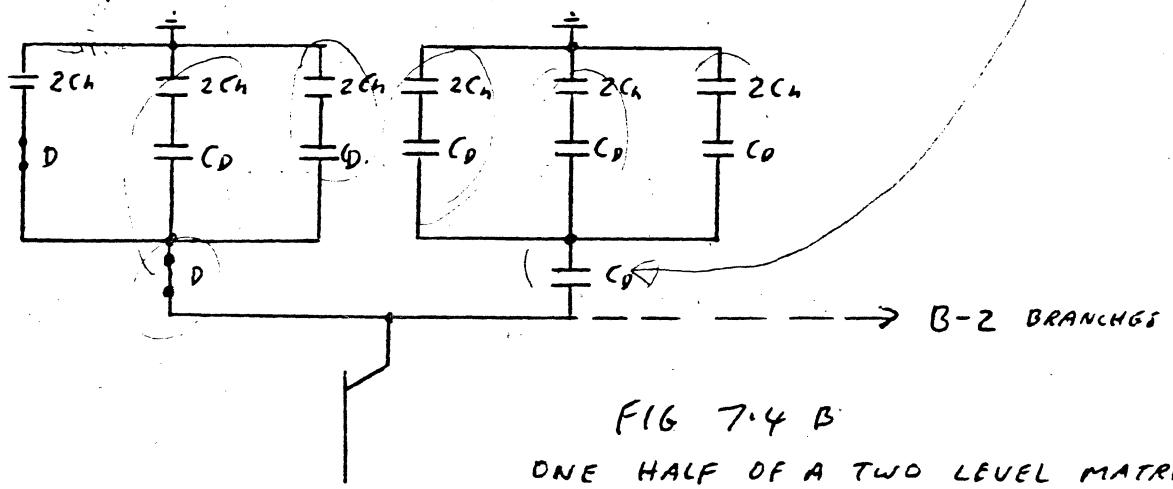


FIG 7.4 B

ONE HALF OF A TWO LEVEL MATRIX OR EXTENDED TO A 'B' LEVEL MATRIX

$$2C_T = 2C_h + (X-1)(C_\epsilon) + \frac{(X)(C_\epsilon)(C_D)}{XC_\epsilon + C_D}(B-1) \qquad \text{(EQ 7.4)}$$

For example let $X \cdot B = N$ heads $\therefore \quad X = \frac{N}{B}$

$$2C_T - 2C_h = \left(\frac{N}{B} - 1\right)C_\epsilon + \left[\frac{\frac{N}{B} C_\epsilon C_D}{\frac{N}{B} C_\epsilon + C_D}\right](B-1) \qquad \text{(EQ 7.5)}$$

This equation can then be solved for the desired number of heads as a function of the two groups which will minimize the head capacitance. Each head's centertap has its own transistor and reverse biasing resistor. These transistors can then be controlled by a decoder operating from a register. The input base level must be corrected for the transistor emitter voltage chosen. In the example this voltage is ground; therefore, the bases will need to be driven negative.

Figure 7.5A shows one method of interfacing T^2L logic blocks. Ground is the best level to return the head to because of the noise usually on the supply voltages. Other configurations are possible that use some reference voltage as long as that reference is quiet electrically. The extra series diodes are considered when making up the bias diagram for the total circuit. This includes the select transistor, all series diodes, any head resistance, the maximum voltage transient (in one direction), any required reverse bias of the Write Drivers, and lastly the variation in base voltage from the Pre Driver Circuit.

The type of diode chosen depends on the write current. Usually a high conductance diode with as low a C_D as possible is best. Also, the leakage current when reversed biased is very essential as it affects noise in the

network during a read which will be discussed next. A IN4448 diode serves well in this position if the reverse leakage is specified.

The second function of the Matrix is to connect the selected head to the Pre Amplifier as well as block the large voltage swings of the write function from damaging or disturbing the Pre Amplifier. This function is not so straight forward as was the circuit for isolation during write.

Consider the circuit of Figure 7.6A. The nodes A and B can be called the main nodes. Branching off from the main node is the Write Driver circuit isolated with a pair of diodes, D_1 and D_2. The Write Driver circuit also includes the Write Damping network. It should be noted that current flow from the reverse bias source R_1 thru the centertapped write damping resistor subtracts from the write current as seen by the head. This reverse bias is necessary in order to isolate both the Write Driver capacitance and the Write Damping resistors from affecting the read function. It can now be seen that any leakage in any reverse biased diode will affect the read signal. The problem with reading is that the read signal is A.C.; therefore, using diodes not only would form a half wave rectifier but silicon diodes would not even conduct. One way this can be accomplished is to force a small current thru the head and diodes such that they form a conducting path to the Pre Amplifier. The currents for both halfs of the head cancel their flux therefore the data is not disturbed magnetically. About 2.0 ma is necessary in order to adequately forward bias the diodes to a sufficiently low series resistance. The Head AC signal now modulates this current which passes the signal to the Pre Amplifier. Resistors R_2 and R_3 are tied to a negative voltage in this example to supply

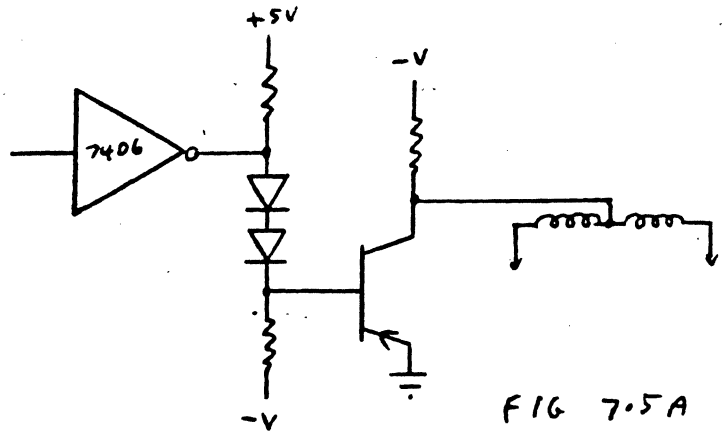


FIG 7.5A

HEAD SELECT CIRCUITS FOR A GROUNDED PNP TRANSISTOR (NPN WRITE DRIVER)

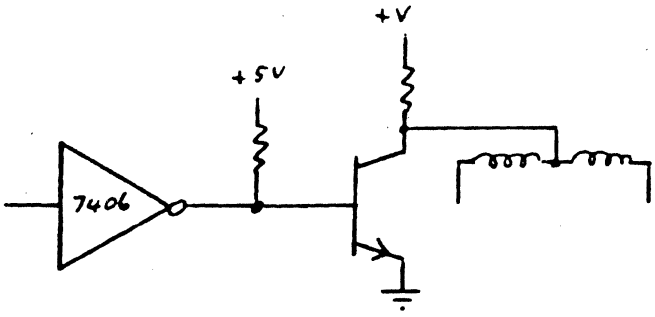


FIG 7.5B

HEAD SELECT CIRCUIT FOR A GROUNDED NPN TRANSISTOR (PNP WRITE DRIVER)

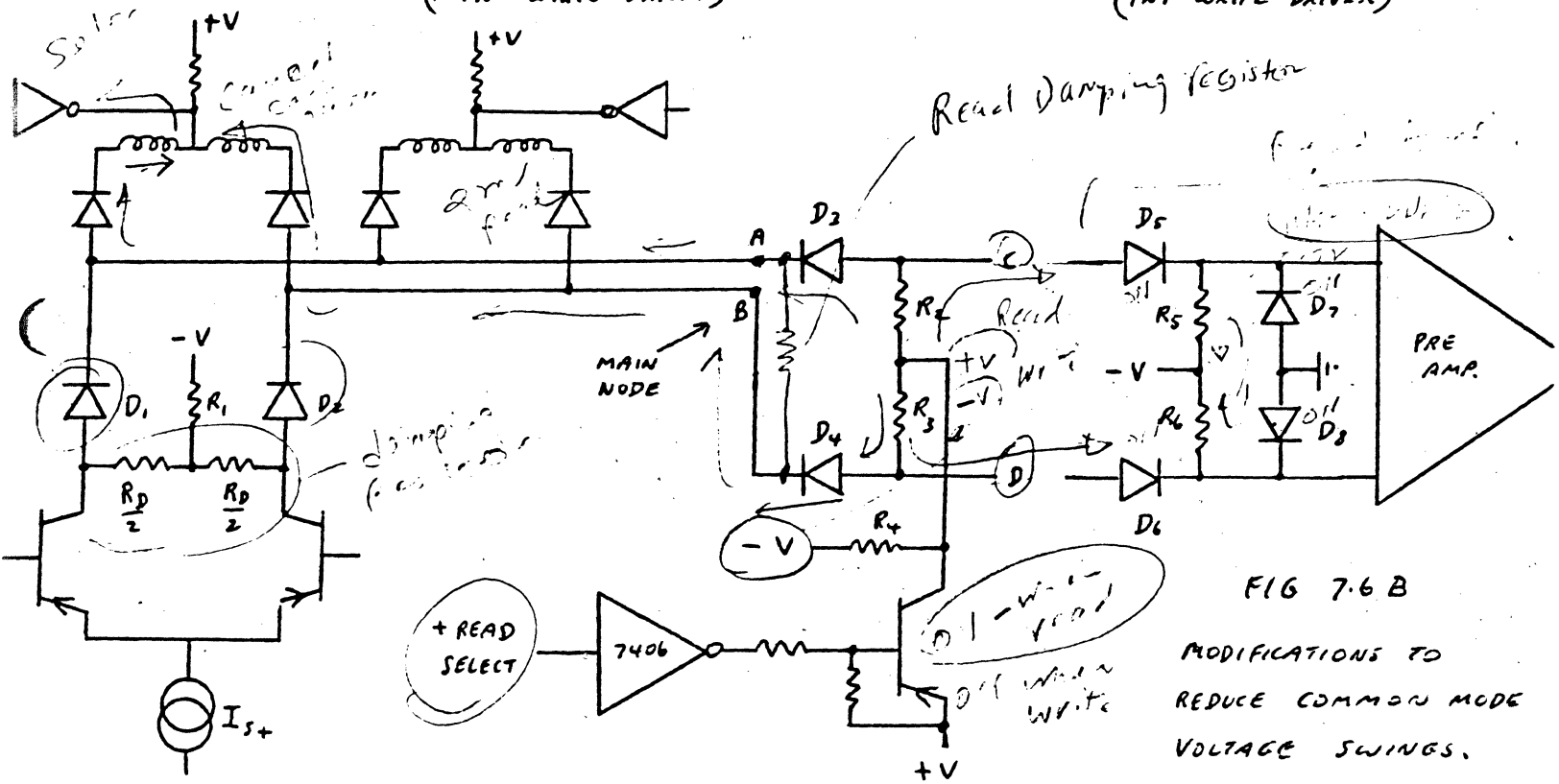


FIG 7.6A

SINGLE LEVEL MATRIX WITH READ CAPABILITY

FIG 7.6B

MODIFICATIONS TO REDUCE COMMON MODE VOLTAGE SWINGS.

the desired current. The DC voltage at this node $_{C,D}$ is equal to

$$V_{CE\ SAT} + \left(I_{bias}\right)\left(\frac{R_4}{2}\right) + 2V_D = V_C \stackrel{?}{=} V_D \qquad \text{(EQ 7.6)}$$

$$\text{WHERE} \qquad I_{bias} = \frac{+V - V_{CE\ SAT_1} - V_C}{R_2} \qquad \text{(EQ 7.7)}$$

When writing the head voltage transient as given in EQ 5.23 needs to be blocked by Diodes D_3 and D_4 ; therefore, R_4 is included to provide the reverse bias. This then means that when writing the Pre Amplifier sees -V on its input. This may not be desirable particularly for some commercial types. The circuit is modified as shown in Figure 7.6B to add another pair of Diodes D_5 and D_6 to block the large write transient blocking voltage. Another pair of resistors now need to be added to supply current thru D_5 and D_6 when reading to forward bias them. Also yet another pair of diodes need to be added to clamp this voltage when writing to a value tolerated by the Pre Amplifier.

The current flows are now much more complicated. The extra current D_5 and D_6 needs to be supplied thru R_2 and R_3. This current splits between D_5 and D_3 as also D_4 and D_6.

We can come close to the real currents as we assume that the diode drops are referenced to the head centertap voltage as shown in EQ 7.8 to the Pre Amplifier input.

$$V_{PA\ IN} = V_{CE\ SAT} + \left(I_{bias}\right)\left(\frac{R_4}{2}\right) + V_D + V_{D_3} - V_{D_5} \qquad \text{(EQ 7.8)}$$

$$\text{WHERE} \quad I_{bias} = \frac{+V - V_{CE\ SAT_1} - \left(V_{CE\ SAT} + \left(I_{bias}\right)\left(\frac{R_4}{2}\right) + V_D + V_{D_3}\right)}{2R_2} \qquad \text{(EQ 7.9)}$$

recognizing that we want the current to split at this point.

Now we can make the Resistor Rs equal to

$$R_S = \frac{V_{BIAS} + (-V)}{I_{BIAS}} \qquad \text{(EQ 7.10)}$$

All this assumes that the V_D drops are equal which is of course not true. These errors will show up as a small unbalance in current in the head as well as an imbalance in voltage at the Pre Amplifier inputs. This latter is disastrous as these voltages are usually several tenths of a volt which the Pre Amp cannot handle without saturating. Going back to the Pre Amplifier circuit of Figure 6.13 and Figure 6.15 or Figure 6.6, we can see a solution. The coupling capacitor in the emitter feedback path effectively isolates the two input mismatches. All we are left with is a small differential unbalance due to the unequal attentuation thru the diodes. This affects the Common Mode Rejection Ratio of the amplifier which needs to be high. The function of Read Damping is accomplished either thru the network or by the addition of another resistor across the output terminals. This is necessary due to the different value of Zeta between Read and Write. A better position would be across the main node. This way the attenuation is lessened. The resistor value for Read Damping is higher than for write damping; therefore, we can leave the Read Damping across the main node for both Read and Write and make the Write Damping resistor for the parallel function to get the lower value required (See Figure 7.8). The attenuation of the head signal is calculated from three simultaneous equations.

$$V_{SIG} = C_1\left(R_L + LS + \overset{A}{R_D} + R_{DAMP} + R_D\right) - C_2\left(\overset{B}{R_{DAMP}}\right) \qquad \text{(EQ 7.11)}$$

$$0.\ V = -i_1 \overset{B}{(R_{Amp})} + i_2 \overset{C}{(R_{Amp} + R_{D_3} + R_2 + R_5 + R_{D_4})} - i_3 \overset{D}{(R_2 + R_5)} \quad \text{(EQ 7.12)}$$

$$0\ V = \cdot C_1(0) - i_2 \overset{D}{(R_2 + R_5)} + i_3 \overset{E}{(R_2 + R_5 + R_{D_5} + R_7 + R_6 + R_{D_6})} \quad \text{(EQ 7.13)}$$

$$\mathcal{L} = (R_7 + R_6) i_3 = (R_7 + R_6) \frac{\begin{vmatrix} A & -B & V_{sig} \\ -B & C & 0 \\ 0 & -D & 0 \end{vmatrix}}{\begin{vmatrix} A & -B & 0 \\ -B & C & -D \\ 0 & -D & E \end{vmatrix}} = \frac{(B.D.\ V_{sig})(R_7 + R_6)}{A.C.E. - B^2E - D^2A} \quad \text{(EQ 7.14)}$$

From the above it can easily be seen that if the Read Damping resistor were in place of R_5 and R_6 then its necessary low value would greatly attenuate the read signal. As it is, only one pair of diodes cause the main attenuation. All following attenuation is small, (R_2 + R_3 and R_5 + R_6), if the bias voltages are high enough.

We know the Pre Amplifier input voltage from EQ 7.8 and 7.9 is the Reading input voltage. When writing the input voltage is clamped by diodes D_7 and D_8. The input voltage during write then is $V_{D_{7\text{or}8}}$. This is common mode and is about -0.7 Volts. It can now be seen that when switching from Read to Write and from Write to Read the amplifier input voltage common mode goes from +0.7 volts to -0.7 volts or a 1.4V change. As long as these voltages are true common mode, the amplifier sees no transient, but due to the tolerances previously mentioned, some difference remains which forces large step changes in input voltage. These must be amplified for the duration of the time constants involved. The case where the Pre Amplifier is driven by current sources is easily calculated.

$$T_{TRANSIENT} = \frac{C_c (\Delta V_{DIFF})}{I_{SOURCE\ AMP}} \quad \text{(EQ 7.15)}$$

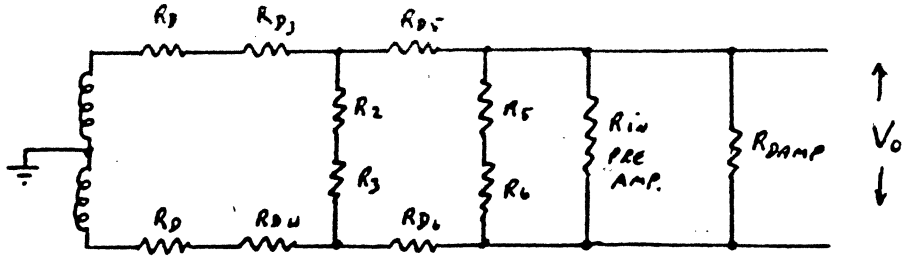


FIG 7·7

ATTENUATION IN THE MATRIX

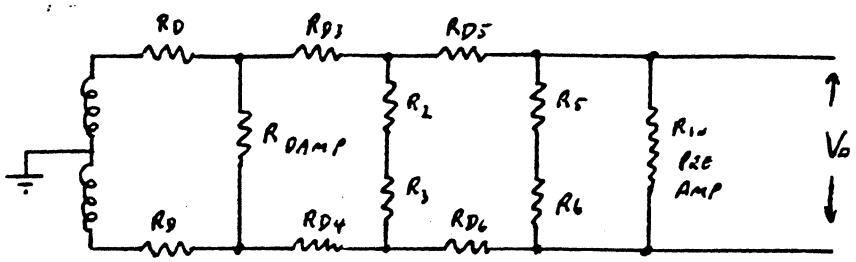


FIG 7·8 A

READ DAMPING AT MAIN NODE

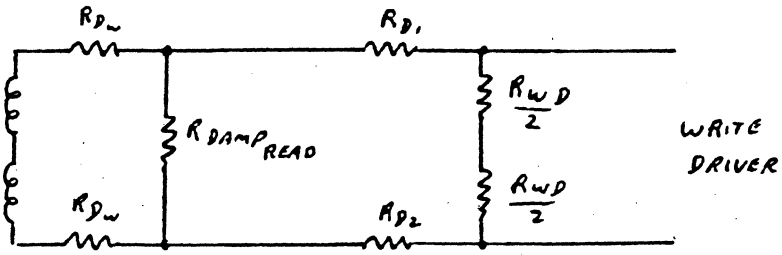


FIG 7·8 B

WRITE DAMPING A FUNCTION OF READ DAMPING IN PARR. WITH A HIGER VALUE R_{WDAMP}

The actual transient time is slightly longer due to the conduction of the previously cut-off transistor during the last portion of the transient (about 0.1 volt). We can calculate ΔV_{diff} in the worse case by assuming all diodes in the upper half of the matrix have high voltage drops and all the diodes in the lower half have low voltage drops for similar currents. Also we can take the tolerances in the resistors in such a direction to accentuate the problem.

$$\Delta V_{Diff} = 3\,V_{D\,max\,(AT\,I_{max})} - 3\,V_{D\,min\,(AT\,I_{min})} \qquad \text{(EQ 7.16)}$$

The associated time must be accounted for in any selection process.

In the multi level matrix the equations for attentuation and ΔV_{diff} need to be modified to include the extra levels of diodes or diode drops.

For a two level matrix as shown in Figure 7.4, for example, Equation 7.8 is modified as in EQ 7.17.

$$V_{PA\,IN} = V_{CE\,SAT} + \left(I_{bias}\right)\left(\frac{R_h}{2}\right) + 2\,V_D + V_{D_3} - V_{DS} \simeq +1.4\,V \qquad \text{(EQ 7.17)}$$

$$\text{Le } \Delta V \text{ common mode is} \quad +1.4\,V - (-0.7) = 2.1\,V \qquad \text{(EQ 7.18)}$$

and

$$\Delta V_{Diff} = 4\,V_{D\,max\,(AT\,I_{max})} - 4\,V_{D\,min\,(AT\,I_{min})} \qquad \text{(EQ 7.19)}$$

similarly we must double both left hand R_D's of Figure 7.8 to get the attenuation.

As can be imagined, the losses in these networks are substantial. Also, the noise induced by the diode noise and the reduction in Common Mode Rejection play a heavy role in reducing the signal to noise ratio. Obviously, such methods can only be used if the head signal is large compared to the total system noise.

A further problem is that the write transients are not completely blocked from the Pre Amplifier. If we replace the reversed biased diodes leading to the Pre Amplifier by capacitors we can easily see how the Pre Amplifier is overdriven. See Fig 7.10.

If R_D and R_D are about 10Ω, the transient is simply

$$\frac{(V_{transient})(10\Omega)}{10\Omega + \frac{1}{Cs}} \qquad \text{(EQ 7.20)}$$

If we assume the fundamental frequency component of the transient is 5MHz then the signal transferred is

$$\frac{(7.0V_{p\,p\,se})(10\Omega)}{10\Omega - j\,\frac{1}{(2\pi)(5\times10^{6})(1\times10^{-12})}} = \frac{70.0V_{p\,p\,se}\,\Omega}{10 - j\,3.18\times10^{4}} = 2.20\,mV \qquad \text{(EQ 7.21)}$$

These matrices can be inverted polarity wise by using NPN write drivers, reversing all diodes and bias voltages, including the Read Select transistor network feeding R_2 and R_3.

A much better method is to provide a separate Pre Amplifier and Write Driver for each head. This has been done in several disc drives since 1977 to great advantage. By mounting an IC containing these circuits on the head or near it most stray capacitances are reduced, common mode unbalances are eliminated and stray field pick up is reduced. Modern Integrated circuitry can perform this function for several heads at once when properly addressed. A general circuit is shown in Figure 7.11.

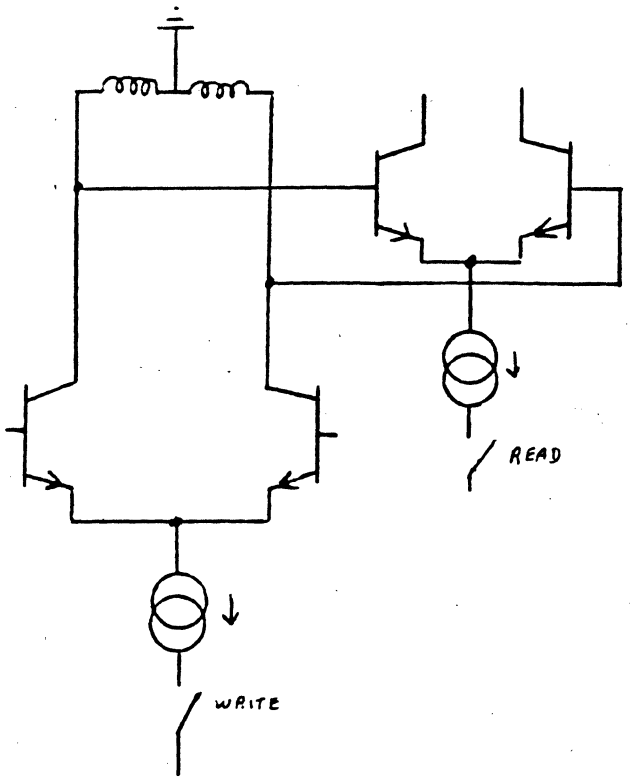


FIG 7.9

SEPARATE HEAD WRITE DRIVER AND PRE AMP.
DIRECT CONNECTION

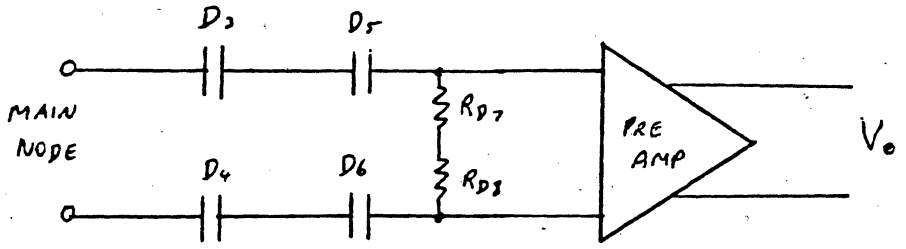


FIG 7.10 A

SHOWING DIODE CAPACITANCES INVOLVED WITH WRITE VOLTAGE NOISE COUPLING

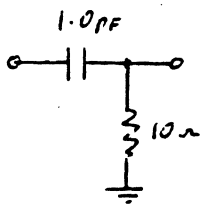


FIG 7.10 B

EQUIVILENT CIRCUIT

The current source can be switched on and off for writing. When reading the Pre Amplifier can be activated by biasing. Immediately we can see one problem and that is that the write transient is seen by the Pre Amplifier bases. Depending on the head inductance and the required write current, this voltage breaks down the base emitter junctions as one is forward biased and the other is zenered which reverses on the next transient. For integrated circuits there are two detrimental results depending on the circuit values.

Zenering an emitter - base junction causes low current β to be degraded. This is of little consequence if we use higher currents to bias the first stage of the amplifier, usually above 0.5 ma. The second effect is when the forward biased base-emitter junction conducts it effectively saturates the associated transistor which turns on a substrate PNP transistor altering the biases. This latter effect is eliminated if the transient duration is small compared to the turn on time of the substrate PNP.

When more than one head is serviced from the same integrated circuit, the multiplexing function is achieved by appropriately shifting the biases to favor the selected circuit. For example, the write drivers can all be in parallel from the same current sources but the base drive can be raised a few volts for the selected write driver while the remainder are held, thus cutting them off, or the current sources can be switched separately. The Pre Amplifier is easily selected if the first stage is our favorite cascode circuit. Here each head is connected to its own first half of a cascode amplifier with its own switched current source. The other stages are *parralled* to this mode with the upper half being a single circuit serving all.

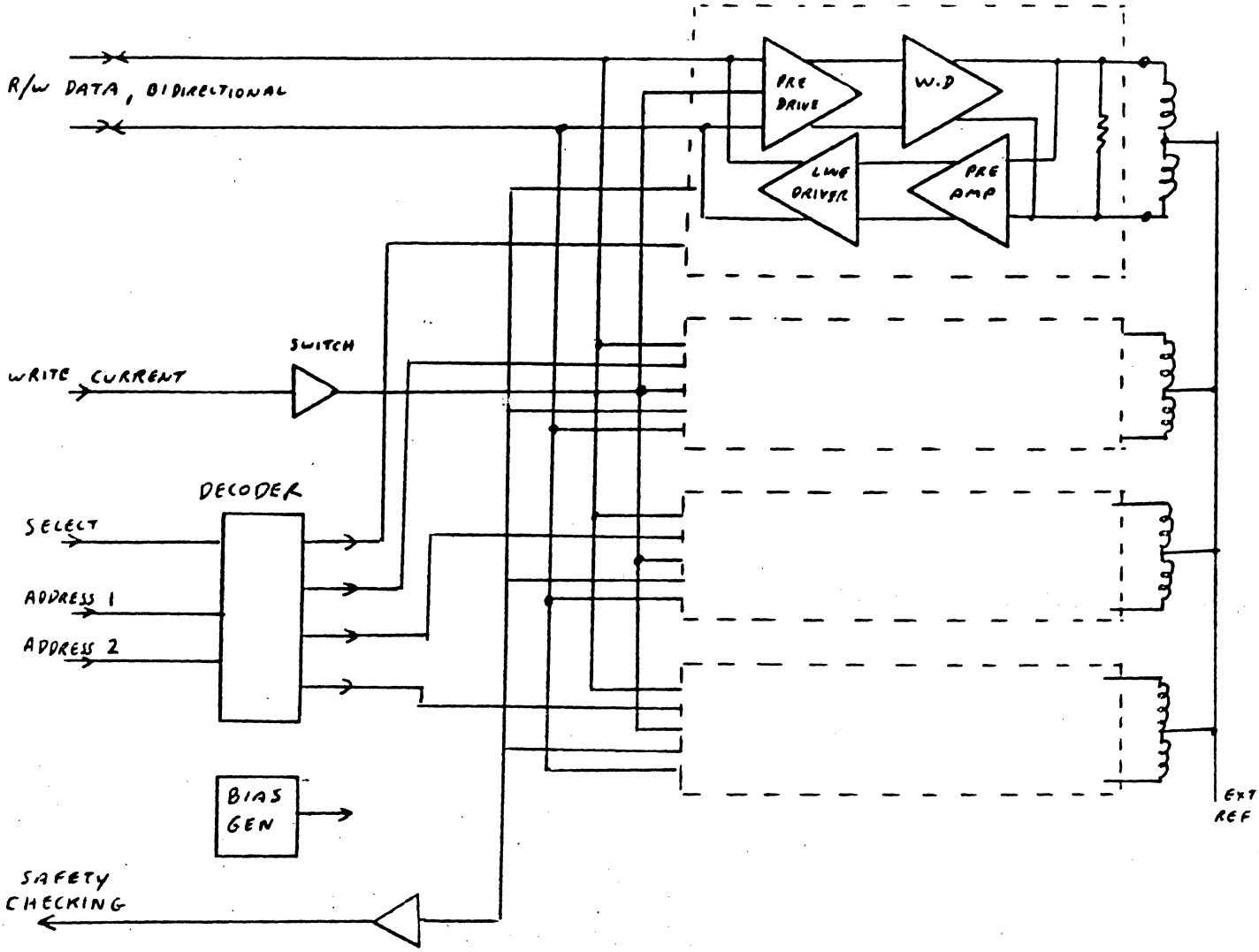


FIG 7.11

INTERGRATED READ WRITE CIRCUIT FOR 4 HEADS WITH ADDRESSING AND SAFETY CHECKING CIRCUITS

When designing such an IC, great care should be exercised in considering internal biases and power dissipation. Figure 7.11 shows a typical circuit presently used for centertapped heads. By proper consideration many IC's can be paralled to address many heads by using parallel circuitry and address lines.

We might profitably consider the serial time required by these multiflexed circuits when handling data. Before the first transition can be recorded, the write current source must be turned on and the current built up to final value in the head. This is typically about 100 ns or more. Worse, when going from writing to reading we must turn off the write current source, turn on the Read bias circuits, recover the Pre Amplifier from the select transient back to the base line and include any following AC coupling in later circuits. Then times can be as great as 5 - 30μs depending on circuit bandwidths.

SAFETY CIRCUITS

This chapter is included at this point in the presentation because we now have completed all the basic circuits that interface with the head. The subject of safety has to do with the recorded data that resides on the disc or tape. Since this data represents the accumulated efforts of some programmer or computer operator then every means must be taken to assure the user that their data is not disturbed due to any malfunction of the circuits themselves. Of course, no amount of checking can protect any data if the machine receives a valid command to write even if it was intended or not. We will confine ourselves to only those malfunctions that are invalid or a result of component failure.

There are several checks that will indicate a possible endangering of data. They are:

1. Write Current and no Write Command
2. Write Command and no Write Current
3. Write Command and no Write Data
4. Write Current during a protected data field
5. Write Current and not directly over the assigned track
6. Write Command and a Read Command simultaneously
7. More than one head selected at once (serial machines)
8. Open heads
9. Shorted heads
10. Centertap Current Sense

We will discuss each of these and propose circuits that can be used to sense the failure.

1. Write Current and No Write Command

If a circuit is provided that senses the presence of write current the output can be 'anded' with a signal indicating no command is present. The sensing circuit depends on the type of head that is used and also the type

of write driver chosen. For two terminal heads driven by saturated switches similar to that shown in Figure 5.4, we have several choices. We could monitor the current drawn from the +V supply by providing a single series resistor. Then thru the use of a suitably biased comparator any current exceeding an acceptable value could be indicated. The problem of sensitivity could be overcome by using a high valued resistor in parallel with a diode as shown in Figure 8.1A.

With this circuit the V+ line previously used to calculate the Write Currents must be modified to subtract one Vbe drop during operation. Sensitivity can now be made as high as needed within the bias and common mode requirements of the comparator. If no appropriately higher voltage is available to operate the comparator a current mirror could be used that makes use of the diode as a reference as shown in Figure 8.1B. Here the current mirror is not a true mirror due to the differences in V_D and Vbe of the two descrete devices, but sufficient current can be guaranteed to operate the following logic block. These circuits could be used with almost all of the write drivers shown in Chapter 7. The complimentary pair driver of Figure 5.1 would need two, one for each source or the emitter voltage itself could be monitored.

With Current Source driven Write Drivers, the Write Driver emitter circuit could be monitored if a small modification is included. This circuit type is shown in Figure 8.2. Diode D_1 is added to isolate the emitters from the +V voltage expected on the collectors of the Current Source. Again, a comparator is made to sense if this voltage ever goes negative by 2 diode drops or less below the Write Driver bases. Note that the current thru R_1 and R_2 must be appropriately subtracted from the current source current value. The reference

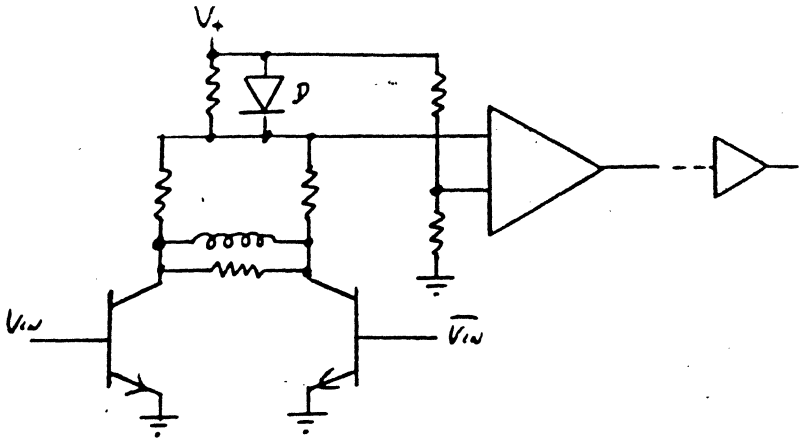


FIG 8.1 A

SATURATED WRITE DRIVER WITH CURRENT SENSE

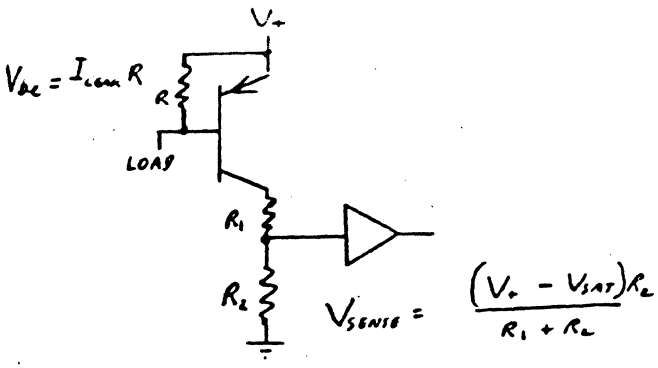


FIG 8.1 B

ALTERNATE LOGIC INTERFACE REPLACING D WITH THE PNP BASE

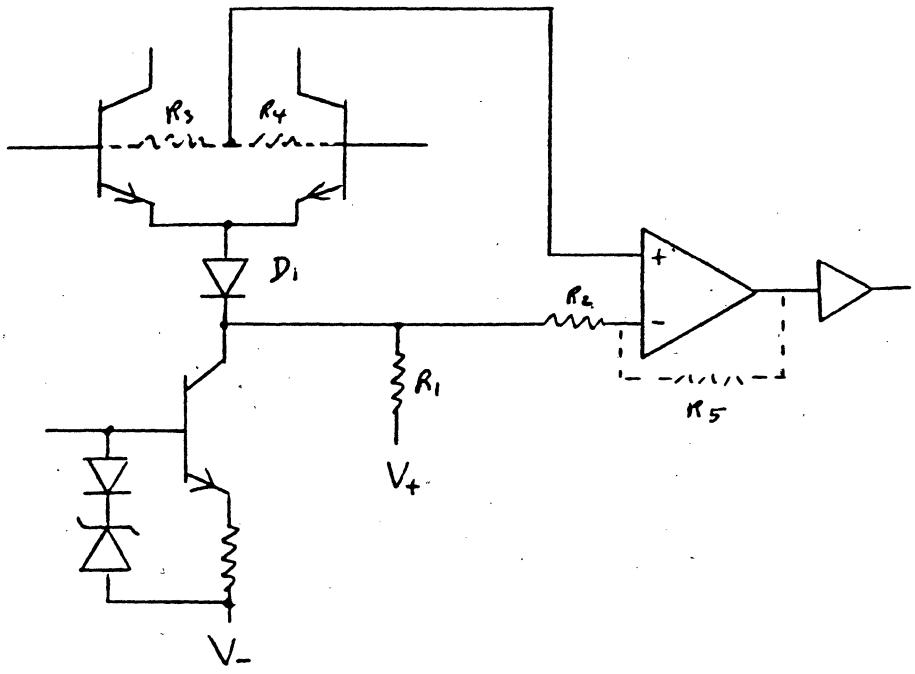


FIG 8.2

CURRENT SENSE FOR CURRENT SOURCE DRIVEN WRITE DRIVERS (NPN)

could be the voltage between the two bases itself which will reduce the worse case calculations. The fact that the base voltages will move up and down with data normally is cancelled out by the 'difference' connection shown dotted R4 R3. Feedback could be applied around the comparator (operational amplifier) R5 if desired. The logic signal resulting from sensing the current can now be 'anded' with the Not-Write command to indicate the unsafe condition. A timing problem now exists that needs addressing. The response time of the current source to the Write Command will be slow on both edges as also the response of the sense curcuit. We are interested, in this safety circuit, if there is write current without a Write Command. When the Write goes off the combined delays of the Write Current Source and the sense circuit will indicate write current well past the trailing edge of the Write Command. This false response needs to be blocked while maintaining the basic function. A simple circuit using a T^2L or DTL logic AND or NAND, Figure 8.3, can cover most delays encountered. The sensitivity of T^2L circuits to slow edges is of no concern due to the use of latches following that hold the fault information.

Similar circuits can be devised using other logic families. In this circuit we take advantage of the construction of a T^2L gate, or a DTL gate. If either input A or B of Figure 8.3B are low all current is removed from the input circuit with none left for the following transistor. When both go high, the the capacitor receives the current until the voltage rises such that the transistor turns on thus initiating a delay in the response of the AND/NAND function. When either A or B go low in response to the eventual fall of the write sense line the capacitor discharges thru R1. When designing for a certain delay the tolerances of Rb and the diode drops and transistor base turn on voltages must be considered. The resistor R6 typically has tolerances of $\pm$ 25%. The delay time becomes close to the following equations when solved for T.

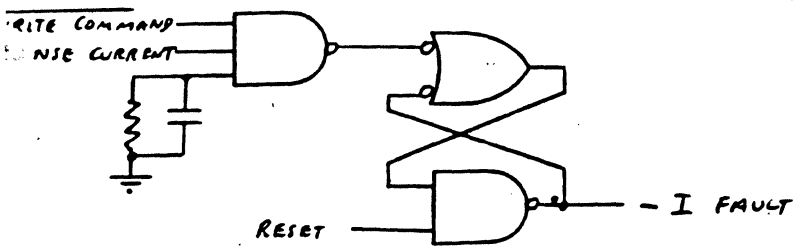


FIG 8.3A

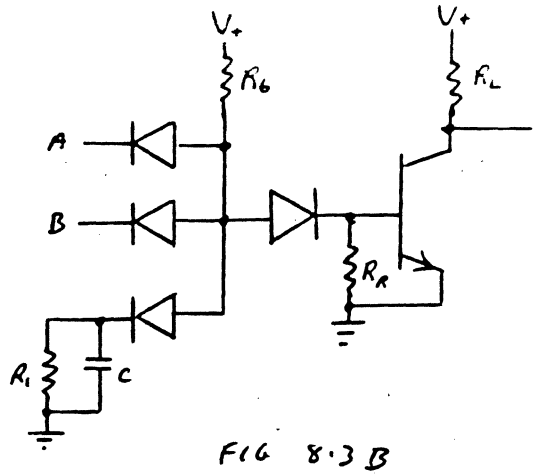


FIG 8.3B

+ EDGE DELAY AND SAFETY SENSE LATCHING CIRCUIT

DTL LOGIC CIRCUIT SHOWING TYPICAL OPERATION

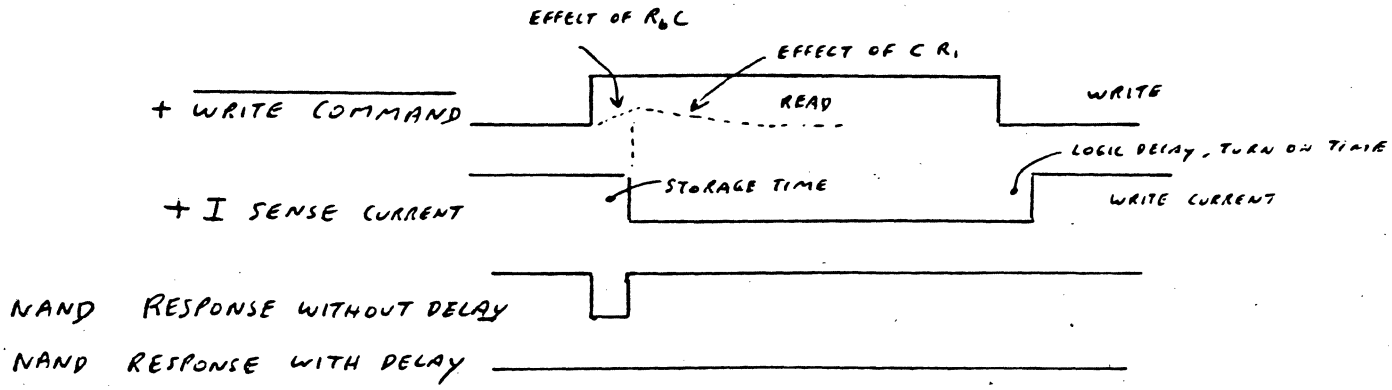


FIG 8.3C

EFFECT OF R_BC DELAY FOLLOWING LAST + EDGE

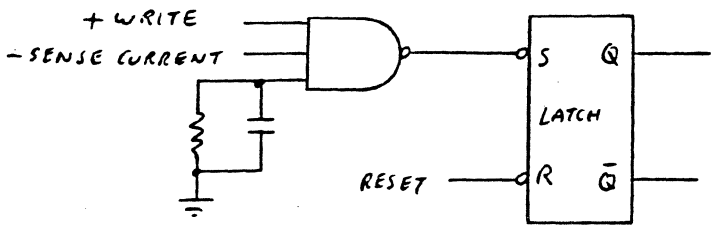


FIG 8.4

WRITE AND NO CURRENT SENSE

$$V_{b_{CW\,MIN}} = 0.75V_{(S)} = \frac{(V_{+\,MIN} - V_{D\,MAX})R_{1\,MIN}}{(S)\left(R_{b\,MAX} + \frac{\frac{R_{1\,MIN}}{C_{MAX}S}}{R_{1\,MIN} + \frac{1}{C_{MAX}S}}\right)} \qquad \text{(EQ 8.1)}$$

$$0.75V_{(S)} = \frac{(V_{+\,MIN} - V_{D\,MAX})(R_{1\,MIN})\left(S + \frac{1}{R_{1\,MIN}C_{MAX}}\right)}{S\left[S + \frac{R_{b\,MAX} + R_{1\,MIN}}{C_{MAX}R_{b\,MAX}R_{1\,MIN}}\right]R_{b\,MAX}C_{MAX}} \qquad \text{(EQ 8.2)}$$

$$0.65V_{(S)} = \frac{(V_{+\,MAX} - V_{D\,MIN})(R_{1\,MAX})\left(S + \frac{1}{R_{1\,MAX}C_{MIN}}\right)}{S\left[S + \frac{R_{b\,MIN} + R_{1\,MAX}}{C_{MIN}R_{b\,MIN}R_{1\,MAX}}\right]R_{b\,MIN}C_{MIN}} \qquad \text{(EQ 8.3)}$$

If we take the inverse Laplace of the last two equations and solve for T, we have the maximum and minimum delays.

$$0.75V_{(S)} = \frac{(V_{+\,MIN} - V_{D\,MAX})R_{1\,MIN}}{R_{b\,MAX}C_{MAX}}\left[\frac{S + \frac{1}{R_{1\,MIN}C_{MIN}}}{S\left(S + \frac{R_{b\,MAX} + R_{1\,MIN}}{C_{MAX}R_{b\,MAX}R_{1\,MIN}}\right)}\right] \qquad \text{(EQ 8.4)}$$

$$\mathcal{L}^{-1}\,0.75V_{(S)} = \frac{(V_{+\,MIN} - V_{D\,MAX})R_{1\,MIN}}{R_{b\,MAX}C_{MAX}}\left\{\left[\left(R_{b\,MAX} + R_{1\,MIN}\right) + \frac{1}{C_{MIN}}\right]e^{-\frac{(R_{b\,MAX}+R_{1\,MIN})t}{C_{MAX}R_{b\,MAX}R_{1\,MIN}}} + \frac{1}{C_{MIN}}\right\}$$

EQ 8.5

$$= \frac{(V_{+\,MIN} - V_{D\,MAX})R_{1\,MIN}}{R_{b\,MAX}C_{MAX}}\left\{e^{-\frac{R_{b\,MAX}+R_{1\,MIN}\,t}{C_{MAX}R_{b\,MAX}R_{1\,MIN}}}\left[1 - \frac{C_{MAX}R_{b\,MAX}}{(R_{b\,MAX}+R_{1\,MIN})C_{MIN}}\right] + \frac{C_{MAX}R_{b\,MAX}}{(R_{b\,MAX}+R_{1\,MIN})C_{MIN}}\right)$$

$$0.65V_{(t)} = \frac{(V_{t\,max} - V_{D\,min})(R_{1\,max})}{R_{6\,max}\,C_{min}}\left[\frac{s + \frac{1}{R_{1\,max}\,C_{min}}}{s\left(s + \frac{R_{6\,min} + R_{1\,max}}{C_{min}\,R_{6\,min}\,R_{1\,max}}\right)}\right] \qquad \text{(EQ 8.6)}$$

$$\mathcal{L}^{-1}\,0.65V_{(t)} = \qquad \text{(EQ 8.7)}$$

With the above circuit, all we need to guarantee is that the maximum time of the write sense current remaining high and the minimum delay do not overlap. If we used this fault signal to shut off faulty write current directly without the latch then the circuit would oscillate. To see why we need only consider that if faulty current is sensed correctly which shuts off the current then the 'sense' will drop after some delay which then permits the faulty current and its following sense to return. All this repeats. Latching the fault indication prevents oscillation.

2. Write Command and No Write Current

This fault condition is sensed using the same sense circuit as in Figures 8.1A or 8.2. The difference is in the following 'and' gate. Figure 8.4 shows the same circuit as Figure 8.3A except that the signals + Write Command and + Sense Current are changed to + Write Command and - Sense Current. Now if current is lost during a 'Write Command' then the fault is sensed. Again a timing problem exists at the leading edge of Write Command because the current

source and the sense circuit need time to respond. The same logic delay circuit will function correctly with the same but appropriately relabled timing diagram (Figure 8.3C).

3. Write Command and No Write Data

When writing we need to be assured that data is being written on the media. What signals are there available to us that truly indicate that we are writing? In tape machines this is an easy task as it is usual practice where there are Read Heads following the Write Heads. In Disc machines where the heads are used for both Read and Write a full revolution of the disc is required before Read Verify is possible. If we assume that the head is in proximity to the disc then we can sense the flux changes to verify that we are writing. Rather than a separate winding around the head core to sense the flux changes, we can sense the voltage transient across the head instead. There are several simple circuits that can do this with any type of head and write driver. They do increase the circuit capacitance which increases the current rise time as this needs to be considered. The design must consider the attenuation R1 and R2, the voltage transient, and the voltage gain of A. Amplifier A should be a limiting amplifier compatible with the T^2L logic of the Resetable Single Shot. If the circuit of Figure 8.5A were implemented the output would respond to only the positive transient; therefore, we would have one output for every other current transition. This is acceptable if the Resetable Single Shot time is greater in the worse case than twice the maximum current transition spacing. If both current transitions responses are required then the amplifier should be a linear differential amplifier with the outputs connected to a full wave rectifier thus achieving a unidirectional pulse for each current transition. Figure 8.6 is such a circuit.

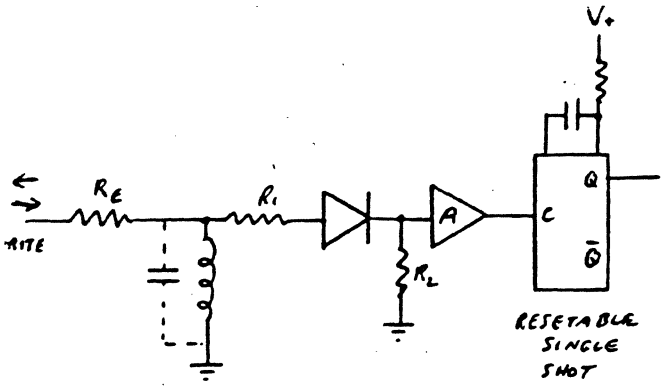


FIG 8.5 A

HALF WAVE TRANSITIONS SENSE
FOR COMPLIMENTARY WRITE DRIVER.

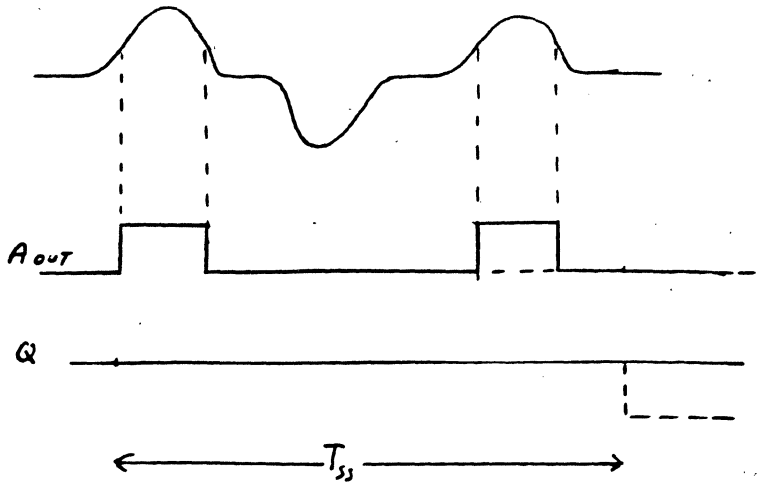


FIG 8.5 B

TIMING DIAGRAM.

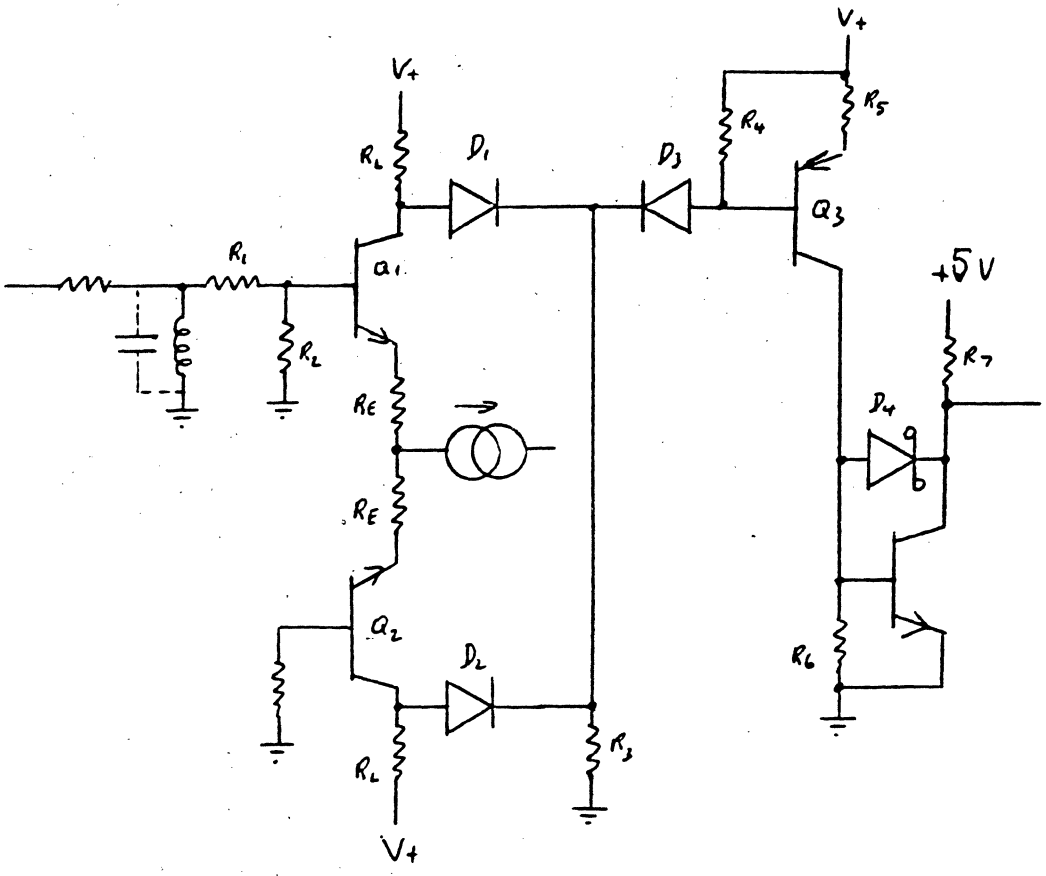


FIG 8.6

COMPLIMENTARY FULL WAVE TRANSITION SENSE

SAFETY CIRCUITS

Although the complimentary write driver is not often used, we will go thru the design.

If the transient voltage expected worse case minimum at point B were 7.V BP, then we could isolate the head from the differential amplifier with an attenuation of 3.5_{max} making, (this reduces the extra capacitance across the head proper with only a slight change to the damping if we keep the impedance high)

$$\frac{R_{2\,min}}{R_{2\,min} + R_{1\,MAX}} = 3.5 \qquad \text{(EQ 8.8)}$$

Giving us a 2V min transient applied to the base.
The $gain_{min}$ of the amplifier could be set for 1 with the actual value of RE_{max} and the I current $source_{min}$ equal to less than the expected $2V_{min}$ transient. This way the differential amplifier will limit on the minimum expected signal transient. The two equations are shown in EQ 8.9 and EQ 8.10.

$$A_{MIN} = 1 = \frac{R_{L\,MIN}}{R_{E\,MAX} + r_e} \qquad \text{(EQ 8.9)}$$

$$R_{E\,MAX} = \frac{2\,V_{SIG\,MIN} - \left(V_{be\,ON\,MIN} - V_{be\,LEAK\,MAX}\right)}{I_{SOURCE\,MIN}} \qquad \text{(EQ 8.10)}$$

SAFETY CIRCUITS

Depending on the speed of operation we try keeping I_{source} to around 2 to 5 ma, and use a transistor with a reasonable F_t. This is dictated by the Gain Bandwidth curves for the transistor. The next part of the circuit considers the value of the quiescent voltage at the base of Q3. This voltage should be such that during no signal the base must be conducting worse case. Assume Vbe = 0.75V as the conducting maximum Vbe required, then

$$\left(I_{source\ min}\right)\left(R_{L\ min}\right) - V_{D\ max\ (load)} \geq 0.75\,V \qquad \text{(EQ 8.11)}$$

When operating, the most positive collector voltage will be when Q, or Q2 are cut off. We designed the amplifier to do this deliberately to ensure limiting in the positive direction. When this occurs the voltage at the cathode of Diode 1 or 2 is about 0.6V below +V. If the value of R3 is kept high so that --

$$\frac{\left(V_{+\,MAX} - V_{D\,1,2,3\,min}\right) R_{3\,max}}{\left(\frac{R_{L\,min}\ R_{4\,min}}{R_{L\,min} + R_{4\,min}}\right)} \geq V_{+\,MAX} - 0.4\,V \qquad \text{(EQ 8.12)}$$

--then transistor Q3 is cut off during the peak of the transient.

The following two stages are designed considering the current of Q3, the current loss thru R6, and keeping Q4 out of saturation. The gain is made high only to accommodate the current loss in R6 until the voltage rises to turn on the base of Q4. If we make R6 2KΩ then our maximum loss of current will be

$$I_{Loss\ MAX} = \frac{V_{be\ on\ max}}{2 \cdot R_{-\,min}} \qquad \text{(EQ 8.13)}$$

$\pm 5\%\ R_{min} = 1.9K$

The value of R5 is chosen as we consider the current required to turn on Q4 and the minimum signal on the base of Q3. Because the diodes D1, or D2 and D3 are always conducting then we should see almost the full limited signal at the base of Q3. Assume we only receive half or 1.0V, then the collector current of Q3 is approximately

$$\frac{1.0V - V_{be\,max_{Q3}}}{R_{5\,max}} \geq I_{loss} + \frac{+5V_{max} - V_{ce\,max}}{R_{7\,min}\ \beta_{4\,min}} \qquad \text{(EQ 8.14)}$$

Once this is accomplished then we know that worse case we can turn on Q4 until the transistor Q3 is cut off when Q4 will then cut off. R6 should conduct any charge on the base of Q4 away before the next transient pulse. We keep Q4 out of saturation (only needed for speed) by using a Schottky diode between base and collector circuit. This circuit works well and is called a Schottky clamp.

The base voltage vs. base current curve has some positive slope, so does the Vsat as a function of Ic. As the base voltage is usually much greater than Vsat then the about 0.4 volt drop of the Schottky diode (which also has a positive slope with increasing current) is sufficient to keep the transistor out of saturation. For example, if Vbe is 0.7V at some value of base current and Vce sat were 0.2 volts for some value of collector current, then 0.7V - 0.4V = 0.3 Volts or about 0.1 volt difference which we can use to reduce the base voltage to around 0.6V or out of saturation. This is a very rudimentary explanation as the solution can only be obtained graphically since all 3 junctions voltages are changing to obtain final balance. EQ 8.14 is correct.

$$V_b = V_{D_s} + V_{ce} \qquad \text{(EQ 8.15)}$$

The above circuit will then output a positive pulse for each transistor (plus or minus) in the worse case since we chose all our design criteria to ensure an output pulse for a minimum transient and we designed the amplifier to limit above this value in order to reduce the current of Q3 for larger transient voltages.

We will next consider the most used Write Driver and generate several circuits that will function as transition detectors for them. One of the problems we should consider is that of the polarity of the sensed transient. If we sense the negative transient at the collectors of the Write Driver Q1 and Q2, then if the head were open on one side, we would still get a negative transient as the current flow will be thru the damping resistor. The affected transistor will saturate but the negative transient will still be present even though of a different amplitude. Remember that we minimized the Collector-Base voltage in order to reduce power dissipation; therefore, there will not be much difference in voltage between the saturated case and the normal case. A better way is to use the transformer action of the head to get the positive transient occurring at the 'off' transistor (for NPN Write Drivers). This then assures us that the whole head is working for if one side were open transformer action could not occur.

The circuit of Figure 8.7 shows a typical connection. The attenuation of R1 and R2 is provided to help keep Q3 out of saturation yet guarantee it does conduct on the positive transient. Again the Schottky diode is provided to make a Schottky Clamp. The circuit includes the Matrix diodes just to show that

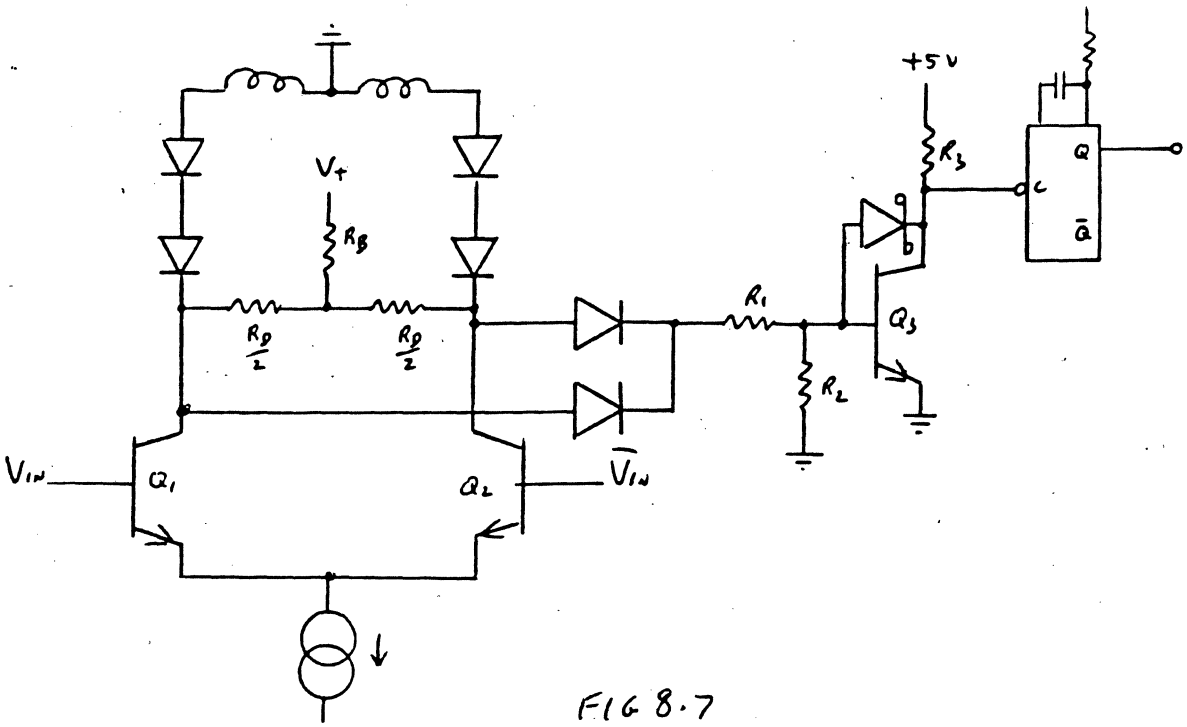


FIG 8.7

CENTER TAPPED, CURRENT SOURCE DRIVEN WRITE TRANSITION SENSE CIRCUIT (NPN)

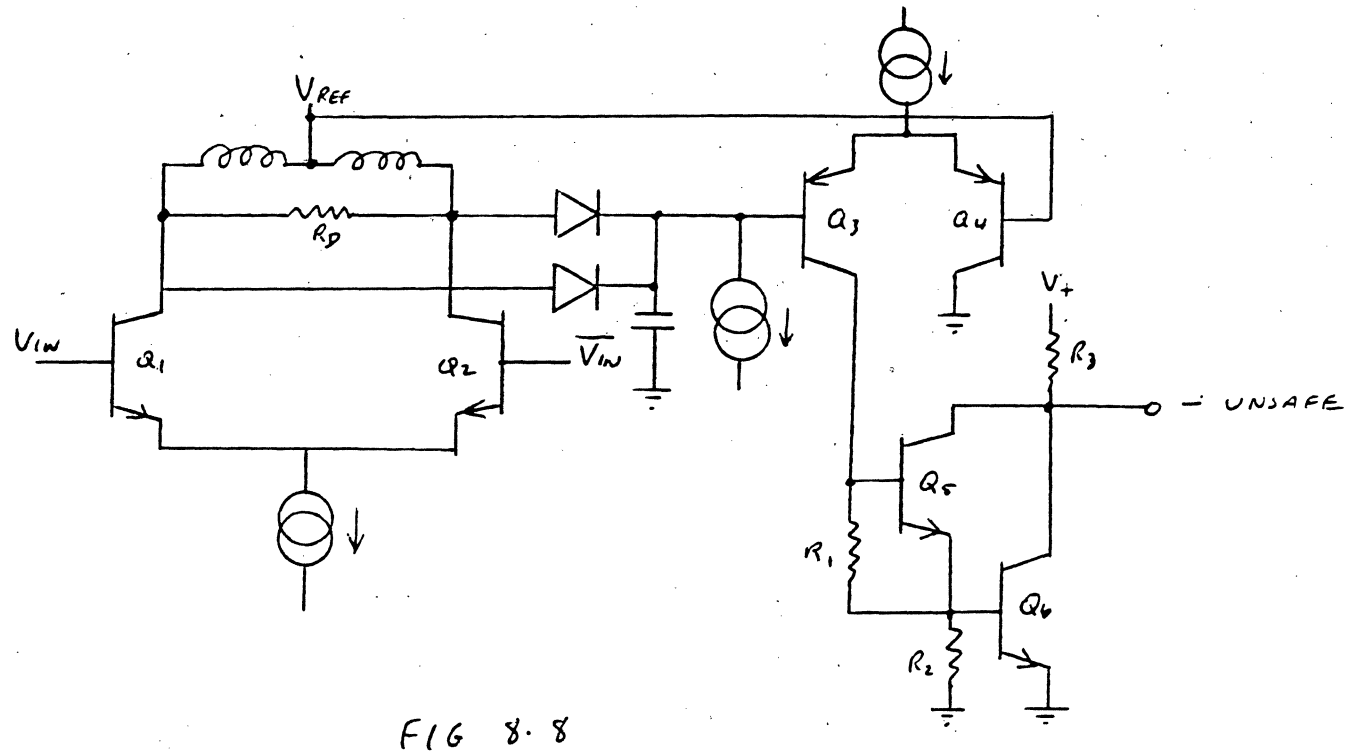


FIG 8.8

COMBINED TRANSITION DETECTOR AND ΔTIME INTERGRATOR

it makes no basic difference except in bias levels. The diodes that pass the head current correctly to the Write Drivers also pass the positive transient therefore the transient will be available at the Write Driver collectors.

If the head transient were 7.0V.BP SE then the attenuation max becomes

$$V_{be\,max} \leq R_{2\,min}\left[\frac{\left(V_{TRANSIENT\ BP\text{-}SE\ min} - 3\,V_{D\,max}\right)}{R_{1\,max} + R_{2\,min}} - \frac{\left(5.0V_{max} - V_{C_3\,min}\right)}{R_{3\,min}\,\beta_{3\,min}}\right] \quad \text{(EQ 8.16)}$$

If the head were shorted then there would be no V TRANS.

If we design using this equation, then we ensure operation as long as the R.C. loading Q_3's base is small compared to the discharge times required between transients. Note also that because of the loading effect of R_1 and R_2 in parallel with Hib of Q_3 the damping Resistor R_D total needs to be raised in value accordingly. This particular configuration lends itself to driving ground referenced logic and is by far the better way to go than circuits that put in emitter degeneration in the emitter of Q_3. In these cases level translation is required such as we did in Figure 8.6.

Another circuit provides a capacitor charging and discharging current such that the end result is a combination including the effects of the Resettable Single Shot. This circuit has been used in head interfacing integrated circuits designed first by IBM. The capacitor is charged thru the diodes on each transition.. This reverse biases Q_3's base. A negative current source steadily discharges the capacitor such that if the time between transitions exceeds the time for discharge to the Reference level Q_3 turns on activating the Darlington Q_5 and Q_6. Thus as long as transitions occur regularly within the time allotted then the output remains high. If they cease or the spacing exceeds the discharge

time then the output goes low. The circuit can be used for a current indication by using the output at the top of R_3 instead of the collector as is usual. This permits other functions to be performed in parallel. We will not go into this design in detail except to note several considerations. First, the diode charging current does affect the damping depending on the state of charge of the capacitor. It also is affected time wise by the negative current source variations and the value of C as well as the β of the PNP transistors as they start to conduct or any leakage if cut off. When this circuit is integrated, these variations can be extensive.

We might consider a circuit that could be used with the Bridge driver. This two terminal head driver's head transient negative voltage is not much different between an open head case and the normal head case. With the upper emitter followers generally controlling the voltage the only Peak difference is the margin provided between the normal transient voltage and the negative swing of the supposed cut off emitter follower. This is not sufficient for an indication. However the positive level of the emitter follower is significant in that it can be measured against a reference. It is the difference between the Vbe lightly conducting and the Vbe heavily conducting. This is usually a few tenths of a volt against a solid reference. A circuit can be devised that can utilize this difference.

If the head were open then the normal path for current is blocked; therefore, the lower half of the bridge will draw current directly from the upper bridge immediately above it, such as from Q_2 thru Q_4 in Figure 8.9.

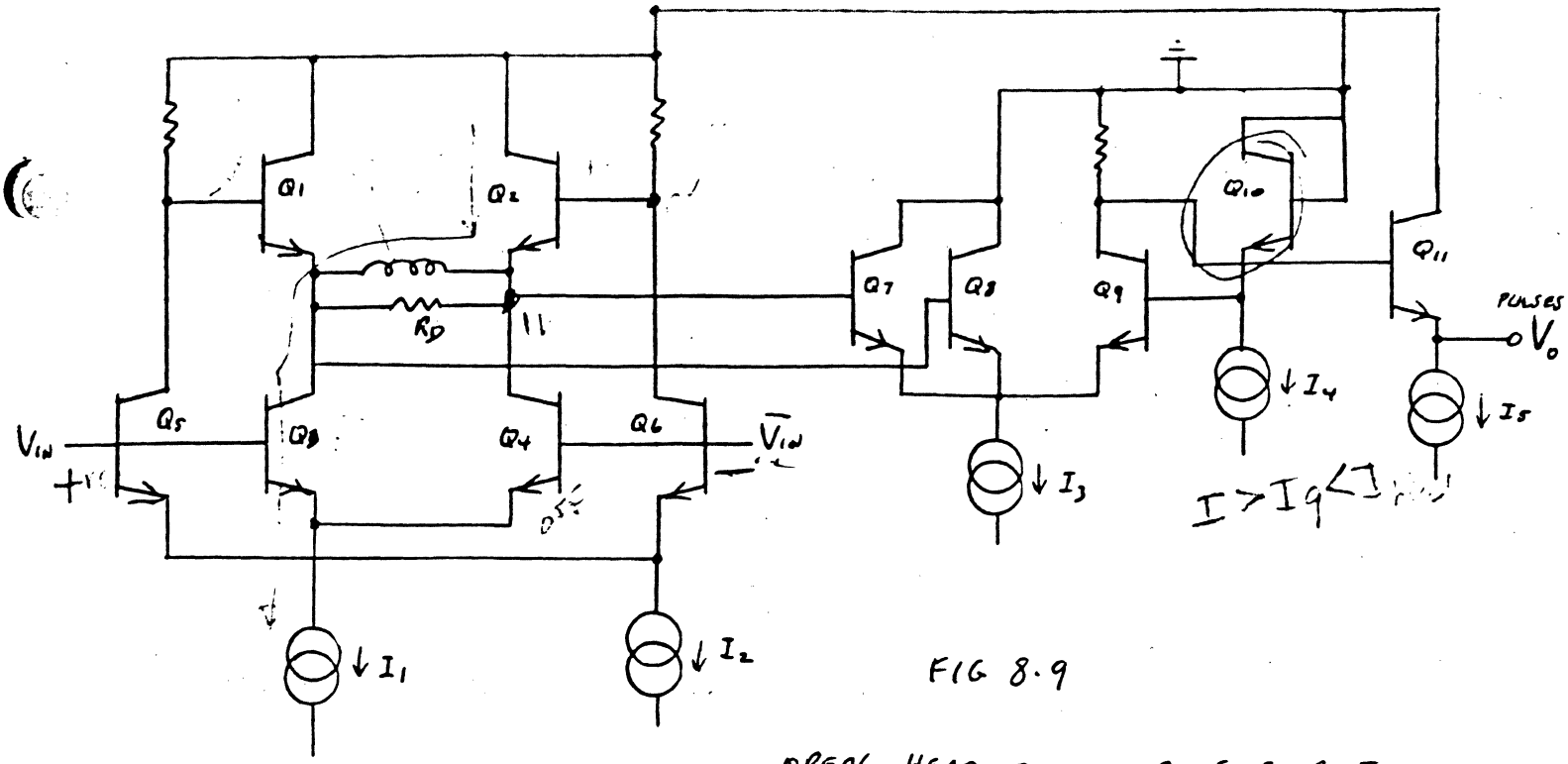


FIG 8.9

OPEN HEAD DETECTOR FOR A TWO TERMINAL HEAD USING A BRIDGE DRIVER

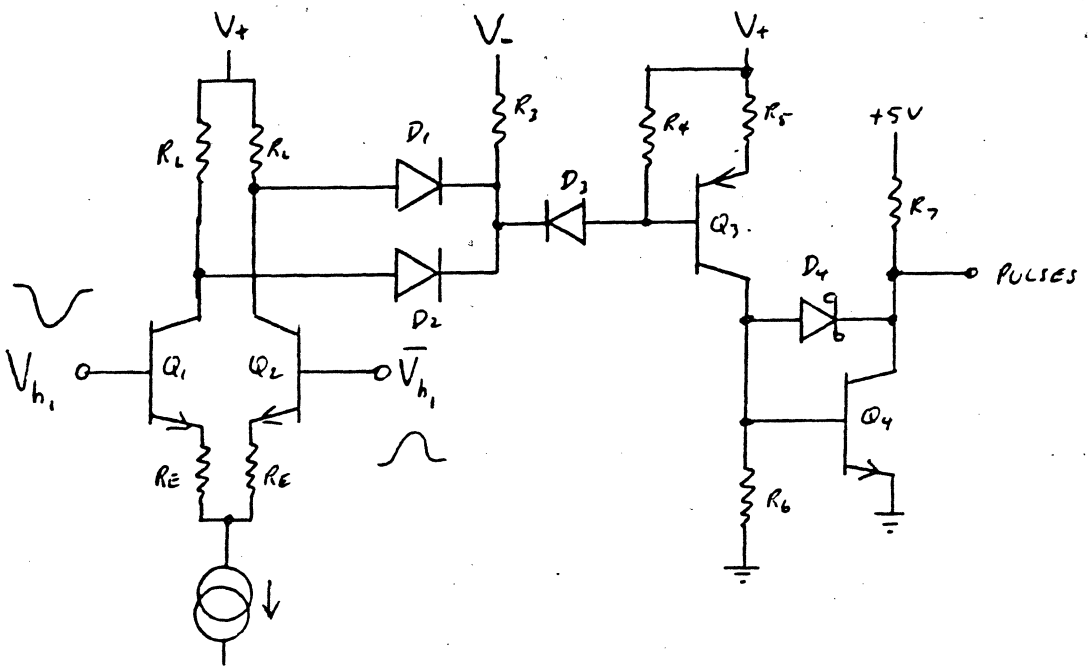


FIG 8.10

INTERMEDIATE AMPLIFIER TRANSITION DETECTOR

The transistor Q10 is provided a slight current of only a few hundred microamperes, I_4, such that the base voltage of Q9 is at a voltage of say 0.6 Volts. If the Write Current I_1 were say 50 ma and the head were open with Vin high then Q3 is conducting from Q1 even though the base of Q1 is low. The base of Q4 is low so Q4 is cut off, but the base of Q2 is high since Q6 is cut off. Since the current thru Q2 is very low, being only leakage, then the Vbe of Q2 is around 0.4 Volts. This makes the difference between the bases of Q7 and Q9 about 0.2 Volts or sufficient to switch. Using low currents for I_3 to reduce base current effects, we have a circuit that will respond to open heads but it will not respond to normal and shorted heads. The voltage on the emitters of Q1 and Q2 will not change in the case of a shorted head except due to rise - fall time crossing effects which are very narrow. A circuit that will totally detect true transitions at the Write Driver requires two detectors, one for open heads and one for normal heads since the shorted head case produces only a small transient voltage. The second circuit simply needs to verify that the head terminal voltages appropriately follow the upper half of the Bridge's base voltages. This can be done using a low gain differential amplifier and full wave rectifier similar to Figure 8.6 and shown in Figure 8.10 for inductive heads. The amplifier isolates the head to reduce the extra capacitance. The output pulse width can be pulse width discriminated in order to isolate the shorted head narrow pulse, resulting from poor rise times, from the normal head wide pulse. This is only necessary if the rise and fall times of the pre driver are not very fast such that there is a glitch at the head terminals at the intended transition edge. Figure 8.11 illustrates the phenomenon.

voltage change in the shorted head case is about half that of a
l case as long the swing on the upper half of the bridge bases is
· exceeds slightly the regular transient. Getting back to Figure 8.10,
signals are really not differential but are negative pulses referenced
vel for each oppositely occurring transition which alternates between
e can use this to generate our full wave rectifed pulses if we limit
a level between the half level expected from shorted heads and the
rmal level. If the base swing were 5.0V, then we need a cut off
round (.75)(5.V) = 3.75V. Choosing the +V supply sufficiently high

$$V_{t\,min} - (I_{1\,max})(R_{L\,max}) \geq V_{HD\ HIGH\ max} \qquad \text{(EQ 8.17)}$$

roceed as before.

$$_{max} = \frac{(.75)(V_{TRANSIENT}) - (V_{b\,e\ ON\ min} - V_{b\,e\ LEAK\ max})}{I_{SOURCE_1\ min}} \qquad \text{(EQ 8.18)}$$

inder of the design follows from EQ 8.11 thru 8.14.

re is a variation of the above circuit if the head is essentially
e, such as the case of a thin film head. Here the voltage waveform
ead is almost a squarewave. Again, if power dissipation is minimized
esign, the down level applied to the upper bases of the bridge is
below that of the IR drop in the resistive head. One circuit that
riminate the normal head from the shorted head or partially shorted

head is to diode couple the low level signal on each head terminal and compare that to a reference. However this circuit cannot distinguish the case of no data applied meaning an open data cable or circuit somewhere earlier in the circuit chain. We must really use the dV/dt of each transition. This can be done using a frequency level detector such as is shown in Figure 8.12 or a + pulse rectifier and amplifier as shown in Figure 8.13.

The circuit of Figure 8.12 should be driven from an intermediate amplifier to isolate the circuit capacitance from the head circuit. The level at Point B is a function of the voltage change at the inputs, the ratio of the two capacitors, and the time constant of C2 and the two resistors. For large capacitor ratios the level at B is fairly smooth, but if we choose a smaller ratio with a short RC time constant, then the circuit will perform as a full wave rectifier with controlled output pulse widths.

The circuit of Figure 8.13 may not require an intermediate amplifier if the coupling capacitor's value is small, and as the RC time constant should be large compared to the transition rate, it requires large resistors which in turn affect the base bias due to base currents. The circuit is basically a full wave biased rectifier. The difference from the base line to the clipping level is determined by the resistor network R1, R2 and R3. The upper portion of the transient is 'dot ored' and is available as a series of squared positive pulses at the output.

A better circuit would result if an intermediate amplifier is used to permit lowered resistances. Lastly, if we consider the transition detectors

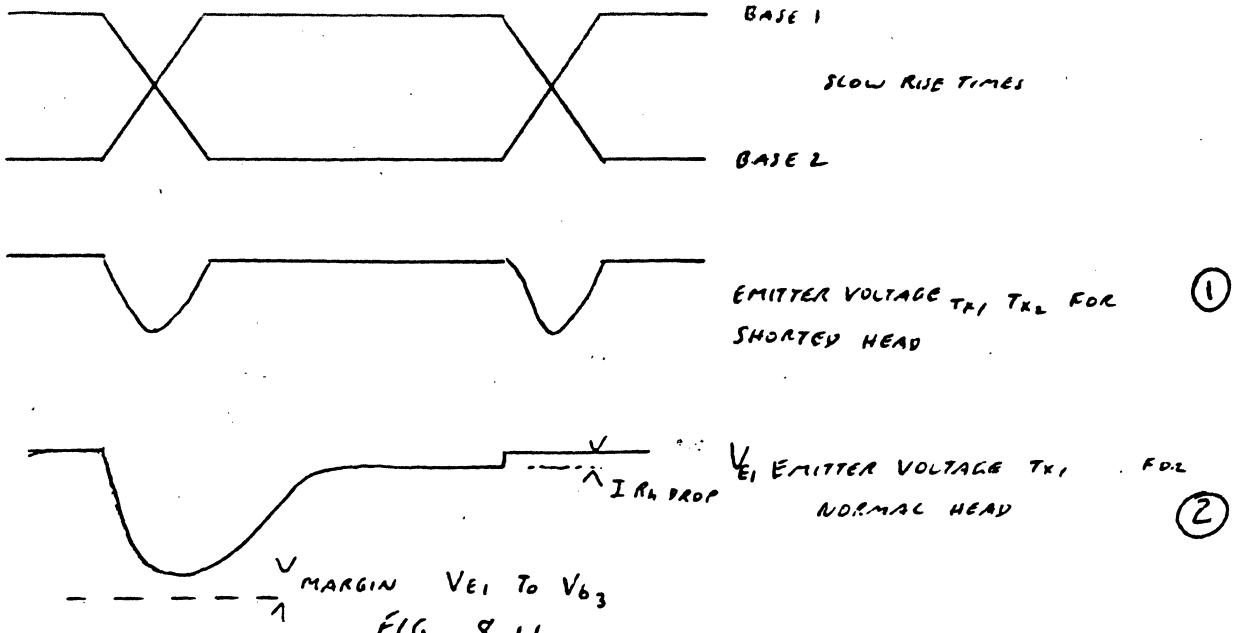


FIG 8.11

UPPER BRIDGE VOLTAGE WAVE FORMS
1 SHORTED HEAD, 2 NORMAL HEAD.

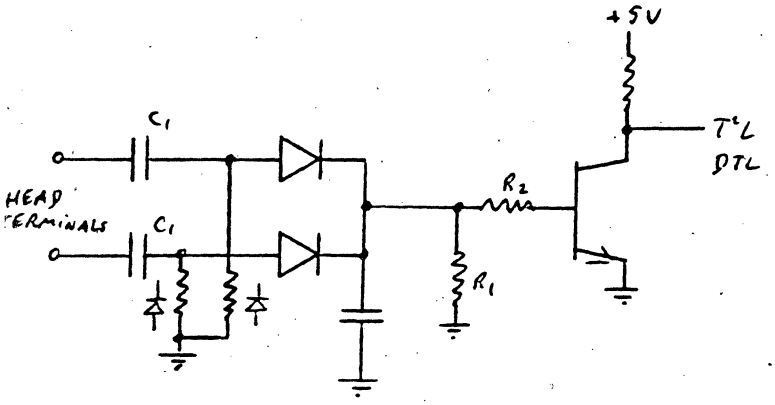


FIG 8.12

T^2L TRANSITION DETECTOR
CAPACITIVE COUPLED
RESISTIVE LOAD OR DIODE CLAMP.
BOTH SHOWN.

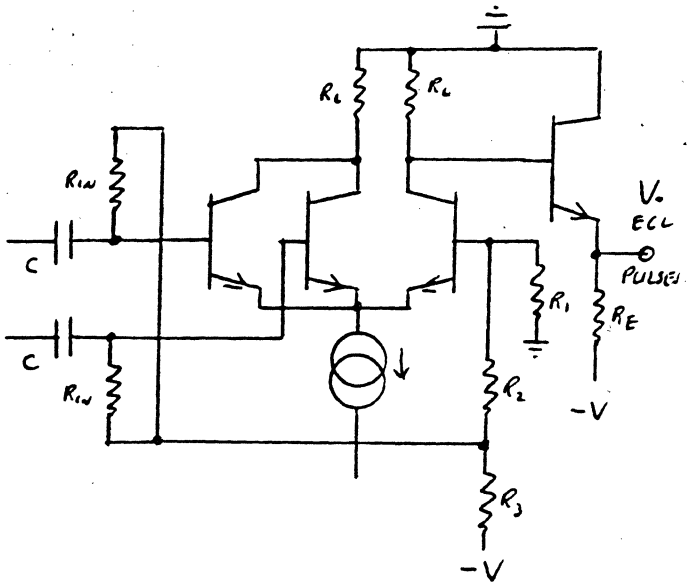


FIG 8.13

CAPACITIVE COUPLED TRANSITION DETECTOR FOR ECL. 'POSITIVE OR'

they really indicate the correct functioning of the entire Write chain. Without write current there would be no transitions neither without a continuous data path and a functioning head. The only thing it does not tell us is if the head is in contact with or in proximity to, the media.

4. Write Current thru a Protected Data Field

There are several variations of this circuit family. The first involves the Memorex invention of the Write Protect feature. A simple circuit operated from a switch blocks the Write Command from the Write Current circuits. The circuit usually indicates back to the control function that the Data is protected. One consideration is to design the logic circuits such that operation of the switch during a Write operation will not disturb the write in progress until the operation is over. A simple gated latch will accomplish this.

A second variation is used in embedded servo type disc drives. These disc files have servo information pre written on the data discs in sectors or interleaved with the data. The same information must be protected in order to maintain correct servo operation. Counting circuits that are indexed to the disc position are usually used. Decoding the count determines the areas where the prewritten servo information is recorded. Circuits are also used that verify correct operation of the counters, such as, frequency sensitive discriminators and phase locked loops. Protection of the servo data is of such inportance that redundant counters are sometimes used with phase detectors to monitor their differences.

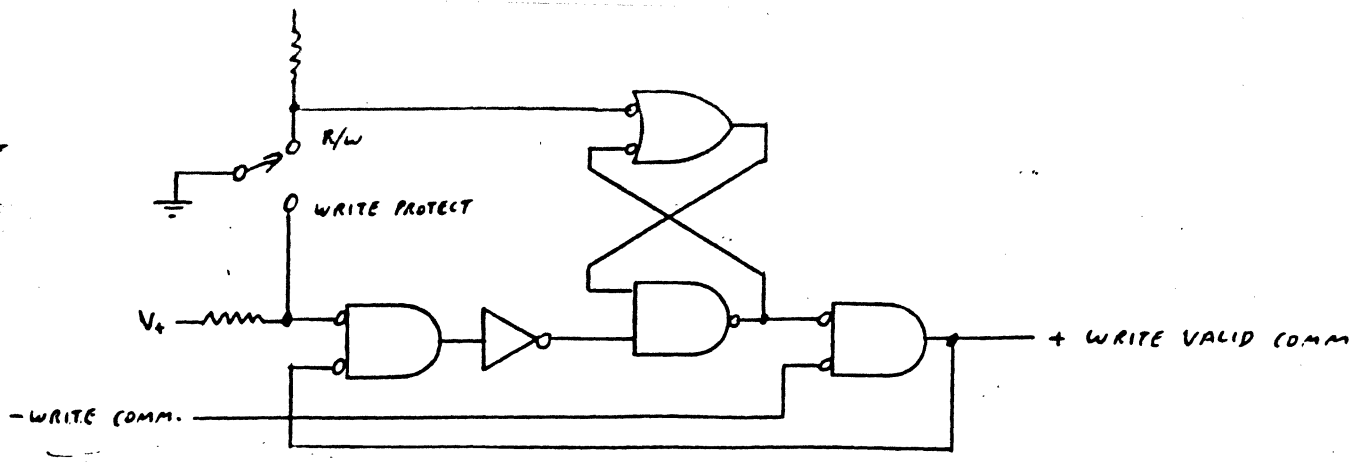


FIG 8.14 A

WRITE PROTECT FEATURE

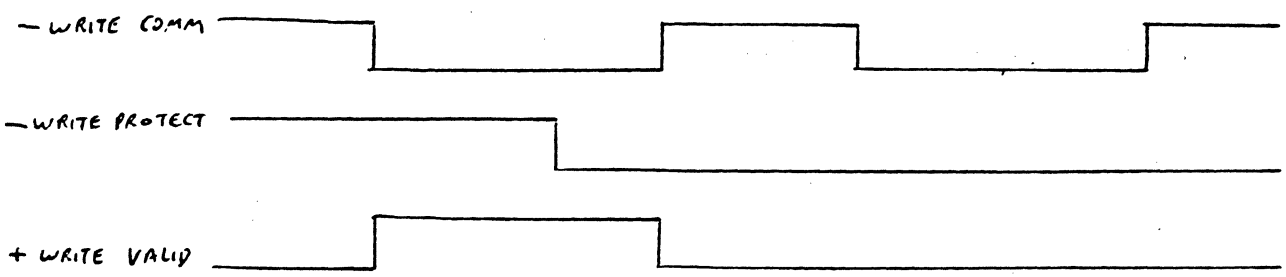


FIG 8.14 B

WRITE PROTECT WAVEFORMS

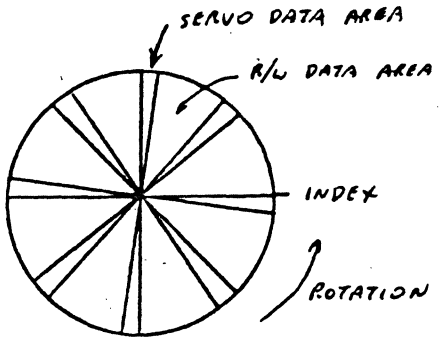


FIG 8.15

EMBEDDED SERVO, 8 SECTORS,

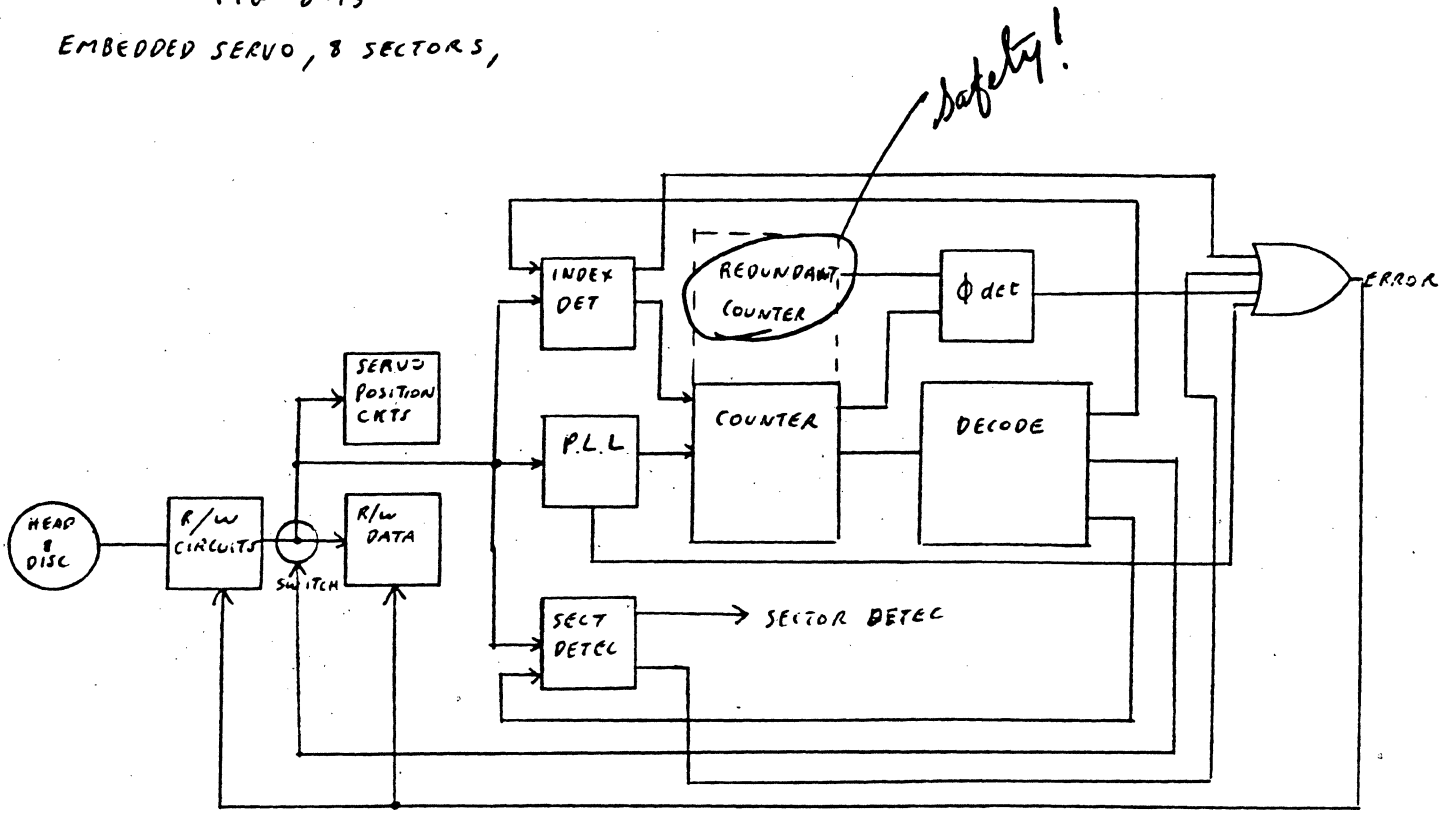


FIG 8.16

SAFETY PHYLOSOPHY IN EMBEDDED SERVO SYSTEMS

5. Write Current and Not Directly over the Assigned Track

Again there are two checks performed to verify the head position in Disc Drives before a Write is allowed. The first requires the successful comparison of the desired address contained in a register and the written address recorded on the disc usually before each record. The second requires circuits for monitoring the heads position with respect to some reference. In track following servo systems there are signals available that are sensitive to the percentage variation from the tracks centerline. As the head deviates from the centerline by an amount exceeding the off track capability of a Read - Write - Read sequence at the track extremes, we require a signal that will terminate a Write function immediately. This is because any data thereafter written will be difficult to read due to the off track, adjacent track and fringing crosstalk or interference. As the head mechanism cannot move instantaneously, some earlier or narrower range is sensed with a time limitation imposed for the system to restore the head to be within these normal limits. Such circuit timing must consider the mass and forces of the moving mechanism. Figure 8.17 illustrates the phenomenon by showing the head centerline movement compared to the track centerline as a function of time. The two limits are shown dotted with the appropriate sense levels and timing.

This figure shows the head returning within limits within the allotted time. If it did not then a signal would be generated that shuts off any Write in progress and notifies the control circuits.

6. Write Command and Read Command Simultaneously

This circuit is strictly a monitor of the control circuits, but it also checks for logic failures in that if a logic gate failed the opposing

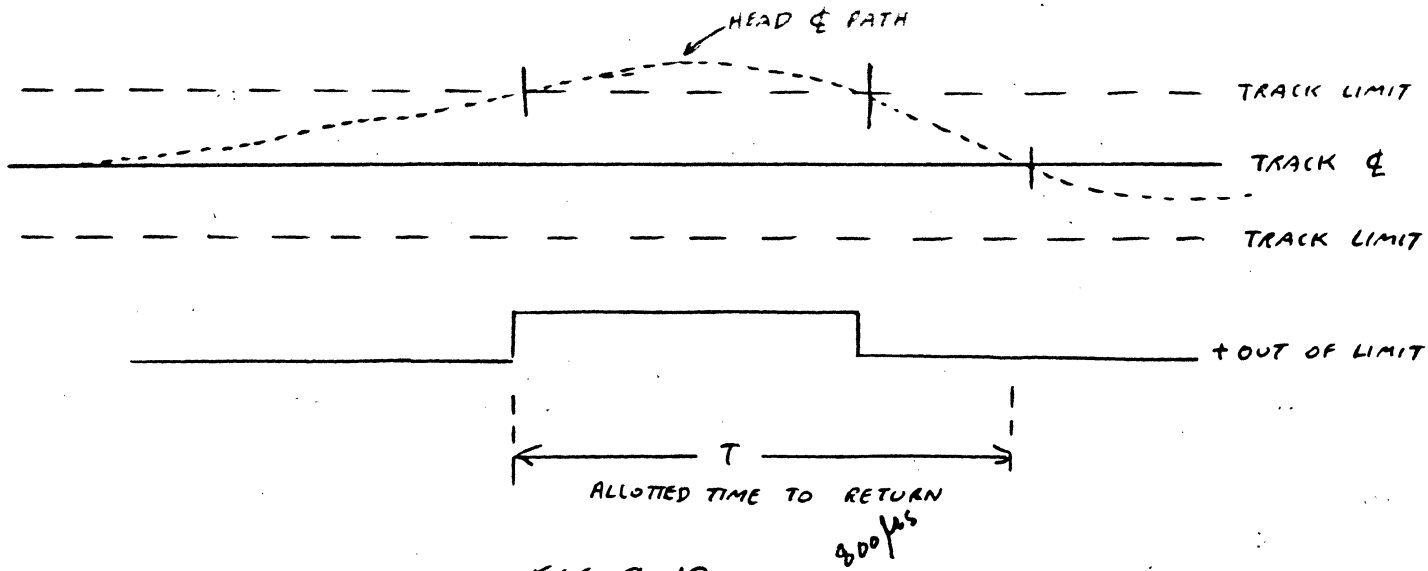


FIG 8-17

OFF TRACK SAFETY LIMITS AND TIME ALLOTTED

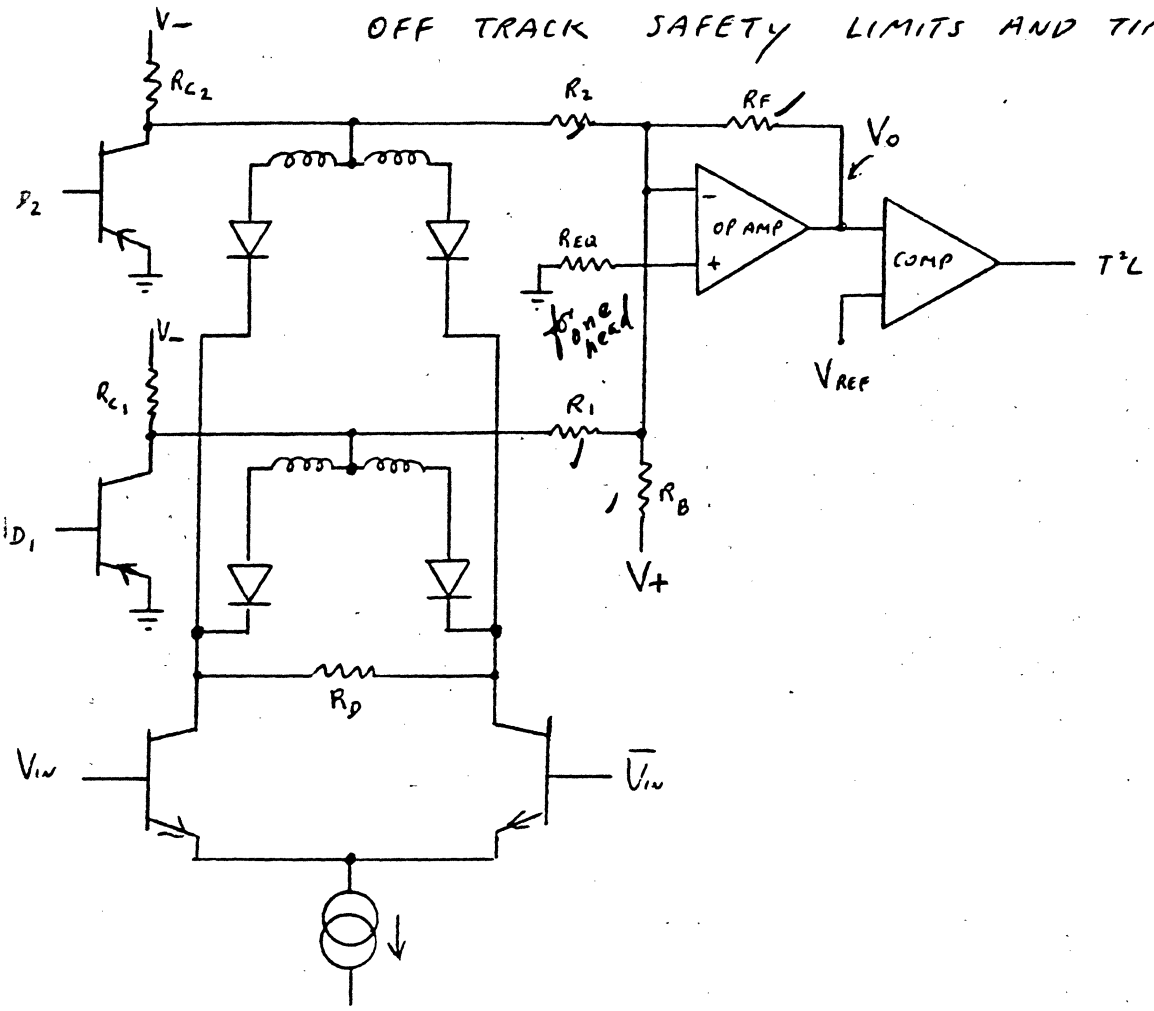


FIG 8-18

CENTERTAP SENSING CIRCUITS

Commands could be issued. A simple 'And' gate at the last logic position of the two commands will suffice. This way the error can be caught up to and including the input to the Analog Read and Write Circuits.

7. More than One Head Selected

Depending on the Matrix configuration, there could be several circuits required. If the number of heads built into a machine is great enough, some designers choose to use two separate Write Circuits, two separate Read Pre Amplifiers and two separate Head Centertap drivers. When this occurs there must be circuits that monitor circuits to verify that one and only one is operating at any one time in serial machines or in parallel machines to see that all such circuits are operating simultaneously.

First let us pursue the Centertap Monitoring Circuits. The engineer has a choice of an 'Exclusive Or' tree or an operational amplifier. The latter is least expensive even in the earlier machines when such a circuit had to be built of discreet components.

Consider Figure 8.18. We have shown two heads with the centertaps capable of being grounded, with NPN Write Drivers and PNP centertap drivers. Reverse bias for the centertap is a negative voltage. The ratio of the resistors R2 to R_{c2} and R1 to R_{c1} is large such that the correct amount of reverse bias is maintained to block the head transients. R_F is chosen to provide a certain worse case guaranteed voltage output from the operational amplifier. R_B is provided to cancel the affect of all but one of the reverse biased centertaps. For example, if Figure 8.18 showed 5 heads with 5 centertap drivers then R_B would compensate for 3 head centertaps leaving one to provide a negative signal and, of course, the other being at ground if selected. The circuit will respond

with a negative output if more than one head centertap is grounded and with a positive output if only one or none are grounded. The design is worse cased as follows where N is the total number of head centertap drivers.

$$V_{o\ max_{NORMAL}} = -R_{F\ max}\left[\frac{-V_{C-\ max}\ (N-1)}{R_{C2\ min} + R_{1\ min}} + \frac{V_{o\ min}}{R_{B\ max}} - \frac{V_{SAT\ max}}{R_{1\ min}}\right] \qquad \text{(EQ 8.18)}$$

$$V_{o\ min_{NORMAL}} = -R_{F\ min}\left[\frac{-V_{C-\ min}\ (N-1)}{R_{C2\ max} + R_{1\ max}} + \frac{V_{+\ max.}}{R_{B\ min}} - \frac{V_{SAT\ min}}{R_{1\ max}}\right] \qquad \text{(EQ 8.19)}$$

$$V_{o\ max_{NONE}} = -R_{F\ max}\left[\frac{-(V_{C-\ max})\ N}{R_{C2\ min} + R_{1\ min}} + \frac{V_{T\ min}}{R_{B\ max}}\right] \qquad \text{(EQ 8.20)}$$

$$V_{o\ min_{NONE}} = -R_{F\ min}\left[\frac{-(V_{C-\ min})\ N}{R_{C2\ max} + R_{1\ max}} + \frac{V_{+\ max}}{R_{B\ min}}\right. \qquad \text{(EQ 8.21)}$$

$$V_{o\ max_{2\ ON}} = -R_{F\ max}\left[\frac{-V_{C-\ max}\ (N-2)}{R_{C2\ min} + R_{1\ min}} + \frac{V_{+\ min}}{R_{B\ max}} - \frac{2\ V_{SAT\ max}}{R_{1\ min}}\right] \qquad \text{(EQ 8.22)}$$

$$V_{o\ min_{2\ ON}} = -R_{F\ min}\left[\frac{-V_{C-\ min}\ (N-2)}{R_{C2\ max} + R_{1\ max}} + \frac{V_{+\ max}}{R_{B\ min}} - \frac{2\ V_{SAT\ min}}{R_{1\ max}}\right] \qquad \text{(EQ 8.23)}$$

Equations 8.18, 8.19, 8.22, and 8.23 require the worse case bias set into the comparator to be centered between the values of Vo min normal and Vo min $_{2\ ON}$. By properly specifying the value of R_B this range could be centered around the Vbe of 2 transistors and the circuit of Figure 18.9 used instead of the comparator as long as the Vo difference is greater than one volt. In the above equations, we neglected the voltage and current offsets and gain effects of the operational amplifiers. These should be included if the difference is less than one volt which would jeopardize the correct biasing of the transistor. Resistor R_{EQ} should be chosen to equal the resistance seen on the negative input in order to correct for balanced base or input currents. For the other two offsets the total resistance can be kept as low as possible within the current limitations of the centertap and bias resistor currents and the voltage attenuation due to the action of R_{C2} and R2 as it relates to the head transient voltage. Usually the voltage change due to one centertap circuit changing state is large so the main offsets are swamp ed.

The other monitoring circuits become just 'And' and 'OR' combinations of the previous circuits outputs in multiples. The block diagrams of Figure 8.20 thru 8.22 show typical examples of some multiple arrangements. Notice that in Figure 8.20 we need both the indication for write current from either current source as well as the indication of more than one source on at a time for serial data machines. For parallel data machines the dotted addition is required to indicate a failure. There are variations of Figure 8.21. As it is shown the outputs of the Resettable Single Shots are combined to indicate the function. This then would be 'Anded' with the delay gate of Figure 8.3A to capture the failure. However, if one of the Resettable Single Shots failed true then we would never sense any failure of transitions. The 'Exclusive OR' gate will

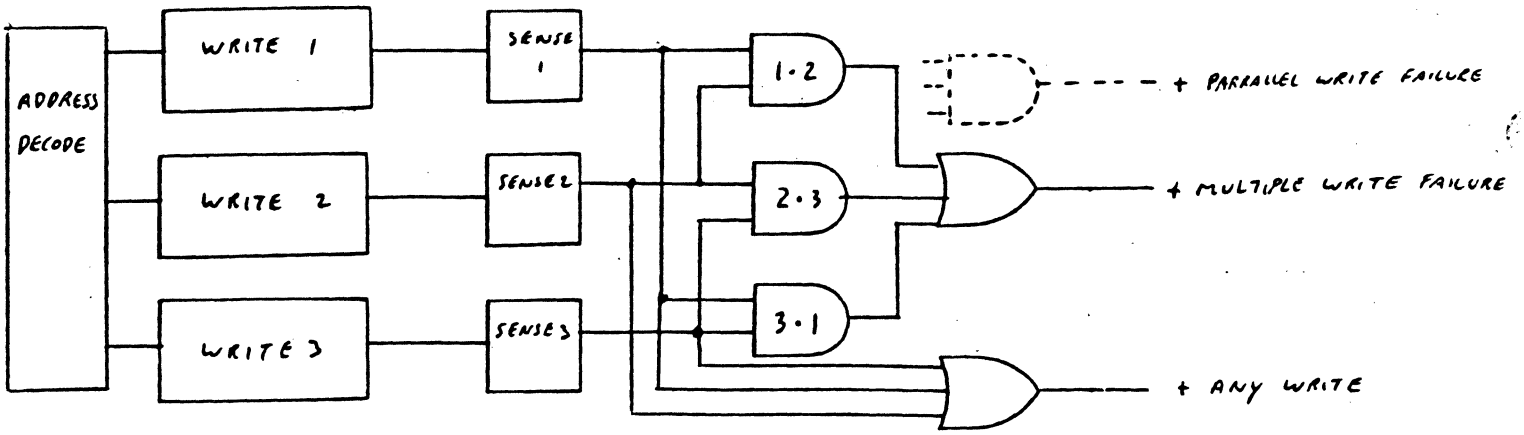


FIG 8·20

SENSING MULTIPLE SOURCES (3)

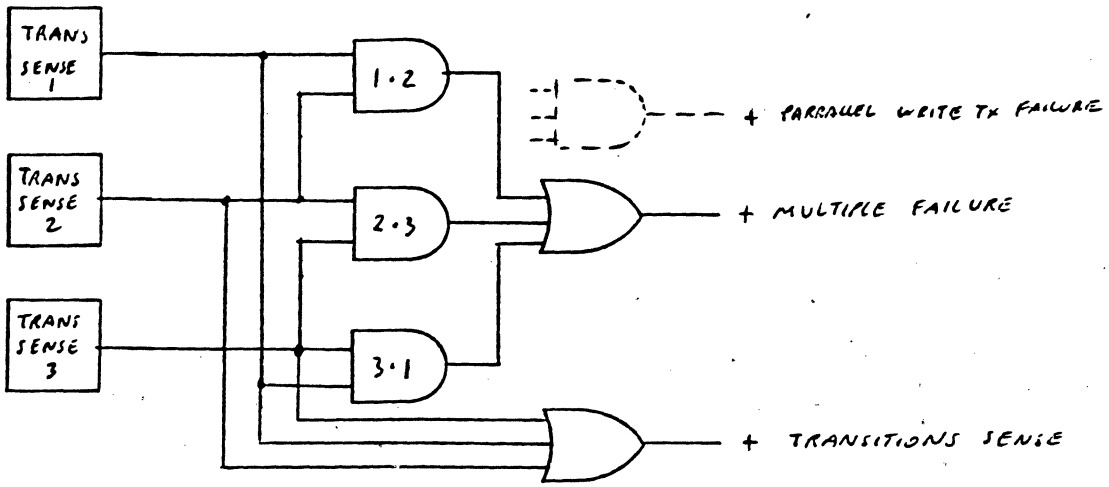


FIG 8·21

SENSING TRANSITION FAILURE AND R.S.S FAILURE

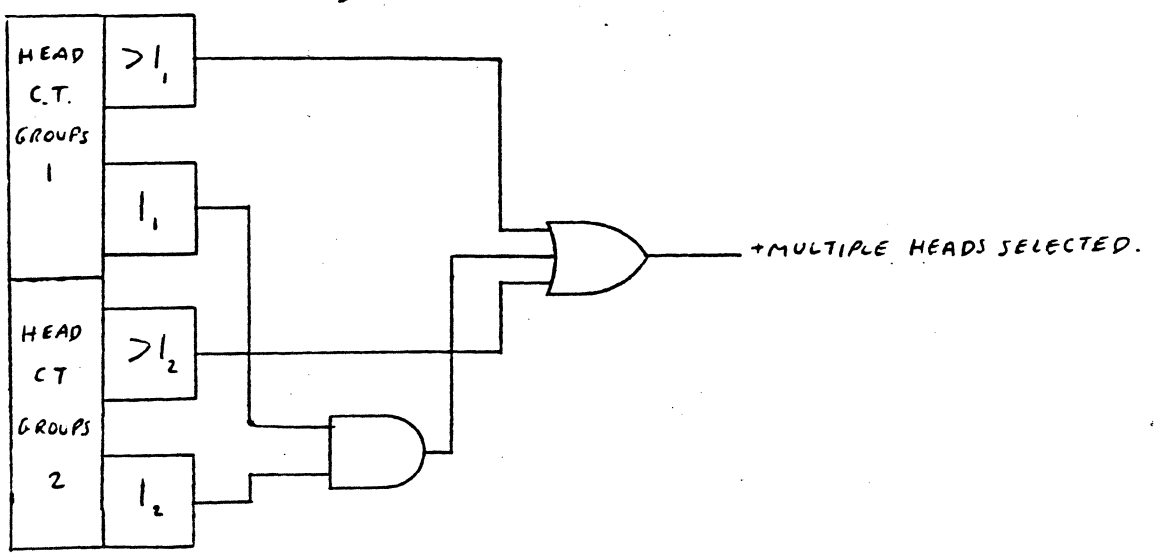


FIG 8·22

SENSING MULTIPLE HEAD SELECTION BETWEEN GROUPS AND WITHIN GROUPS

n this kind of failure. Usually we only try to sense no more than two al failure conditions, beyond that would greatly complicate the design.

The last arrangement shown in Figure 8.22 is for multiple groups of heads y one of which is permitted at a time. Here we require two groups of rational amplifiers, one to sense more than one centertap as we discussed l the other to sense one head centertap. The only difference in the design the value of R_B also taking into consideration that there are now two loads the R_C resistors instead of one for both cases. The extra requirement for ne 'One Sense' can be eliminated if the centertap selection of one head in ach group simultaneously can be accommodated, such as in parallel data machines nd for machines where discrimination occurs in the Write Driver Channel and he Read Channel. In this last case a matrix can be formed connecting head centertaps and Write - Read circuits in some convenient or cost effective manner (see Figure 8.23).

This matrix gives rise to a common terminology used throughout this industry of X and Y. The X circuits are the most expensive and are therefore minimized in number. The cheapest are the centertap circuits and are maximized where possible. Hence the centertap circuits are often referred to as Y select circuits.

8. Open Heads

   We discussed this situation in the section on transition sensing.

9. Shorted Heads

   This also was discussed in the transition sensing section.

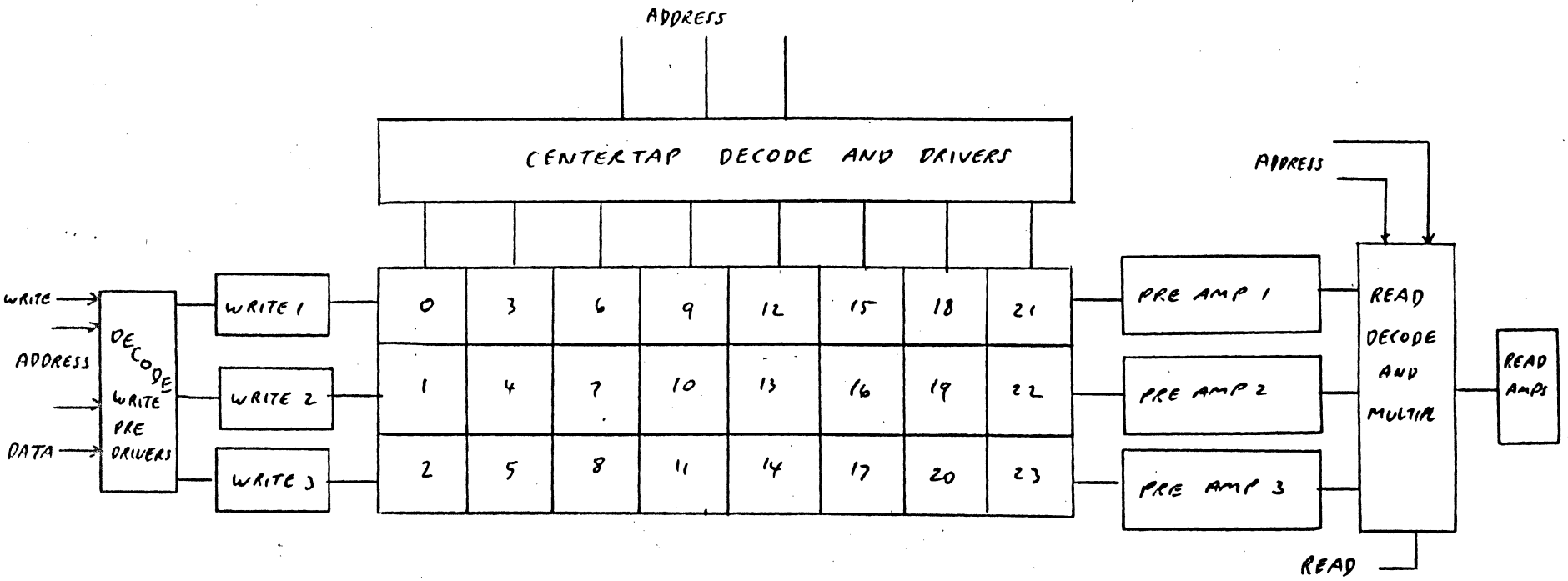


FIG 8.23

X - Y MATRIX ARRAINGEMENT OF HEADS AND CIRCUITS

10. Centertap Current Sense

In the cases where the centertapped head or a group of centertapped heads are driven or connected directly to an Integrated circuit collector, there is a distinct possibility of a short between the collector junction and the substrate. This situation exists due to the usual practice of tying the centertaps together to a common voltage or ground. The favored method of sensing this failure is to monitor the current in the centertap line. A simple series resistor and a comparator suffices. The bias on the comparator allows up to normal write current and any combined leakages to flow without changing the output state of the comparator.

There are many other circuits or conditions that could be monitored. These will evidence themselves to designers as they consider all the possible paths erasing current can take thru the head from whatever source or all the combinations that can prevent data from being correctly written such as any data encoding circuits or clocking circuits. They all need to be monitored and latched for the protection of stored data.

One side-light to the half open centertapped head case not previously discussed is the condition during selection by the centertap driver. Consider for a moment what we discussed back in Capter 7 on Matrices. We were very careful to balance the Read biasing currents so that the flux generated cancels. When a head coil opens on one side the bias current now is unbalanced and hence can partially erase the media. However, the flux is very small and is not usually of sufficient magnitude, if properly designed, to bring the media particle back into the open portion of the hysterisis loop. There is one other circuit parameter that we have neglected and that is the single ended matrix capacitance. See Figure 8.24. When a head is reverse biased, the cable, head, and diode capacitance is charged with some number of coulombs. If the centertap selection

transistor is turned on quickly, this permits a discharge current to flow thru the head half winding that normally is balanced out that now is of sufficient magnitude to cause the media to be brought out of saturation thus actually writing a disturbance on the media. Because of this effect it is doubly important to minimize head capacitance, and maximize the discharge rate while maintaining sufficient drive to saturate the centertap driving transistor during writing. This then refers back to the circuit used to drive the bases of the centertap drivers which we did not discuss at that time.

Once we have sensed a safety related failure we must now determine what we can do to minimize the damage to the recorded data from either timing or circuit related failures. The best thing we can do is to prevent the flow of head current either by blocking the current path at any point or several points or by sinking the current off to ground before it can reach the head. The first type usually are limited to shutting off the current sources and the centertap drivers logically. These precautions work well for command related failures or timing failures, but they do not work if either the current source transistors shorted or the centertap transistors shorted or some other current path became established. Again, shutting off all current paths can block the current thru the head either at its source or at its sink despite a failure of one of the sources or sinks (centertap circuit). We can provide a circuit that clamps the current source to a potential that reverse biases the Write Driver emitters. Part of the precautions already exist when we added a series diode to protect the Write Driver emitter from the Current Sensing circuit during the off condition. Refer back to Figure 8.2, R1 and D1. If we connect a large

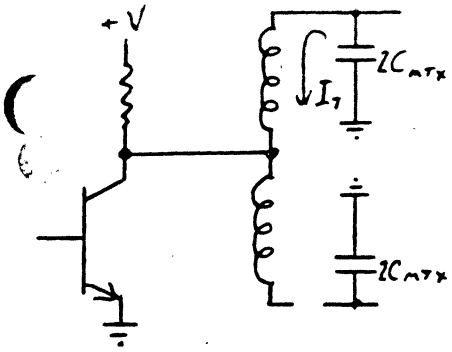


FIG 8.24

SHOWING UNBALLANCED TRANSIENT HEAD CURRENT AT SELECT TIME

FIG 8.25

CURRENT CLAMPING ON RECEIPT OF ERRONIOUS WRITE CURRENT OR SENSING A FAILURE. OR WHEN NOT WRITING

transistor to the collector of the Current Source Transistor and tie its emitter to a potential more positive (in this case) than the bases of the Write Drivers, we can turn this transistor on immediately upon sensing any failure. This is achieved by an 'OR' tree driven from each of the safety circuit storage latches. Some engineers like to turn it on every time we are not writing in order to quickly discharge the current source lines. The transistor and its base current drive must be capable of handling any current resulting from shorting the current source transistor which means the current would be that obtained thru the emitter resistor in series with the difference in potential between the source supply and the clamp transistor emitter reference. This must be worse cased as failure to do this might allow forward bias one of the Write Drivers. A circuit is shown in Figure 8.25 for the NPN Write Driver with the centertap referenced to ground. Other configurations are left to the reader.

A typical design would ensure the following equations are met.

$$I_{Q_4\,MAX} = \frac{V_{-\,MAX} - V_{SAT_4\,MIN}}{R_{1\,MIN}} \qquad (\text{shorted } Q_3) \qquad \text{(EQ 8.24)}$$

$$\frac{V_{-\,MIN} - V_{be_4\,MAX}}{R_{3\,MAX}} = I_{R_3\,MIN} \geq \frac{I_{Q_4\,MAX}}{\beta_{Q_4\,MIN}} \qquad \text{(EQ 8.25)}$$

$$2V_D - V_{SAT_{DRIVER}} \geq V_{be_4\,MAX} \qquad \text{(EQ 8.26)}$$

$$V_{b_4\,REV} \geq 0.0V \qquad \text{(EQ 8.27)}$$

$$\frac{V_{-\,MAX}}{R_{3\,MIN}} \leq \frac{V_{+\,MIN} - 2V_{D\,MAX}}{R_{7\,MAX}} \qquad \text{(EQ 8.28)}$$

If all these 5 equations are met simultaneously then the clamp will operate correctly even in the event of a shorted current source. Notice we are worse than worse case, but totally safe as we used V- min and V- max as simultaneous conditions in EQ 8.24 and 8.25 respectively.

This concludes the current related safety considerations.

IT MIGHT BE PROFITABLY CONSIDERED THAT ANY CHANGE IN CIRCUIT CONFIGURATION MAY USUALLY BRING ITS OWN CURRENT PATHS THAT COULD DESTROY DATA. A CASE IN POINT IS THE CURRENT USE OF INTEGRATED CIRCUITS AND CENTERTAPPED HEADS. IF A COLLECTOR TO SUBSTRATE SHORT OCCURED THEN THERE IS A PATH FOR CURRENT FROM THE CENTER TAP THRU ONE HALF OF THE HEAD TO THE COLLECTOR AND THENCE INTO THE SUBSTRATE WHICH IS THE MOST NEGATIVE POTENTIAL. THIS THEN WILL ERASE THE DATA ANY TIME THERE IS A SUITABLE DIFFERENCE IN POTENTIAL BETWEEN THE CENTER TAP AND THE SUBSTRATE. THIS IS EASILY MONITORED BY SENSING THE CURRENT IN THE CENTER TAP LINE DURING READ AND WHEN GREATER THAN THE WRITE CURRENT DURING WRITE

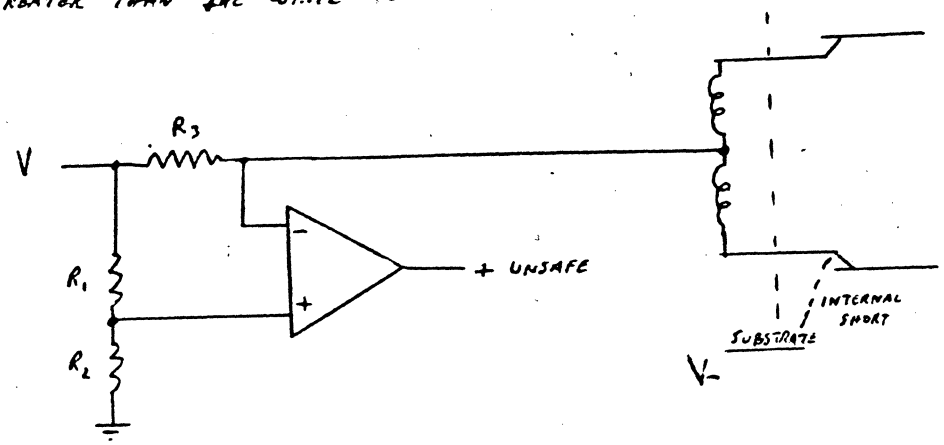


FIG 8-26 A

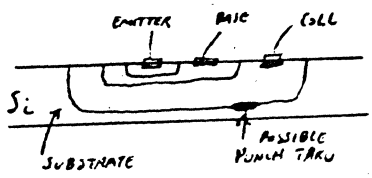


FIG 8-26 B

FAULT $\Rightarrow$ $\dfrac{V\,R_2}{R_1 + R_2} > I_4\,R_3$

SAFE $\dfrac{V\,R_2}{R_1 + R_2} < I_4\,R_3$

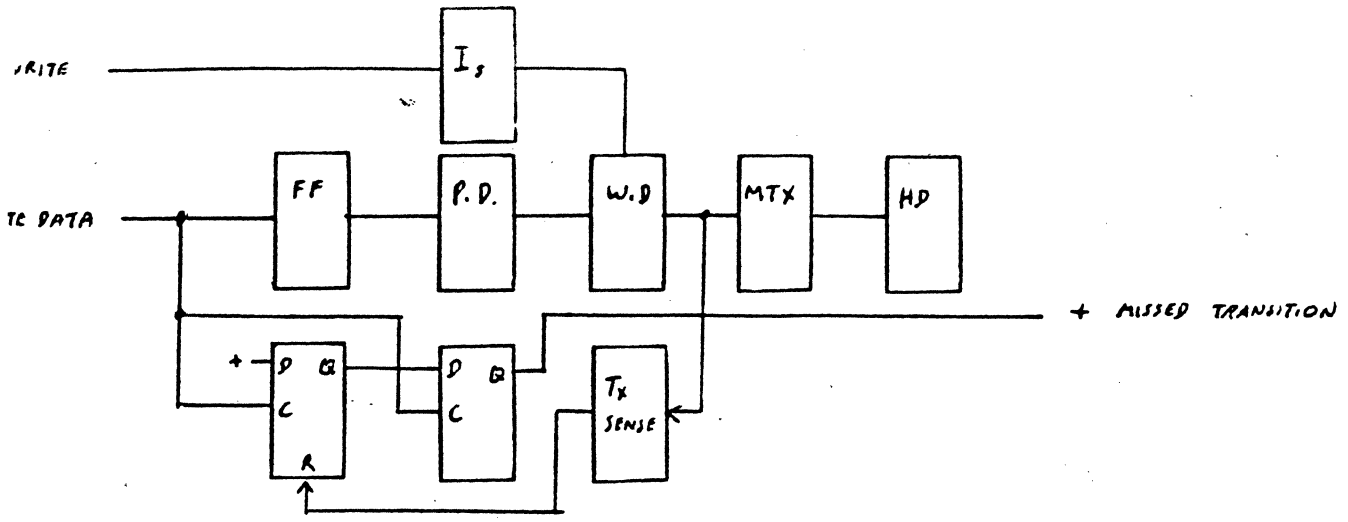


FIG 8.27 MISSED TRANSITION CIRCUIT THAT IS NOT TIME CONSTANT DEPENDENT

IF DATA WERE MISSING THE CIRCUIT WILL NOT RESPOND

IT DOES REQUIRE A SECOND DATA MONITOR CIRCUIT.

USEFUL IN INTERGRATED VERSIONS TO PREVENT OBSOLESCENCE.

DUE TO DATA CODE AND DATA RATE.

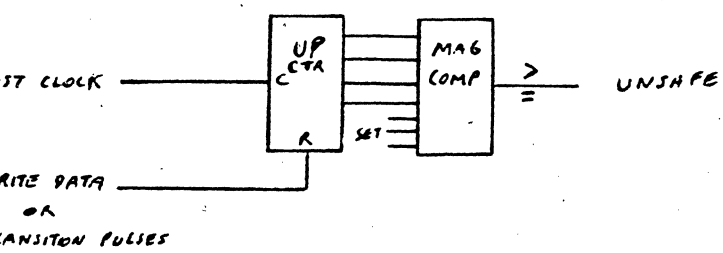


FIG 8.28 CODE RELATED TRANSITIONS UNSAFE CIRCUIT

THE COUNTER AND MAGNITUDE COMPARATOR IS SET FOR THE MAX CLOCK SPACING OF THE CODE USED

FOR EXAMPLE, IF THERE CAN BE 5 CLOCK SPACES BETWEEN TRANSITIONS AS PERMITTED BY THE CODE THEN THE MAGNITUDE COMPARITOR SHOULD BE SET FOR '5', TO USE THE GREATER OUTPUT ONLY, OR TO 6 IF THE EQUAL AND GREATER OUTPUTS ARE 'OR'ED TOGETHER.

DETECTORS

This chapter deals with the detectors or the circuitry used to sense the transitions recorded on the media. We will discuss the intervening amplifiers in the next chapter as the type of amplifier depends on the type detector chosen. The detector in turn depends entirely on the head signal and the region of operation. Let us refer to Figure 9.1. This figure is our old friend the bit density curve with some added segments. Historically, all magnetic recording was done using the left hand side or the 'good' resolution portion, Region 1. Here the various frequency or density components result in very similar amplitudes. Therefore a string of transitions produces signals of fairly equal amplitudes with little interaction. The process of detection is to sense the peak which results from the instant of maximum time rate of change of flux of the transition and output a pulse, the leading edge of which, corresponds to the center of the transition. Circuitry for doing this consists of some amplitude reference and a peak sensing circuit. The amplitude reference serves to eliminate noise associated with the signal base line, Figure 9.2. As the signal is bipolar there needs to be two references or some means of changing the signal to unipolar. A Full Wave rectifier fills this latter function quite well. The peak sensing circuits are required to be very accurate. For this reason, amplitude sensing circuits fail. Consider the peak of sine wave. The change of the amplitude as a function of time is very poor, changing only a few percent over a considerable number of degrees. Amplitude sensitive circuits then will have poor time resolution of the peak. They are also sensitive to the variations in amplitude of each transition in a chain for no pulse is of the same amplitude as its neighbor due to variations in media and head-media mechanics.

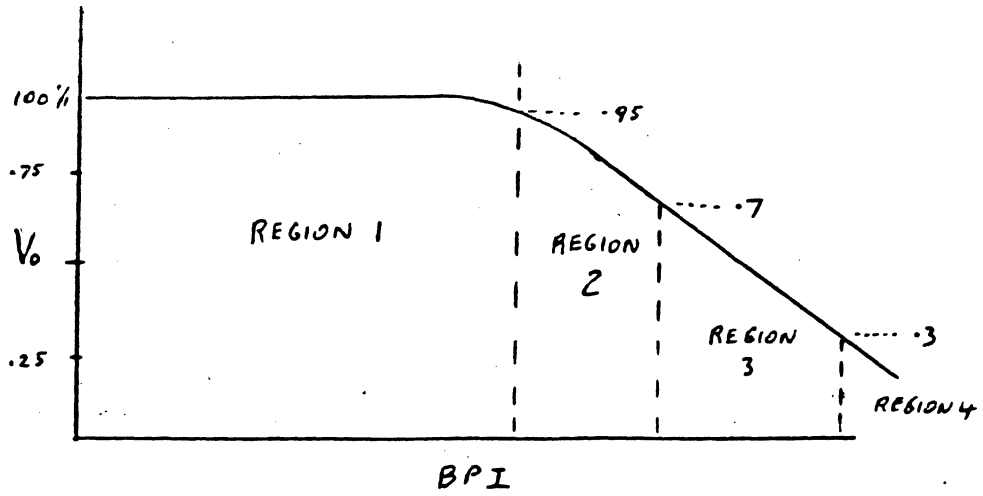


FIG 9.1 BIT DENSITY CURVE / DETECTOR REGIONS
DIVISIONS NOT FIRM.

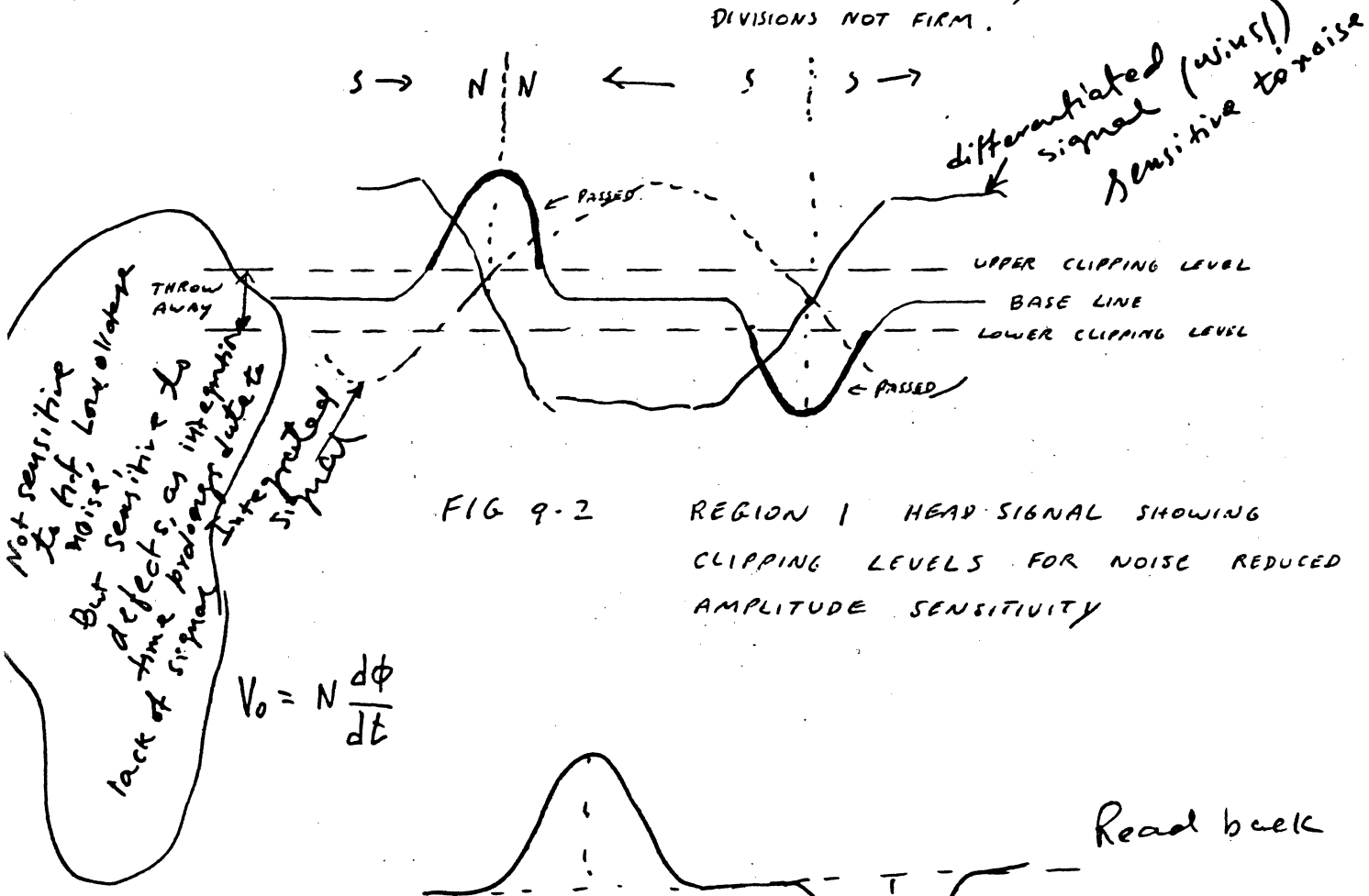
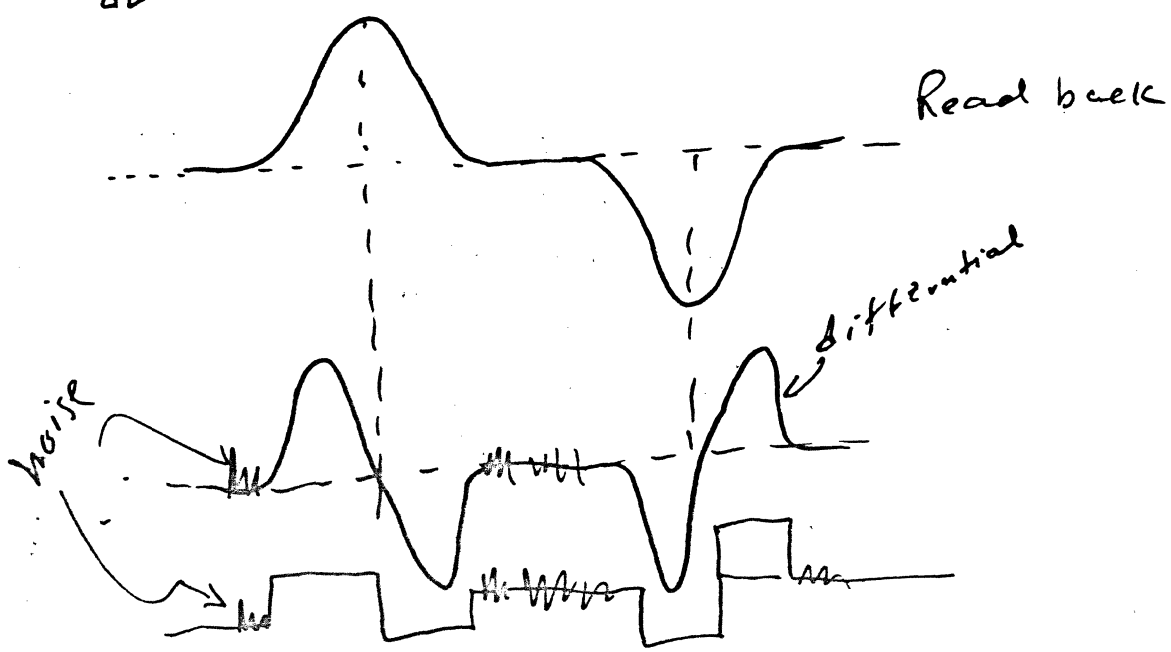
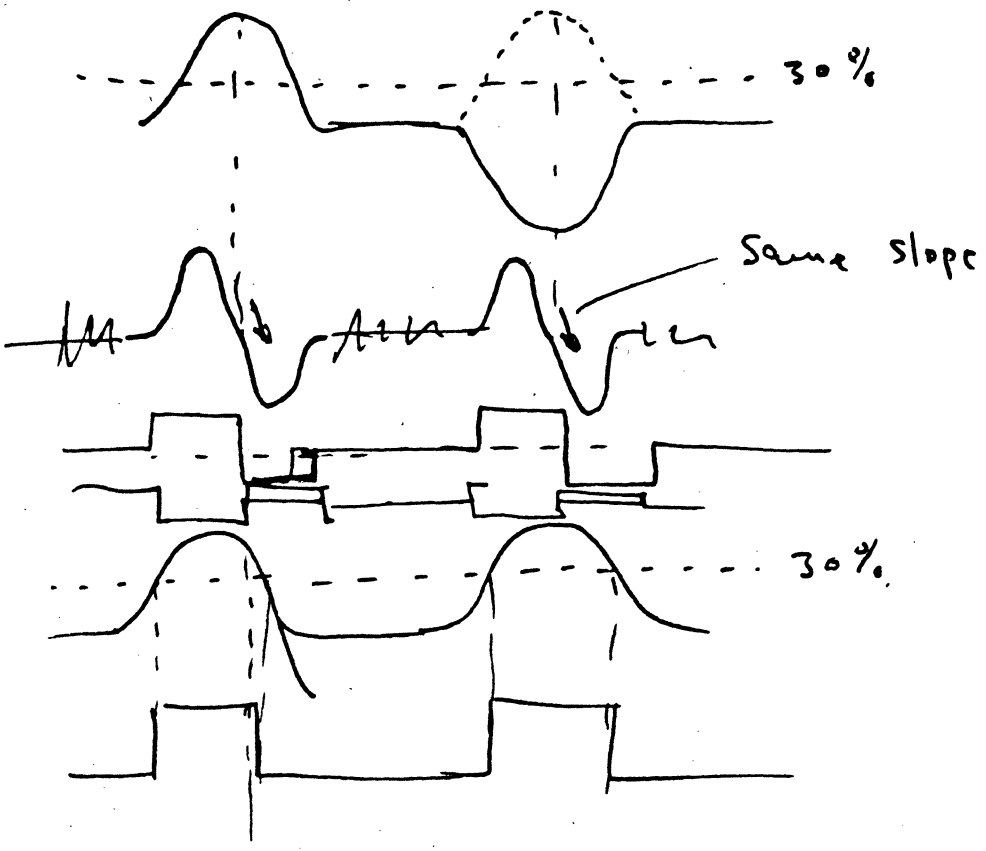


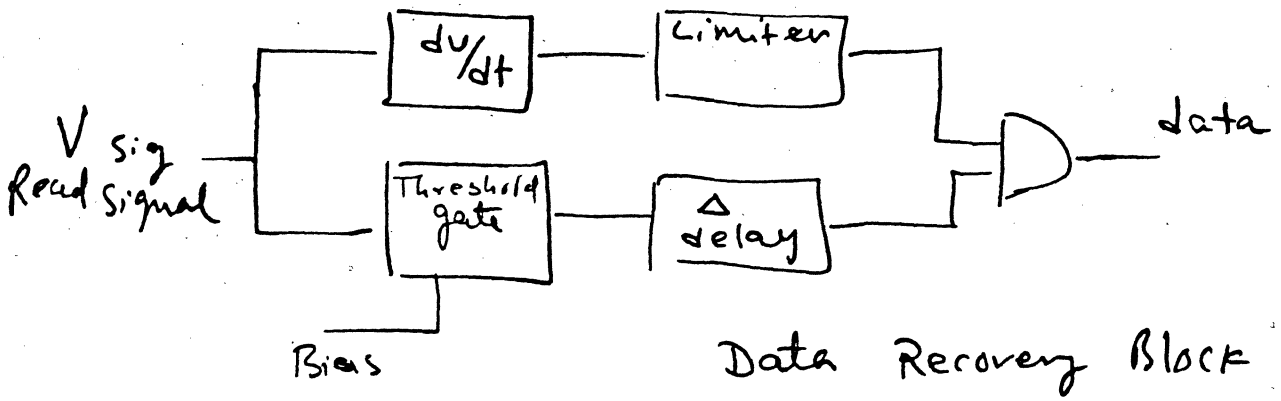
FIG 9.2 REGION 1 HEAD SIGNAL SHOWING CLIPPING LEVELS FOR NOISE REDUCED AMPLITUDE SENSITIVITY

$$V_o = N\frac{d\phi}{dt}$$

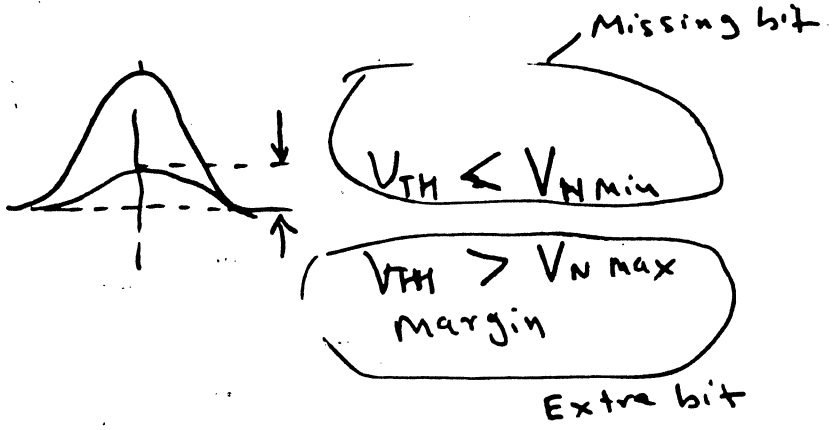


Detector





Data Recovery Block diagram


Missing bit
$V_{TH} < V_{H\,min}$
$V_{TH} > V_{N\,max}$
Margin
Extra bit

Two other signal processing circuits resolve the dilemma. They are the differentiator and the intergrater. Here the slowly changing peak value becomes a fast changing base line crossing as the slope of the peak changes from one polarity thru zero to the other. This result is ideal as we can build base line crossing sensing circuits very easily. The question now becomes which of the two is better. First the differentiator. The circuit or amplifier has a zero at the origin. The gain then increases from zero at zero frequency to infinity at infinite frequency. Our head signal is complex in that it can be described as a fundamental sine wave with many harmonics. Noise also enters the picture. As we pass the bandwidth of interest, the gain continues to increase, therefore, all the noise of higher frequencies will be amplified accordingly while the signal of interest remains at its lower gain. The result is a decrease in signal-to-noise ratio. Practical circuits have finite bandwidth as they roll off due to stray capacitances introducing a pole. We see then that practical differentiators have signal-to-noise ratio problems but are limited to the bandwidths of the circuitry.

The intergrator on first look is ideal as its output results from a pole at the origin, or zero frequency, which produces decreasing gain for increasing frequency. If we were only dealing with a sine wave, the resulting waveform would always be symmetrical; however, we are dealing with a complex waveform containing many harmonics and noise which also vary in amplitude from pulse to pulse. Ideally, then we would have a base line crossing for each transition as the area under the curve alternates. But the signal's non-uniformity will result in variations in the time of the base line crossing from the actual peak time of the input signal.

DETECTORS

When compared to the noise shifted differentiated signal base line crossing and the quiet but time shifted base line crossing of the integrated signal, we are forced to choose the lesser of two evils. The differentiator has dominated the application mainly due to the way the higher frequency noise is rolled off by judicious choice of the pole location of practical circuits. The operation of the intergrator in the presence of defects accentuates the inaccuracies. Compromises can be worked out using carefully placed zeroes to minimize its sensitivity to defects and amplitude variations, but no practical recording channel has succeeded in mass production using the approach.

We will now develop a circuit that will perform the detection function previously described. The block diagram for such a detector is shown in Figure 9.3. The full wave rectifier can be built from the differential signal with a pair of diodes. See Figure 9.4. Because the differential signal contains two positive peaks for each pair of pulses, the diodes will pass two positive peaks each one corresponding to a transition. Depending on the biasing, the output dc potential will be one diode drop below the input base line or dc potential. The driving circuit must ensure the two dc levels to be equal or unsymmetry will result. Where the input and output are referenced to ground the diodes will subtract one diode drop from the signal before passing it on. Figure 9.5 illustrates this function. We could take advantage of this phenomenon by making the diode drop equal to the amount of signal around the base line (clipping level) that we want to remove to reduce sensitivity to base line noise. Under these circumstances the signal itself must be amplified to an amplitude such that V diode equals the percentage of the signal we want to remove. For example, if we want to remove $\pm$ 10% of the signal around the base line, the input signal must be amplified to $2\left(\frac{100\%}{10\%}\right)(V_D)$in V_{pp} diff. (EQ 9.1)

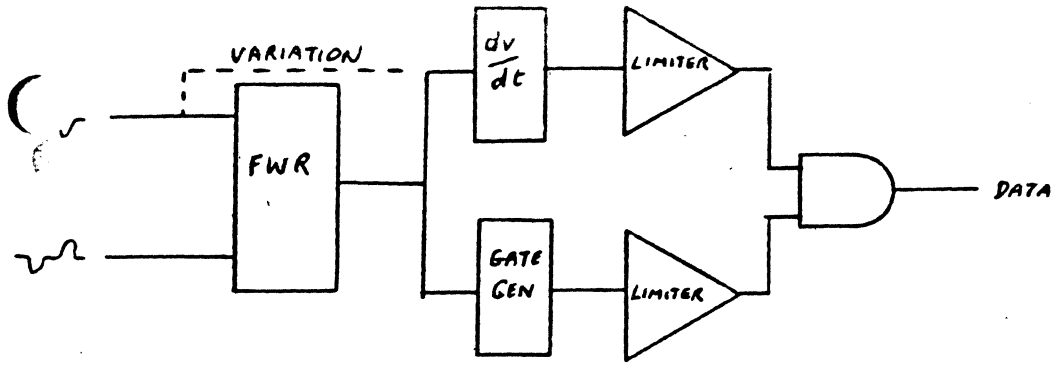
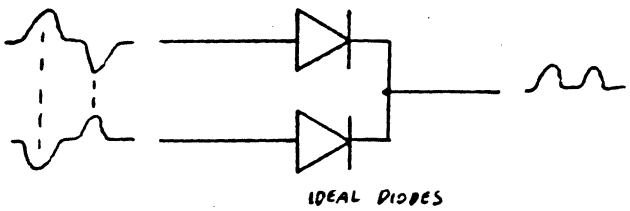
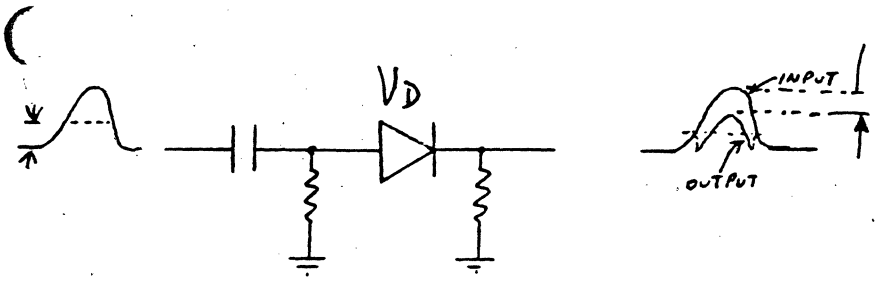


FIG 9.3 REGION I DETECTOR
BLOCK DIAGRAM
DIFFERENTIATOR CAN BE DRIVEN DIRECTLY OR FROM THE F.W.R.



no perfect diode

FIG 9.4 FULL WAVE RECTFIER



Diode drop V_D = 0.7
0.4 schottky

FIG 9.5 RECTIFIER CAUSED AMPLITUDE AND WIDTH REDUCTION

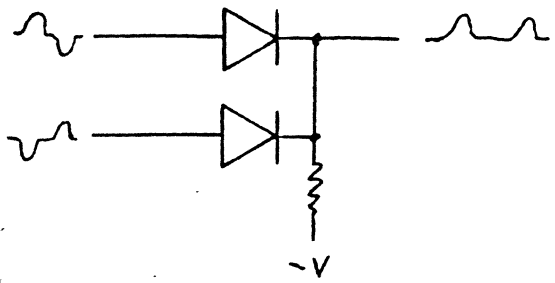


FIG 9.6 RECTIFIED SIGNAL HAS FULL AMPLITUDE BUT SHIFTED ONE DIODE DROP

Such a circuit will function but is not easily changed if we want to change the clipping level percentage except by changing the amplitude of the input signal or the references. The signal swings are very large, as in our example: 2(10)(.7V) = 14.Vpp diff. is required. (EQ 9.1)

When the full Wave Rectifier is DC coupled with an appropriately negative return voltage as shown in Figure 9.6, then the output amplitude is equal to the input amplitude but is one diode drop below the input signal voltage. We need not get confused if we reverse the diodes to pass the negative peaks instead. Input DC balance is required for correct symmetrical operation.

We could perform this function with active circuits such as is shown in Figures 9.7 or 9.8. The advantage of the emitter follower connection is the lowered impedance provided by the transistor. It has one disadvantage and that is the reverse base-emitter breakdown voltage, BV_{ber}, restricts the input signal peak-peak value for one emitter is conducting while the other is cut off. Again, input balance is required. The addition of the third emitter follower permits any percentage of clipping. It is similarly restricted to low amplitudes of around 7.V_{pp} SE max due to BV_{ber}. Another consideration is that of the transfer curves at around 0.1 volts difference where there is partial conduction of the transistors. The operation is very similar to a positive 'OR' circuit meaning only the transistor with the most positive base will conduct. It is this characteristic that permits Full Wave Rectification or Biased Rectification. The input amplitude must be such that the 0.1V uncertainty is small compared to the signal of interest but within the restrictions imposed by the emitter-base breakdown voltage.

We will now turn our attention to the Gate Limiter. Its purpose is to provide a gate to operate the 'and' circuit at the appropriate time to allow the pulse resulting from the differentiated signal peak to pass while blocking all noise related signals. The output of the Full Wave Rectifier is limited either thru direct amplification or that resulting from positive feedback such as in a Schmidt Trigger. The latter circuit has the advantage of reducing the effect of noise around the threshold of the Schmidt, whereas the former will be noisy around the bias point. The gate threshold level results from either the Schmidt threshold or the amplifier bias point. Figure 9.9 illustrates the waveform relationships we want to design into the total circuit.

If we design a limiting amplifier that limits around the clipping level we desire, then we have performed the function we need. From Figure 9.9 the clipping level value is determined as a percentage of the input signal magnitude. Let us use 15% of 10.Vpp diff. This becomes 5.Vpp SE or 2.5V_{BP} out of the Full Wave Rectifier and 15% is 375.mV. We next need to guarantee the reference of the input signal to the Full Wave Rectifier such that we have control over the percentage. We also need to include the tolerances of the diodes or transistors used to build the Full Wave Rectifier. We can do this with the circuit shown in Figure 9.10.

Transistors 1 and 2 perform the Full Wave Rectification. Transistors 3 and 4 are a high gain differential amplifier. The input reference is ground. The bias reference is 375.mV with transistor 5 providing the same or similar Vbe drop as transistors 1 and 2 if current sources 1 and 3 are equal. This function is harder to perform with discreet components, but is very easy today

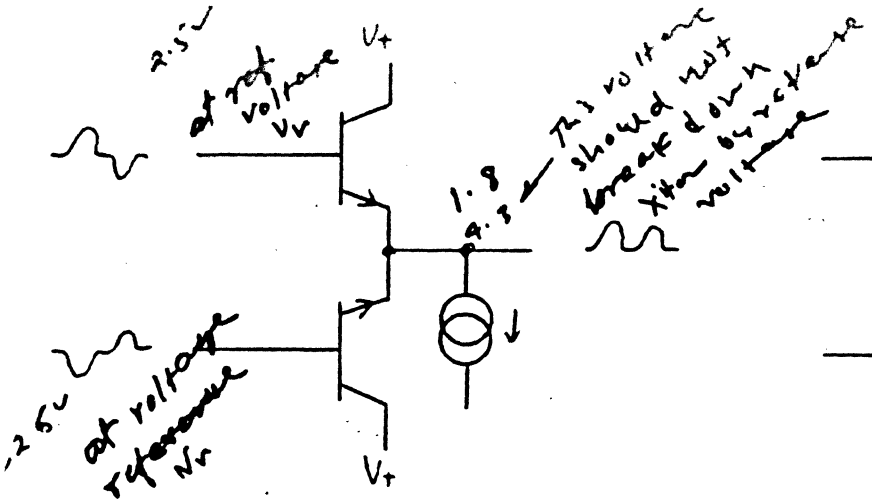


FIG 9.7 TRANSISTOR FULL WAVE RECTIFIER

same as 9.6, but with ~~diode~~ Transistor

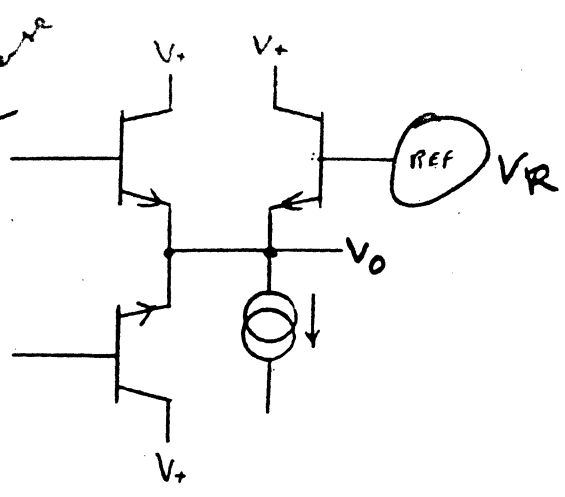


FIG 9.8 BIASED TRANSISTOR CLIPPING FULL WAVE RECTIFIER

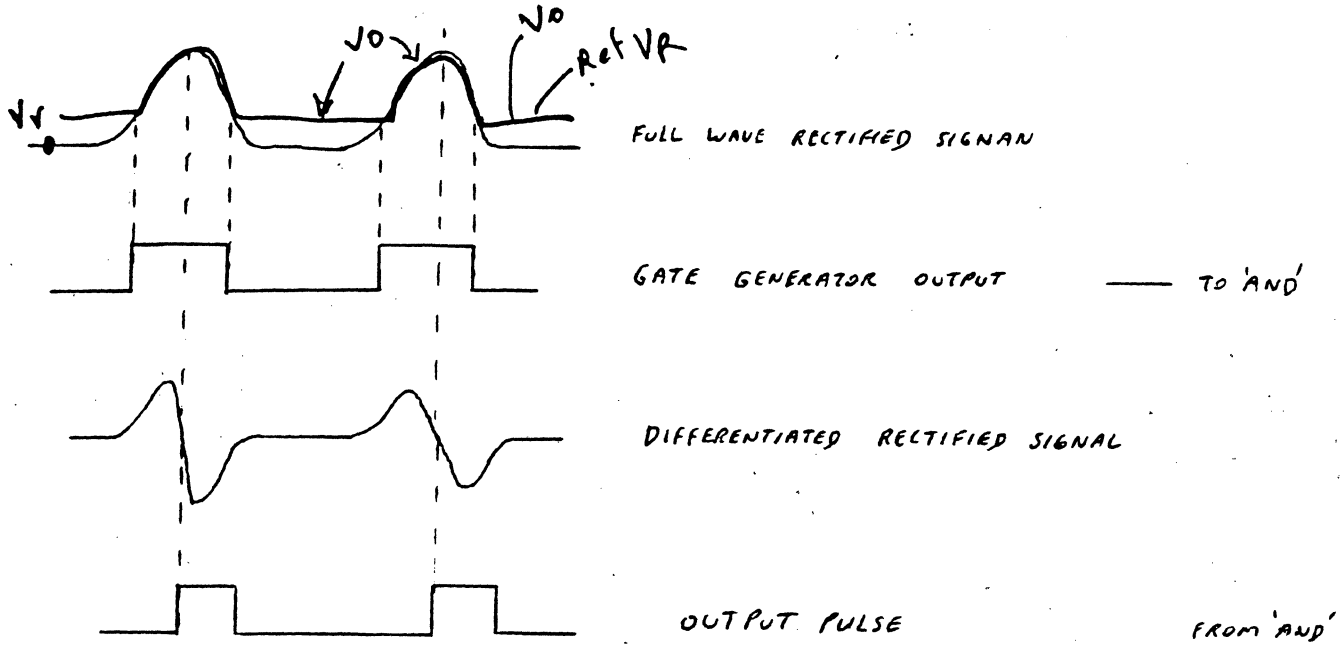


FIG 9.9

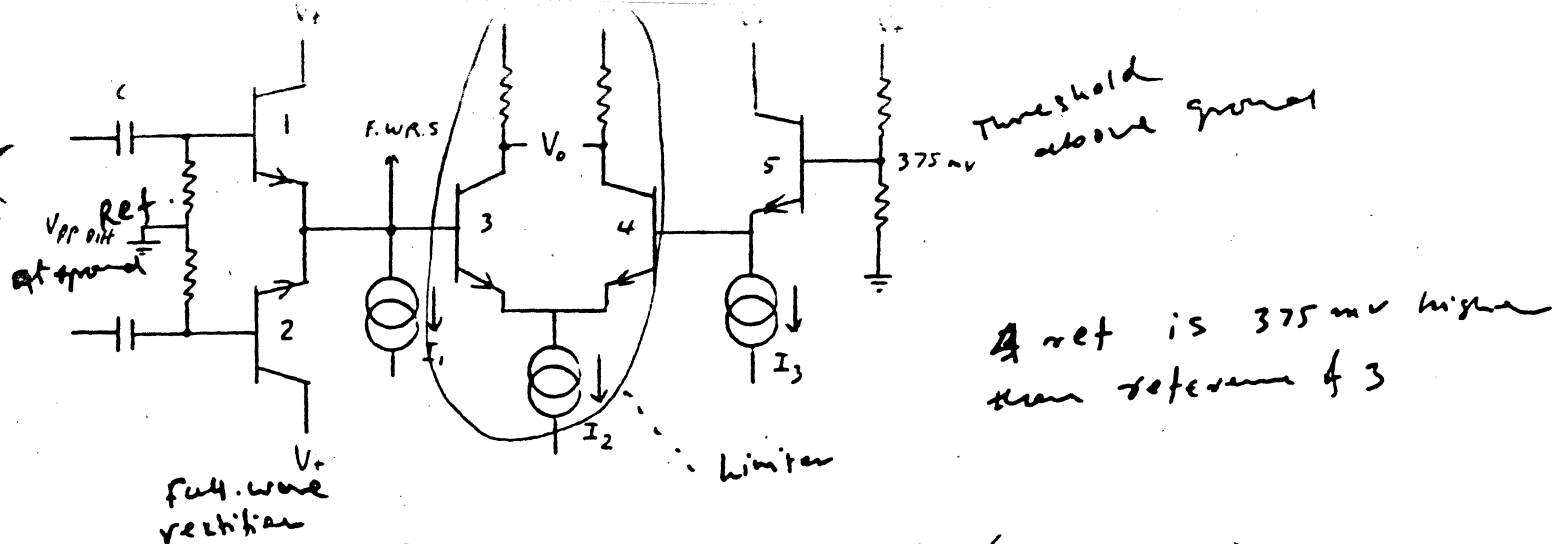


FIG 9.10 GATE GENERATOR (15% CLIPPING)

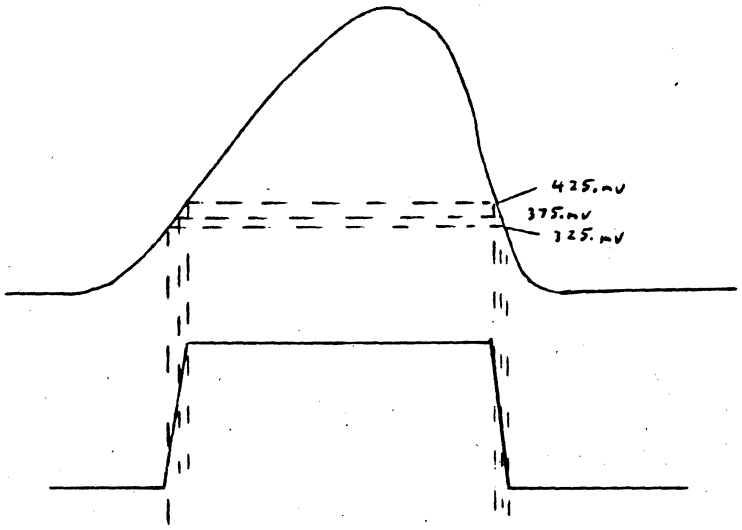


FIG 9.11 LIMITED SIGNAL V_o

when we can use the inherent matching in integrated circuits. The uncertainty of the bias point depended heavily on the variations in Vbe from transistor to transistor or diode to diode. The output of the amplifier becomes the output gate. If the signal swing is insufficient then further amplification is necessary especially if the slope of the gate edges is poor. For our 2.5V BP amplifier input, we know that the output will be a squarewave, Figure 9.11, centered around the 325 mV to 425 mV, but modified by the bandwidth of the amplifier. The more precise we want the gate edges defined, the higher the gain we require. With limiting it is not the output amplitude swing that defines the slope, but the gain divided into the input signal slope for the specified output swing. For example, if our gain were 100 and the output swing was limited to 2 volts then the input equivalent change would be $\frac{2.V}{100} = .02$ volts for a full swing of the output. This then would indicate that the output gate edge would be similar to the time it took for the input to change from 365 to 385 mV if the amplifier bandwidth is adequate. With proper biasing the output gate levels could be made compatible with some logic family which would make the following 'and' gate simple to construct.

The considerations then are the bias point stability, the gate edge slope and logic compatibility. There is a propagation delay consideration particularly when using multiple gain stages in series to obtain the required gain.

The Schmidt version is not too much different. Positive feedback is provided to the bias point in order to interfere least with the Full Wave Rectified signal fig.9.12 The feedback resistor R_F is connected between the out-of-phase output and the bias reference network. The percentage feedback is determined by the signal swing on the collector and the resistor divider

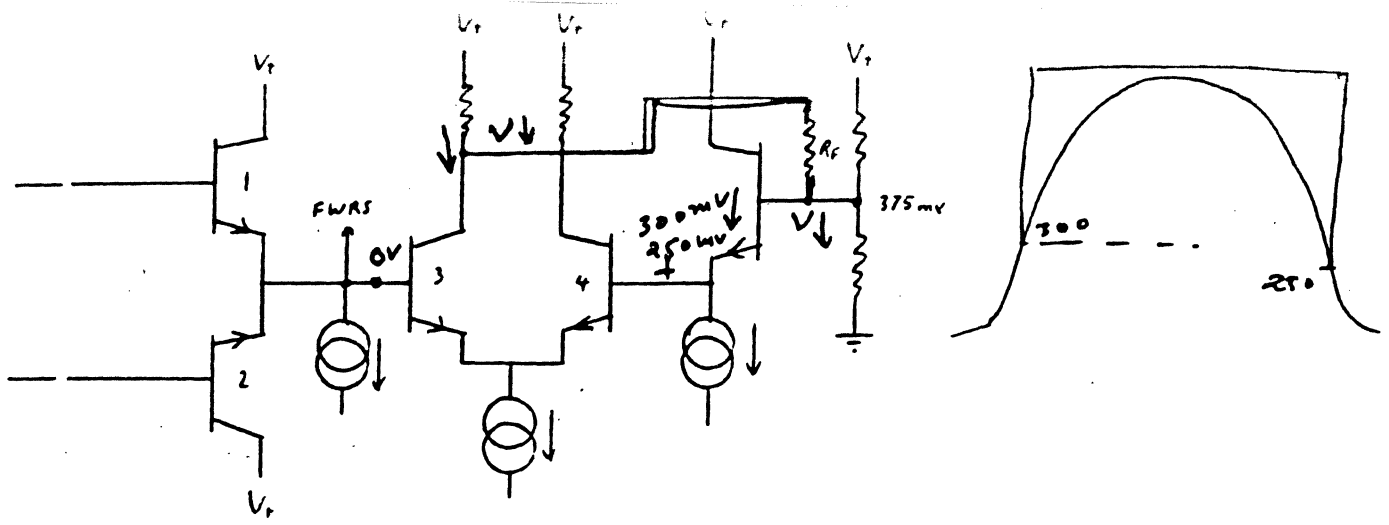


FIG 9.12 SCHMIDT CONNECTION

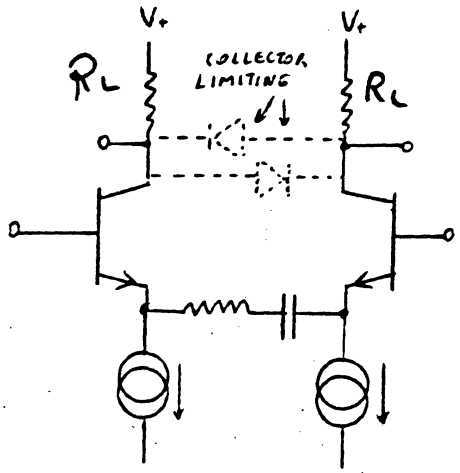


FIG 9.13 A

BALANCED DIFFERENTIATOR

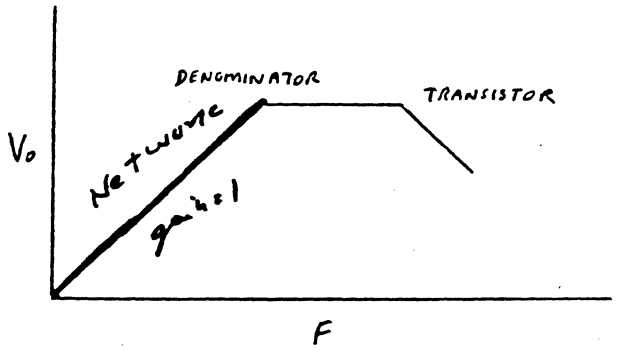
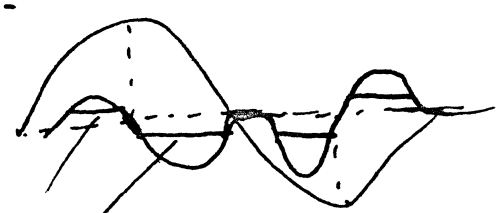


FIG 9.13B

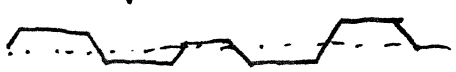
BODE PLOT OF BALANCED DIFF.

$$A = \frac{2R_L}{R_E + 2R_m + 2r_e + \frac{1}{cs}}$$

=



Clamping diode



network to the input signal. This circuit is easier and cheaper since it provides a large gain, fast slopes and a more stable bias reference than a multiple amplifier chain required to get the same slope. Also a second advantage is the shorter propagation delay and freedom from noise while traversing the bias point.

The differentiator may take several forms ranging from passive differentiator to active. Operational amplifiers are usually not suitable due to their restricted bandwidth. We can build discreet active differentiators fairly easily. Consider a differential amplifier with a capacitor in the series feedback path or a resistor and capacitor.

The gain equation simply becomes from Fig 9.13 A

$$A = \frac{2R_L}{R_E + 2r_e + \frac{1}{CS}} = \frac{2R_L CS}{(R_E + 2r_e)CS + 1} \qquad \text{(EQ 9.2)}$$

When rearranged, we can see both the zero at the origin, which we desire, and a pole which would be helpful to restrict the bandwidth related noise. There are other poles resulting from the internal transistor and stray capacitances. The circuit is inherently AC balanced in the emitter but does require DC current source balance and load resistor balance if output DC levels are important.

The same circuit can be built using an inductive load with similar results. Here a single current source could be used if we wanted to. See Fig. 9.14.

$$A = \frac{2LS}{R_E + 2r_e} \qquad \text{(EQ 9.3)}$$

Notice the absence of the pole. We do not escape as easily since all inductors have stray capacitance and series resistance. The Bode plot will show a roll

due to this capacitance and it will be second order with a zero close to the origin, but not at the origin.

$$A = \frac{2\left(\frac{(LS + R_s)\frac{1}{CS}}{LS + R_s + \frac{1}{CS}}\right)}{R_E + 2r_e} \qquad \text{(EQ 9.4)}$$

$$= \frac{2(LS + R_s)}{(R_E + 2r_e)(LCS^2 + R_sCS + 1)}$$

Some have tried to make the AC unbalance due to the separate inductors tolerances more balanced by winding the two inductors on a single core using bifilar circuits. If we were to use a simple RC coupling network as a differentiator we would have to contend with the attenuation. With the above amplifiers we can adjust the gain for a net gain instead of a loss. Of the two active differentiators considered, the first has a serious difficulty with DC stability if it is to drive a sensitive threshold circuit. The inductor version has a very low IR drop therefore is insensitive to variations in the current source or load resistor and current source balance. One way to retain the advantages of the first circuit is to add a balancing circuit to the current source and make the following stage differential with a large common mode input range. Balance is simply obtained by the use of a potentiometer in series with the two sources in one of several configurations. Some are shown in Figure 9.15.

Regardless we can use a differential following stage to provide the limiting function. The emitter followers are required to minimize Miller Effect. The circuit shown in Figure 9.16 consists of a differentiator followed by a differential amplifier. Here we choose not to use a Schmidt trigger as we

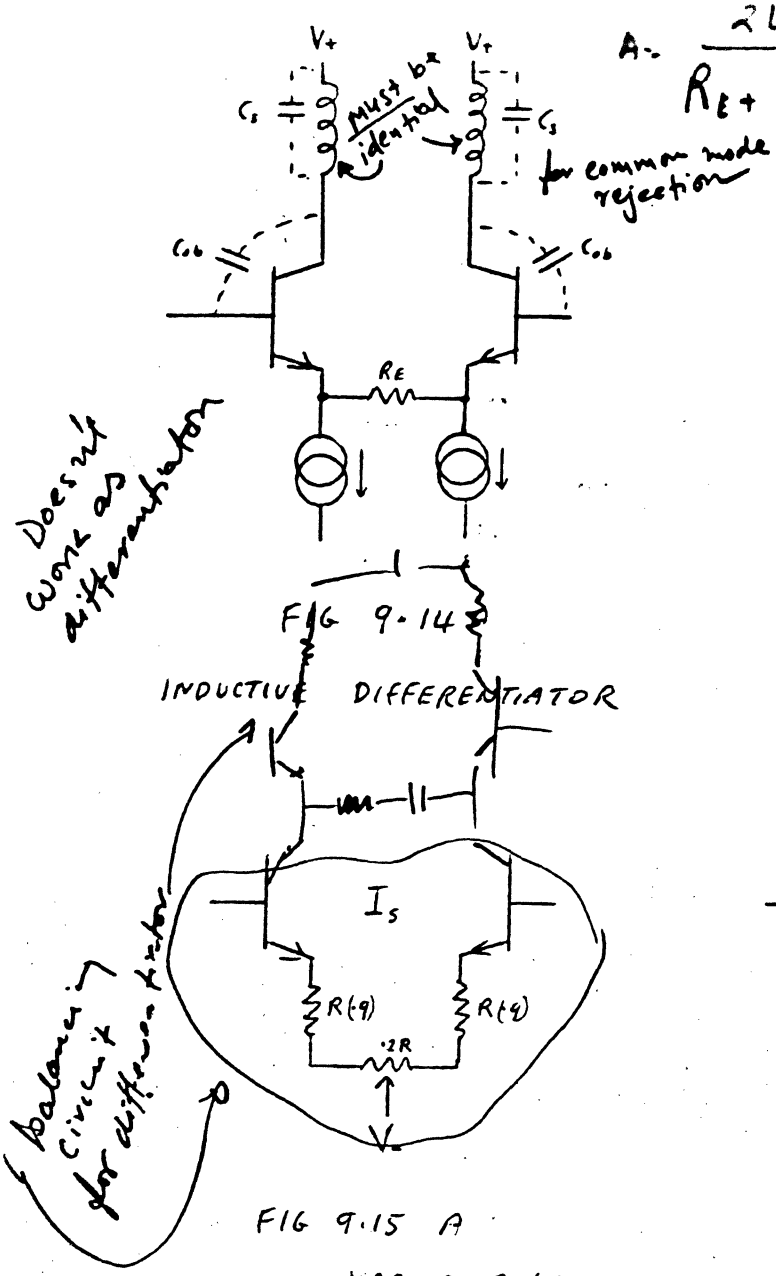


FIG 9.14 A

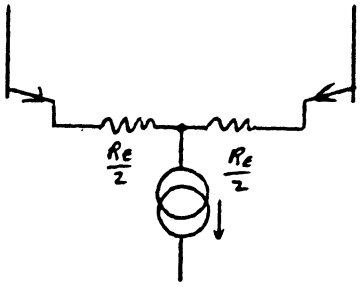


FIG 9.14 B

ALTERNATE EMITTER CIRCUIT

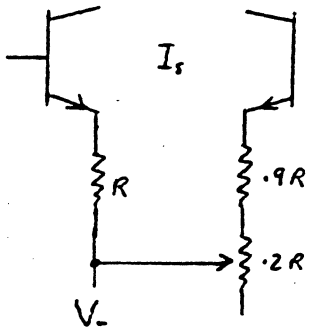
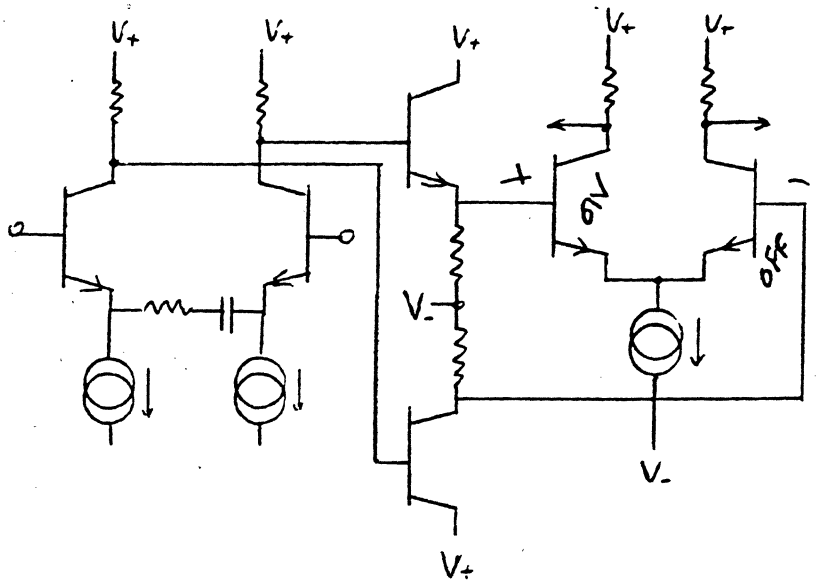


FIG 9.15 A

FIG 9.15 B

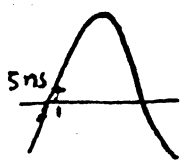
VARIOUS BALANCED CURRENT SOURCES FOR THE DIFFERENTIATOR.



$$\text{gain} : \frac{R_L}{R_m + r_e} = \frac{R_L}{R_m + \frac{2\alpha}{I_s}}$$

FIG 9.16 DIFFERENTIATOR AND LIMITER

want our output edges to correspond exactly to the center of the differentiated pulse peak which is the base line crossing out of the differentiator. The amount of gain required depends on the accuracy we require. If we were dealing with a sine wave input we could easily calculate the required gain. Assume the input signal were a $(2.5 - .375)V_{BP}$ sine wave, $0\text{-}180^{0}$, of 1.MHz timing or $\frac{0.5\mu s}{180}$ for 2.77 ns/degree. If our logic 'and' gate had a minimum rise time of 5.0 volts in 5.55 ns then we should provide sufficient gain to make 2.0 degrees of input signal at the base line equal to 5.0V logic level. Therefore, the minimum gain of the limiter and differentiator should be:



$$A \simeq \frac{5.0\,V}{(2.5-0.375)(\sin 2^{\circ})} = 67.4 \text{ min}$$

(EQ 9.5)

OR

$$\frac{V_{Logic}}{\sin\left(\frac{360^{\circ}\cdot F\cdot Tr}{\pi}\right)\cdot V_{in\,min\,pp}}$$

Tr = rise time of Logic family.

Going back over our bandwidth restrictions for the collector load resistor, we should provide this gain in several stages instead of one.

To get our gain we would probably require three stages. The gain per stage is $\sqrt[3]{67.4}$ or 4.06 per stage. We can do this fairly easily with basic emitter coupled amplifiers similar to that shown in Figure 9.17. Because the input to the differentiator is single ended, we need another gain of 2. If the differentiator is designed for a gain of 2 then we need 3 other limitors with a gain of 4.06 min each. The next problem we need to solve is the cascading bias required with DC coupled amplifiers. With discreet amplifier construction we could alternate NPN - PNP amplifiers and thus maintain a reasonable power supply voltage. This is probably the best way to proceed as AC coupling requires a knowledge of the low frequency band-pass requirements and hence the data code's spectrum.

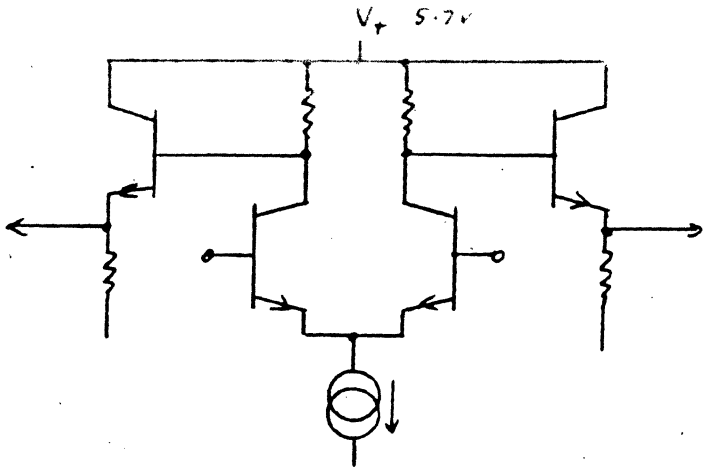


FIG 9-17 LIMITING AMPLIFIER

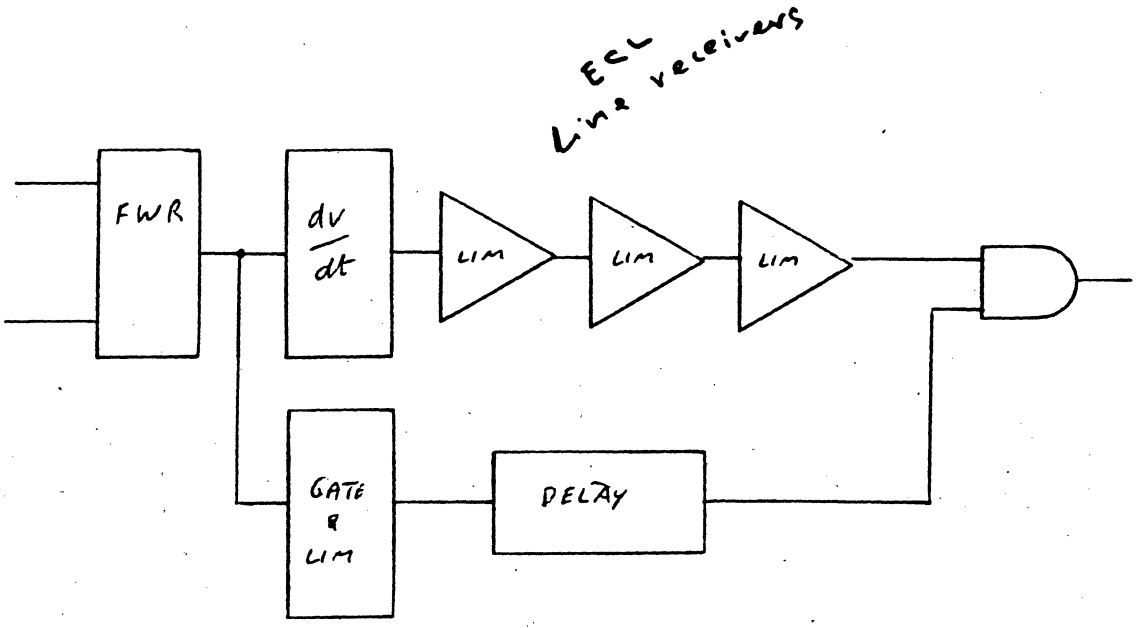


FIG 9-18 BLOCK DIAGRAM SHOWING DELAY USED TO CENTER LEADING EDGE OF DATA IN GATE (AT MINIMUM INPUT AMPLITUDE)

Notice the timing requirement to center the propagation delayed differentiated pulse in the Gate square wave. If the Gate generator used an equal number of stages of limiting gain then the delays should be about equal. This becomes more important when we consider that the input signal actually varies in amplitude as a function of disc diameter (in disc machines) or media-coating variations. This variation must be considered when calculating the total gain required for both limiters. The total gain must be calculated based on the absolute minimum signal from the head as modified by any intervening gain stages using their minimum gain.

Let us go thru a design using the ECL logic family as our output. This is chosen due to the levels and speed but particularly the non requirement for cascading bias.

Going back to our circuit of Figure 9.12, we can make a small change to make it compatible with the ECL family. See Figure 9.19. The change required is in the base bias network for transistors 1 and 2 with regard to the bias required on the base of transistor 8 and the improvement required to reduce the effect of power supply variations on the clipping level by using the diode D as a partial regulator. Notice also that the Differentiator does not care about the use of the clipping bias as it is AC coupled and will not be affected by the difference in DC potential as does the Gate Generator. We will choose the bias values such that the input signal swing will not permit saturation of the Differentiator and Gate Generator collector stages.
If the input signal maximum is 10.Vpp differential, the nominal is 7.0Vpp differential and the minimum is 5.Vpp differential we then know that the Full Wave Rectified signal is 2.5 V_{BP}SE max and 1.25 V_{BP-SE} min. We will

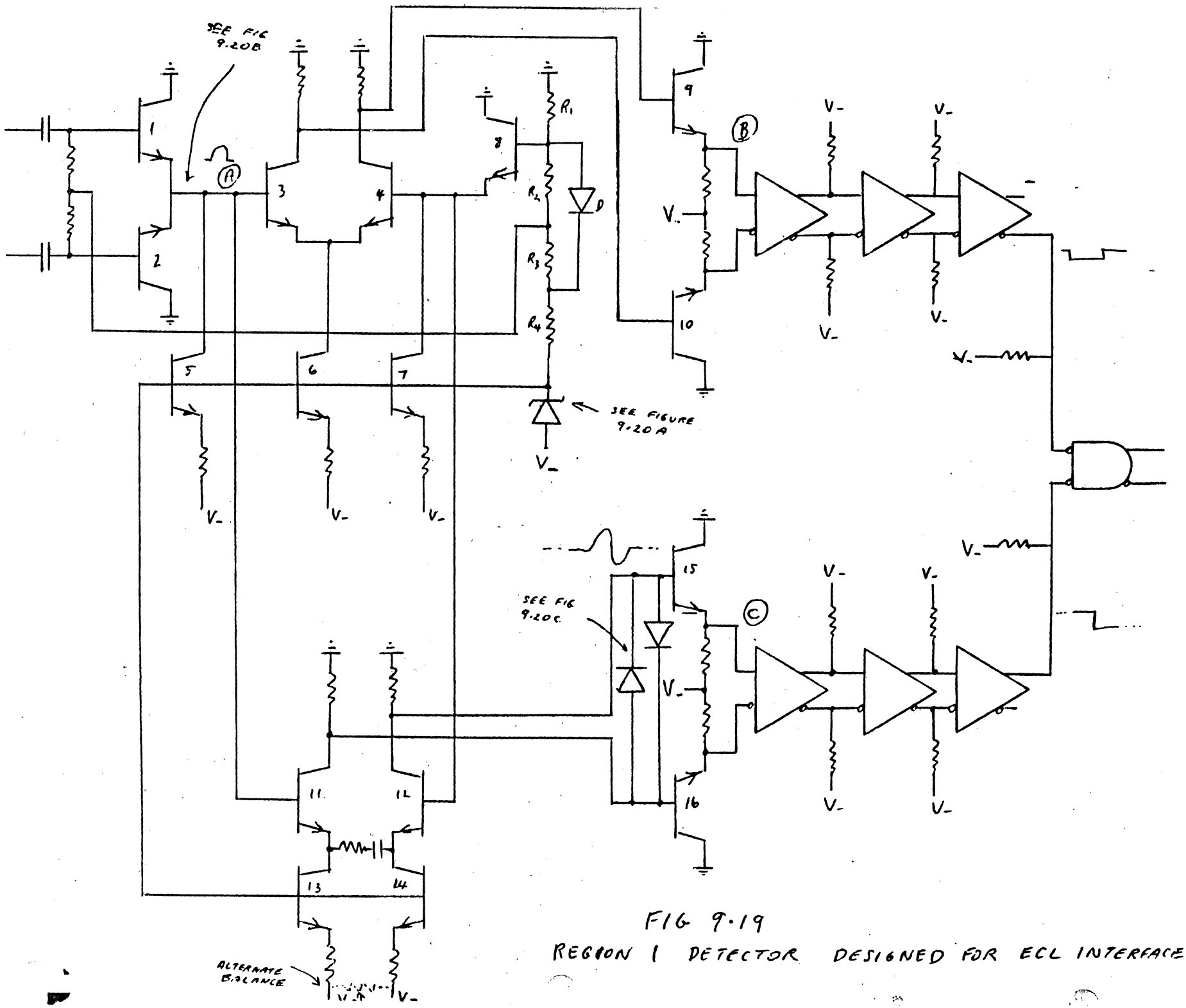


FIG 9.19
REGION I DETECTOR DESIGNED FOR ECL INTERFACE

retain our 15% clipping level. The bias for Transistors 1 and 2 should be:

$$V_{C_1\,MIN} + V_{BP\text{-}SE\,MAX} = -1.2V - 2.5V = 3.70V \qquad \text{(EQ 9.6)}$$

Choose -4.3V to allow for tolerances of 5%. The bias for Transistor 8 should be:

$$-4.3V + 15\%\left(\frac{7V_{PP\,diff}}{4}\right) = -4.562V \qquad \text{(EQ 9.7)}$$

If we do this then we have built in a one diode drop margin for the collectors 3 and 4. If we choose the resistor divider network to operate off -10 volts and have current of 10.ma to maintain some stability due to base currents then:

$$\frac{15.V_{NOM} - V_{D\,NOM}}{10.ma} = R_1 + R_4 = 1.43K \qquad \text{(EQ 9.8)}$$

$$R_1 = \frac{4.3}{10.ma} = 430\Omega \qquad \text{(choose } 432\Omega\ 1\%\text{)} \qquad \text{(EQ 9.9)}$$

$$R_4 = 1.43K - 430\Omega = 1.0K \qquad \text{(choose } 1.00K\ 1\%\text{)} \qquad \text{(EQ 9.10)}$$

If we allow 5 ma for the diode, a IN4448, we should have approximately 0.7V drop. This leaves 5.ma for R_2 and R_3 to develop 0.375 volts.

$$R_2 = \frac{0.262\,V}{5.ma} = 52\Omega \qquad \text{(choose } 51.1\Omega\ 1\%\text{)} \qquad \text{(EQ 9.11)}$$

$$R_3 = \frac{0.7 - 0.262}{5.ma} = 87\Omega \qquad \text{(choose } 86.6\Omega\ 1\%\text{)} \qquad \text{(EQ 9.12)}$$

DETECTORS

In order to minimize the disturbance to the bias due to base current variations in the operation of the Full Wave Rectifier, we should replace R_1 with a 4.3V 5% zener. This way all margins are met while maintaining a low impedance at the bases of transistors 1, 2 and 8. We now need to lower R_4 to allow 20.ma zener current as well as allowing a higher current thru the diode (10.ma). We can also add a zener at the bottom to control the current sources (see figure 9.20A).

$$R_{4_{REVISED}} = \frac{V_- - V_Z - V_\gamma}{I_Z} = \frac{15.V - 2(4.3) - 0.75V}{20.\ ma} = 282.5\Omega \quad (use\ 280\Omega\ 1\%) \qquad \text{(EQ 9.13)}$$

The value of Resistors R2 and R3 should be halved to accommodate 10.ma instead of 5.ma. Let R2 = 24.9Ω and R3 = 43.2Ω

If we choose the CA 3045 transistor array we will obtain an added bonus of Vbe matching to 5.mv which will help. The 20.V breakdown between substrate and collector junction is adequate, as is the 15V collector to emitter breakdown. The base emitter breakdown needs examining for transistors 1 and 2 as they will see the full swing of the input during rectification. The maximum input difference is 5.0 Vpp SE or just equal to the minimum specified BV_{be} for the devise. This is close but is acceptable. If the max input swing were larger we wou ld either have to find a transistor with a higher BV_{be} or use series diodes for rectification preceded by normal emitter followers in order to minimize distortion. We would also have to refigure the bias on the Gate Generator. Either way we are faced with a matching problem in order to minimize the unbalance of the rectifier halves. See Figure 9.20B for the variation .

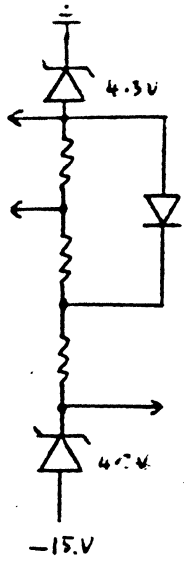


FIG 9.20 A

BIAS SUBSTITUTE IN FIG 9.19

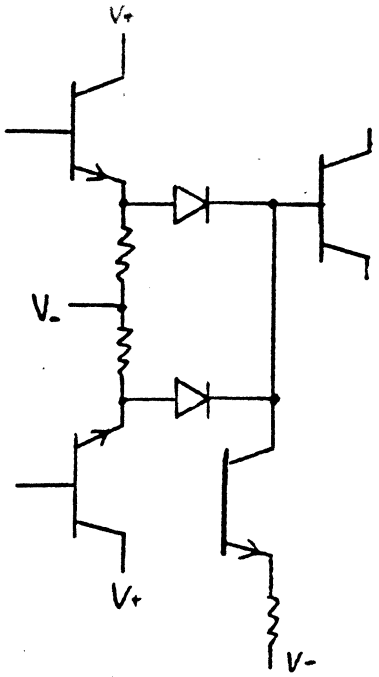


FIG 9.20 B

F.W. RECTIFIER SUBSTITUTE FOR HIGH LEVEL SIGNALS

$V_s > 2BV_{ebR}$

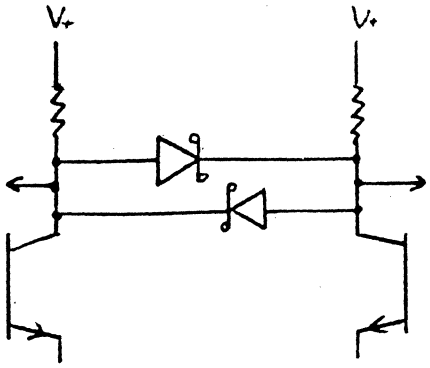


FIG 9.20 C

LIMITING SUBSTITUTE TO PROVIDE HIGHER GAIN AT BASE LINE CROSSINGS YET NOT SATURATE THE COLLECTORS.

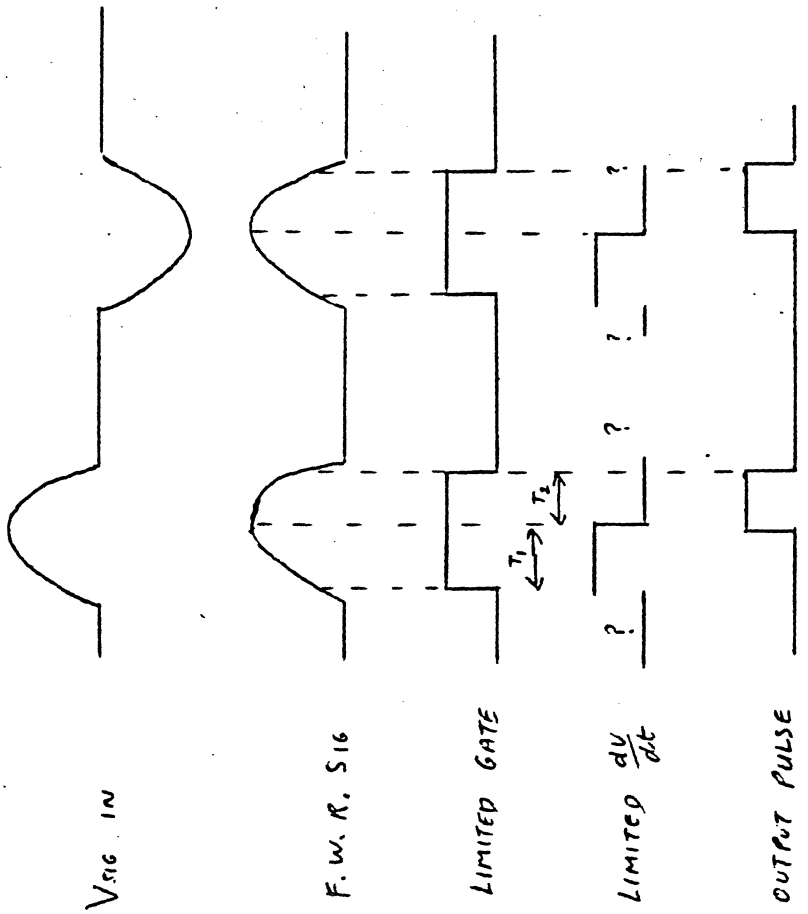


FIG 9.21

WAVE FORMS FOR A REGION I DETECTOR UNCERTAIN AREAS OF THE DIFFERENTIATED LIMITED SIGNAL ARE OUTSIDE THE AREA OF THE GATE

(ADJUST TO MAKE $T_1 = T_2$ TO CENTER BIT NEGATIVE EDGE IN GATE)

Continuing with our design. We will choose the current for the emitter follower rectifiers based on the stray capacitance on the node A.

$$C_T = 2C_{eb} + C_{ob}(2+A) + C_{cs} + C_w$$
$$= 2(0.6pF) + (0.58pF)(2+12) + 2.8pF + 10.pF = 22.12pF \quad \text{(EQ 9.14)}$$

For a 5.0 MHz sinewave input the required current must exceed

$$\frac{(2.5V_{BP-SE})(22.12\,pF)}{\frac{1}{2\pi(5\times10^6\,Hz)}} = 1.73\times10^{-3}\,amps \quad \text{(EQ 9.15)}$$

This should operate well if we choose 3.0 ma to allow for rectifier linearity, capacitance and worse case variations.

The current source $_{5,7}$ then becomes

$$R_{E_{5,7}} = \frac{V_{Z\,min} - V_{be\,max}}{3.0\times10^{-3}\,a} = \frac{4.085V - 0.75V}{3.0\times10^{-3}\,a} = 1.11\times10^3\,\Omega \quad \text{(choose 1.1K } \Omega\text{)} \quad \text{(EQ 9.16)}$$

The collector resistors for both the Gate and Differentiator should be related to the Bandwidth. For our 5.MHz signal we need a minimum of 50.MHz Bandwidth in order to minimize phase distortion and amplitude reduction.

$$R_L = \frac{1}{2\pi F C_T} = \frac{1}{(2\pi)(5\times10^7)(1.092\times10^{-11})} = 2.914\times10^2\,\Omega \quad \text{(EQ 9.17)}$$

Where $C_T = (2+A(C_{ob}) + C_{cs} = (14)(.58pF) + 2.8pF = 10.92pF$

If we choose 200 Ω we can obtain a 1.2 volt swing with 6.ma $I_{s(6)}$.

The gain becomes:

$$A = \frac{R_L}{\frac{26}{I_{ma}} + R_m} = \frac{200}{\frac{26}{3} + 5} = 14.63\ \Omega \qquad \text{(EQ 9.18)}$$

This gain is slightly higher than the 12 we assumed in equations 9.14 and 9.17, but we have plenty of margin in both cases.

The current source resistor should be:

$$R_1 = \frac{V_{2min} - V_{be\ max}}{6.0 \times 10^{-3}\ a} = \frac{4.085\ v - 0.77\ V}{6.0 \times 10^{-3}} = 552.5\ \Omega \quad (\text{use } 549\ \Omega\ 1\%) \qquad \text{(EQ 9.19)}$$

Next we can choose an emitter follower resistor 9, 10, 15, 16 to drive the following limiter stages if we allow 3 ma again we could use

$$\frac{V_{-min} - V_{be\ max} - V_{D\ max}}{3\ ma} = \frac{14.25\ V - 0.75\ V - \frac{1.2\ V}{2}}{3 \times 10^{-3}\ a} = 4.30\ K \quad (\text{choose } 4.32\ K\ 1\%)$$

(EQ 9.20)

The design of the differentiator is next. Going back to Equation 9.2 we would like to place the pole at an order of magnitude above the highest frequency of operation F_H. (We, may modify this later when we consider noise degradation. (The Rule of Thumb usually makes the phase angle equal to 70^0 at F_H).

We would also like the gain at F_H to be such that the collector signal is linear and centered around -1.2V in order to drive the following amplifier. With a 2.5V peak rectified input signal we would require a gain loss to guarantee a linear swing. We can solve the dilemma by providing clipping at the collectors to reduce the amplitude swing while retaining the slope around the base line. A pair of back-to-back Schotky diodes (IN 5711) will function

well as they have no storage time constraints. (Figure 9.20C) Linearity in the emitter circuit is maintained by making the impedance and current such that neither transistor is cut off. For our 200Ω collector resistors, two 6.0 ma current sources will give us a -1.2V collector voltage on each. The emitter circuit must have an impedance at 5.0 MHz of greater than

$$Z_{min} = \frac{V_{SIG\ BP\text{-}SE\ MAX}}{6.0\ ma} = \frac{2.5V}{6.0\times10^{-3}} = 433\Omega\ MIN \qquad \text{(EQ 9.21)}$$

If Xc is greater than 433Ω then that will satisfy the requirement.

$$C = \frac{1}{2\pi F_H X_C} = \frac{1}{(2\pi)(5\times10^6)(4.33\times10^2)} = 7.35\times10^{-11}\ Fds \qquad EQ\ 9.22$$

If we chose 62 PF then we should cover the worse case capacitor value since we calculated the minimum current source current at 6.ma.

The gain at 5.0 MHz becomes . if RE = 0

$$A = \frac{2R_L}{2(r_e + R_m) + R_E - j\frac{1}{2\pi F_H C}} = \frac{2(200)}{2\left(\frac{26}{6} + 5\right) - j\frac{1}{(2\pi)(5\times10^6)(6.2\times10^{-11})}} \qquad \text{(EQ 9.23)}$$

$$= \frac{400}{513.74\angle -87.9^\circ} = 0.778\angle 87.9^\circ$$

The pole is located at (EQ 9.24)

$$\frac{1}{2\pi R_T C} = 1.37\times10^8\ Hz \quad \text{OR } 137\ MHz$$

We could calculate the degradation in phase due to the real transistor parameters using the hybrid π model, but the frequencies are higher than we will be interested in (above 50 MHz).

We can add current balance by using the potentiometer as part of the emitter source resistors. The remaining gain for both limiters needs to be calculated. For a minimum input signal of 5.Vpp Diff we have 1.25V $_{BP\text{-}SE}$ out of the rectifier to make this signal be a 1.V square wave with rise

and fall times equal to the logic families characteristics of 1.1 ns then at 5.0 MHz 1.1 ns represents

$$(5\times10^{6}\ \text{Hz})(1.1\times10^{-9}\ \text{sec})(360^{0}) = 1.98^{0}$$

Therefore

$$A_{req} = \frac{1.0V}{(1.25V)(\sin 1.98^{0})} = 23.4 \qquad \text{(EQ 9.25)}$$

This indicates that the gain already provided with transistors 3 and 4 is not sufficient. Since we also need to match the propagation delay of both channels and the gain of the differentiator is only .778, then we need to have the differentiator channel gain equal to

$$A_{CHAN} = \frac{1.0\,v}{(0.778)(1.25\,v)\sin 1.98^{\circ}} = 29.76 \qquad \text{(EQ 9.26)}$$

We need a gain of greater than 29.76 to achieve the same accuracy. Since this gain cannot be achieved in one stage using a MC10116 line receiver as an amplifier-limiter then two will be used. This is because $\sqrt{29.76} < A(10116)_{MIN}$

The final design needs only two series MC 10116 line receivers for each channel. The output 'and' gate can be a MC 10105 positive 'or-nor' gate. Notice that this design is based entirely on a percentage of the nominal signal therefore amplitude plays a dominant role in the detection process. If signal amplitude is lost then data is lost.

The waveforms should be reviewed and are shown in Figure 9.21. The output pulse width is not controlled and may be equal to any value between

1.1 ns and 50.ns depending on the input signal amplitude. If the differentiated pulse is not centered in the gate then noise can occur at the edges of the gate due to the noise of the differentiated signal. The clipping level may be adjusted to a different percentage of the nominal signal to change the width of the nominal gate. To properly characterize the circuit a plot of pulse centering as a function of amplitude should be made, also as a function of frequency at certain fixed amplitudes.

This completes the design of a detector that can handle signals in the left hand portion, Region 1, of the bit density curve.

There are a large number of circuits that could perform this function depending on the availability of components and integrated circuits. In putting together the circuits for each block, all we need to consider are the various interfaces, their voltage, current and timing requirements.

Now what of the other areas of the bit density curve? We can evaluate them by looking at the waveforms in each area for a string of random data.

In Region 2 the signal is modified due to the pulse interaction at the higher densities. Resolutions can go down to around 70%, meaning that for three transitions in a row, but isolated on either side, the center pulse barely crosses the base line. This is illustrated in Figure 9.22. Similarly, we show the triple pulse waveform for each of the four Regions. If we were to use the detector we have just designed for Region 2 signals, the clipping level would have to be lowered to such an extent that the base line noise would pass thru. Or in other words, the minimum amplitude of the center pulse is less than the noise. In terms of clipping level percentages in Region I we could

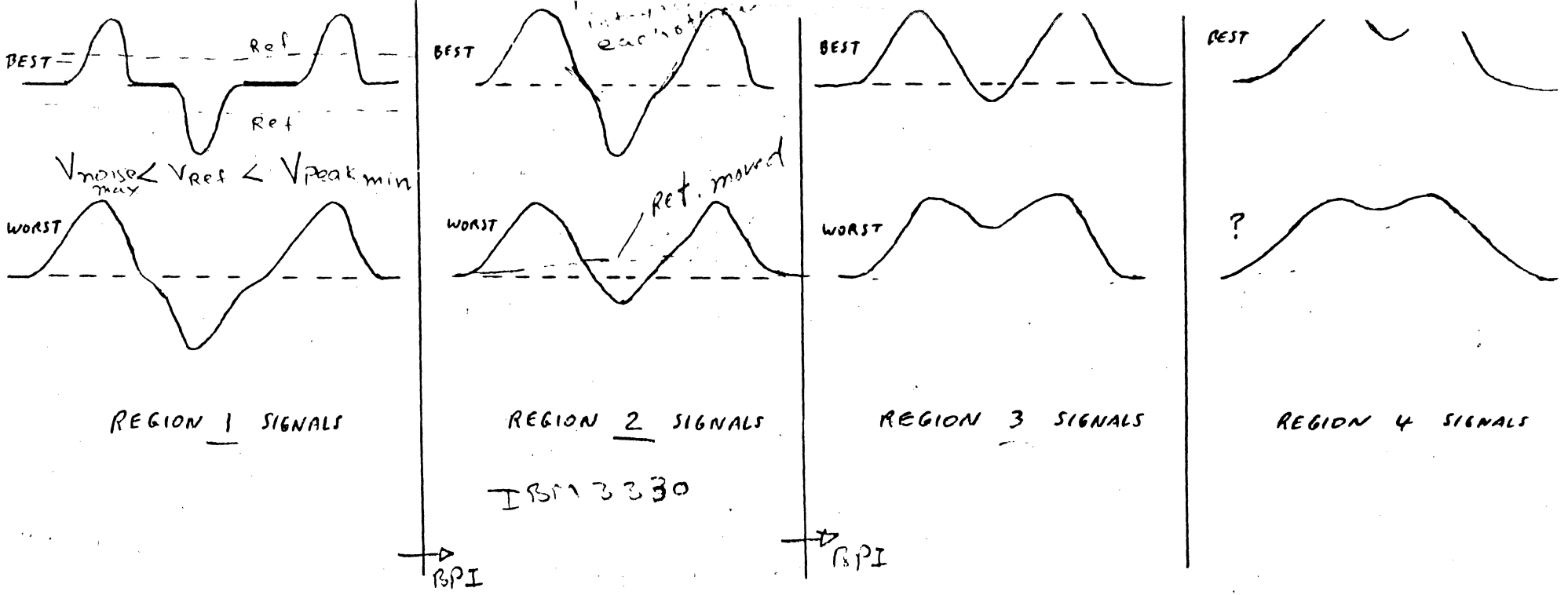


FIG. 9.22

HEAD RESPONSE TO TRIPPLES IN VARIOUS REGIONS

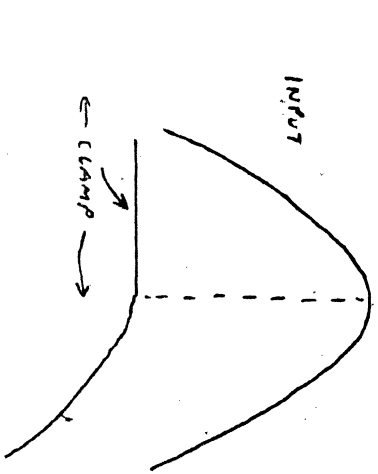
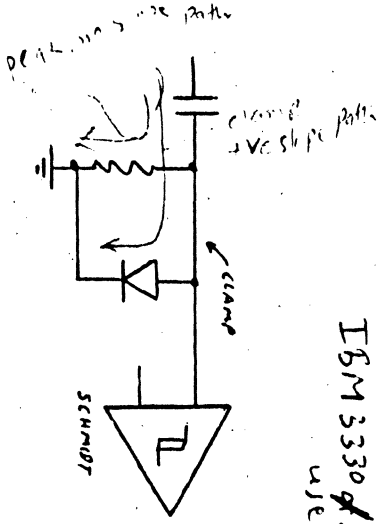


FIG 9.23 B

PEAK REFERENCING AMPLITUDE SENSE CIRCUIT.



+ CLAMPING. FOLLOWED BY A - AMPLITUDE SENSE. OR ΔV

IBM 3330 & 2350 use

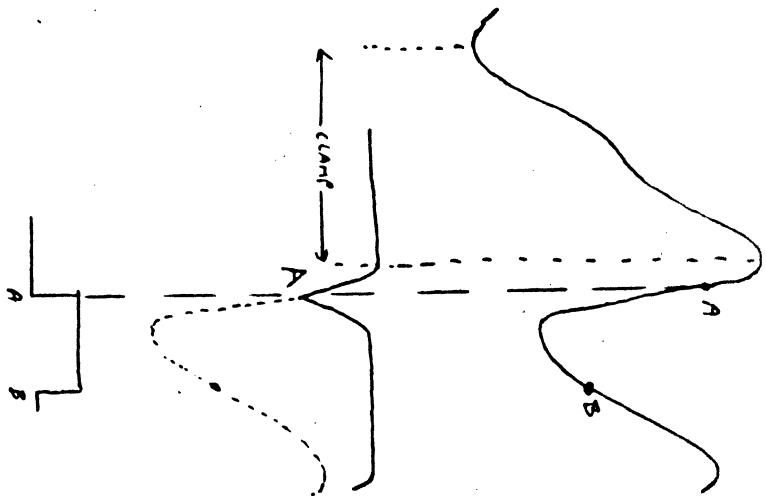


FIG 9.23 A

have a clipping level range of from 50 to 15% or 35% where 50% represents the value where we just lose a pulse and 15% of the level where we just pick up the noise. In Region 2 this range becomes negative; in other words, the clipping level value for loosing the center pulse is below the value to pick up noise.

Since we must operate some detectors in these other regions, we will discuss methods for accomplishing this. The first clue comes from the way we visually determine the position of a transition pulse. Each pulse has some leading slope, some zero slope peak, a trailing slope of the opposite polarity, and some peak-to-peak amplitude difference. In Region I the amplitude is always in reference to the base line therefore some fixed or moving reference around the base line will suffice. Here in Region 2, or worse, the base line becomes a moving target depending on the transition pattern. We can take two approaches. The first is where each signal peak is clamped to a reference and the amplitude is measured opening a gate if the signal exceeds some delta. Clamping is achieved by a diode as shown in Figure 9.23. As the input signal approaches its positive peak, the capacitive current is shunted to ground thru the diode then as we pass the peak this current reverses, the diode reverse biases and the RC time constant is restored. This allows the negative slope of the signal to be measured by the comparator. This voltage is chosen to represent some percentage of the signal. As can be seen there will be a time delay before the gate is opened following the true peak or transition center. This must be allowed for in the design. Further problems are the forward bias of the diode and the capacitance on that node. The main problem though is the non-uniformity of the pulse amplitudes. If the second positive pulse were of a lessor amplitude than the first then the circuit will not clamp to the second peak. All of these problems are addresssed in the design that follows.

The choice of the coupling capacitor is dictated by the stray capacitance f the following network. If C is very large compared to the stray capacitance ıen we have control of the signal transfer. Secondly, if the diode drop is mall compared to the signal peak-peak amplitude, we retain control. Also, f the diode has very short minority carrier life time, this also helps educe the switching time between the forward conducting clamp and the peak ollowing long RC time constant. Another problem is the variation in amplitude. his can be solved by taking advantage of the alternating nature of all read signals polarities. If we provide a deliberate positive current during the periods of time immediately following the comparator sensing level, we can ensure that when the waveform again changes to a positive slope there will always be sufficient charge to clamp during the positive sloped portion of the signal. This is illustrated in Figure 9.24 where the point B represents the point where the second half comparator operates and point A the first sensing point. All this clearly shows that two such circuits are necessary to cover both positive and negative (positive on the opposite half of a differential signal).

Consider our 5.MHz H_F signal we used earlier. The diode should be a hot carrier type to eliminate the minority carrier lifetime and thus speed up the reverse recovery of the diode. If the forward drop is 0.4 volts then our signal should be 10 to 20 times this value or 4-8 V_{pp} min. Let us choose 10.$V_{pp\ diff}$ as an amplitude requirement nominal. If the stray capacitance is around 10 PF then the coupling capacitor should be large or 20 - 100 times that value or lets use 1000.PF. The value of R should be such that it is large compared to X_C at the lowest sinusoidal frequence (F_L) of interest. If that were 2.5 MHz then let

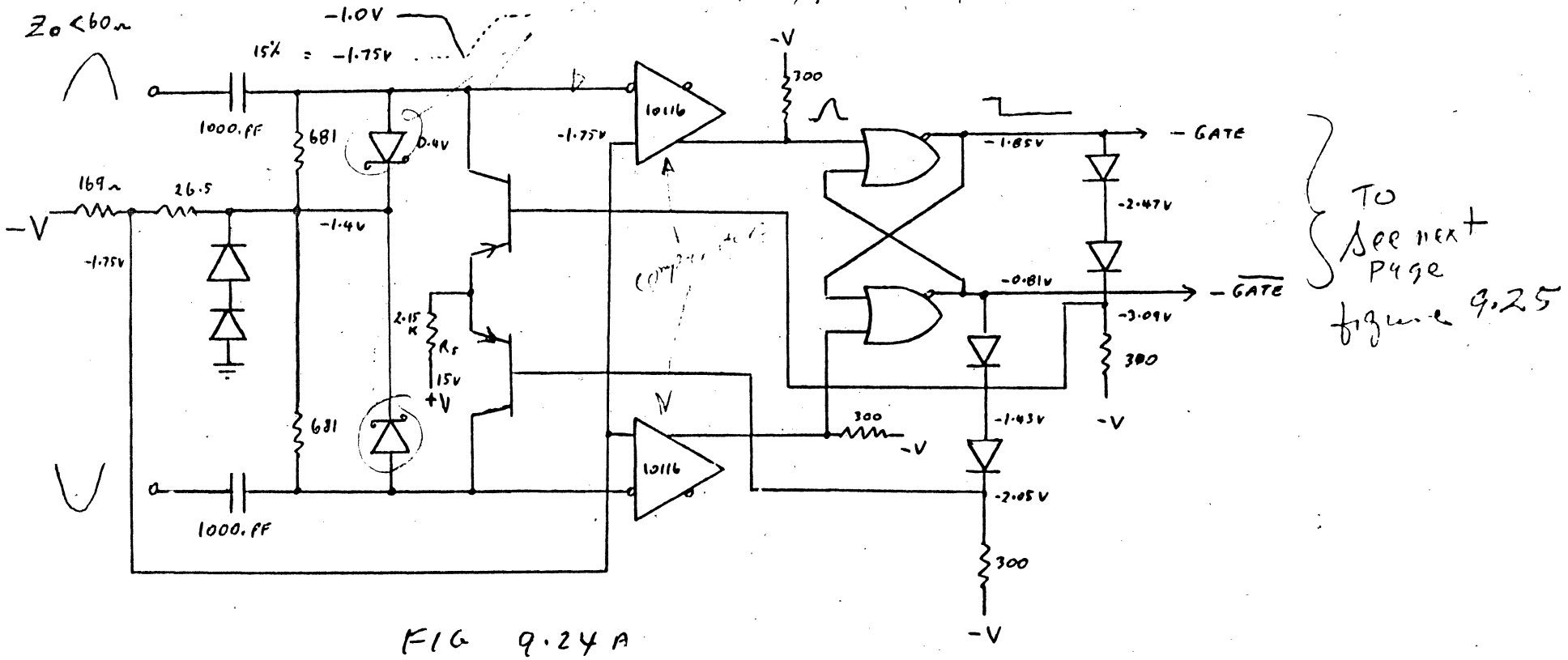


FIG 9.24 A

Δ V GATE GENERATOR CIRCUIT

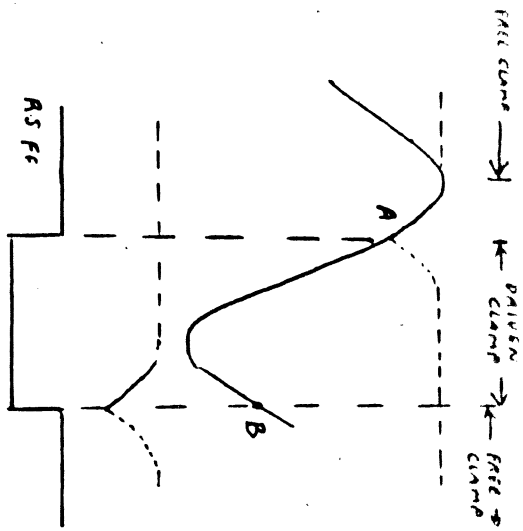


FIG 9.24 B

Δ V GATE GENERATOR WAVEFORMS SHOWING FREE AND DRIVEN CLAMP ACTIONS

$$R = 10\,X_c = \frac{10}{2\pi F_L C} = \frac{10}{(2\pi)(2.5\times10^{6}\,Hz)(10^{-9}\,F)} = 6.366\times10^{2}\,\Omega \qquad \text{(EQ 9.27)}$$

Let us choose 681Ω as a standard value. This forces the driving impedance to be less than 60Ω which we can obtain by an emitter follower.

Because the signal at the comparator input is less than a few volts, it is well within the capability of an ECL line receiver. The basic design is shown in Figure 9.24. If we choose a comparator reference of 15% of the $5.0V_{pp\text{-}SE}$ signal nominal then the comparator should be set to

$$V_{Ref} = -2V_D + V_{HC} - (0.15)(5.0V) = -1.4V + 0.4V - 0.75V = -1.75V \qquad \text{(EQ 9.28)}$$

In order to try to keep this value close to the 2 diode reference, which we provided in order to keep the comparator signal within the range of the line receiver, we should make them part of the comparator reference. If we used a simple resistive divider as part of the negative supply to the diodes, then we maintain control and the comparator reference becomes relatively insensitive to the diode drop variations. The PNP switch should supply a current to the clamp equal to that expected from the signal. THIS IS SIMPLY —

$$I = \frac{V_{SIG\ pp\text{-}SE}}{R} = \frac{5.0V}{681\Omega} = 7.34\ ma \qquad \text{(EQ 9.29)}$$

Let us choose 10 ma. The bases of the PNP switch must be driven from a level such that the collectors do not saturate. The standard ECL levels are -0.81 $_{max\ H1}$ and -1.85 min low. Two diode drops will guarantee a voltage more negative than the collectors. The IN4448 diode shows a minimum drop

of 0.62 Volts at 5.ma; therefore, the two extremes will be (worse case most positive) -

$$V_{bH} = -0.81V - 2(0.62V) = -2.05V \qquad \text{PNP OFF} \qquad \text{(EQ 9.30)}$$

$$V_{bL} = -1.85V - 2(0.62V) = -3.09V \qquad \text{PNP ON} \qquad \text{(EQ 9.31)}$$

Now all we need to do is guarantee a minimum of 5.ma thru the diode. Using a -5.2V ± 5% power supply, the resistor is calculated from

$$\frac{(V_s - 5\%) - V_{bL}}{5.0\times10^{-3}A} = \frac{-4.94 + 3.09V}{5\times10^{-3}A} = 370\Omega \text{ MAX} \qquad \text{(EQ 9.32)}$$

If we used 300Ω, we have plenty without exceeding the current from the logic.

Now we know the current supplied from the PNP switch was chosen at 10.ma. The emitter resistor can be calculated as a nominal or we can ensure worse case that we have our 7.34 ma -- let us do the latter.

$$R_s = \frac{V_{+min} + V_{b\,max\,on} - V_{be\,max}}{7.34\times10^{-3}} = \frac{(15-0.75) - 3.09V - 0.8V}{7.34\times10^{-3}} \qquad \text{(EQ 9.33)}$$

$$= 2.253\times10^{3}\Omega \quad \text{(choose 2.15K 1%)}$$

We can now calculate the current required to maintain the two diode -1.4V reference as the PNP switch current subtracts. Using a -5.2V supply and a 10.ma residual current for the diodes we get

$$10.ma + \frac{V_{+max} + V_{b\,max\,on} - V_{be\,min}}{R_{5\,min}} = \frac{15.75V + 3.85V - 0.65V}{2.128\times10^{3}} \qquad \text{(EQ 9.34)}$$

$$= 18.9\,ma$$

Therefore the resistor total required from -5.2V is

$$\frac{V_{-min} - 2(V_{b\,max})}{18.9\,ma} = \frac{4.94V - 2(0.62V)}{1.89\times10^{-2}A} = 195.7\Omega \qquad \text{(EQ 9.35)}$$

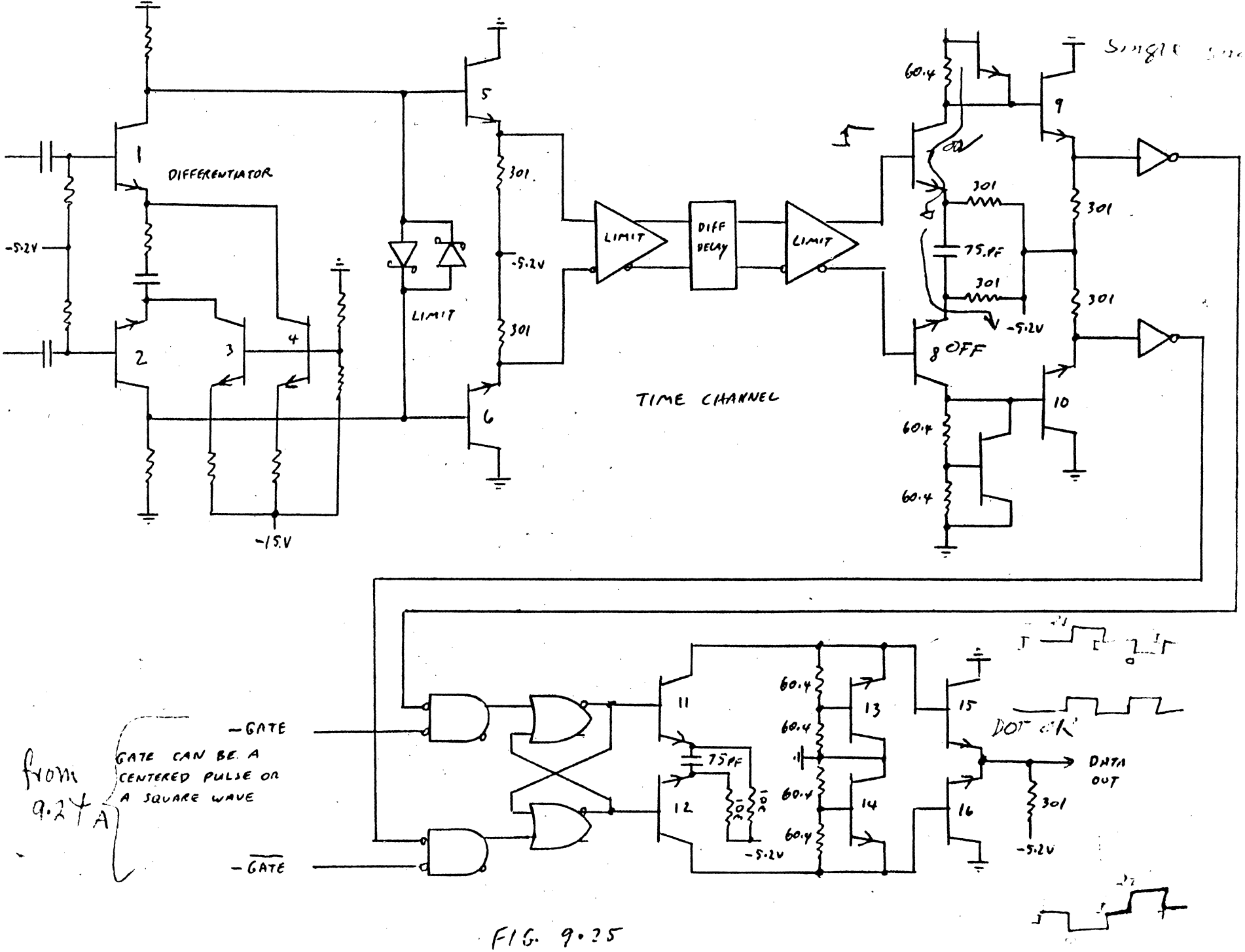


FIG. 9.25

TIME CHANNEL FOR REGION 2 DETECTOR

To get our reference for the comparators we need to find

$$\frac{-1.75V_{REF} + 2(0.624)}{1.89\times10^{-2}\,\Omega} = \frac{0.51}{1.84\times10^{-2}} = 26.98\,\Omega \quad (\text{choose } 26.5\,\Omega) \qquad \text{(EQ 9.36)}$$

This leaves

195.7 - 26.5 Ω = 169.2 remaining (choose 169 Ω 1%)

Now we have a squarewave resulting from the amplitude sensing of the negative slope following a peak. We need to 'and' this with a delayed pulse representing the absolute peak. But now we need a pulse representing each peak separated as to polarity one for each side. Note also that the Gate is now the total width of the transition spacing instead of just a pulse which goes away after the amplitude is lost. We need a circuit that accepts the first pulse and ignores any following until the change in polarity is sensed.

We will now design the rest of the detector to go with the Gate Generator we just designed. The differentiator is similar to the one we designed for the Region 1 detector only we will drive the two bases differentially instead of from a rectified signal. This way we retain polarity sensing. If we choose the base bias of -5.volts then the entire design can be repeated except for a stage that follows the limiting amplifiers. Notice our preoccupation with ECL compatibility. The differentiator outputs an edge for each zero crossing with undetermined edges in between due to the signal returning to the noisy base line of our example in Figure 9.22. This edge is unipolar for each peak of the same polarity thus we can separate the pulses. A circuit for generating these pulses is called a split Bidirectional Single Shot. Although

true Single Shot the input timing makes it behave that way. It takes age of a single capacitor in the emitter feedback path just as in our differentiator only here the transistors are either conducting or f which makes it an overdriven differentiator. It functions by forcing nitter current source to flow ONTO the capacitor until its charge es such that the transitor bias changes to a forward bias. Figure 9.26 the waveforms. The collector 7 current changes from the source value ro on cut off while the capacitor is changing its charge. When the ter 7 voltage falls to the value necessary to turn the transitor 7 back ncreasing the collector 7 current back to the source value. When the site edge occurs the current which normally flowed thru the transistor w flows thru transistor 7 as well as its own source current, thus doubling collector 7 current until transistor 8's bias allows it to turn back on. ause of the double collector current, a pair of clamp transistors 9 and 10 added in order to keep the collector 7, 8 out of saturation.

The design follows the ideas presented in the waveforms. Let us choose current source of 10.ma in order to maintain circuit speed.

$$R_{s\,max} = \frac{V_{-min} - V_{Logic\ H\ min} - V_{be\ max}}{10.0\ ma} = \frac{4.94V - 0.98V - 0.85V}{10^{-2}} \quad \text{(EQ 9.37)}$$

$$= 311\Omega \qquad (\text{choose } 301\Omega\ 1\%)$$

we want a minimum pulse width of 50.ns, then the capacitor should be larger n

$$C = \frac{(I_{min})(T_{min})}{2 \cdot \Delta V_{b\ min\ logic}} = \frac{(1\times10^{-3}a)(5\times10^{-8}sec)}{1.65V - 0.98V} = 7.46\times10^{-11}\ Fds \quad \text{(EQ 9.38)}$$

We should choose 75 to 82 pF since the pulse time is slightly altered by the transistor current as the off transistor starts to conduct. The collector load resistor is chosen to give greater than an ECL logic level change max of 1.85-0.81 volts or 1.04V; let us choose 1.2 Volts at 10 ma. The two resistors associated with the clamp are set to develop 0.6V across the emitter-base resistor and 0.6 volts across the base-collector resistor at 10 ma; therefore, they should be 60Ω each (choose 60.4Ω).

The remainder of the circuit is built using ECL blocks. The circuit for ignoring subsequent same polarity pulses is simply an RS latch followed by a Bi-directional Single Shot. The only difference to the design from the Split Bidirectional Single Shot we just designed is when we tie the two emitter followers together thus performing the positive dot or function. We will use 300Ω resistors for the emitter return resistors as we calculated before. The delay line should be inserted in the limited differentiated signal path preferably between the two amplifiers in order to preserve as much symmetry as possible. A differential delay line is preferable.

Now that was a lot of circuitry but we had to perform the functions required. To summarize, we needed a Gate Generator capable of operating on peak-to-peak differences instead of a base line related reference. The circuit chosen introduced an amplitude dependent delay and a bidirectional gate equal to the timing between transitions plus or minus some error. This forced the pulses resulting from the limited differentiated signal to require a delay and to have separated positive and negative peak sensed pulses. We solved both these problems with a differential delay line which maintained most of the symmetry. (If a single ended delay line is used then the symmetry can be recovered by careful adjustment of the bias of the following differential amplifier. Note

ı of 2 is lost by going single ended and back to differential.)
ted pulses was obtained with a unique circuit called a Split
nal Single Shot. We gated these two polarity-determined-pulses
thru an 'AND' circuit. The noise related pulses that might follow
mate 'peak' pulse were ignored by setting and resetting an R.S.
ı the first pulse thru the gate. (Subsequent pulses do nothing.)
is now in square wave form, an edge for EACH transition, just like the
rent was which was used to write the data. This was converted back
to Zero Pulses by the use of another Bidirectional Single shot
time with a positive dot 'or' output. When testing this circuit
must be adjusted to permit all acceptable signal amplitudes to
hout altering the time of the peak sensed pulse. Alteration may occur
eak sensed pulse split the edge of the gate. The comparator sense
y be lowered if the noise related to a base line is absent. This may
the low frequency signal is high enough to keep the entire bit
within the poor resolution area of the Bit Density Curve. The above
ctly a function of the code used and will be discussed in a later

second Region 2 detector can be built using a simpler Gate Generator
code guarantees that the signal will not return to the base line.
of the circuitry essentially does not change. We could say that a
detector is the most complex of all the possible detectors. The
on is simply to alter the signal by passing it thru a lead network,
ll differentiator. The reason is that there exists in Region 2 a
shoulder on the lowest frequency signal which, if differentiated,
a droop. It is this droop that is noise sensitive. If noise enters

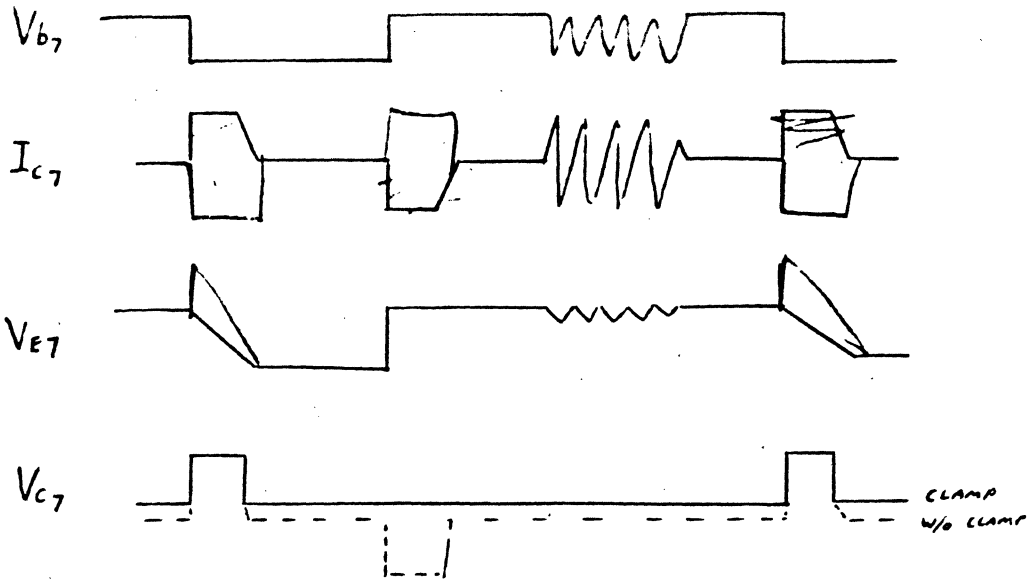


FIG 9.26

OPERATIONAL WAVEFORMS FOR THE SPLIT BIDIRECTIONAL SINGLE SHOT OF FIG 9.25

$$C = \frac{I_{r \cdot c} \cdot t_{r \cdot c}}{2 \cdot \Delta V_b}$$

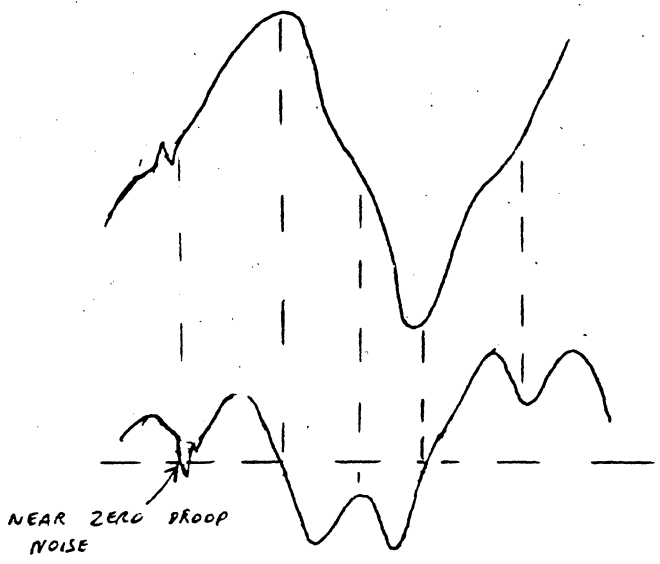


FIG 9.27

SHOULDER CAUSED DROOP, NOISE CAUSES DROOP TO CROSS BASE LINE

the signal, and it does, the slope of the shoulder is changed either to zero or oppositely such that a zero crossing is obtained which results in an erroneous output pulse. It is this shoulder caused droop (Figure 9.27) that forces code related bandwidth limited Region 2 signals to still require a Gate Generator. Such is the case with the industry wide MFM code when used in this region. If we use a lead network instead of a SECOND FULL differentiator then the droop is reduced and the output can be limited to create a polarity related gate. We still require the FIRST differentiator because we need the precise time of the true peak. A lead network would distort the pulse timing. The circuit is shown in Figure 9.28.

The shoulder can be shown to contain considerable 3rd Harmonic. The lead network need only attenuate the third harmonic to achieve the desired result. Let us design for a F_L of 2.5 MHz. We want to reduce the differentiated 3rd harmonic by 6 db more than the F_L signal. This, of course, depends on the amount of shouldering we have on the lowest density signal. The pole associated with the network can be established from

$$R_E - jX_c = 200 \angle \theta \qquad \text{(EQ 9.39)}$$

$$R_E - j\frac{X_c}{3} = \frac{2}{3} 200 \angle \alpha \qquad \text{(EQ 9.40)}$$

The 3rd harmonic gain of the differentiator is 3 times the gain at the fundamental; therefore, we want half of that to get our 6 db loss at the 3rd harmonic.

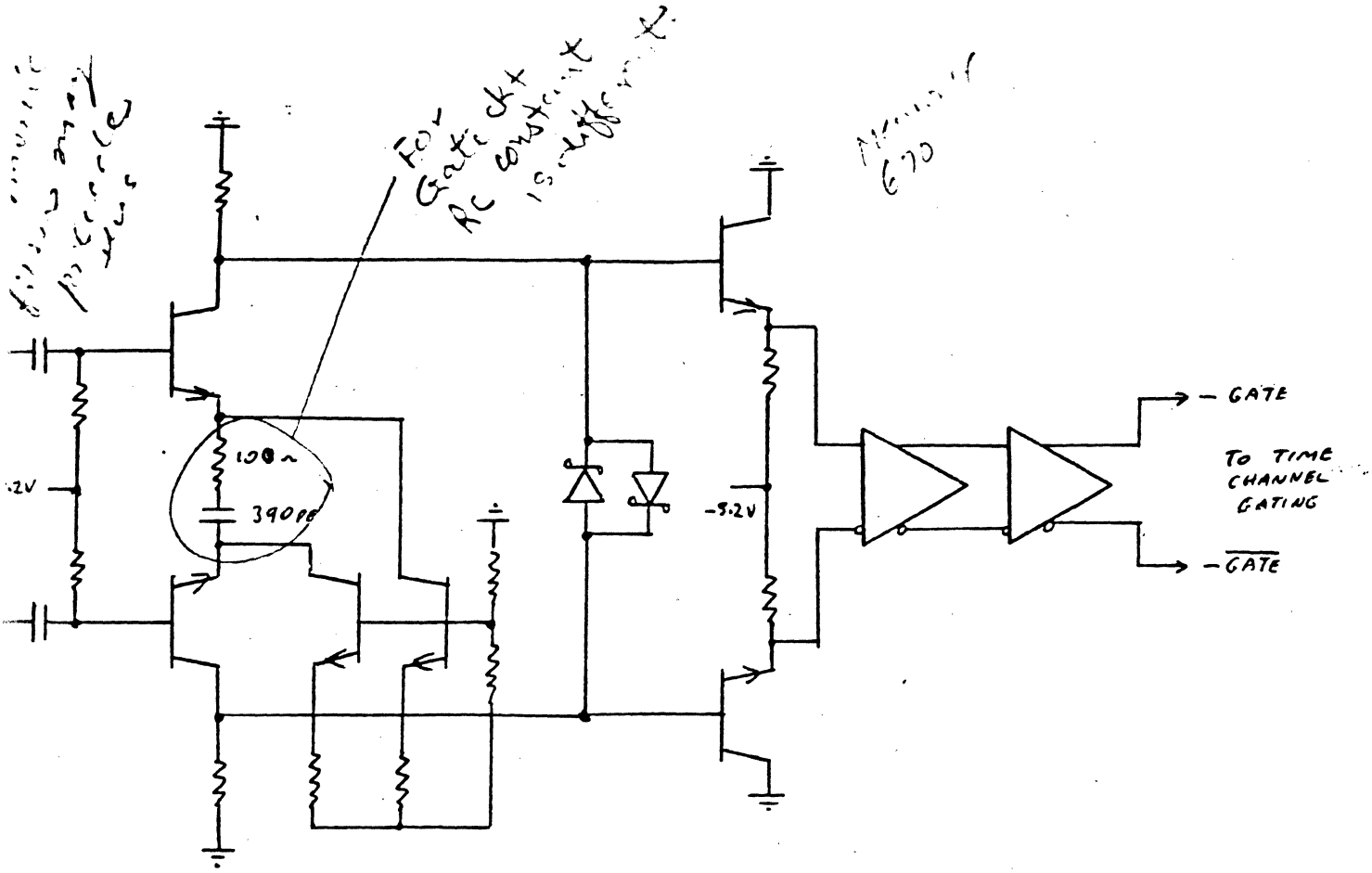


FIG 9.28

LEAD NETWORK GATE GENERATOR FOR REGION 2

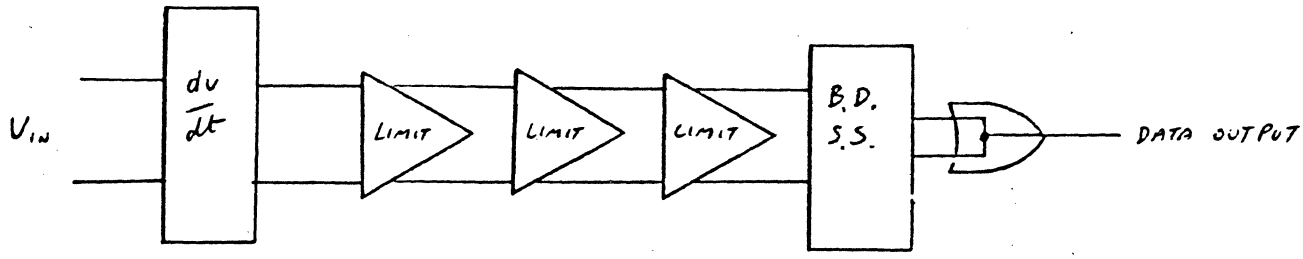


FIG 9.29

REGION 3 DETECTOR (SIMPLEST OF ALL)

$$\Theta = \tan^{-1}\frac{X_c}{R_E} \qquad \alpha = \tan^{-1}\frac{X_c}{3R_E}$$

(EQ 9.41, 2)

$$X_c = 200\sin\Theta = 200\sin\left(\tan^{-1}\frac{X_c}{R_E}\right) \quad \text{(EQ 9.43)}$$

$$R_E = \frac{X_c}{\tan\Theta} = \frac{200\sin\left(\tan^{-1}\frac{X_c}{R_E}\right)}{\tan\Theta} \quad \text{(EQ 9.44)}$$

$$R_E = \frac{X_c}{\tan\alpha} = \frac{133.33\sin\left(\tan^{-1}\frac{X_c}{3R_E}\right)}{\tan\alpha} \quad \text{(EQ 9.45)}$$

$$R_E = \frac{200\sin\left(\tan^{-1}\frac{X_c}{R_E}\right)}{\frac{X_c}{R_E}} = \frac{133.33\sin\left(\tan^{-1}\frac{X_c}{3R_E}\right)}{\frac{X_c}{3R_E}} \quad \text{(EQ 9.46)}$$

$$1 = \frac{200\sin\left(\tan^{-1}\frac{X_c}{R_E}\right)}{X_c} = \frac{3(133.33)\sin\left(\tan^{-1}\frac{X_c}{3R_E}\right)}{X_c} \quad \text{(EQ 9.47)}$$

$$200\sin\left(\tan^{-1}\frac{X_c}{R_E}\right) = 400\sin\left(\tan^{-1}\frac{X_c}{3R_E}\right) \quad \text{(EQ 9.48)}$$

$$\therefore \sin\left(\tan^{-1}\frac{X_c}{R_E}\right) = 2\sin\left(\tan^{-1}\frac{X_c}{3R_E}\right) \quad \text{(EQ 9.49)}$$

$$\sin\left(\tan^{-1}A\right) = 2\sin\left(\tan^{-1}\frac{A}{3}\right) \quad \text{(EQ 9.50)}$$

substituting the identity of $\tan^{-1}A = \sin^{-1}\frac{A}{\sqrt{1+A^2}}$ (EQ 9.51)

we can get $\sin\left(\sin^{-1}\frac{A}{\sqrt{1+A^2}}\right) = 2\left(\sin\left(\sin^{-1}3\frac{A}{\sqrt{1+\frac{A^2}{9}}}\right)\right)$ (EQ 9.52)

$$\frac{A}{\sqrt{1+A^2}} = \frac{2}{3}\,\frac{A}{\sqrt{1+\frac{A^2}{9}}} \quad \text{(EQ 9.53)}$$

$$\text{or} \quad A = \sqrt{\frac{5}{3}} = 1.2909944 \qquad \text{(EQ 9.54)}$$

$$\therefore \quad \tan^{-1}A = 52.238756^{\circ} = \theta \qquad \text{(EQ 9.55)}$$

$$\tan^{-1}\frac{A}{3} \quad 23.283731^{\circ} = \alpha \qquad \text{(EQ 9.56)}$$

$$X_C = 200 \sin 52.238756^{\circ} = 158.11\ \Omega \qquad \text{(EQ 9.57)}$$

$$R_E = \frac{158.11}{\tan 52.238756} = 122.47\ \Omega \qquad \text{(EQ 9.58)}$$

for proof we will verify that $\frac{Xc}{3}$ produces 133.33Ω at 23.383°

$$Z_{3F_L} = \frac{158.11}{3 \sin 23.283731} = 133.33\ \Omega \text{ QED} \qquad \text{(EQ 9.59)}$$

Our capacitor is

$$C = \frac{1}{(2\pi)(2.5\times10^{6}\text{Hz})(158.11\Omega)} = 4.02\times10^{-10}\text{F} \qquad \text{(EQ 6.60)}$$

Choose 390 PF.

The resistor becomes, using our 6.0 ma current source,

$$R = 122.11\Omega - 2\left(\frac{26}{6} + 5\right) = 103.4\Omega \quad \text{(choose } 100\Omega\text{)}$$

All other circuit values are as we calculated them before.

The limiting gain required for a 5.0 V_{pp} diff input signal is from EQ 9.25

$$A_{min} = \frac{1.0V}{(1.25V_{BP\ SE})(\sin 1.98^{\circ})} = 23.4 \text{ as before}$$

Therefore we need $\frac{23.4}{2}$ = 11.7 in the following stages.

The MC 10116 has a minimum gain of around 8; therefore, we need two series stages as before. This circuit is less noisy due to the clamping operation of the first circuit for Region 2 but, as stated before, can only be used with signals that do not return to the base line between transitions but have a reasonably small shoulder.

The circuits of Region 3 are are far simpler, in fact, they are the cheapest of all. The drawback is, of course, the much poorer resolution and the attendent bit shift. We did discuss ways of reducing the bit shift by using Write Precompensation, but in this region that method has diminishing returns due to amplitude loss. We will discuss LATER other methods of compensating, but they are restricted to certain codes.

The block diagram is simply a differentiator followed by a series of limiters and a Bidirectional Single Shot. No gating and no need to line up pulses in the gate. See Figure 9.29.

Again we will consider our 5.0 MHz, 10 $V_{pp\ max}^{diff}$ linear input signal only we will drop down to 1.0$V_{pp\ diff}$ for the minimum. This is consistent with practice in this region. The differentiator is the same as before -- we do not need to change a thing (except remove the delay line from the Region 2 version). All design criteria is the same since we want to use ECL logic. The only thing we need to do is to refigure the total minimum gain required.

(EQ 9.61)

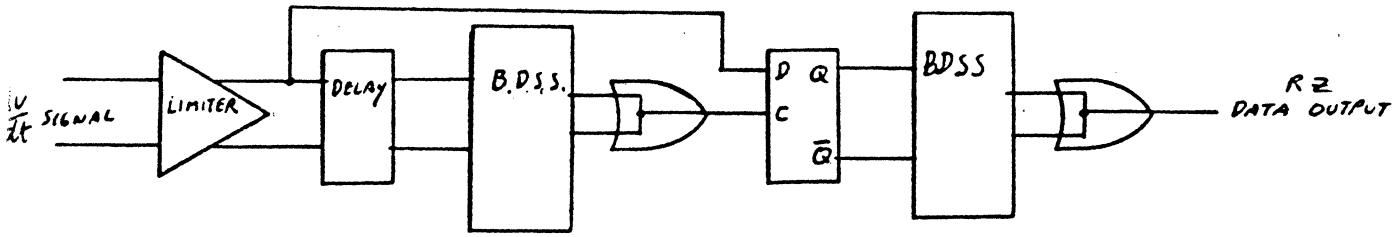
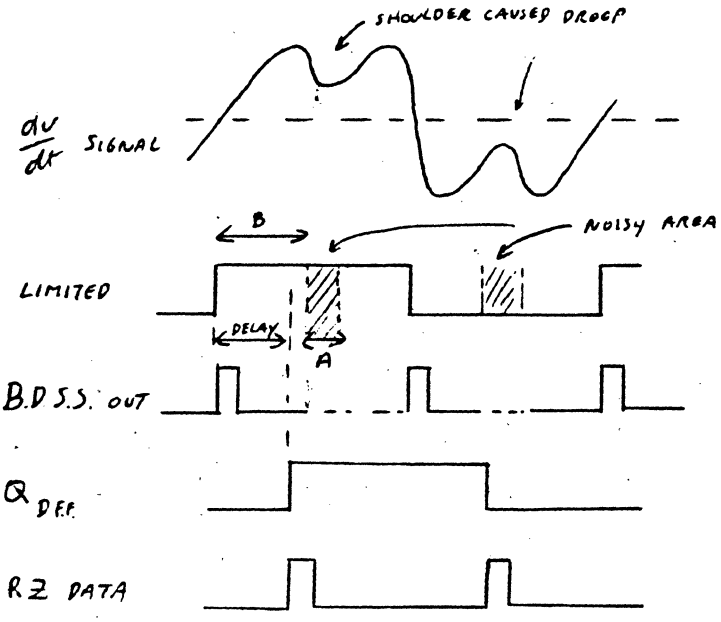


FIG 9.30 A

REGION 2 (RIGHT PORTION ONLY) DATA DETECTOR



As long as.
B > A,

FIG 9.30 B

owing the gain of the differentiator is $.778 \angle 89.9^{o}$ from EQ 9.23, we quire

$$\frac{115.77}{.778} = 148.8 \text{ in the limiters} \qquad \text{(EQ 9.62)}$$

is is too much for two limiters with a gain of 8 each therefore we need 3. e spare gain then $\frac{512}{148.8}$ = 3.4 more than we need, but it does not hurt at ll since the signal is limited.

In Region 4 the detector is the same as in Region 3 only the bit shift r peak shift is so bad that other methods must be used to determine the presence r absence of a peak in a particular time slot or bit cell. This will be dis- ussed when we talk about clocking circuits.

There is no reason to assume that the preceding circuits could not be lesigned to interface with T^2L logic or any other logic family and most vere prior to 1965.

Another type of detector is used in Region 2, particularly the right hand side of Region 2. As can be seen the shouldering which is due to 3rd harmonic content is worse to the left and better to the right which is opposite for bit shift. We can take advantage of the lesser shouldering by introducing a circuit that is tolerent of some shouldering. Refer back to Figure 9.27. As can be seen, the worry is when the noise content carries the differentialed signal back across the base line thus generating a false bit. In the circuit of Figure 9.30A a delay line has been added in series with a Bidirectional Single Shot and applied to the clock of a 'D' flip flop. This then provides a clock for each zero crossing regardless of its legitimacy. The operation can be deduced with the aid of Figure 9.30 B

As can be seen if the delay is established to be greater than the noisy droop area A and less than area B or the certainty area then the 'D' flip flop will reproduce the limited signal, delayed, without the noisy area for a noise generated clock will just clock in the same polarity.

$$A < \text{Delay} < B \qquad \text{(EQ 9.63)}$$

The engineer must be able to guarantee the above equation for all head - disk variations over the proposed production. The remainder of the circuit follows as before, with the use of a Bidirectional Single Shot to generate RZ data.

There is another form of differentiator that can be designed that provides a poorer response to the third harmonic than the more traditional differentiator. Usually high frequency roll off is provided at the expense of true differentiation by placing the unavoidable pole such that the phase angle at the highest frequency is 70°. For example, in a system where

$F_H = 2\ F_L$ we choose Xc of the differentiator as 100Ω at F_H

therefore

$$Z_{EMITTER\ F_H} = \frac{100\Omega}{\sin 70^\circ} = 106.4\Omega \qquad \text{(EQ 9.64)}$$

$$R_{EMITTER} = \frac{100\Omega}{\tan 70^\circ} = 36.39\Omega \qquad \text{(EQ 9.65)}$$

$$Z_{EMITTER\ F_L} = \frac{200\Omega}{\sin\left(\tan^{-1}\frac{200}{36.39}\right)} = 203.28 \qquad \text{(EQ 9.66)}$$

$$Z_{EMITTER\ 3F_L} = \frac{200}{3}\left(\frac{1}{\sin\left(\tan^{-1}\frac{200}{3(36.39)}\right)}\right) = 75.95\Omega \qquad \text{(EQ 9.67)}$$

for a gain f_L to $3f_L$ of $\frac{203.28}{75.95} = 2.676$ (EQ 9.68)

Then we can see the worsened effect of the shoulder. The new differentiator produces a fixed phase angle of +90° with a sine function in magnitude that can be judiciously placed to our advantage.

This circuit was invented by a student named TSAI HWA CHEN* in response to an engineer's complaint of the foregoing effects. Mr. Sordello provides the following derivation from figure 9.31:

From the identities $\sin\theta = \frac{e^{j\theta} - e^{-j\theta}}{2j}$ (EQ 9.69)

$$T = e^{-sc} \rightarrow e^{-j\omega T} \quad \text{(EQ 9.70)}$$

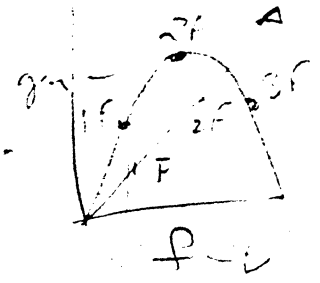
$$V_o = V_{in}\left(1 - e^{-j\omega T}\right) \quad \text{(EQ 9.71)}$$

$$\frac{V_o}{V_{in}} = 1 - e^{-j\omega T} = 1 - e^{-\frac{j\omega T}{2}} \cdot e^{-\frac{j\omega T}{2}} \quad \text{(substitute)}$$

$$\frac{e^{\frac{+j\omega T}{2}}}{e^{\frac{+j\omega T}{2}}} - \frac{e^{\frac{-j\omega T}{2}}}{e^{\frac{+j\omega T}{2}}} = \frac{2j}{2j}\left(\frac{e^{\frac{+j\omega T}{2}}}{e^{\frac{+j\omega T}{2}}} - \frac{e^{\frac{-j\omega T}{2}}}{e^{\frac{+j\omega T}{2}}}\right)$$

$$= 2j\left(\frac{e^{\frac{+j\omega T}{2}} - e^{\frac{-j\omega T}{2}}}{2j}\right)\left(e^{-\frac{j\omega T}{2}}\right) = \underbrace{2j\left(\sin\frac{\omega T}{2}\right)}_{\text{shaping term}}\left(e^{-\frac{j\omega T}{2}}\right) \quad \text{(EQ 9.72)}$$

Differentiator



*"Use of delay lines in Reading Manchester Codes," IEEE Trans. on Computers, Vol. C17 = #9, Sept. '68, Pages 827-845.

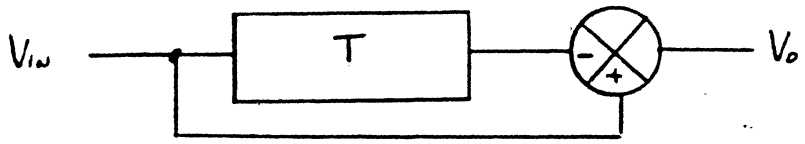


FIG 9.31

DELAY LINE DIFFERENTIATOR

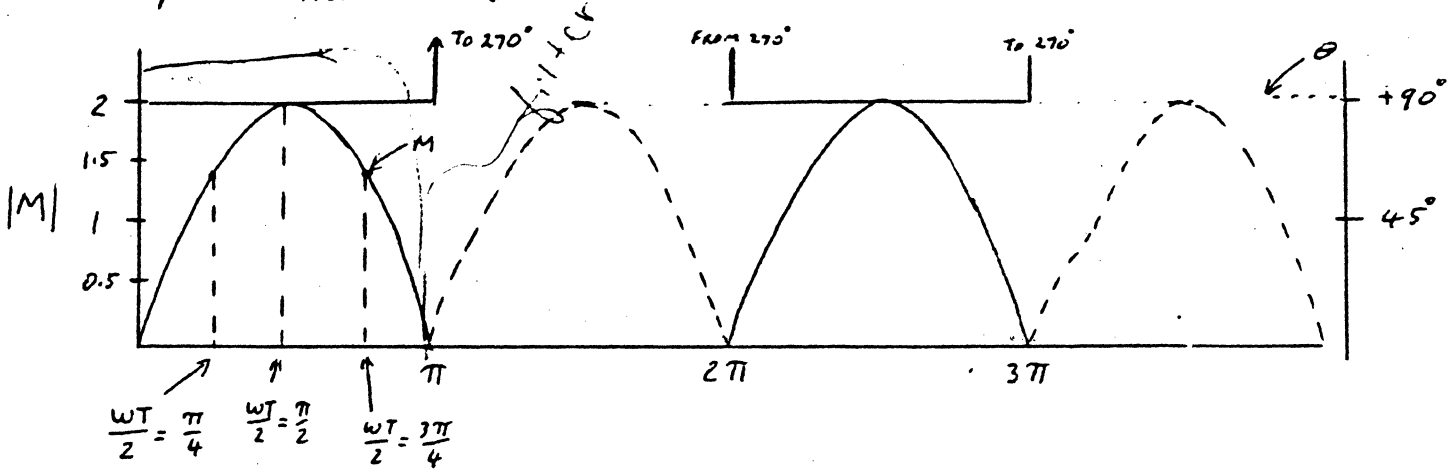


FIG 9.32 A

MAGNITUDE AND PHASE PLOT OF THE DELAY LINE DIFFERENTIATOR WITH DELAY REMOVED

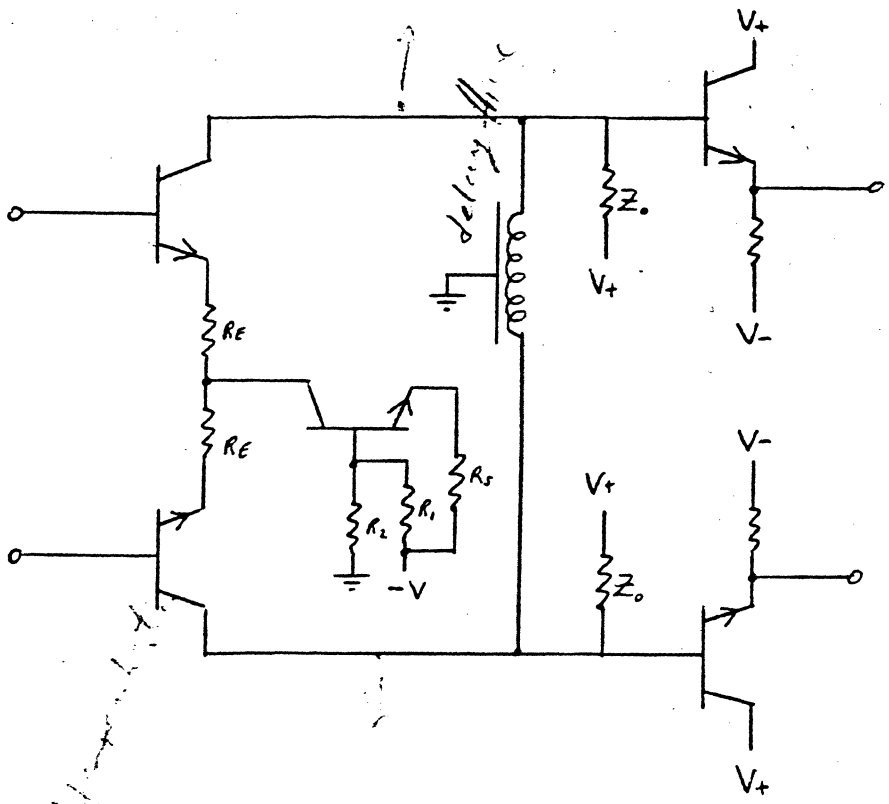


FIG 9.32 B

SCHEMATIC OF A BALANCED DELAY LINE DIFFERENTIATOR

which is a gain with a sinusoidal response at a phase angle of 90° and delayed by $\frac{T}{2}$.

If for a Region 2 system where $F_H = 2\ F_L$ we can choose to place F_H at $\frac{\pi}{2}$ then, (FIG 9.32 A)

$$\frac{\omega_H T}{2} = \frac{\pi}{2} \qquad \text{(EQ 9.73)}$$

Where $W_H = 2\pi F_H$ then we can calculate the delay required.

$$\frac{2\pi F_H T}{2} = \frac{\pi}{2} \quad \text{or} \quad T = \frac{1}{2FH} \qquad \text{(EQ 9.74)}$$

This says that at $3F_L$ and at F_L the magnitude response is 0.707 or

$$\frac{2\pi F_L T}{2} = \frac{\pi}{4} \quad \text{and} \quad \frac{2\pi 3F_L}{2} = \frac{3\pi}{4} \qquad \text{(EQ 9.75)}$$

Using this circuit the gain at $3F_L$ becomes the same as at F_L for a $\frac{1}{2.676}$ improvement over the older method.

Because the accuracy of the peak itself in the presence of the 3rd harmonic is enhanced the limiter stage gain needs to be raised by 2.676 or more when using this differentiator due to the reduction in the 3rd harmonic content. A second filter is required in order to suppress the higher lobes of the magnitude response. This filter precedes the differentiator.

The differentiator circuit is shown in Figure 9.31.B

The delay line is placed across the collectors thus producing the function of subtraction differentially. The collector resistor load

is fixed at Zo of the delay line thus absorbing reflected energy either way.

This concludes the chapter on detectors.

10. LINEAR AMPLIFIER

The linear amplifiers are used to provide the linear gain between the preamplifier and the detector. They include stages of gain, selection, filtering, phase correction, AGC gain control, and signal shaping. This chapter will deal with all these circuit functions.

Reviewing the block diagram, Figure 5.1, we can see the placement of these amplifiers. Figure 10.1 shows a typical functional block diagram.

We can now discuss the type of amplifier required based on what we know about the detector requirements. All detectors that include a fixed amplitude reference as a criteria for opening a gate require Automatic Gain Control (AGC) such that the input to the Detectors remains within some bounds. Those that do not have a fixed reference for the gate and those with no gate do not require AGC. There are some circuits that use a amplitude determined reference for the clipping level instead of a fixed value such as we used in Figure 9.19. This could just as easily be derived from the input amplitude by adding a filter to the diode isolated Full Wave Rectified signal such as is shown in Figure 10.2.

The dynamic range of the Detector input would be wider due to the variations in head signal and amplifier tolerances. The AGC restricts the dynamic range of the detector input to a more reasonable value which reduces the amplifier power dissipation associated with very large signal swings.

Detectors in Region 3 do not require AGC. They do, however, require a very large gain but most of this can be in limiting stages as previously discussed. The Bandwidth of the Linear amplifiers requires careful control in order to properly pass the head signal while eliminating extraneous noise. They are also called upon to correct for non linear phase delays and to provide some

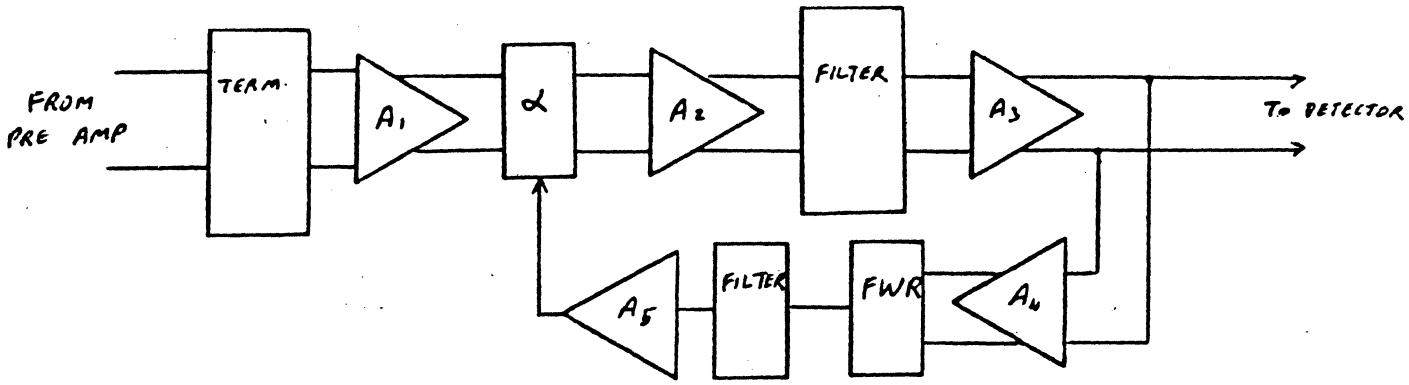


FIG 10.1 BASIC LINEAR AMPLIFIER BLOCK DIAGRAM INCLUDING A.G.C.

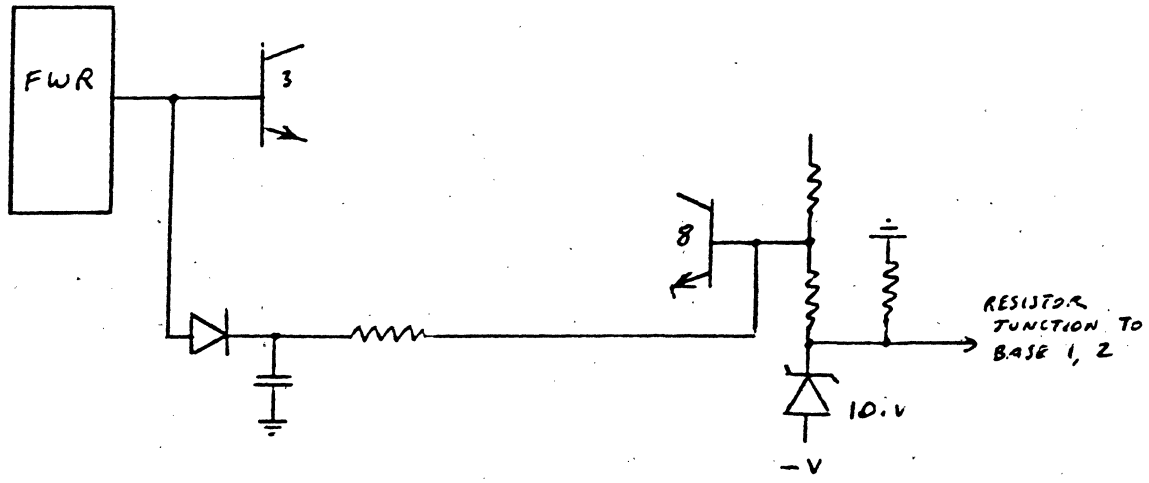


FIG 10.2 SIGNAL AMPLITUDE CONTROL OF THE CLIPPING LEVEL. (SEE FIG 9.19 FOR THE REST OF THE CIRCUIT)

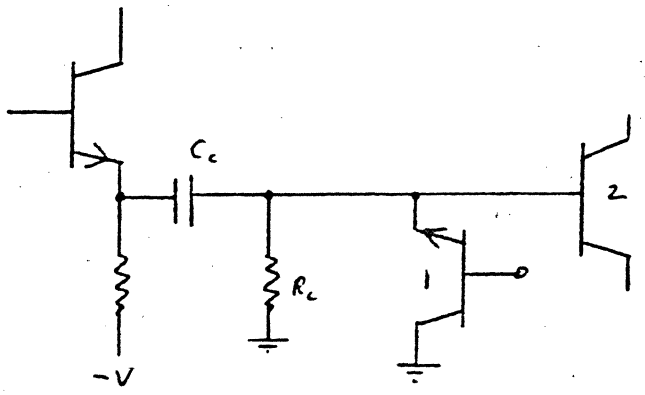


FIG 10.3A

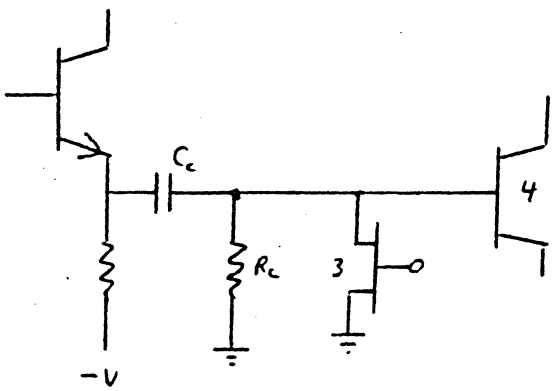


FIG 10.3B

TRANSIENT RECOVERY OF COUPLING CIRCUITS

spectral shaping in some cases. The code used determines the width of the Bandwidth while the data transfer rate or transition rate determine the upper fundamental frequency of interest. Most codes have a DC content thus any AC coupling after the differentiator is detrimental to the bit timing. This occurs as the base line moves to make the area above and below the base line equal. When this happens the zero crossings are lost resulting in time shifted crossings. You will notice that all the designs of the differentiator that we discussed in Chapter 9 are DC coupled with balancing circuits, to eliminate offsets, following actual differentiation.

AC COUPLING

The linear amplifier, however, can be AC coupled as long as the low frequency cut off is below the frequency at which there is significant energy. This raises a problem as the τ of such coupling circuits is large making recovery from a DC transient or shift very long. We have this problem any time we change heads or when switching from a Write to a Read. If τ were 10 μs then for the base line to be fully recovered we require 50.μs or 5τ. We can take advantage of the common mode rejection of a differential amplifier for common mode shifts in DC level, but unfortunately almost all transient shifts are non symmetrical therefore differential. We must reduce this recovery time substantially as it forces an increase in the formatting time lost between records. There are two circuits that can be used that reduce τ for the duration of a switching transient yet allows the full T for data handling. These are shown in Figure 10.3A and B.

LINEAR AMPLIFIER

In Figure 10.3A a chopper transistor is used in the inverted connection to short the coupling resistor for the duration of the transient. The charge on the coupling capacitor then can quickly reach the level required. After just a few microseconds the base drive is removed restoring the τ of the coupling circuit. The emitter follower pull down current must be great enough to charge the capacitor to the most negative transient value within the allotted time. Care must also be given to the emitter current - base current curves of the chopper transistor as there is still an offset, but this should now be common mode therefore the recovery from the chopper offset appears much shorter and less noticeable.

The second circuit shown in Figure 10.3B uses a Fet as a voltage controlled resistor. There are two considerations associated with its use. First the drain voltage of the transient may force the Fet into the current source mode which is past the knee of the V_D - I_D curves which increases the recovery time; and second, the capacitive coupling of the gate switching transient to the drain can leave an undesirable transient, but this is also common mode or nearly so.

The τ of the coupling circuit must be at least ten times that for the lowest fundamental frequency. This must include the effect of all the series coupling capacitors, base and emitter, up to the bases of the differentiator. For example, if we had 5 such coupling circuits each with a τ of 10.μsec then a single one would have a -3db frequency F_1 of 15.91 KHz, but 5 in series would be down 15 db at 15.91 KHz. The real -3db frequency would be

$f = \frac{1}{2\pi\tau}$

$$(F_{3db\ single}) \tan\left[\sin^{-1}\left(\frac{1}{\log^{-1}\frac{3.01}{20N}}\right)\right] = \tan\left[\sin^{-1}\left(\frac{1}{\log^{-1}\frac{3.01}{(20)(5)}}\right)(15.91)\right] = 41.22 \text{ KHz} \quad (\text{EQ } 10.1)$$

UPPER FREQUENCY ROLL OFF

The upper 3db roll off is controlled by the upper transition rate.

UPPER FREQUENCY ROLL OFF (Continued)

Generally, the -3db point occurs near 1.5 times half the transition rate. This is necessary as we need to include the harmonics associated with the shoulders. The degree of roll off should be a function of the noise spectrum and usually is between 18-24 db per octave. A primary consideration is the effect on phase linearity which we will discuss shortly. The type of filter depends on the amount of roll off required, the amount of phase correction required and the degree of signal shaping required. Most amplifiers use either the Butterworth type filter because of its maximally flat magnitude response, or the Butterworth Thompson filter which is a compromise between a maximally flat time delay AND MAGNITUDE response. The Bessel filter is also used because of its maximally flat time delay characteristics, however, it has poor roll off characteristics. A newer approach is to use cosine filters to shape the signal before detection in order to improve the PW50. Generally filters are concerned with reducing noise while retaining the signal except as last discussed. The amplifier must not contribute to the roll off significantly as this type of roll off is uncontrolled due to stray capacitances, transistor junction capacitances, and Miller capacitance. For this reason, we follow the general rule of 10 times the required filter bandwidth for the complete amplifier. This means that each stage should have a bandwidth greater than $10\sqrt{N}$ times the upper 3db point where N is the number of stages in series.

GAINS

The amount of gain required in the linear portion can be calculated from the minimum head signal expected out of the Pre Amplifier and the minimum

GAINS

signal required at the input to the Detector.

$$A_{MIN} = \frac{V_{IN\ MIN\ PP\text{-}DIFF\ @\ DETECTOR}}{(V_{HEAD\ MIN\ PP\text{-}DIFF})(A_{MIN\ PRE\ AMP})} \quad \text{(EQ 10.2)}$$

We must then determine the maximum signal that will appear at the Detector input resulting from a maximum head signal times the maximum Pre Amplifier gain times the maximum Linear Amplifier gain.

$$V_{O_{MAX\ PP\text{-}DIFF}} = (V_{HEAD\ MAX\ PP\text{-}DIFF})(A_{MAX\ PRE\ AMP})(A_{L.A.\ MAX}) \quad \text{(EQ 10.3)}$$

If we compare the results of EQ 10.3 to the restrictions to the upper input voltage to the Detector, we will see if we need AGC or not, or if we need to increase the linear range of the Detector input. For Region 1 and 2 circuits it is preferential to use AGC which allows us to reduce the power dissipation of the last linear stages. It is also preferred that the signal level at the Detector inputs be established at at least -6db below the tolerable distortion limit of the Detector's first input stage. This allows a ±6 db margin to handle sudden amplitude changes without detrimental distortion. For Region 1 and Region 2, fixed reference Detectors AGC is required unless the delta signal amplitude worse case is less than the Detector limits. Such is highly unlikely. As we discussed before it is also preferable to break the gain requirements up into several stages of low gain rather than one or two of high gain because of Bandwidth requirements. Commercial differential video amplifiers can serve well in these positions except for the last stage or stages due to the signal output swing requirements for accuracy vs. the IC's specification. We designed for 5.0 V_{pp} nominal into the Detector because of the 0.1V linear region of a

AR AMPLIFIERS

S (Continued)

ent switch vs. the percentage reference. As long as the signal remains
ar where we require the peak this will always be true.

CTION

Often the head signal originates from several sources, such as, groupings of widely separated heads, fixed heads and moving heads, or d-Verify heads. Such arrangements can be suitably handled by providing arate loading and separate amplification at the first stage. The ividual first stages can have different gains to accommodate differing el signals. Selection is usually accomplished by a collector dot of individual amplifiers with a switched current source. Such a circuit is own in Figure 10.4. There is a commercial device available, MC1445, that rforms the same function though the input dynamic range is limited to a w hundred millivolts including offsets. As can be seen this circuit can be signed for any signal amplitude gain, Bandwidth, or bias levels. The design cludes the usual considerations of linearity etc. Let us assume that the put signal is referenced to ground at 200Ω differential with a -0.7V DC mponent across 402Ω to ground each phase. This signal maximum can be 150 mV DIFF with a 50 mV maximum DC offset. Bandwidth requirements are DC to .MHz. Let us have a gain of 2. This will allow us to connect the Preamplifier near Amplifier. First we can calculate R_3 from the known value of R_1 and = 402Ω to obtain the desired termination.

$$R_3 = \frac{(R_1+R_2)(Z_T)}{R_1+R_2-Z_T} = \frac{(402+402)(200)}{402+402-200} = 266.2\ \Omega \qquad \text{(EQ 10.4)}$$

(use 261Ω 1%)

SELECTION (Continued)

Now that we have fixed the collector resistors we need to determine the emitter current as that in series with the total emitter resistance RE + r_e + Rm determines the linearity of the stage.

We calculated a 200 mv pp input maximum; therefore we need greater than

$$\frac{V_{IN\,pp\,max}}{R_T} = \frac{0.200\,V}{100\,\Omega} = 2.0\,ma \qquad \text{(EQ 10.8)}$$

This then means that we need twice that current to reduce the variables associated with r_e at this current level. Choose 4.0 ma.

The main resistor RE can now be determined.

$$R_E + r_e + R_m = \frac{R_L}{A} = \frac{200}{2}\,\Omega \qquad \text{(EQ 10.9)}$$

$$\text{Therefore} \quad R_E = \frac{200}{2} - \frac{26}{4} - 5 = 88.5\,\Omega \qquad \text{(EQ 10.10)}$$

Just to see the effect of the emitter resistance, let us tabulate the gain change as a function of current I, using

$$A_{\pm} = \frac{R_L}{R_E + \frac{26}{I_E} + R_m} \qquad A_T = \sqrt{A_+ \cdot A_-} \qquad \text{(EQ 10.11)}$$

SELECTION (Continued)

With a 150 mV_{pp} DIFF input signal + 50 mV offset we can have a 200 mV_{pp} differential input maximum.

The collector resistor is based on the 50 MHz Bandwidth and the stray capacitance. Notice that this is the stage Bandwidth not the Amplifier Bandwidth.

We will calculate the capacitive load as C_T.

$$C_T = C_{ob_1}(1+A) + C_{ob_2} + C_w = 0.58(1+2) + 0.58 + 10.\,pF = 12.32\,pF \quad \text{(EQ 10.5)}$$

Miller capacitance must be included as our source impedance is $\frac{200}{(2)(2)} = 50\Omega$ which represents the termination and the input cable in parallel single ended.

Choose $R_L = 200 = 2\,R_{E_T}$ and $\beta_{Min} = 20$

$$F_{\substack{Miller \\ -3db}} = \frac{1}{(2\pi)(R_L)(C_{ob})(1+A)} = \frac{1}{(2\pi)(200)(5.8\times10^{-13})(1+2)} = 4.573\times10^{8}\,Hz \quad \text{(EQ 10.6)}$$

OR F due to stray capacitance

$$f_{-3db} = \frac{1}{(2\pi)(R_L)(C_T)} = \frac{1}{(2\pi)(200)(1.232\times10^{-11})} = 6.459\times10^{7}\,Hz \quad \text{(EQ 10.7)}$$

Which means that the stray capacitance dominates with a pole at 64.6 MHz which satisfies our requested Bandwidth.

SELECTION (Continued)

interested in the peak input into the differentiator as that determines the slope at the zero crossing. We might have improved our chances by settling for a higher Z_E. This should have been calculated with 2 Z_E instead of ZE. Although the ratio of r_e to Z_E is much smaller than our present one thereby maintaining linearity over a wider input swing and thereby making the loss in gain unnecessary.

The linearity is determined solely in the emitter circuit (if linear resistors are used in the collector).

The rest of the circuit can be designed with the tools we have already established and therefore we will not take the space to repeat them. We could emphasize the bias of the current switch used to select the amplifier input to the collector dot. The most positive base must never cause the negative swing of the amplifier emitters above it to saturate due to V_{be} and IR_E drops.

TOTAL AMPLIFICATION

With all the previous background we will now design an amplifier to connect between the Preamplifier and the Detector.

If the range of the Preamplifier output was 100 $V_{min\ PP\ DIFF}$ to 150 $mV_{PP\ MAX}$, including offset, and we wanted a nominal of 7.5 $V_{PP\ DIFF}$ into our Detector (5.0_{MIN} - $10.0_{\ MAX}$VPP DIFF) then we can establish the gain required and the number of stages. From Equation 10.2 we calculate we need a gain of 50 MIN and 66.66 MAX or 60.0 nominal. This is obviously too great for presently available video amplifiers for the output swing required, therefore we need to break it up. $\sqrt{60} = 7.74$ but $\sqrt[3]{60} = 3.914$ which is much more easily manageable.

TION (Continued)

I	A+	A-	A_{TOTAL}	DIFF DISTORTION
4.0	2.000	2.000	2.000	0
3.5	1.9815	2.0145	1.9979	0.105%
3.0	1.9575	2.0263	1.9916	0.42%
2.5	1.9249	2.0361	1.9797	1.01%
2.0	1.8779	2.0442	1.9592	2.04%
1.5	1.8045	2.0512	1.9238	3.81%
1.0	1.6736	2.0573	1.8555	7.22%
0.5	1.3745	2.0625	1.68371	15.81%
0.25	1.0126	2.0649	1.4460	27.70%
0.125	.66334	2.0660	1.17066	41.46%

this means is that the distortion at the peak input signal amounts to an antaneous gain change of about 2.04% differentially and about 6.1% single d on one collector, the positive peak, and 2.21% on the other collector, tive peak. Now we see the reason for increasing the current. The factor is not sacred but strictly depends on the ratio of r_e to RE & Rm. If we e 8.ma then the distortion would be much less differentially and particularly he positive peak. Now the designer has to choose between the power dissi- on resulting from the higher currents and the amount of distortion that can olerated. This will be increasingly more important when we consider the er level stages and the differentiator.

Notice back when we calculated the gain impedance for the differentiator .23 and for the Gate Generator where no RE was used we were not concerned collector peak distortion since we threw away the peaks with limiters, but ere interested in the linear portion around the base line. But we <u>are</u>

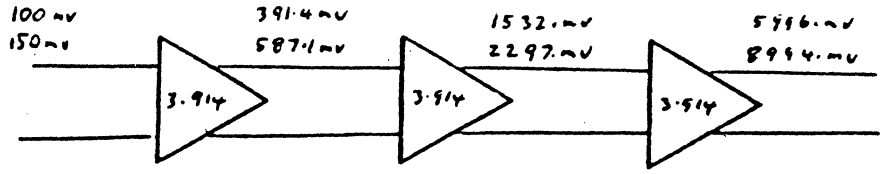


FIG 10.5

THREE STAGE LINEAR AMPLIFIER

50. MHZ BW, $A_T = 60$

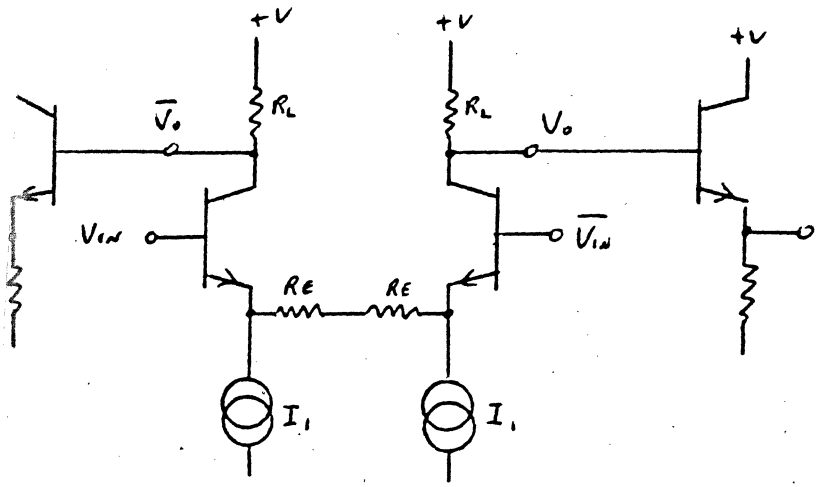


FIG 10.6

BASIC AMPLIFIER

TOTAL AMPLIFICATION

Or we could use a μa 733 for the first stage with a gain of 10 followed by a high level amplifier with a gain of 6 either would be acceptable but as we would like to add some other function, we will go with the 3 stage circuit. The amplitude MIN and MAX is shown for each stage in Figure 10.5.

The bandwidth of the amplifier is to be 50 MHz; therefore we need $\sqrt{3}\,)50 = 86.6$ MHz at each stage. The correct formula is $\left(50.\left(\frac{1}{\tan\left[\sin^{-1}\left(\frac{1}{\log^{-1}\frac{3.01}{20N}}\right)\right]}\right)\right) = 98.\text{MHz}$

$$C_T = C_{ob_1}(1+A) + C_{ob_2} + C_w = 0.58_{pF}(1+3.914) + 0.58_{pF} + 4.0_{pF} \qquad \text{(EQ 10.12)}$$
$$= 13.43\,pF$$

$$R_L < \frac{1}{(2\pi)\,B_w\,C_T} = \frac{1}{(2\pi)(8.66\times10^7)(1.343\times10^{-11})} \qquad \text{(EQ 10.13)}$$
$$= 2.47\times10^2 \quad (\text{choose } 243\,\Omega\ 1\%)$$

$$R_{E_{TOTAL}} = \frac{240\,\Omega_{MIN}}{3.914} = 61.32\,\Omega \qquad \text{(EQ 10.14)}$$

This requires some thought for if we calculate the currents required to maintain a maximum of 5% distortion we must have only $\sqrt[3]{5\%}$ distortion in each stage or 1.7% each. The distortion is worst for the positive half cycle; therefore, we will use that value as our 1.7% MAX.

$$1.7\% \text{ of } 61.32\,\Omega = \Delta r_e = 1.042\,\Omega \qquad \text{(EQ 10.15)}$$

TOTAL AMPLIFICATION (Continued)

$$\frac{-4.988 \pm 2.33\text{NO}'}{2(1.042)} = -13.58\text{m}, +8.79\text{m} \qquad \text{(EQ 10.23)}$$

The correct root is 13.58 ma which gives $1.041\Omega = \Delta r_e$

Stage 3 is $1.042\, I_{1m}^2 - 1.042\left(\frac{2297.\text{mv}}{2(61.32\Omega)}\right) I_{1m} - 26\left(\frac{2297.\text{mv}}{2(61.32)}\right)$ (EQ 10.24)

$$= 1.042 I_{1m}^2 - 19.51\, I_{1m} - 486.96$$

$$\frac{-19.51 \pm \sqrt{(19.51)^2 - 4(1.042)(-486.96)}}{2(1.042)} \qquad \text{(EQ 10.25)}$$

$$\frac{-19.51 \pm 49.09}{2(1.042)} = -32.92\text{m}, +14.19\text{m} \qquad \text{(EQ 10.26)}$$

The correct root is 32.92 ma which gives $\Delta r_e = 1.042\,\Omega$

As can be easily seen the small built-in error in the equation is when we used 2(61.32) as R_T when we know the real resistance is $\frac{26}{I_{INST}} + 5 + R_E$ to determine I_2.

The value of RE for each of the 3 stages is:

Stage 1: $R_{E_1} = 61.32 - 5 - \frac{26}{I_{1\,nom}} = 61.32 - 5 - \frac{26}{6.17} = 52.1\,\Omega$ (EQ 10.27)

Stage 2: $R_{E_2} = 61.32 - 5 - \frac{26}{I_{1\,nom}} = 61.32 - 5 - \frac{26}{13.58} = 54.4\,\Omega$ (EQ 10.28)

Stage 3: $R_{E_3} = 61.32 - 5 - \frac{26}{I_{1\,nom}} = 61.32 - 5 - \frac{26}{32.92} = 55.5\,\Omega$ (EQ 10.29)

TOTAL AMPLIFICATION

Therefore $\frac{26}{I_1 - I_2} - \frac{26}{I_1} = 1.042\,\Omega$ (EQ 10.16)

where I_2 is the current resulting from $\frac{V_{IN\ PP\ DIFF}}{2R_{eT}} = \frac{V_{IN\ PP\ DIFF}}{122.64\,\Omega}$

rearranging $0 = 1.042 I_{1_m}^2 - 1.042 I_{1_m} I_{2_m} - 26 I_{2_m}$ (EQ 10.17)

Stage 1 : $1.042 I_{1_m}^2 - 1.042\left(\frac{150mV}{2(61.32)}\right) I_{1_m} - 26\left(\frac{150mV}{2(61.32)}\right)$ (EQ 10.18)

$= 1.042 I_1^2 - 1.274 I_1 - 3.18\times10^{-3}$

roots $\frac{-1.274 \pm \sqrt{(1.274)^2 - 4(1.042)(-3.18\times10^{-3})}}{2(1.042)}$ (EQ 10.19)

$\frac{-1.274 \pm 1.158\times10^{-1}}{2.084} = -6.17\,m, +10.3\,m$ (EQ 10.20)

The correct root is 6.17.ma which gives $\Delta r_e = 1.042\ \Omega$

Stage 2 is $1.042 I_{1_m}^2 - 1.042\left(\frac{587.1mV}{2(61.32\Omega)}\right) - 26\left(\frac{587.1mV}{2(61.32\Omega)}\right) = 0$ (EQ 10.21)

$1.042 I_{1_m}^2 - 4.988 I_{1_m} - 1.244\times10^{-2}\,m$

roots $\frac{-4.988 \pm \sqrt{(4.988)^2 - 4(1.042)(-1.244\times10^{-2})}}{2(1.042)}$ (EQ 10.22)

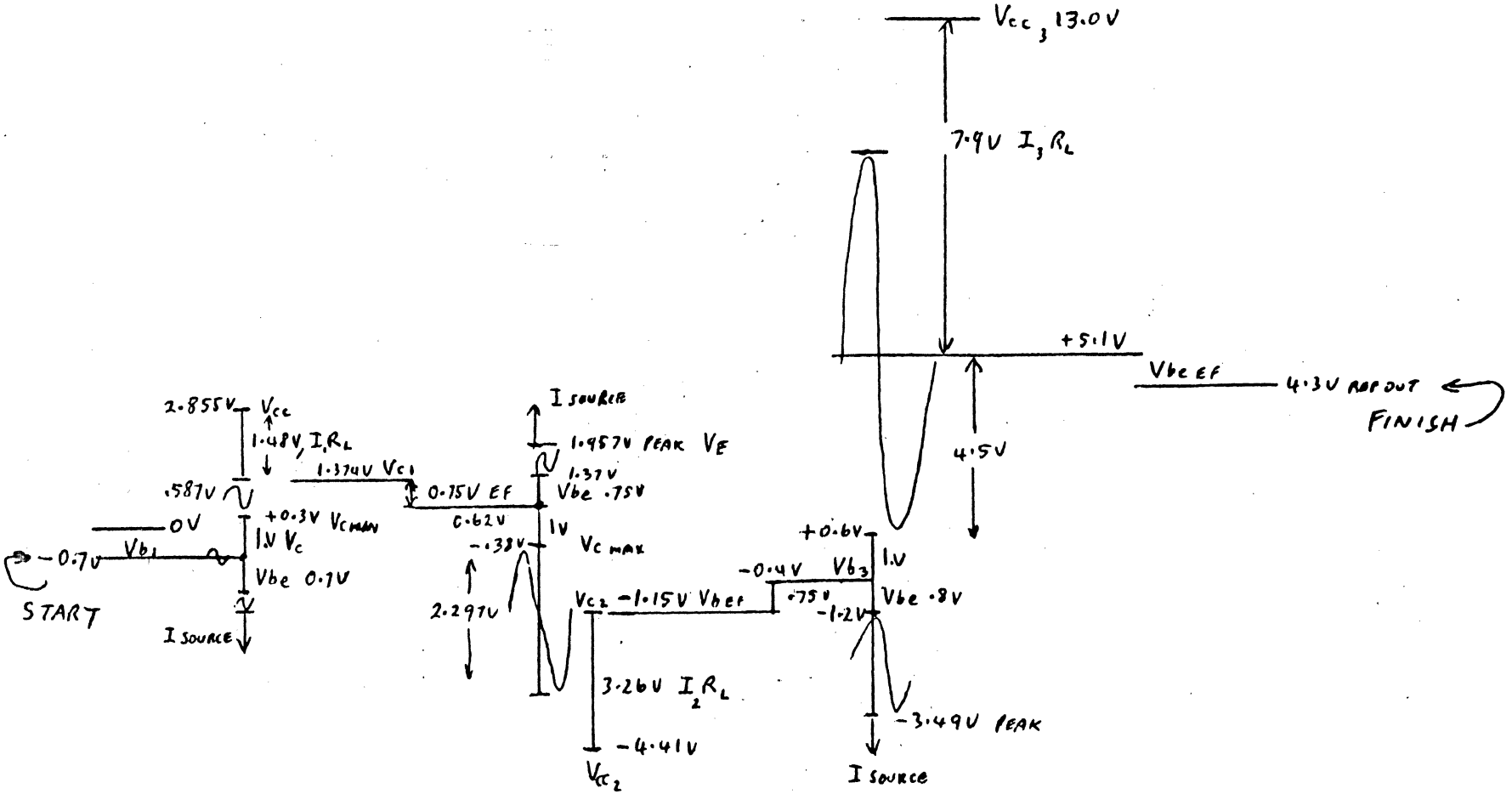


FIG 10.7 INITIAL BIAS DIAGRAM FOR 3 STAGE NPN, PNP, NPN AMPLIFIER

TOTAL AMPLIFICATION (Continued)

If we used the two current source version as planned then the total value of RE will be double or

104.21, 108.81, and 111.06 Ω respectively for R_E .

Now that we know the values of the nominal gain and minimum current as for Fig 10.6 we can now calculate current sources, supply voltages etc. To do this we first need to make a bias diagram. We should also use a PNP stage in the middle in order to minimize the power supply requirements. The Bias Diagram is shown in Figure 10.7.

If we DC couple the bases but AC couple in the emitters we can reduce the number of decoupling circuits.

We make the bias diagram by establishing all the voltages including AC and DC associated with the collector, base and emitter circuit starting with the base. We will allow 6.db margin for the signal swings in every case in order to reduce the possibility of clipping. This is the same as using the differential swing as if it were the single ended swing. We will also allow 1 volt margin between the base and the collector.

Judging by the Bias Diagram, we can easily build the amplifier using ±15V supplies with ± 6.2 zener references for the current sources as no collector will be forward biased.

Stage 1 and stage 2 will require a series resistor to develop the correct bias for stage 2 and 3.

TOTAL AMPLIFICATION (Continued)

We can now draw the schematic for the amplifier and it is shown in Figure 10.8.

$$R_{1\,MAX} = \frac{(15V-5\%)-(6.2V+5\%)-V_{be\,max}}{6.17ma\ min} = \frac{14.25V-6.51V-0.75V}{6.17ma} = 1.132K \quad \text{(EQ 10.30)}$$

choose 1.10K 1% for 6.29ma I min

$$R_{4\,MAX} = \frac{(15V-5\%)-(6.2V+5\%)-V_{be\,max}}{13.58ma\ min} = \frac{14.25V-6.51V-0.80V}{13.58ma} = 511\Omega \quad \text{(EQ 10.31)}$$

choose 511Ω 1% for 13.44ma I min

$$R_{7\,MAX} = \frac{(15V-5\%)-(6.2V+5\%)-V_{be\,max}}{32.92ma\ min} = \frac{14.25V-6.51-0.9V}{32.92ma} = 207.7\Omega \quad \text{(EQ 10.32)}$$

choose 200.Ω 1% for 33.86ma I min

The maximum currents are calculated as:

$$I_{S_1\,MAX} = \frac{(15V+5\%)-(6.2V-5\%)-V_{be\,min}}{(1.1K-1\%)} = \frac{15.75V-5.89V-0.70V}{1.089K} = 8.41ma_{MAX} \quad \text{(EQ 10.33)}$$

$$I_{S_2\,MAX} = \frac{(15V+5\%)-(6.2V-5\%)-V_{be\,min}}{(511-1\%)} = \frac{15.75V-5.89V-0.75V}{505.89\Omega} = 18.00ma_{MAX} \quad \text{(EQ 10.34)}$$

$$I_{S_3\,MAX} = \frac{(15V+5\%)-(6.2V-5\%)-V_{be\,min}}{(200-1\%)} = \frac{15.75V-5.89V-0.80V}{198\Omega} = 45.7ma_{MAX} \quad \text{(EQ 10.35)}$$

To find R_3 max we need to use the following:

$$R_{3\,MAX} = \frac{(6.2V-5\%)-V_{cc_1\,min}}{2(I_{S_1\,MAX})} = \frac{5.89V-2.85V}{2(8.41ma)} = 180.7\Omega_{MAX} \quad \text{(EQ 10.36)}$$

(use 178Ω 1%)

$$R_{6\,MAX} = \frac{(6.2V-5\%)-V_{cc_2\,min}}{2(I_{S_2\,MAX})} = \frac{5.89V-4.62V}{2(18.00ma)} = 35.27\Omega_{MAX} \quad \text{(EQ 10.37)}$$

(use 34.8Ω 1%)

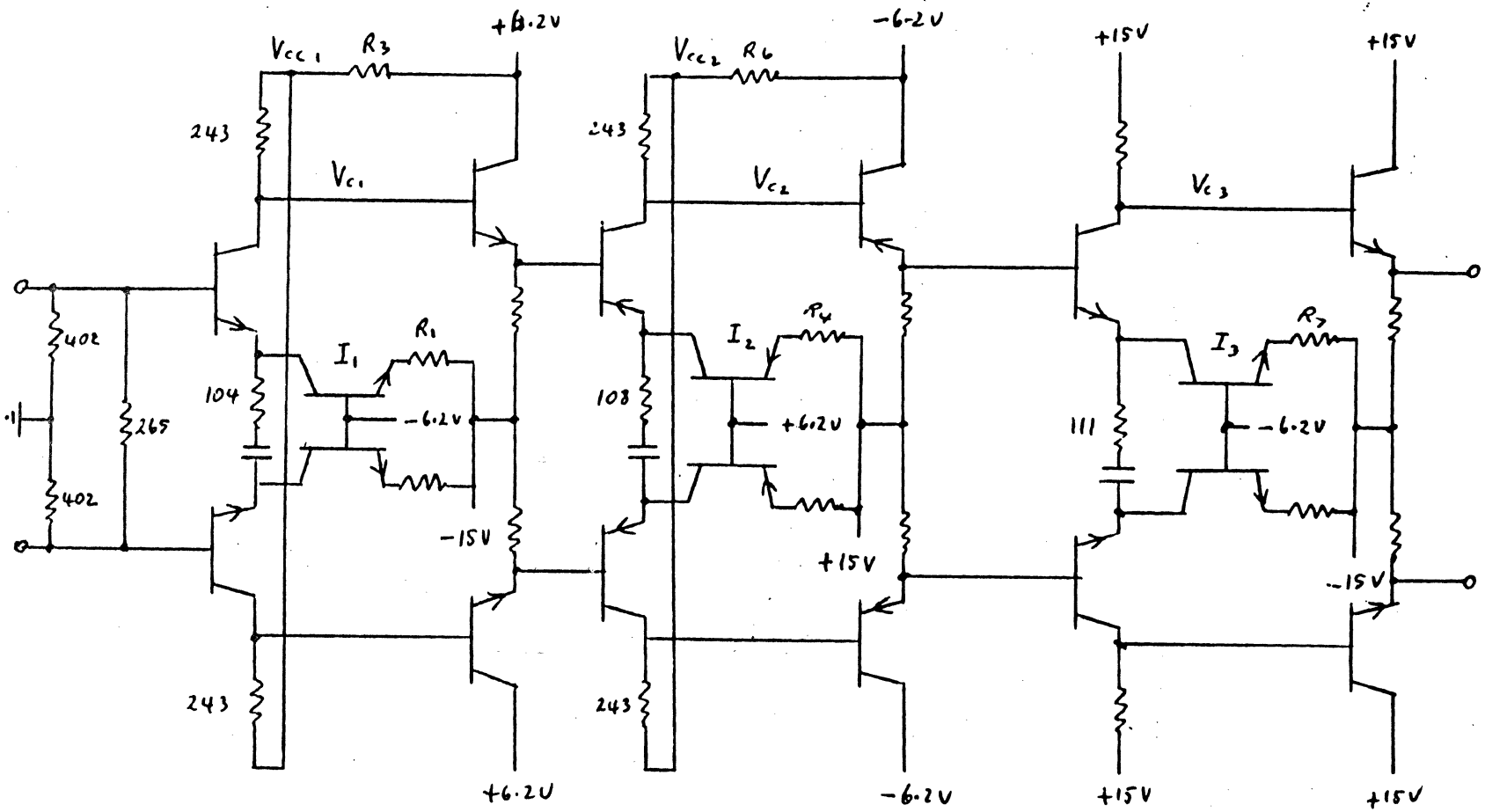


FIG 10.8 SCHEMATIC OF 3 STAGE AMPLIFIER

TOTAL AMPLIFICATION (Continued)

We must eliminate R6 to lower the base voltage of stage three. This is mostly because of the large tolerance on I_{S3}.

$$V_{C_2\ modified_{MAX}} = -(6.2V + 5\%) + I_{S_2\ min}(R_{L\ min}) \quad \text{(EQ 10.43)}$$
$$= -6.51V + (13.44\ m)(240.5\ \Omega) = -3.277\ V$$

$$V_{C_2\ modified_{MIN}} = -(6.2V - 5\%) + I_{S_2\ MAX}(R_{L\ MAX}) \quad \text{(EQ 10.44)}$$
$$= -5.89V + (18.00\ m)(245.4\ \Omega) = -1.47\ V$$

The signal swing at the collector is 2.297V $_{PP\ DIFF}$ and following our 6.db margin rule this makes the minimum emitter peak voltage at stage 3.

$$-V_{C_2\ MAX} - \Delta V_{be_{2-3\ MAX}} - 2.297V = -3.277 - 0.05V - \frac{2.297V}{2} \quad \text{(EQ 10.45)}$$
$$= -4.47\ V\ Peak.$$

which solves the emitter problem but we still have a collector problem in Stage 3.

The base of stage 3 is +0.48V MAX peak, but the collector is

$$3.033V - \frac{8.99V}{2} \text{ or } -1.463V \quad \text{(EQ 10.46)}$$

or an overlap of 0.48V + 1.463V = 1.94V

We need to raise the collector supply to 17V to handle the collector bias problem. That is hard to obtain in most cases as ± 15Volts are standard supplies.

TOTAL AMPLIFICATION (Continued)

By using the maximum value we guarantee the collectors will not saturate. There will be a difference in Vc for each stage as the worse case is calculated. Let's do that.

$$V_{C_1\,MAX} = (6.2V + 5\%) - 2(I_{S_1\,min})(R_{3\,min}) - I_{S_1}(R_{L\,min}) = \quad \text{(EQ 10.38)}$$
$$= 6.51V - 2(6.29mA)(176.2\Omega) - 6.29mA(240.5\Omega) =$$
$$= 6.51 - 2.216 - 1.51 = 2.784\ V$$

$$V_{C_2\,MAX} = -(6.2V + 5\%) + 2(I_{S_2\,min})(R_{6\,min}) + I_{S_2\,min}(R_{L\,min}) \quad \text{(EQ 10.39)}$$
$$= -6.51V + 2(13.44mA)(34.8\Omega) + 13.44mA(240.5\Omega)$$
$$= -6.51V + 0.925V + 3.23V = -2.335\ V$$

$$V_{C_1\,min} = (6.2V - 5\%) - 2(I_{S_1\,MAX})(R_{3\,MAX}) - I_{S_1\,MAX}(R_{L\,MAX}) \quad \text{(EQ 10.40)}$$
$$= 5.89V - 2(8.41mA)(179.78\Omega) - (8.41mA)(245.43\Omega)$$
$$= 5.89V - 3.02V - 2.06V = 0.81\ V$$

$$V_{C_2\,min} = -(6.2V - 5\%) + 2(I_{S_2\,MAX})(R_{6\,MAX}) + I_{S_2\,MAX}(R_{L\,MAX}) \quad \text{(EQ 10.41)}$$
$$= -5.89V + 2(18.0mA)(35.15\Omega) + 18.0mA(245.43\Omega)$$
$$= -5.89V + 1.26V + 4.41V = -0.22\ V$$

With the last result we see we need a couple more volts or so to keep the last stage out of saturation which might be accomplished by using +15V for Vcc.

$$V_{C_3\,MIN} = (15 - 5\%) - (I_{S_3\,MAX})(R_{L\,MAX}) = 14.25V - (45.7mA)(245.43\Omega)$$
$$= 3.033\ V \quad \text{(EQ 10.42)}$$

which is too low.

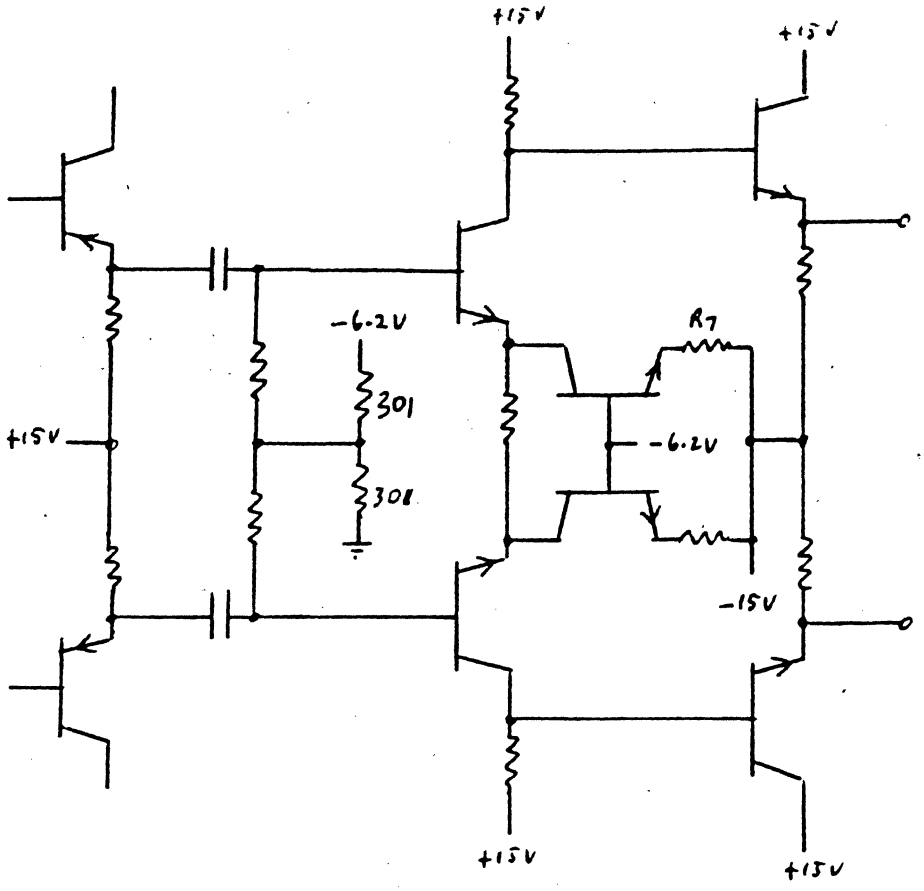


FIG 10.8 B

MODIFIED STAGE 2-3 COUPLING DUE TO BIAS PROBLEMS OCCASIONED BY CURRENT REQUIRED FOR LINEARITY.

TOTAL AMPLIFICATION (Continued)

. . . . . Perhaps we can reduce the tolerance on current I_{S3} from

33.86 ma → 45.7 ma = $\triangle$11.84 ma (EQ 10.47)

The best solution is to AC couple the bases of stage 3 to remove the almost 4 volt tolerance. We can return the base resistors to a nominal -3.1V by dividing the -6.2V supply. This will permit all stages to be totally linear. The τ of the coupling stage must take into account the base current associated with the 33- 45 ma.

Now we have designed a three-stage, high-level amplifier. Before we add filters and phase compensate it let us turn our attention to an AGC stage.

obvious if we consider that any common mode voltage influences the coupling circuits and depending on the common mode rejection ratio of the following amplifier we end up with a differential voltage change that disturbes the signal base line. As there is almost always a non-linearity somewhere we should avoid circuits that control the gain by controlling emitter current.

In Figures 10.9A thru 10.9F, the gain control is achieved by a contolled resistance by either current or voltage. In each case the range of resistance is large but the input swing is limited due to the characteristics of the devices. In the case of diodes, we can examine the V_D-I_D curves. Here we see that the signal voltage will be superposed on the curve which does affect the resistance instantaneously; therefore, the actual diode resistance is a function of the signal voltage as well as the control current thru R_2. Fortunately, our signal is differential. When one diode is conducting more due to a positive going signal, the opposite diode is conducting less for the same reason which if we keep the swing small the total resistance, differential, remains almost constant. The input swing then should be kept below 100.mv MAX $_{PP\ DIFF}$. The circuit of Fig. 10.9A is driven by a voltage source therefore the attenuation is simply

$$\alpha = \frac{R_D}{R_1 + R_D} \qquad \text{(EQ 10.51)}$$

$$I_{CONTROL} = \frac{V_{CONTROL} - V_D}{R_2} \qquad \text{(EQ 10.52)}$$

AGC STAGES

We approach this design problem by first calculating the total $gain_{MAX}$ required for a minimum signal at the head, minimum pre amplifier gain, minimum linear amplifier gain to provide the maximum input to the Detector.

$$(V_{HEAD\ MIN})(A_{P.A.\ MIN})(A_{L.A.\ MIN}) = V_{IN\ DET\ MAX} \qquad \text{(EQ 10.48)}$$

This gain assures us of linearity margin and correct operation of the detector. The next number we need is the maximum signal output assuming the amplifier does not limit and using the maximum gain.

$$(V_{HEAD\ MAX})(A_{P.A.\ MAX})(A_{L.A.\ MAX}) = V_{O\ MAX\ LINEAR} \qquad \text{(EQ 10.49)}$$

The amount of controllable attenuation required then is simply

$$\frac{V_{IN\ DET\ MAX}}{V_{O\ MAX\ LINEAR}} = \alpha_{MIN} \qquad \text{(EQ 10.50)}$$

If this number is greater than 0.5 then we really do not need AGC as the Detector dynamic range will handle it if the gain is lowered to make EQ 10.48 = VIN DET MIN instead. Assume α = .10 then we need an attentuator with at least a 10:1 range. The type of attenuator depends on the signal amplitude and on the signal bandwidth. Figure 10.9 shows several types that have been used. Contrary to the radio business, our AGC circuits must not introduce a common mode voltage change. The reason for this is

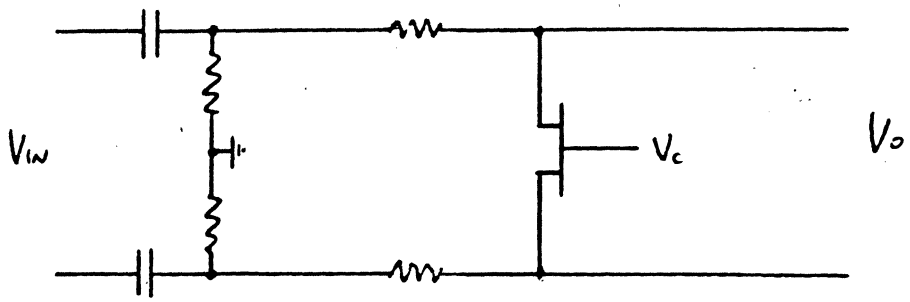


FIG 10.9C

FET RESISTANCE CONTROLLED ATTENUATOR

$V_o \leq 0.10 V_{pp}$ DIFF

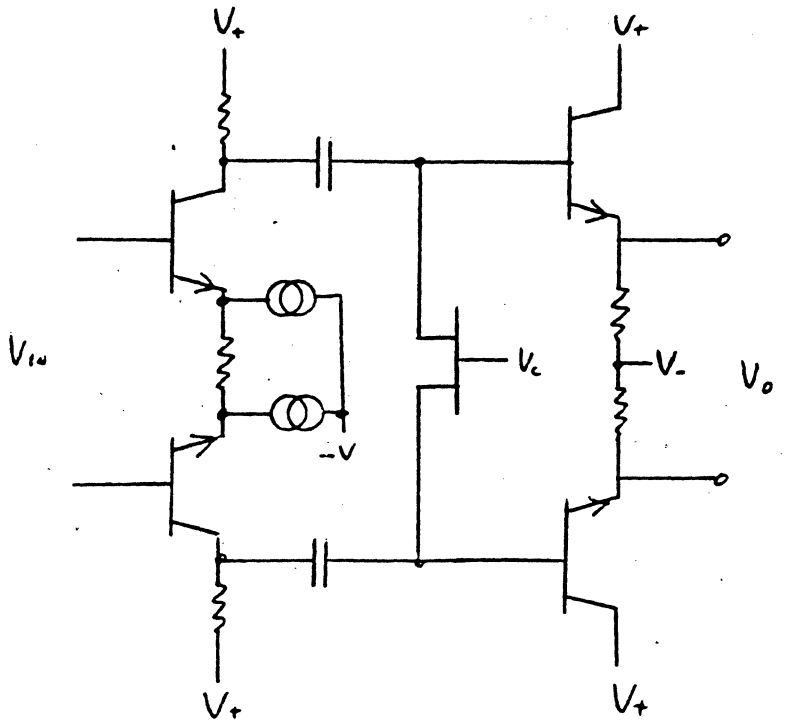


FIG 10.9D

FET RESISTANCE GAIN CONTROL

COLLECTOR STAGE

$V_o \leq 0.10 V_{pp}$ DIFF

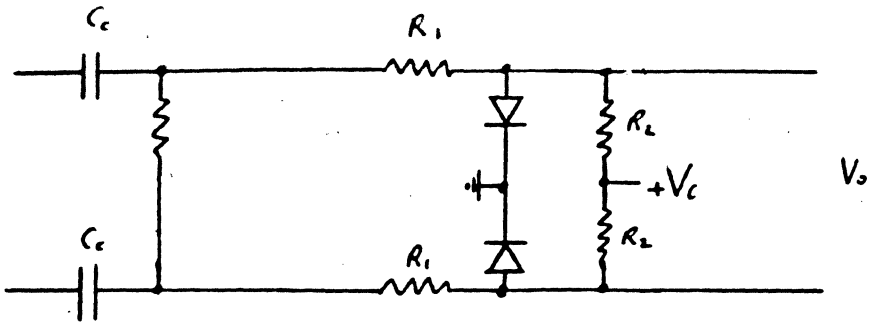


FIG 10.9 A

DIODE RESISTANCE CONTROLLED ATTENUATOR (SERIES)

$V_o \leq 0.10\ V_{pp}$

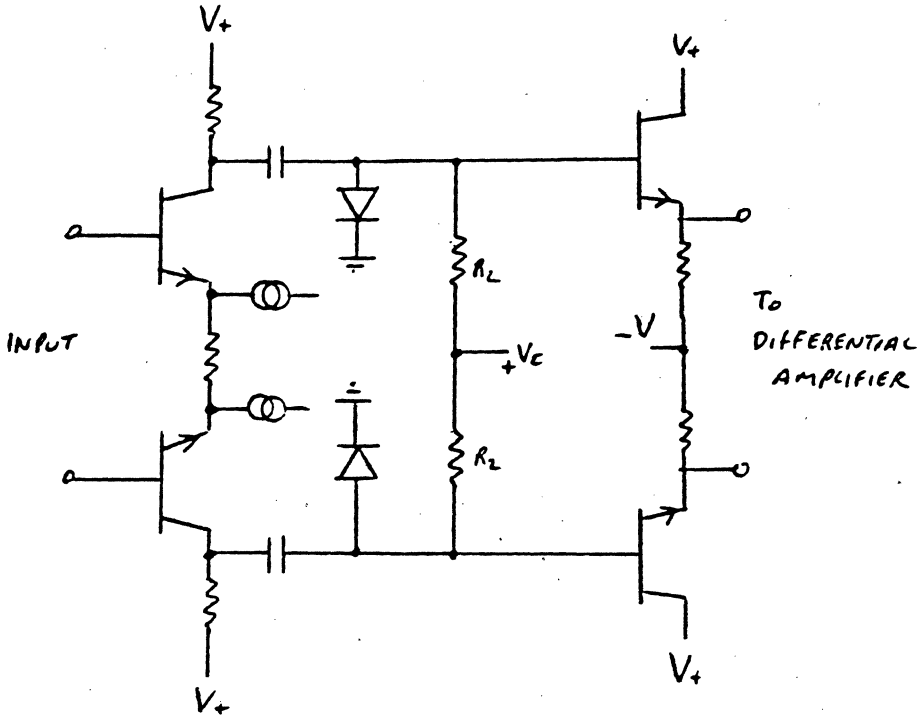


FIG 10.9 B

DIODE RESISTANCE CONTROLLED GAIN

$V_o \leq 0.10\ V_{pp}$

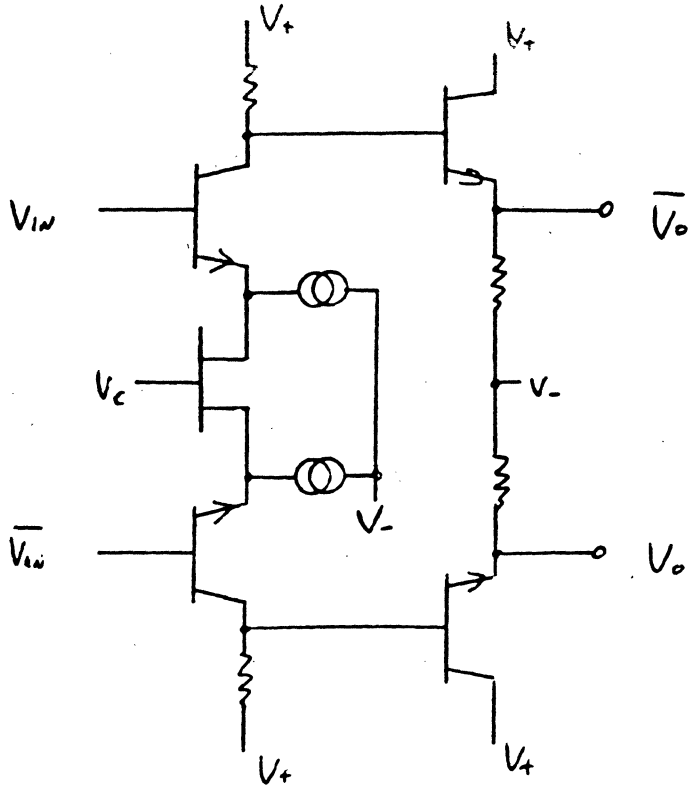


FIG 10.9 E

FET GAIN CONTROL IN EMITTERS

$V_{IN} \leq 0.10\,V_{PP}$ DIFF

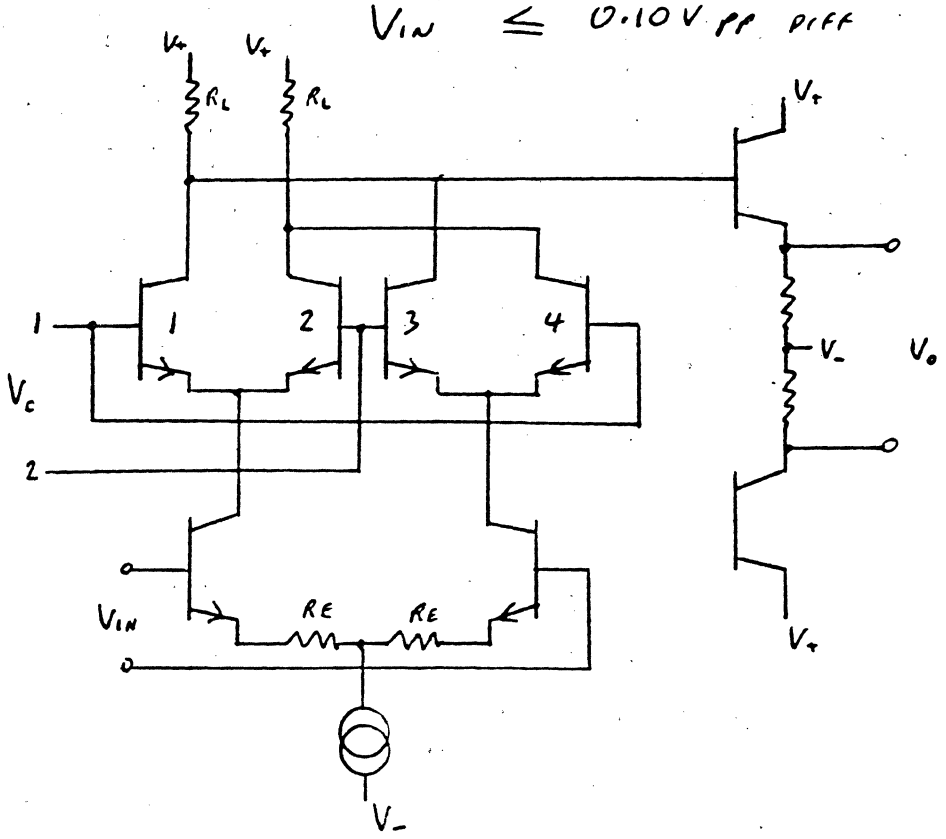


FIG 10.9 E

BASIC MULTIPLIER GAIN CONTROL

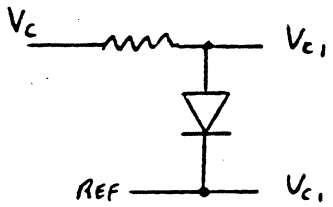


FIG 10.9 F

QUADRANT CONTROL

In figure 10.9B the attenuation circuit is driven by a current source. Here the gain of the stage is a function of the parallel combination of R_1 , R_2 and R_D

$$A = \frac{R_1 R_2 R_D}{\left(R_E + r_e + R_m\right)\left(R_1 R_2 + R_1 R_D + R_2 R_D\right)} \qquad \text{(EQ 10.53)}$$

In both cases there is a common mode output voltage change that affects the output base line. Both circuits must be followed by a very good common mode rejection amplifier. The diode capacitance must also be considered as it affects bandwidth which will change as a function of the control current τ changes.

In Figure 10.9C and D a Fet is used as the controlling resistor. The equations are the same as for the diode versions, however, the controlled resistor is a function of voltage. If we examine the Fet curves we again find a signal swing restriction. As long as the drain to source voltage remains below about 100 mV PP DIFF then we remain in the resistive portion of the curve. Beyond that the resistance is pinched and we go into a current source mode where the resistance is very high thus distorting the signal waveform. We still have a common mode problem due to the gate signal being capacitively coupled into the source and drain that may not be common mode. Also the gate capacitance affects the bandwidth which is changed by the changing resistance $\left(\tau = R_{DS(V)} C_{DG}\right)$

In Figure 10.9E a different type of attenuator is shown. These circuits are multipliers and care must be used in predetermining which quadrants are used. With careful balancing these circuits can be made to exhibit no change

in output DC or common mode voltage as a function of the control voltage. As this circuit is basically a 4 quadrant multiplier by biasing V_{C1} to be always equal to or more positive than V_{C2} only two quadrants are used. The circuit functions by mixing the two 180° out of phase signals such that one subtracts from the other resulting in reduced amplitude. The reason we must confine ourselves to only two quadrants is that the gain slope changes with the control voltage polarity as shown in Figure 10.10. The reversal would cause a malfunction of the AGC closed loop operation. Because of the signal subtraction process very careful balancing and phase control must be used in the signal path.

The DC collector voltage level is maintained by causing the current lost on one side to be made up by current from the opposite side as the control voltage is varied.

The gain of the circuit is a function of the control voltage and the balance within the circuit

$$A = \frac{2 R_L K V_c}{R_E + 2 r_e + 2 R_m} \qquad \text{(EQ 10.54)}$$

Where K is a constant depending on the matching of the diode, the transistors emitter-base diode, and resistor R4.

This circuit has a constant bandwidth only if it is correctly balanced. Any unbalance will cause unequal phase delays, therefore, altering the bandwidth as a function of control voltage.

Of the three different types given here, lets choose the FET version.

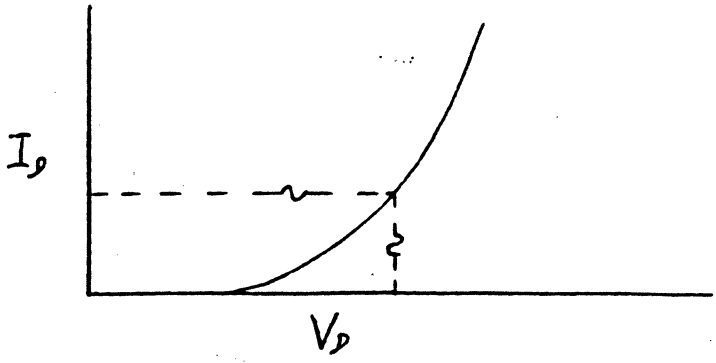


FIG 10.9 H

DIODE TRANSFER CURVE

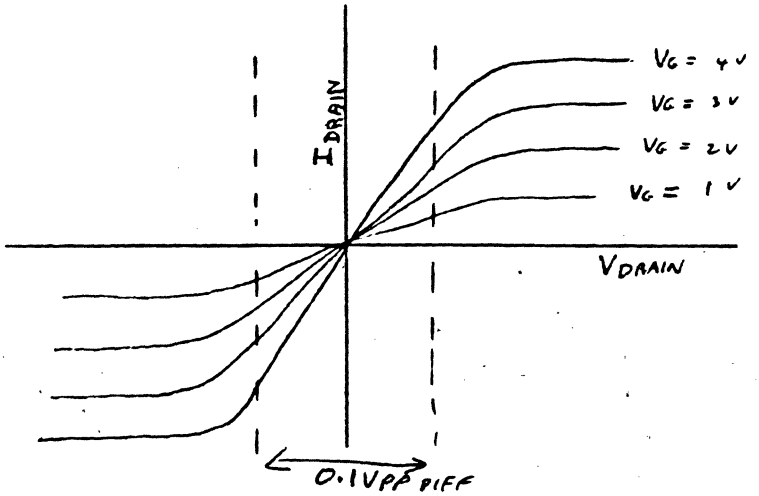


FIG 10.9 I

FET TRANSFER CURVES

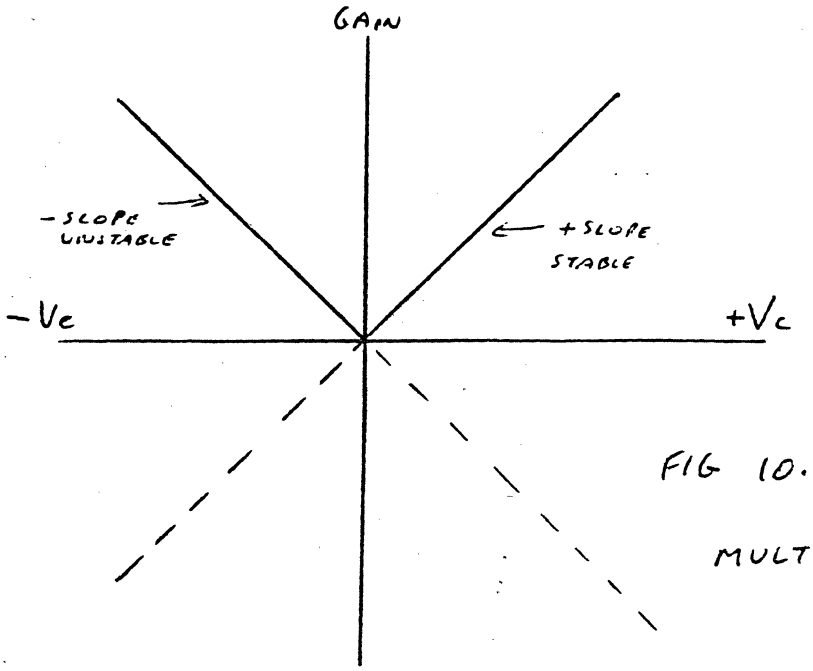


FIG 10.9 J

MULTIPLIER TRANSFER CURVES

If $R_{DSON\ MAX}$ were 100Ω then the maximum attenuation achievable is, from EQ 10.55

$$\frac{100\Omega}{454\Omega + 100\Omega} = 0.18 \quad \alpha = -14.8\ dB \qquad \text{(EQ 10.59)}$$

TO COVER WORSE CASE TOLERANCES ECT, WE SHOULD INCREASE α TO 10 OR MORE

To get an attenuation of 10 we need two in isolated series. We could then lower the R_1 to keep the same attenuation while maintaining above 50 MHz bandwidth.

$$\frac{1}{\sqrt{10}} = \frac{1}{3.162} = 0.3162 \qquad \text{(EQ 10.60)}$$

$$R_{1\ MIN} = 100\Omega\left(\frac{1 - 0.3162}{0.3162}\right) = 216.2\ \Omega\ MIN \qquad \text{(EQ 10.61)}$$

$$R_{1\ MAX} = \frac{1}{(2\pi)(5\times10^{7})(3\times10^{-12} + 4\times10^{-12})\sqrt{2}} = 321.5\Omega \qquad \text{(EQ 10.62)}$$

As the stage bandwidth calls for $\sqrt{2}$ (50 MHz) then choose 300 Ω for better dynamic range.

The isolation can be obtained with an emitter follower or an intervening gain of 3.162, thus maintaining the signal-to-noise ratio as much as possible. The circuit is shown in Figure 10.11. Transient coupling recovery can be added immediately following the first coupling capacitors as shown in Figure 10.3A or B or it can be added following both coupling PAIRS OF capacitors if needed.

ere are several other considerations. If a junction FET is used, care st be used to see that the gate circuit is not forward biased. This is sponsible for the capacitive coupling in the two examples shown. We could liminate the capacitors of Figure 10.9D and use a MOS FET as long as the C difference is zero. If not then currents will flow altering the DC uiescent point causing a differential shift in the output. A series apacitor in either the drain or source lead will eliminate the problem.

To use the circuit we must first determine the amount of attenuation equired. If the attenuation required results in a large resistor R_1 then he bandwidth degradation must be calculated at both extremes.

If this is intolerable then the attenuator must be broken up into two stages with some gain in between if necessary to keep the signal to noise ratio high. In any event isolation prevents interaction.

From Fig 10.9C

$$\alpha_{MAX} = \frac{R_{DS\ ON\ MAX}}{R_{1\ MIN} + R_{DS\ ON\ MAX}} \qquad \text{(EQ 10.55)}$$

$$R_{1\ MIN} = R_{DS\ ON\ MAX}\left(\frac{1-\alpha}{\alpha}\right) \qquad \text{(EQ 10.56)}$$

$$F_2 = \frac{1}{2\pi\ T} = \frac{1}{2\pi\ R_{1\ MAX}\ C_{D\text{-}S}} \qquad \text{(EQ 10.57)}$$

If in our example we want a 50 MHz bandwidth and $C_{D\text{-}S}$ is 4.PF then we want

$$R_1 < \frac{1}{2\pi\ F_2\left(C_{DS} + C_\omega\right)} = \frac{1}{(2\pi)(5\times10^{7})(3\times10^{-12}+4\times10^{-12})} \qquad \text{(EQ 10.58)}$$

$$= 454\ \Omega$$

LINEAR AMPLIFIERS

We could have made the second stage of attenuation like that shown in Figure 10.9D with similar results but we would need to calculate the attenuation again as it now involves R4 as well. THE FET GATE OVAS would ALSO BE DIFFERENT One of the advantages of the series type of attenuator circuits is that the voltage across the FET can be maintained below the 100 mv pp max while the input can exceed it. Lets look at the maximum signal levels as we maintain the 100 mv limit:

$V_{FET\ MAX} = 100\ mV$ AS STATED

$\alpha_{MIN} = \frac{R_{ds\ on}}{R_{SERIES} + R_{ds\ on}} = \frac{100}{300+100} = 0.25$ (CHANGED α VALUE)

$V_{IN_1\ MAX} = \frac{100.\ mV}{\alpha} = \frac{100.\ mV}{\frac{100\ \Omega}{300+100\ \Omega}} = 400.\ mV$ (EQ 10.63)

WITH A PRECEEDING GAIN OF $\frac{1}{\alpha} = 4$ AND A 100 mV MAX VOLTAGE ACROSS THE FIRST FET, V_{IN_2} WOULD BE

$V_{IN_2} = A(100.\ mV) = (4.0)(100\ mV) = 400\ mV$ (EQ 10.64)

$V_o = V_{IN_2}(\alpha) = (400.\ mV)(0.25) = 100.\ mV$ (EQ 10.65)

which maintains the output FET voltage at its maximum. The minimum input signal occurs when the FET's are just off while maintaining the 100 mv output FET voltage.

$V_{IN_{MIN}} = \frac{100.\ mV}{A} = \frac{100.\ mV}{4.0} = 25.\ mV$ (EQ 10.66)

Therefore the input range under AGC control would be 25 to 400 mv or 16 : 1 which satisfies both bandwidth and our required attenuation.

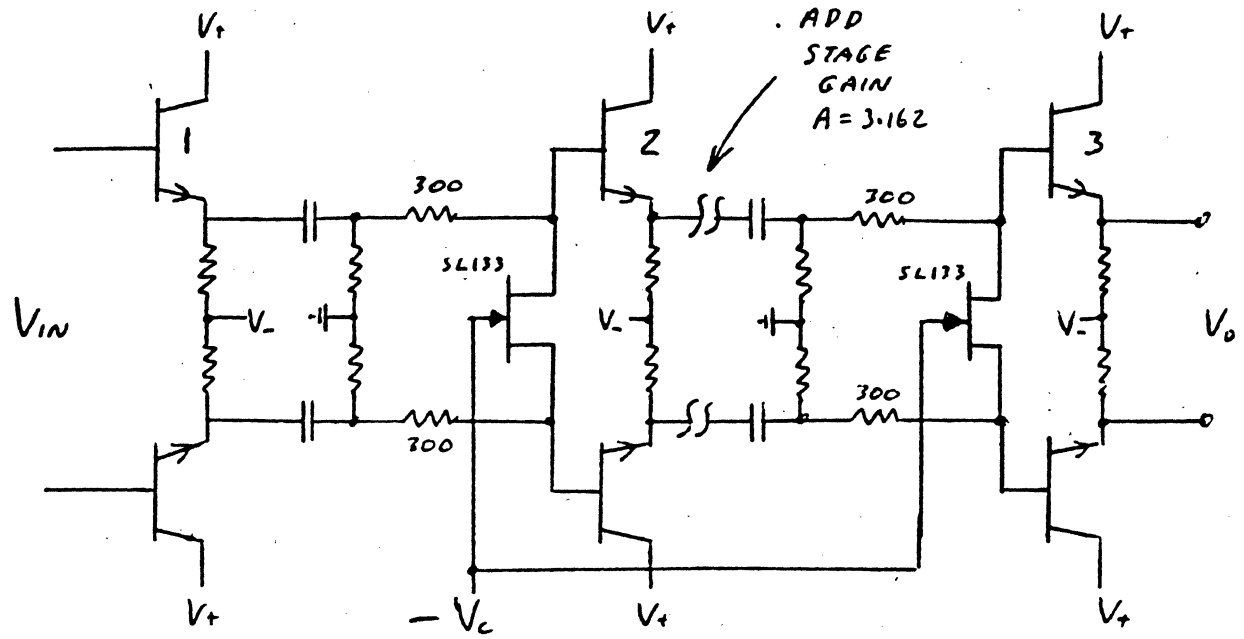


FIG 10.11

ISOLATED ATTENUATORS $\alpha = \alpha_1 \alpha_2$

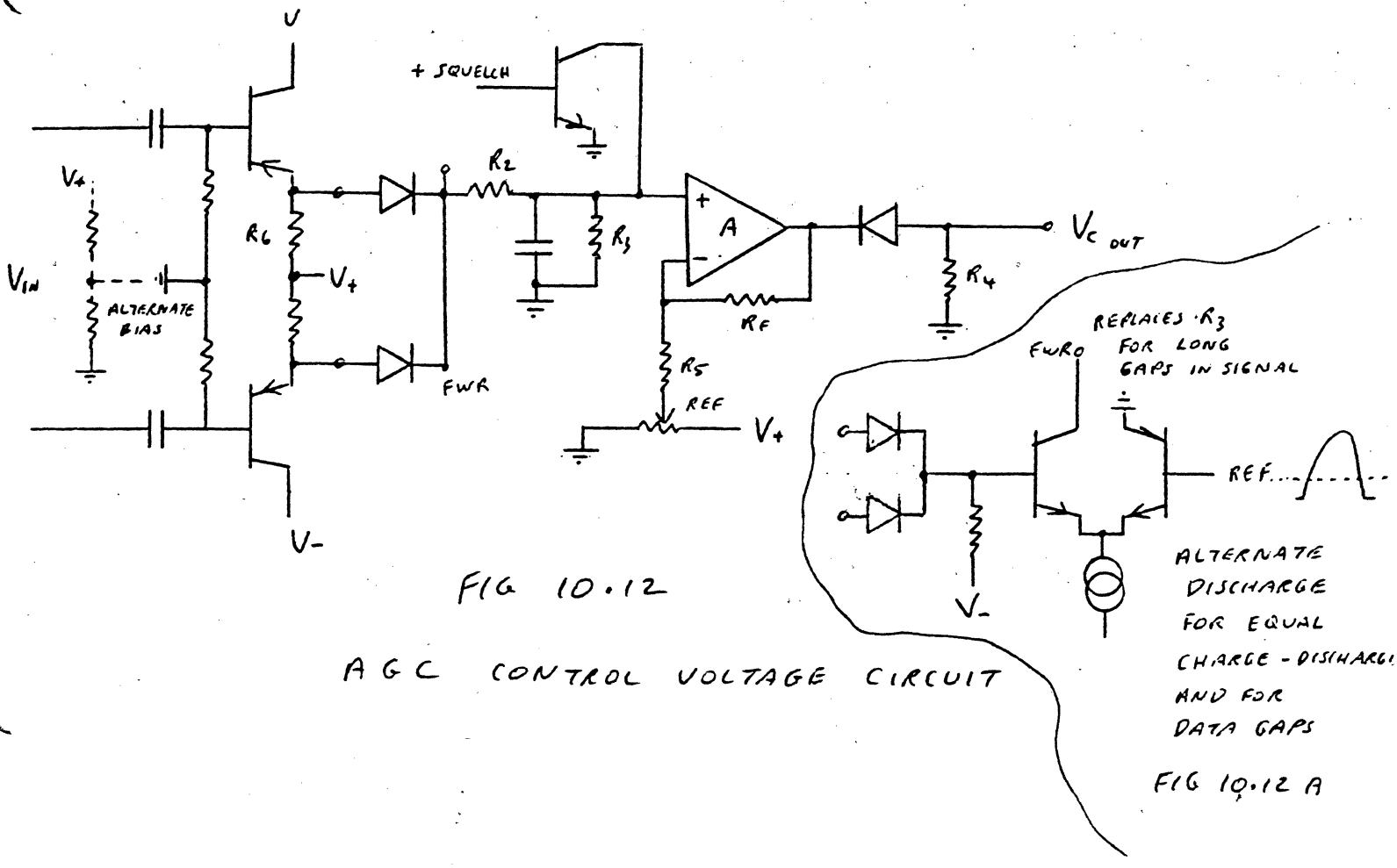


FIG 10.12

AGC CONTROL VOLTAGE CIRCUIT

FIG 10.12 A

$$\tau_{Attack} = \frac{(Z_o + Z_D + R_2) R_3 C}{Z_o + Z_D + R_2 + R_3} \qquad \text{(EQ 10.69)}$$

And the decay τ is

$$\tau_{DECAY} = R_3 C$$

if the input current to the Amplifier is very small and the diode leakage current is small also. The diode minority carry lifetime does affect the decay τ.

The actual gain required is (EQ 10.70)

$$A = \frac{(V_{Pinch\ off\ max})}{\Delta V_{sig}} \left(\frac{Z_o + Z_D + R_2 + R_3}{(Z_o + Z_D + R_2) R_3} \right)$$

We should now turn our attention to the temperature affects since our following gain A is so high. Notice that we used a pair of PNP emitter followers to drive our Full Wave Rectifier. The base emitter diode nearly compensates for the rectifier diodes, but not completely due to the large difference in currents caused temperature. Also the current thru R1 must be large compared to the current thru R2 in order to maintain PNP emitter follower linearity. This means that the temperature of the PNP transistors is higher than the surrounding components therefore its Vbe will be less and the voltage into the operational amplifier will be more negative.

This requires a divider in the return ground lead from R6 shown dotted in Figure 10.12. Or we can change the PNP emitter followers to NPN and use a Vbe multiplier to compensate, as shown in Figure 10.13, for both the two

CLOSED LOOP AGC

Before we close the control loop for the AGC circuits, we must develop a DC voltage that is a function of the output of the linear amplifier. This can be obtained from a filtered full wave rectifier such as shown in Figure 10.12. There are several features of this circuit that need to be discussed since there are many other ways this could be implemented.

The gain control desired is a function of the gain in the loop. If we desire a 1% output voltage variation then the peak-to-peak single ended signal will vary 0.5% Vpp Diff and the Base to Peak rectified signal will vary 0.25% Vpp Diff. For our 7.5 Vpp Differential output signal nominal this means that we need to have

$$\Delta V_o = \left(V_{pp\ DIFF}\right)\left(\frac{\%\ control}{100\ (4)}\right) = \left(7.5\ V_{PP\ DIFF}\right)\frac{1\%}{100\ (4)} = 18.75\ mV \quad \text{(EQ 10.67)}$$

to control the full range of attenuation. If our FETs pinch off voltage is -5.0V MAX then we need a gain of

$$A_{MIN} = \frac{V_{PINCH\ OFF\ MAX}}{\Delta V_o} = \frac{5.0\ V}{18.75\ mV} = 266.66 \quad \text{(EQ 10.68)}$$

This would be the case if resistor R2 were zero but there is an attenuator formed by R2 and R3 which causes us to raise this gain. Now R2 is there in order to slow down the the attack of the AGC to a sudden increase in signal amplitude. This is very desirable for two reasons. First we do not want to respond to noise caused amplitude variations and second, it permits us to achieve stability of the closed loop circuit using the Nyquist criteria. The attack τ is

junctions as well as the temperature difference. For example if we need to compensate two diode junctions at a difference of $10^{o}C$, we need to provide an additional $(10^{o})(2\ mv/o_{C}) = 20.mv$ correction. This is interesting as it is nearly the same as the control signal range of 18.74 mv which emphasizes the point. We still have V supply variations to contend with or if we stabilize it with a zener we need to concern ourselves with the zener temperature behavior as well as its zener impedance.

The last output diode is inserted to protect the junction FET (Fig. 10-11) from positive excursions which would forward bias its Gate junction. A MOS FET would not need the diode. If we used enhancement mode FETs we would need to reverse the polarity. The potentiometer or a fixed resistive divider is added to the negative input to adjust for the charged capacitor signal amplitude. That value can be calculated from the following for a PNP emitter follower.

$$V_{REF} = V_{SIG\ BP\ SE} + V_{be_{PNP}} - V_{D} + V_{comp} \qquad \text{(EQ 10.71)}$$

Note that this will be very broad due to tolerances of the two junctions which justifies the potentiometer. The squelch transistor is added to discharge the AGC capacitor at the beginning of a read function following the selection transient, thus reducing the time to discharge from the transient using the discharge τ.

The amplifier phasing allows for an N type J FET. If we put an attenuator stage ahead of our linear amplifier then we can close the loop.

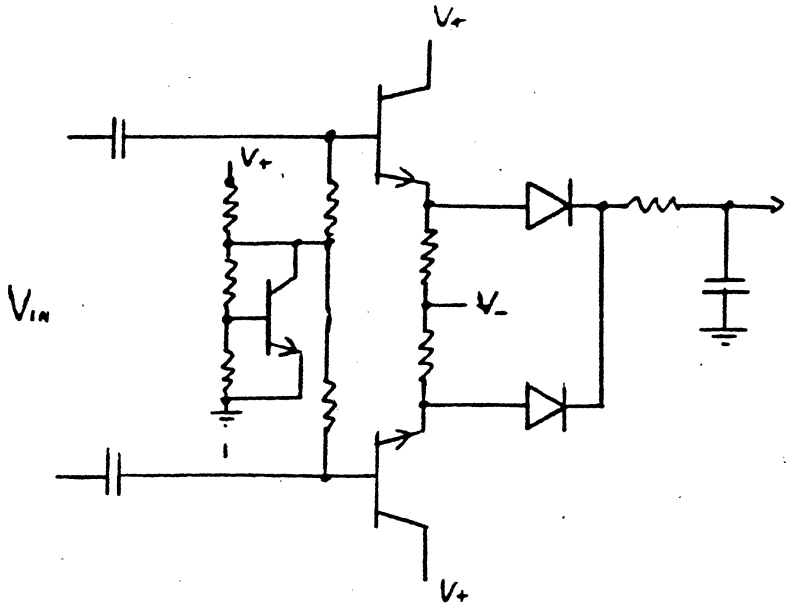


FIG 10.13

ALTERNATE BIAS, TEMPERATURE COMPENSATED.

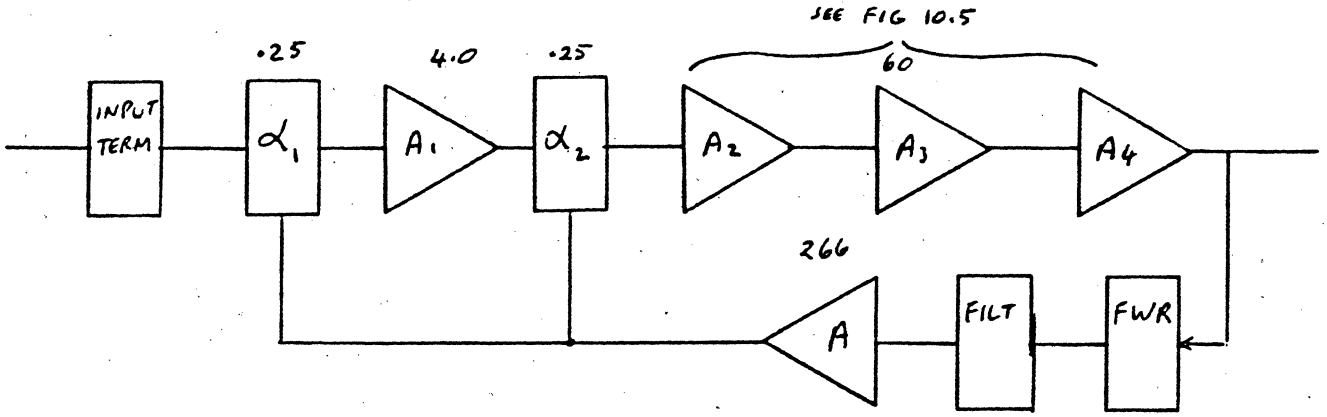


FIG 10.14 A

AGC AMPLIFIER BLOCK DIAGRAM

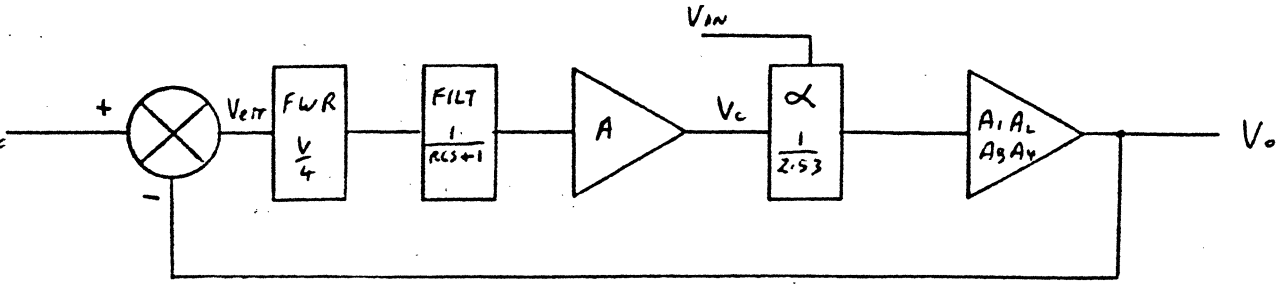


FIG 10.14 B

EQUIVILENT CIRCUIT FOR AGC

LINEAR AMPLIFIERS

FILTERS

Since our ability to determine the peak of the pulse resulting from a magnetic transition depends on differentiation then we need to concern ourselves with noise. The input stage including the head is the main source of this noise. Some of the noise is white, while the remainder is pink as it results from both the head impedance times the amplifier noise current plus any diode currents and the media noise. Particulate media is the main culprit. The frequencies of this latter noise falls in the bandpass of interest and beyond. We can improve the signal-to-noise ratio by filtering out that noise above the bandwidth of interest. We also know that the head signal contains harmonics which are required in order to maintain the signal PW_{50} and therefore resolution. For example, if we lost the 3rd harmonic then the PW_{50} would be widened, and the resolution would drop, and the voltage time rate of change at the peak would be lessened giving poorer peak detection. The filter roll off characteristics then are important to us. There are several different filter types that we could choose from besides the constant K and M derived types. The best candidates are the Butterworth, Butterworth Thompson, and the Bessel. The Chelbyshev has ripple in both phase and gain, therefore, is useless to us unless we want to use the ripple as some kind of correction for existing anomalies. The Butterworth has very desirable amplitude characteristics which are maximally flat in the pass band and roll off with a well-defined corner depending on the number of elements. The Bessel filter has a very long drawn out roll off which does affect the amplitude of frequencies somewhat removed from the poor corner. The Butterworth Thompson is a

$$\alpha = \frac{1}{12.65} = 0.079$$

$$V_{REF} = \frac{7.5\,V_{PP\ DIFF}}{4} \qquad \text{(EQ 10.72)}$$

$$V_{err} = 4\,V_{REF} - V_o \qquad \text{(EQ 10.73)}$$

$$V_o = V_{IN}\left(V_{CONTROL}\,\alpha\right)(60)(3.162) \qquad \text{(EQ 10.74)}$$

$$V_{CONTROL} = V_{err}\left(\frac{1}{4}\right)\left(\frac{1}{RCs+1}\right)(266.6) \qquad \text{(EQ 10.75)}$$

$$V_{err} = 4\,V_{REF} - V_{IN}\left(V_{err}\right)\left(\frac{1}{4}\right)\left(\frac{1}{RCs+1}\right)(266.6)\left(\frac{1}{12.65}\right)\left(\frac{1}{5.0}\right)(60)(3.162) \qquad \text{(EQ 10.76)}$$

$$V_{err} = \frac{4\,(7.5\,V_{PP\ DIFF})}{\left(1 + V_{IN}\left(\frac{1}{4}\right)\left(\frac{1}{RCs+1}\right)(266.6)\left(\frac{1}{12.65}\right)\left(\frac{1}{5}\right)(60)(3.162)\right.} \qquad \text{(EQ 10.77)}$$

$$V_{err} = \frac{30}{1 + V_{IN}\,(199.98)\left(\frac{1}{RCs+1}\right)} \qquad \text{(EQ 10.78)}$$

$$V_o = V_{IN}\left(V_{err}\right)\left(\frac{1}{4}\right)\left(\frac{1}{RCs+1}\right)(266.6)\left(\frac{1}{12.65}\right)\left(\frac{1}{5.0}\right)(60)(3.162) \qquad \text{(EQ 10.79)}$$

$$= V_{IN}\left(V_{err}\right)(999.59)\left(\frac{1}{RCs+1}\right) \qquad \text{(EQ 10.80)}$$

$$V_o = \frac{(V_{err})(199.918)\,V_{in}}{RCs+1} = \frac{V_{err}\left(1.99918\times10^{2}\right)V_{IN}}{RCs+1} \qquad \text{(EQ 10.81)}$$

$$= \left[\frac{30\ \ V_{in}}{1 + V_{IN}\left(199.918\right)\left(\frac{1}{RCs+1}\right)}\right]\left[\frac{1.9918\times10^{2}}{RCs+1}\right] \qquad \text{(EQ 10.82)}$$

$$V_o = \frac{30\left(1.9918\times10^{2}\right)V_{in}}{RCs+1 + V_{IN}\left(199.918\right)} \qquad \text{(EQ 10.83)}$$

not correct

which is unconditionally stable

$$V_o = \frac{5.9975\times10^{3}}{RCs+1+1.99918\times10^{2}} = \frac{5.9975\times10^{3}}{RCs+20.9918} \qquad \text{(EQ 10.84)}$$

(too crowded)

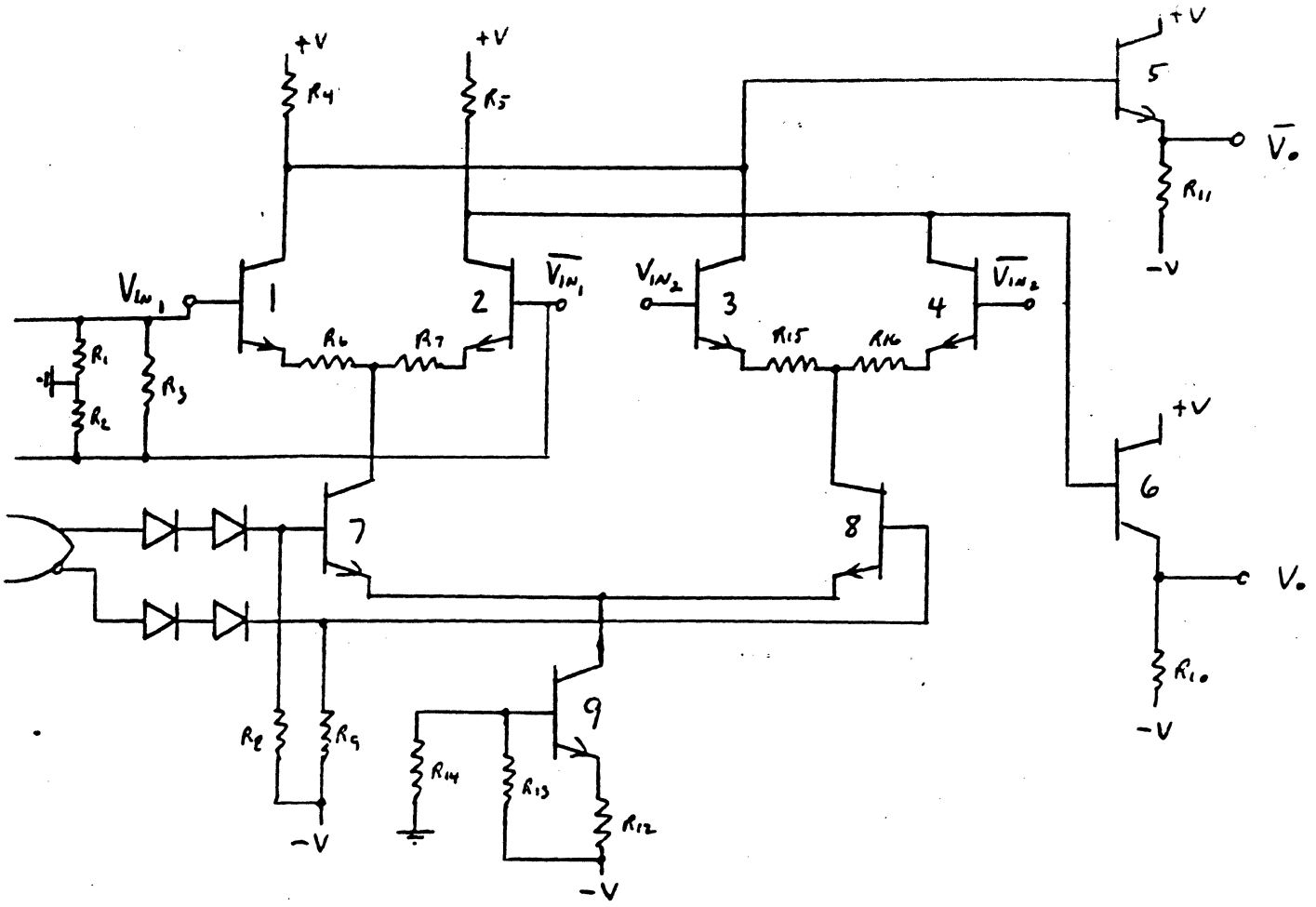


FIG 10.4 A

TWO CHANNEL SELECTOR FOR LOW LEVEL SIGNALS (100mV)

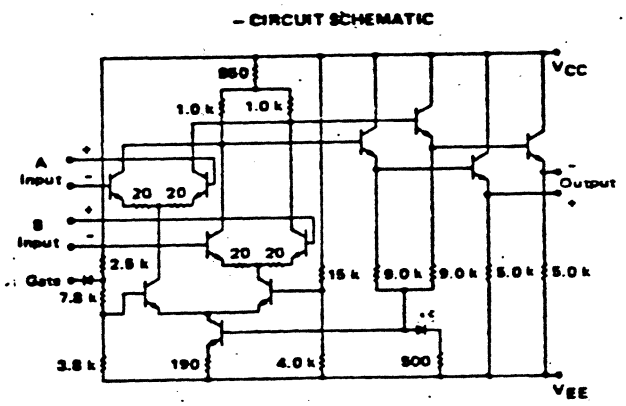


FIG 10.4 B

COMMERCIAL TWO CHANNEL SELECTOR FOR LOW LEVEL SIGNALS (100mV)
MC 1445

compromise between the Butterworth and the Bessel filter. In regards to Phase and Group Delay the three filters are rated differently. The Bessel filter has maximally flat group delay and the Butterworth doesn't. The Butterworth Thompson is again a compromise. Now it is obvious that flat magnitude characteristics are desirable due to the relationship among the pertinent harmonics. The Maximally flat phase and group delay characteristics are not so obvious. If we were to take a fundamental cosine wave and add to it a third harmonic such that the peak of both start together as in Fig. 10.27,

$$V = A\cos\omega t + B\cos 3\omega t \qquad \text{(EQ 10.86)}$$

then we will obtain a waveform very similar to our head signal in Region 2 containing shoulders. Now if we were to repeat our graphical analysis with the third harmonic shifted by a constant angle ϕ then we can

$$V = A\cos\omega t + B\cos(3\omega t + \phi) \qquad \text{(EQ 10.87)}$$

see there is peak and shoulder distortion. The peak distortion includes both amplitude and peak position Fig. 10.16. Now our main concern is the peak as it defines the center of the bit or transition; therefore, if we cause unequal phase delay, then we lose peak timing information accuracy. As can be seen from Figure 10.15, if our filter introduces amplitude reduction of the third harmonic, then the signal PW_{50} widens and if our filter introduces unequal group delay then we have peak shift.

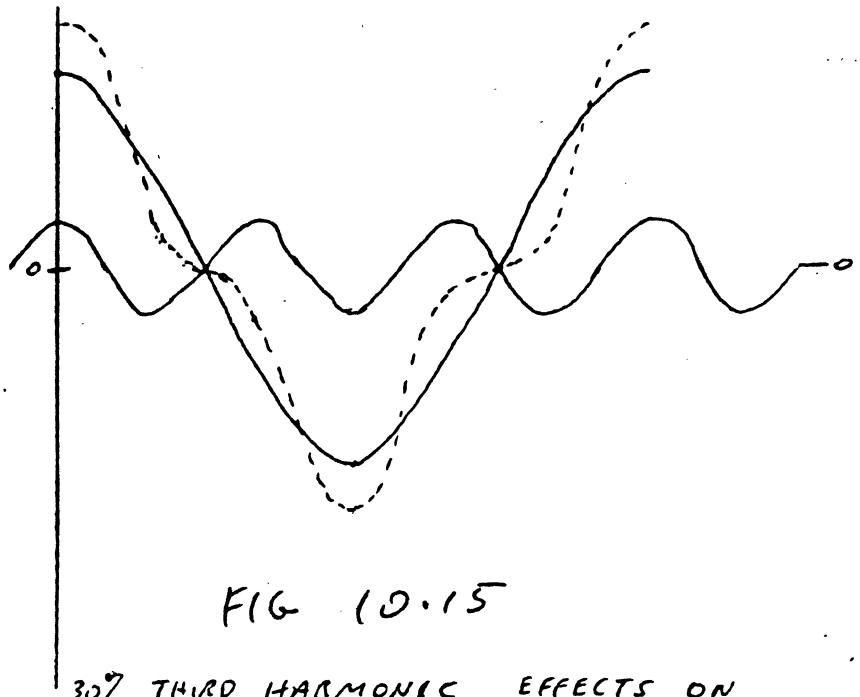


FIG 10.15

30% THIRD HARMONIC EFFECTS ON FUNDAMENTAL (SYMETRICAL SHOULDERS)

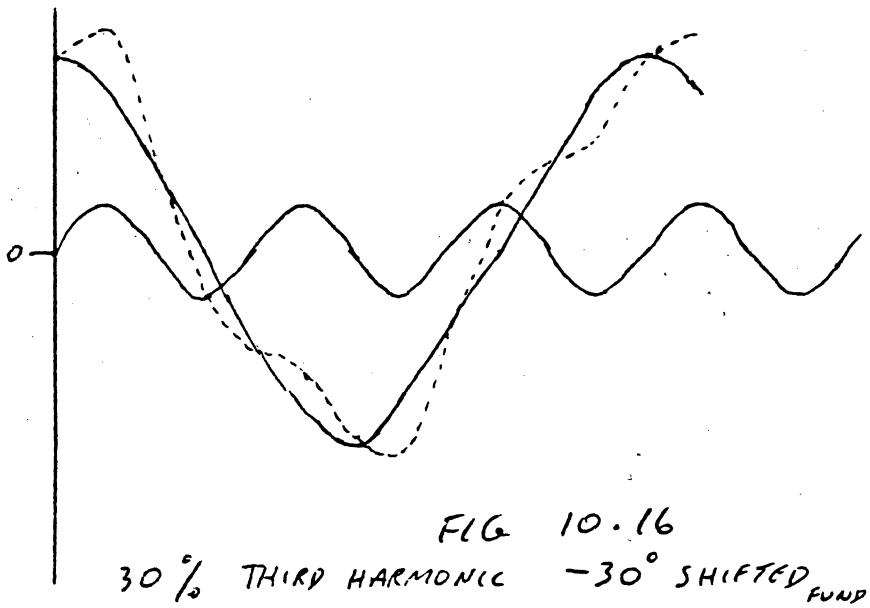


FIG 10.16

30% THIRD HARMONIC -30° SHIFTED$_{FUND}$ EEFECTS ON FUNDAMENTAL (NON SYMETRICAL SHOULDERS)

It is obvious that the Butterworth filter preserves the amplitude but distorts the group delay. The Bessel filter will widen the PW50 a little but maintains the peak in position. Most filters designed for disc drives use the Butterworth filter with a phase correcting filter in series. Some use the Bessel but wonder why the shoulders climb up the waveform as shown in Fig. 10.16. The answer lies not so much in the amplifier but in the head. If we go back to Chapter 3 where we discussed the head circuit and Chapter 6 where we discussed the Read Circuit we can see that the head is a two pole filter as shown again in Figure 10.17.

The output voltage is determined from the series paralleled network.

$$V_o = \frac{V_{IN}\left(\frac{\frac{1}{CS}R}{\frac{1}{CS}+R}\right)}{LS+\left(\frac{\frac{1}{CS}R}{\frac{1}{CS}+R}\right)} = \frac{V_{IN}\left(\frac{1}{CS}R\right)}{RLS+\frac{LS}{CS}+\frac{R}{CS}} \qquad \text{(EQ 10.88)}$$

$$V_o = \frac{V_{IN}\left(\frac{1}{LC}\right)}{S^2+\frac{1}{RC}S+\frac{1}{LC}} = \frac{V_{IN}\ \omega_n^2}{S^2+2\zeta\omega_n S+\omega_n^2} \qquad \text{(EQ 10.89)}$$

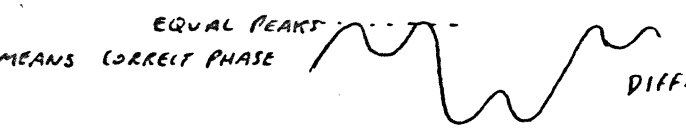
The phase characteristics of this circuit are not linear, or maximally flat group delay, therefore, phase distortion is added to the head signal. When we design filters to provide the characteristics we need, the head circuit forces a different compromise. The use of phase correcting filters allows use of the Butterworth filter without degradation of the PW50 or the peak position. There are other approaches that are presently being pursued which involve spectral shaping which narrow the PW50 while maintaining the the peak position. These approaches permit higher transition densities by

eliminating or greatly reducing the peak shift due to pulse crowding at a small cost of increased noise. Curves can be generated relating the improvements and degradation as a function of the degree of slimming. We will not pursue this form of filter here but it might be a worthwhile study as it has definite advantages. (Mr. D. Huber is very familiar with this approach.) The design of these filters has been made easy by several authors of Filter Synthesis books.

In chapter 13 of Louis Weinberg's Network Analysis and Synthesis published by McGraw Hill in 1962 and republished by Kreger Publishing Co. in 1975, he gives extensive tables for these and other filters as either conventional filters and as equal dissipation filters. An analysis of the various filter characteristics by Eggen and McAllister is published in Electro Technology, August 1966.

The Phase correction filters or Phase equalizers are the subject of several texts. Chapter 17 of Electronic Designers Handbook by Landee Davis and Albrecht published by McGraw-Hill, 1957, is a good source.

Because the head circuit is part of the total gain and phase response, the determination of the amount of phase correction required must be obtained from the signal itself rather than as input sine wave to the amplifier. There are two sources. The first is the position of the shoulders on the head signal. If they are symmetrical around the base line then the phase is correct. If the shoulders are not symmetrical but are above and below the baseline, then correction is required. The amount can be determined by the position of the shoulder compared to a graph, but this is rather sloppy as it neglects the phase distortion of the differentiator.



LINEAR AMPLIFIERS

The best method requires the use of a current passing near the gap of the head which generates a voltage according to the relationship $\frac{KND\phi}{dt}$.

The flux generated by the current in the wire is loosely coupled to the head core which causes a voltage to be developed in the head coil that can be amplified. The phase measurements at various frequencies can then be plotted if we remember to subtract the 90^0 associated with the flux to voltage conversion. The oscillator must be a true sine wave type with very low distortion. Function generators have substantial harmonics and cannot be used. The series resistor is equal to the Z_0 of the generator therefore I wire is

$$I_{wire} = \frac{V_{sig\ gen}}{Z_0} = \frac{V \sin \omega t}{Z_0} = K_1 \phi_{wire} \qquad \text{(EQ 10.90)}$$

$$V_{o\ head} = K_2 N K_1 \frac{dI}{dt} = K_1 N K_2 \frac{d\left(\frac{V \sin \omega t}{Z_0}\right)}{dt} \qquad \text{(EQ 10.91)}$$

$$= \frac{\omega N K_1 K_2 V \cos \omega t}{Z_0}$$

Care must be taken to keep track of the phase expected thru the amplifier stage by stage including the linear differentiator of the Detector. For constant group delay, the phase must be a direct function of frequency.

$$\Theta_{F_1} = \frac{F_1}{F_2} \Theta_{F_2} \qquad \text{(EQ 10.92)}$$

ANY FIXED DELAYS SHOULD BE SUBTRACTED FIRST. (CABLE OR DELAY LINE)

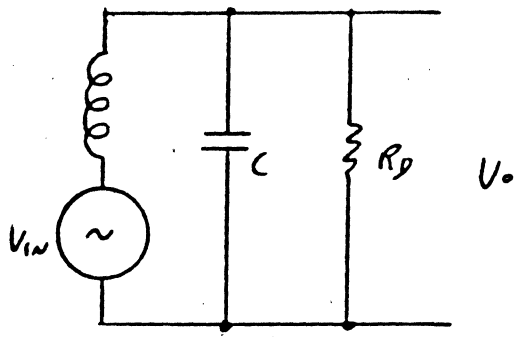


FIG 10.17

HEAD SIGNAL FILTER

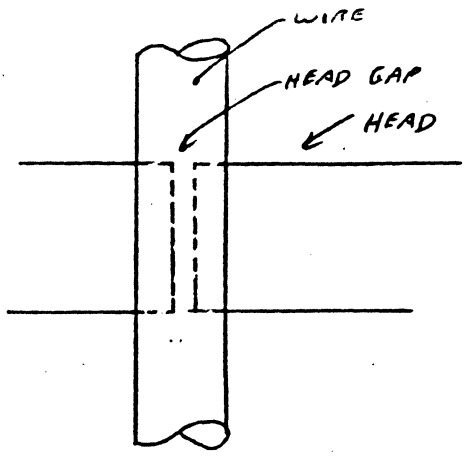


FIG 10.18B

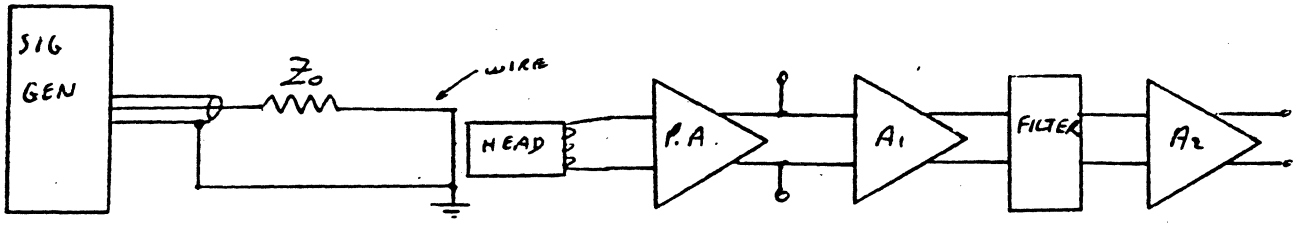


FIG 10.18A

TEST CIRCUIT FOR HEAD } RESPONSE AND PHASE LINEARITY FOR THE TOTAL AMPLIFIER AND FILTER.

FIG 10.19 TYPICAL WAVE FORM WITH POOR RESOLUTION F_1 AND F_2

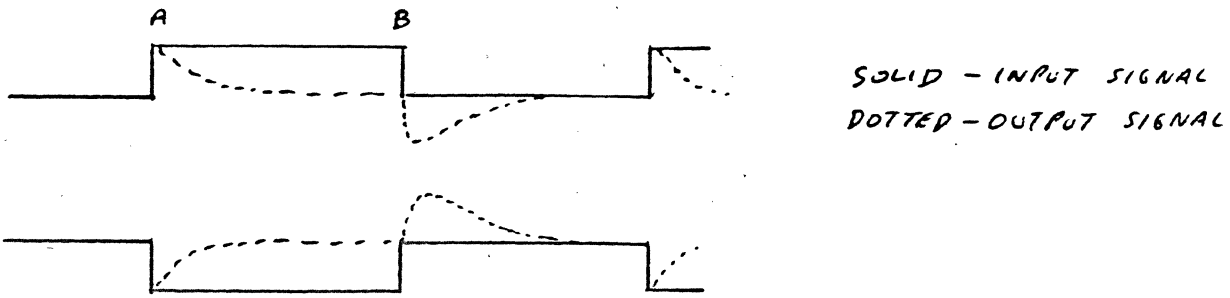


FIG 10.20 SIGNAL ENVELOPE WITH POOR RESOLUTION AFTER AGC ACTION

The design of the phase correction circuits must force compliance to EQ. 10.92 for all the frequencies possible with the code used, including the third harmonic. (OF ANY SIGNAL THAT EXHIBITS SHOULDERING)

Notice that the voltage output is an increasing function of frequency as shown in EQ. 10.91; therefore, care must be taken to maintain linearity. To plot the magnitude one must divide by F first. This will result in a very good check on the Read damping factor if the Bandwidth of the Pre Amplifier is wider than 10 times the self resonance of the head. In this case the plot must be taken at the output of the Pre Amplifier so as to not include the effects of the filters. Any series coupling capacitors must be taken into consideration.

The above measuring technique is very valuable and has been used for many years. If the amplifier bandwidth is less than 10 times $F_{RES_{HEAD}}$ then a graphical solution can be obtained if the gain and phase characteristics of the Pre Amplifier are known.

One last problem that can be discussed is the affect on the AGC circuits of a signal in Region 2. Here the various head signals have amplitudes as a function of frequency. If a signal was composed of a series string of groupings of frequencies that are wider than the $\frac{\tau}{5}$ of the AGC filter then we have introduced an amplitude modulation not present in the original signal. Consider the case of two frequencies, one at the 90% point on the BPI curve, Fig. 4.3, and the second at the 70% point as shown in Fig. 10.19, for a 20% amplitude difference. The AGC circuit on encountering the 90% amplitude signal will reduce the gain then on entering the area of the 70% signal will increase the gain again. The result is a signal with

double the amplitude modulation in the area just following the change. This is shown in Figure 10.20.

Obviously this is undesirable due to the low amplitude remaining just following the 90% to 70% change at B. Also the high amplitude at A will affect the linearity range required of the amplifier. (SHOWN dotted)

One solution to this dilemma is to provide a filter before the full wave rectifier, but not in the main signal path to the detector, that will correct the amplitude differences. Here phase distortion and amplitude distortion are of no concern only equality of peak amplitude. Going back to the resolution -- our signal will have a resolution of $\frac{70}{90}$ or 77.7%. We can introduce a frequency sensitive impedance in the emitter circuit to control the gain just as we did for the differentiator of Fig. 9.13A.

If we do, we can write some equations that relate resolution to the gain required.

$$\text{Resolution} = \frac{V_{HF}}{V_{LF}} = \frac{G_{LF}}{G_{HF}} = \frac{Z_{e\,HF}}{Z_{e\,LF}} \qquad \text{(EQ 10.93)}$$

$$\frac{Z_{e\,HF}}{Z_{e\,LF}} = \frac{\frac{X_{C\,HF}}{\sin \theta_{HF}}}{\frac{\frac{F_H}{F_L} X_{C\,HF}}{\sin \theta_{LF}}} = \frac{\frac{X_{C\,HF}}{\sin\left(\tan^{-1} \frac{X_{C\,HF}}{R}\right)}}{\frac{\frac{F_H}{F_L} X_{C\,HF}}{\sin\left(\tan^{-1} \frac{F_H X_{C\,HF}}{F_L R}\right)}} \qquad \text{(EQ 10.94)}$$

$$= \frac{F_H}{F_L}\left[\frac{\sin\left(\tan^{-1}\frac{F_H}{F_L}\frac{X_{C\,HF}}{R}\right)}{\sin\left(\tan^{-1}\frac{X_{C\,HF}}{R}\right)}\right] \qquad \text{(EQ 10.95)}$$

substituting $\sin\alpha = \frac{\tan\alpha}{\sqrt{1+\tan^2\alpha}}$ (EQ 10.96)

$$\frac{\frac{F_L}{F_H}\left[\frac{\frac{F_H}{F_L}\frac{X_{C\,HF}}{R}}{\sqrt{1+\frac{F_H}{F_L}\frac{X_{C\,HF}}{R}}}\right]}{\frac{\frac{X_{C\,HF}}{R}}{\sqrt{1+\left(\frac{X_{C\,HF}}{R}\right)^2}}} = \frac{\sqrt{1+\left(\frac{X_{C\,HF}}{R}\right)^2}}{\sqrt{1+\left(\frac{F_H}{F_L}\frac{X_{C\,HF}}{R}\right)^2}} \qquad \text{(EQ 10.97)}$$

$$\therefore \frac{Z_{HF}^2}{Z_{LF}^2} = Res^2 = \frac{1+\left(\frac{X_{C\,HF}}{R}\right)^2}{1+\left(\frac{F_H}{F_L}\frac{X_{C\,HF}}{R}\right)^2} \qquad \text{(EQ 10.98)}$$

$$Res^2\left[1+\left(\frac{F_H}{F_L}\frac{X_{C\,HF}}{R}\right)^2\right] = 1+\left(\frac{X_{C\,HF}}{R}\right)^2 \qquad \text{(EQ 10.99)}$$

$$1-Res^2 = \left(\left(\frac{F_H}{F_L}\right)^2 Res^2 - 1\right)\left(\frac{X_{C\,HF}}{R}\right)^2 \qquad \text{(EQ 10.100)}$$

$$\frac{X_{C\,HF}}{R} = \sqrt{\frac{1-Res^2}{\left(\frac{F_H}{F_L}\right)^2 Res^2 - 1}} \qquad \text{(EQ 10.101)}$$

From Equation 10.101 we can calculate the ratio of Xc to R at the highest frequency which in turn gives the required correction if that ratio is used to obtain

$$\Theta_{HF\ feedback} = \tan^{-1} \frac{X_{C\,HF}}{R} \qquad \text{(EQ 10.102)}$$

$$A_{corr_{HF}} = \frac{R_L \sin \Theta_{HF}}{X_{C\,HF}} \qquad \text{(EQ 10.103)}$$

The result then is an AGC system that does not introduce any modulation to the output waveform but retains the original resolution. One point of interest is that in disc drives where the resolution is a function of radius then the resolution must be taken from a compromise track between, but not necessarily half way between, the inner and outer radius where the resultant modulation has minimum effect.

We have now discussed the linear amplifier in which we included the Region of operation in our discussion as to the blocks required. We found that where a percentage amplitude is not required for detector operation then a simple amplifier and phase corrected filter is all that is necessary. Where a percentage amplitude is required for detector operation, such as in gate generators, then AGC or some kind of amplitude controlled clipping or gate sensing level is required as well as the phase corrected filters. We also provided a means to maintain the poor signal amplitude characteristics while using AGC. The latter circuit is also useful for driving amplitude controlled clipping or gate level sensing circuits instead of a fixed level.

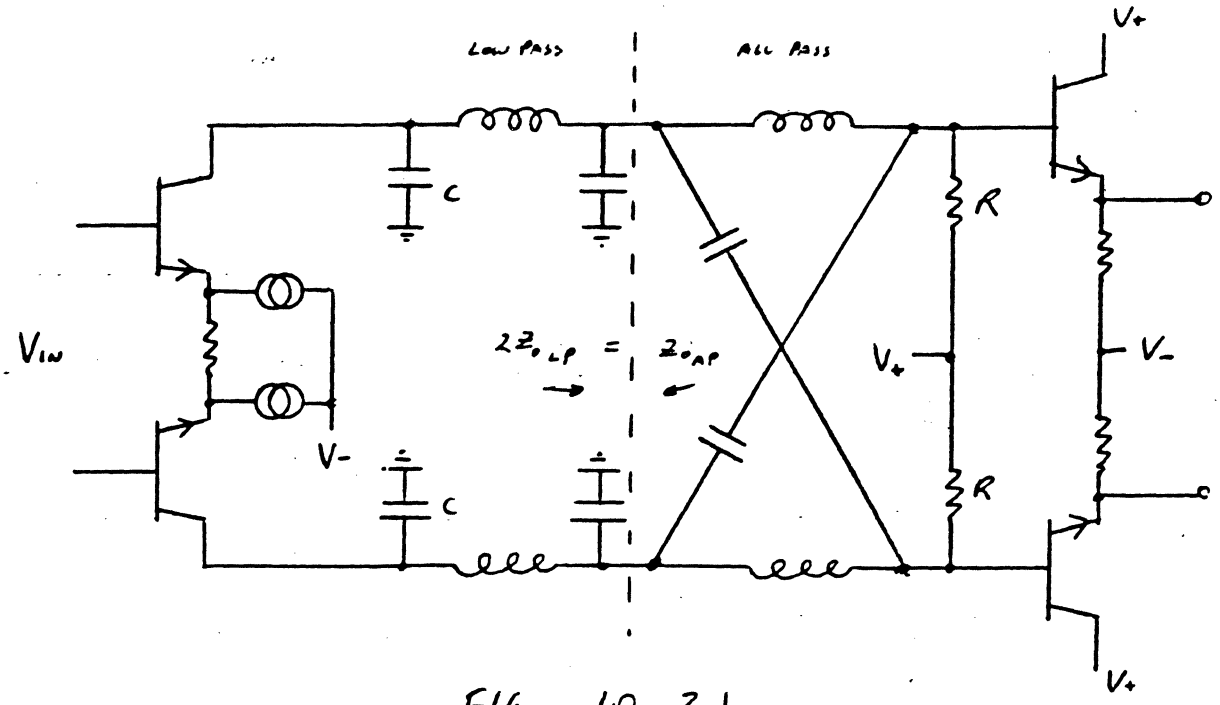


FIG 10.21

COMBINED GAIN, 3 POLE (BUTTERWORTH) FILTER, ALL PASS LATTICE FILTER. (CURRENT INPUT STYLE)
BUTTERWORTH FILTER HAS Z_0 OF R
ALL PASS LATTICE HAS Z_0 OF 2R
TWO CAPACITORS C PROVIDE SINGLE ENDED AS WELL AS DIFFERENTIAL FILTERING.

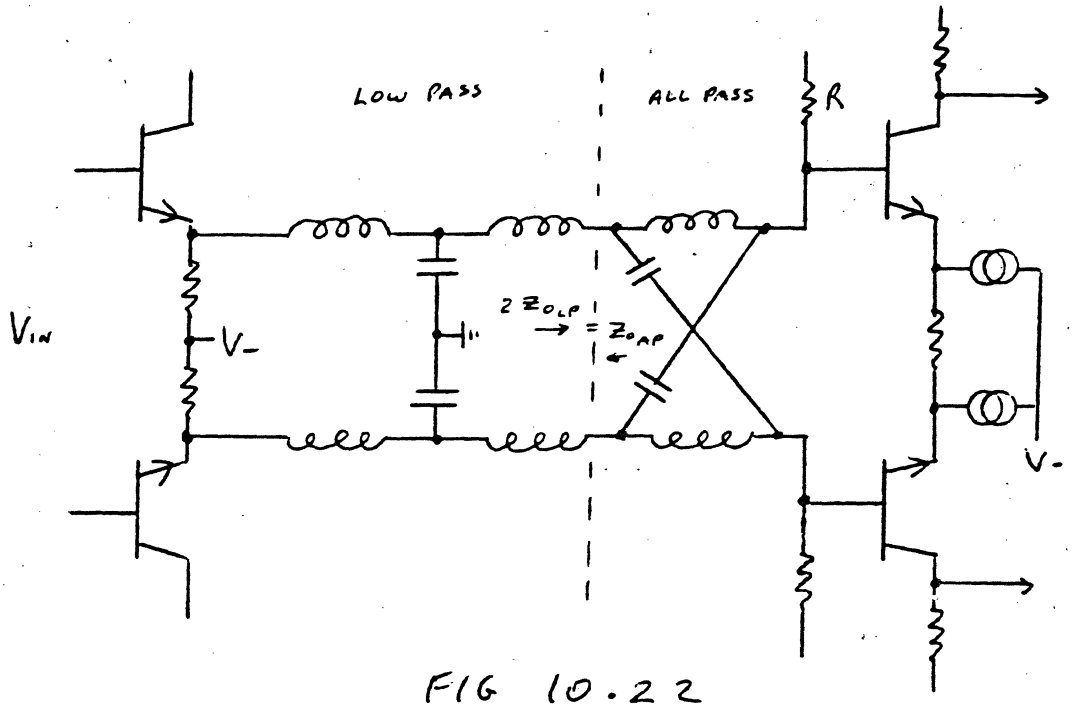


FIG 10.22

COMBINED GAIN, 3 POLE BUTTERWORTH FILTER, ALL PASS LATTICE FILTER, (VOLTAGE INPUT STYLE)

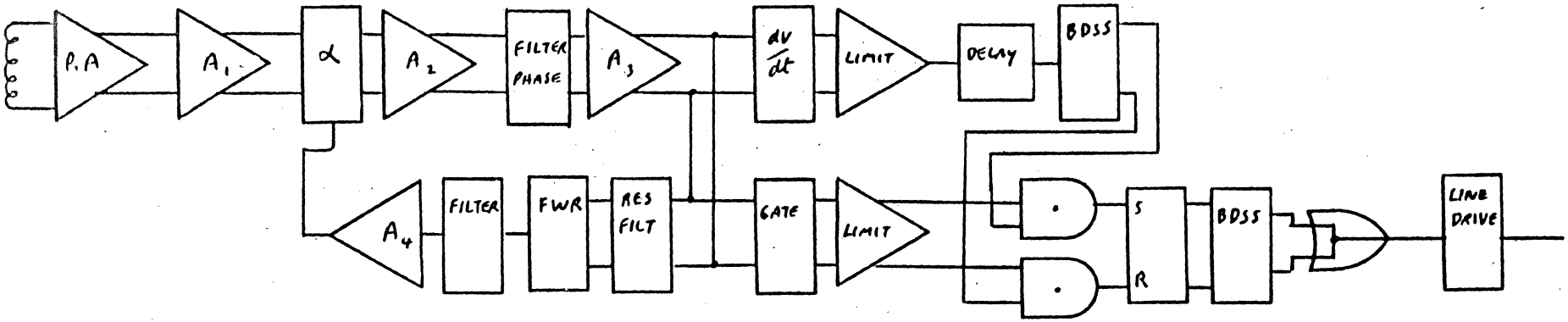


FIG 10.25 REGION 1 OR 2 BLOCK DIAGRAM OF ENTIRE READ CHANNEL

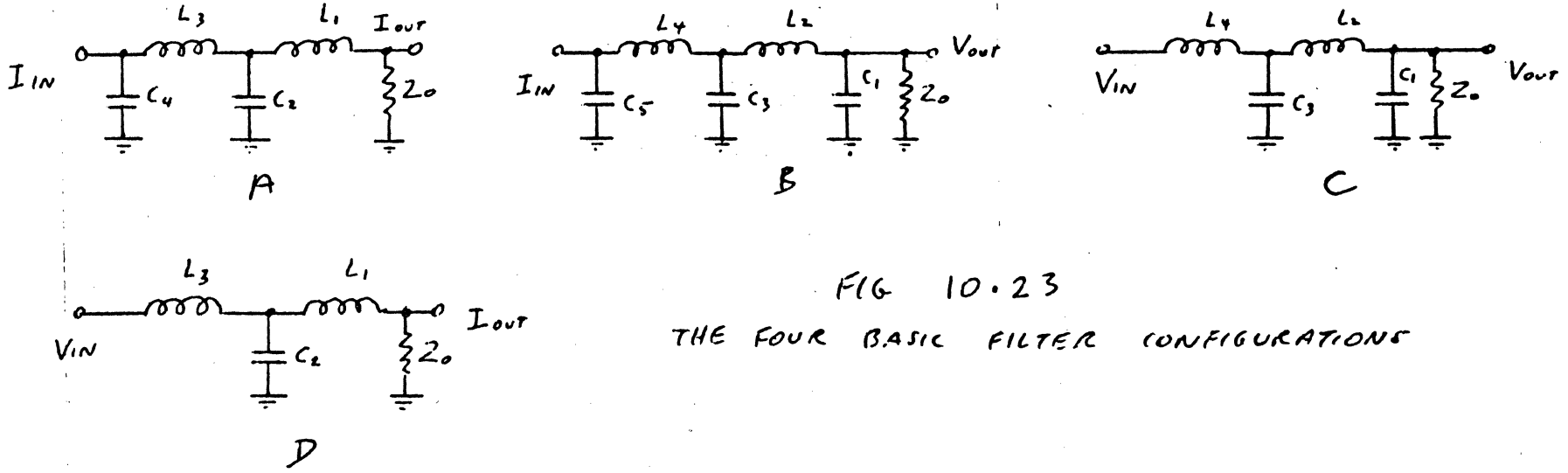


FIG 10.23
THE FOUR BASIC FILTER CONFIGURATIONS

LINEAR AMPLIFIERS

Figures 10.21 shows a typical filter amplifier block. It can be used to replace one of the three blocks we designed back in Fig. 10.8. The filter design is for a current input which we have with a common emitter amplifier using emitter feedback. If we chose a voltage input filter then we must use either an emitter follower driver as shown in Figure 10.22 or by loading the collectors of an amplifier with a load resistor equal to the filter impedance then we can convert a current to a voltage. Note that the voltage input filter must be terminated on both ends, therefore, the gain is half unless the impedance is doubled (see Figure 10.24).

Figure 10.23 shows the four basic types of filters. Each must be terminated with its characteristic impedance Z_0. The type is determined by the input and the number of poles. Figure 10.23A shows a current input and four (even) poles, therefore, the output will be a current feeding Z_0. Figure 10.23B is again a current input with five (odd) poles, therefore the output is a voltage feeding Z_0. The next figure 'C' is of a voltage input filter with four (even) poles, therefore, it has a voltage output to Z_0. Similarly, Figure 10.23D is a voltage input with 3 poles (odd), therefore, a current output feeding Z_0.

Any filter may be used depending on the design. The current input type is handy as it can be used directly in the collector of our standard linear amplifiers thus minimizing the number of transitors required. The function of the Phase correction filter can also be made a part of the low pass filter by making its Z_0 equal to two times the Z_0 of the low pass filter. This is shown in Figures 10.21, 10.22, and 10.24. Although there are several forms of the All Pass filter, the most deisrable is that shown in Fig. 10.26A and B. Two types are shown. Each of these can be matched to the low pass Z_0. The first, A, provides a shift of 180^0 as a function of frequency

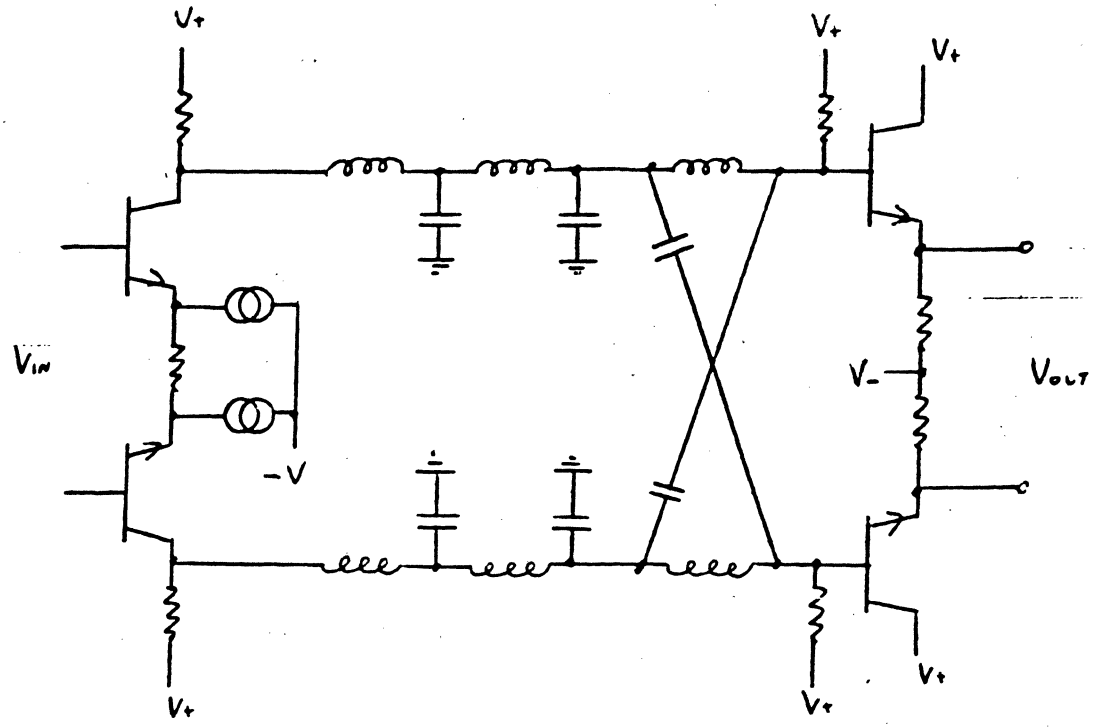


FIG 10.24
VOLTAGE INPUT 4 POLE FILTER VARIATION

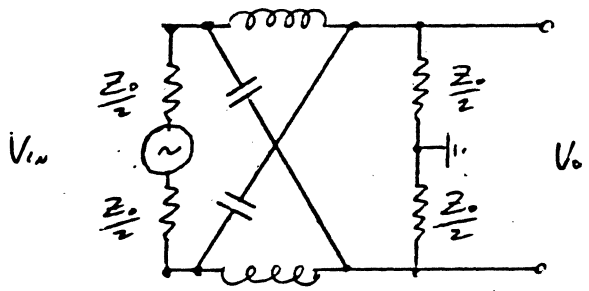


FIG 10.26 A 180° ALL PASS LATTICE

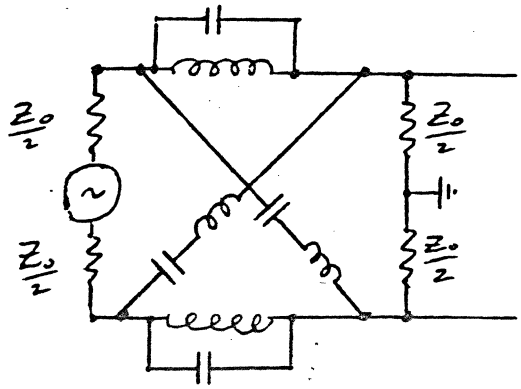


FIG 10.26 B 360° ALL PASS LATTICE

The second provides a 360^{o} change as a function of frequency and can be altered as to the rate of change depending on the ratio of its elements. This is discussed in the reference. The number of poles of the low pass filter depends on the slope of the roll off required. But it also affects the phase error rate of change which forces either a 360^{o} All Pass Lattice or less poles in the Low Pass. Such is the case in many designs where the low pass uses only three poles. Sometimes some degree of phase correction can be performed by using either or both lead and lag circuits in the emitter feedback path.

One of the variations is the use of separate or different filters for the gate and peak sensing channel, as shown in Fig. 10.27A and B.

One last type of filter that has some usage is one derived from the delay line differentiator discussed in Chapter 9; only this time the two are added directly. The derivation is obtained from the block diagram of Fig. 10.28

$$\frac{V_o}{} = V_{in}\left(1 + e^{-j\omega T}\right) \qquad \text{EQ 10.104}$$

$$\frac{V_o}{V_{in}} = 1 + e^{-j\omega T} = 1 + e^{-j\frac{\omega T}{2}} \cdot e^{-j\frac{\omega T}{2}} = \frac{e^{j\frac{\omega T}{2}}}{e^{j\frac{\omega T}{2}}} + \frac{e^{-j\frac{\omega T}{2}}}{e^{j\frac{\omega T}{2}}} \qquad \text{EQ 10.105}$$

$$= \frac{e^{j\frac{\omega T}{2}} + e^{-j\frac{\omega T}{2}}}{e^{j\frac{\omega T}{2}}} = 2\left(\frac{e^{j\frac{\omega T}{2}} + e^{-j\frac{\omega T}{2}}}{2}\right) e^{-j\frac{\omega T}{2}} \qquad \text{EQ 10.106}$$

and from $\cos\theta = \dfrac{e^{j\theta} + e^{-j\theta}}{2}$ Eq.10.107

we get $2\left(\cos\dfrac{\omega T}{2}\right) e^{-j\frac{\omega T}{2}}$ Eq.10.108

which is a filter with no phase shift except a fixed delay. It is used particularly in spectral shaping or in circuits that require no phase shift. An implementation of the filter is shown in Fig. 10.30.

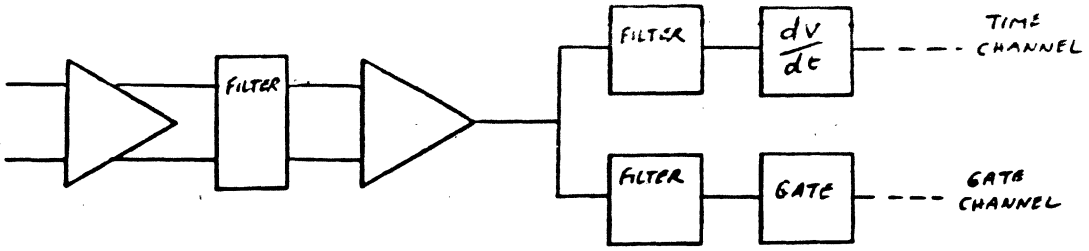


FIG 10.27A

BANDPASS SHAPING FOR EACH CHANNEL

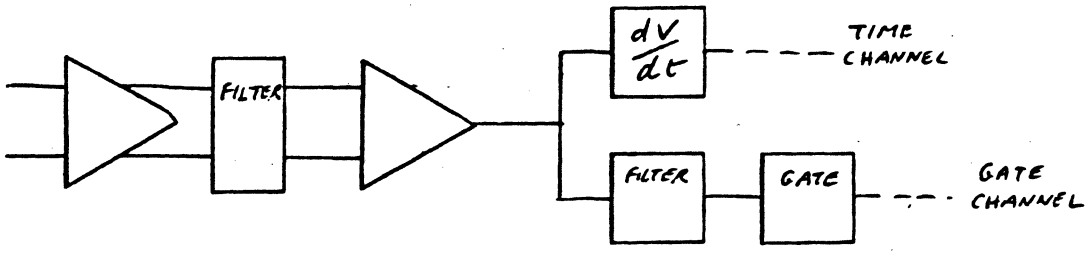


FIG 10.27B

BANDPASS SHAPING FOR GATE CHANNEL (less noisy?)

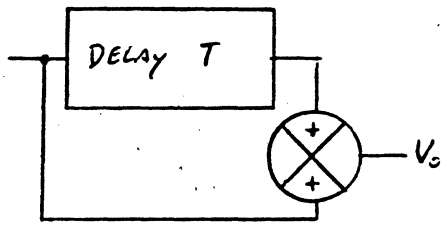


FIG 10.28

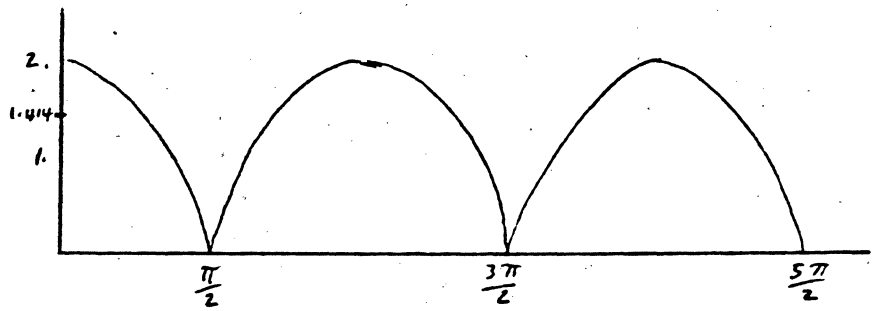


FIG 10.29

MAGNITUDE OF $2\left(\cos\frac{\omega T}{2}\right)e^{-j\frac{\omega T}{2}}$

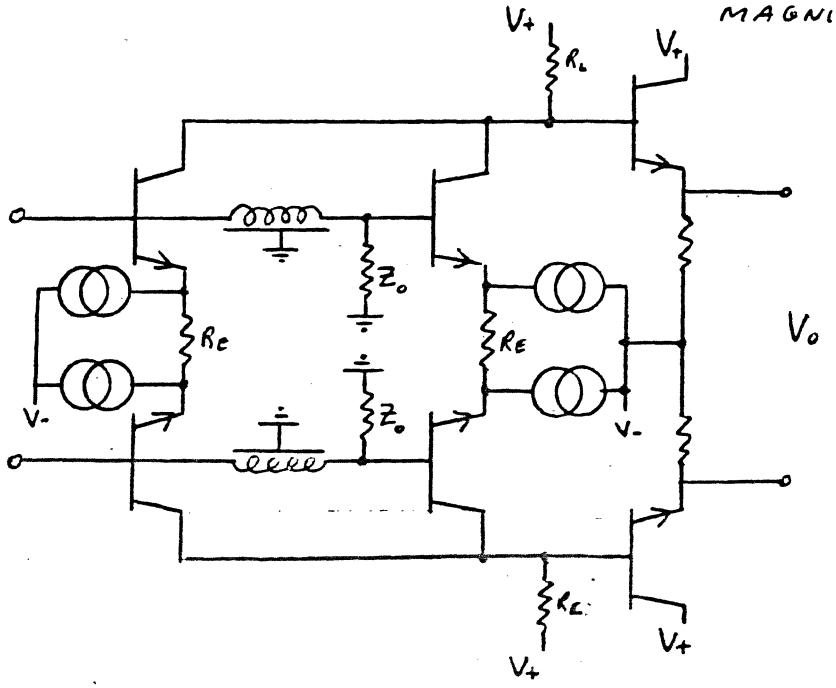


FIG 10.30

IMPLEMENTATION OF COSINE MAGNITUDE FILTER

CORRECTED 1/15/81

There is a class of full wave rectifiers that needs no temperature compensation. These are the differential type where the Vbe and/or diode drops are symmetrical and therefore cancel. The circuit of Fig. 10.31 can be designed for any compatible level as long as the input has no offset and is a true differential input. The attack and decay can be tailored, but squelch is difficult due to the lack of a reference, unless a MOS Fet is used directly across the capacitor. The chopper transistors will not work due to the base current. The optional capacitor, C_2, around the op amp may be added, in addition to the capacitor , C_1. One nicety is that C_2provides equal attack and decay, while C_1 provides the sample storage which is at a higher bandwidth than the op amp can handle. The second circuit, Fig. 10.32, uses a multiplier configuration. Again, the same comments regarding squelch and C_1 and C_2. Careful balance is required of the two current sources, I_1 and I_2, for correct operation.

Between the AGC control of amplitude with fixed percentage gate references and fixed gain with a level controlled reference type detectors, the AGC versions are preferred due to their being under closed loop control, while the signal level controlled gate reference is open loop, meaning that under worse case conditions the reference can wander all over the place.

Back in the section on AGC we presented equations, 10.69, to describe the attack and decay of the AGC. This becomes very important when we consider the signal amplitude envelope resulting from a read. Up to now we have mostly only considered the individual pulses, or just a few in a row. Here we need to discuss the effect of variations in amplitude as a function of magnetic coating thickness and dispersion. Quite often the amplitude modulation is significant and in order to recover all the transitions written each and every pulse must be detected.

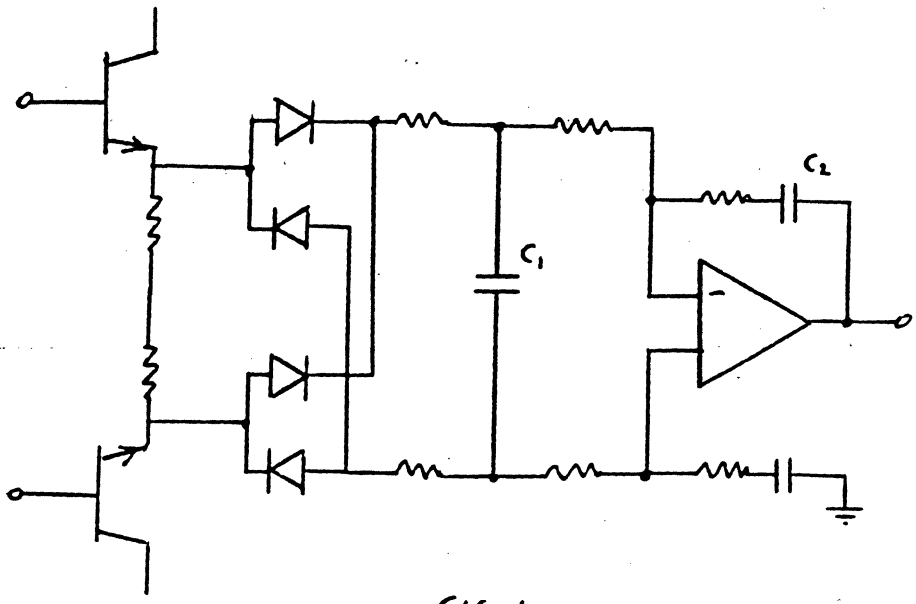


FIG 10.31

DIFFERENTIAL FULL WAVE RECTIFIER AND FILTER

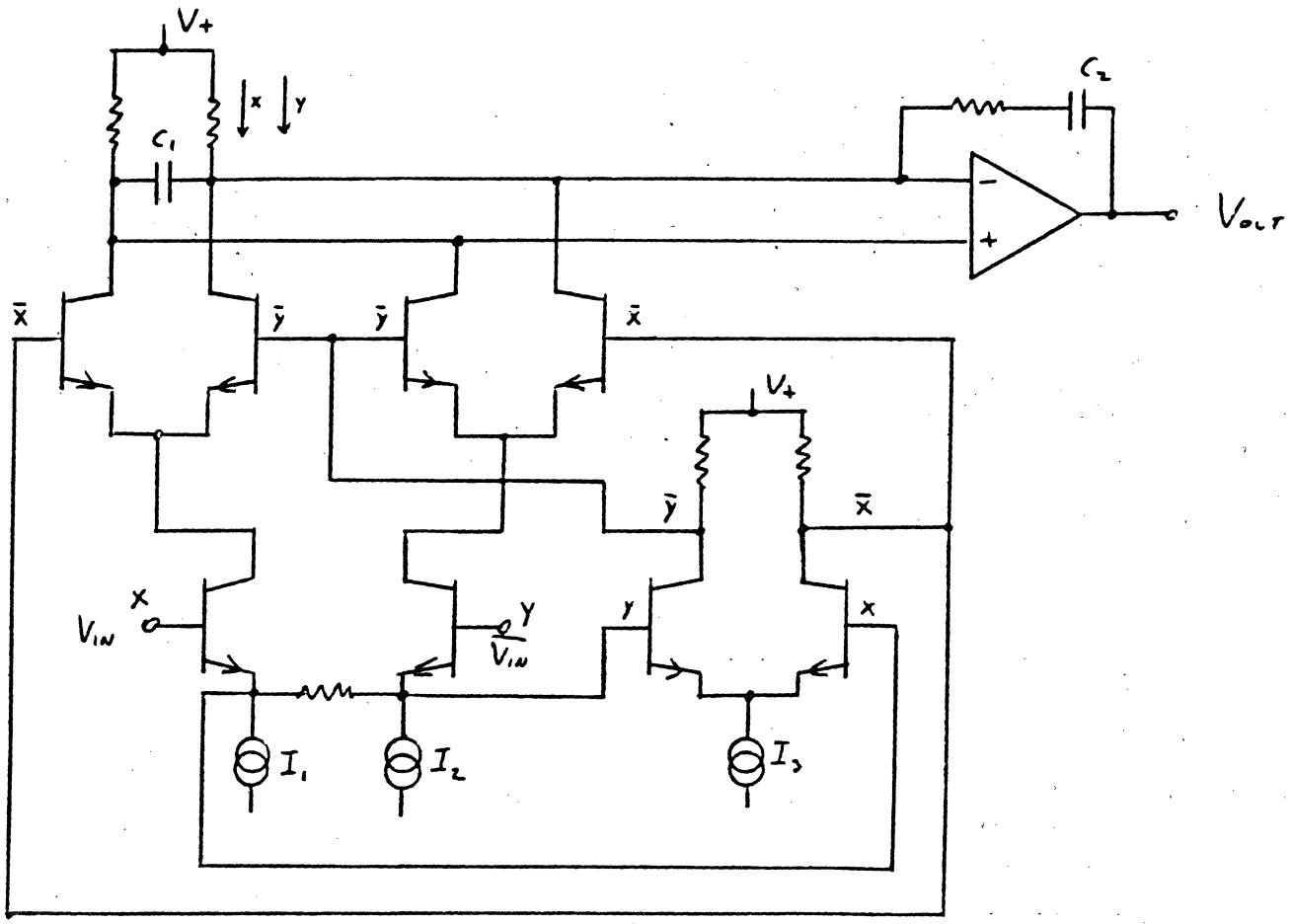


FIG 10.32

MULTIPLYER, FULL WAVE DIFFERENTIAL RECTFIER AND FILTER

In Fig. 10.33, at A, there is a 50% amplitude reduction and at B a 100% loss due to a hole in the media. We should be able to recover transition pulses down to around 15% remaining. As we discussed the detectors we can see that the time channel (if it has sufficient gain, or, in other words, if this 15% were used as the minimum signal in order to calculate the minimum gain required) will always sense these pulses. The problem is the Gate Generator. For Region 1 we can see that the clipping level must include the lowest amplitude expected but not low enough to add false pulses due to noise. In Region 2 the Gate Generator becomes even more of a problem due to the compromise between the center bit of a triple, Fig. 9.22, and the remaining signal in a defect, such as A, Fig. 10.33. In order to assist the detection of these pulses, then, the AGC must be able to follow the modulation. As most AGC circuits in use in radio have fast attack and slow decay, it is easily seen that we cannot tolerate this behavior. If our AGC circuits were designed to have equal attack and decay of sufficient bandwidth to follow the types of defects we want to allow, or are forced to use, then the clipping level, or sense level (depending on the Region), need not extend to the lowest levels near the noise. The τ, then, must allow the defect to be traversed with minimal change in amplitude. Being a Type 0 loop, it is obvious there must be some error in order to achieve the *nessisary* gain change. Therefore the actual output amplitude change should follow the dotted lines of Fig. 10.34.

The same comments apply if the designer wants to use the amplitude controlled clipping level approach despite the fact that it is open loop. Much effort is lost by attributing loss of recovered data to the clipping or sense level based on the average amplitude of the envelope instead of allowing for the defect

caused amplitude reductions. With a correctly operating envelope circuit the clipping level, Region 1, or the sense level, Region 2, may be raised so that the detector is less susceptible to the noise, yet is fully able to sense all the pulses.

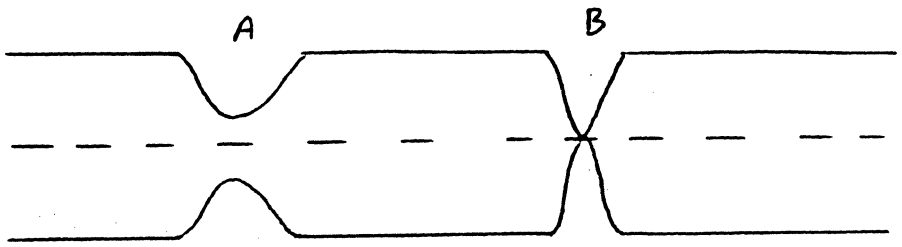


FIGURE 10.33

TYPICAL SIGNAL ENVELOPE WITH DEFECTS

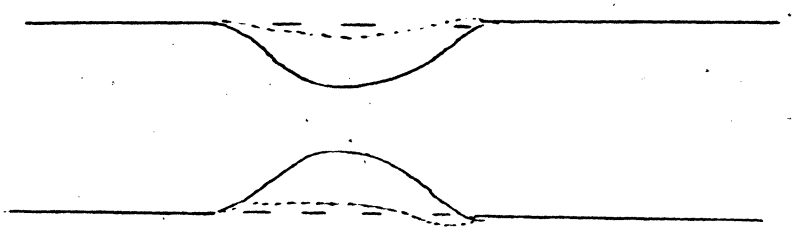


FIG 10.34

PREFERED EQUAL T AGC'd SIGNAL ENVELOPE

11 DATA CLOCKING - PHASE LOCKED LOOPS

In early disc drives and tape drives all data clocking was handled by a separate clock track, as shown in Fig. 11.1. As the data density increased the tape skew in tape machines and separate head vibration in disc machines forced a move towards self clocking data codes. Some relief was obtained by breaking up the clock signal into four phases in quadrature and selecting the phase closest to the data on a per record basis. These early data streams contained long strings of no transitions. Therefore they were difficult to use for generating their own clock. Attempts were made, using HiQ ringing amplifiers, to fill in the spaces and gaps; but these all suffered from frequency pulling if the tuned circuits were not exactly tuned to the incoming data frequency. For example, if the data transitions were continuous, then the output phase was a function of the difference between the LC tuned frequency and the data frequency. During periods of no transitions, the clock was equal to the LC tuned frequency. Therefore the phase error would accumulate until the transitions recurred. Variations in disc tape speed prevents exact tuning of these circuits. Fig. 11.2 shows a typical circuit.

The self clocking data codes restricted the maximum spacing between transitions which permitted the use of either single shot controlled data recovery and clocking or, better yet, phase locked loop controlled clocking and recovery. An example of the single shot type is shown in Fig. 11.3A, where the incoming transition pulses include alternate clock and data. The regularly occurring clock transitions establish a gate for the following data transition if it is present. Correct phasing is always established following any cell not containing a data transition such as at D_2 in Fig. 11.3B.

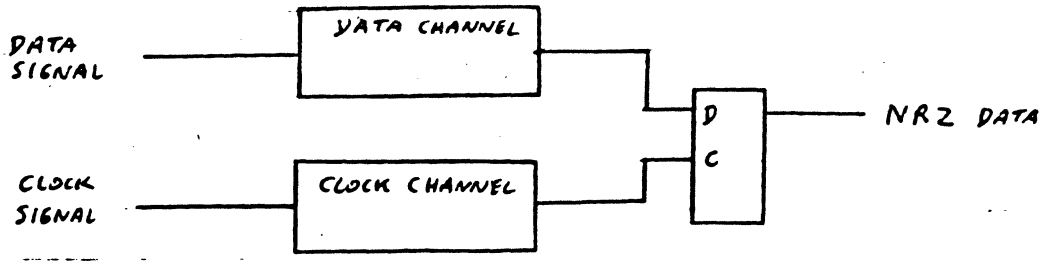


FIG 11.1

SEPARATE CLOCK TRACK DATA CLOCKING

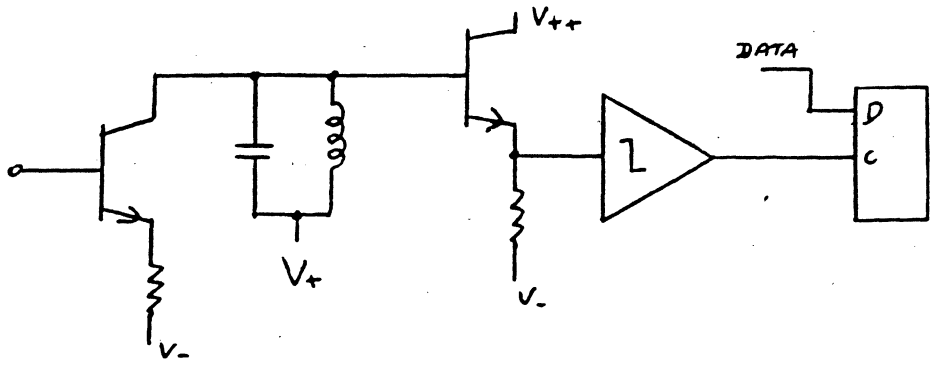


FIG 11.2

TYPICAL CLOCK RINGING AMPLIFIER

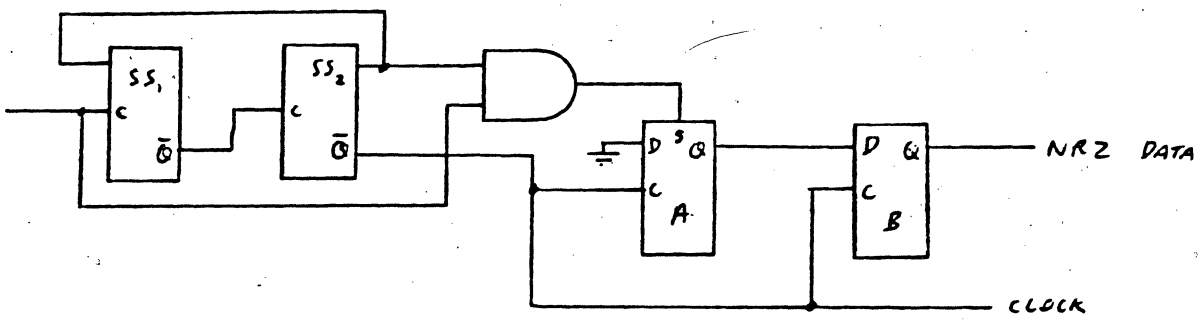


FIG 11.3 A

DUAL SINGLE SHOT CLOCKING OF F.M. DATA

By far the better clock generating circuits are the phase locked loops. There are four basic types. The type number for the closed loop is obtained from the characteristic equation. This equation is derived from the basic block diagram shown in Fig. 11.4. The type number is equal to the number of poles at the origin of $G(s)$.

The equation for the Basic Phase Locked Loop is given in EQ 11.1.

$$\frac{\phi_o}{\phi_{in}} = \frac{G(s)}{1 + G(s)} = \frac{K_s K_\phi K_F K_A K_o \frac{1}{s}}{1 + K_s K_\phi K_F K_A K_o \frac{1}{s}} \qquad \text{(EQ 11.1)}$$

$$C.E. = 1 + K_s K_\phi K_F K_A K_o \frac{1}{s}$$

where K_s is the sample rate expressed as $\frac{2\pi F_{in}}{2\pi F_{osc}}$ (EQ 11.1A)

K_ϕ is the phase detector gain expressed as either volts per radian or amps per radian, depending on the circuit used

K_F is the gain of the filter in standard LaPlace notation

K is the gain of the amplifier expressed as either volts per volt, volts per amp, amps per volt, or amps per amp - again depending on the circuit used to interface the filter to the oscillator

K is the gain of the oscillator which can be expressed as radians per second per volt or radians per second per amp, depending on the circuit

The frequency to phase conversion is simply $\frac{1}{S}$ -- a mathematical integration.

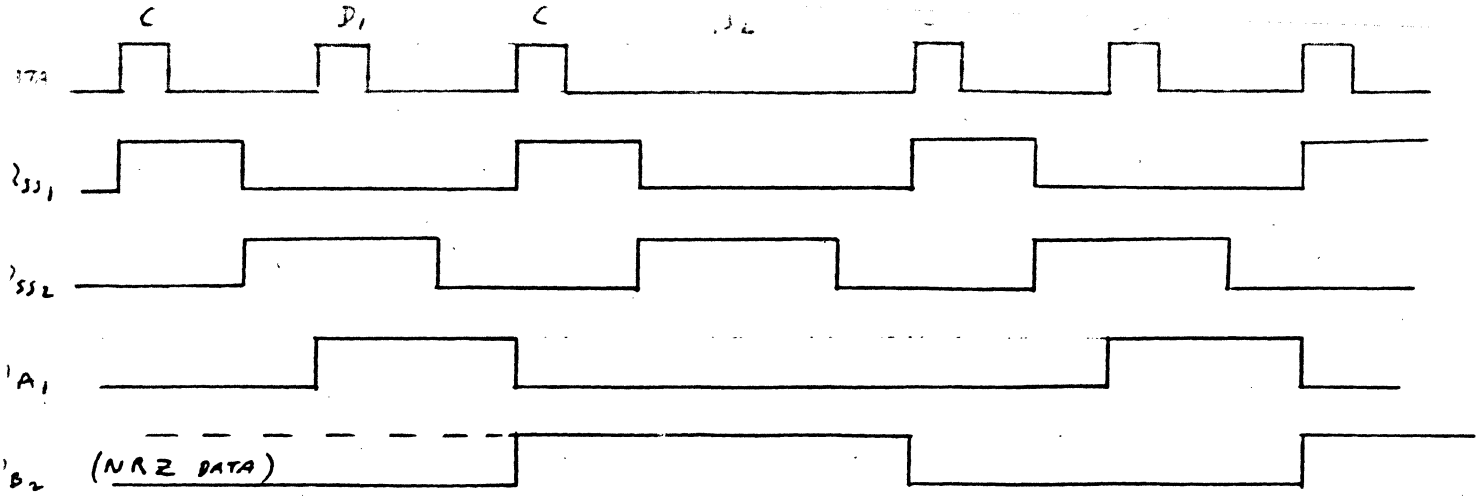


FIG 11.3B

DUAL SINGLE SHOT CLOCKING OF F.M. DATA

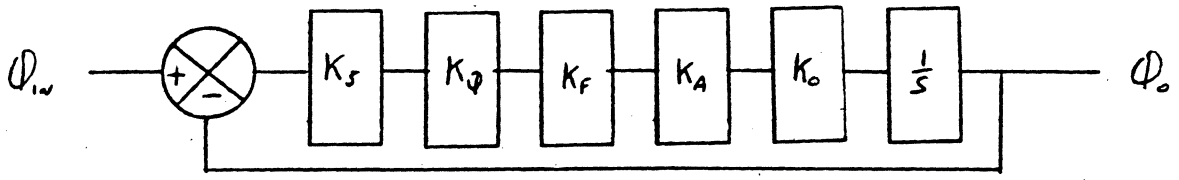


FIG 11.4

BASIC BLOCK OF A PHASE LOCKED LOOP

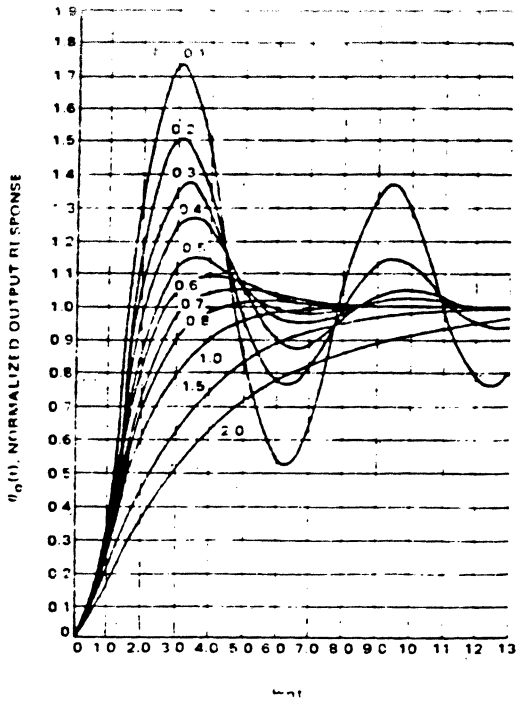


FIGURE – Type 1 Second Order Step Response

11.5

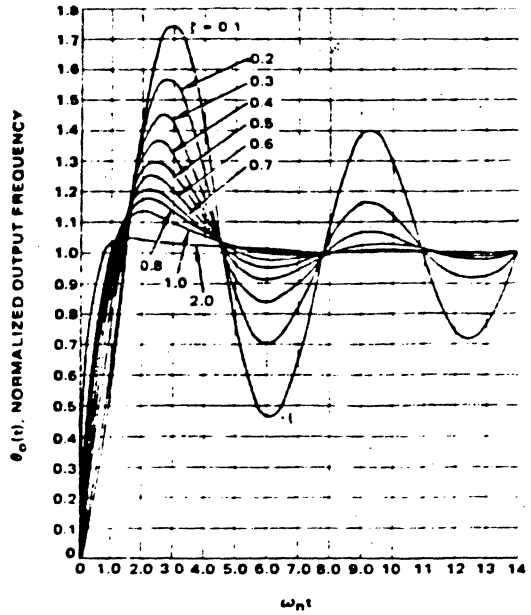


FIGURE – Type 2 Second Order Step Response

11.6

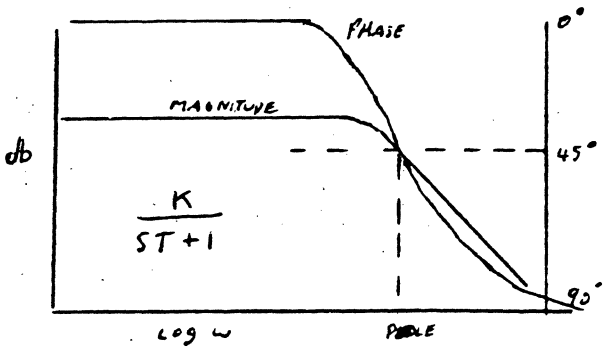


FIGURE 11.7

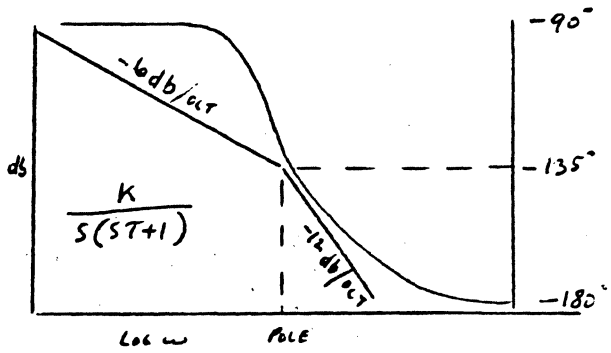


FIG 11.8

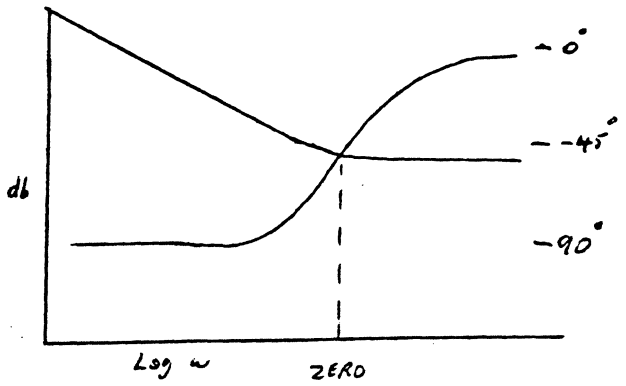


FIG 11.9

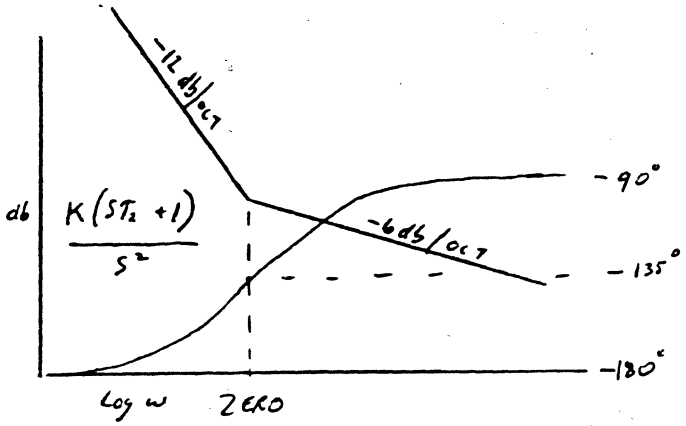


FIG 11.10

The type number is determined by the number of poles at the origin of G(s) for unity feedback and H(s)G(s) for non unity feedback. Thus, the single "s" in the denominator of EQ 11.2 indicates a Type 1 loop.

$$G(s) = \frac{1}{s(Ts+1)} \qquad \text{(EQ 11.2)}$$

$$G(s) = \frac{(T_z s + 1)}{s^2 (T_p s + 1)} \qquad \text{(EQ 11.3)}$$

and a Type 2 for EQ 11.3.

The order of the loop is determined from the highest order of the characteristic equation, which is the denominator of EQ 11.1, as 1 + G(s) = 0 $(C.E.)$

Picking up the equation G(s) of EQ 11.2 in EQ 11.4,

$$C.E. = 1 + \frac{1}{s(Ts+1)} = 0 \qquad \text{(EQ 11.4)}$$

we get $$s(Ts+1) + 1 = Ts^2 + s + 1 \qquad \text{(EQ 11.5)}$$

which states the circuit to be second order as S is squared. Similarly, EQ 11.3 would be a third order when evaluated.

We can now evaluate the error conditions for various inputs.

$$\Phi_{error} = \Phi_{IN} - \Phi_{out} = \Phi_{IN} - \Phi_{err}(G(s)) \qquad \text{(EQ 11.6)}$$

$$\Phi_{err} = \frac{\Phi_{IN}}{1 + G(s)}$$

For a step change in input phase, in radians, $\mathcal{L}(\phi_{IN}) = \frac{\phi}{s}$ (EQ 11.7)

Therefore

$$\phi_{err_{step}} = \frac{\phi}{s(1+G(s))} \quad \text{evaluated at } t=\infty \text{ or } s=0$$

For a Type 0 loop where G(s) is

$$G(s) = \frac{K}{Ts+1}$$

$$\phi_{err_{s.s}} = \lim_{s\to 0}\left[\frac{s\,\phi_{(s)}}{s\left(1+\frac{K}{Ts+1}\right)}\right] = \lim_{s\to 0}\left[\frac{s(Ts+1)(\phi_{(s)})}{Ts^2+(1+K)s}\right] \quad \text{(EQ 11.8)}$$

which IS A CONSTANT $\frac{\phi_s}{1+K}$

For a Type 1 loop, where G(s) is

$$G(s) = \frac{K}{s(sT+1)}$$

$$\phi_{err_{s.s.}} = \lim_{s\to 0}\left[\frac{s\,\phi_s}{s\left(1+\frac{K}{s(Ts+1)}\right)}\right] = \lim_{s\to 0}\left[\frac{s\phi_{(s)}(Ts+1)}{Ts^2+s+K}\right] \quad \text{(EQ 11.9)}$$

which is ZERO

For a Type 2 loop where G(s) is

$$G(s) = \frac{K(T_2s+1)}{s^2(T_ps+1)}$$

$$\phi_{err_{s.s.}} = \lim_{s\to 0}\left[\frac{s\,\phi_{(s)}}{s\left(1+\frac{K(T_2s+1)}{s^2(T_ps+1)}\right)}\right] = \lim_{s\to 0}\left[\frac{s\,\phi_{(s)}\,s(Ts+1)}{Ts^3+s^2+K(T_2s+1)}\right] \quad \text{(EQ 11.10)}$$

which is 0.

Similarly, for a Type 3 loop where G(s) is

$$G(s) = \frac{K(T_2 s+1)(T_3 s+1)}{s^3(Ts+1)}$$

$$\phi_{err_{s.s.}} = \lim_{s\to 0}\left[\frac{s\phi_0}{s\left(1+\frac{K(T_2 s+1)(T_3 s+1)}{s^3(Ts+1)}\right)}\right] = \lim_{s\to 0}\left[\frac{s\phi_0 s^2(Ts+1)}{s^3(Ts+1)+K(T_2 s+1)(T_3 s+1)}\right] \quad \text{(EQ 11.11)}$$

which is, again, 0.

For a ramp change in input phase

$$\mathcal{L}\,\phi_{in_{ramp}} = \frac{\phi_R}{s^2}$$

Starting with a Type 0,

$$\phi_{err_{ss}} = \lim_{s\to 0}\left[\frac{s\phi_R}{s^2\left(1+\frac{K}{Ts+1}\right)}\right] = \lim_{s\to 0}\left[\frac{s\phi_R(Ts+1)}{s^2(Ts+1)+Ks^2}\right] \quad \text{(EQ 11.12)}$$

which is continually increasing towards infinity.

For a Type 1

$$\phi_{err_{ss}} = \lim_{s\to 0}\left[\frac{s\phi_R}{s^2\left(1+\frac{K}{s(Ts+1)}\right)}\right] = \lim_{s\to 0}\left[\frac{s\phi_R s(Ts+1)}{s^3(Ts+1)+Ks^2}\right] \quad \text{(EQ 11.13)}$$

which is a constant $\frac{\phi_R}{K}$

corr 12/16/80

For Type 2

$$\Phi_{err} = \lim_{S\to 0}\left[\frac{S\Phi_R}{S^2\left(1+\frac{K(T_2S+1)}{S^2(TS+1)}\right]}\right] = \lim_{S\to 0}\left[\frac{S\Phi_R\, S^2(TS+1)}{S^4(TS+1)+KS^2(T_2S+1)}\right] \quad \text{(EQ 11.14)}$$

which is zero .

And for a Type 3

$$\Phi_{err} = \lim_{S\to 0}\left[\frac{S\,\Phi_R}{S^2\left(1+\frac{K(T_2S+1)(T_3S+1)}{S^3(TS+1)}\right]}\right] = \lim_{S\to 0}\left[\frac{S\Phi_R\, S^3(TS+1)}{S^5(TS+1)+KS^2(T_2S+1)(T_3S+1)}\right]$$

(EQ 11.15)

which is zero.

We can also evaluate the various types for an accelerated phase changing input where

$$\mathcal{L}(\ddot{\psi}_A) = \frac{\Phi_A}{S^3} \quad \text{(EQ 11.16)}$$

wherein we find that the various errors are for a Type 0 and Type 1 infinite, Type 2 constant, and Type 3 zero, which can be determined by the reader from

$$\Phi_{err_{S.S.}} = \lim_{S\to 0}\left[\frac{S\,\Phi_A}{S^3(1+G_{(S)})}\right] \quad \text{(EQ 11.17)}$$

cont' 12/16/80

We can make a table for the various conditions and types

	Type 0	Type 1	Type 2	Type 3
Step FREQ	$\frac{\Phi_s}{K}$ / ∞	0 / A_s	0 / 0	0 / 0
Ramp FREQ	∞ / ∞	$\frac{\Phi_R}{K}$ / ∞	0 / A_R	0 / 0
Accel. FREQ	∞ / ∞	∞ / ∞	$\frac{\Phi_A}{K}$ / ∞	0 / A_A
Step Φ_s	$\frac{\Phi_s}{K}$	0	0	0
Ramp Φ_R	∞	$\frac{\Phi_R}{K}$	0	0
Accel. Φ_A	∞	∞	$\frac{\Phi_A}{K}$	0

FREQ / PHASE

if we remember that to convert from phase to frequency we multiply by S, as shown in EQ 11.18, and from frequency to phase we divide by S they then cancel and we have our same table.

$$F_{err_{ss}} = \lim_{S \to 0} \left[\frac{S \Phi_{(s)} S}{S\left(S\left(1 + \frac{K}{S(TS+1)}\right)\right)} \right] \qquad \text{(EQ 11.18)}$$

for a Type 1 step change in frequency.

When we choose between the various types, then, it is desirable to have all input variations result in zero phase error. This only occurs for a Type 3, but in practice the 'accelerated phase' condition, if it occurs, is only for a short time. Therefore, as the Type 2 is easier to build, it is preferred. This is borne out in testing, comparing the two in disc drives.

As we developed the equations for G(s) for the various types the reader may have noticed the addition of zeros for Types 2 and Types 3. This needs explanation. The Nyquist criteria requires the phase shift to be more positive than -180°at the point of the zero db gain crossing of the open loop G(s)H(s). For Type 0

Corr 12/16/80

and Type 1 the maximum phase shift is -90^0 and -180^0 respectively, at infinity. For Type 2 the phase shift starts out at -180^0 and heads towards -270^0 following the pole. To make it stable, a zero has to be added before the pole. For Type 3 the starting phase is -270^0 heading towards -360^0 due to the pole. The addition of two zeros before the pole brings the phase above the -180^0 required for stability. In all cases the pole is not required but is usually present due to stray effects.

The addition of these poles and zeros introduces another parameter called 'order.' The order of the circuit is determined from the order of Characteristic Equation, or the denominator of EQ 11.1, (C.E.)

$$1 + G(s)H(s) = 0 \qquad \text{(EQ 11.19)}$$

For example, if H(s) for unity feedback and G(s) were $\frac{K}{s(\tau s+1)}$ then

$$s(\tau s+1) + K = 0 = \tau s^2 + s + K \qquad \text{(EQ 11.20)}$$

which is a second order equation, hence the term 'Type 1 second order' when referring to that circuit (see also EQ 11.5). First order and second order circuits are well described in the literature and there exist many curves and equations relating their behavior. Third order and above are more difficult to predict, except as the entire equation is evaluated on a computer. The tradeoffs are not easily seen, as with the second order circuits.

For example, second order circuits can be described in terms of ζ and ω_n and are easily changed to obtain the required responses. Figs. 11.5 and 11.6 show the step response of Type 1 second order and Type 2 second order, respectively.

The open loop Bode plots for several configurations are shown in Figs. 11.7 through 11.12. These are not the only ones possible but are representative. In each case notice the stability criteria constraints. In Figs. 11.7, 8, 9, 10, and 11 the circuits would be unconditionally stable and in Fig. 11.12 it is stable only if the gain goes to zero well before the phase reaches -180^{o}.

In Fig.11.11 stability is only achieved if the gain goes to zero between the zero and pole. Due to stray capacitances the stability of Figs. 11.10 and 11.12 are questionable but predictable. The circuit used to obtain Fig. 11.10 is quite popular and is often used in the trade publications.

The -3db bandwidth in radians/sec is given for a Type 1 second order system as

$$\omega_{-3db_1} = \omega_n \sqrt{1 - 2\zeta^2 + \sqrt{2 - 4\zeta^2 + 4\zeta^4}} \qquad \text{(EQ 11.21)}$$

and for a Type 2 second order circuit as

$$\omega_{-3db_2} = \omega_n \sqrt{1 + 2\zeta^2 + \sqrt{2 + 4\zeta^2 + 4\zeta^4}} \qquad \text{(EQ 11.22)}$$

The settling time for a step response (within 5%) for a Type 1 second order system is approximately

$$t_{settling} \simeq \frac{4}{\zeta \omega_n} \qquad \text{(EQ 11.23)}$$

The curves of Figs. 11.5 and 11.6 correctly predict this behavior.

The Bode plot of Fig. 11.11 requires some further treatment as the presence of the added pole makes it a Type 2 third order, which is not so easily discussed. The Characteristic Equation is

$$C.E. = 1 + \frac{K(T_2 s + 1)}{s^2(T_p s + 1)} = 0 \qquad \text{(EQ 11.24)}$$

$$T_p s^3 + s^2 + K T_2 s + K = 0 \qquad \text{(EQ 11.25)}$$

We will return to this later, after we have demonstrated the difficulty.

Our best approach is to provide circuits to fill the blocks and then design several loops as examples.

Φ Detectors

The phase detector takes two forms, either the non-harmonic type or the harmonic type. The first is the kind usually used in frequency synthesizers for continuous waveforms. There are several forms. We will restrict ourselves to only the digital forms, as they fit the circuit blocks we might use. The second are insensitive to missing cycles or pulses such as occur in a data stream. The first kind develop false errors if a cycle is missed. Since we need both kinds, we will develop several of each. The test of a phase detector's function is the phase transfer curve, which relates the detector's response to various phase errors.

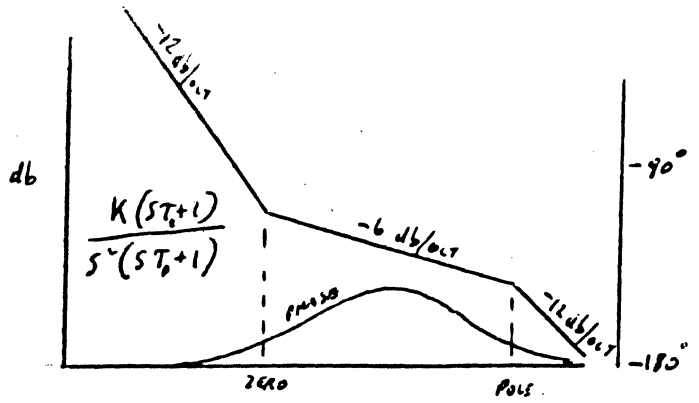


FIG 11.11

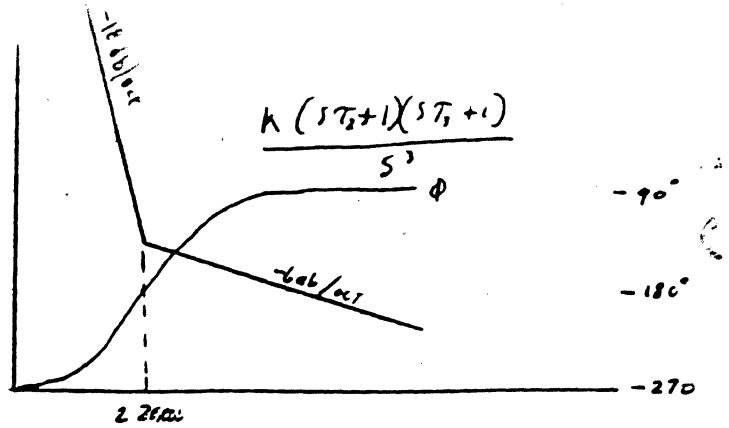


FIG 11.12

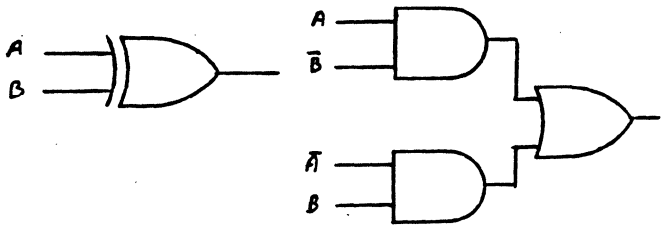


FIG 11.13 A

QUADRATURE NON HARMONIC Φ DETECTOR

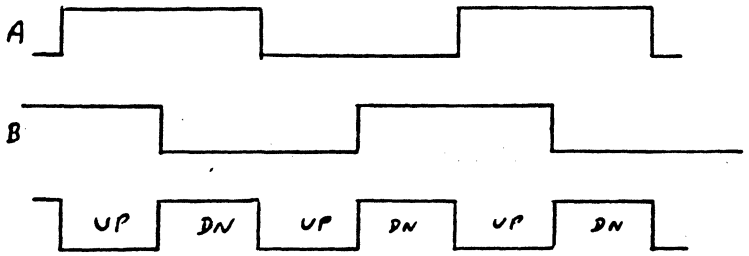


FIG 11.13 B

WAVEFORMS FOR QUADRATURE DETC

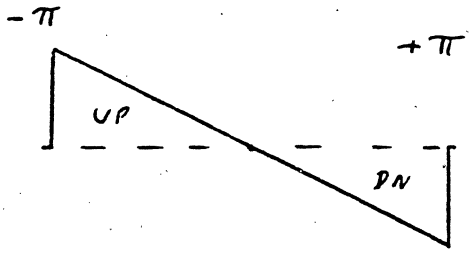


FIG 11.13C

PHASE TRANSFER CURVE FOR QUADRATURE DETECTOR

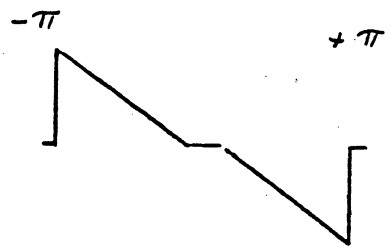


FIG 11.14 C

PHASE TRANSFER CURVE FOR 'D' NON HARMONIC Φ DETC.

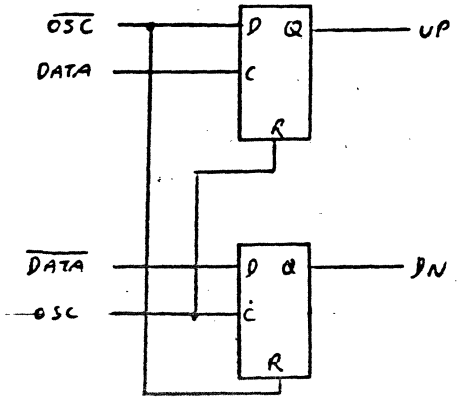


FIG 11.14 A

'D' NON HARMONIC Φ DETECTOR.

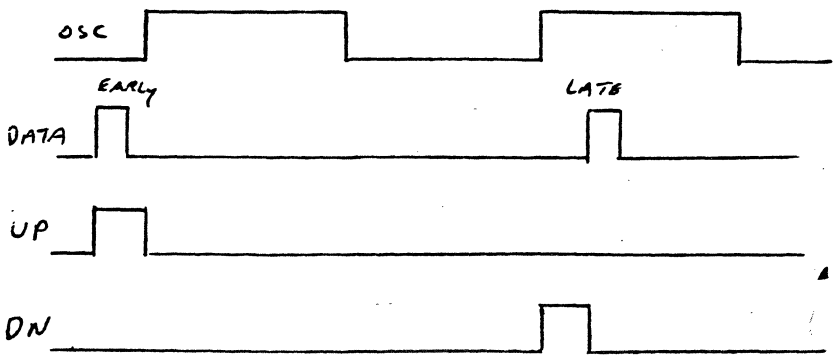


FIG 11.14 B

'D' NON HARMONIC Φ DETC WAVEFORMS

Non Harmonic

There are two forms of these. The first always produces both UP and DOWN errors. The average of these two errors becomes the error signal. The phase reference is $\pi/2$ radians; therefore they are called quadrature phase detectors. Examples of these are the exclusive "OR" circuit or the multiplier configuration which is a current exclusive or output device.

In Fig. 11.13B we can see the operation. Any phase shift to the left (early) or to the right (late) causes a shift in the area of the UP or the DOWN error which, when filtered, produces the desired error. The circuit has no dead band as a result of the two errors always being present (see Fig. 11.13C).

The second form are the "in phase" versions. They produce an error referenced to the edge of the waveforms. The circuit of Fig. 11.14A is one of these. This circuit is useful but suffers from some dead band due to setup and propagation times. Also, the filter must be able to handle very narrow pulses when the phase errors are near the phase reference. As also the logic family chosen must be able to handle the pulse widths (Fig. 11.14C). The circuit of Fig.11.15A does not exhibit dead band and is therefore preferred. There are commercial versions of these available; the Motorola MC4044 and 12040 being typical. These have similar waveforms to those discussed. Again, dead band and logic speed need consideration, particularly when the logic response times are an appreciable part of the duty cycle as this increases the tolerances or phase jitter, which can be referred to as spurious sidebands, in the closed loop operation.

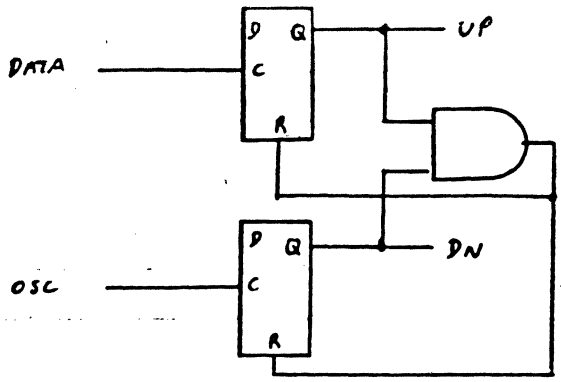


FIG 11.15 A

DEAD BAND ELIMINATION VERSION

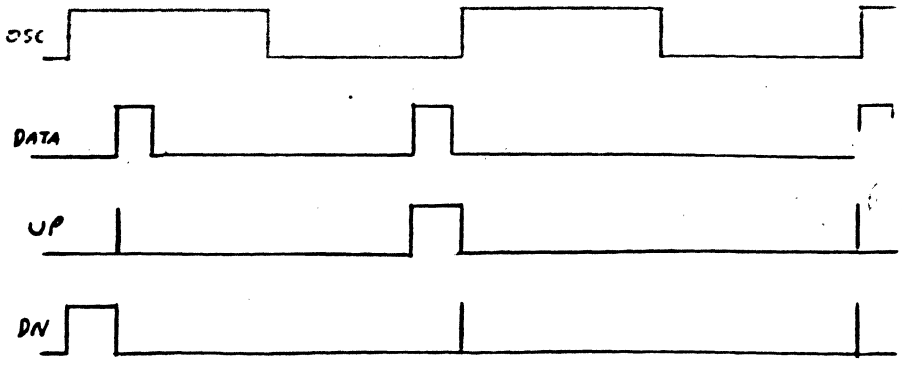


FIG 11.15 B

WAVEFORMS FOR NON DEAD BAND Φ DETECTOR.

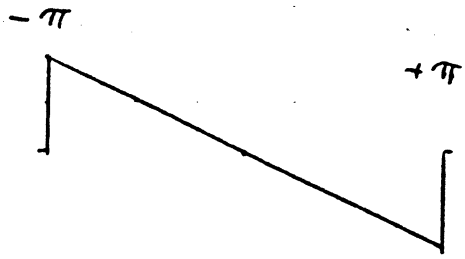


FIG 11.15 C

PHASE TRANSFER CURVE FOR NON DEAD BAND Φ DETC

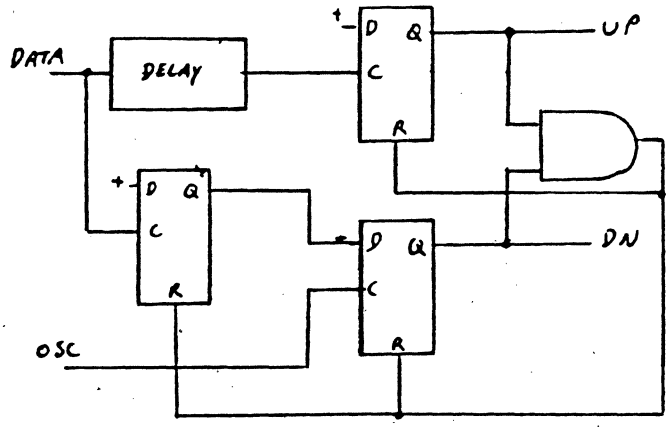


FIG 11.16 A

HARMONIC Φ DETECTOR

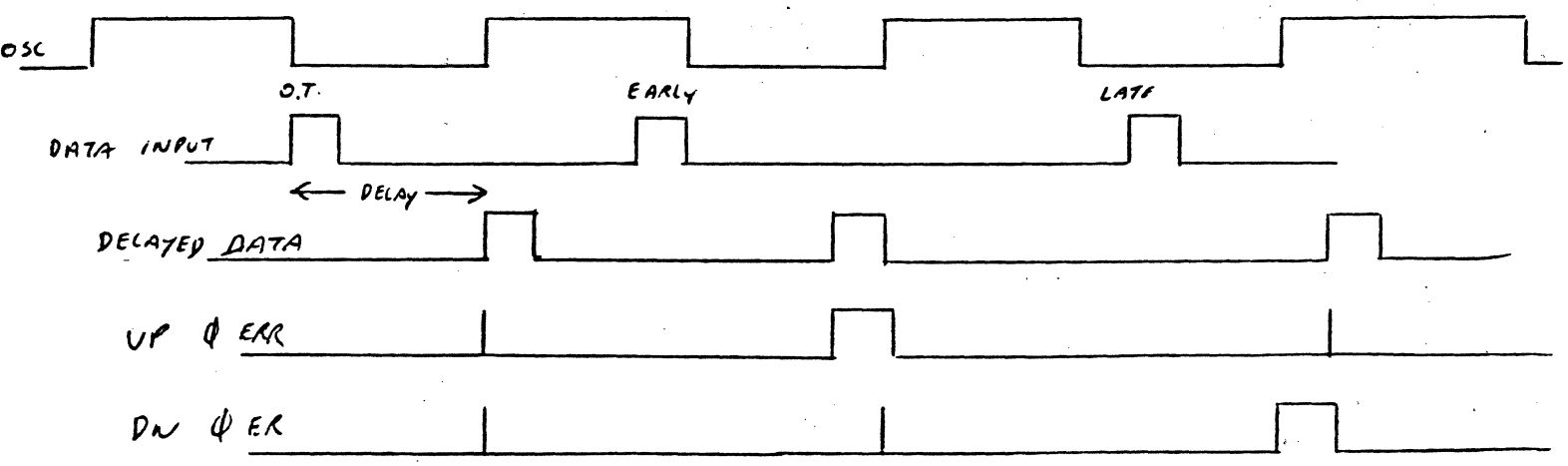


FIG 11.16 B HARMONIC Φ DETECTOR WAVE FORMS

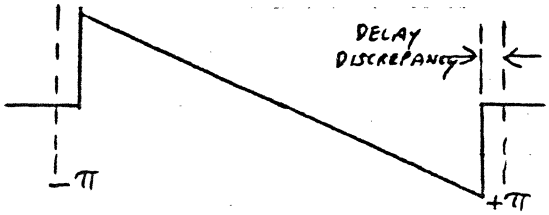


FIG 11.16 C

PHASE TRANSFER CURVE

Harmonic Phase Detectors

These detectors are required for data synchronization due to the nature of the data. A stream of ones and zeros require insensitivity to the missing data. The detectors already discussed fail in that they produce false "DOWN" error at the missing data time. The design of this class of phase detector includes circuits that allow the phase detector to work for one cycle following an input pulse.

One version of this is shown in Figs. 11.16A, B, and C. The circuit is similar to the non dead band version just discussed except that a gate has been added to condition the lower "D" F.F. clocked by the oscillator. The delay must be equal to, or slightly less than, one-half period of the oscillator plus 1 logic delay "C-Q" and 1 "D" setup time. The difficulty of this approach is that the input frequency has some tolerance due to tape speed or rotating discs. Therefore the phase transfer curve has a truncation at the leading edge 'A'. If the total delay were greater than a half period plus the other two delays then there could occur a false down error of π or greater, depending on the location of the following INPUT pulse. The circuit of Figs. 11.17A, B, C, is no better off as it also requires a delay. Here the incoming data sets both the UP and the DOWN error simultaneously. The UP is reset by the fixed delay and the DOWN is reset by the oscillator. The resultant error is the difference in area of the two wafeforms. The reference ϕ is the output of the delay line. The delay required to reset the UP FF must be equal to or less than one half an oscillator period. If it is greater, then the phase transfer curve is distorted in that the DOWN error is shortened at the previous oscillator edge instead of the correct edge for a late pulse. Fig. 11.18. For delays shorter

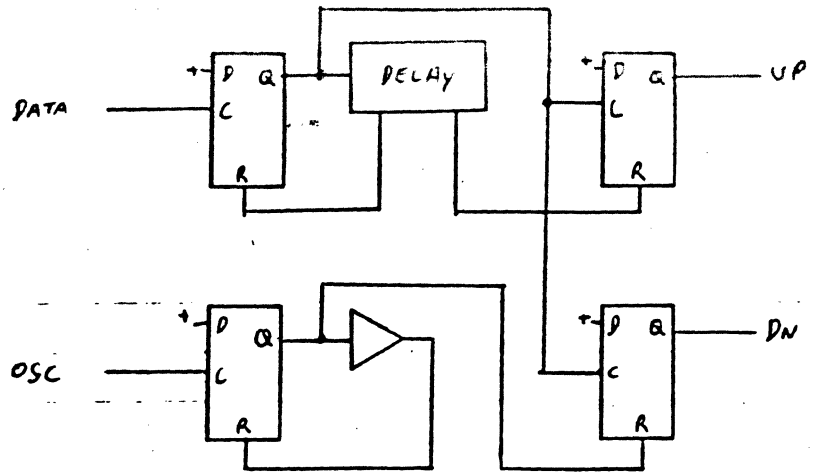


FIG 11.17A

SECOND VERSION HARMONIC Φ DETECTOR

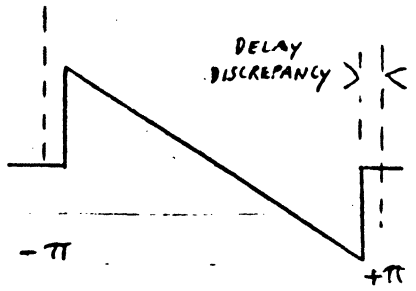


FIG 11.17C

PHASE TRANSFER CURVE

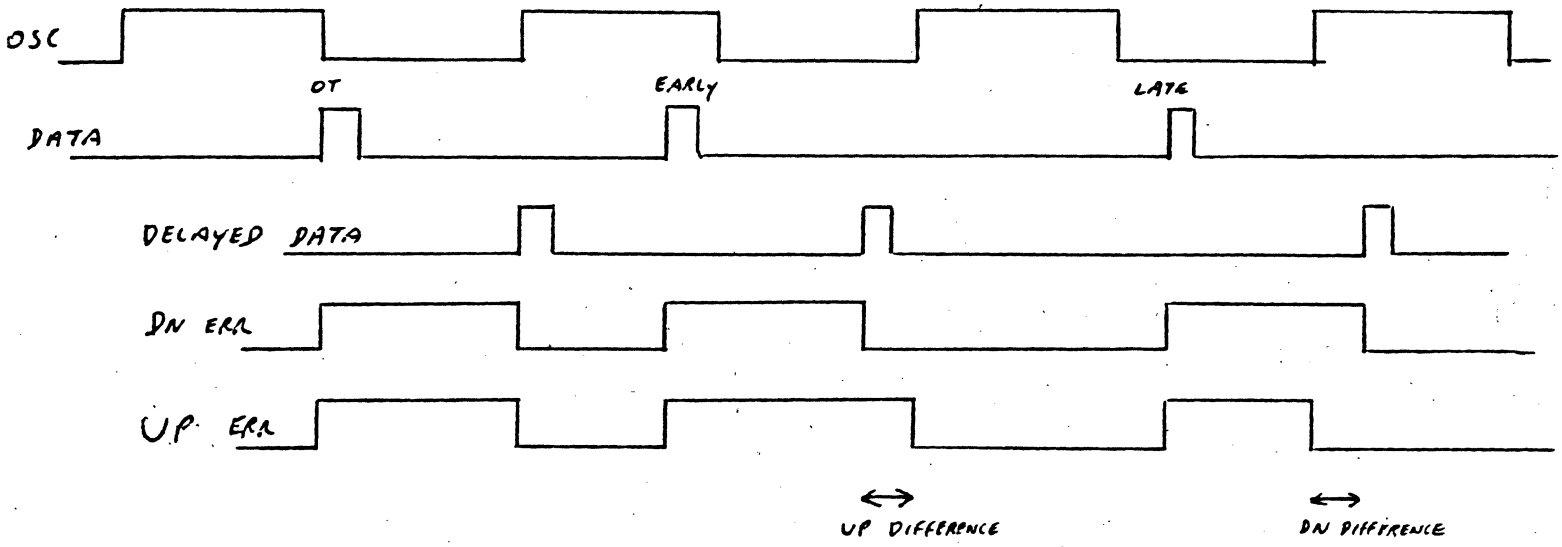


FIG 11.17 B

SECOND VERSION HARMONIC Φ DETECTOR WAVEFORMS

alf the oscillator period the phase transfer curve is truncated . As
seen, the need for a conditioning circuit causes the phase transfer curve
less than ideal due to the fixed delays versus the variable oscillator
d and/or the input frequency.

tor Interface

output of all the phase detectors illustrated so far are voltage pulses.
interface to the filter sometimes calls for a current. If this is the
irement then the voltage output must be converted to a current. Where
ow pulse widths are expected, as will occur in the circuits of Figs. 11.14,
.5, and 11.16, the current conversion circuit must have very wide bandwidth.
rent switches of both polarities are often used, such as in Fig. 11.19, or
mall capacitor can be added to a resistive convertor to "store" some of the
rgy of narrow pulses before integration, such as in Fig. 11.20. This will
come more obvious as we discuss filter circuits. The gain of these detectors
e $\frac{V \text{ Logic Swing}}{\pi \text{ radians}}$ for the voltage circuits and $\frac{I_s}{\pi \text{ rad}}$ for the current
itch forms.

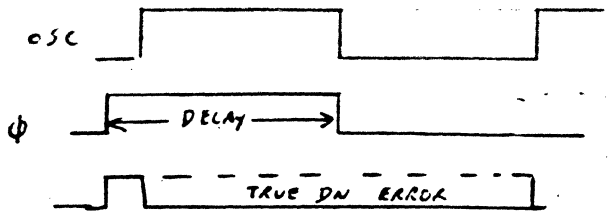
other type phase detector is the sample and hold. It requires a time varying
itage driven by the oscillator, which is always of the same slope, and a sample
rcuit. These are inherently Harmonic detectors in that the sample is always
itiated by the incoming data. The pulse width of the sample gate must be
all compared to the oscillator half period. Also the ramp must be symmetrical
ound some reference. Some phase locked loops are built around a ramp oscil-
tor which automatically provides the time varying ramp. One problem with
slope ramps is that a very fast return edge is required. This could be a
y fast capacitor discharge (Fig. 11.22) or it could be a 180^0 phase reversal

of a symmetrical triangular wave (Fig. 11.23). The latter type are easily obtained from the oscillator by using an amplifier similar to that which we developed to handle two separate inputs as in Fig. 10.4.

The sample gate must be able to handle the full swing of the ramp and pass the charging or discharging currents into the hold capacitor within the period of the sample. There are two kinds. The first is illustrated in Fig. 11.24 and is a transformer driven diode bridge. It can handle both the positive and the negative portions of a ground referenced symmetrical ramp, as well as the discharge and charge currents of the holding capacitor for bidirectional samples.

Another form that is currently popular is the analog switch shown in Fig. 11.25. This circuit has series resistance and therefore requires careful consideration of the RC time constants of the hold capacitor and $R_{DS_{on}}$ of the Fet. The phase transfer characteristics of these circuits is shown in Fig. 11.26. The limitations are the sample period and the bandwidth restrictions to the fast return slope.

There are no commercial versions of harmonic phase detectors available. However, the non harmonic types already referred to can be made harmonic by the addition of the en able gate structure shown in Fig. 11.16A that is made to block the oscillator input in the absence of data via an AND gate or the reset input to the lower "D" FF.



TRUNCATED DOWN ERROR DUE TO

DELAY > $\frac{T_{OSC}}{2}$

FIG 11.18

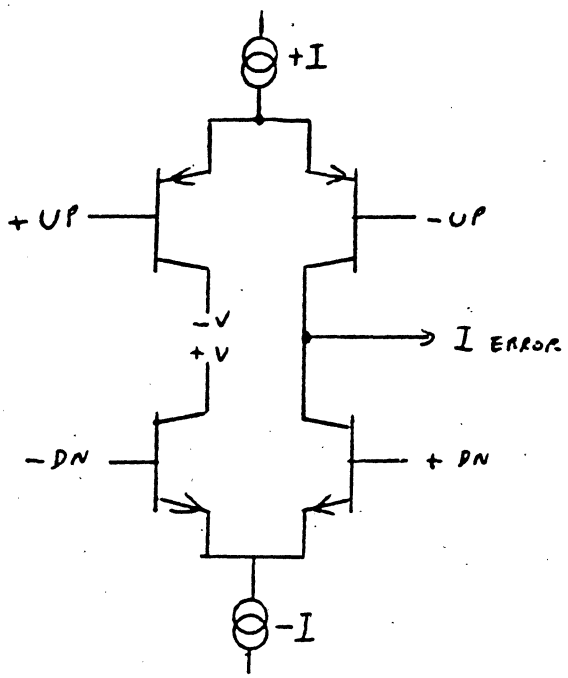


FIG 11.19

CURRENT SWITCH INTERFACE

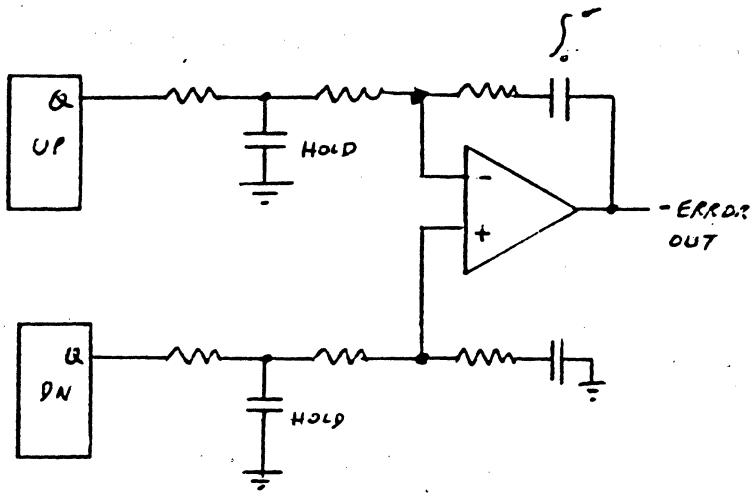


FIG 11.20

VOLTAGE-CURRENT & FILTER

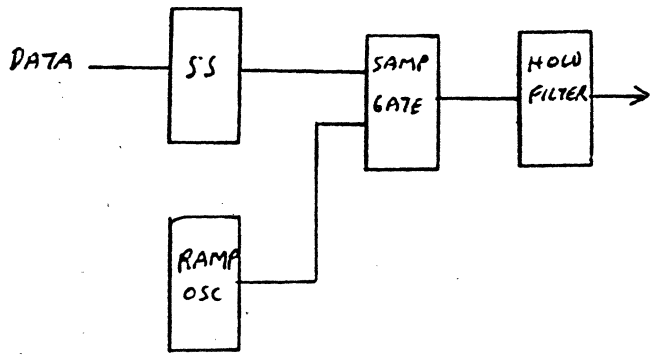


FIG 11.21

SAMPLE GATE Φ DETECTOR

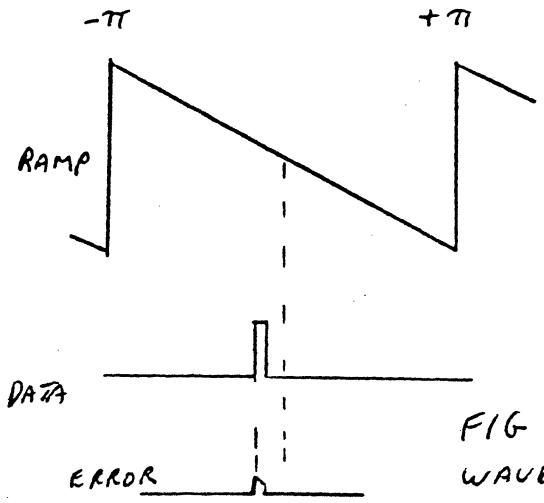


FIG 11.21 A

WAVEFORMS

Oscillators

These are all either voltage or current controlled oscillators. Their purpose is to produce an output frequency that is a function of some control input. They can take the form of controlled multivibrators, controlled LC oscillators, or controlled sawtooth oscillators. There are a large number of commercial types available and many other circuits using discreet components that can be built. Except for the linear LC type oscillator, their frequency period is subject to the noise around a threshhold amplitude where the level of charge on a capacitor is used as one extreme of the oscillator output swing. Some commercial types require very careful power supply filtering or isolation in order to reduce their susceptibility to injection locking, even though separate pins are provided for the oscillator power and the output driver powers inputs. Very careful layout and component placement is required for best stability or minimum phase jitter. This is particularly true for the control input which is the error voltage or current.

A voltage controlled multivibrator may be constructed from a bidirectional SS circuit with positive feedback. The circuit is shown in Fig. 11.27 and can be designed for either ECL outputs or T^2L output, depending on the positive supply and the resistor ratios R_1, R_2 used for the clamp. Sensitivity can be altered by changing the ratio between R_3, R_4, and R_5. As the value of R_3 is lowered, the change in frequency as a function of the control voltage is smaller.

The frequency is determined by the clamp voltage, the current source values, and the capacitor value.

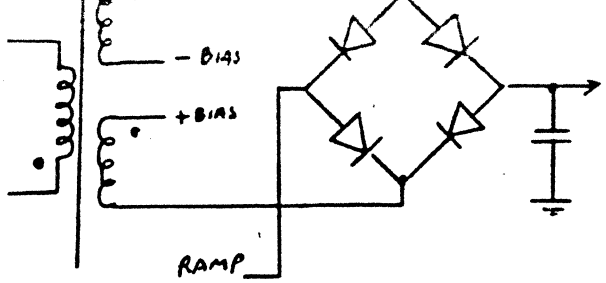


FIG 11.24
SAMPLE GATE CIRCUIT

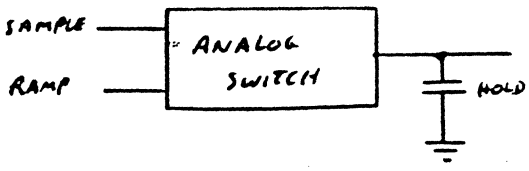


FIG 11.25
ANALOG SWITCH SAMPLE GATE

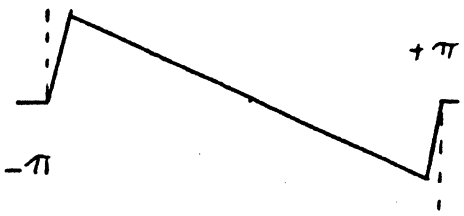


FIG 11.26
PHASE TRANSFER CURVE FOR RAMP SAMPLE GATE Φ DETECTOR

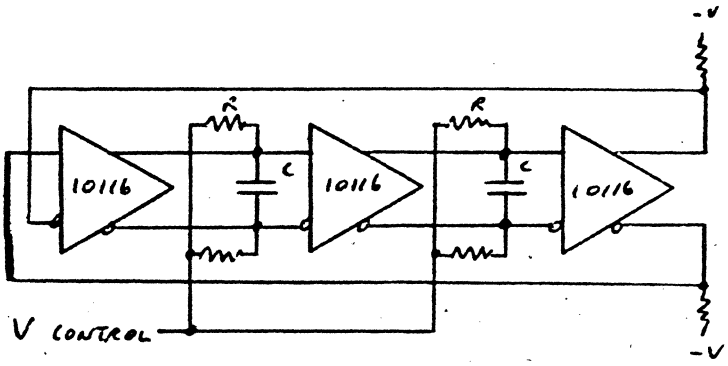


FIG 11.28
DESCRETE + IC TYPE MULTIVIBRATOR

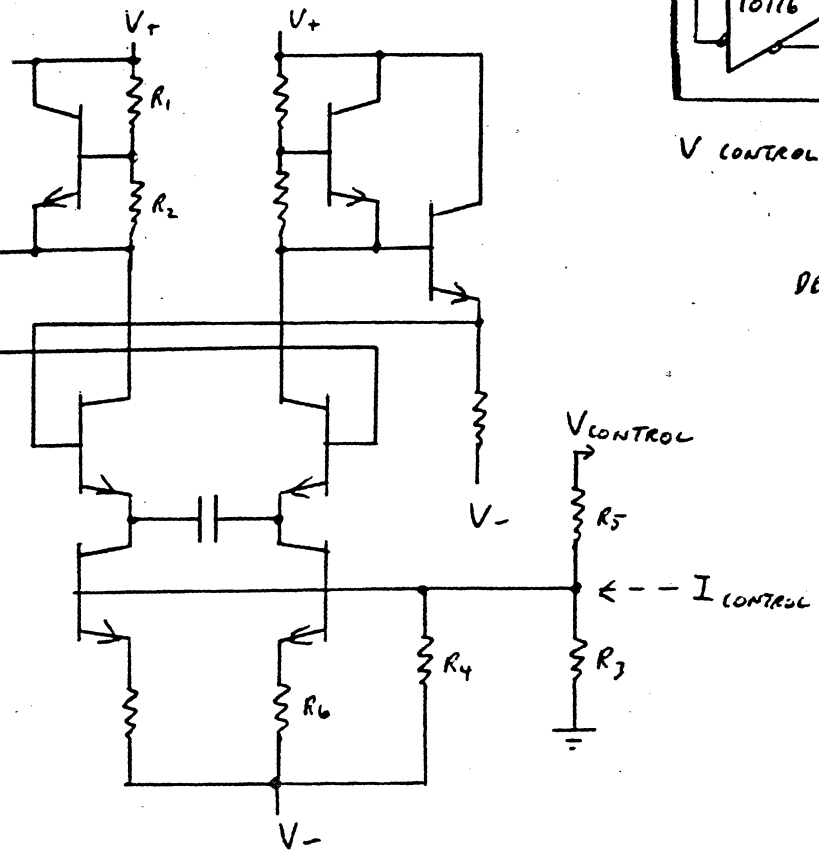


FIG 11.27
MULTIVIBRATOR OSCILATOR
WITH VOLTAGE CONTROL OR I CONTROL

$$F = \frac{1}{2T} \simeq \frac{I_{source}}{V_{be}\left(\frac{R_1}{R_2} + 1\right) C} \qquad \text{(EQ 11.26)}$$

and

$$I_{source} = \frac{\frac{(V_{IN} - V_T) R_T}{R_5} - V_{be}}{R_6} \qquad \text{(EQ 11.27)}$$

$$I_{source} = \frac{\frac{\left[V_{IN} - \frac{(V_-)(R_3)}{R_3 + R_4}\right] \frac{R_3 R_4}{R_3 + R_4}}{R_5} - V_{be}}{R_6} \qquad \text{(EQ 11.28)}$$

Another ECL oscillator can be built using the line receiver 10116. Here the discharge current for the capacitors is provided from the four emitter return resistors supplied from the control voltage.

A simplified schematic is shown in Fig. 11.29A. If we refer to the waveforms of 11.29B, then we can see the operation of the oscillator. The most positive level is clamped by the output emitter followers. The actual voltage is determined by the Vbe drop resulting from the emitter current. The other side of the capacitor would normally be pulled down to around -1.8 volts, but the capacitor will not allow this change until sufficient charge is accumulated via the resistors and the control voltage. The non-conducting base, B3, voltage will be held steady at -0.6 volts while the conducting base, B7, will be pulled above 0 volts by the action of the capacitor. The capacitor discharges to the point where the base voltages are equal, which initiates the reversal. The gain of the other stages increases the slope of the RC waveforms around the transition region in order to improve stability. One of the greatest concerns is that the base voltage of the conducting transistor is above ground while its collector is around -1.0 volts, clearly saturated. INTERNAL + CLAMPING FORCES THIS LEVEL MORE NEGATIVE RESULTING IN A MORE LINEAR OPERATION BUT FREQUENCY IS AFFECTED BY THE CLAMPING ACTION

A sawtooth oscillator can be built using discrete components. One of its biggest problems is the flyback circuit and the time for flyback. It is shown in Fig. 11.30A. Saturation storage time can be minimized by using gold doped transistors or schottky clamps. The flyback transistor base pulse width is controlled by the propogation time thru the comparator and can be extended by the use of a capacitor C_2.

The frequency is controlled by the current source Q1 as controlled by the control voltage Vc. The peak of the waverform exceeds the comparator voltage due to the comparator response time. The base drive for Q_2 is increased by the emitter follower Q_3. As can been seen, there are a lot of tolerances or dependencies that affect the frequency. These type oscillators can only be used

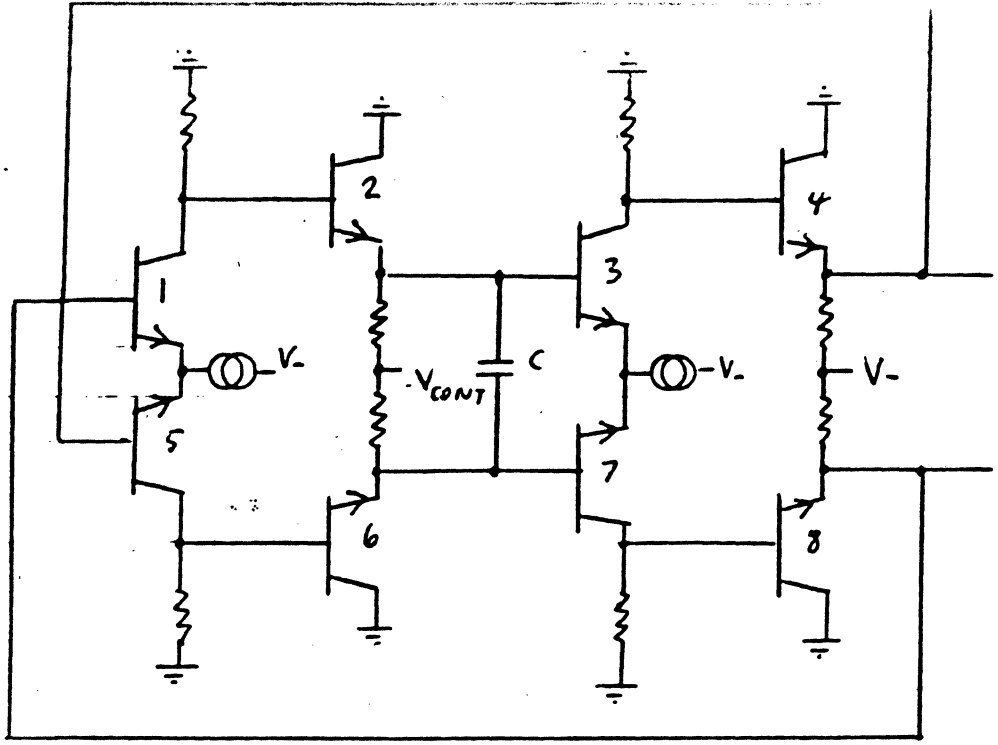


FIG 11·29A

SIMPLIFIED CIRCUIT DIAGRAM

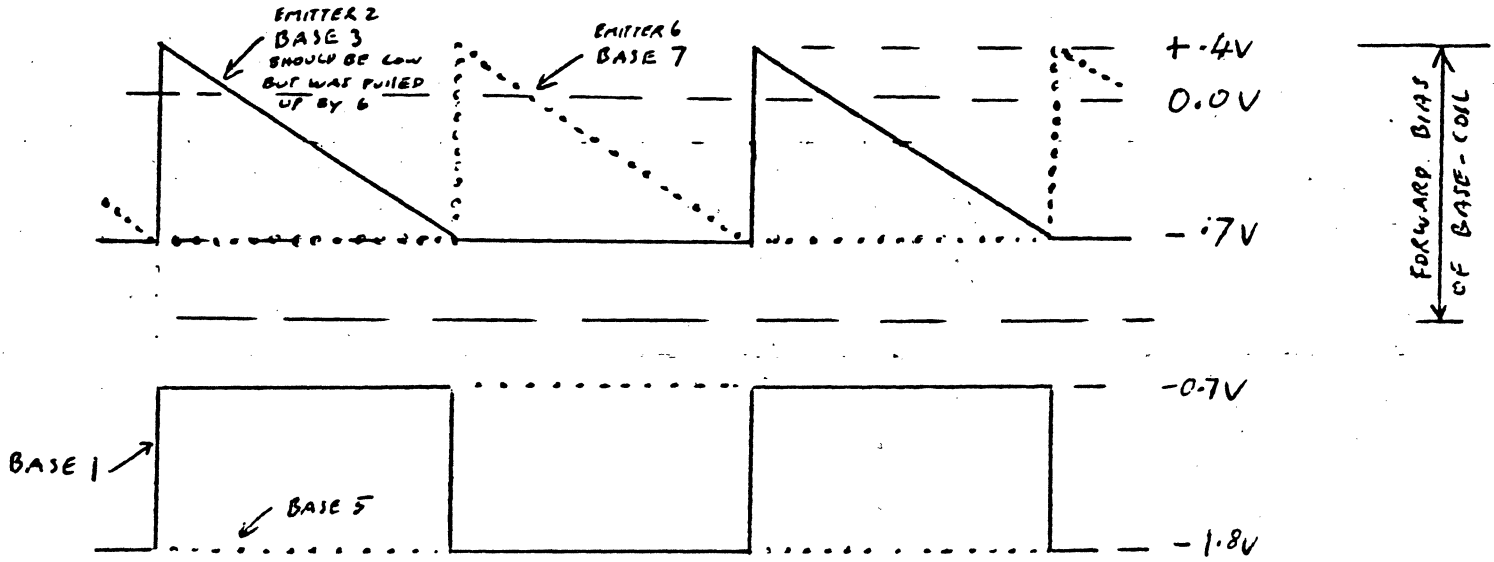


FIG 11·29B

BIAS LEVELS OF BASE AND COLLECTOR. (CLAMPING IGNORED)

THE ABOVE IS TO SHOW THAT THIS CIRCUIT HAS POOR FREQUENCY CONTROL AS THE DISCHARGE IS ALSO A FUNCTION OF THE FORWARD BIASED BASE – COLLECTOR JUNCTION AND IS THEREFORE DEVISE AND TEMPERATURE SENSITIVE.

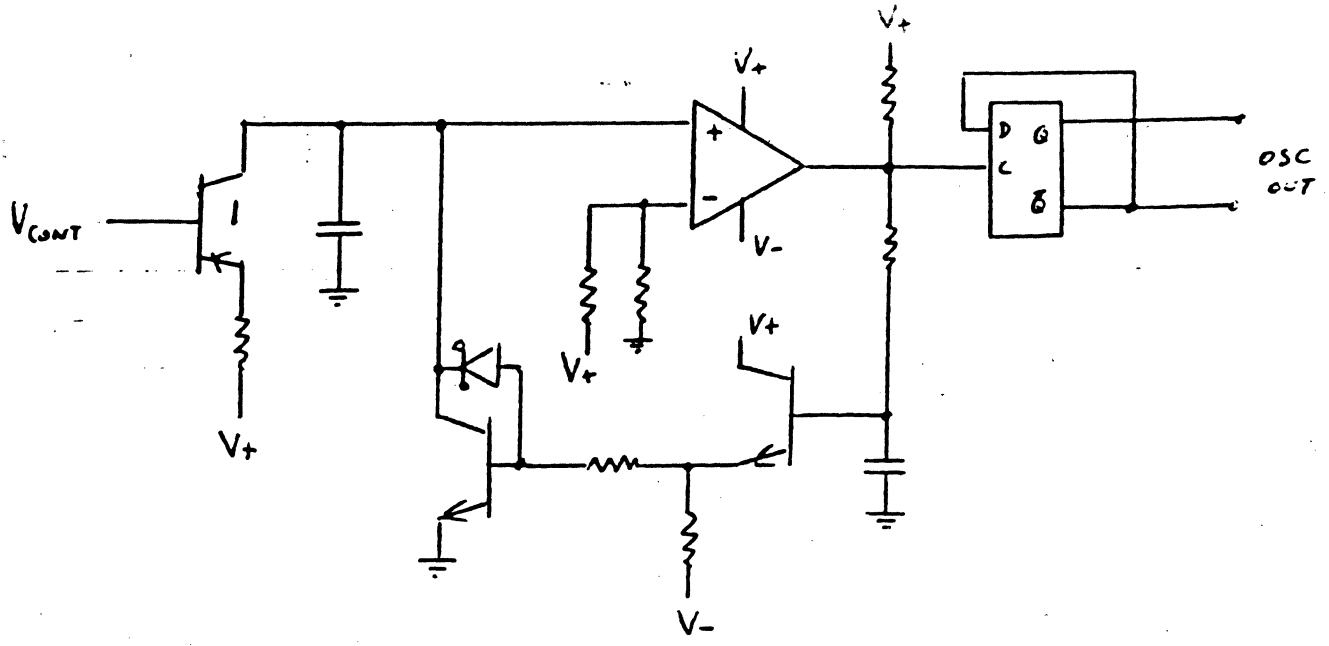


FIG 11·30 A

RAMP OSCILATOR VOLTAGE CONTROLED FREQUENCY

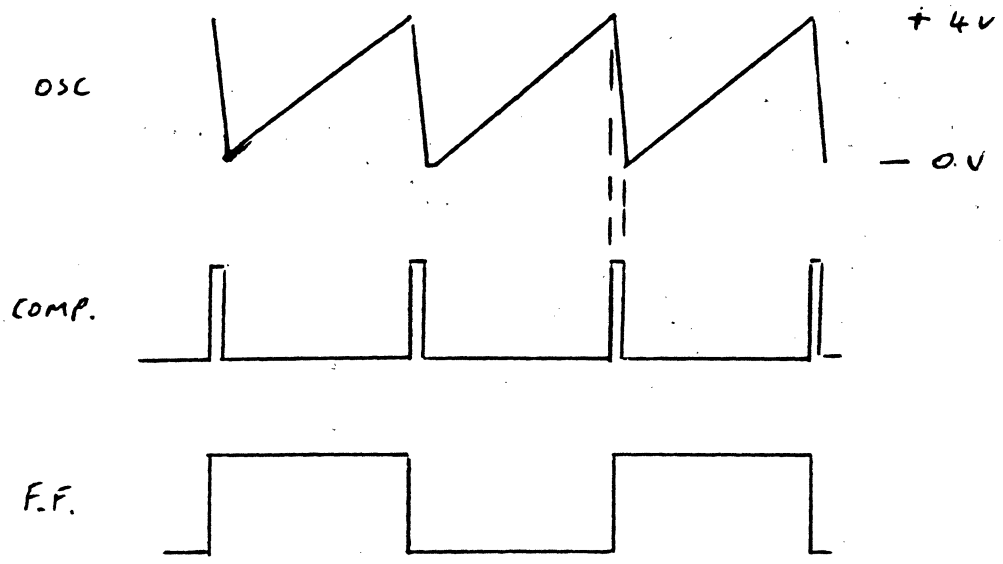


FIG 11·30 B

RAMP OSCILATOR WAVE FORMS

for "low" frequency work meaning below a few Mega Hertz. Above this, the flyback time takes an appreciable portion of the cycle thereby altering the phase transfer curve of the phase detector. Some of the commercial versions of square wave oscillators include the M(4024; 1648, 1658, 74S124, and 74LS124). Of these the MC1648 is an LC version oscillator that requires a voltage variable capacitor to control the frequency. Data for these are contained in their respective data sheets and will not be discussed here.

The gain of the oscillators is expressed in radians per second per volt, or radians per second per amp. depending on the type filter.

FILTERS The purpose of the filter is primarily to provide some bandwidth limitations while providing the desired poles and zeros for stability. If we look at $G_{(s)}$, it contains a single S term in the denominator from the frequency to phase conversion. This by itself provides a pole at the origin making a type 1 loop without adding any other components. The response time of the loop to a step change in phase is shown for second order systems back in Figs. 11.5 and 11.6. Knowing the overshoot permitted and the repsonse time for settling, the bandwidth can be obtained from the graphs. The filters take three basic forms.

The first filter, Fig. 11.31A, interfaces the logic blocks producing the up and down errors as voltage pulses. Its transfer function is stated below. The effect of C_1 is to capture the narrow error pulses that the OP amp cannot respond to.

$$V_{up} = I_1\left(\frac{R_1}{2} + \frac{1}{C_1 s}\right) - I_2\left(\frac{1}{C_1 s}\right) \qquad \text{(EQ 11.29)}$$

$$O_v = -I_1\left(\frac{1}{C_1 s}\right) + I_2\left(\frac{1}{C_1 s} + \frac{R_1}{2}\right) \qquad \text{(EQ 11.30)}$$

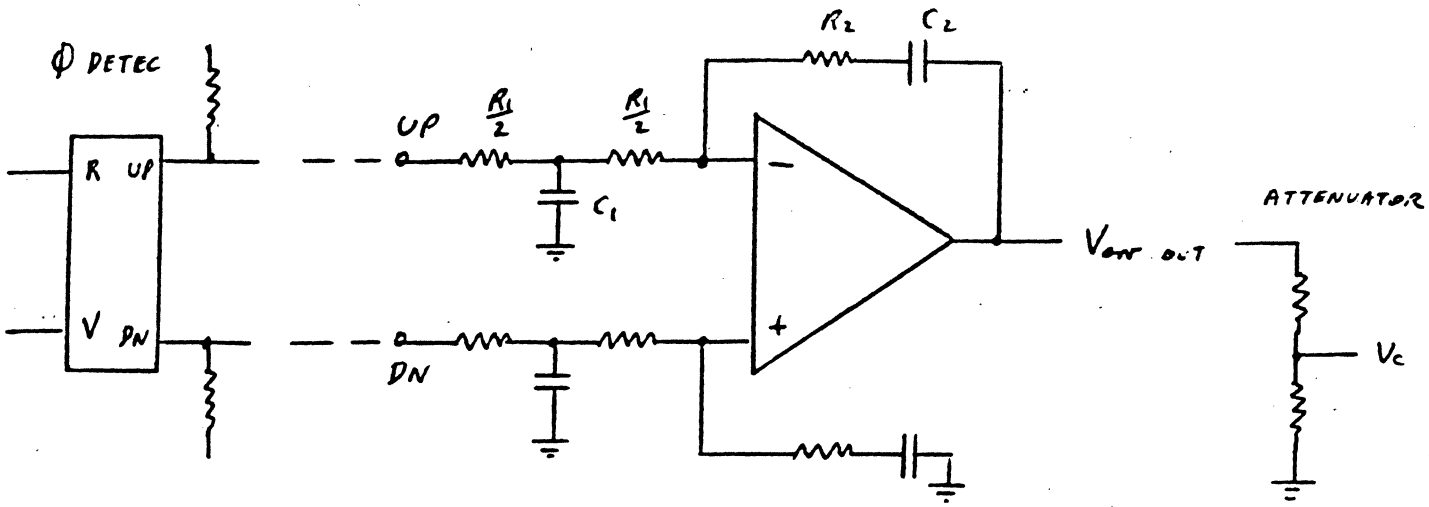


Fig 11.31 A

VOLTAGE PULSE φ ERROR FILTER
TWO POLES ONE ZERO

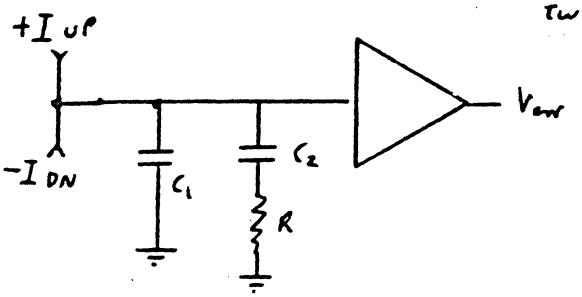


FIG 11.31 B

CURRENT SWITCH φ ERROR
FILTER 2 POLES 1 ZERO

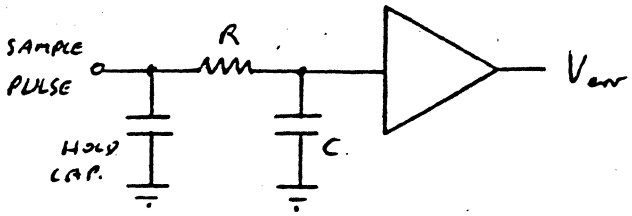


FIG 11.31 C

SAMPLE AND HOLD FILTER
SAMPLE + 1 POLE

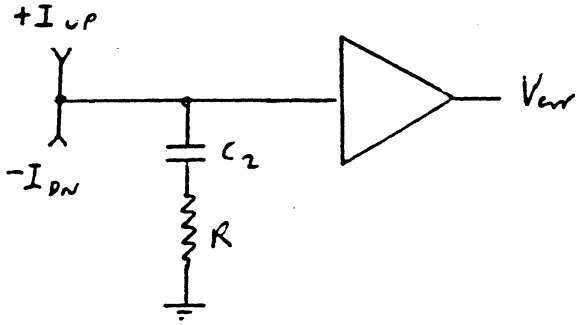


FIG 11.31 D

CURRENT SWITCH φ ERROR
FILTER 1 POLE 1 ZERO

$$\therefore I_2 = \frac{\begin{vmatrix} R_1 + \frac{1}{C_1 S} & V_{IN} \\ -\frac{1}{C_1 S} & 0 \end{vmatrix}}{\begin{vmatrix} R_1 + \frac{1}{C_1 S} & -\frac{1}{C_1 S} \\ -\frac{1}{C_1 S} & \frac{R_1}{2} + \frac{1}{C_1 S} \end{vmatrix}} = \frac{\frac{V_{IN}}{C_1 S}}{\left(\frac{R_1}{2} + \frac{1}{C_1 S}\right)\left(\frac{R_1}{2} + \frac{1}{C_1 S}\right) - \frac{1}{C_1^2 S^2}} \quad \text{(EQ 11.31)}$$

$$= \frac{V_{IN}}{CS\left(\frac{R_1^2}{4} + \frac{R_1}{C_1 S}\right)} = \frac{V_{IN}}{R_1\left(\frac{R_1}{4} CS + 1\right)} \quad \text{(EQ 11.32)}$$

$$I_F = \frac{V_o}{R_2 + \frac{1}{C_2 S}} = \frac{V_o C_2 S}{R_2 C_2 S + 1} \quad \text{(EQ 11.33)}$$

For an ideal OP. amp $\quad I_2 = I_F \quad$ (EQ 11.34)

$$\therefore \frac{V_{IN}}{R_1\left(\frac{R_1}{4} CS + 1\right)} = \frac{V_o C_2 S}{R_2 C_2 S + 1} \quad \text{(EQ 11.35)}$$

$$\frac{V_o}{V_{IN}} = K_F = \frac{R_2 C_2 S + 1}{R_1 C_2 S\left(\frac{R_1 C_1}{4} S + 1\right)} \quad \text{(EQ 11.36)}$$

Notice that this filter has a zero at $\frac{1}{R_2C_2}$, a pole at $\frac{1}{R_1C_1}$, and a gain of $\frac{1}{R_1C_2}$.

The filters of Fig. 11.31B can be analyzed by considering the voltage out is

$$V_o = I_{err} Z_F = I_{err}\left(\frac{\frac{1}{C_1 S}\left(\frac{1}{C_2 S} + R\right)}{\frac{1}{C_1 S} + \frac{1}{C_2 S} + R}\right) = \quad \text{(EQ 11.37)}$$

$$Z_F = \frac{R C_2 S + 1}{C_2 S + C_1 S + R C_1 C_2 S^2} \quad \text{(EQ 11.38)}$$

$$= \frac{R C_2 S + 1}{S\left(C_1 + C_2\right)\left(\frac{R C_1 C_2}{C_1 + C_2} S + 1\right)} \quad \text{(EQ 11.39)}$$

Which is a pure, integrator, a zero at $\frac{1}{RC_2}$, a pole at $\frac{RC_1C_2}{C_1+C_2}$, and a gain of $\frac{1}{C_1+C_2}$. As can be imagined, the added S in the denominator can be used to change a type 1 to a type 2 loop.

The filter of Fig. 11.31C can be similarly analyzed. Here the imput is a voltage such as might be stored on a holding capacitor C_1

$$V_o = \frac{\frac{q_1}{C_1} \cdot \frac{1}{C_2 S}}{\frac{1}{C_1 S} + R + \frac{1}{C_2 S}} = \frac{q S\left(\frac{1}{C_1 + C_2}\right)}{S\left(R C_1 C_2 S + C_1 + C_2\right)\left(\frac{1}{C_1 + C_2}\right)} \quad \text{(EQ 11.40)}$$

$$= \frac{V}{\left(C_1 + C_2\right)\left(\frac{R C_1 C_2}{C_1 + C_2} S + 1\right)} \quad \text{(EQ 11.41)}$$

Again there is a single pole at $\frac{1}{RC_1C_2}$, and a gain of $\frac{1}{C_1+C_2}$. This would be fine for a type 1 system as it does not introduce a pole at the origin. The filter of Fig. 11.31D is for use with a current input

$$V_o = I_{in} Z_f$$

$$Z_f = \frac{1}{C_2 s} + R = \frac{RC_2 s + 1}{C_2 s} \qquad \text{(EQ 11.42)}$$

which is a single zero at $\frac{1}{RC_2}$ and a pole at the origin. AND A GAIN OF $\frac{1}{C_2}$ This particular filter is not reliable at high frequencies due to stray capacitance, which makes it look like EQ 11.39 where C_2 is C_1 of EQ 11.42 and C_1 is the stray capacitance in parallel.

There are many other variations that could be desired depending on our requirements. For our examples we will use the filters of Fig. 11.31A and B. The filters can be used if the pole associated with the stray capacitance is far removed from the zero. The last remaining block is the sample rate block. Usually with a data stream there are at least two (a maximum and a minimum) pulse rate that are subharmonics of the oscillator frequency, Therefore, there is a maximum gain and a minimum gain to be specified. Both cases should be calculated. The loop performance usually requires the maximum peak overshoot response to occur at the maximum gain therefore this value must be used to establish the loop conditions. The sample rate gain is

$$\frac{\omega_{sample}}{\omega_{osc}} = \frac{2\pi F_{sample}}{2\pi F_{osc}} = \frac{F_{sample}}{F_{osc}} \qquad \text{(EQ 11.43)}$$

for the filters we will be using, or if we used the zero order hold circuits

such as sample and hold the gain is

$$\frac{\sin \pi \left(\frac{\omega}{\omega_s}\right)}{\pi \frac{\omega}{\omega_s}} \Big/ -\pi \frac{\omega}{\omega_s} \qquad \text{(EQ 11.44)}$$

where W is the radians per second of interest and WS is the radians per second of the sample rate.

The design of a Type 2 phase locked loop is shown in Fig. 11.32. We will first specify the loop. The input shall be Fm =2.5 MHz. It shall have a maximum peak overshoot of no more than 5% and a response time of 2.0 μs to within ± 5%. The oscillator shall run at 5.0 MHz and shall be able to capture within ± 20% of 5.0 MHz and be able to follow frequency excursions of ± 5%.

If we look at the graph of Fig. 11.6, we see that a ζ of 2 meets the criteria of 5% maximum overshoot. The error comes within -5% at around a $W_n t = 0.5$. Since we want to settle within ± 5% in 2 μs, then

$$W_n = \frac{0.5}{t} = \frac{0.5}{2 \times 10^{-6}} = 2.5 \times 10^5 \text{ RADIANS} \qquad \text{(EQ 11.45)}$$

The gain of the oscillator is determined from the curve, Fig. 11.33, for the MC 1658. If we choose a nominal control voltage of -0.7 volts, the F.C. product is 950. For a frequency of 5.0 MHz, the capacitor must be

$$\frac{950 \text{ MHz} \cdot \text{pf}}{5. \text{ MHz}} = 190. \text{ pf} \qquad \text{(EQ 11.46)}$$

which can be made up of 180 pf + 10 pf or 180 plus a variable capacitor to fine tune it in. To obtain ± 20% range, the error voltage excursions must

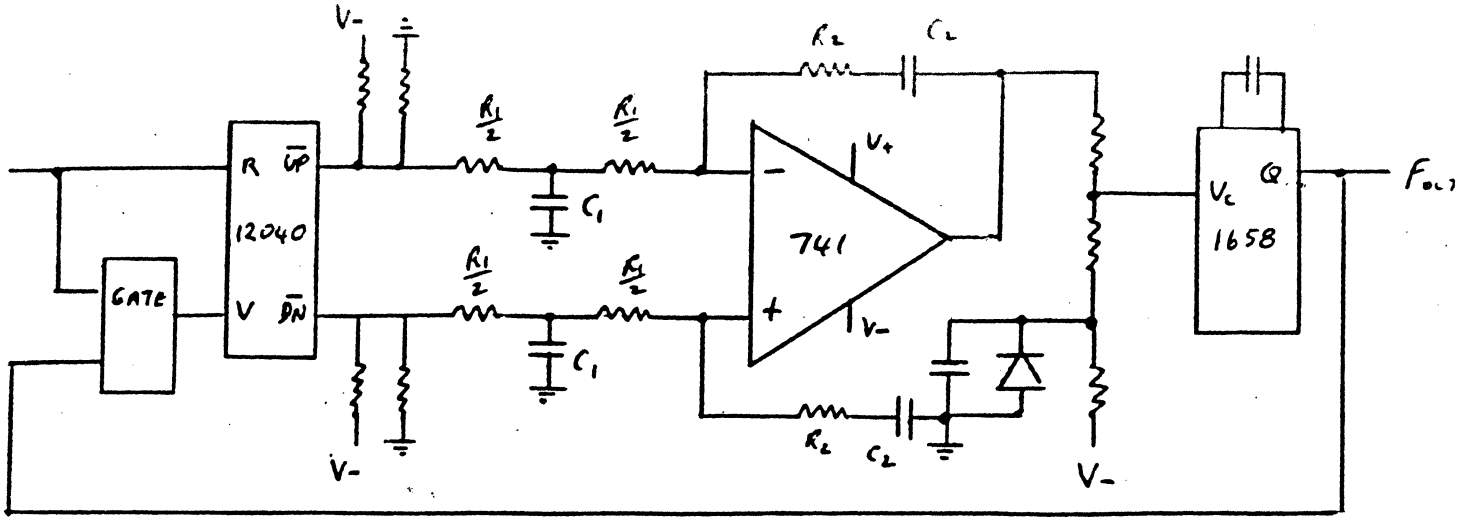


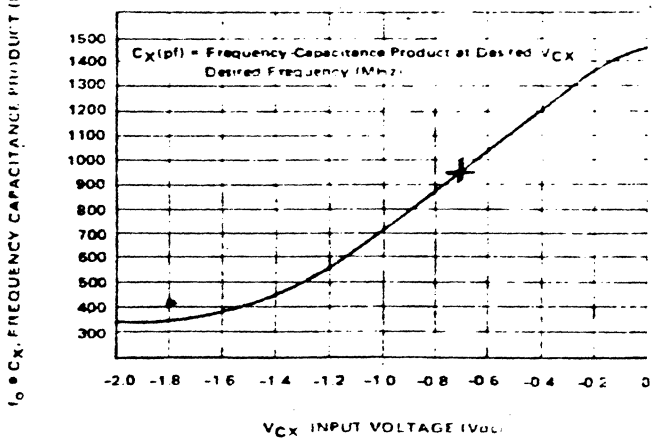
FIG 11.32

COMPLETE PHASE LOCKED LOOP ECL VERSION
PROVISION IS MADE FOR HARMONIC OPERATION

FIGURE 8 – FREQUENCY-CAPACITANCE PRODUCT versus CONTROL VOLTAGE (V_{CX})

11.33

MC 1650 OSCILATOR



include

$$950 \pm (950)(.2) = 950 \pm 190 \qquad \text{(EQ 11.47)}$$

on the curve. These control voltages represent at 1140, up to -0.45 volts and down to 760 at -0.95 volts. The gain Ko of the oscillator is in radians per second per volt

$$K_o = \frac{2\pi\,(1200 - 710)\ \text{MHz}\cdot\text{pf}}{(190.\text{pf})(1.0 - 0.4)\text{v}} = 2.7\times10^{7}\ \text{rad/sec}\cdot\text{volt} \qquad \text{(EQ 11.48)}$$

The gain of the phase detector is

$$\frac{\Delta V_{\text{Logic}}}{\pi} = \frac{1.737 - .885\ \text{V}}{\pi} = \frac{0.8525}{\pi} = 0.2713\ \text{V/rad} \qquad \text{(EQ 11.49)}$$

The gain of the K sample is

$$\frac{2.5\ \text{MHz}}{5.0\ \text{MHz}} = \frac{1}{2} \qquad \text{(EQ 11.50)}$$

The gain of the filter is Kf

$$K_F = \frac{R_2 C_2 S + 1}{R_1 C_2 S\left(\frac{R_1 C_1}{4} S + 1\right)} \qquad \text{(EQ 11.36)}$$

To get ζ and Wn specified, we need to write the entire equation. We will include an attenuator Koe as we may need it.

$$\frac{\phi_o}{\phi_{in}} = \frac{G(s)}{1 + G(s)} = \frac{K_\phi K_s K_F K_\alpha \frac{K_o}{s}}{1 + K_\phi K_s K_F K_\alpha \frac{K_o}{s}} \qquad \text{(EQ 11.51)}$$

$$\frac{\phi_o}{\phi_{in}} = \frac{\left(0.2713\ V/rad\right)\left(\frac{1}{2}\right)\left(\frac{R_2C_2S+1}{R_1C_2S\left(\frac{R_1C_1}{4}S+1\right)}\right)\left(K_d\right)\left(\frac{2.7\times10^7\ rad/sec\ volt}{S}\right)}{1+\left(0.2713\right)\left(\frac{1}{2}\right)\left(\frac{R_2C_2S+1}{R_1C_2S\left(\frac{R_1C_1}{4}S+1\right)}\right)\left(K_d\right)\left(\frac{2.7\times10^7\ rad/sec\ volt}{S}\right)} \quad \text{(EQ 11.51) continued}$$

$$= \frac{\left(3.662\times10^6\right)\frac{K_d}{S}\left(\frac{R_2C_2S+1}{R_1C_2S\left(\frac{R_1C_1}{4}S+1\right)}\right)}{1+\left(3.662\times10^6\right)\frac{K_d}{S}\left(\frac{R_2C_2S+1}{R_1C_2S\left(\frac{R_1C_1}{4}S+1\right)}\right)}$$

If we look at the characteristic equation of the denominator, we can determine the response

$$C.E. = R_1C_2S^2\left(\frac{R_1C_1}{4}S+1\right) + 3.66\times10^6\,\frac{K_d}{S}\left(R_2C_2S+1\right) \quad \text{(EQ 11.52)}$$

$$\frac{R_1^2C_1C_2S^3}{4} + R_1C_2S^2 + 3.66\times10^6\,K_d\,R_2C_2S + 3.66\times10^6\,K_d \quad \text{(EQ 11.53)}$$

which is a third order equation which was forced on us by the inclusion of C_1 used to improve the repsonse to very narrow phase errors. If we ignored this C_1 and rewrote the equation making sure in our design that the pole $\frac{4}{R_1C_1}$ is more than a decade above the zero, we can proceed with a second order solution.

$$K_F = \frac{R_2 + \frac{1}{C_2S}}{R_1} = \frac{R_2C_2S+1}{R_1C_2S} \quad \text{(EQ 11.54)}$$

now rewriteing equation 11.52

$$C.E. = 1 + \frac{3.662\times10^6}{S}K_d\left(\frac{R_2C_2S+1}{R_1C_2S}\right) = 0 \quad \text{(EQ 11.55)}$$

$$R_1 C_2 s^2 + 3.66\times10^6 K_a (R_2 C_2) s + 3.66\times10^6 K_a = 0 \qquad \text{(EQ 11.56)}$$

$$s^2 + \frac{(3.662\times10^6 K_a)(R_2 C_2)}{R_1 C_2} s + \frac{3.662\times10^6 K_a}{R_1 C_2} = 0 \qquad \text{(EQ 11.57)}$$

Therefore, $$\sqrt{\frac{3.662\times10^6 K_a}{R_1 C_2}} = \omega_n = 2.5\times10^5 \qquad \text{(EQ 11.58)}$$

and $$2\zeta\omega_n = \frac{(3.662\times10^6 K_a)(R_2 C_2)}{R_1 C_2} \qquad \text{(EQ 11.59)}$$

$$(2)(2)(2.5\times10^5) = 10^6 = \frac{3.662\times10^6 K_a R_2}{R_1} \qquad \text{(EQ 11.60)}$$

Therefore, $$\frac{K_a R_2}{R_1} = \frac{10^6}{3.662\times10^6} = 0.273 \qquad \text{(EQ 11.61)}$$

going back to $$\omega_n^2 = \frac{3.662\times10^6 K_a}{R_1 C_2} = (2.5\times10^5)^2 \qquad \text{(EQ 11.62)}$$

or $$\frac{K_\alpha}{R_1 C_2} = \frac{(2.5\times10^{3})^2}{3.662\times10^{6}} = 1.706\times10^{4} \qquad \text{(EQ 11.63)}$$

Now we know that $$\frac{K_\alpha}{R_1} = \frac{0.273}{R_2} = (1.706\times10^{4})C_2 \qquad \text{(EQ 11.64)}$$

If we choose $C_2 = 1.0 \times 10^{-8}$ farads, then

$$R_2 = \frac{0.273}{(1.706\times10^{4})(10^{-8})} = 1.60\times10^{3} \text{ or } 1600.\,\Omega \qquad \text{(EQ 11.65)}$$

which also means that $$\frac{K_\alpha}{R_1} = (1.706\times10^{4})(10^{-8}) = 1.706\times10^{-4}\,\Omega \qquad \text{(EQ 11.66)}$$

so if K_α were 1, then R_1 would be

$$R_1 = \frac{1}{1.706\times10^{-4}} = 5.86\times10^{3} = 5860.\,\Omega \qquad \text{(EQ 11.67)}$$

Therefore we may need an attenuator for interface purposes (AS R_1 IS TOO LARGE). Now we have all the parameters we need for design. We should review the change in voltage out of the OP amp to see if it can drive the attenuator for the frequency range since we need some interface to the IC.

If $K_\alpha = \frac{1}{5}$ then $R_1 = 1172\,\Omega$ (EQ 11.67A)

$$5(\Delta V_{CONT}) = (0.95 - 0.45)5 = 2.5\ V \text{ swing total} \qquad \text{(EQ 11.68)}$$

Now we know R_1, R_2, and C_2, we can designate the locations of the zero.

$$\frac{1}{R_L C_L} = \frac{1}{T_2} = \frac{1}{(1.60\times10^3)(10^{-8})} = 6.25\times10^4 \text{ rad} \qquad \text{(EQ 11.69)}$$

$$\text{or} \quad \frac{6.25\times10^4}{2\pi} = 9.94\times10^3 \text{ Hz} \quad \text{or} \simeq 10.\text{KHz}$$

We can now draw the Bode diagram knowing all the information we have. We now need an equation to describe the zero gain crossing Wcn of the -40 db/decade (-12 db/oct) slope determined by the S^2 in the denominator.

$$\sqrt{K_T} = \sqrt{K_s K_\varphi K_F K_a K_o} = \omega_{cn} \text{ in radians} \qquad \text{(EQ 11.70)}$$

$$= \sqrt{\left(\tfrac{1}{2}\right)(0.2713)\left(\frac{1}{R_1 C_2}\right)\left(\frac{1}{5}\right)(2.7\times10^7)}$$

$$= \sqrt{\tfrac{1}{2}\left(0.2713\right)\left(\frac{1}{(1.172\times10^3)10^{-8}}\right)\left(\frac{1}{5}\right)(2.7\times10^7)} = 2.5\times10^5 \qquad \text{(EQ 11.71)}$$

which checks with EQ 11.45. The zero is located at 6.25 x 10^4 rad, the phase margin at the 0 db crossing of the open loop is 86.4 as obtained from Fig. 11.33B and

$$\Phi_m = \tan^{-1}(\omega_{odb} T_2) = \tan^{-1}(10^6)(1.6\times10^{-5}) \qquad \text{(EQ 11.72)}$$

$$= \tan^{-1} 16 = 86.4^\circ$$

We should investigate the phase error that results from using the OP amp. The high frequency gain of the OP amp is $\frac{R_2}{R_1} = \frac{1.60 \times 10^3}{1.172 \text{ ro}^3} = 1.36$

Bode plot for type 2 second order phase lower loop

Fig 11.33 B

The manual for a μa 741 shows a phase shift, closed loop, at a gain of 1.36 of around 45° at 0.7 MHz or 4.4 M rad above our ^ Odb crossing which dictates a loss of about 12.8° phase margin and some increase in overshoot and settling time. The same goes for any pole we might add by placing a C_1 in position.

$$\text{If } \frac{1}{T_p} = \frac{4}{R_1 C_1} = 10^7 \qquad \text{(EQ 11.73)}$$

$$C_1 = \frac{4}{10^7 (1.172 \times 10^3)} = 3.41 \times 10^{-10} \text{ fds}$$

which is the largest capacitor, we could add half way along the R_1 or at the 586 Ω to 586 Ω junction without subtacting more than 5° from the phase margin. This would place the spread between the zero and pole of $\frac{wp}{wz} = \frac{107}{6.25 \times 10}4 = 160$ which is adequate. Now the phase jitter is described by the equation for a step ϕ error of π

$$\phi_{o(s)} = \frac{\phi_{in(s)}}{s}\left[\frac{G(s)}{1+G(s)}\right] = \frac{\pi}{s}\left[\frac{\frac{1}{2}(0.2713)\left(\frac{1}{1.172\times10^{-3}}\right)\left(\frac{1}{s}\right)\left(2.7\times10^{7}\right)\left(\frac{1.6\times10^{-5}s+1}{s^2}\right)}{1+\frac{1}{2}(0.2713)\left(\frac{1}{1.172\times10^{-3}}\right)\left(\frac{1}{s}\right)\left(2.7\times10^{7}\right)\left(\frac{1.6\times10^{-5}s+1}{s^2}\right)}\right] \qquad \text{(EQ 11.74)}$$

$$= \frac{\pi}{s}\left[\frac{6.25\times10^{10}\left(\frac{1.6\times10^{-5}s+1}{s^2}\right)}{1+6.25\times10^{10}\left(\frac{1.6\times10^{-5}s+1}{s^2}\right)}\right] = \frac{\pi}{s}\left[\frac{6.25\times10^{10}\left(1.6\times10^{-5}s+1\right)}{s^2+\left(6.25\times10^{10}\right)\left(1.6\times10^{-5}\right)s+6.25\times10^{10}}\right] \qquad \text{(EQ 11.75)}$$

$$= \frac{\pi}{s}\left[\frac{\omega_n^2\left(1.6\times10^{-5}s+1\right)}{s^2+2\zeta\omega_n s+\omega_n^2}\right] = \frac{\pi}{s}\left[\frac{6.25\times10^{10}\left(1.6\times10^{-5}s+1\right)}{s^2+1.0\times10^{6}s+6.25\times10^{10}}\right] \qquad \text{(EQ 11.76)}$$

$$\frac{\pi}{S}\left[\frac{\omega_n^2\left(\frac{2\zeta\omega_n}{\omega_n^2}S+1\right)}{S^2+2\zeta\omega_n S+\omega_n^2}\right] \qquad \text{(EQ 11.77)}$$

The solution is (EQ 11.78)

$$\mathcal{L}^{-1}\phi_{(o)(s)} = \pi\left\{1+e^{-\omega_n\zeta t}\left[\frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_n\sqrt{1-\zeta^2}\,t-\cos\omega_n\sqrt{1-\zeta^2}\,t\right]\right\}$$

As can be seen as long as we keep the bandwidth up to improve the response to a step change in phase, we introduce Ø errors in the oscillator which always occur when reading written transitions due to both noise and pulse interferences as discussed before. The ideal solution would be a system that would lock up fast and then revert to a low bandpass loop while reading transistions associated with the customers data. This is accomplished by having a preable prior to the data area to be used for locking the phase locked loop then changing the location of the zero, perferably, to lower the bandpass, and hence the jitter.

Now in the example just cited, the zero is the result of R_2 and C_2 around the OP amp. We could change the location of the zero without changing the gain. If we examine $G_{(s)}$, the gain of the filters is $\frac{1}{R_1C_2}$ which means we could lower the zero by only changing R_2. We could do this with a Fet if we could accept the

transient associated with the C_{iss} of the Fet. If we did this, we could see from the resulting Bode plot that we really need to lower the gain at the same time, in other words we need to lower Wcn the same amount that we lower Wz, but $Wcn = \sqrt{K_t}$ therefore, we need to look elsewhere to do the job or allow greater time for lockup or allow a compromize.

There is a better filter that can easily be used that allows both gain and Tz to be changed by only changing one component. If we use the filter of Fig. 11.31B in conjunction with current convertors for the phase detectors, we get interesting results.

$$G(s) = K_s K_\phi K_F K_\alpha \frac{K_o}{S} \qquad \text{(EQ 11.79)}$$

Now the gain of K_ϕ is in $\frac{amps}{radian}$ or I source/π radians. The gain of the filter is given in EQ 11.39.

$$G(s) = \left(\frac{1}{2}\right)\left(\frac{I_{amps}}{\pi}\right)\left(\frac{RC_2S+1}{S(C_1+C_2)\left(\frac{RC_1C_2}{C_1+C_2}S+1\right)}\right)\left(K_\alpha\right)\left(\frac{2.7\times10^7 \text{ rad/acc.V}}{S}\right) \qquad \text{(EQ 11.80)}$$

Therefore, the characteristic equation is

$$C.E. = 1 + G(s) = 1 + \left(4.297\times10^6\right)\left(I_{amps}\right)\left(K_\alpha\right)\left[\frac{RC_2S+1}{S^2(C_1+C_2)\left(\frac{RC_1C_2}{C_1+C_2}S+1\right)}\right] \qquad \text{(EQ 11.81)}$$

$$= S^2(C_1+C_2)\left(\frac{RC_1C_2}{C_1+C_2}S+1\right) + 4.297\times10^6 I K_\alpha\left(RC_2S+1\right) \qquad \text{(EQ 11.82)}$$

$$= RC_1C_2S^3 + \left(C_1 + C_2\right)S^2 + 4.297\times10^6\left(I\right)\left(K_a\right)RC_2S + 4.297\times10^6\, I\, K_a \quad \text{(EQ 11.83)}$$

which is a third order equation again. We can take two approaches. The first is the same as before where the pole is widely separated from the zero by at least 100.

$$\text{Now} \quad T_2 = RC_2 \quad \text{and} \quad T_p = \frac{RC_1C_2}{C_1+C_2} \quad \text{(EQ 11.84)}$$

$$100 = \frac{T_2}{T_p} = \frac{RC_2\left(C_1+C_2\right)}{RC_1C_2} = \frac{C_1+C_2}{C_1}$$

$$\frac{C_2}{C_1} = 100-1 = 99 \quad \text{(EQ 11.85)}$$

If we take the same parameters as before, the filter would be that of Fig. 11.31D which has an impedance as given in EQ 11.42.

$$G_{(s)} = \left(4.297\times10^6\right)\left(I\right)\left(K_a\right)\left[\frac{RC_2S+1}{C_2S^2}\right] \quad \text{(EQ 11.86)}$$

$$C.E. = 1 + \left(4.297\times10^6\right)\left(I\right)\left(K_a\right)\left(\frac{RC_2S+1}{C_2S^2}\right) \quad \text{(EQ 11.87)}$$

$$= S^2 + 4.297\times10^6\left(I\,K_a\right)RC_2S + \frac{4.297\times10^6\, I\, K_a}{C_2} \quad \text{(EQ 11.88)}$$

$$\omega_n^2 = \left(2.5\times10^5\right)^2 = \frac{4.297\times10^6\, I\, K_a}{C_2}$$

$$\omega_n^2 = (2.5\times10^5)^2 = \frac{4.297\times10^6\, I\, K_\alpha}{C_2} \qquad \text{(EQ 11.89)}$$

$$I\, K_\alpha = \frac{C_2 (2.5\times10^5)^2}{4.297\times10^6} = (1.454\times10^{-4})\, C_2 \qquad \text{(EQ 11.90)}$$

$$2\zeta\omega_n = (2)(2)(2.5\times10^5) = 10^6 = \frac{4.297\times10^6\, I\, K_\alpha R \chi_\alpha}{\chi_2} \qquad \text{(EQ 11.91)}$$

If we let $C_2 = 10^{-8}$ farads as before, then

$$I\, K_\alpha = 1.454\times10^{-4} \qquad \text{(EQ 11.92)}$$

$$10^6 = (4.297\times10^6)(1.454\times10^{-4})\, R \qquad \text{(EQ 11.93)}$$

$$\therefore R = \frac{10^6}{(4.297\times10^6)(1.454\times10^{-4})} = 1.60\times10^3\ \Omega \qquad \text{(EQ 11.94)}$$

which looks familiar. We could raise the current by changing C_2 and C_1 as $\frac{I}{C_2}$ as part of the gain. The current source could be 145. μa without an attenuator a 1.45 ma with a 10:1 attenuator which should be better unless we run into linearity and range problems. C_1 would be

$$\frac{C_2}{99} = 101.\ pF \qquad \text{(EQ 11.95)}$$

Now we can calculate Wcn = $\sqrt{K_t}$ as

$$\omega_{cn} = \sqrt{\frac{(4.297\times10^{6})(1.45\times10^{-4})}{10^{-8}}} = 2.5\times10^{5}\ \text{RADIANS} \qquad \text{(EQ 11.96)}$$

$$T_2 = RC_2 = (1.6\times10^{3})(10^{-8}) = 1.6\times10^{-5} \qquad \text{(EQ 11.97)}$$

which all gives the exact same Bode diagram as Fig. 11.33β with a loss of 5^{o} phase margin at the zero db crossing except we do not need an OP amp. To get our lower Bandpass for less jitter, lets change Tz to 1.6×10^{-4}. To lower the Wz and the gain, we simply raise the resistor R of the filters by 10 and change the current of the current sources by $(10)^{-2}$ from 1.45 ma to 14.5 μa. This would work very well as the pole associated with C_1 does change therefore, the loss of phase margin is less. There is no disturbance to the error voltage by changing R unless the R switch introduces an error due to stray coupling. The whole Bode plot moves down by 1 decade meaning that the settling time is now 20 μs insted of 2 μs which is ideal after syncronization.

The second version is to make the separation of the pole and the zero 16 in order to get as good a phase margin as possible. Unfortunatly, we cannot obtain

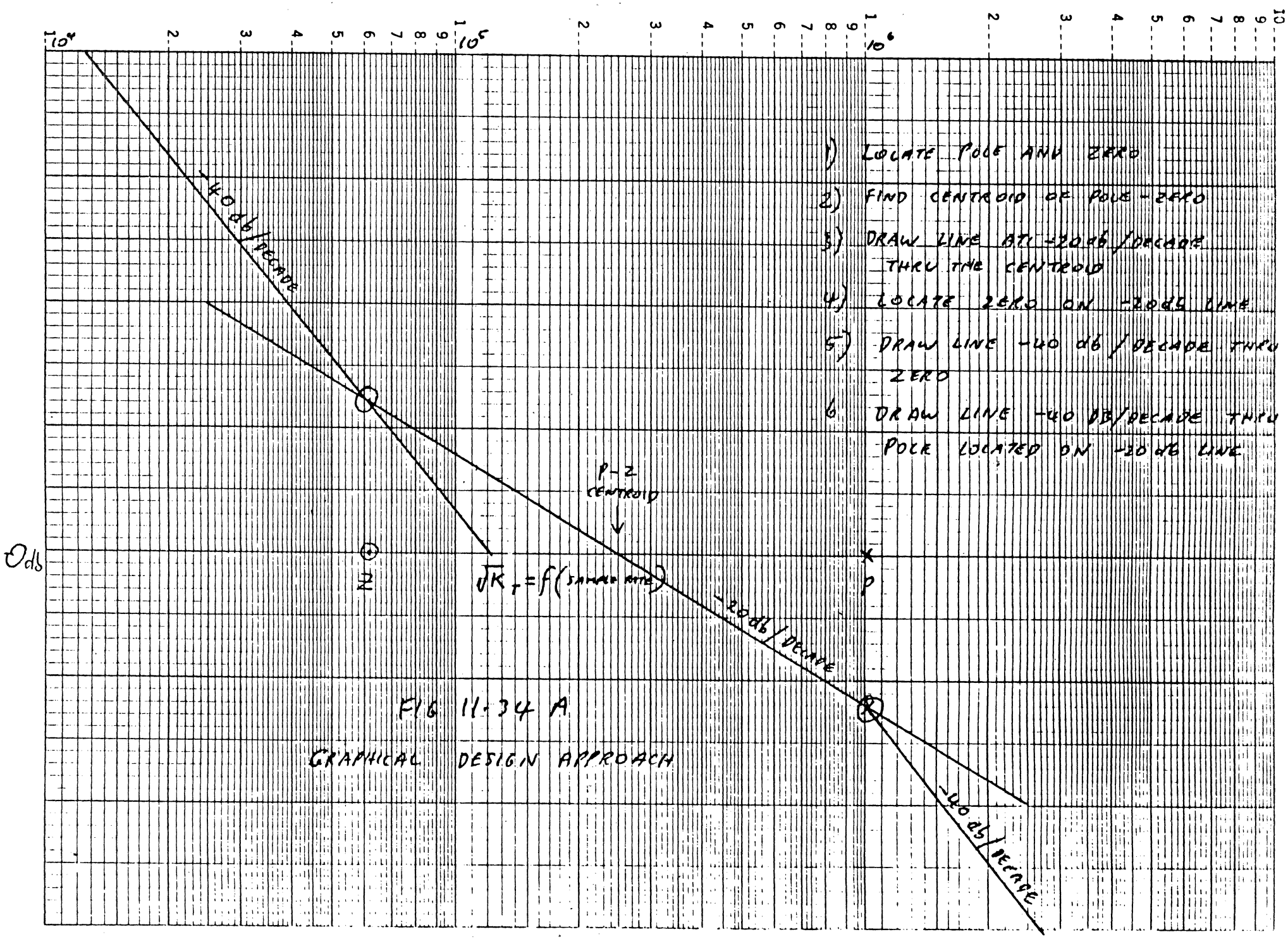


FIG 11-34 A

GRAPHICAL DESIGN APPROACH

Therefore, at the centroid we have a maximum phase margin. If the data recorded has several frequencies such as 1F, 2F, and 1-1/3F, then the 1F and 2F, in radians, should be placed on either side of the zero crossing equally spaced on log paper Fig. 11.34. This will provide the best possible phase margin for all frequencies of interest. If the spread between frequencies is large, then a wider spacing is required between the pole and zero. Notice that the gain K of the loop is changed by the data (sample) rate. Therefore, all frequencies of interest should be given the best possible phase margin. Such will be the case if, in our example, starting at EQ 11.86 the sample rate were used that corresponds to the lowest sample rate, then for all higher sample frequencies the system should be stable unless the pole is exceeded (associated with RC_1). It should be noted that the phase margin, hence ζ, is a function of the spacing of the pole and zero.

Since the solution of a third order equation is not so straight foward we will try another approach. (from G Winner and R. Spencer)

$$G(s) = \frac{T_z s + 1}{\frac{T_p}{K} s^3 + \frac{s^2}{K} + T_z s + 1}$$

$$G(j\omega) = \frac{1 + j\omega T_z}{1 - \frac{\omega^2}{K} + j\omega\left(T_z - \omega^2\left(\frac{T_p}{K}\right)\right)}$$

$$G(j\omega) = \frac{\left(1 + \omega^2 T_z^2\right)^{\frac{1}{2}}}{\left(\left(1 - \frac{\omega^2}{K}\right)^2 + \omega^2\left(T_z - T_p\frac{\omega^2}{K}\right)^2\right)^{\frac{1}{2}}}$$

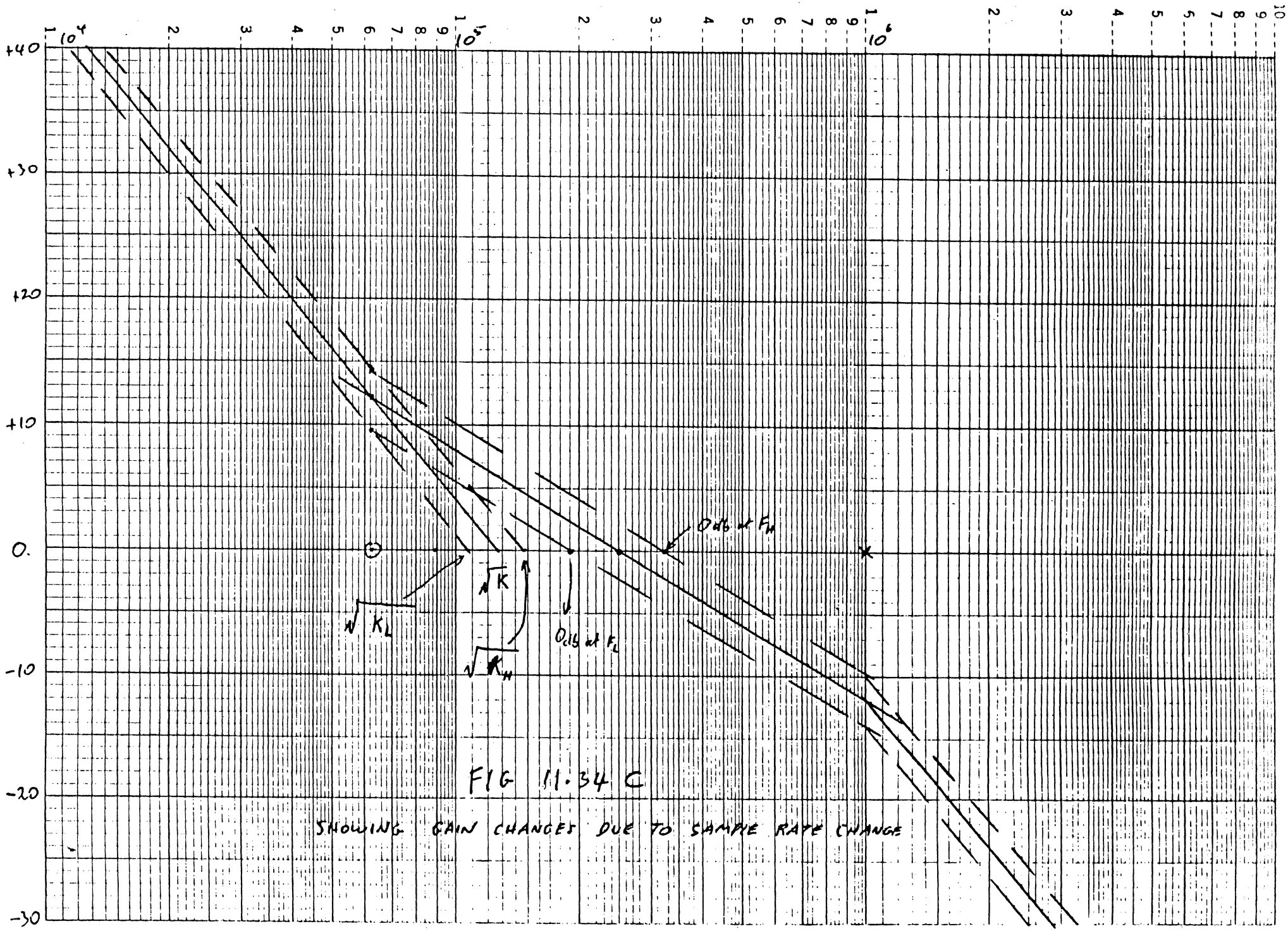


FIG 11.34 C

SHOWING GAIN CHANGES DUE TO SAMPLE RATE CHANGE

FIG 11.34B G(s) OPEN LOOP BODE PLOT FOR EQ 11.80

a phase margin sufficient for a ζ of 2, so lets see what we get if we keep the zero at the same location.

$$C_1 = \frac{C_2}{16-1} = \frac{10^{-8}\ Fds}{15} = 6.66\times10^{-10} = 666.\ pf \qquad \text{(EQ 11.98)}$$

substituting in EQ 11.81

$$C.E. = 1 + G_{(s)} = 1 + (4.297\times10^{6})(I_a)(K_2)\left[\frac{RC_2s+1}{s^2(C_1+C_2)\left(\frac{RC_1C_2}{C_1+C_2}s+1\right)}\right] \qquad \text{(EQ 11.99)}$$

$$= s^2(C_1+C_2)\left(\frac{RC_1C_2}{C_1+C_2}s+1\right) + 4.297\times10^{6} I_a K_2 (RC_2s+1) \qquad \text{(EQ 11.100)}$$

$$= s^3(RC_1C_2) + (C_1+C_2)s^2 + 4.297\times10^{6} I K_2 RCs + 4.297\times10^{6} I K_2 \qquad \text{(EQ 11.101)}$$

The usual method of making the Bode plot is to locate the centroid between the zero and the pole on the W axis on semi log paper and draw a line with a slope of -20 db per decade passing thru zero db. Then on this line locate the zero and pole and draw lines -40 db/decade passing thru each, the pole and zero making, the line thru the zero extend to the 0 db axis. The intersection of this line with the 0 db axis is equal to $\sqrt{K_t}$. The phase bulge extends from -180^0 on the left upwards peaking at the centroid and trailing back to -180^0 to the right according to the equation.

$$\theta = -180^\circ + \tan^{-1}\omega T_2 - \tan^{-1}\omega T_p \qquad \text{(EQ 11.102)}$$

rewriting

$$|G(j\omega)|^2_{\text{normalized}} = \frac{\left(1 + \frac{\hat{\omega}^2}{\hat{\omega}_z^2}\right)}{\left(1 - \hat{\omega}^2\right)^2 + \frac{\hat{\omega}^2}{\hat{\omega}_z^2}\left(1 - \frac{\hat{\omega}^2}{X}\right)^2}$$

$$\hat{\omega} = \frac{\omega}{\sqrt{K}} \qquad \hat{\omega}_z = \frac{\omega_z}{\sqrt{K}} \qquad X = \frac{\omega_p}{\omega_z}$$

There are three terms that cause peaking in the frequency domain

$$1 + \frac{\hat{\omega}^2}{\hat{\omega}_z^2} \quad , \quad \left(1 - \hat{\omega}^2\right)^2 \quad , \quad \text{and} \quad \left(1 - \frac{\hat{\omega}^2}{X}\right)^2$$

@ $\omega = \omega_z$, @ $\omega = \sqrt{K} = \omega_{3dB}$, @ $\omega = \sqrt{KX}$

Note that the $(1 - \frac{\hat{\omega}^2}{X})^2$ term would cause peaking at a much higher frequency than the other two terms. Hence, for ~~nominal~~ MINIMAL peaking, we need to balance the affects of the other two terms. For closed loop bandwidth where

$$|G(j\omega)|^2 = \frac{1}{2} \quad @ \quad \omega = \omega_{3dB}$$

and was solved numerically and shown in Fig. 11.35. Closed loop peaking

$$\text{when} \quad \frac{d|G(j\omega)|^2}{d\omega} = 0 \quad @ \quad \omega = \omega_{PK}$$

$$|G j\omega| = \text{maximum} \quad @ \quad \omega = \omega_{PK}$$

The solution is shown in Fig. 11.36, 37, and 38. For various values of x we can tabulate the peaking resulting from ignoring the S^3 term and applying the above derived corrections. These are tabulated in Table 11.1.

Table 11.1

X	Ignoring S^3	Corrected for S^3	Improvement
4	4.5 db	4.44 db	.06 db
10	2.12 db	1.74 db	.39 db
16	1.56 db	1.08 db	.78 db
20	1.37 db	.86 db	.51 db
25	1.21 db	.70 db	.51 db

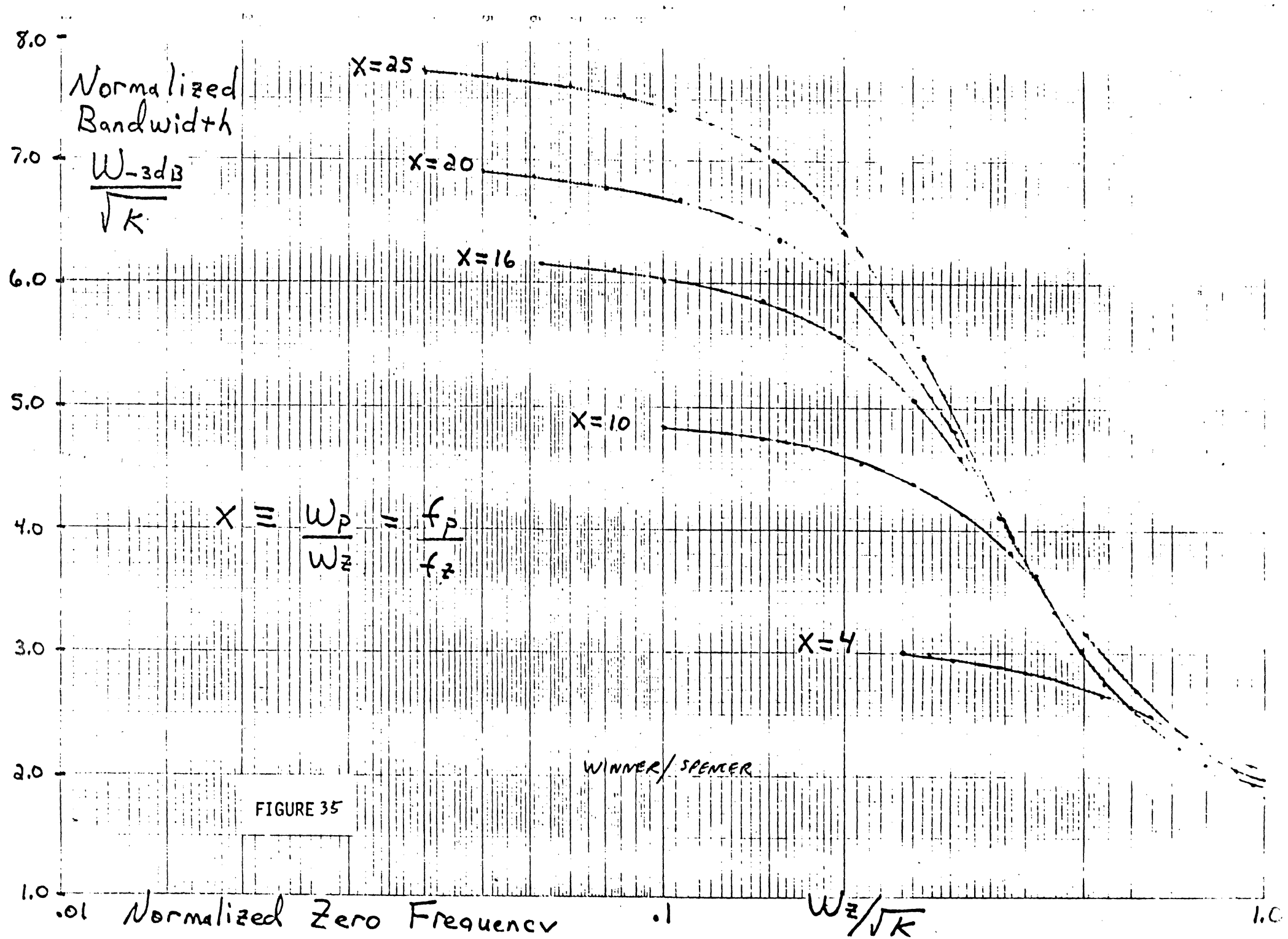


FIGURE 35

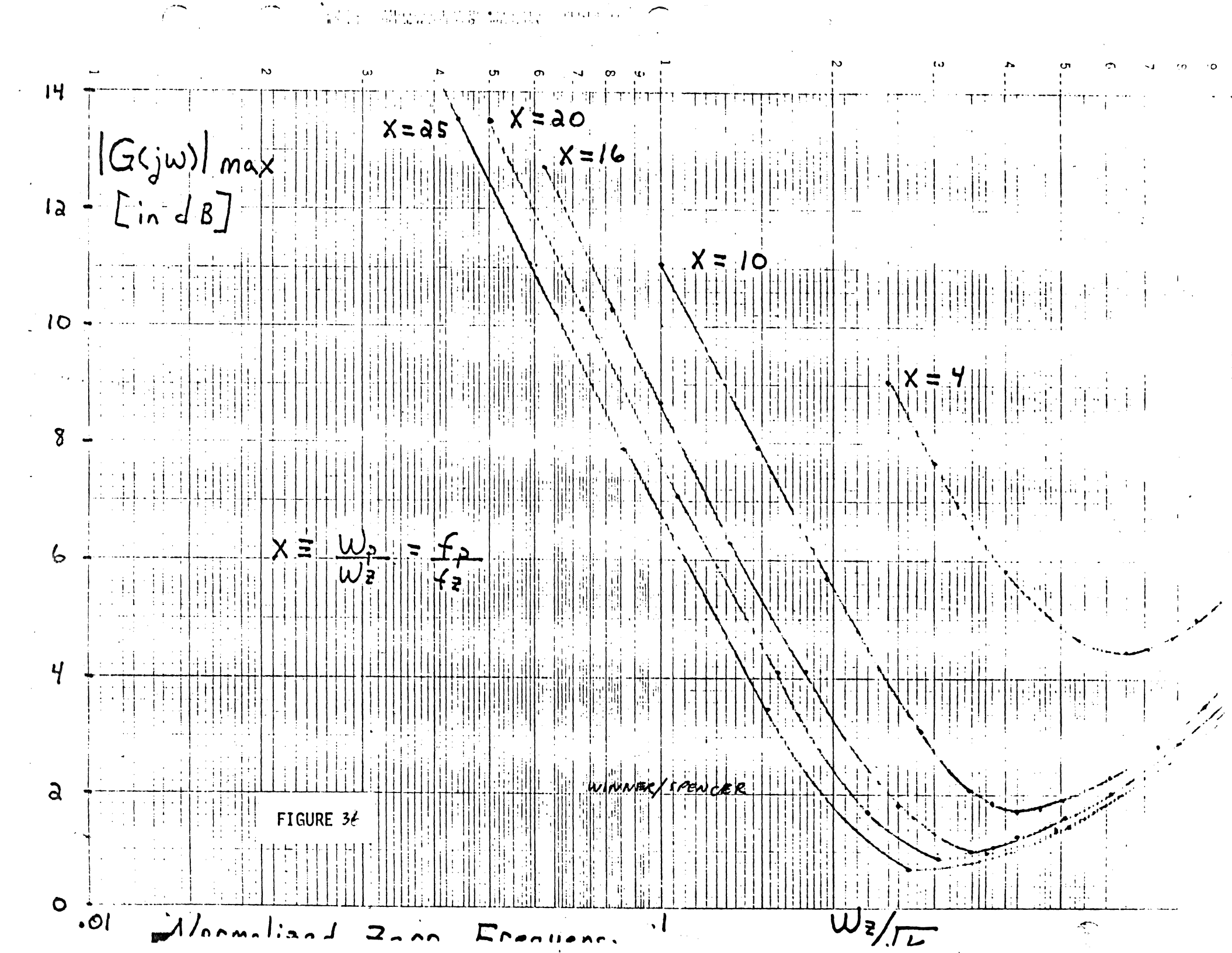

$|G(jw)|_{max}$ [in dB]
X = 25
X = 20
X = 16
X = 10
X = 4
$X \equiv \frac{w_p}{w_z} = \frac{f_p}{f_z}$
WIANNER/SPENCER
14
12
10
8
6
4
2
0
.01
.1
w_z/w_c
Normalized Zero Frequency

FIGURE 36

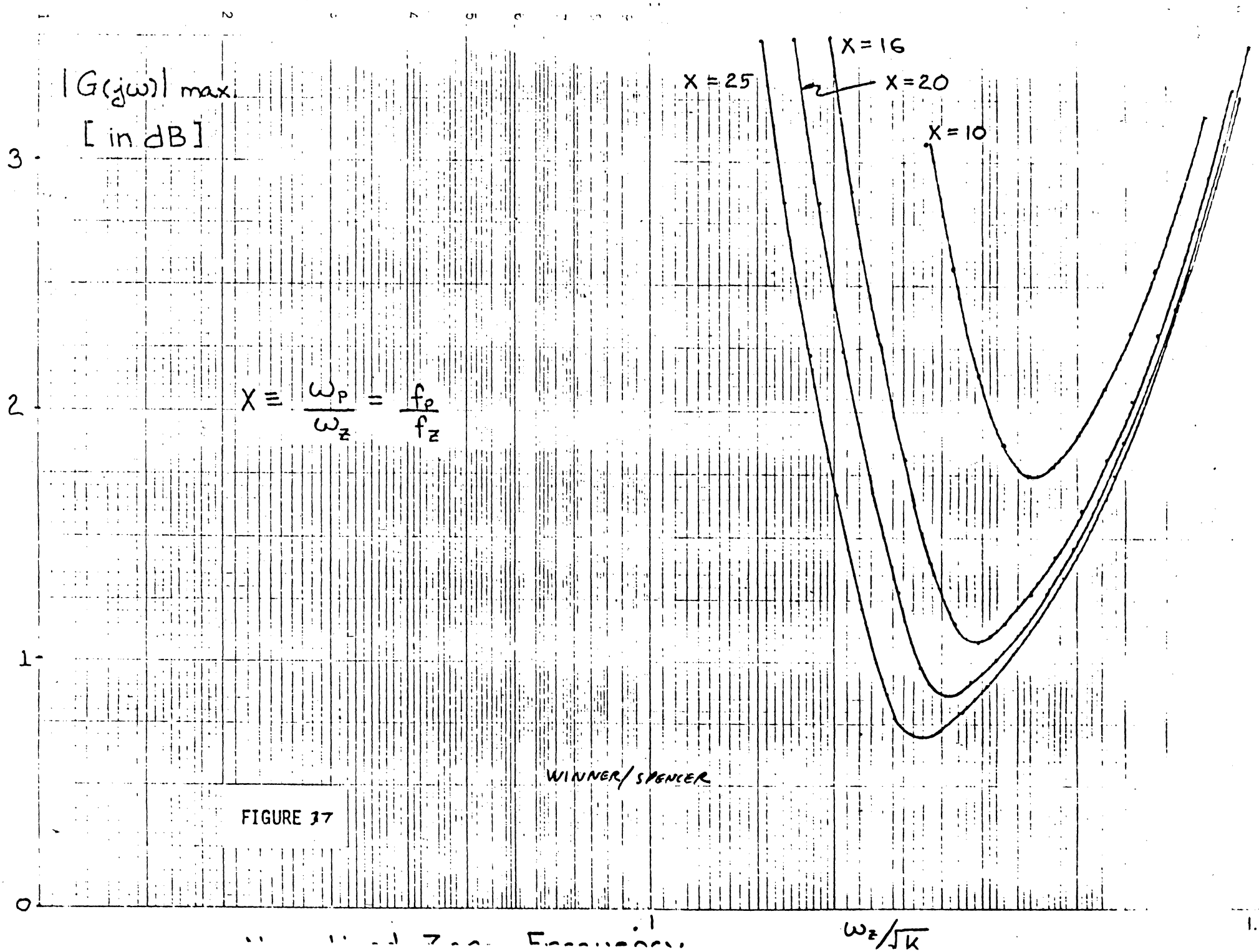


FIGURE 37

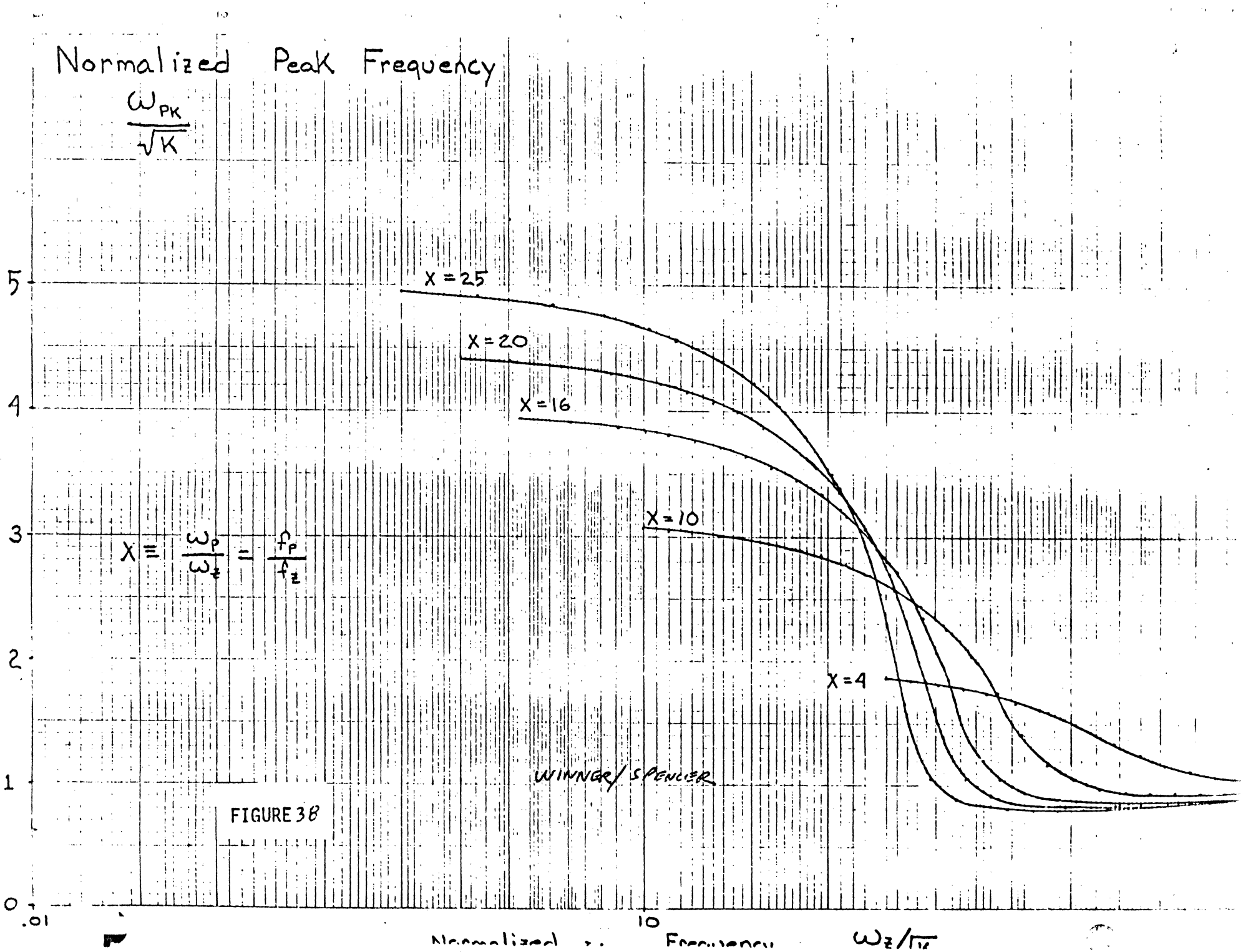

Normalized Peak Frequency
$\frac{\omega_{PK}}{\sqrt{K}}$
X = 25
X = 20
X = 16
X = 10
X = 4
$X \equiv \frac{\omega_p}{\omega_z} = \frac{f_p}{f_z}$
WINNER/SPENCER
5
4
3
2
1
0
.01
10
Normalized Frequency ω_z/ω_{PK}

FIGURE 3B

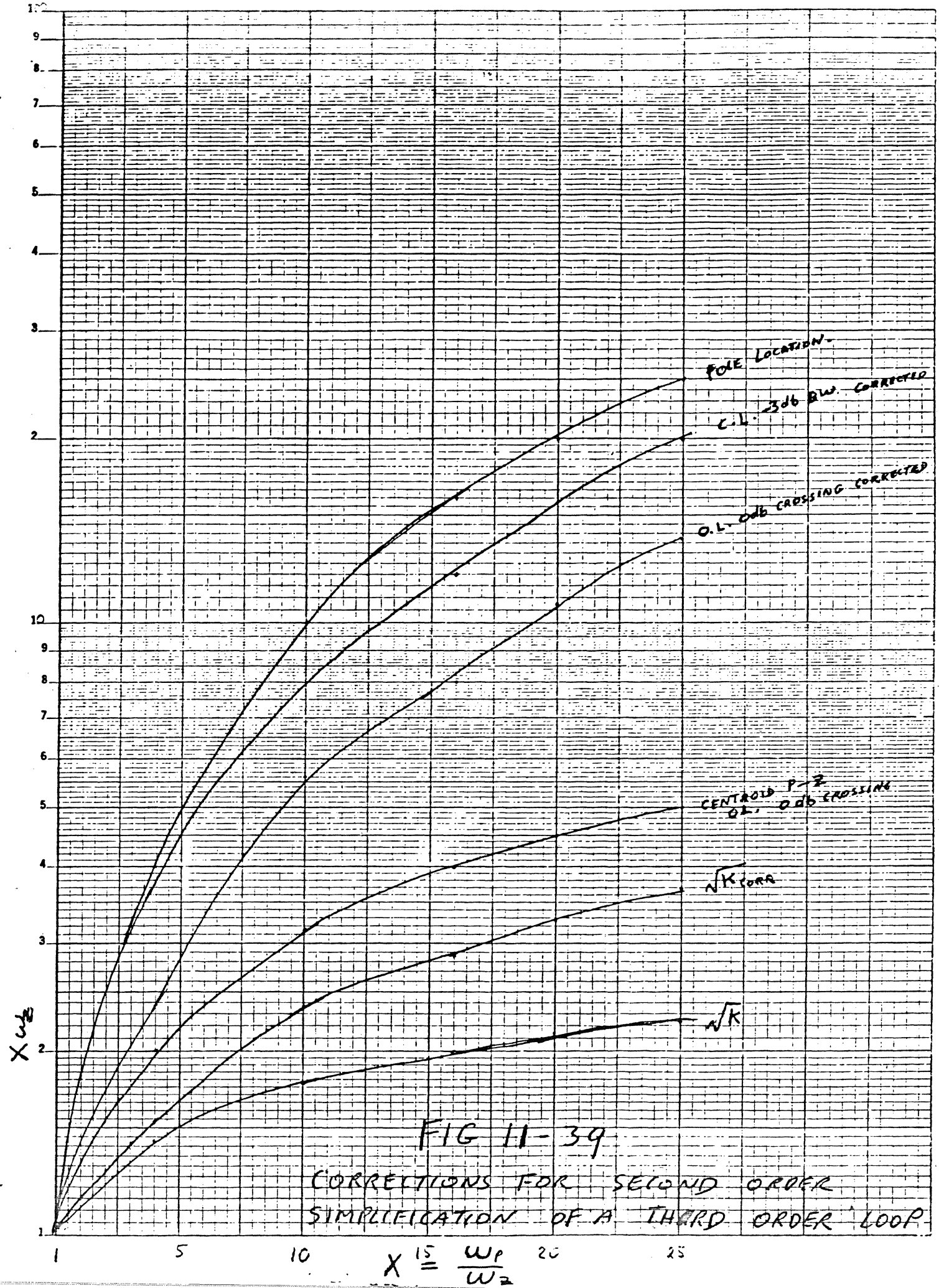


FIG 11-39

CORRECTIONS FOR SECOND ORDER SIMPLIFICATION OF A THIRD ORDER LOOP

As can be seen, the further apart the pole and zero (X →) the more the amount of error decreases which is what we did when we made the W pole 100 times the W zero. The improvement is noted in the right hand column.

To obtain minimum peaking in the frequency domain, either K should be increased or Wz should be lowered from the values that would occur if we ignored the cubic. These are listed in Table 11.2

Table 11.2

X	W_z normal	W min peaking
4	.707	.648
10	.562	.425
16	.500	.340
20	.473	.310
25	.447	.280

The results of these tables are plotted in Fig. 11.39. They show the correction as well as the centroid approach values as a function of X and $X \cdot \omega_z$

As can be seen, the gain required for best operation is higher than one would expect using the graphical design approach. It also makes the Odb crossing further towards the pole which would reduce the phase margin. Using the classical approach, this would spell trouble and would, therefore, be avoided. The real problem is in predicting the behavior of the loop when it is third order or above. It also presents an easy solution to known problems such as can be achieved with second order circuits.

To complete the design the Wo = $\sqrt{K}$ was located at 1.23 x 10^5, therefore, K = 1.513 x 10^{10} for the first solution, Fig. 11.34B, then when we modify it for $\frac{wp}{wz}$ = X = 16, we get the new value of K from Fig. 11.39.

$$\sqrt{K}_{corrected} = 2.86\ \omega_{a_{16}} = (2.86)(6.25\times10^4) = 1.78\times10^5 \qquad \text{(EQ 11.110)}$$

$$\text{therefore } K = (1.78\times10^5)^2 = 3.195\times10^{10} \qquad \text{(EQ 11.111)}$$

and from equation 11.80 we get

$$3.195\times10^{10} = K_s K_\phi K_F K_a \frac{K_o}{S} \qquad \text{(EQ 11.112)}$$

$$= \frac{1}{2}\left(\frac{I}{\pi}\right)\left(\frac{RC_2S+1}{S(C_1+C_2)\left(\frac{RC_1C_2}{C_1+C_2}S+1\right)}\right)\left(K_a\left(\frac{2.7\times10^7 \text{ rad}}{S \text{ volt sec}}\right)\right) \qquad \text{(EQ 11.113)}$$

$$= \frac{1}{2}\left(\frac{I}{\pi}\right)\left(\frac{1}{6.66\times10^{-13}+10^{-8}}\right)(K_2)(2.7\times10^{7}) \qquad \text{(EQ 11.114)}$$

$$\therefore\ K_2 I = \frac{3.195\times10^{10}}{\left(\frac{1}{2}\right)\left(\frac{1}{\pi}\right)\left(\frac{1}{6.66\times10^{-13}+10^{-8}}\right)(2.7\times10^{7})} = 7.93\times10^{-5} \qquad \text{(EQ 11.115)}$$

$$= 79.3\ \mu a$$

For a reasonable current, we could use an attenuator of say 10:1 so the current can be raised to 793 μa for the higher gain. This we should do anyway because when we lower the zero and hence the gain, we will require only 7.93 μa (from $\frac{Wz}{10}$ changes $\sqrt{K}$ by $\sqrt{10^2}$). We do need to concern ourselves with the bandwidth of the current switches at these current levels. The last item is the voltage swing on the capacitor for the frequency lock range requirements.

$$10\left(\Delta V_{cw}\right) = 10\left(0.95v - 0.45\ v\right) = 5.0v \sim \pm 2.5\ V \qquad \text{(EQ 11.116)}$$

This we cannot do with the ECL interface from the phase detectors, Fig. 11.40A, therefore, further voltage translation is required before applying the error pulses to the current switches. We require the modification on both phase locked loops as Fig. 11.40A only allows ± 0.5 v error range and the first circuit required ± 1.25 v, EQ 11.68, and the second required ± 2.5 v, EQ 11.116. Let us address the problem of the low current. If we look at the terms for gain, we see $\frac{1}{C_1+C_2}$. If we raise C_2 to 10^{-7} farads instead of 10^{-8}, we can raise the current. We would also need to change the value of R_f to relocate the zero back to 6.25×10^4 radians, and the pole by changing C_1 accordingly to .0066 μF.

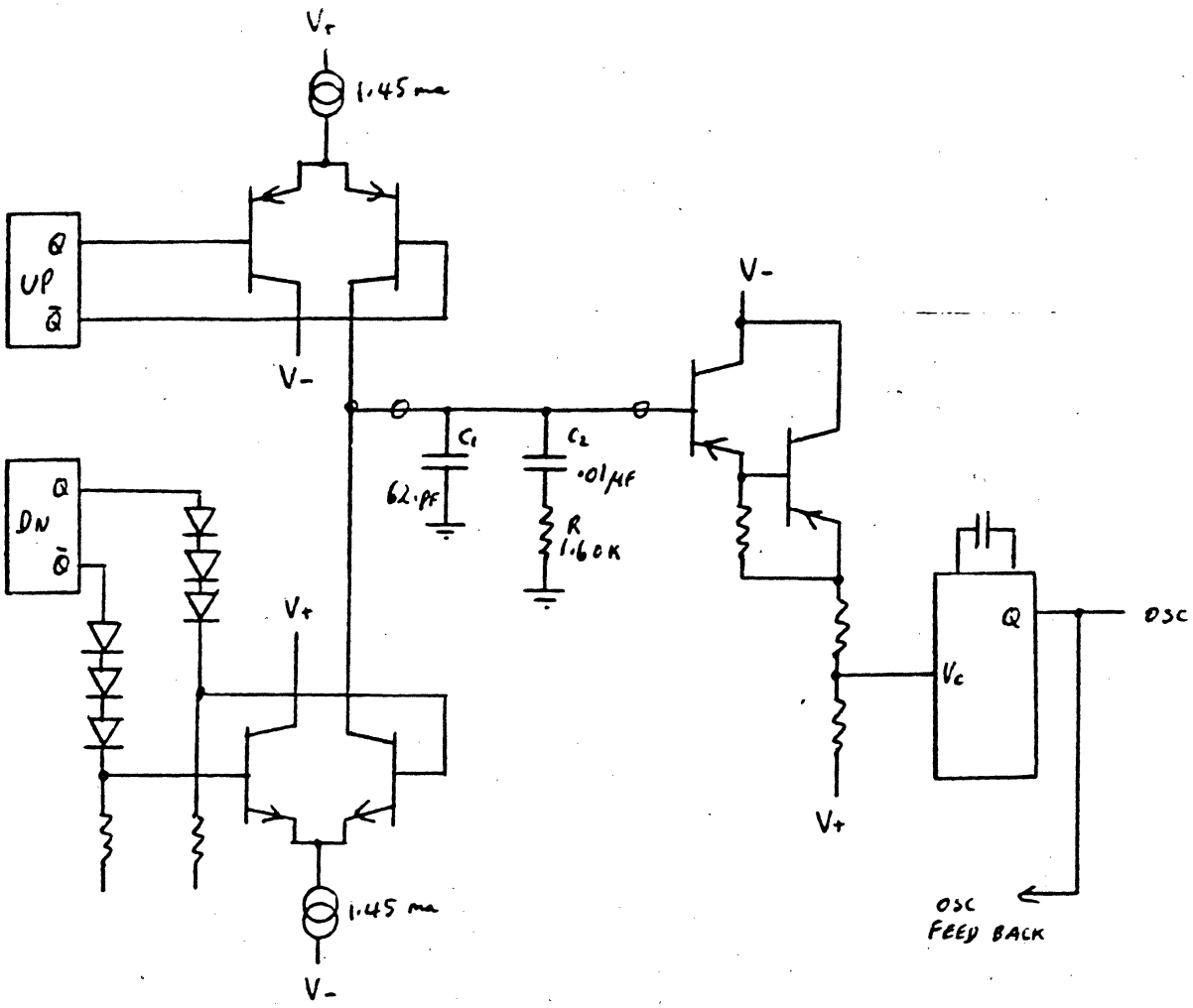


FIG 11.40 A

$\frac{\omega_p}{\omega_z} = 160$ VERSION DESCRIBED IN EQ 11.79 — 11.97

ECL VERSION ($V_+ = GND$)

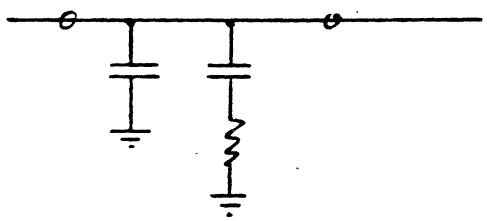


FIG 11.40B

$\frac{\omega_p}{\omega_z} = 16$ VERSION FILTER ONLY

EQ 11.98 – 102

EQ 11.110 – 118

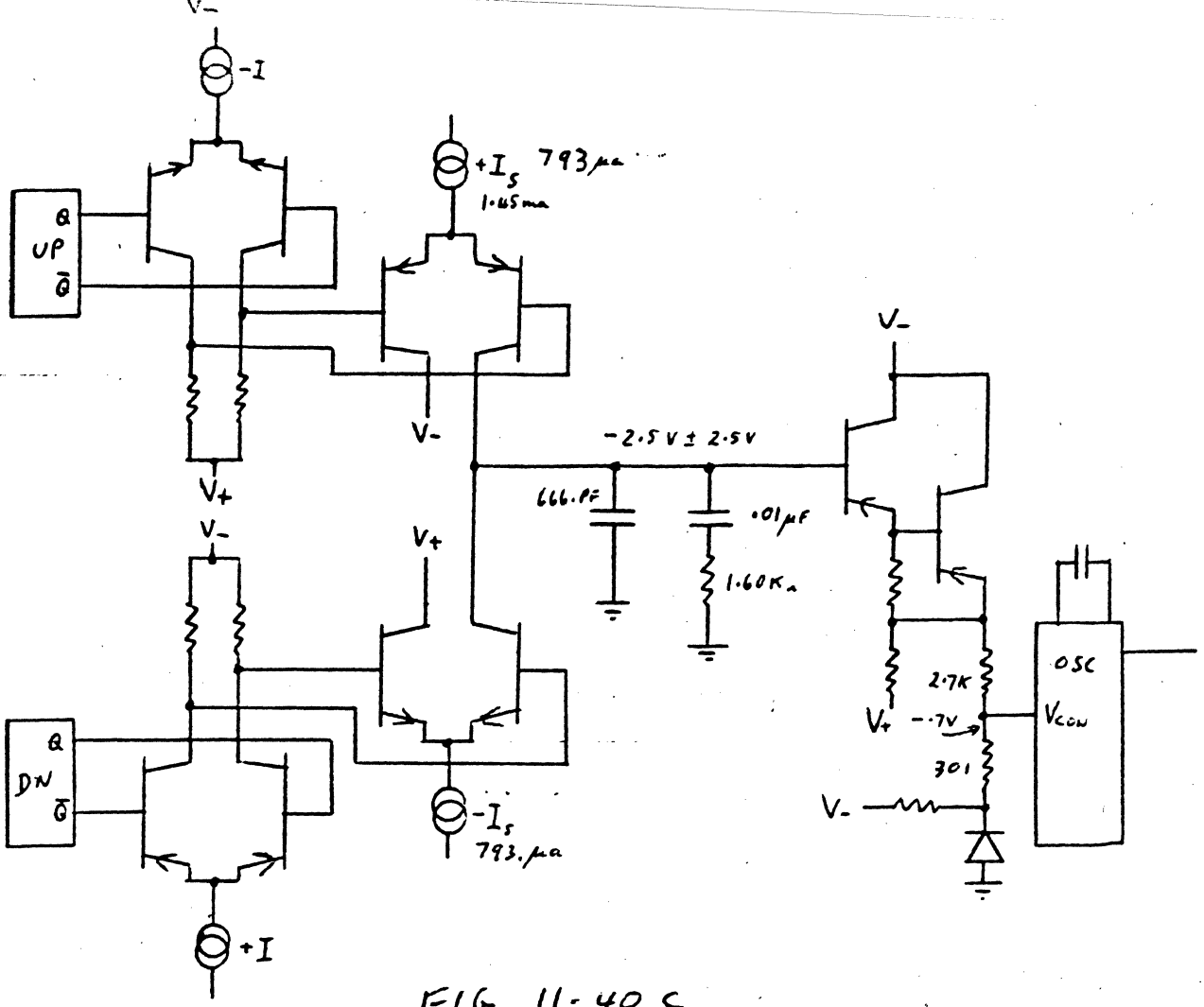


FIG 11-40 C

Q DETECTOR INTERFACE MODIFICATION REQUIRED BY DYNAMIC RANGE.

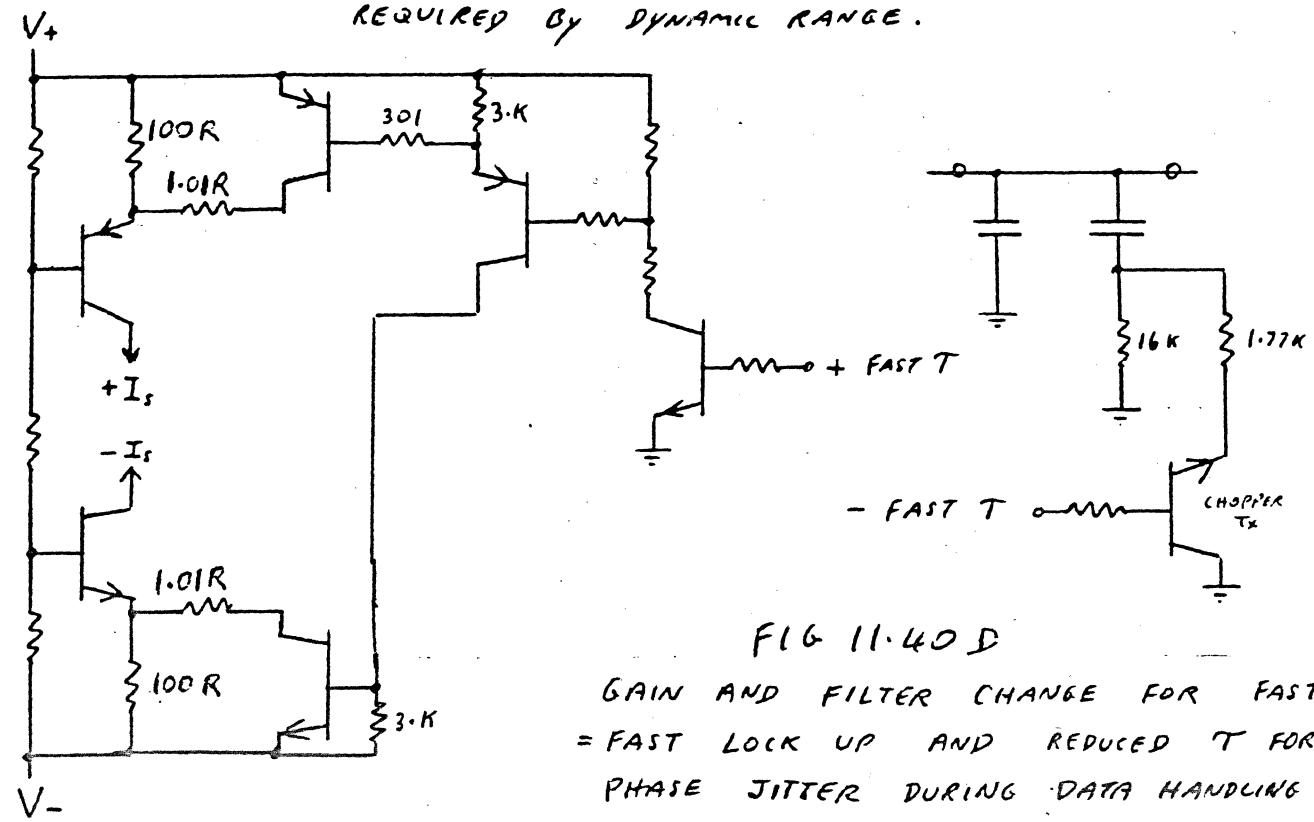


FIG 11.40 D

GAIN AND FILTER CHANGE FOR FAST T = FAST LOCK UP AND REDUCED T FOR MINIMUM PHASE JITTER DURING DATA HANDLING

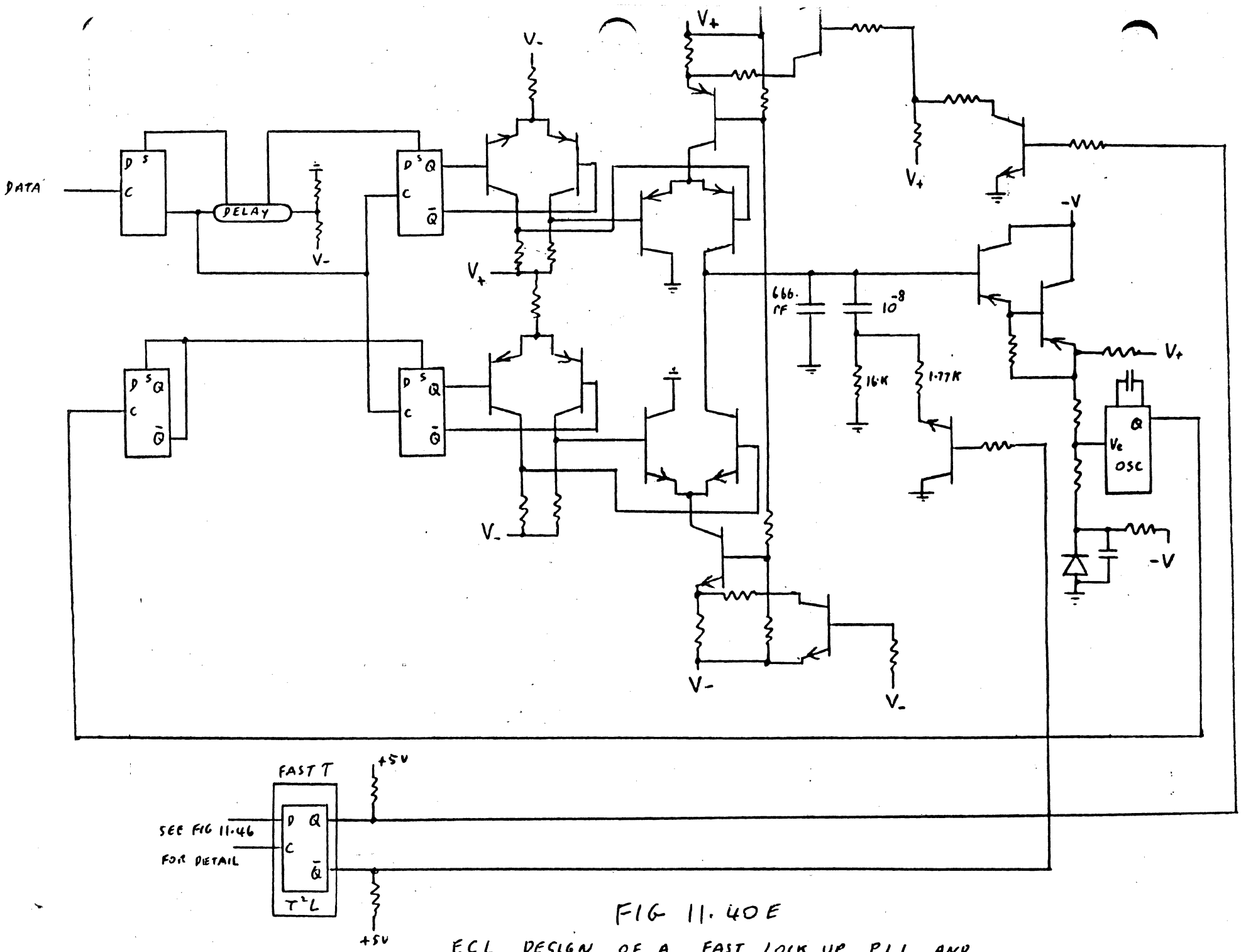


FIG 11.40E

ECL DESIGN OF A FAST LOCK UP P.L.L. AND LOW JITTER · DATA HANDLING

$$R = \frac{1}{(6.25\times10^{-4})(10^{-7})} = \frac{1}{\omega_2 C_2} = 160.\Omega \qquad \text{(EQ 11.117)}$$

The gain $K_\alpha I$ becomes

$$\frac{3.195\times10^{10}}{\left(\frac{1}{2}\right)\left(\frac{1}{\pi}\right)\left(\frac{1}{.66\times10^{-4}+10^{-7}}\right)(2.7\times10^{7})} = 7.93\times10^{-4} \qquad \text{(EQ 11.118)}$$

Now if K_α were $\frac{1}{10}$ as before, the currents are 7.93 ma for the high gain, fast lock up case, and 79.3 ua for the low gain, loop which is easier to handle. We have now gone thru a series of compromizes in order to come up with a viable design for a phase locked loop. With each compromize, we pointed out the difficulty and a possible solution. The final design is shown in Fig. 11.40 E and were numbered this way in order to emphasize the development of the design.

We have now discussed three type 2 phase locked loops. The first using the filter of Fig. 11.31A, EQ 11.79 as shown in Fig. 11.32. The second phase locked loop using the filter of Fig. 11.31D, EQ 11.42 as detailed in Fig. 11.40A and C. We then discussed the third order effects and their adjustments if we built a phase locked loop using the filter of Fig. 11.31B, EQ 11.39 as developed in Fig. 11.40B, C, and D. We might profitably discuss a type 1 loop although its use is limited due to its phase error due to the difference between the free running frequency and the input frequency. This error is easily visualized when we think that with a type 1 there is no intergration of the error. The error that is stored on the filter capacitors leaks away, therefore, it must be constantly replenished which requires a constant phase error to maintain the oscillator on frequency.

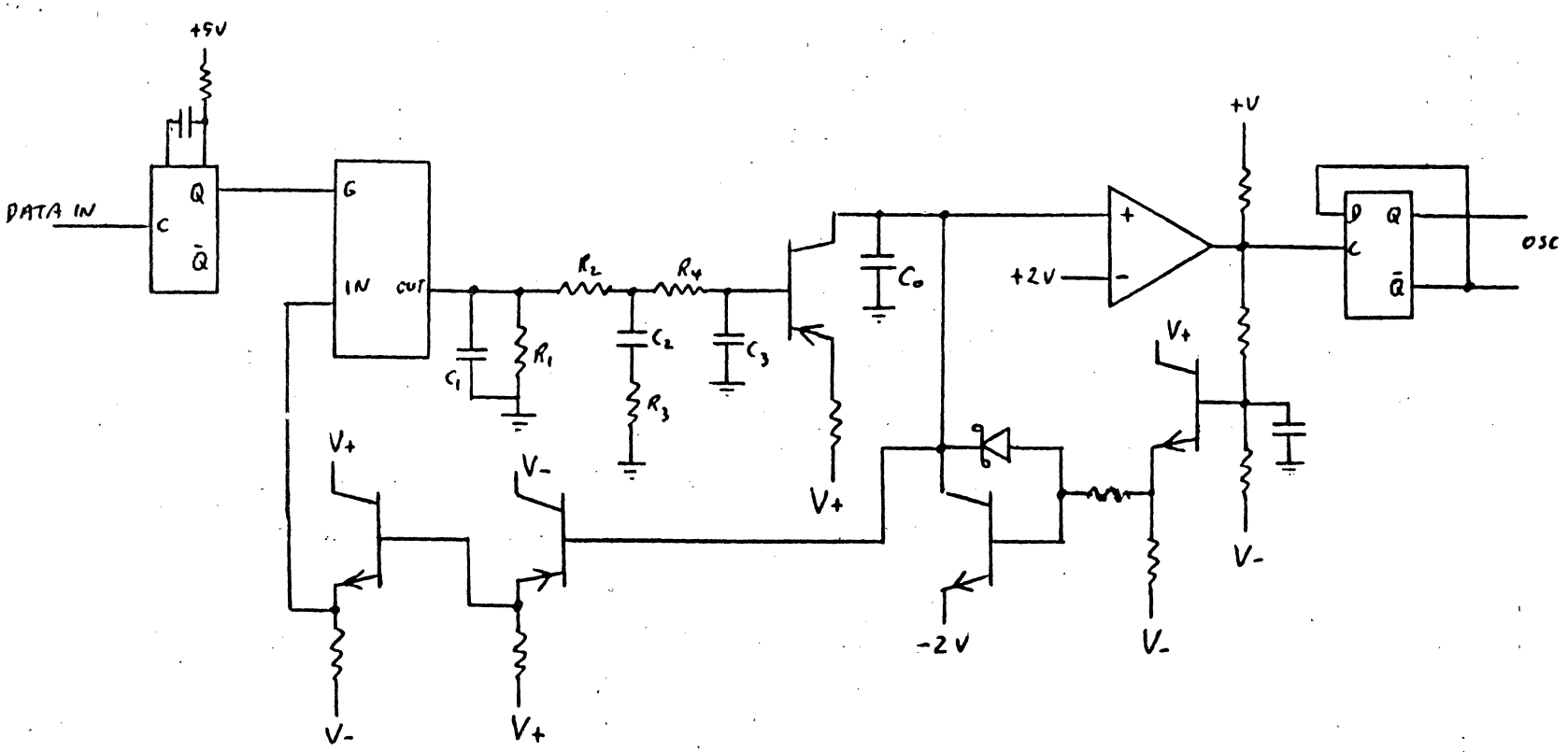


FIG 11.41

TYPE I PHASE LOCKED LOOP USING A RAMP OSCILATOR AND A SAMPLE AND HOLD FILTER

The gain of the phase detector is $\frac{\Delta Vosc}{(2)\ \pi}$ volts/rad. The oscillator output is a sawtooth wave from -2.0 v to +2.0 v as is seen by the reference clamp voltage at the emmitter of Q2 and the +2 v reference on the comparator. The gain is

$$W_{osc} = (F_{osc})(2\pi) = \frac{2\pi}{T_{osc}} = \frac{2\pi\ I_{source}}{(4.0V)\ C_o} \qquad \text{(EQ 11.119)}$$

The current I_{source} is

$$I_{source} = \frac{V_{err} - V_{bc} + V_+}{R_s} \qquad \text{(EQ 11.120)}$$

Therefore, the gain of the oscillator is radians per volt second is

$$W_{osc} = \frac{(V_{err} - V_{bc} + V_+)\ 2\pi}{(R_s)(4.0V)(C_o)} \qquad \text{(EQ 11.121)}$$

The filter is in two parts. The first is the sample and hold network, R_1 in parallel with C_1, and the second is R_2 in series with C_2. Assuming the design of the analog switch and the input pulse width permits full charge or discharge of the hold capacitor, we can write the equation 11.122 if Ks= sample rate gain, Ksh is the sample and hold filter gain.

$$G_{(S)} = K_s\ K_\phi\ K_{SH}\ K_F\ \frac{K_o}{S} \qquad \text{(EQ 11.122)}$$

Therefore,

$$\frac{\Phi_o}{\Phi_{IN}} = \frac{G_{(S)}}{1 + G_{(S)}} = \frac{K_s\ K_\phi\ K_{SH}\ K_F\ \frac{K_o}{S}}{1 + K_s\ K_\phi\ K_{SH}\ K_F\ \frac{K_o}{S}} \qquad \text{(EQ 11.123)}$$

A more complete discussion of the phase locked loop can be found in the Bell System Journel, vol 41, March 1962, pp 559-633.

As we look at the basic system block diagram, Fig. 5.1, we see two phase locked loops. One called the "VFO" and the second the "PLO." These are just names that have come into use to designate the function and are not really discriptive. The "VFO" meaning variable frequency oscillator is intended to designate the phase locked loop used to clock the read data from the recording channel into the host system. The "PLO," phase locked oscillator, is designated as the phase locked loop used for general timing and is locked to the servo data recorded on the servo disc. Since both these data, readwrite and servo, contain harmonic information, meaning the data contains missing bits, both circuits require the use of harmonic phase detectors. The only time the non-harmonic type are used is if there is required some other frequency multiplication that cannot be handled with the original two loops that use the output of either the "VFO" or "PLO" as its input. The "PLO" and "VFO" loop bandwidths are necessarily different. The read data handling "VFO" having the wider bandwidth of the two.

We now need to discuss circuits that utilize the "VFO" as the clocking oscillator. These circuits must assign a read pulse into its proper valued time slot and block any extra pulse that the code may generate that are for self clocking purposes only. There are two kinds. The first are for codes with encoded clocks, and the second is for codes that assign each transition a value.

Declocking circuits with RZ to NRZ convertors (Data Separators). All the FM codes insert clock bits into the data stream to make the data self clocking meaning that a data clock can be easily regenerated from the mixed data itself.

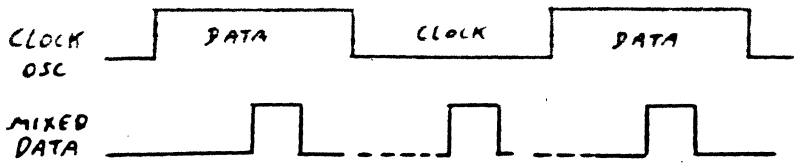


FIG 11.42 A

DATA DECLOCKING TIMING

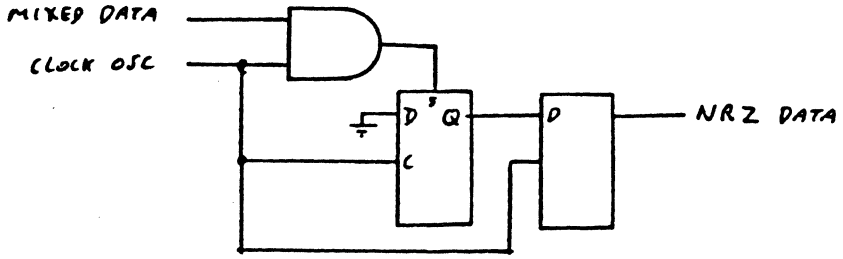


FIG 11.42 B

DATA DECLOCKING PRINCIPLES

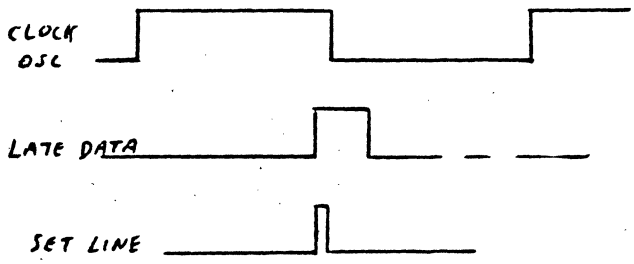


FIG 11.43

PULSE GLITCHING OF LATE DATA

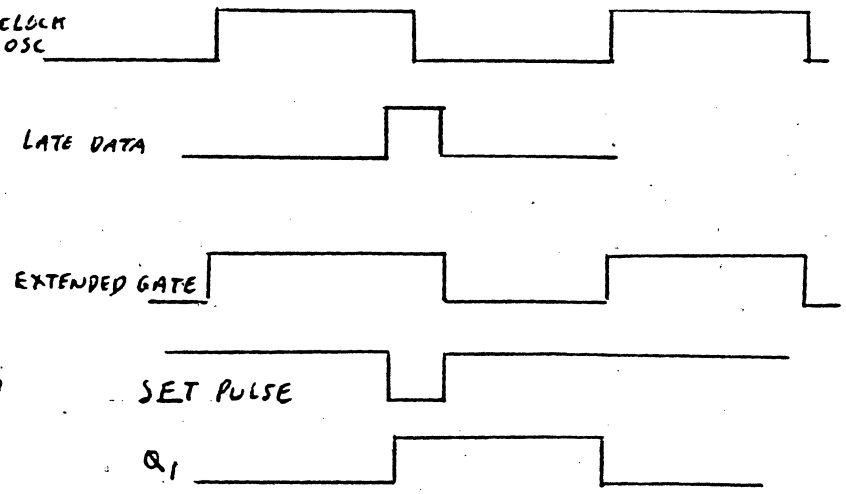


FIG 11.44 B

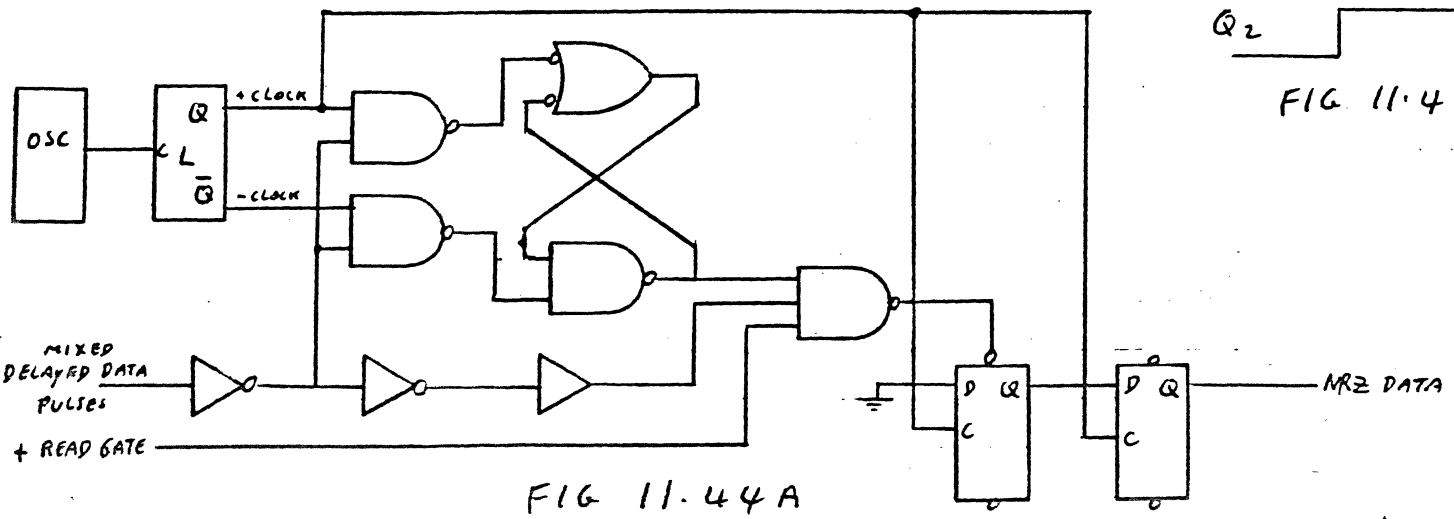


FIG 11.44 A

DATA WINDOW EXTENDER AND NRZ CONVERTOR (T²L VERSION)

In Fig. 11.42A we show a typical sequence. At "A" a data bit can be gated out to the following registers while the clock bit B is blocked from transmission. This is simple enough except as we consider that noise and pulse interference i the analog data shift the pulses from their desired central position. Now these pulses have been deliberately moved to have the leading edge of the pulse centered in the positive "window" by the use of a delay line. This is necessary due to the P.L.L. locking the leading edge of data to the leading edge of the oscillator output (in phase). The problem is when these pulses are shifted due to interference, such that the leading edge is within the window but the trailing edge overlaps into the "clock" window. When this happens, the pulse is split and it is possible that it is split such that the output is insufficiently wide to set the following shift register. This is shown in Fig. 11.43. Further if it is wide enough, the propogation time thru the the first stage may be such that the set up time of the second stage may be insufficient for shifting correctly. These problems may be solved by adding a circuit called a "window extender." The principle is to prevent the fall of the clock until the data pulse has cleared the "window." A circuit for doing this is shown in Fig. 11.44. The circuit works by blocking the change of the FF "L" as long as the data, now inverted by K, is present on both nand gates A and B. When the data goes away, the appropriate nand is conditioned thus setting or resetting the R.S.F.F., C-D. Because the positive "window" has to pass thru both A and C before it is gated by G, the inverted data from K is reinverted by E and further delayed by F such that the total delay is equal in both paths. The pulse from G sets the front stage of the shift register H. The clock for H becomes the actual clock meaning that the register is shifted just before the data is accepted thru the window for each cycle. This permits all propogation delays to be over and settled before shifting. It does not matter to the circuit if a clock pulse blocks Nands A

or B as they are rejected by G anyway. This is a very useful circuit and is required for all FM codes unless logic is devised to overcome the original limitations of both pulse width and propogation delays. The shift regisiter itself is used to convert from RZ to NRZ which is used to transmit data between units of a system.

RZ to NRZ Convertors for Valued Pulses

The design problem with this type circuit is that there is no interlea ed "clock pulse" cycle to use for housekeepping activities such as we used for extending the data window. Obviously the circuits that must be devised for this application must be edge sensitive only, as each cycle is a complete window in itself. The problem is at both extremes of the window unless the designer disallows a small area at each extreme of the window. If this is done, then a pulse located near the edge will be missed entirely and not just and mispositioned. Careful attention to logic delays and extra circuits for parallel use are required to ensure no missing data. Fig. 11.45A, B, and C show both the waveform application and a circuit. The data input is slimmed into a very narrow pulse by A which sets B. An overlap of the set pulse and the possible clock edge is possible therefore, that data bit will be maintained in B causing an error if the next pulse is missing. The FF C is only used for delay to account for propogation thru B and the set up of block D. Again, if the delay from C is not long enough, then the late bit into B will not propogate on the cycle but will possibly show up in the next position due to the overlap of the set and clock lines of B. Block E is now added to allow for the completed set pulse into B before B is clocked. This is about the best that can be done using current logic for this application. What would be great is to have a block that has two independent clocks, one for data and the other for clock.

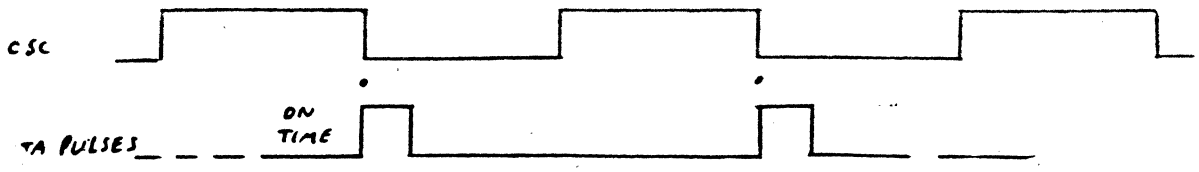


FIG 11.45 A

VALUED DATA PULSES AND OSCILATOR TIMING

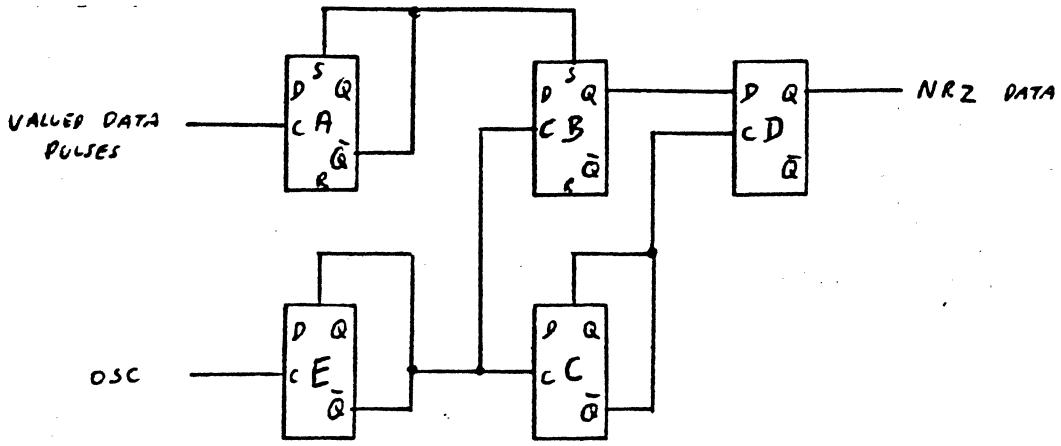


FIG 11.45 B

VALUED DATA PULSE RZ TO NRZ CONVERTOR IMPLEMENTED IN ECL.

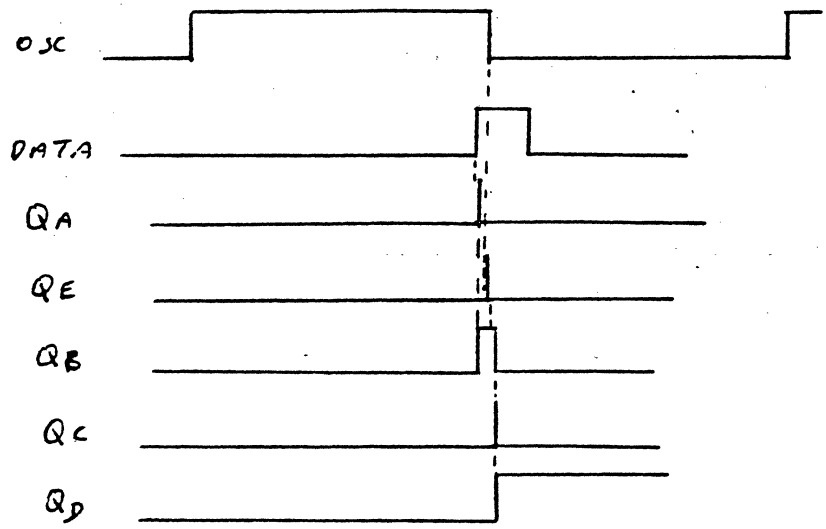


FIG 11.45 C

TIMING SHOWING DELAYS

Starting Circuits

Before we leave the subject of Phase Locked Loops, we should discuss starting circuits. These circuits are used to stop the oscillator briefly and restart it in phase with the incomming data. This is useful if the oscillator is already running near or at the incomming frequency which minimizes the phase correction required. If the incomming data pulses were inputted at random phases to the oscillator, then there is a distinct possibility that the oscillator will slip phase, meaning the phase error exceeds 180^0 therefore it will lock up on the next cycle. For mixed clock and data systems, this will cause clock pulses to be missinterpreted as data pulses. Also, it takes longer for the loop to stabilize after a 360^0 slip. In disc files it is now customary to lock the "VFO" to the "PLO" during non-Read cycles and then switch to read pulses during Read. The circuits are also complicated by using non harmonic phase detectors when locking to the "PLO," and using a harmonic phase detector when locking to data. The block diagram is shown in Fig. 11.46. Here we have both features of the oscillator clamp controlled by the changing edge of the Read Gate and the input data (either Read Data or "PLO") as well as the 'High Bandwidth' switch used for fast sync up. This latter is usually referred to as 'Fast Tau'. As can be seen, the oscillator is clamped to one half cycle (ϕ) until the counter is satisfied by counting input "data" after the Gate edge. For correct operation, the clamp must be able to charge the capacitor during some minimum interval. For some type oscillators, this minimum interval is one half cycle, therefore, the counter must count to two and hold. For others, a single count and hold will suffice. The Fast τ must also be synchronized with the data due to the gain change. The circuitry for doing this includes the current gain change and the filter change (zero-pole) switches illustrated in Fig. 1140E. The control blocks are simply a regular single shot

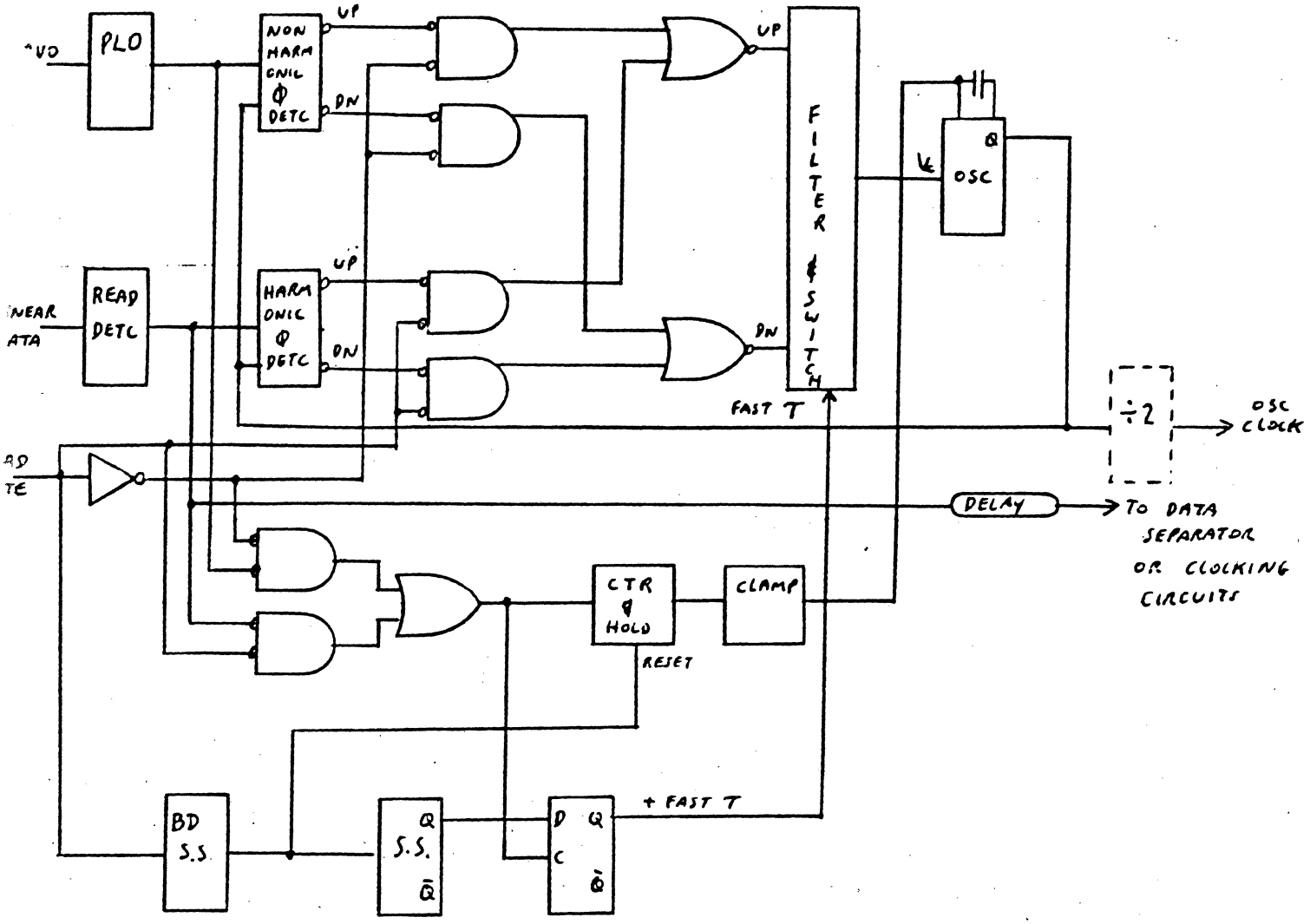


FIG 11.46

IN PHASE START CIRCUITRY AND FAST T SWITCH LOGIC BLOCK DIAGRAM

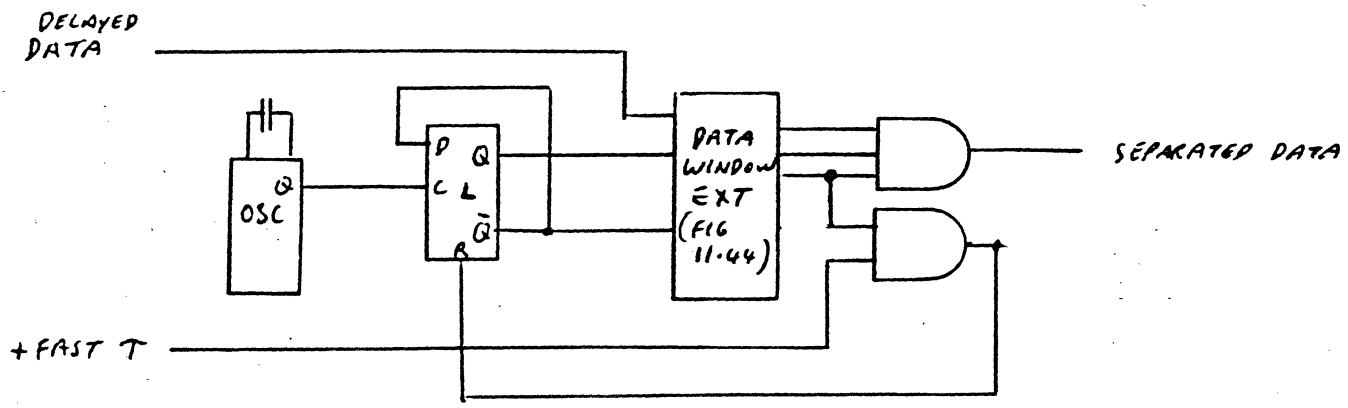


FIG 11.47

OSCILATOR PHASING WITH FORMATTED DATA (FOR FM CODES)

triggered by the change of the Read Gate followed by a 'D' F.F. Thus the + High B.W.' or + 'Fast T' signal will be synchronized with the first data pulse following the Read Gate change and is reset by the first data pulse following the fall of the single shot output.

Data Format Requirements

From the above we can see that it takes time for the various circuits to be ready to properly handle the data. The Write requires establishment of the Write Current only before valid transition can be written on a previously selected head. With Read, the amplifier requires recovery from any select transient, the AGC requires establishment and lastly the "VFO" require synchronization. Thus the preable written prior to each record must include the foregoing, plus some data pattern recognizable as 'the grouping' just prior to the actual record. This grouping can be a single transition phased to occur in the data window or some pattern. With FM codes, some means must be provided to allow differencing between clock pulses and data pulses. This is usually accomplished during the sychronization time of the VFO by making all transitions, clock transitions. Circuitry can be added to ensure that during this period no data is clocked out of the data separator. If it does, then the phase of the oscillator requires reversal. This is easy to do for FM code since the oscillator is required to run at a frequency that includes both data and clock pulse cycles serialy for phase detector use, but at half this rate for separation use; therefore, an intervening F.F. is added to divide by two. If, during the synchronization period, a "data" were clocked out, it would be routed to the reset line of the FF to reverse its phase. Such a curcuit is shown in Fig. 11.47.

CODES

Ever since the first binary recording devices were invented, there have been codes developed to represent the data. The first codes were simply a train of pulses. The early magnetic storage devices recorded these pulses using linear recording, then later with saturation recording of each pulse. As data density increased on subsequent machines, it was realized that there was no need to return the surface magnetization to zero nor was it necessary to erase the old information before making a new record. This permitted the non return to zero, NRZ, codes compared to the earlier RZ or Retrun to zero. NRZ codes had long spaces between transitions due to strings of ones beginning with a transition ~~These codes mainly being from the railroad industry used a pulse for a zero~~ and ending with a transition with nothing in between. ~~and no pulse for a one~~. This was alternatly inverted for each one bit for our industry and became known as NRZI.

All of the data for drum memories, tape, and disc files used NRZI codes for many years. The data was assigned its position value by using clock tracks recorded on the tape or disc as a separate channel. When the tape skew or disc-head arm circumferential vibration increased such that it was no longer possible to correctly clock the data, it became necessary to clock with self clocking codes. In Fig. 12.1 we illustrate the RZ code (~~implemented inverted in order to be consistant with subsequent codes~~). In Fig. 12.2 we show NRZI for the same data pattern. Also included is a typical read linear waveform. A variation of the NRZI is the inclusion of a ninth bit for byte identity which was subsequently used in the 2305 drum file to make the code "self clocking." This is shown in Fig. 12.3.

The self clocking codes were essentially Frequency Modulation codes. More correctly Pulse Position Modulation that used only two positions. The four codes were subdivided into several types. The first, F.M., consisted of always writing a clock transition at the bit cell boundary then if the data

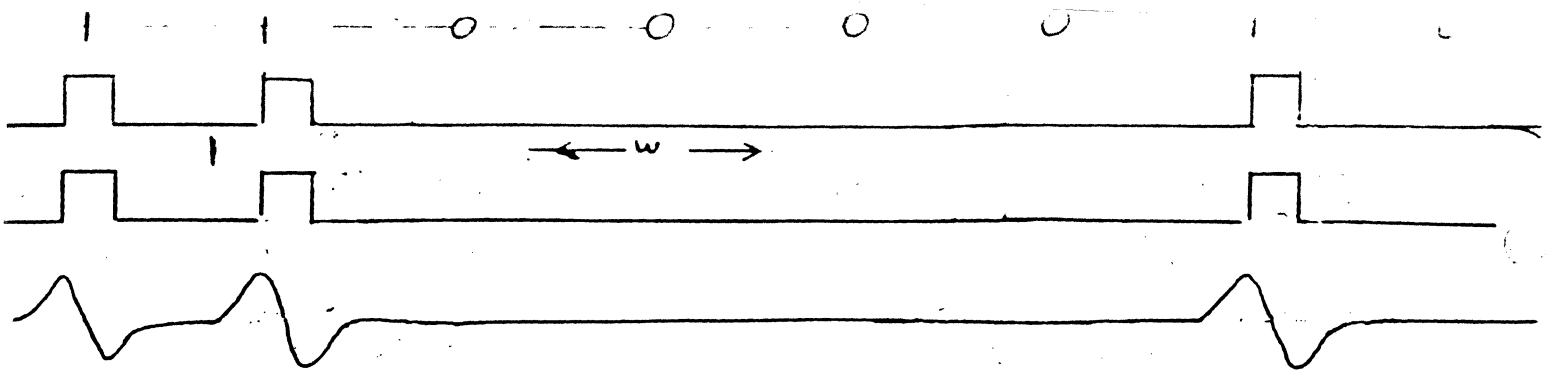


FIG 12·1

RZ CODE, DATA, I_{WRITE}, V_{READ}. $1(0, \infty)$

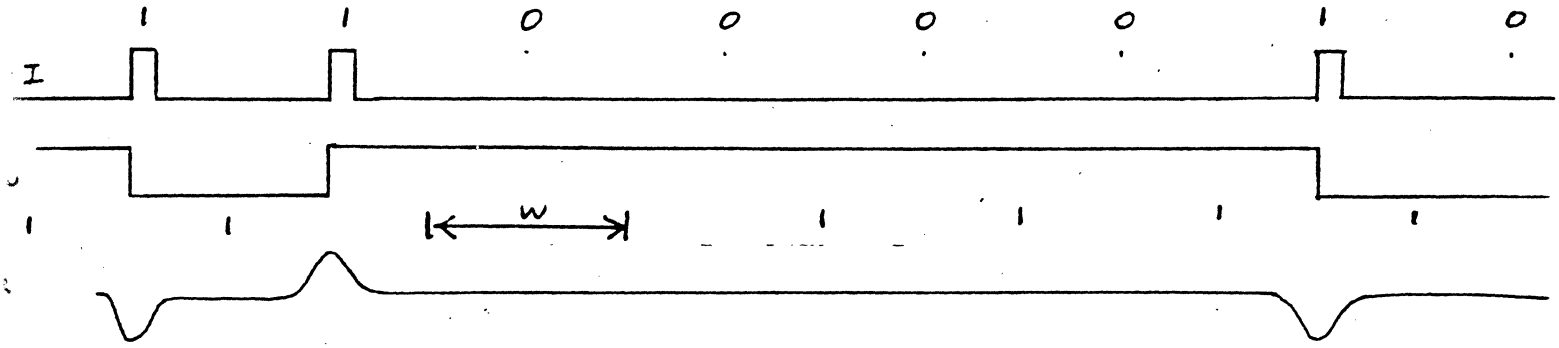


FIG 12·2

NRZI CODE. DATA, I_{WRITE}, V_{READ}. $1(0, \infty)$

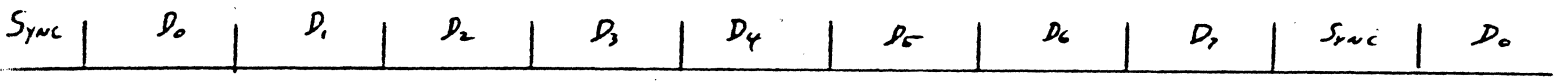


FIG 12·3

MODIFIED NRZI CODE, SHOWING DATA, SYNC CELLS

$\frac{8}{9}(0, 8)$

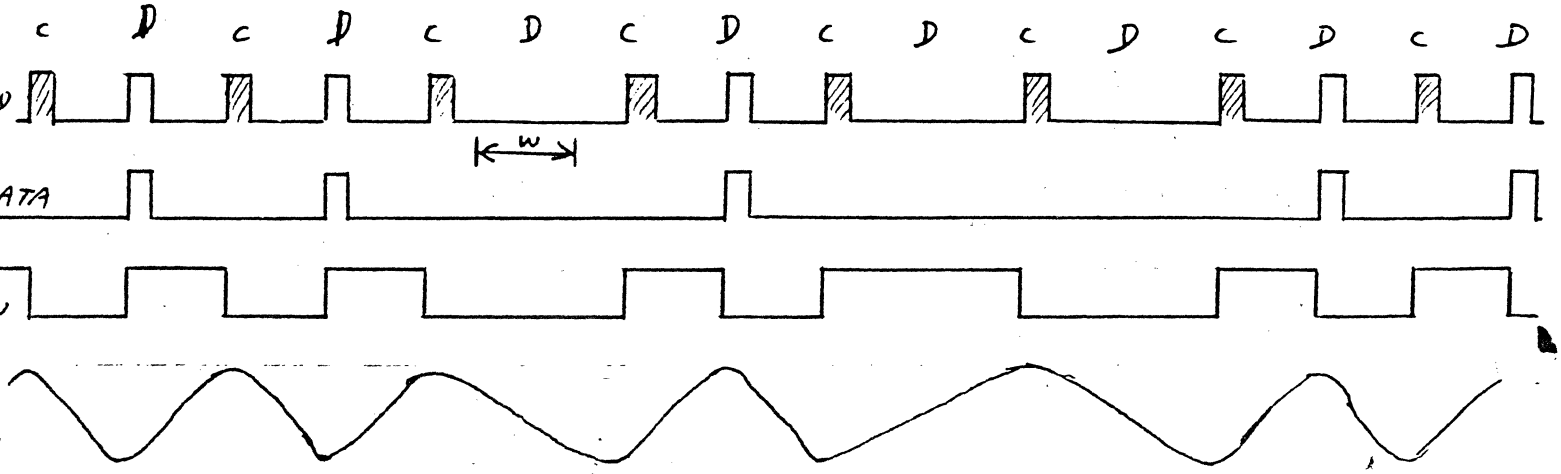


FIG 12·4

FM CODE CODED, DATA, I_{WRITE}, V_{READ} $\frac{1}{2}(0, 1)$

requires it a transition at the cell center for a one bit. Thus, there were twice as many transitions for an all one pattern as there were for NRZI, but it was self clocking. Because of this density increase, the early machine that used this code operated in Region 3 of the B.P.I. curve. Fig. 12.4 shows a typical pattern. In Fig. 12.5, we show the PM or Phase Modulation code. It also required the same density as FM, but it differed in that they recorded the zero bits as well. The direction of the read back TRANSITION OR pulse was always of the same polarity for a one bit and the opposite for a zero bit. In the event two bits of the same value followed together, the code required that a middle transition be recorded to return the flux phase such that the second bit could be recorded in its proper direction. It is easily seen that now we have clock bits added that can have either polarity. These extra transitions occur at the cell boundary and therefore, require the same clock decoding as FM codes and hence, the same recording density increase. It was then determined that there were other self clocking codes that could be developed that only required the standard one for one density. The first of these were the Modified FM codes. These were first invented by Mr. W. Pouliart et al in 1954 and subsequently reinvented in a slightly different form by Mr. A. Miller, Mr. W. Woo, and again by Mr. Jacoby.

The basic code is shown in Fig. 12.6 wherein a set of rules were established for writing. The first rule was: All data bit ones are recorded. The second was that a clock transition will be written at the cell boundary only if there was a data zero in the preceding and following data cell. This latter requirement is mixed up in some subsequent literature. A variation of the MFM code is the M^2FM and there are several of them. One has the following set of rules. Write all data ones. Second--write a clock transition at the cell boundary only if there are two cells containing zeros

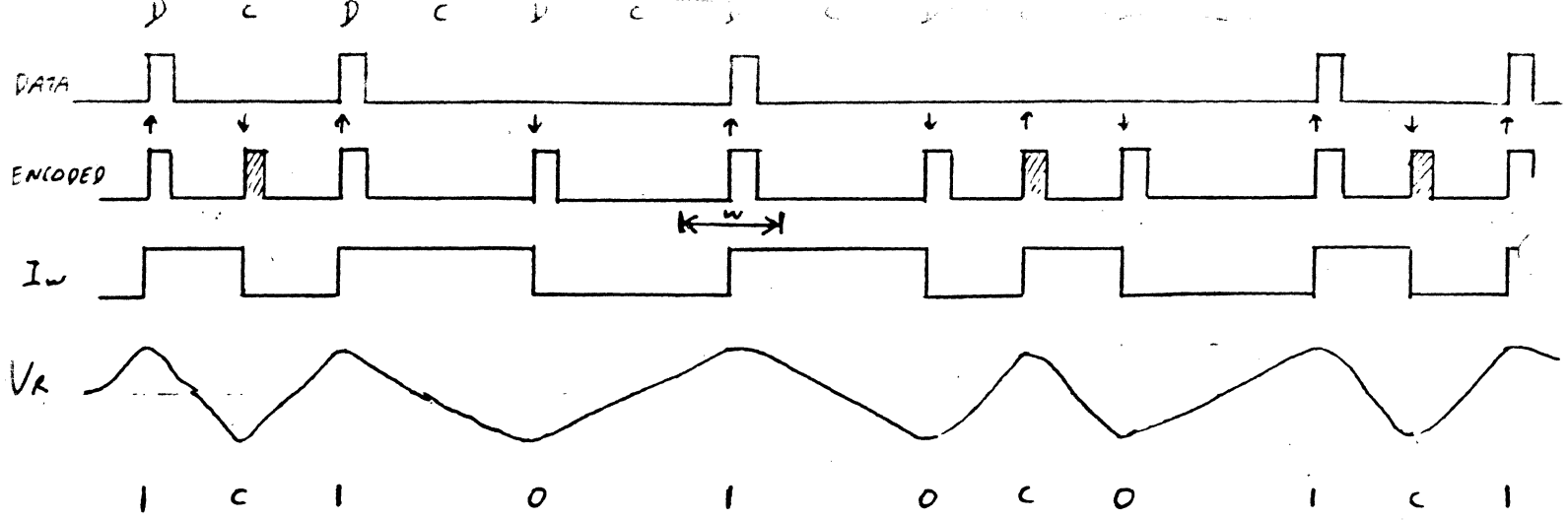


FIG 12.5

PM CODE DATA, ENCODED, IWRITE, VREAD, VALUE, ½ (0, 1)

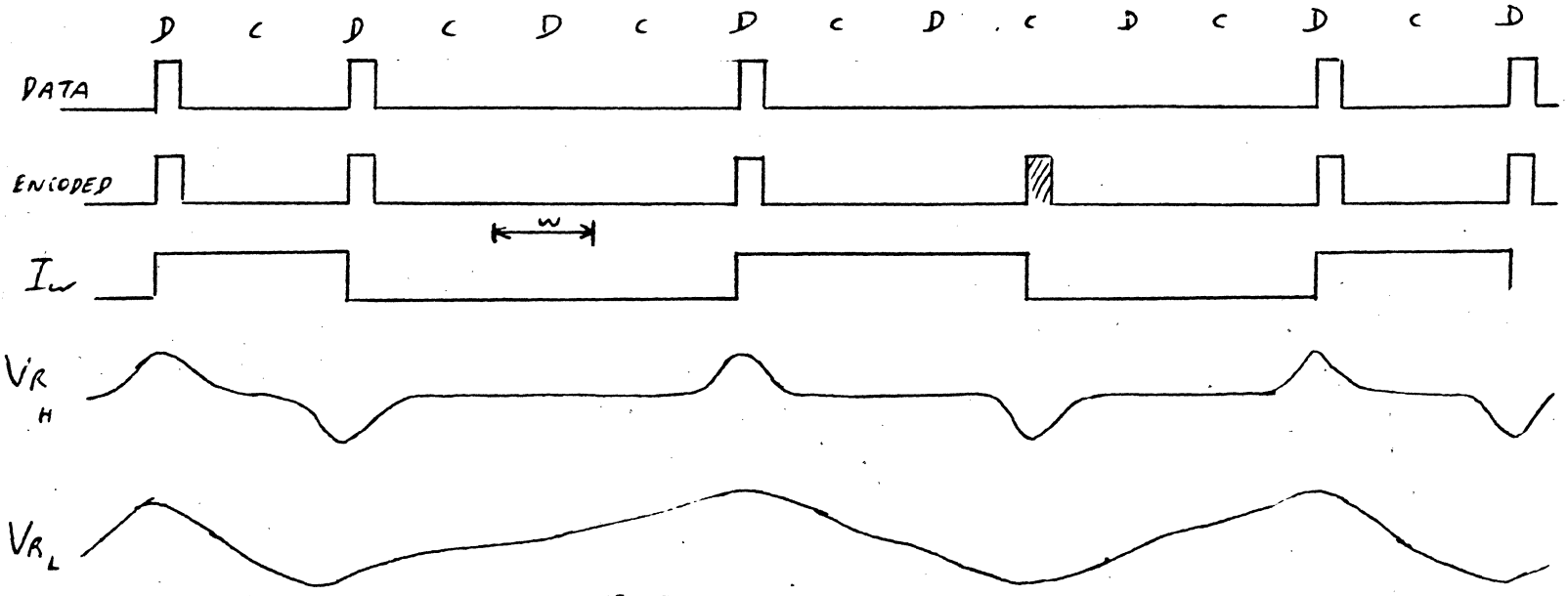


FIG 12.6

MFM CODE DATA, ENCODED, IWRITE, VREAD HIGH RESOLUTION ½ (1, 3) VREAD LOW RESOLUTION

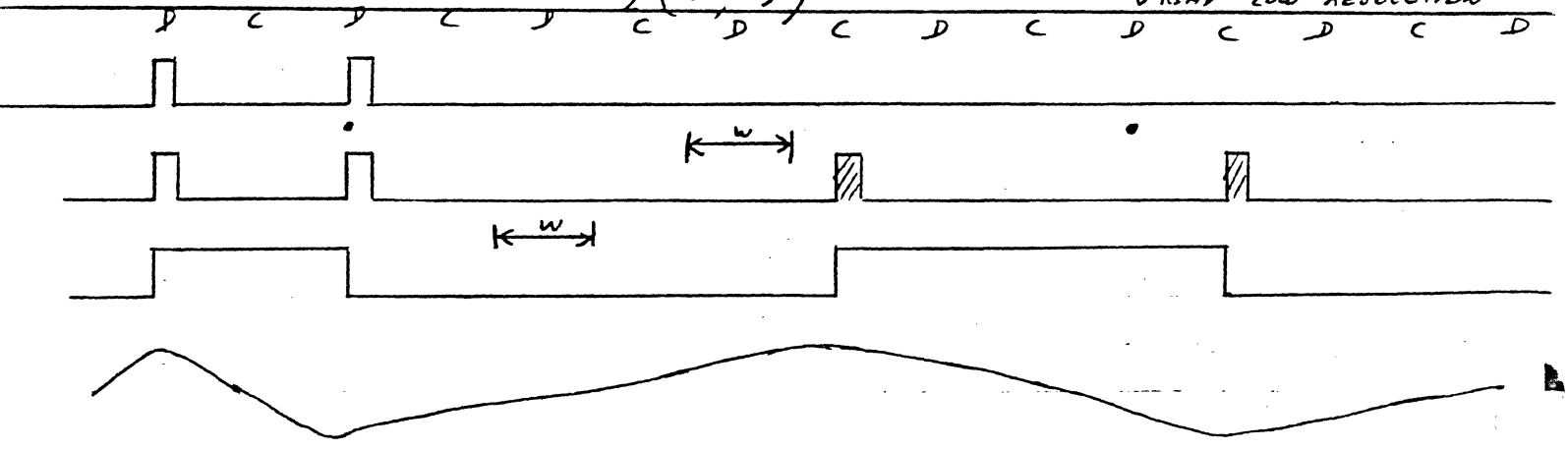


FIG 12.7

M²FM CODE DATA, ENCODED, IWRITE, VREAD ½(1, 7)

Also never write two clocks adjacent to each other.

on either side, see Fig. 12.7. There is a benefit to this that is not as obvious. Back with the FM code it can be seen that for a data stream containing ones and zeros when encoded, the 'one' bits are always bounded by clock bits and will, therefore, have some symmetrical interference which results in small peak position shift. The clock transitions on the other hand have no such symmetry and will therefore exhibit severe bit shift especialy in Region 3. This fact was taken advantage of in early machines by using an unsymmetrical window, 40% for the data ones, which are shifted least, and 60% for clocks, which are shifted most, Fig. 12.8. Now with M^2FM, a similar condition occurs. The data, which is always written, has the highest density, therefore, will be shifted most. The clock transitions, however, are always spaced a mimimum of two cell times away from its neighbor, therefore, suffers the least shift. Therefore, a window of 60% or thereabouts, Fig. 12.9, is assigned to the data ones and a 40% window assigned to the clock. This is a real advantage and similar codes are presently used in the floppy disc market.

The next codes are the group codes or substitution codes so named because a group of input is encoded into a differing group for recording. During Read, the process is reversed. There is a huge variety of these codes. The first assembly of order into the growing literature on these codes came with an article by A. Patel in the IBM Journal, July 1975, page 366, and quoted in Computer Design, August 1976, page 85. Here Mr. Patel introduced a symbology that can be used to describe all codes, although not uniquely. The input data may be introduced as either one bit at a time or may, depending on the code, be in groups. The number of input bits in the group is m, and may range from one to m bits. The second part is the number of cells in which the number of input bits may be assigned values. This number is designated as n, and may range from one to n. These two numbers are expressed

as a fraction, thus, two input bits may be assigned into four cell positions and will, therefore, be designated

$$\frac{m}{n} = \frac{2}{4} = \frac{1}{2} \qquad \text{(EQ 12.1)}$$

The remaining code designations refer to the minimum and maximum run length of zeros. The minimum number of zeros is designated as d, and the maximum number of zeros is k, thus any code can be described

$$\frac{m}{n}\ (d,k) \qquad \text{(EQ 12.2)}$$

To see how we use this designation, let us try it on several of the codes we have previously introduced. Our NRZI code can be written

$$\frac{m}{n}\ (d,k) = 1\ (o,\infty) \qquad \text{(EQ 12.3)}$$

where m=1, n=1 meaning that for every input bit there is a unique cell assigned. d is zero meaning that each cell can have a one bit, and k=∞ indicates that an all zero record can be written without any transitions. The code similar to NRZI where a ninth sync bit is added for every eight bits can be written

$$\frac{m}{n}\ (d,k) = \frac{8}{9}\ (0,8) \qquad \text{(EQ 12.4)}$$

which would be easier to handle thru the amplifiers. The FM code would be

$$\frac{m}{n}\ (d,k) = \frac{1}{2}\ (0,1) \qquad \text{(EQ 12.5)}$$

meaning that there are two cell positions for every input data. Each cell can be filled and only one cell in a row may be left zero.

The PM code is designated the same. For this reason we still need further description to identify any particular code. The MFM code is written

$$\frac{m}{n}\ (d,k) = \frac{1}{2}\ (1,3) \qquad \text{(EQ 12.6)}$$

where any bit is associated with two cell positions, one of which must be left

zero with no more than three cells in a row left zero. The version of M^2FM described earlier can be written

$$\frac{m}{n}(d,k) = \frac{1}{2}(1,7) \qquad \text{EQ 12.7}$$

The $\frac{m}{n}$ and d are the same as in EQ 12.6, but the k=7 comes from the sequence 1c0c0c0c1

There are several advantages for using codes. Some, as was Mr. A. Patel's, designed to minimize the dc content of the code in order to write thru a transformer of a rotating head system. Others are designed to maximize the amount of data stored for a given transition density or to ensure readability such as self clocking codes.

Another distinct difference in codes may be the rule by which they were written although observing the recorded waveform it would be impossible to tell them apart. For example, the FM code can be designated as the following.

Write all input ones at the cell boundary and write all clocks mid cell. The second may be written. Write all input ones at the cell center and write all clocks at the cell boundary. Unless the observer knew the phase of the writing circuits, he could not tell them apart, yet they are distinct and different.

For MFM there are eight separate encoding rules that produce the same recorded result. It is interesting to note that three of them have been patented and a fourth cannot as it is now in the public domain. They are: 1) Write all one bits at the cell center, write a clock bit at the leading boundary of the cell if preceded and followed by a zero. 2) Write all one bits at the cell center, suppress all clock bits at the leading boundary of a cell except those preceded and followed by a zero. 3) Write all one bits at the leading boundary of a cell, write a clock bit at the cell center if preceded and followed by a zero. 4) Write all one bits at the leading

boundary of a cell, suppress all clock bits at the cell center except those preceded and followed by a zero, and so on as the boundary is changed to the trailing boundary instead of the leading boundary.

The group codes differ in that they are substitution codes with sometimes very elaborate rules as to run length. The most familiar of these is the GCR code, or Group Coded Recording, used in the tape industry

$$\frac{m}{n}\,(d,k) = \frac{4}{5}\,(0,2) \qquad \text{EQ 12.8}$$

The code conversion is listed in Table 12.1 below

TABLE 12.1 GCR CODE	Data Value	Recorded
	0000	11001
	0001	11011
	0010	10010
	0011	10011
	0100	11101
	0101	10101
	0110	10110
	0111	10111
	1000	11010
	1001	01001
	1010	01010
	1011	01011
	1100	11110
	1101	01101
	1110	01110
	1111	01111

should introduce a third measure for a code. This is the Density Ratio where

$$Dr = \frac{\text{Data Density}}{\text{Maxtransition Density}} = \frac{T\ min}{T} = \frac{m}{n}(d + 1) \qquad \text{EQ 12.9}$$

r the codes we have introduced, we can tabulate the various parameters for mparison. (see table 12.2)

ABLE 2.2

Code	m	n	d	k	Dr	$\frac{m}{n}$ WINDOW
NRZI	1	1	0	∞	1	1
FM	1	2	0	1	0.5	.5
PM	1	2	0	1	0.5	.5
MFM	1	2	1	3	1	.5
GCR	4	5	0	2	0.8	.8
3PM	3	6	2	11	1.5	.5
2/3(1,7)	2	3	1	7	1.333	.6666
1/2(2,7)	2	4	2	7	1.5	.5

Now one of the purposes of using codes is to increase the information content for the same number of transitions. As can be seen in Table 12.2, Dr, the density ratio is one or less for the first five codes listed. Mr. G. Jacoby published a code in 1977 that allows an increase of 50% or a Dr of 1.5. He called it 3PM. It limited the minimum number of zeros to two in order to reduce the transition density. With this code, he substituted a 6 cell "word" for a three bit data input. Keeping the restriction of a MAXIMUM number of zeros, created a number of inconsistencies that occur as data are preceded or followed by certain patterns that would violate the rule of a two zero MAXIMUM. These he solved with alternate patterns. The listing is given in Table 12.3

as taken from his paper.

TABLE 12.3 3PM CODE

Input	Adjacent Word Influence		p_6'	Output					
	Preceding	Following		P_1	P_2	P_3	P_4	P_5	P_6
000	X	0	0	0	0	0	0	1	0
000	X	ŧ	0	0	0	0	0	0	1
001	X	X	0	0	0	0	1	0	0
010	X	X	0	0	1	0	0	0	0
011	X	0	0	0	1	0	0	1	0
011	X	ŧ	0	0	1	0	0	0	1
100	X	X	0	0	0	1	0	0	0
101	0	X	0	1	0	0	0	0	0
101	ŧ	X	1	0	0	0	0	0	0
110	0	0	0	1	0	0	0	1	0
110	ŧ	0	1	0	0	0	0	1	0
110	0	ŧ	0	1	0	0	0	0	1
110	ŧ	ŧ	1	0	0	0	0	0	1
111	0	X	0	1	0	0	1	0	0
111	ŧ	X	1	0	0	0	1	0	0

where ŧ=influence, 0=no influence, X=don't care, and P_6' is the previous word's P_6 as altered for this word. For a further explanation, see the paper entitled "A New Look Ahead Code for Increased Data Density" GV Jacoby. IEEE Sept Proceedings on Magnetics 1977, vol 13, No 5, p 1202. Another code that is useful is designated as (Newman, Fisher)

$$\frac{m}{n}(d,K) = \frac{2}{3}(1,7) \qquad \text{EQ 12.10}$$

which allows an increase of one-third in density, Dr = 1.3333. Again it is a substitution code with changes depending on adjacent word interference that would alter the minimum run length. The changes are implemented by accepting four input bits at a time instead of the usual two bits and encoding them uniquely

TABLE 12.4

DATA		CODE		
NOW	FUTURE	NOW	FUTURE	
00	00	001	XXX	
00	01	001	XXX	
00	10	001	000	double group
00	11	010	000	double group
01	00	010	XXX	
01	01	010	XXX	
01	10	010	XXX	
01	11	010	XXX	
10	00	100	XXX	
10	01	100	XXX	
10	10	100	XXX	
10	11	100	XXX	
11	00	101	XXX	
11	01	101	XXX	
11	10	000	100	double group
11	11	101	000	double group

2/3 (1,7) code (X = don't care)

The first IBM disc drive to use a group code was the 3370. This code is designated as

$$\frac{m}{n}\,(d,k) = \frac{1}{2}\,(2,7) \qquad \text{EQ 12.11}$$

The code compression Dr = $\frac{3}{2}$ = 1.5. To utilize the code, input data may be accepted 2, 3, or 4 bits at a time depending on the content. All possible combinations can be made up from those listed in Table 12.5. This code is attributed to Mr. Franazek of IBM.

TABLE 12.5

DATA	CODE AB	AB	AB	AB
10	01	00		
010	10	01	00	
0010	00	10	01	00
11	10	00		
011	00	10	00	
0011	00	00	10	00
000	00	01	00	

1/2 (2,7) code

As there are so many codes possible of both the block and the merging types, we will not cover the remainder, but will simply give some equations that when solved will describe some of them. These equations were described by Dr. T. Campbell. The number of code words, Cw (n,d), is given by

$$Cw\,(n,d) = \sum_{i=1}^{R_{max}} \frac{(n-di)!}{i!\,(n-(d+1)i)!} \qquad \text{EQ 12.12}$$

where $R_{max} \leq \frac{n}{d+1}$

also $m \leq \log_2 \{Cw(n,d)$ EQ 12.13

for conventional Block codes. When merging is allowed, as with the 3PM codes, the following is given for cases where $d \geq 2$

$$Cw'(n,d) = \sum_{j=0}^{R'} \frac{(n-d-dj)!}{j!\,(n-d-(d+1)j)!} \qquad \text{EQ 12.14}$$

where $R' \leq \frac{n-d}{d+1}$

PRECOMPENSATION

We first introduced this subject block in Chapter 5. Basicly it consists of deliberatly writing a transition such that the resultant movement of the peak due to superposition or pulse interference will move the peak back to its 'on time' position. Thus, if pulse interference makes a pulse peak late that transition is deliberatly written early such that the resultant peak shift places the peak at its true unshifted position. It has been shown that the worst peak shift occurs for two adjacent transitions with no transitions on either side. If we wrote two transitions T seconds apart and we measured the time between the two resultant peaks as 1·2 T, then the peak shift of each transition is 0·1T. In order to correctly write the two transitions, the first has to be written late by greater than 0·1T and the second has to be written early by greater than 0·1T. This is because when the early and late transitions are written at .8333T, the pulses are closer together and their pulse interference produces a peak timing of greater than T. The process is an itterative one and is best calculated using about 10% greater shift than predicted for the plus and minus Precompensation, then recalculate the predicted peak seperation. We left this subject to this chapter because we now have an understanding of the clocking effects of the window allowed and the need to minimize the movement of a pulse so as to keep it in its assigned window. Further, the implementation

FIG 12.8 F.M. WINDOW UNSYMETRY

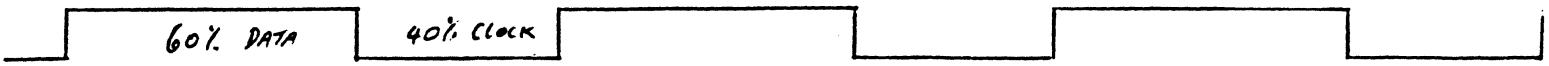


FIG 12.9 M²FM WINDOW UNSYMETRY

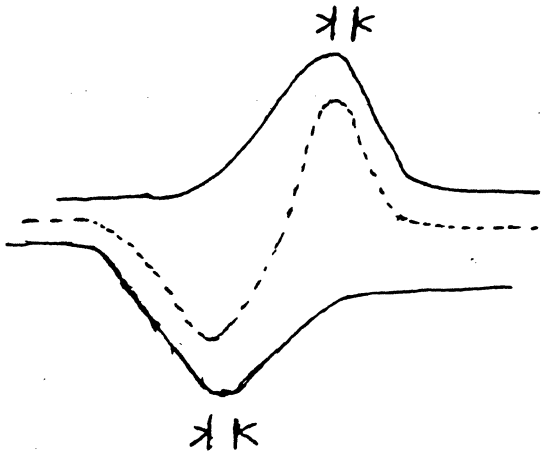


FIG 12.10

PEAK SHIFT DUE TO PULSE INTERACTION

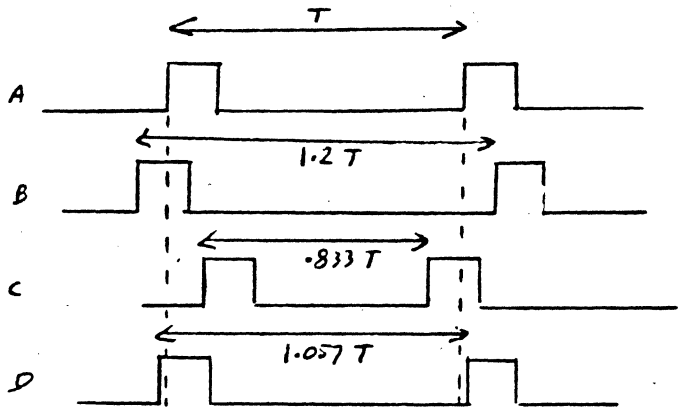


FIG 12.11

A = WRITE ON TIME
B = READ BACK SHIFTED
C = WRITE PRE COMPENSATED
D = READ BACK NEARLY ON TIME

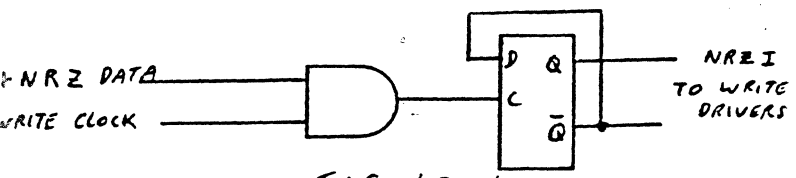


FIG 12.12

NRZI ENCODER

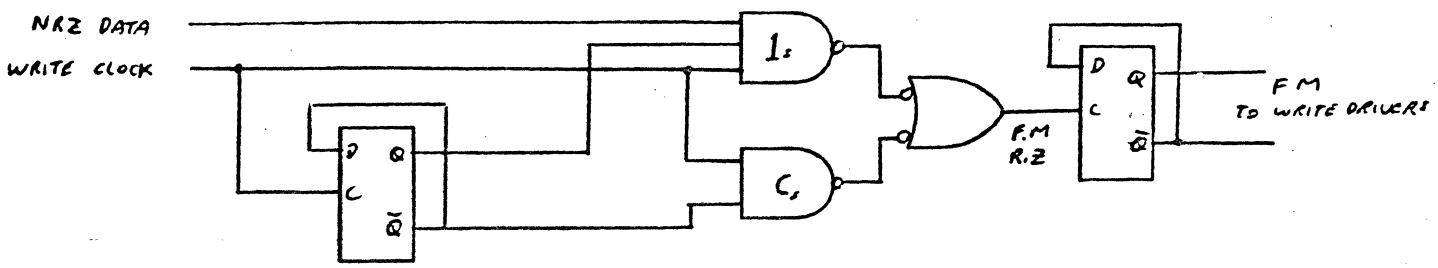


FIG 12.13

F.M. ENCODER

is best incorporated in the encoder circuitry . See Fig. 12.10 and 12.11. Sometimes a particular code cannot be optimized for peak shift with only a single value of plus and minus Precompensation. Certain patterns may produce a lessor amount of peak shift that would be overcompensated if the single value were used. It is then therefore necessary to install a multi level Precompensation depending on the signal degradation and the degree of window margin allowed. This is calculated the same way as before using the new peak se paration values.

CIRCUITS

We might profitably consider a few encoding circuits recognizing that the decode is just the opposite.

NRZI

The NRZI encoders are trivial being only single 'and' gate as shown in Fig. 12.12. Driving the reguired FF to produce the current reversals for each bit, hence the need to 'and' the write clock to produce RZ code first. The decode is shown in Fig. 11.45B of the previous chapter.

FREQUENCY MODULATION

To encode NRZ data into FM for writing requires the use of a phase switch as well as the write clock to produce the write current reversals. The rule is, write all clocks, C^{s}, and write all, 1^{s}. The decoder is the same as Fig. 11.44. To add Precompensation requires a memory shift register in order to look ahead and behind the data being written. We can write a truth table to indicate the shift direction, but as the previous transition and the following transition will always be a clock bit; the data will be bounded and will not shift. The clocks, however, will shift, therefore, our table is simple.

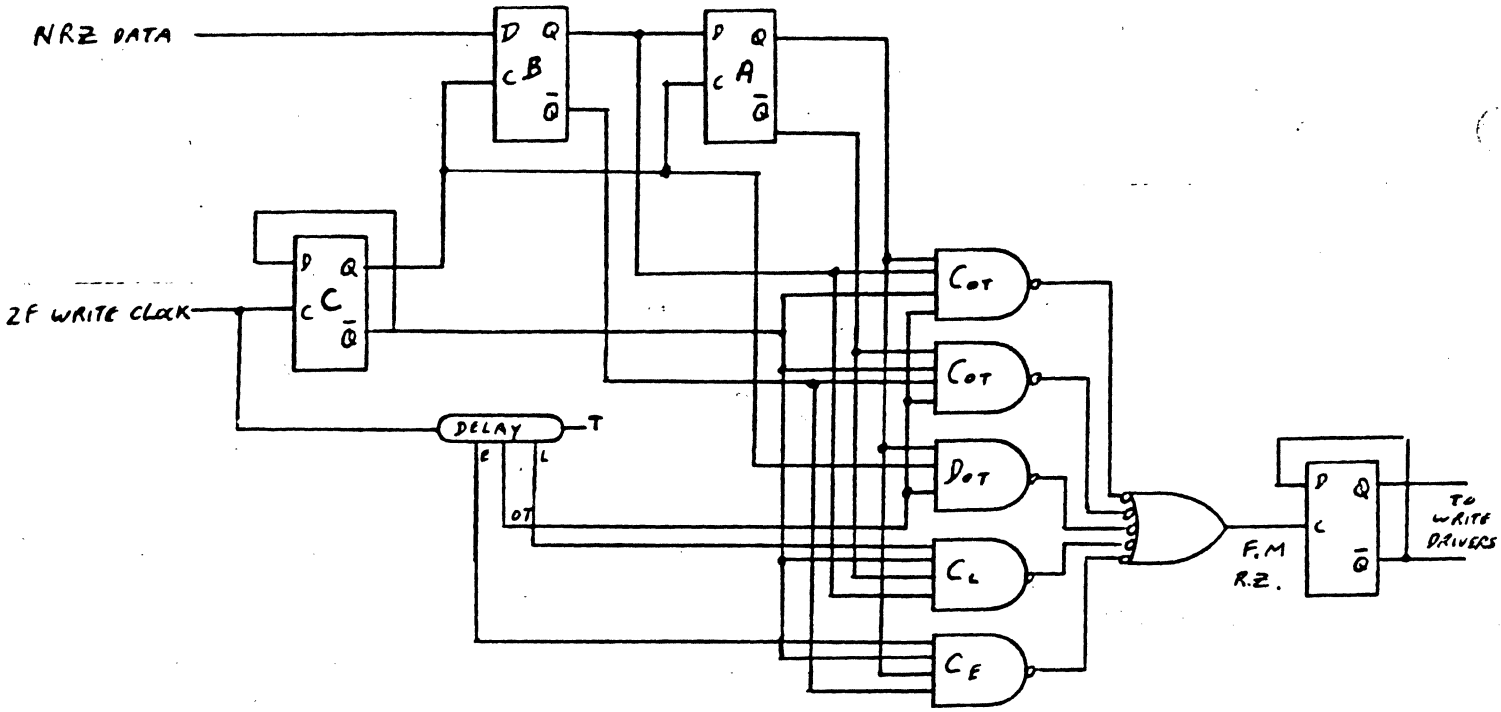


FIG 12·14

F. M. ENCODER WITH PRE COMPENSATION

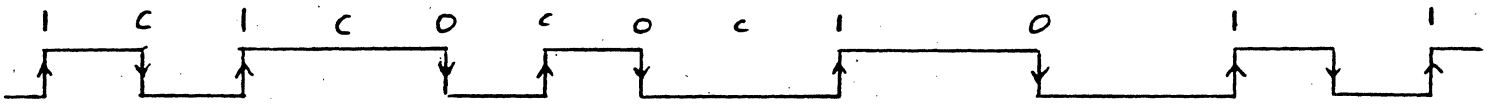


FIG 12·15 P. M ENCODED SEQUENCE

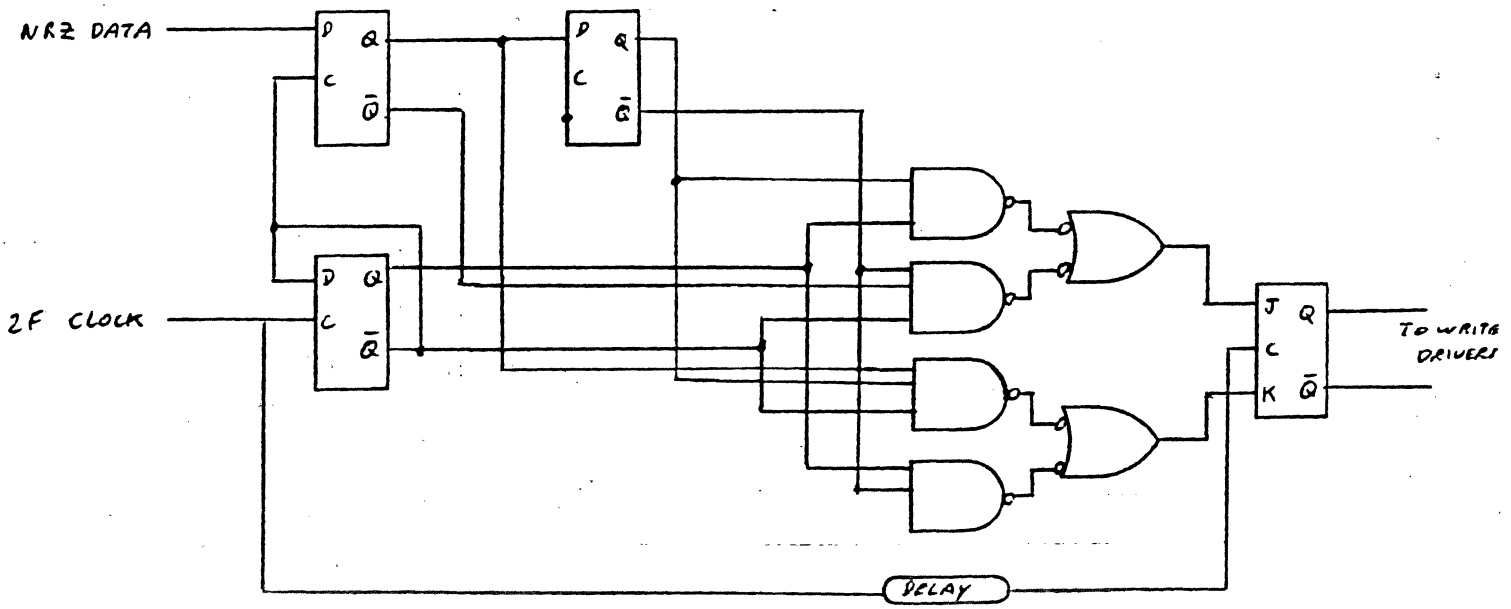


FIG 12·16 PHASE MODULATION ENCODER

TABLE 12.6

	NOW		FUTURE		
	A		B		
Early ⟵	1	C	1	Late ⟶	Clock on Time
	1	C	0		Clock Early
	0	C	1		Clock Late
	0	C	0		Clock on Time

Clock on time = $A \cdot B + \bar{A} \cdot \bar{B}$

Clock Early = $A \cdot \bar{B}$ EQ 12.15

Clock Late = $\bar{A} \cdot B$

This is implemented in Fig. 12.14. Note that the and gates must have the same propogation delay or unwanted shift will alter the data timing. The data bit A is written first on time followed by the clock bit, then the register shifts and repeats.

PHASE MODULATION

Implementing straight Phase Modulation requires truth tables in order to anticipate the phase reversals required for the clock bit which can have either polarity. This can be implemented with a J.K.-F.F. Refer to Fig. 12.15 and table 12.7. The sequence is shift-data-clock-shift.

$$J = \bar{Q}_1\bar{Q}_2\,\overline{1F} + Q_1\bar{Q}_2\,1F + Q_1Q_2\,1F$$

Q_2 \ Q_1		1		
		1		1

1F

EQ 12.16

$$K = \bar{Q}_1\bar{Q}_2\,1F + \bar{Q}_1Q_2\,1F + Q_1Q_2\,\overline{1F}$$

Q_2 \ Q_1			1	
	1		1	

1F

EQ 12.17

TABLE 12.7

NOW Q_1	PAST Q_2	1F	J	K	
0	0	0	1	0	
0	0	1	0	1	
0	1	0	0	0	no change
0	1	1	0	1	
1	0	0	0	0	no change
1	0	1	1	0	
1	1	0	0	1	
1	1	1	1	0	

$$J = Q_1\ 1F + \bar{Q}_1\bar{Q}_2\overline{1F}$$

$$K = Q_1\bar{Q}_2\overline{1F} + \bar{Q}_1\ 1F$$

All this can be implemented as shown in Fig. 12.16. The decode of phase modulation uses a slightly different detector than we have used in the past. Depending on the resolution, where we do not need a gate generator, the circuit used for a split Bi Directional Single Shot will suffice directly out of the differentiated and limited signal. If a gate generator is used, then the split Bi Directional Single Shot is the last block. The phasing must be compatable (+ = 1) because the zeros are realy superfluous; we could ignore them and just use the ones or we could use them and repeat the circuitry . Lets do the later as an exercise. To provide Precompensation for P.M., we again need a truth table only this time we need to look at four levels to anticipate the peak shift. The sequence is shift-data-clock-shift.

TABLE 12.8

PAST A	NOW B	CLOCK ↓	FUTURE C	1F	
0	0	+	0	0	+ Cot, J
0	0		0	1	Zot K
0	0	₀	1	0	no change
0	0		1	1	Z early K
0	1	₀	0	0	no change
0	1		0	1	Dot J
0	1	-	1	0	-Cot K
0	1		1	1	D late J
1	0	+	0	0	+Cot J
1	0		0	1	Z late K
1	0	₀	1	0	no change
1	0		1	1	Z ot K
1	1	₀	0	0	no change
1	1		0	1	D early J
1	1	-	1	0	–C ot K
1	1		1	1	D ot J

$$J = \bar{A}\,\bar{B}\,\bar{C}\,\overline{1F} + \bar{A}\,B\,\bar{C}\,1F + \bar{A}\,B\,C\,1F + A\,\bar{B}\,\bar{C}\,\overline{1F} + A\,B\,\bar{C}\,1F + A\,B\,C\,1F \qquad \text{EQ 12.18}$$

$$K = \bar{A}\,\bar{B}\,\bar{C}\,1F + \bar{A}\,\bar{B}\,C\,1F + \bar{A}\,B\,C\,\overline{1F} + A\,\bar{B}\,\bar{C}\,1F + A\,\bar{B}\,C\,1F + A\,B\,C\,\overline{1F} \qquad \text{EQ 12.19}$$

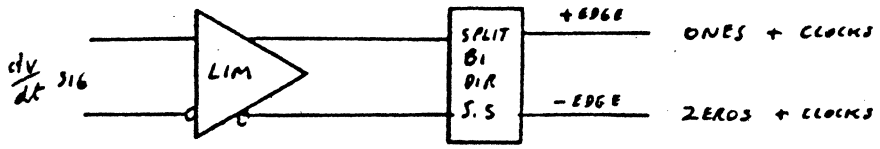


FIG 12·17A REGION 3 DETECTOR REVISION FOR PHASE MODULATED SIGNALS

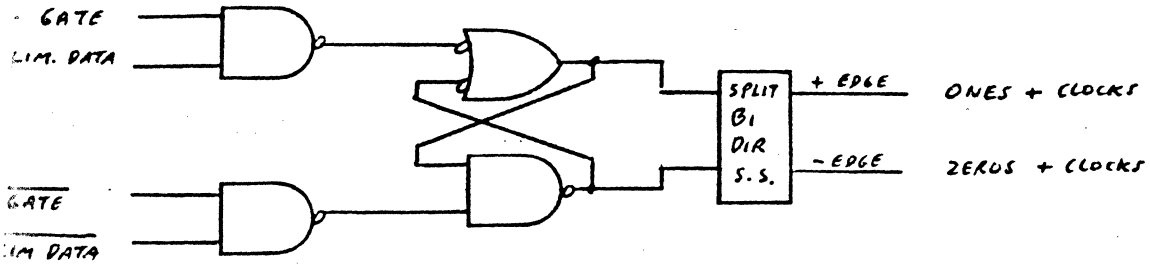


FIG 12·17B REGION 1 OR 2 DETECTOR REVISION FOR PHASE MODULATED SIGNALS

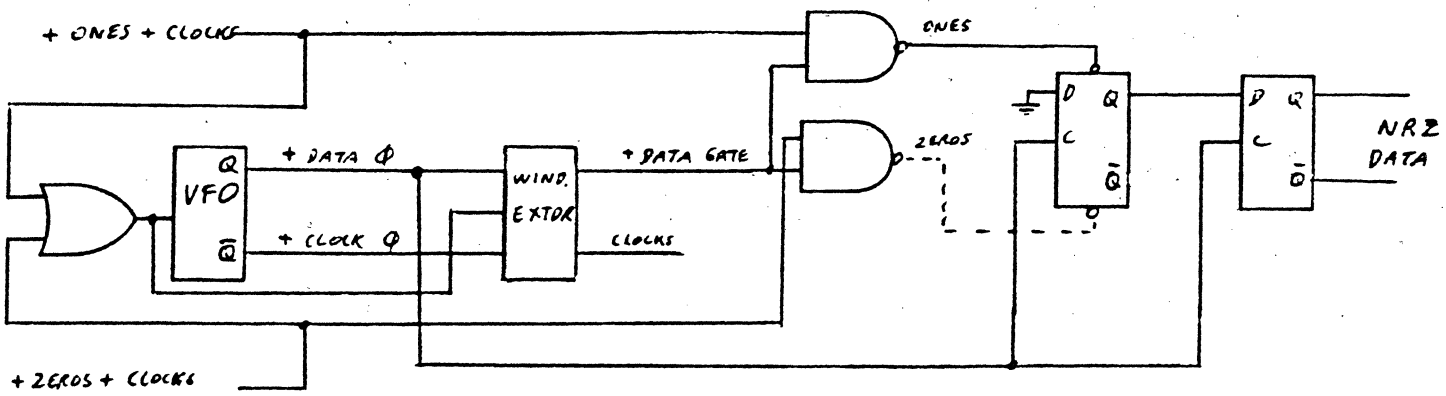
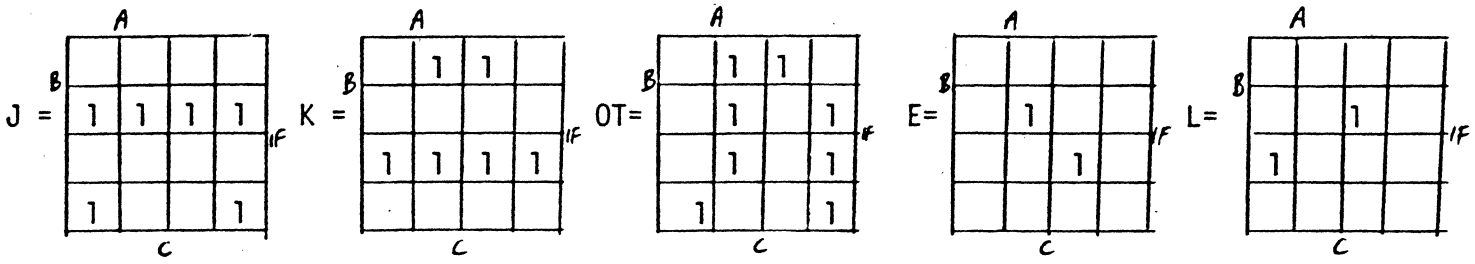


FIG 12·18 PHASE MODULATION DECODER

$$O.T. = \bar{A}\,\bar{B}\,\bar{C}\,\overline{1F} + \bar{A}\,\bar{B}\,\bar{C}\,1F + \bar{A}\,B\,\bar{C}\,1F + \bar{A}\,B\,C\,\overline{1F} + A\,\bar{B}\,\bar{C}\,\overline{1F} + A\,\bar{B}\,\bar{C}\,1F + A\,B\,C\,\overline{1F} + A\,B\,C\,1F \qquad \text{EQ 12.20}$$

$$E = \bar{A}\,\bar{B}\,C\,1F + A\,B\,\bar{C}\,1F \qquad \text{EQ 12.21}$$

$$L = \bar{A}\,B\,C\,1F + A\,\bar{B}\,\bar{C}\,1F \qquad \text{EQ 12.22}$$



The decode of these Veich diagrams is as follows

$$J = B\cdot 1F + \bar{B}\,\bar{C}\,\overline{1F} \qquad \text{EQ 12.23}$$

$$K = \bar{B}\cdot 1F + BC\,\overline{1F} \qquad \text{EQ 12.24}$$

$$\text{On Time} = \bar{B}\,\bar{C}\,\overline{1F} + \bar{A}\,\bar{C}\,1F + A\,C\,1F + B\,C\,\overline{1F} \qquad \text{EQ 12.25}$$

$$\text{Early} = \bar{A}\,\bar{B}\,C\,1F + A\,B\,\bar{C}\,1F \qquad \text{EQ 12.26}$$

$$\text{Late} = \bar{A}\,B\,C\,1F + A\,\bar{B}\,\bar{C}\,1F \qquad \text{EQ 12.27}$$

The implementation is shown in Fig. 12.19. Great care should be used to control the logic delays in the clock paths to prevent timing problems.

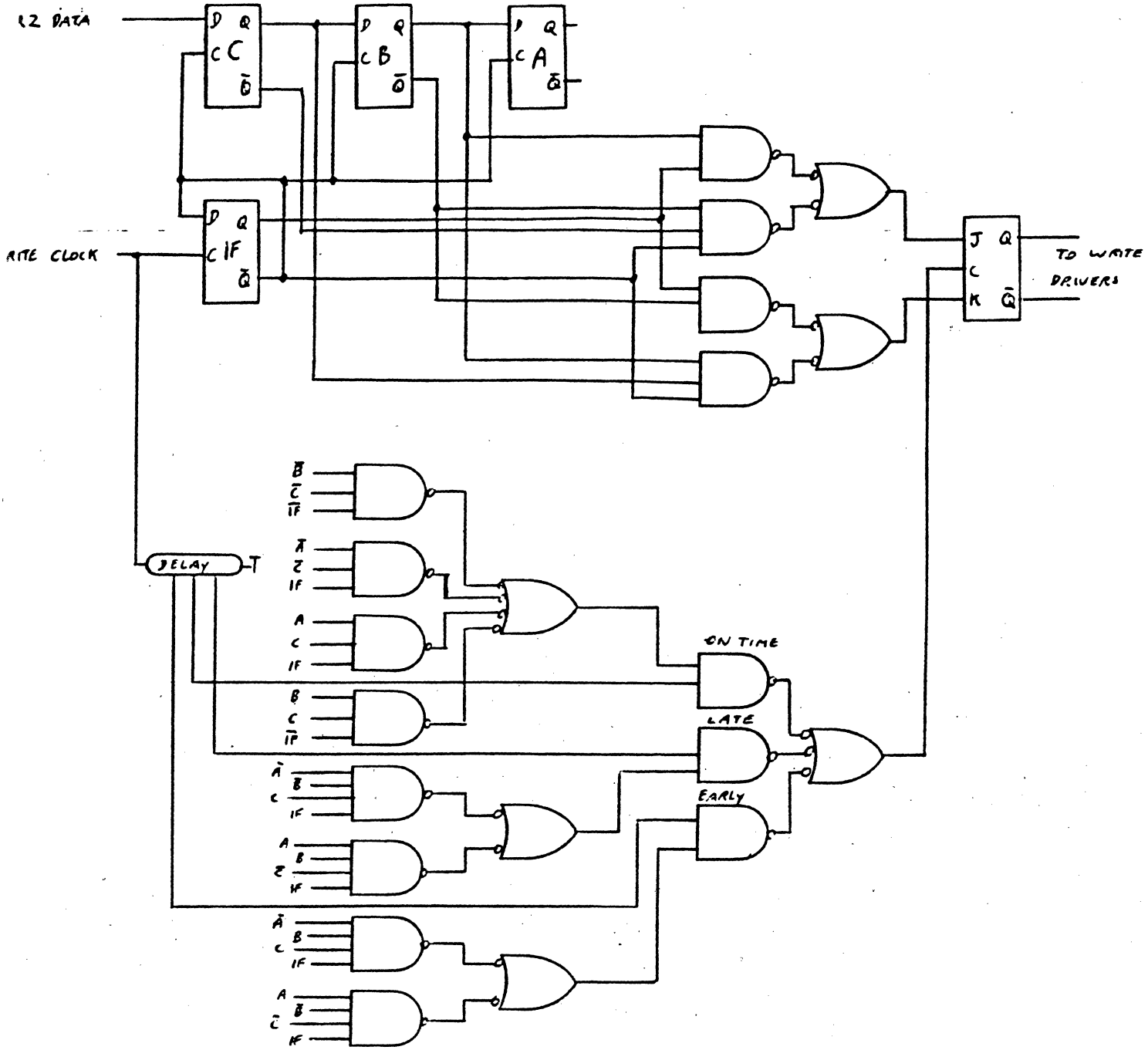


FIG 12.19

PHASE MODULATION ENCODER WITH PRE COMPENSATION

F.M.

With this code we need to know the past and future data, therefore, we ;ed a three level truth table.

ABLE 12.9

PAST A	CLOCK B (NOW)	FUTURE C	
0	0	0	clock = $\bar{A}\,\bar{B}\,\bar{C}$
0	0	1	clock = $\bar{A}\,\bar{B}\,C$
0	1	0	data = $\bar{A}\,B\,\bar{C}$
0	1	1	data = $\bar{A}\,B\,C$
1	0	0	-
1	0	1	-
1	1	0	data = $A\,B\,\bar{C}$
1	1	1	data = $A\,B\,C$

CLOCK =

B \ A				
			1	1

C

DATA =

B \ A				
	1	1	1	1

C

The decode is clocks = $\bar{A}\,\bar{B}$, data = B. Therefore, we can delete the C FF or "future" and deal only with the past and present. The sequence is shift, clock, data, shift. The implementation is given in Fig. 12.20. Note that the set up and timing paths prevent unequal logic delays from hurting bit timing. To add precompensation, we require a four level truth table (Table 12.10). We will shift-data-clock-shift.

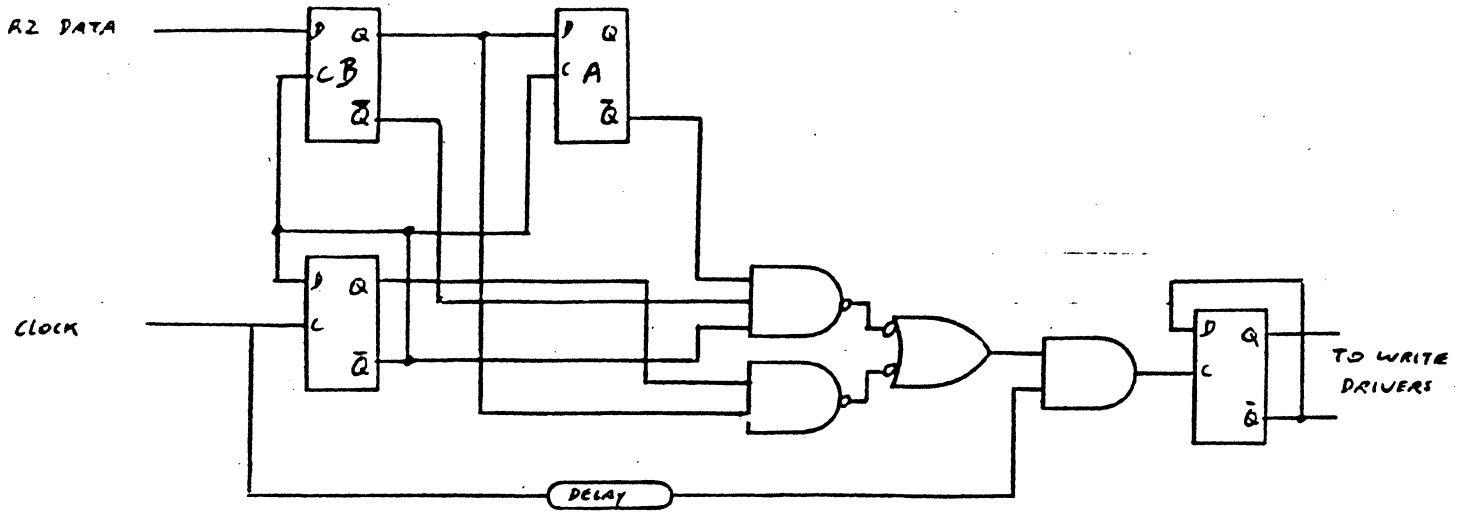


FIG 12·20 M.F.M ENCODER

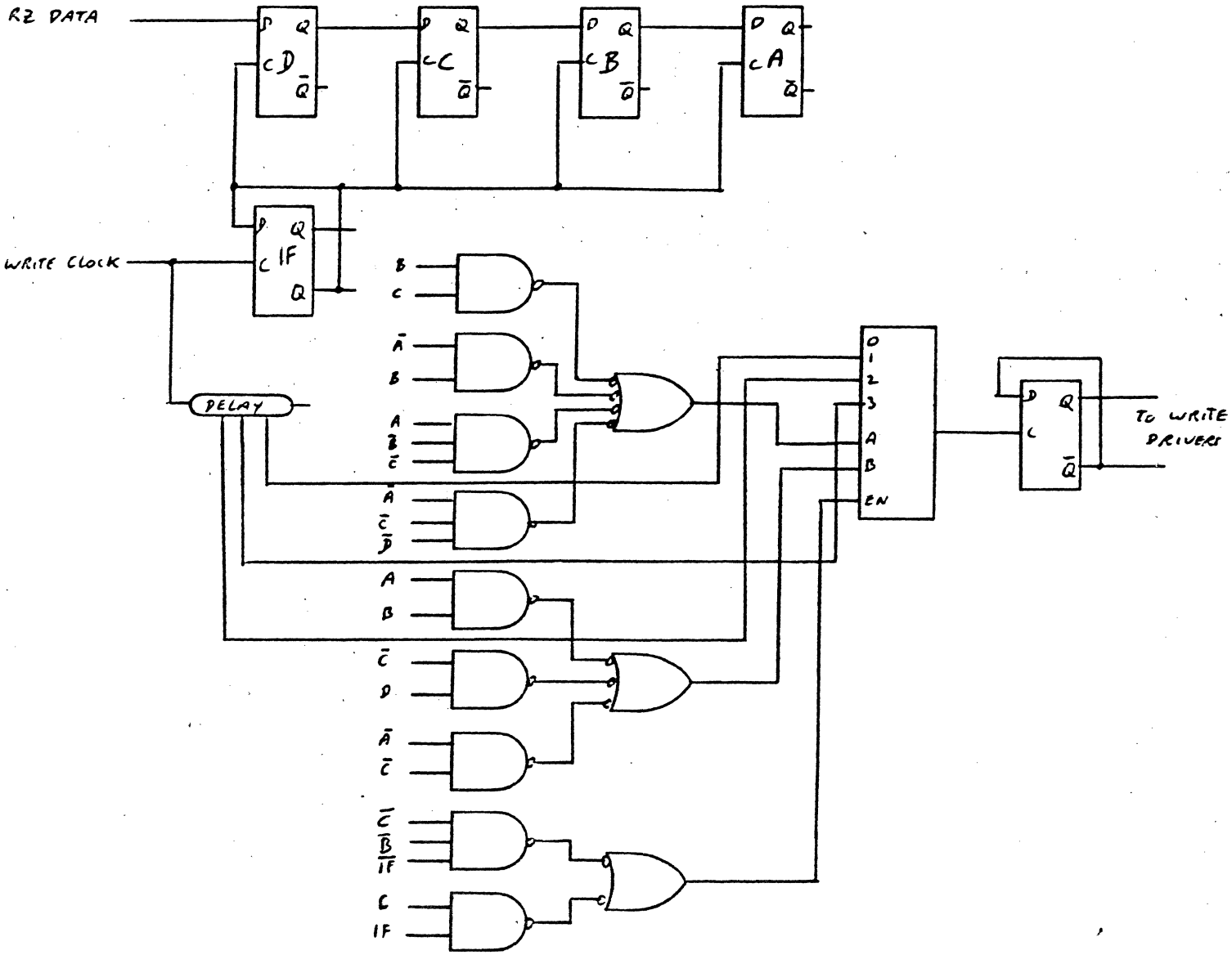


FIG 12·21 M.F.M ENCODER WITH PRE COMP. (3 LEVEL)

TABLE 12.10

PAST A	NOW B	CLOCK C ↓	FUTURE D		
0	0	0	0	0	clock OT
0	0	0	1	1	clock early
0	0	1	0	2	-
0	0	1	1	3	-
0	1	0	0	4	data OT
0	1	0	1	5	data OT
0	1	1	0	6	data late
0	1	1	1	7	data late
1	0	0	0	8	clock late
1	0	0	1	9	clock OT
1	0	1	0	10	-
1	0	1	1	11	-
1	1	0	0	12	data early
1	1	0	1	13	data early
1	1	1	0	14	data OT
1	1	1	1	15	data OT

$$\text{clock OT} = \bar{A}\,\bar{B}\,\bar{C}\,\bar{D} + A\,\bar{B}\,\bar{C}\,\bar{D} = 0 + 9 \qquad \text{EQ 12.28}$$

$$\text{clock early} = \bar{A}\,\bar{B}\,\bar{C}\,D = 1 \qquad \text{EQ 12.29}$$

$$\text{clock late} = A\,\bar{B}\,\bar{C}\,\bar{D} = 8 \qquad \text{EQ 12.30}$$

$$\text{data OT} = \bar{A}\,\bar{B}\,\bar{C}\,\bar{D} + \bar{A}\,B\,\bar{C}\,D + A\,B\,C\,\bar{D} + A\,B\,C\,D = 4+5+14+15 \qquad \text{EQ 12.31}$$

$$\text{data early} = A\,B\,\bar{C}\,\bar{D} + A\,B\,\bar{C}\,D = 12 + 13 \qquad \text{EQ 12.32}$$

$$\text{data late} = \bar{A}\,B\,C\,\bar{D} + \bar{A}\,B\,C\,D = 6 + 7 \qquad \text{EQ 12.33}$$

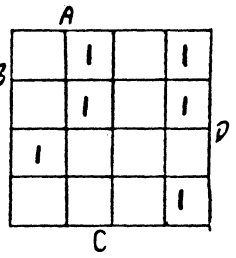
that is a lot of logic, therefore, we could do an intermediate step that addresses only on time, early, and late with a gate for data or clocks. Data becomes B and clocks $\bar{B}$ $\bar{C}$, therefore, the enable is $B \cdot IF + \bar{C}\,\bar{B} \cdot \overline{IF}$

On time is = $\bar{A}\,\bar{B}\,\bar{C}\,\bar{D} + A\,\bar{B}\,\bar{C}\,D + \bar{A}\,B\,\bar{C}\,\bar{D} + \bar{A}\,B\,\bar{C}\,D + A\,B\,C\,\bar{D} + A\,B\,C\,D$ EQ 12.34

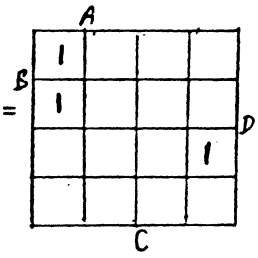
Early = $\bar{A}\,\bar{B}\,\bar{C}\,\bar{D} + A\,B\,\bar{C}\,\bar{D} + A\,B\,\bar{C}\,D$ EQ 12.35

Late = $A\,\bar{B}\,\bar{C}\,\bar{D} + \bar{A}\,B\,C\,\bar{D} + \bar{A}\,B\,C\,D$ EQ 12.36

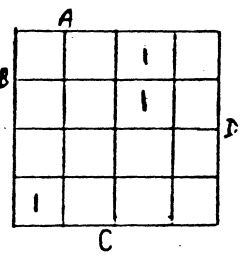
OT =



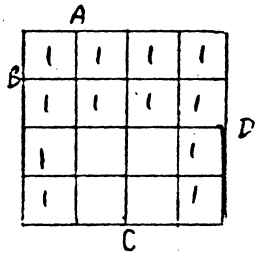
Early =



Late =



Enable



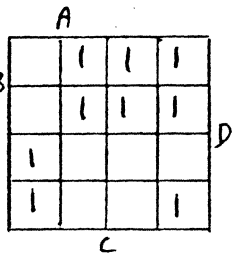
2'	2°	
0	0	-
0	1	Late
1	0	Early
1	1	On Time

TABLE 12.11

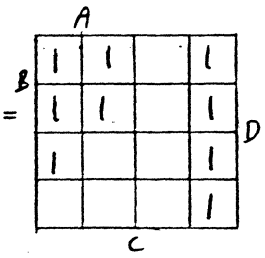
This is still a lot of decode. We could assign a two level state to each such as 1 = late, 2 = early, and 3 = on time.

2° = late + on time
2' = early + on time

2° =



2' =



$$2^{\circ} = BC + \bar{A}B + A\bar{B}\bar{C} + \bar{A}\bar{C}\bar{D} \qquad \text{EQ 12.37}$$

$$2' = AB + \bar{C}D + \bar{A}\bar{C} \qquad \text{EQ 12.38}$$

which is much easier. This can be implemented with a four line multiplexer.

We might profitably study five level Precompensation. As can be imagined, the decode is a little more difficult. We will base the decode on two levels of early and late. A 0110 pattern gives the greatest peak shift L^2, E^2 and a 01110 lessor L, E. We will again follow the shift-data-clock-shift sequence.

TABLE 12.12

early	Past A	B	Now C (clock ↓)	D	Future E				A	B	C (clock ↓)	D	E	
	0	0	0	0	0	C	OT	7	1	0	0	0	0	C OT 7
	0	0	0	0	1	C	E	1	1	0	0	0	1	C E^2 3
	0	0	0	1	0		-		1	0	0	1	0	-
	0	0	0	1	1		-		1	0	0	1	1	-
	0	0	1	0	0	D	OT	7	1	0	1	0	0	D OT 7
	0	0	1	0	1	D	OT	7	1	0	1	0	1	D OT 7
	0	0	1	1	0	D	L^2	6	1	0	1	1	0	D L^2 6
	0	0	1	1	1	D	L	2	1	0	1	1	1	D L 2
	0	1	0	0	0	C	L	2	1	1	0	0	0	C L 2
	0	1	0	0	1	C	OT	7	1	1	0	0	1	C OT 7
	0	1	0	1	0		-		1	1	0	1	0	-
	0	1	0	1	1		-		1	1	0	1	1	-
	0	1	1	0	0	D	E^2	3	1	1	1	0	0	D E 1
	0	1	1	0	1	D	E^2	3	1	1	1	0	1	D E 1
	0	1	1	1	0	D	OT	7	1	1	1	1	0	D OT 7
	0	1	1	1	1	D	OT	7	1	1	1	1	1	D OT 7
			↑ clock								↑ clock			

The number conversion idea will help reduce the logic, therefore, we will make a new table , 12.13.

.13

-
Early
Late
$Early^2$
-
-
$Late^2$
On Time

n write the boolian equation for each condition

$\bar{A}\bar{B}\bar{C}\bar{D}\bar{E} + \bar{A}\bar{B}\bar{C}\bar{D}E + \bar{A}\bar{B}\bar{C}D\bar{E} + \bar{A}\bar{B}\bar{C}DE + \bar{A}\bar{B}C\bar{D}\bar{E} + \bar{A}\bar{B}CDE + \bar{A}B\bar{C}\bar{D}\bar{E} +$

$\bar{A}B\bar{C}\bar{D}E + \bar{A}B\bar{C}D\bar{E} + \bar{A}B\bar{C}DE + \bar{A}BC\bar{D}\bar{E} + \bar{A}BCDE + AB\bar{C}\bar{D}\bar{E} + A\bar{B}\bar{C}\bar{D}E + A\bar{B}\bar{C}D\bar{E} + A\bar{B}\bar{C}DE + A\bar{B}C\bar{D}\bar{E} + A\bar{B}CDE + AB\bar{C}\bar{D}\bar{E} + AB\bar{C}\bar{D}E + AB\bar{C}D\bar{E} + ABC\bar{D}\bar{E} + ABCDE = 111$ EQ 12.38

$\bar{A}\bar{B}CD\bar{E} + A\bar{B}CD\bar{E} = 110$ EQ 12.39

$\bar{A}BCD\bar{E} + \bar{A}BC\bar{D}E + A\bar{B}\bar{C}\bar{D}E = 011$ EQ 12.40

$\bar{A}\bar{B}CDE + \bar{A}B\bar{C}\bar{D}\bar{E} + A\bar{B}CDE + AB\bar{C}\bar{D}\bar{E} = 010$ EQ 12.41

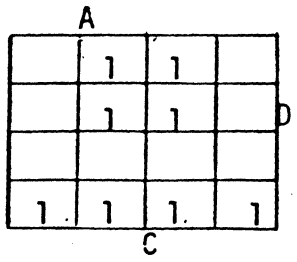
$\bar{A}\bar{B}\bar{C}\bar{D}E + AB\bar{C}\bar{D}\bar{E} + AB\bar{C}D\bar{E} = 001$ EQ 12.42

we can write the equations for the bits as a function of powers of 2.

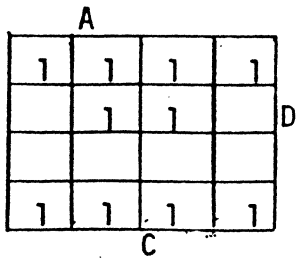
7 + 3 + 1
7 + 6 + 3 + 2
7 + 6 EQ 12.43

The Veich Diagrams are as follows:

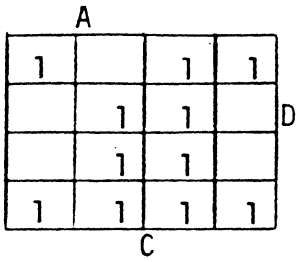
$2^\circ = \bar{E}\cdot$



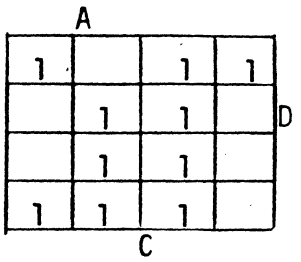
$+ E\cdot$



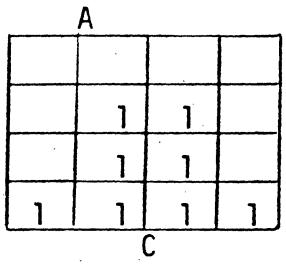
$2' = \bar{E}\cdot$



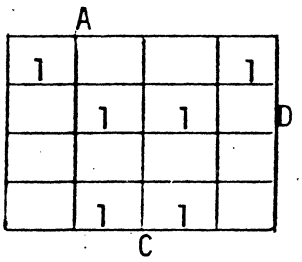
$+ E\cdot$



$2^2 = \bar{E}\cdot$



$+ E\cdot$



The boolian reduction of these are given below

$$2^\circ = \bar{E}(BC + \bar{B}\bar{D}) + E(\bar{D} + BC) = BC + \bar{E}\bar{B}\bar{D} + E\bar{D} \qquad \text{EQ 12.44}$$

$$2' = \bar{E}(\bar{C}\bar{D} + \bar{B}C + \bar{A}C + CD) + E(CD + \bar{A}C + A\bar{B}\bar{D} + B\bar{C}\bar{D})$$
$$= \bar{E}\bar{C}\bar{D} + \bar{E}\bar{B}C + \bar{A}C + CD + EA\bar{B}\bar{D} + EB\bar{C}\bar{D} \qquad \text{EQ 12.45}$$

$$2^2 = \bar{E}(\bar{B}\bar{D} + CD) + E(\bar{B}C\bar{D} + BCD + B\bar{C}\bar{D})$$
$$= \bar{E}\bar{B}\bar{D} + \bar{E}CD + E\bar{B}C\bar{D} + EBCD + EB\bar{C}\bar{D} \qquad \text{EQ 12.46}$$

$$\text{Enable} = IFC + \bar{B}\bar{D}\bar{I}F \qquad \text{EQ 12.47}$$

These three inputs can be fed into a multiplexer together with the appropriate delayed 2F clock pulses as shown in Fig. 12.22.

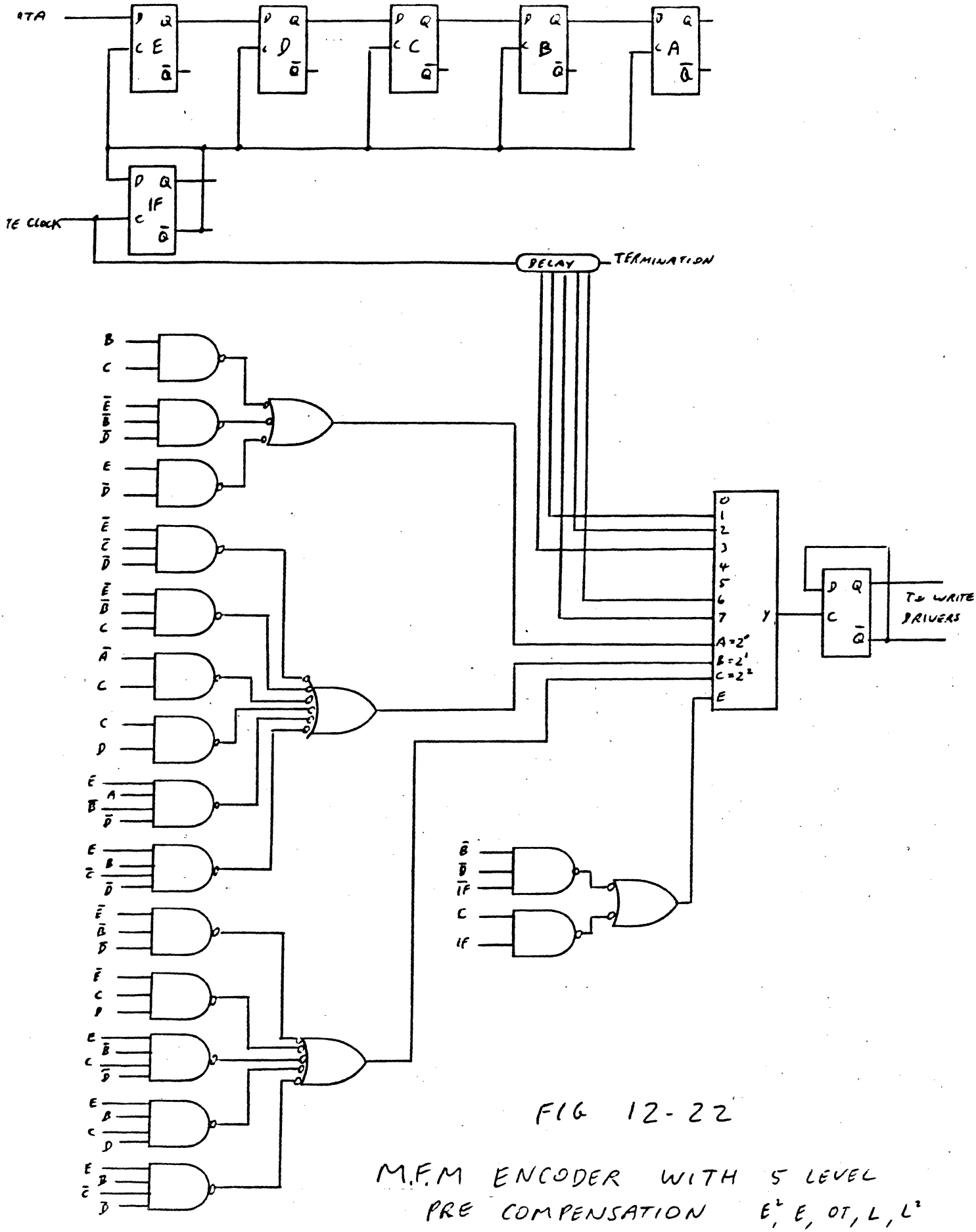


FIG 12-22

M.F.M ENCODER WITH 5 LEVEL PRE COMPENSATION E^2, E, OT, L, L^2

Timing considerations may require the four input gates to be clocked into a register before application to the multiplexer. As the enable conditions are unique or in other words require separate clock and date pulses, the phasing must be considered and is shown included.

We will next discuss the implementation of one of the group codes. As there are several ways of doing this, we will discuss the full logic approach as it is the most complex. The other approaches use memory look-up tables and micro processors. These will be left to the designers as they are much easier to implement. We will choose a 2/3 (1,7) code as one example. The nature of this code requires two input data bits to be encoded into three cells. the conditions are satisfied by developing a table of assigned values.

TABLE 12.13

00 = 001
01 = 010
10 = 100
11 = 101

Now we can easily see that if we had a sequence 0010 we would write 001100 which violates the requirement of one zero between transitions, 2/3 (1,7), therefore, we must make a new table that assigns an alternate 000 symbol to be substituted where necessary. This alternate symbol can be called into use if the next two bits require it. When making the table, the time sequence of the bits and the code must be kept in mind as the following implementation requires correct sequencing.

TABLE 12.14

INPUT DATA time→		ENCODED TRANSITIONS time→			WRITTEN time→
NOW	NEXT	NOW	NEXT	SUBSTITUTIONS	
00	00	001	001		001
00	01	001	010		001
00	10	001	100	001000	001000
00	11	001	101	010000	010000
01	00	010	001		010
01	01	010	010		010
01	10	010	100		010
01	11	010	101		010
10	00	100	001		100
10	01	100	010		100
10	10	100	100		100
10	11	100	101		100
11	00	101	001		101
11	01	101	010		101
11	10	101	100	000100	000100
11	11	101	101	101000	101000

Now if we examin the substituted code, we see that the widest spacing between transitions is for the data sequence 0011,1110 which is written as 010000000100 which gives 7 zeros in a row maximum, hence, 7 in the code description 2/3 (1,7). When the code is implemented, the upcomming data is examined to see if a substitution is required. If so, then all four input bits are taken and written

$$\text{Cell 1} = A\bar{C} + AD + A\bar{B}$$

$$\text{Cell 2} = \bar{A}B + \bar{A}CD$$

$$\text{Cell 3} = AB\bar{C} + ABD + \bar{A}\bar{B}\bar{D} + \bar{A}\bar{B}\bar{C} \qquad \text{EQ 12.52}$$

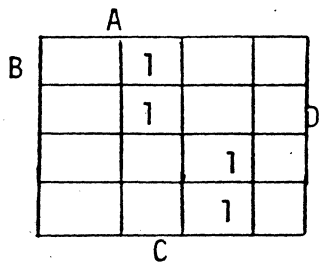
$$\text{Cell 4} = B\bar{D}$$

$$\text{Cell 5} = \text{Cell 6} = 0$$

The conditions for taking four input bits instead of two are given in EQ 12.53.

$$\text{Four} = \bar{A}\bar{B}C\bar{D} + \bar{A}\bar{B}CD + AB\bar{C}\bar{D} + ABCD \qquad \text{EQ 12.53}$$

Four =



$$= ABC + \bar{A}\bar{B}C \qquad \text{EQ 12.54}$$

This can be fed into a down counter set lines to generate an overflow pulse that occurs every 2 or 4 bits (Set 1 or Set 3). The overflow pulse can load a six bit shift register for the output cell information. The NRZ to RZ encoder is added at the output of the 6 bit shift register followed by the write current reversal FF. The entire implementation is shown in Fig. 12.23. The Read Decoder is implemented following the NRZ output of the Read Detector. If we refer to Table 12.15, we can write the equations for decoding the data. We will write it in Octal to save space.

$$A = 1 + 5 + 1.0 \qquad \text{EQ 12.55}$$

If we look at the Binary written code transition cells, 5 and 6 are always zero, therefore, we only need to decode transition cells 4, 3, 2, and 1 for write purposes.

$$\text{Cell 1} = A\bar{B}\bar{C}\bar{D} + A\bar{B}\bar{C}D + A\bar{B}C\bar{D} + A\bar{B}CD + AB\bar{C}\bar{D} + AB\bar{C}D + ABCD \qquad \text{EQ 12.48}$$

$$\text{Cell 2} = \bar{A}\bar{B}CD + \bar{A}B\bar{C}\bar{D} + \bar{A}B\bar{C}D + \bar{A}BC\bar{D} + \bar{A}BCD \qquad \text{EQ 12.49}$$

$$\text{Cell 3} = \bar{A}\bar{B}\bar{C}\bar{D} + \bar{A}\bar{B}\bar{C}D + \bar{A}\bar{B}C\bar{D} + AB\bar{C}\bar{D} + AB\bar{C}D + ABCD \qquad \text{EQ 12.50}$$

$$\text{Cell 4} = ABC\bar{D} + (\bar{A}\bar{B}\bar{C}\bar{D} + \bar{A}\bar{B}\bar{C}D + \bar{A}B\bar{C}\bar{D} + \bar{A}B\bar{C}D + \bar{A}BC\bar{D} + \bar{A}BCD + A\bar{B}\bar{C}\bar{D} + A\bar{B}\bar{C}D + A\bar{B}C\bar{D} + A\bar{B}CD + AB\bar{C}\bar{D} + AB\bar{C}D) \qquad \text{EQ 12.51}$$

of which the bracketed terms are redundant since that part is written separately as transition 1 of the next pair of input bits.

1 =

B \ A				
	1			
	1	1		
	1	1		
	1	1		

C D

2 =

B \ A				
			1	1
			1	1
			1	

C D

3 =

B \ A				
	1			
	1	1		
				1
			1	1

C D

4 =

B \ A				
	X	1	X	X
	X		X	X
	X	X		X
	X	X		X

C D

from which we can reduce the conditions to the following

as six cells, if not, then only two bits are taken and the three cells written. Since the implementation will be done with shift registers, it might be profitable to rewrite the table in shift register form. This is given in Table 12.15.

TABLE 12.15 $\frac{2}{3}$ (1,7) Code in Shift Reg. Form

BINARY				OCTAL	
INPUT DATA		WRITTEN CODE		INPUT DATA	WRITTEN CODE
D C	B A	654	321		
0 0	0 0	XXX	100	0 0	4
1 0	0 0	XXX	100	2 0	4
0 1	0 0	000	100	1 0	0 4
1 1	0 0	000	010	3 0	0 2
0 0	1 0	XXX	010	0 2	2
1 0	1 0	XXX	010	2 2	2
0 1	1 0	XXX	010	1 2	2
1 1	1 0	XXX	010	3 2	2
0 0	0 1	XXX	001	0 1	1
1 0	0 1	XXX	001	2 1	1
0 1	0 1	XXX	001	1 1	1
1 1	0 1	XXX	001	3 1	1
0 0	1 1	XXX	101	0 3	5
1 0	1 1	XXX	101	2 3	5
0 1	1 1	001	000	1 3	1 0
1 1	1 1	000	101	3 3	0 5

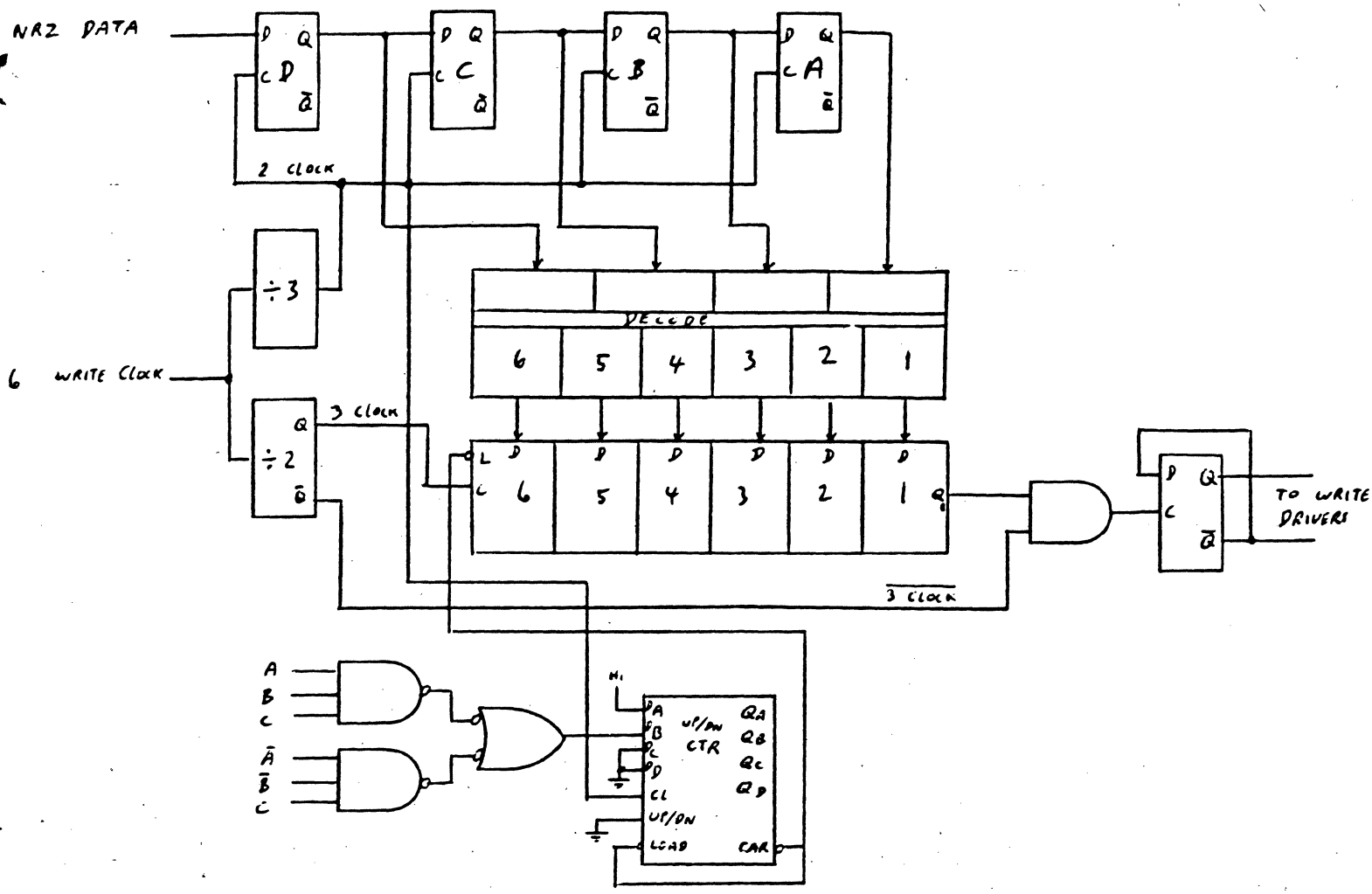


FIG 12.23

$\frac{2}{3}$ (1, 7) WRITE ENCODER

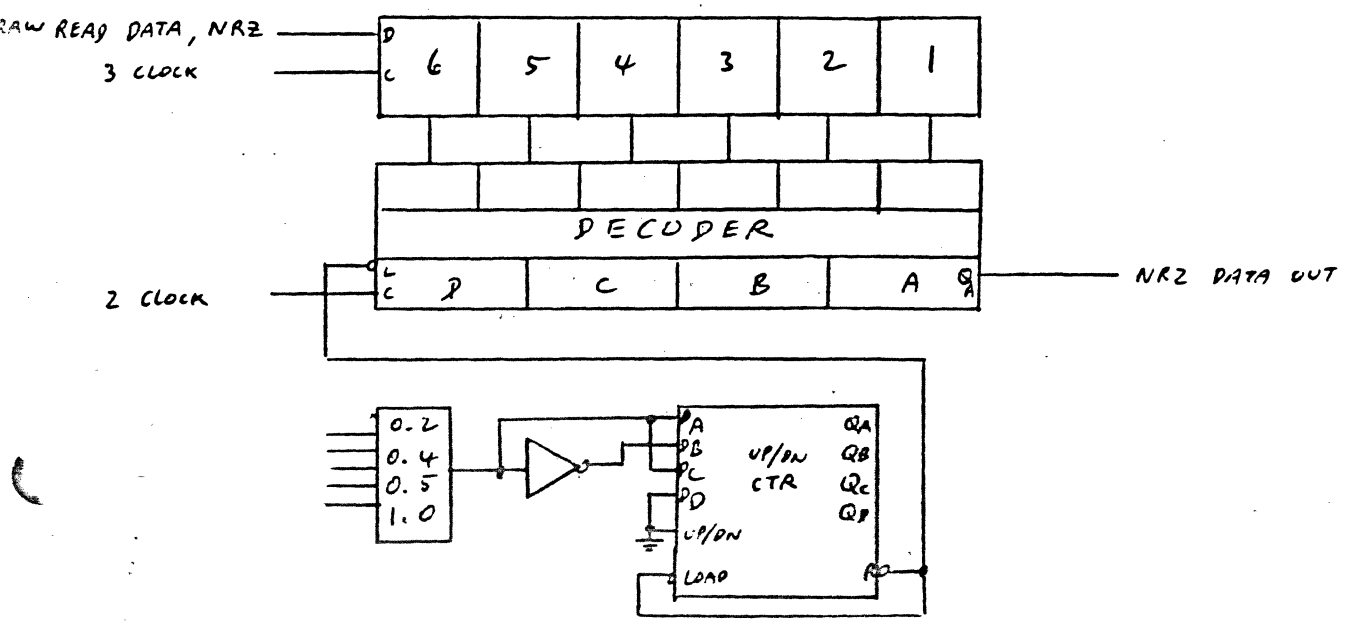


FIG 12.24 $\frac{2}{3}$ (1, 7) READ DECODER

B = 2 + 5 + 1.0 with (0.2 suppression) EQ 12.56

C = 0.4 + 0.2 + 2 + 1 +1.0 + 0.5 EQ 12.57

D = 4 + 0.2 + 2 + 1 + 5 + 0.5 EQ 12.58

but since C and D are replaced in all except the double combinations, then we are free to ignor them or as redundant bits to reduce the logic.

C = 0.4 + 0.2 + 1.0 + 0.5 EQ 12.59

D = 0.2 + 0.5 EQ 12.60

Again, the 2 or 4 bit grouping is controlled by C or D which can activate the "B" bit of a down counter as before as we can keep track of the output decode. See Fig. 12.24. The decode itself is easily implemented by using two 3 line decoders as in Fig. 12.25.

The preceding logic implementation shows one way of generating the encoding or decoding logic. The method is straight forward and should be applicable for the run length limited codes regardless of the number of substitutions or conditions. In this code we had 2 levels meaning we had our information in either two or four bit byte. Therefore, $\frac{m}{n}$ was either $\frac{2}{3}$ or $\frac{4}{6}$. As a contrast, the $\frac{1}{2}$ (2,7) code presently used in the IBM 3370 is a three level code meaning $\frac{m}{n} = \frac{1}{2}, \frac{2}{4}$, or $\frac{3}{6}$.

The next consideration is the format or the preamble to the data. Notice that the decode on Read Back depends on the phase of the clocking circuits. it is imperitive that only one of the 3 phases be used. This can be controlled

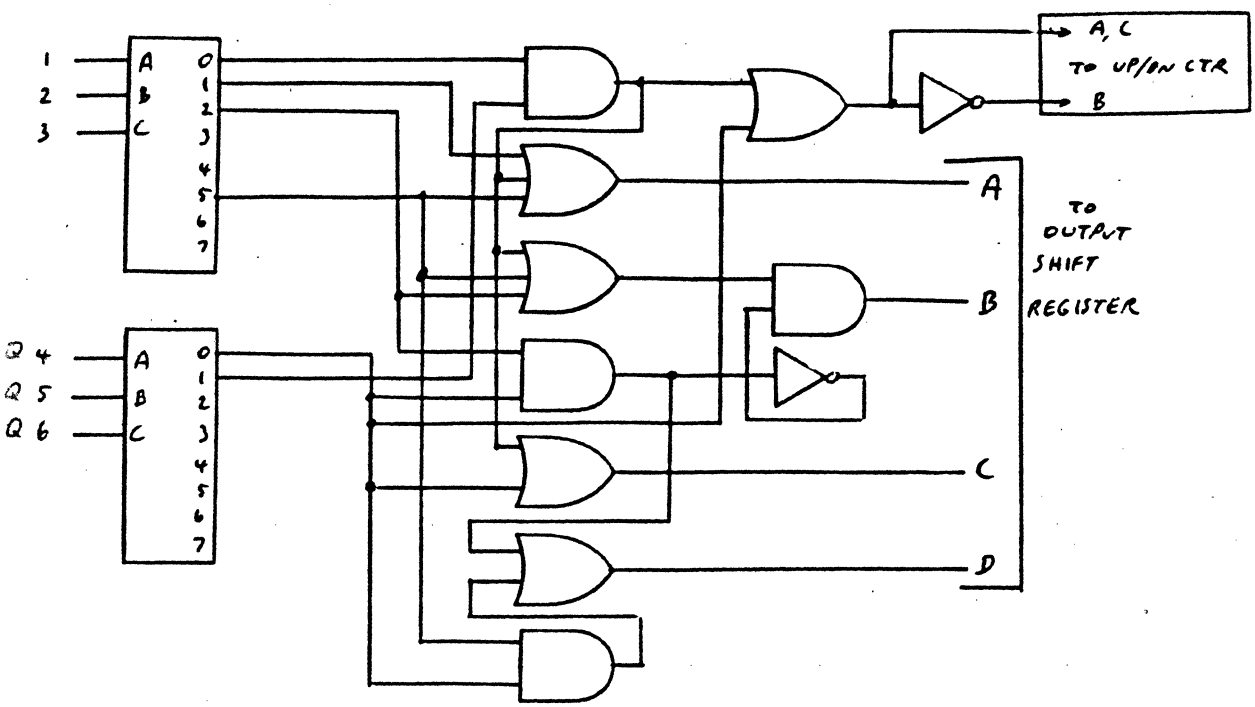


FIG 12-25

$\frac{2}{3}(1, 7)$ DECODE USING 2 - 3 LINE DECODERS

by recording say all zero's meaning the transition will be 001001001 as a time sequence for VFO sync. then we can follow up with a double character such as 001000, which is line 3 of Table 12.15. When this pattern is recognized, the count of the UP/DN counter can be set to five, thus, starting the decode in the correct phase.

ERROR CORRECTING CODES

Ever since the IBM 3330, Disc Drive error correcting codes have been used to detect errors and correct certain kinds of errors. These have become necessary due to the media coating thickness reductions which allow pinholes or small oxide conglomerates which cause missing or extra bits to disturb the data. These codes usually are designed to detect error spreads greater than the correctable spreads. For example, a code may be designed to correct 5 bits in length, but only detect 6 or greater. As might be expected, the ability of the code to detect certain sequences of errors is limited, thus, it is possible that certain sequences may be undetected. These are predictable and a probability is assigned to this occurrance. It is not the intention here to develop these codes or go into the math behind them. There is considerable literature on this subject and the reader is refered to those listed to start. The Fire, Hamming, Goppa, and Read Solomon codes have been used for some time. These only correct single burst errors or errors occurring over a short span. There are other codes that can correct multi-burst errors meaning that groups of errors can occur in a single record separated by a considerable number of bits. These are the interleaved BCH and RS codes. The reader is refered to an invited paper presented by Dr. E.R. Berlekamp published in the IEEE Proceedings, Vol 68, May 1980, p. 564-593.

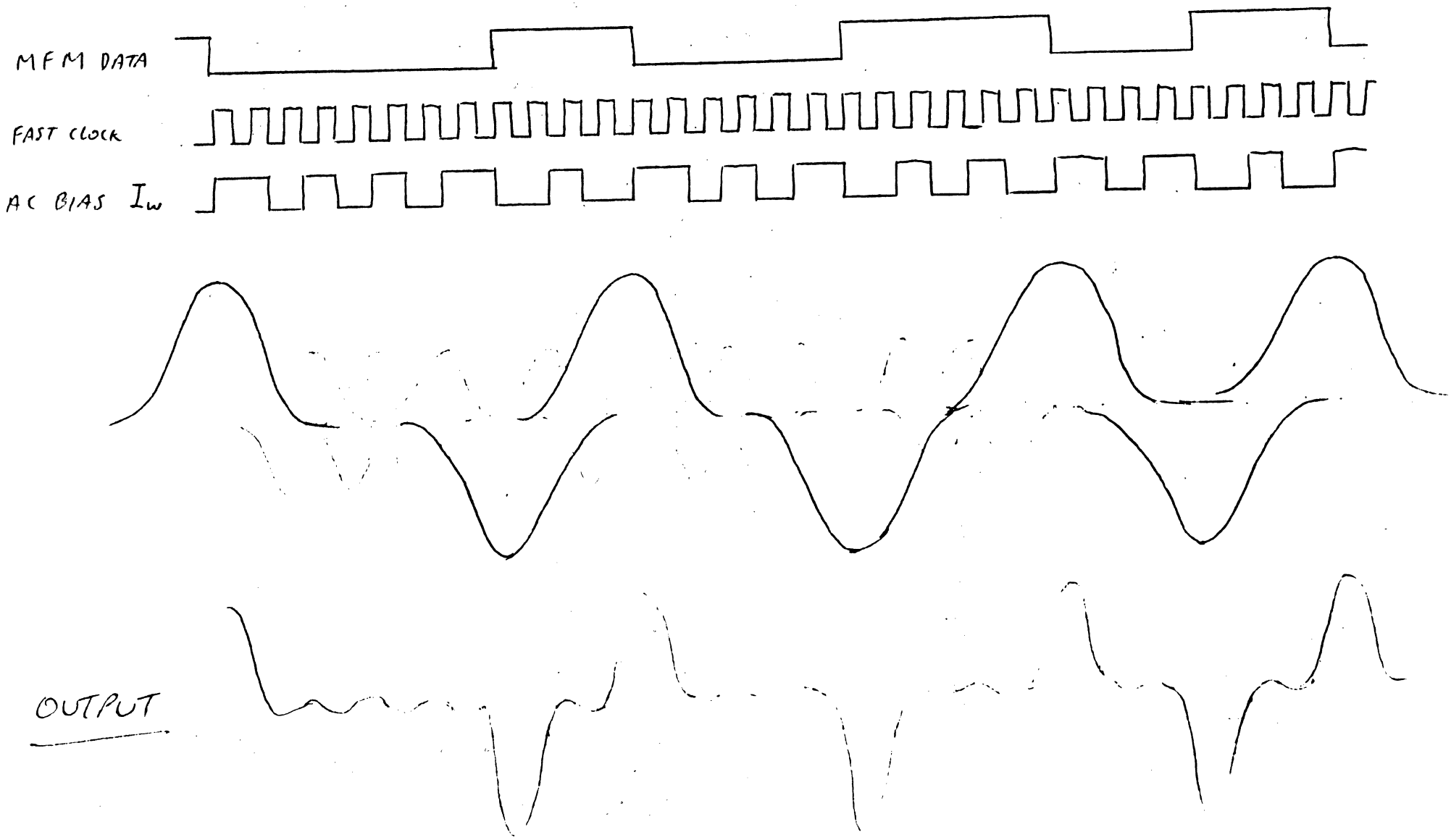


FIG 12-26 A.C. BIAS WRITE WAVEFORMS USED TO REDUCE THE P.W.50 OF THE READ BACK PULSE

13. RECORDING CHANNEL TESTING

Following the design and implementation of a particular recording channel and head combination, the designer must prove his design can meet the machine specifications for error performance. Typically, a channel has been specified as contributing one error in y x 10^{z} bits transferred. Meaning that over a period of time the total number of bits in error divided by the total number of bits transferred during the Read mode becomes a measure of the channels performance. This may be designated for all heads and tracks (for disc files), using a random number generator for the head and track addresses, or it may just be the total sequential file. Usually the random access method is used as it affords the greatest sensitivity of the test, including both inner and outer tracks for all heads. In the past this number was obtained using brute force methods; in other words, actually transferring, say an order of magnitude greater number of bits and counting the errors. Games have been played by testers wherein they automatically add two bits of error to the total saying that statistically an error could have been made just before the test started and another just after the test finished. Also, only soft errors are counted. The hard errors are ignored. The definition of soft and hard errors is the subject of some controversy. Usually soft errors are designated as those errors which do not repeat at the same physical location either following a single write or allowing multiple write updates. The one used must be specified. Hard errors are the repeaters. These can be attributed to disc or media defects at a location designated by head number, track number, and bit count. The two types of errors are specified separately.

With the introduction of error correcting codes, the definition changed for the two types of errors. Soft errors are defined as correctable errors, or those that are within the error correcting codes capability to correct. Hard errors are designated as those that cannot be corrected, or that fall outside the codes correction capability. A third category then contains the media defects which are still not counted in the error performance but may be specified for the total machine. For example, a machine may have a specification that places a maximum on the number of defects allowed either by total machine, by surface, or by track, or some combination. Read errors are caused by the failure of the recorded transition, if it exists, to generate the correct output bit in the correct sequential time slot of a data stream. There are several mechanisms for this, the dominant ones being noise and interpulse interference or bit shift. Both of these were discussed in Chapter 4, particularly as shown in Figs. 4.11 a, b, and c. This, of course, is true only if the correct channel design has been made, including heads, amplifiers, filters, detector type, and the clocking circuit.

Following work done by Dr. E. Katz and later including Dr. T. Campbell, at Memorex, a method was disclosed that greatly reduced the time required to characterize a channel's error performance. This method permitted the separation of the two dominant error mechanisms for the first time, which permitted optimization of the various channel characteristics for best performance.

Obviously, brute force design and testing costs money, particularly if over design is used in one area in order to compensate for poor design in another. This is now unnecessary. Using a circuit designed by Mr. M. Monett, a curve ~~is~~ can be generated that totally describes the error performance of a machine.

The curve has been designated as a Marginalized Variable Frequency Oscillator (MVFO) curve. This is an unfortunate choice, as the name really describes a totally different device; both devices will be discussed in this chapter. For want of a better name, we will designate the curve in question as a Quantified Phase Error Curve (QPEC). Basically, the curve is generated by plotting the total number of bits whose phase shift exceeds a set value, usually expressed in seconds, for a series of values. The reference phase is obtained from a phase locked loop with a low bandwidth. The logic output of the phase detector is compared to a chosen time delay. Any bit's phase shift, either up shift or down shift, that exceeds the chosen time delay, is counted on a counter for a given number of transferred bits. For short settings, a single revolution or a single record is used. For long settings, hundreds or thousands of revolutions of the disc are used, depending on the accuracy and sensitivity of the test. Fig. 13.1 illustrates a typical setup. The capacity of the counter is a factor in the choice of total transferred bits for any given setting.

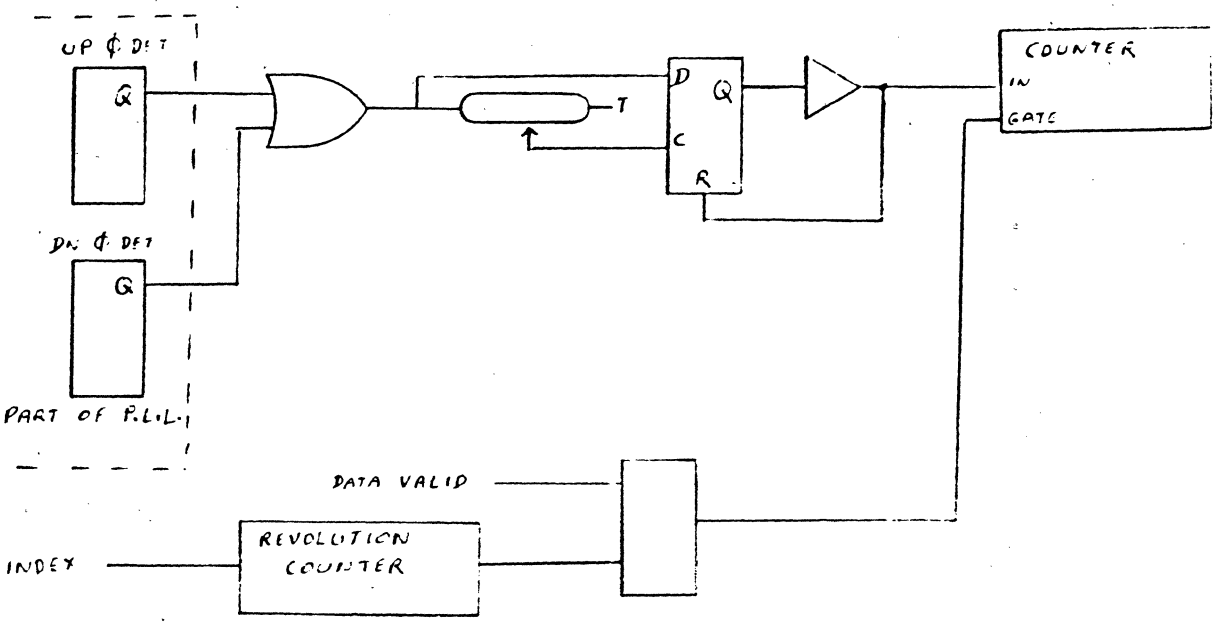


Fig 13.1

QUANTIFIED PHASE ERROR DETECTOR CIRCUIT

Notice that the early part of a record is blocked using "Data Valid." This is to ignore the excessive phase errors during lock up time at the beginning of the record format. Also the type of phase detector used determines the block designated as an "or" following the phase detectors. For simple phase detectors, or non-harmonic phase detectors, as shown in Fig. 11.14A, the circuit shown is adequate. Care must be taken in allowing for "D" setup time in the output block. For harmonic phase detectors, where the up and down errors are turned on simultaneously, an "exclusive or" must replace the "or." Then the difference between the two is fed to the delay line and "D" flip flop. The operation of the two versions is illustrated in Fig. 13.2a and 13.2b.

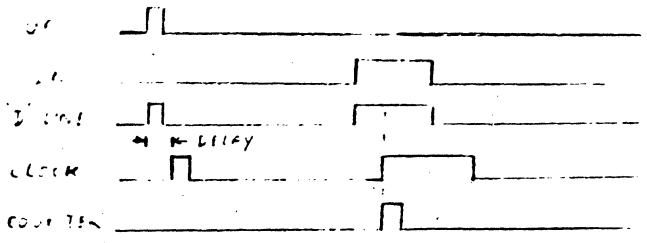


Fig.13.2a
Typical Non-Harmonic Ø Error Discriminator Waveforms

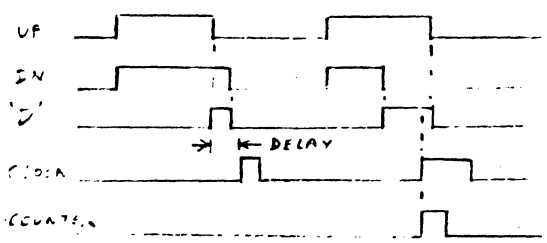


Fig. 13.2b
Typical Harmonic Ø Error Discriminator Waveforms

A typical plot can be made as a function of the total number of pulses out divided by the total number of transitions recorded vs. the delay line setting.

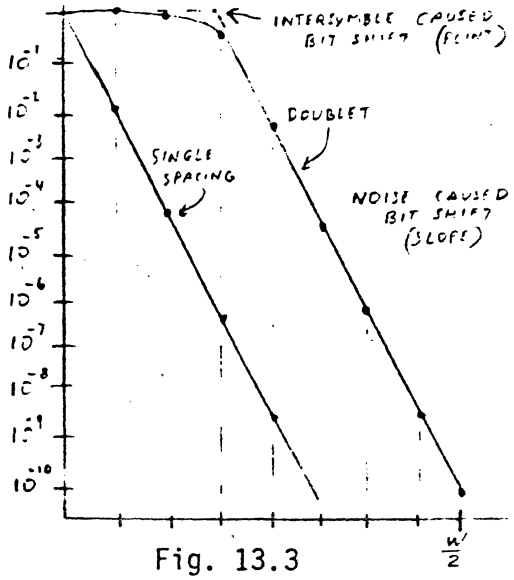


Fig. 13.3
Typical QPEC for Repetitive Doublets

Obviously at the 0_{ns} setting, almost all transitions recorded exceed the setting. There are several features of this curve that depend on the bit pattern, the resolution, and the system noise. It is usual to use the bit pattern that provides the greatest amount of

PULSE INTERFERENCE OR BIT SHIFT. This is the repetitive doublet pattern where two transitions follow at the minimum spacing with at least one zero between the doublets. Usually a single zero or non-transition suffices, depending on the pulse interaction. When counting the number of bits transferred, one must only count the transitions and relate those through the code used. For example, the MFM code is designated as ½(1,3). The maximum doublet pattern would give transitions for every 3 bits transferred: $\overset{123}{011}\,\overset{123}{011}\,0$. The actual cell content on the disc would be c 0c1c1c 0c, or 2 transitions for 6 cells. Since there are 2 cells per bit transferred, the 6 is divided by 2 to give the 3 for 2/3 track capacity; or for 10^5 bits track capacity, we will have 6.66×10^4 transitions recorded. For a different code, such as the 2/3 (1,7) code, we take two data bits and occupy 3 cells. There must be one vacant cell between transitions, therefore, to get the minimum doublet patterns we must have four data bits transferred per two transitions: 101 000 101 000 for ½ track capacity. For example, if the track held 10^5 bits of data, there would only be 5×10^4 transitions recorded. If we look at Fig. 13.3 again we see that the line for the single frequency is a single slope, whereas in the doublet case it is forced over but the slope is the same. What we see here is the predicted bit shift of the doublet. The corner of the curve is located at the superposition caused bit shift value. The slope is a direct function of the channel signal to noise ratio, meaning the head, electronics, and disc noise plus signal to the head, electronics, and disc noise. Notice that such a plot is totally representative of the entire recording channel and, hence, becomes a measure of the error performance of that head disc combination. The intersection of the curve at the time value equal to the one half of the cell window width, $W_{½}$, gives the error performance. In the figure, that is 10^{-10} for the maximum doublet, or worse case pattern. In order to save time,

10^{10} bits need not be transferred. The slope of the curve can just be extended from some lesser value to the 10^{-10} level as the curve in this region is a straight line.

Now we have a tool for measuring the performance of a channel we need only find the worst head and the worst media acceptable under the specifications, plot their QPEC curve to obtain the minimum machine performance for on track conditions. Similarly, we could do the same for the off track conditions at the normal 5 psycometric corners of temperature and humidity and power supply variations to predict the worst machine performance. Obviously, if the results are unsatisfactory, then the specifications need to be tightened for some parameter or component until the performance specification is met. This is much simpler than in the past.

While we are discussing specifications, we need to discuss the head and disc/media . Of the many head parameters that can be measured several stand out as being meaningful. These were all discussed in Chapter 4. Of the electrical parameters, amplitudes of the frequency extremes, resolution, and write over are the most significant. It is intersting to note that as amplitude increases, poor resolution can be accepted for the same error performance. This is predicted from the relationships of the QPEC curve. The better amplitude results in a better S/N ratio which means a steeper slope, while a poorer resolution results in a further shift of the corner to the right. With the intercept point fixed at 10^{-10} and W/2, then various combinations of amplitude and resolution can be accepted. This is taken advantage of in the head specification to allow an increase in head electrical test yield, knowing that these combinations will function well. An example of a head specification curve relating amplitude and resolution is shown in Fig. 13.4.

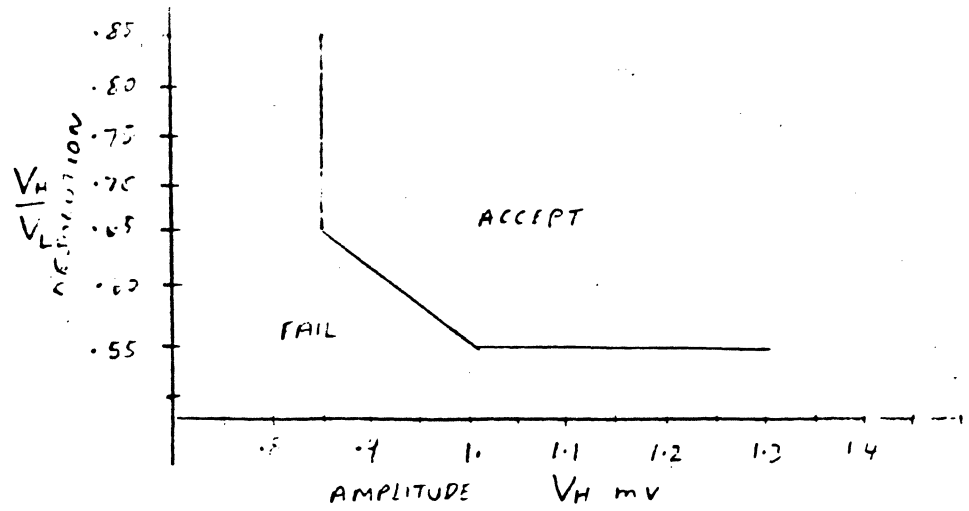


Fig. 13.4 Acceptable Amplitude and Resolution Combinations

It is desirable to have the recording channel cost effective. The design then becomes a compromise between bit shift or resolution controlling parameters and total S/N ratios. Obviously bit shift is controlled by head and disc parameters and S/N is controlled by head-electronics and disc parameters. Therefore a successful and cost effective design hinges on balancing the cost increases necessary to reduce bit shift and those necessary to reduce noise. Since heads and discs are in both camps, then it is probable that equal window spacing be given to bit shift and noise. If the design is based on one third bit shift, one third noise, and one third allowed for manufacturing variables and degradation during machine life, then a satisfactory arrangement has been reached. See Fig. 13.5

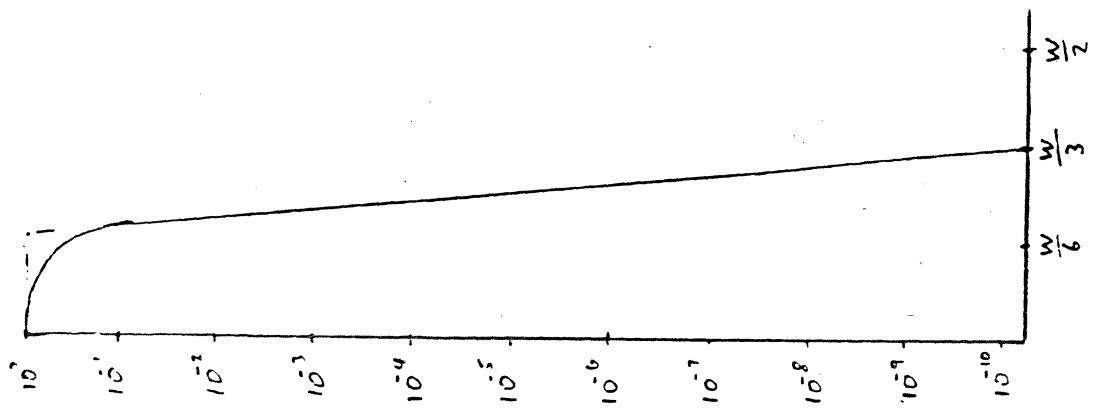


FIG 13.5

There are several variations in the QPEC curve that should be discussed. These relate to defects and/or anomalies of the disc/media surface. In Fig. 13.6 a QPEC curve is shown that illustrates the effect of a small agglomerate

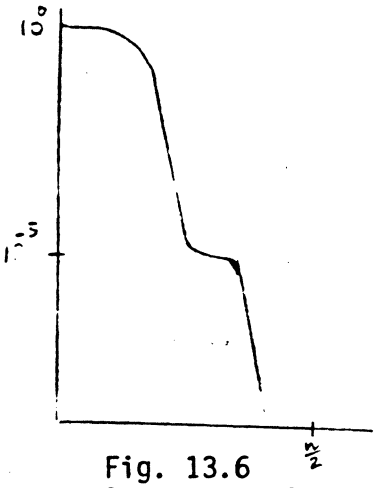


Fig. 13.6
QPEC Curve Showing a
Media Anomaly

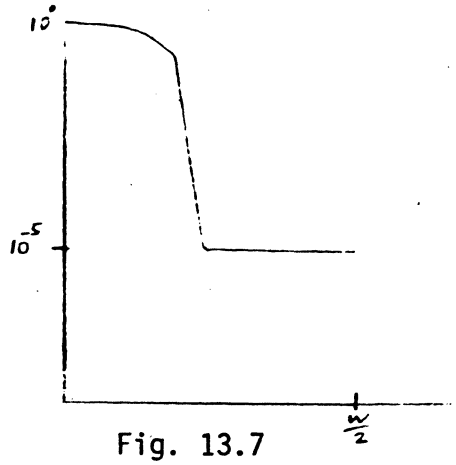


Fig. 13.7
QPEC Curve Showing a
Media Defect

in the media. This causes a bit or a few bits to have a different local bit shift than the remainder of the track, but the S/N ratio remains the same, hence the curve continues down at the same slope. The feature of Fig. 13.7 is caused by a scratch in the media in proximity to a recorded transition. The scratch causes a dQ/dt signal, called an extra bit or drop in. The pulse adds to an existing transition and phase shifts the transition beyond the W/2 limit, thus causing the curve to extend to the right at one count per revolution or more. The variation shown in Fig. 13.6 could also be caused by this mechanism if the shift is small.

During manufacturing testing of a large number of machines, the QPEC curve is not cost or time effective. There are other techniques that permit definitive testing. It has been customary to do several things, parameter or circuit wise, to the channel to remove some of the margin built into the

channel. This is done by altering a circuit or circuits that are not normally part of the machine, but are a later part of the recording channel or it may be accomplished by altering a part of the machine that can be independently tested. Candidates for these are primarily the clocking circuit and, secondarily, the detector. The clocking circuit is best altered by only changing the width of the window used for gating transitions. For example, for FM. PM, or MFM codes the clocking window is a squarewave. One half for data and the other half for redundant clocks. By using a delay line and an 'and' gate (Fig. 13.8),

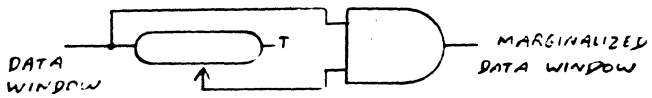


Figure 13.8a

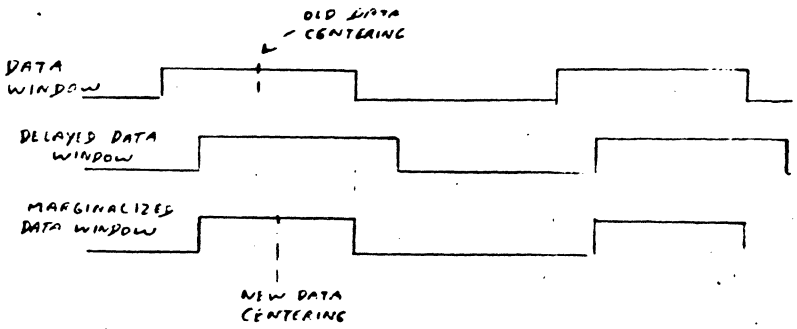


Fig. 13.8b

and then realigning the window to center it to an on time transition, the DATA window can be marginalized. Hence, when operating, transitions with less than a given amount of ± shift are passed and those exceeding that value are lost and cause an error. The error checking is done on the record itself. This is the source of the name Marginalized VFO as MVFO. The reader can now see the difference between the two circuits and functions hence our renaming the QPEC curve. The two functions are different. The QPEC circuit counts all ~~functions~~ TRANSITIONS that have phase shifts greater than many settings to generate a curve while the MVFO circuit blocks all transitions with phase shifts

ater than a single setting thus causing an error in a data channel.
o different functions for two different results. Of course, the QPEC cir-
it could be used at a single setting to get an error indication, but as the
annel also provides record position as byte count information of the error location
t would require extra circuitry if the QPEC circuit were used for defect
ogging as well as just a channel functionality test.

At this point it might be well to point out a method devised by Mr. F. Sordello that graphically predicts the performance of a recording channel if an isolated pulse can be measured for Pw_{50} and the channel S/N ratio is known. The method uses an Arc Tangent pulse drawn on a sheet of paper with a normalized amplitude and time scale.A. Drawn on the same sheet and centered is the differentiated pulse shown to scale^B and again at 10 X horizontal scale^C similarly centered.

P.C. a separate sheet

Fig. 13.9

Using this sheet and a table of noise amplitude probabilities, both intersymbol interference caused bit shift and noise caused bit shift can be predicted. The first is done by algebraically^D adding the contribution of a second inverted Arc Tangent pulse (Fig. 13.10),

$$e_o = \frac{1}{1+\left(\frac{t}{\frac{t_{50}}{2}}\right)^2} \; ; \; \frac{de_o}{dt} = -\left(\frac{1}{t_{50}}\right)\left(\frac{4\frac{t}{\frac{t_{50}}{2}}}{\left[1+\left(\frac{t}{\frac{t_{50}}{2}}\right)^2\right]^2}\right)$$

$\frac{d^2}{dt}$ "ARROW" PULSE

Fiducials 4/3/80

∓	6.0	± 0.029586	∓ 16.0 ± 0.00076
∓	6.0	± 0.01758	
∓	7.0	± 0.0112	∓ 18.0 ± 0.00068
∓	8.0	± 0.007574	
∓	9.0	± 0.005354	∓ 20.0 ± 0.00041
∓	10.0	± 0.003921	
∓	12.0	± 0.002283	
∓	14.0	± 0.001443	

$\frac{t}{\frac{t_{50}}{2}}$	$t_{50}\frac{de_o}{dt}$
-4.0	0.05536
-3.4	0.09621
-3.0	0.1200
-2.4	0.2101
-2.0	0.3200
-1.4	0.6392
-1.8	0.4005
-1.6	0.5050
-1.2	0.8062
-1.0	1.0000
-0.9	1.099
-0.8	1.190
-0.7	1.261
-0.6	1.298
-0.55	1.297
-0.50	1.280
-0.40	1.184
-0.30	1.010
-0.20	0.7396
-0.10	0.3421
-0.05	0.1490

HORIZONTAL SCALE (TIME) EXPANDED BY 10X

see separate sheet

see separate sheet

Fig. 13.10

Fig. 13.11

placed at the correct spacing of $\frac{t_s}{\frac{P_{W50}}{2}}$ then doing the same to the differentiated

Arc Tangent pulse. The peak amplitude reduction is shown on the first curve. The intersymbol interference caused bit shift is shown as the difference in time between the original differentiated pulse zero base line crossing and that of the second or modified differentiated zero base line crossing.

The contribution of bit shift from noise can be determined using a curve relating RMS noise amplitude to a gausian distribution.

The gausian distribution is $P(X) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{V^2}{2\sigma^2}}$ EQ 13.1

From this equation we can generate a table of noise amplitude as a function of probability of occurance.

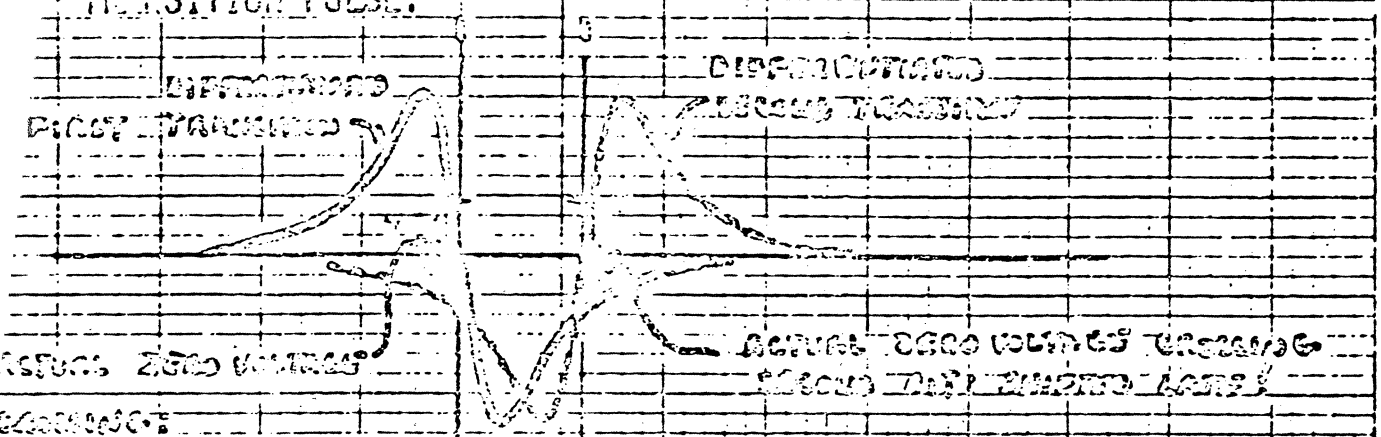
(See MISCHA - SCHWARTZ - McGRAW HILL, 1959, Page 373-390)

CALCULATION OF PATTERN INDUCED BIT SHIFT

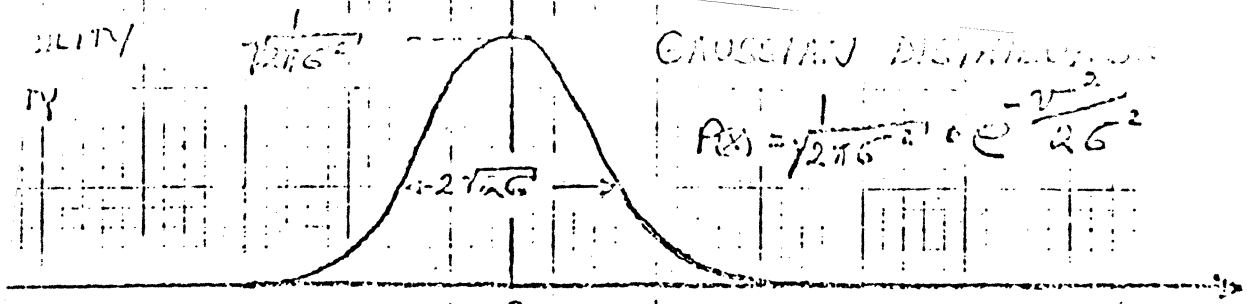
DETERMINE WORST BIT PATTERN FOR BIT SHIFT. GENERALLY IT IS A PATTERN THAT HAS MAXIMUM SPACING, THEN 2 TRANSITIONS AT MINIMUM SPACING, THEN MAXIMUM SPACING AGAIN.

SINCE THE THRESHOLD DETECTOR IN THE READ CHANNEL LOOKS FOR THE ZERO VOLTAGE CROSSING OF THE DIFFERENTIATED PULSES, THE BIT SHIFT IS DUE TO THE COMPOSITE SIGNAL'S ACTUAL ZERO VOLTAGE CROSSING.

USING THE GRAPH OF THE DIFFERENTIATED ISOLATED PULSE, ADD ALGEBRAICALLY ALL OF THE CONTRIBUTIONS OF EACH DIFFERENTIATED TRANSITION PULSE.



FROM BIT SHIFT IS [illegible] NANOSECONDS.



σ = STANDARD DEVIATION OF GAUSSIAN AMPLITUDE FUNCTION

FOR NOISE . . .

σ RELATES TO THE RMS VALUE!

FOR MORE READ Pg 373 -390

INFORMATION, TRANSMISSION, MODULATION, AND NOISE - MISCHA SCHWARTZ - McGRAW HILL - 1959

RMS NOISE = σ	PROBABILITY OF OCCURANCE NOISE WILL
0	1.0
1	1/0.32
2	1/21.74
3	1/370.4
4	1/15,625
5	1/1,724,137.9
→ 6	1/500,000,000
6.36	1/10,000,000,000
7	1/400,000,000,000
8	1/80,000. TRILLION

F. J. Sedlak
6/80

RMS NOISE =	; Probability of Occurance
0	1
1	1/0.32
2	1/21.74
3	1/370.4
4	1/15625
5	1/1724137.9
6	1/500 000 000
6.36	1/10 000 000 000
7	1/400 000 000 000
8	1/80 000 000 000 000

Table 13.1

For 10^{-10} probability of occurance, the noise pulse amplitude is

6.36 X RMS Noise, EQ 13.2

The method of measuring the signal to noise ratio is important. The measurements are taken after the differentiator in order to relate it directly to the differentiated pulse for peak shift measurement purposes.

$$\frac{dS}{dN} = \frac{\frac{d\,V\,sig\ RMS}{dt}}{\frac{d\,Vn\ RMS}{dt}} \qquad \text{EQ 13.3}$$

where $\frac{d\,V\,sig}{dt}$ RMS is the RMS value of a differentiated signal resulting from evenly spaced transitions at the minimum spacing allowed by the code. In disc drives this also means at the inside track. $\frac{d\,Vn\ RMS}{dt}$ is the RMS value

of the differentiated noise including electronic and disc/media noise at the same location.

The peak value of noise voltage becomes

$$(6.36)(Vn\ RMS) \qquad \text{EQ 13.4}$$

or in terms of the $\frac{ds}{dn}$ ratio and converting the RMS sig voltage to base to peak at the same time we get

$$Vn_{Peak} = \frac{6.36}{\sqrt{2}\ \frac{ds}{dn}} = \frac{4.497}{\frac{ds}{dn}} \qquad \text{EQ 13.5}$$

which is the peak noise expressed as a fraction of the signal peak.

When we used the d(RMS signal) value to get our d(S/N) ratio, we are really in error when we use this value with a differentiated isolated pulse as the amplitude of the isolated pulse is greater. All that occurs is an error in favor of poorer performance which is acceptable.

Going back to our graph we now locate this amplitude fraction on the expanded scale (fig 13.12)

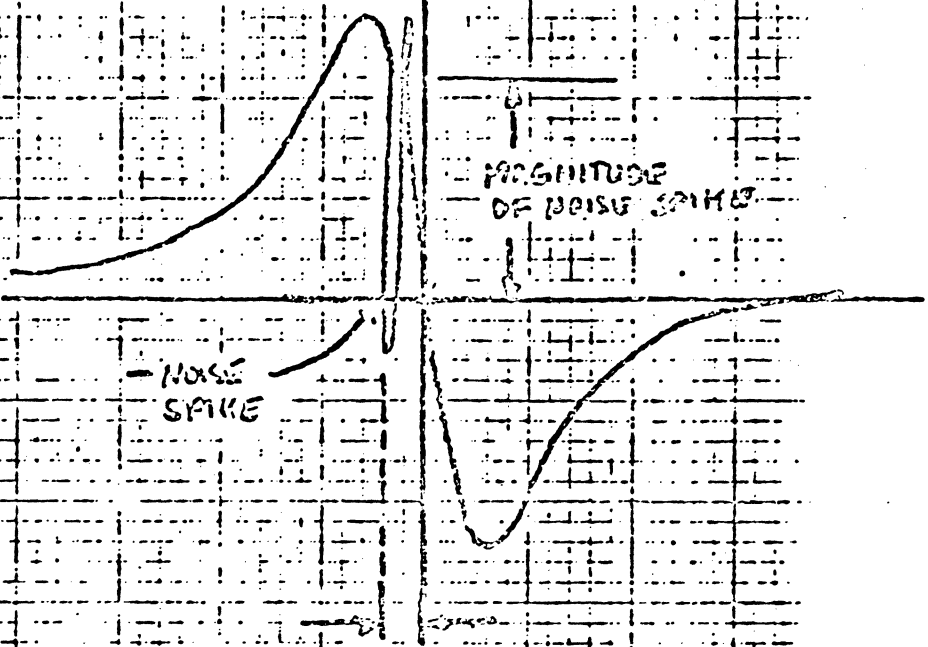
see separate sheet

Fig. 13.12

differentiated pulse. The value of the peak shift can be read off on the horizontal scale and multiplied by $\frac{Pw_{50}}{2}$ which converts it into bit shift in seconds

WHAT IS THE MAGNITUDE OF A NOISE SPIKE THAT OCCURS EVERY 10^{10} TIMES?

$\sigma = 6.36$



∴ NOISE SPIKE AMPLITUDE IS 6.36 X RMS NOISE

BASE-PEAK SNR = $\sqrt{2}$ · RMS SNR = $1.414 \cdot 20 = 28.28:1$

∴ FOR NOISE SPIKE @ $\sigma = 6.36$

NOISE BECOMES = $\frac{6.36}{28.28} = 0.225$ OF SIGNAL

USING ISOLATED DIFFERENTIATED TRANSITION GRAPH

GRAPH SHOWS SIGNAL TO BE ± 130 UNITS

∴ $0.225 \cdot 260$ UNITS = 58.5 UNITS

ON EXPANDED SCALE = $.155 \cdot \frac{650}{2} = 0.155 \cdot \frac{110}{2}$

= 8.525 NANO SEC.

Now to get the total bit shift due to intersymbol interference and noise, we just add the two values. According to our earlier rule of thumb each of these two values should be about one third of the half window width each which predicts performance at the required error rate. (including a margin of one third half window.)

The Memorex Disk Drive Family

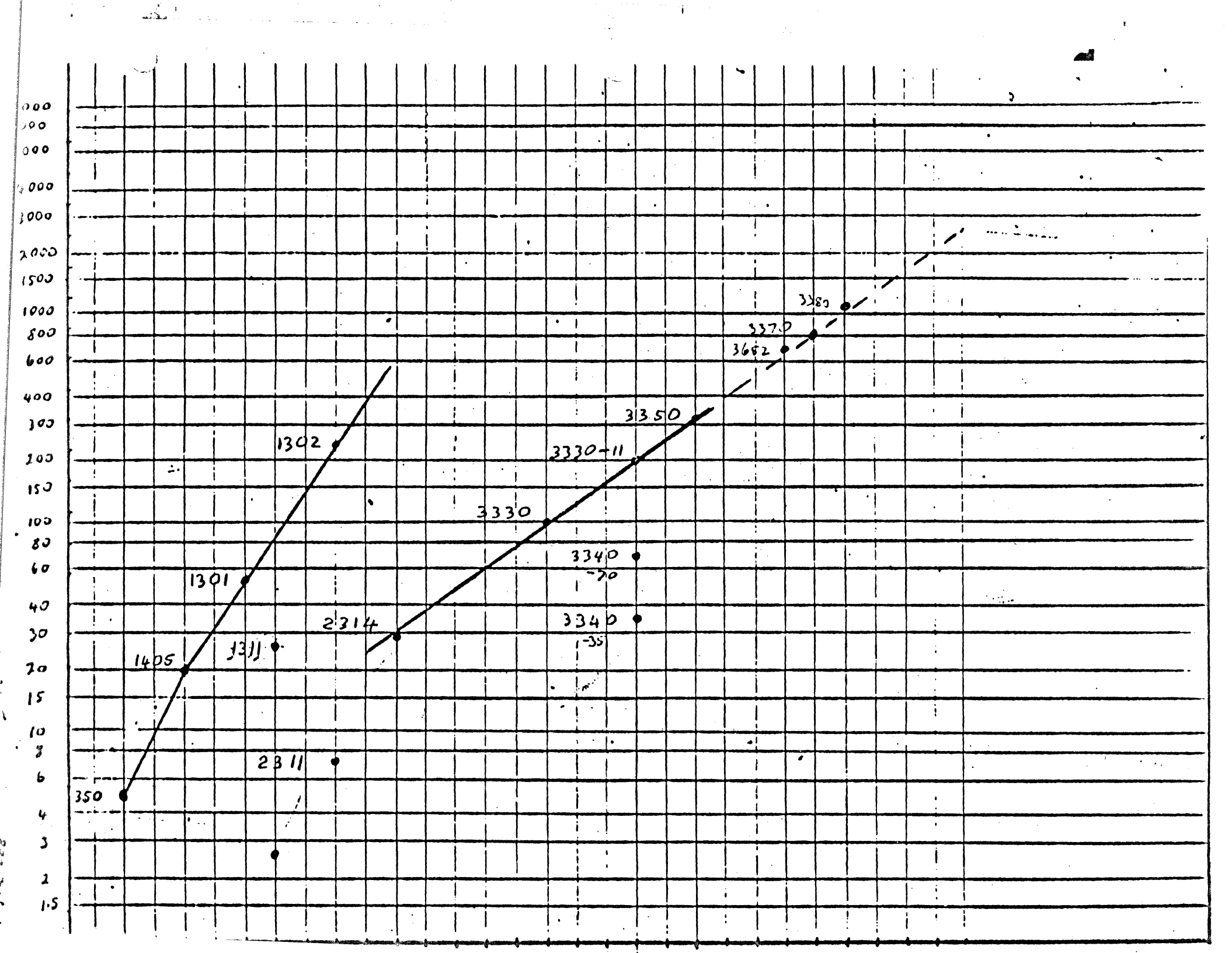
DRIVE #	INTERFACE	CONTR UNIT	REQUIRE EXTERNAL VFO PLO	ROT (RPM) SPEED	DATA TR. RATE (KILOBYTES)	REVOLUTION TIME	NO. OF CYLINDERS	# REC. SURFACES	FIXED HEADS		MO. ADS		BYTES/ TRACK	TRACK DENSITY	MAX BIT DENSITY	SEEK TIME		INTERFACE LEVELS		DATA ERROR RATE	SEEK ERROR RATE	BIT CELL TIME
									QTY.	TOTAL CAPACITY	QTY.	TOTAL CAPACITY				MINIMUM	AVERAGE	CONTROL	DATA/CLK			
660	IBM 2314	MRX 661	YES	2400	312 KB/S	25 MS	203	20			20	29 MB	7,250	100 TPI	2228 BPI	12 MS	35 MS	-2.5V +1.5V	0V +3V	10^9 SOFT 10^{12} HARD	10^9 RANDOM	400 NS
3670	IBM 3330-1	MRX 3673	YES	3600	806 KB/S	16.67 MS	404	19			19	100 MB PER SPINDLE	13,030	192 TPI	4040 BPI	5 MS	27 MS	0V 1.7V (8T23/24)	0V 1VT	10^9 SOFT 10^{12} HARD	10^9 RANDOM	155 NS
3675	IBM 3330-2	MRX 3673	YES	3600	806 KB/S	16.67 MS	808	19			19	200 MB PER SPINDLE	13,030	370 TPI	4040 BPI	5 MS	27 MS	0V 1.7V (8T23/24)	0V 1V	10^9 SOFT 10^{12} HARD	10^9 RANDOM	155 NS
677-30	CDC 9766		NO	3600	1,209 KB/S	16.67 MS	823	19			19	309.5 MB	20,160	384 TPI	6060 BPI	6 MS	38.33			10^9 SOFT 10^{12} HARD	10^9 RANDOM	103 NS
677-0	CDC 9760	TELEFILE	NO	3600	806 KB/S	16.67 MS	411 815	19			19	100 MB 200 MB	13,440	192 TPI 370 TPI	4040 BPI	6 MS	28.5 MS	75107 75110	75108 75110	10^9 SOFT 10^{12} HARD	10^9 RANDOM	155 NS
677-1	DEC-DCL	DEC-DCL	NO	3600	806 KB/S	16.67 MS	411 815	19			19	100 MB 200 MB	13,440	192 TPI 370 TPI	4040 BPI	6 MS	28.5 MS	7404 7436	75108 75113	10^9 SOFT 10^{12} HARD	10^9 RANDOM	155 NS
677-2	IBM 3330-2	ISSI	YES	3600	806 KB/S	16.67 MS	411 815	19			19	100 MB 200 MB	13,440	192 TPI 370 TPI	4040 BPI	6 MS	28.5 MS	0V 1.7V (8T23/24)	-1.6V -0.9V	10^9 SOFT 10^{12} HARD	10^9 RANDOM	155 NS
612	CDC 9760		NO	3600	1,209 KB/S	16.67 MS	350	2 4 6			4 8 12	28 MB 57 MB 85 MB	20,160	300 TPI	5636 BPI	7 MS	32 MS	75107 75110	75107 75110	10^{10} SOFT 10^{12} HARD	10^9 RANDOM	141 NS
601	CDC 9760 +FIXED HEADS	MSC 1000	NO	2964	885 KB/S	20.24 MS	350	7	30 60	0.5 MB 1.0 MB	4 8 12	25 MB 50 MB 75 MB	17,290	300 TPI	5636 BPI	7 MS	32 MS	75107 75110	75107 75110	10^{10} SOFT 10^{12} HARD	10^9 RANDOM	141 NS
3640	IBM 3340-B2	IBM 3343 MRX 3643	YES	2964	885 KB/S	20.24 MS	349 699	3 7	30	0.5 MB	6 12	35 MB 70 MB	8,368			7 MS	20 MS	0V 1.7V (8T23/24)		10^{10} SOFT 10^{12} HARD	10^9 RANDOM	141 NS
3643	IBM 3340-A2	IBM 3830-2 ISC OR IFA	NO	2964	885 KB/S	20.24 MS	349 699	3 7	30	0.5 MB	6 12	35 MB 70 MB	8,368			7 MS	20 MS	0V 1.7V (8T23/24)	0V 1.7V (8T23/24)	10^{10} SOFT 10^{12} HARD	10^9 RANDOM	141 NS
3650	IBM 3350-B2	IBM 3350-A2 MRX 3653	YES	3600	1198 KB/S	16.6 MS	555	15	60	1.14 MB	30	317.5 MB PER SPINDLE	19,069	480 TPI	6350 BPI	10 MS	25 MS	0V 1.7V (8T23/24)	0V -0.6V	10^9 CORRECT 10^{10} RECOVER	10^9 RANDOM	104 NS
3653	IBM 3350-A2	IBM 3830-2 ISC (370 MOD 45) MRX 3674	NO	3600	1198 KB/S	16.6 MS	555	15	60	1.14 MB	30	317.5 MB PER SPINDLE	19,069	480 TPI	6350 BPI	10 MS	25 MS	0V 1.7V (8T23/24)		10^9 CORRECT 10^{10} RECOVER	10^9 RANDOM	104 NS
3644	IBM 3344	IBM 3343 MRX 3643	NO	2964	885 KB/S	20.2 MS	2784	15	60	1.0 MB	30	280 MB PER SPINDLE	8,368	480 TPI	5840 BPI	10 MS	25 MS	0V 1.7V		10^9 CORRECT 10^{10} RECOVER	10^9 RANDOM	141 NS
3652	IBM	MRX 3655	YES	3600	1198 KB/S	16.6 MS	1110	15	120	2.26 MB PER SPINDLE	30	635 MB PER SPINDLE	19089	935 TPI	6350	6 MS	22 MS	0V 1.7V (8T23/24)		10^{10} SOFT 10^{12} HARD	10^9 RANDOM	104 NS

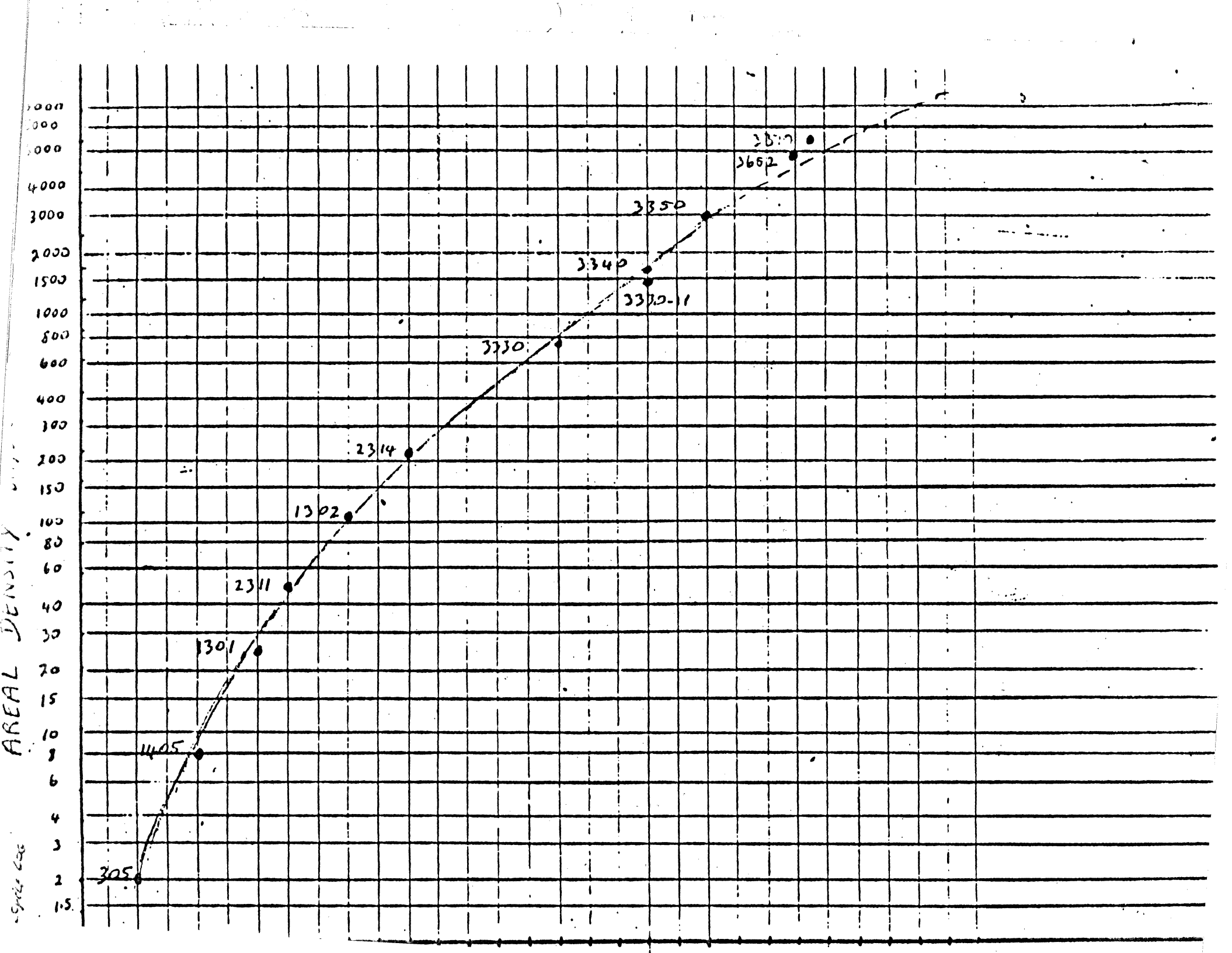
MACHINE	ANNOUNCED	IBM FCS	TPI	BPI	DENSITY	IR	OR	RPM	LATENCY	RATE BITS
350	1956	1957	20	100	2000			1200	25 MS	77.6 KB/S
1405		1959	40	200	8000			1200		155 KB/S
1301	1961	1961	50	500	25000			1800	33 MS	625 KB/S
1311		1962	50	1000	50000			1500	20 MS	700 KB/S
1302	1963		100	1100	110000			1800	33 MS	1.25 MB/S
230-1/2										
2311	4/64	1964	100	2200/1100	110000	4.468"	6.506	2400	12.5 MS	1.25 MB/S
2305										
2314	4/65	1966	100	4400/2200	220000	4.46	6.50	2400	12.5 MS	2.5 MB/S
2319		1969		4400/2200		4.46	6.50	2400	12.5 MS	2.5 MB/S
3330	6/70	1971	192	4040	775680	4.24	6.38	3600	8.4 MS	6.45 MB/S
3330-11	7/17/73	3/74	370	4040	1494800	4.24	6.44	3600	8.4 MS	6.45 MB/S
3340	3/74	11/73 (125) 3/74	300	5630	1690500	4.06	6.60	2964	10.1 MS	7.08 MB/
Sys 32			300	5635	1690500			2964	10.1 MS	7.08 MB/S
3344		4/76				4.69	5.863			
3350		4/76	480	6250	3000000			3600	8.4 MS	9.58 MB/
3370	1979	1/80	635	11900 TPI (7930)	7556500			2964	10.1 MS	14.827 MB/
3375	1980	10/81						2964	10.1 MS	14.827 MB
3380	1980							3600	8.4 MS	24.0 MB

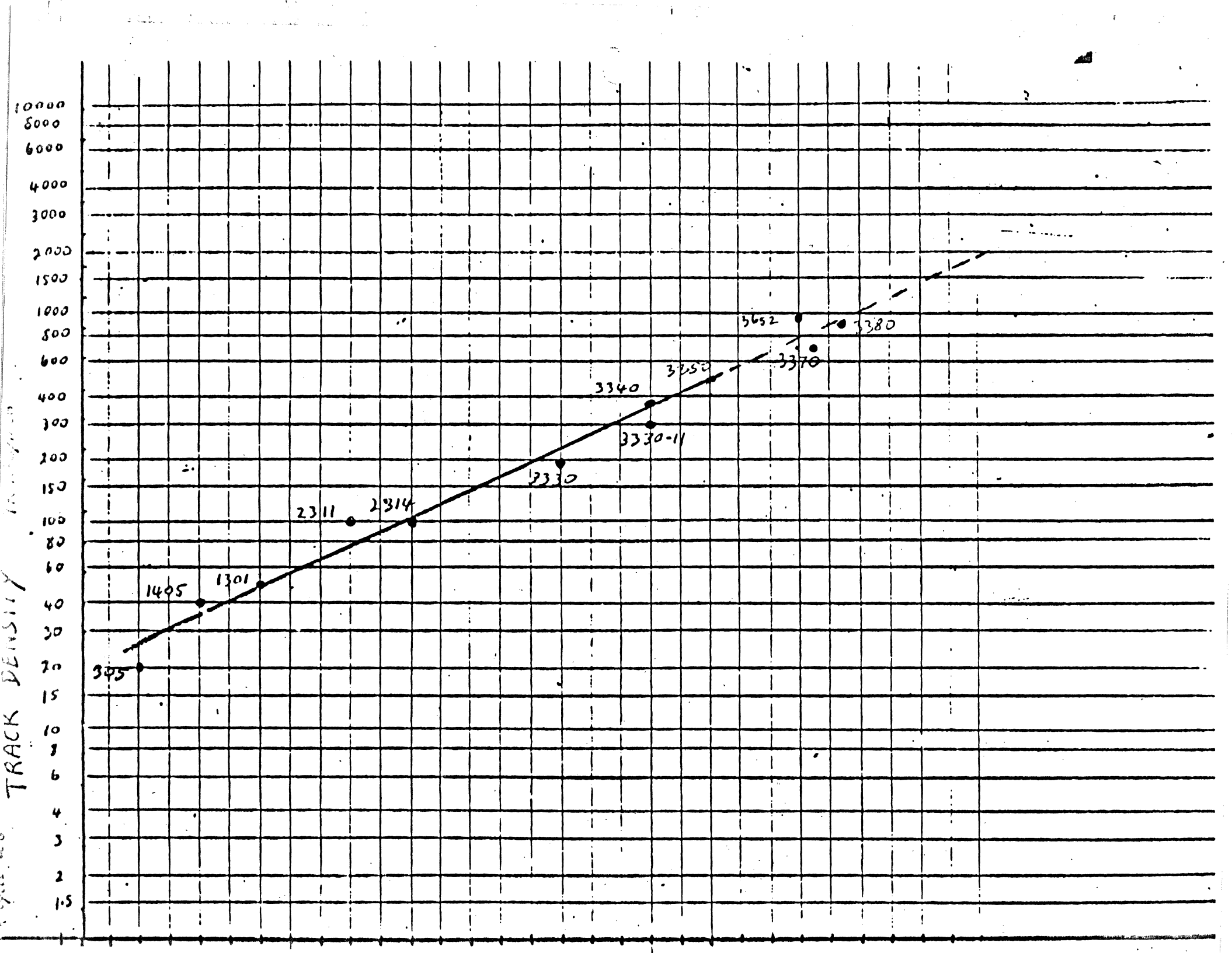
MACHINE	MECH	CODE	COMP	HEIGHT	ELEMENT	METHOD	CYLINDERS	CONTROL	GAP LENGTH	MRX PCS	TRSP	TW
350	Elec Servo	NRZ1	0	1	Yes	Clock Track	----	1.2	1000			
1405		NRZ1	0	500	Yes			.8				
1301	Dual Hydra	NRZ1	0	250	Yes	Clock Track		500				
1311	Hydraulic	NRZ1	0	125	Yes			200				
1302	Dual Hydra.	NRZ1	0	125	Yes	Clock Track		200				
230-1/2	Fixed	NRZ1	0									
2311	Hydraulic	FM	0	125	Yes	Hard Sep	200	200	200(M)	9/69		
2305		NRZ1	0			VFO						
2314	Hydraulic	FM	0	80	Yes	VFO	300	100	100(F)	12/69	10 Mill	7.0
2319	Hydraulic	FM	0		Yes	VFO	200				10 Mill	7.0
3330	Ser VC Lin	MFM	7	50	No	VFO	404	50	50(F)		5.2	4.3
3330-11	Ser VC Lin	MFM	7	35	No	VFO	808	50	50(F)		2.7	2.0
3340	Cou VC Lin	MFM	0	22	No	VFO	348/696	35	50(F)		3.34	2.6
Sys 32	VC Rot Lin	MFM	0	22	No	VFO			50(F)		3.34	2.6
3344						VFO			50(F)		2.08	1.45
3350	Voice Coil Lin Motor	MFM	5	20	No	VFO	555	35	50(F)	4/77	2.08	1.45
3370	L.V.C.	½(2,7)	0	15	No	VFO	746 +4 + 746 +4		25(TF)		1.57	1.33
3375	L.V.C.	½(2,7)	0		No	VFO			(TF)			
3380	L.V.C.	½(2,7)	0		No	VFO			(TF)			

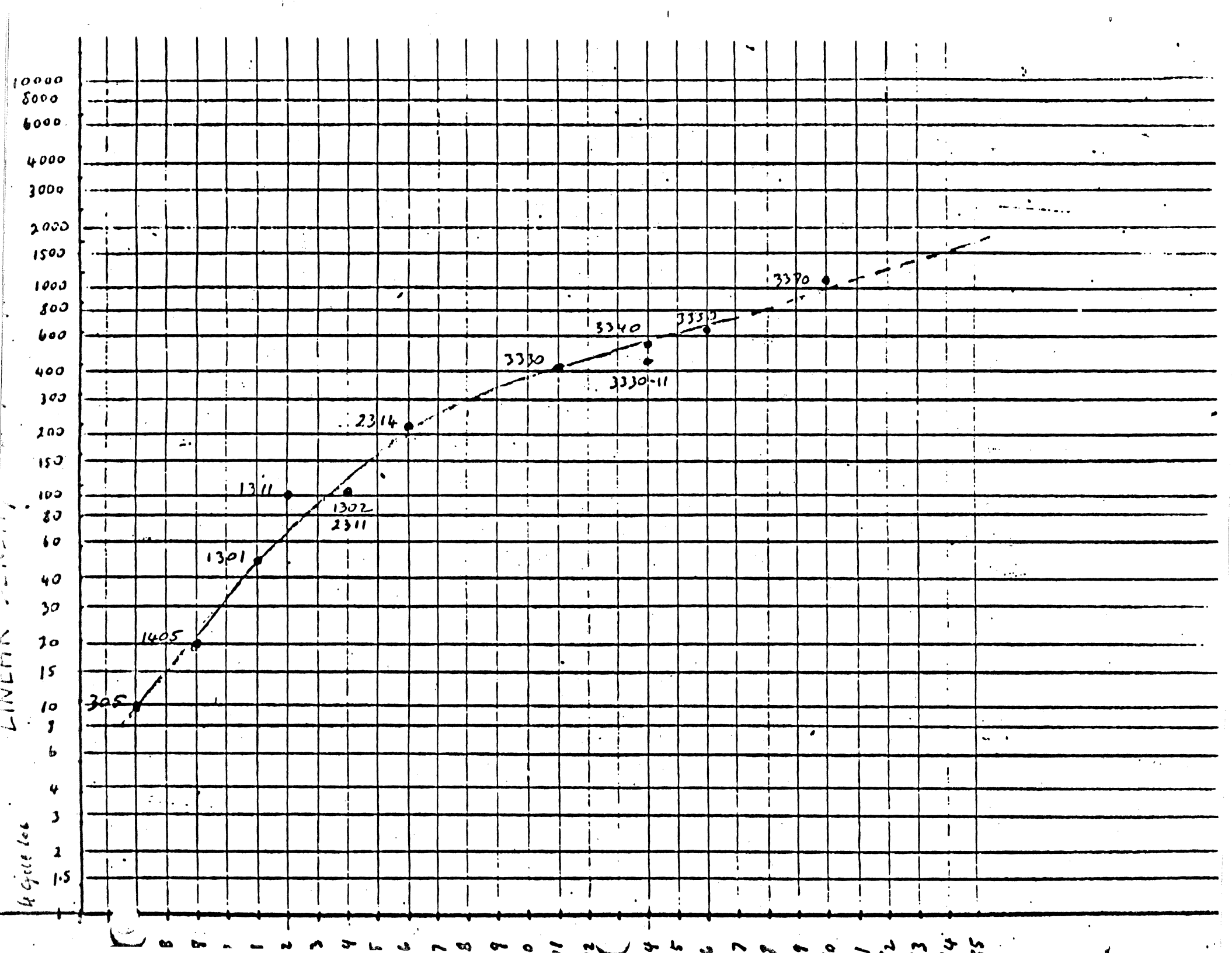
1409			800 MS			F						
1301			180 MS	48	50	F	24 + 24	56 MBy	2600			
1311			150 MS			R		3.65 MBy	3500			
1302				48	50	F	24 + 24	234 MBy	5208			
230-1/2						F						
2311	25	100	75 MS	10	10	R	10	7.25 MBy	3906	3675	3675	
2305						F						
2314	25	100	75 MS	20	20	R	20	29.3 MBy	7812	7200	7200	
2319			60 MS	20	20	R	20	29.3 MBy				
3330	10	55	30 MS	20	19	R	19	100 MBy	13440	13030	13030	
3330-11			30 MS	20	19	R	19	200 MBy	13440	13030	13030	
3340			25 MS	7/13 / 30	6	R	6-12 / 30	35/70 MBy	8960	8368	8368	
Sys 32						F						
3344	10	50	25 MS		30	F		280 MBy	8960	8368	8368	
3350	10	50	25 MS	31 / 60	30	F	30 / 60	317.5 MBy	19968	29069	19069	
3370			20 MS	26	6 R/W+½S/ACT	F	12/ACT	571.3 MBy	37632 R/W (25088)1.5Ser+	744 B1/Cyl 24 Alt/Cyl	512/B1	
3375			19 MS			F		819 MBy				
3380			16 MS					1260 MBy				

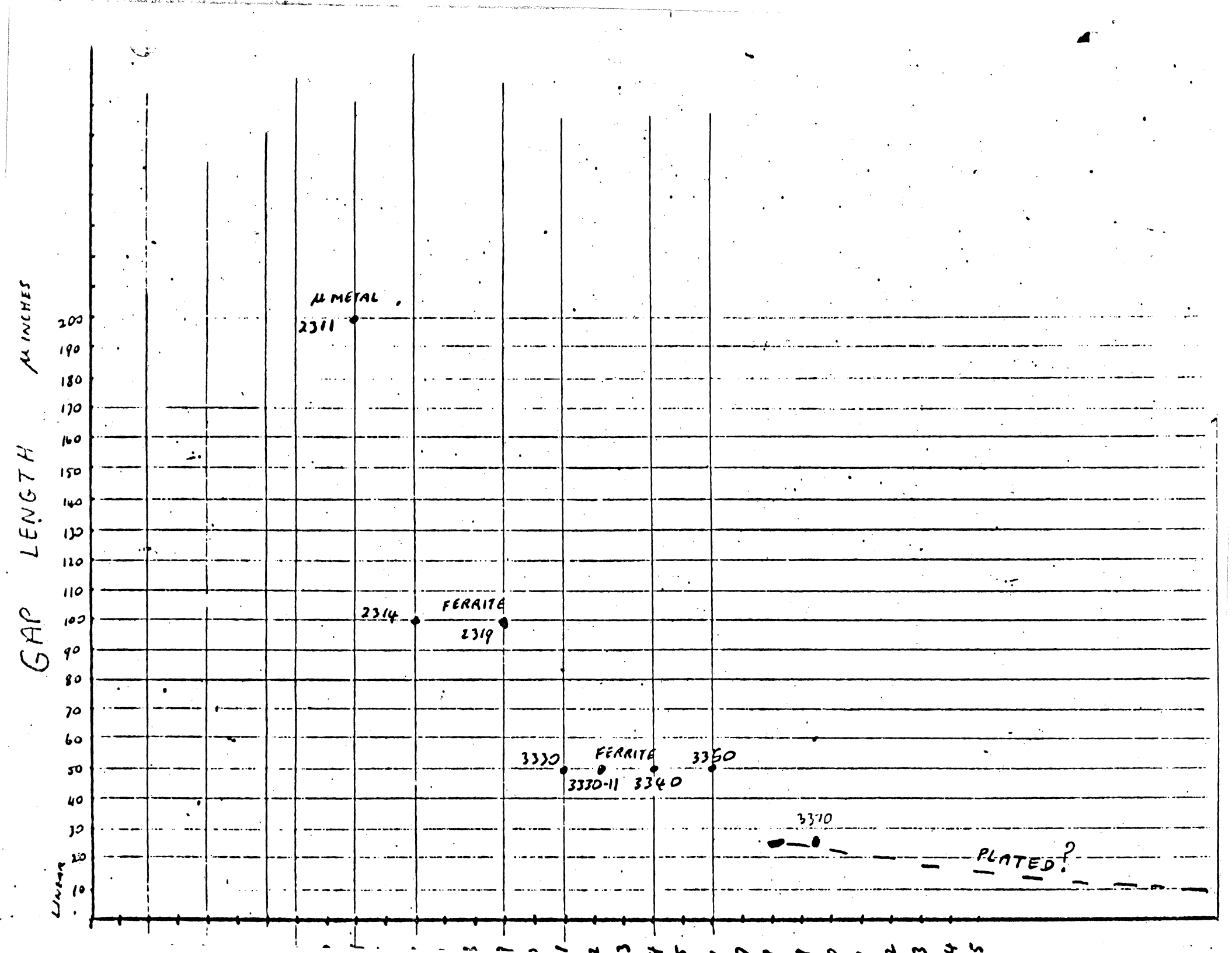
ITE	CLOCK	CODE	ECC	AVE RESOLUTION	ELECTRONIC	MACHINE
1956	CLOCK	NRZI	NO	1.0	CLIPPING	350
7						
8						
9	CLOCK	NRZI	NO	1.0	CLIPPING	1405
60						
1	CLOCK	NRZI	NO	1.0	CLIPPING	1301
2	CLOCK	NRZI	NO	1.0	CLIPPING	1311
3						
4	CLOCK	NRZI	NO	1.0	CLIPPING	1302
5	SELF	FM	NO	2.5	LIMIT	2311
6	SELF	FM	NO	2.5	LIMIT	2314
7						
8						
9						
70						
1	SELF	MFM	11 BIT	1.5	ΔV GATE	3330
2						
3						
4	SELF	MFM	11 BIT	1.5	GATE	3330-11
5	SELF	MFM	4 BIT	1.3	CLIPPING	3340
6	SELF	MFM	4 BIT	1.3	ΔV GATE	3350
7						
8						
9	SELF	½ (2,7)	16 BIT	1.1	CLIPPING	3370
80		SELF CLOCKING				
1		GROUP CODES				
2						
3						
4						
5						

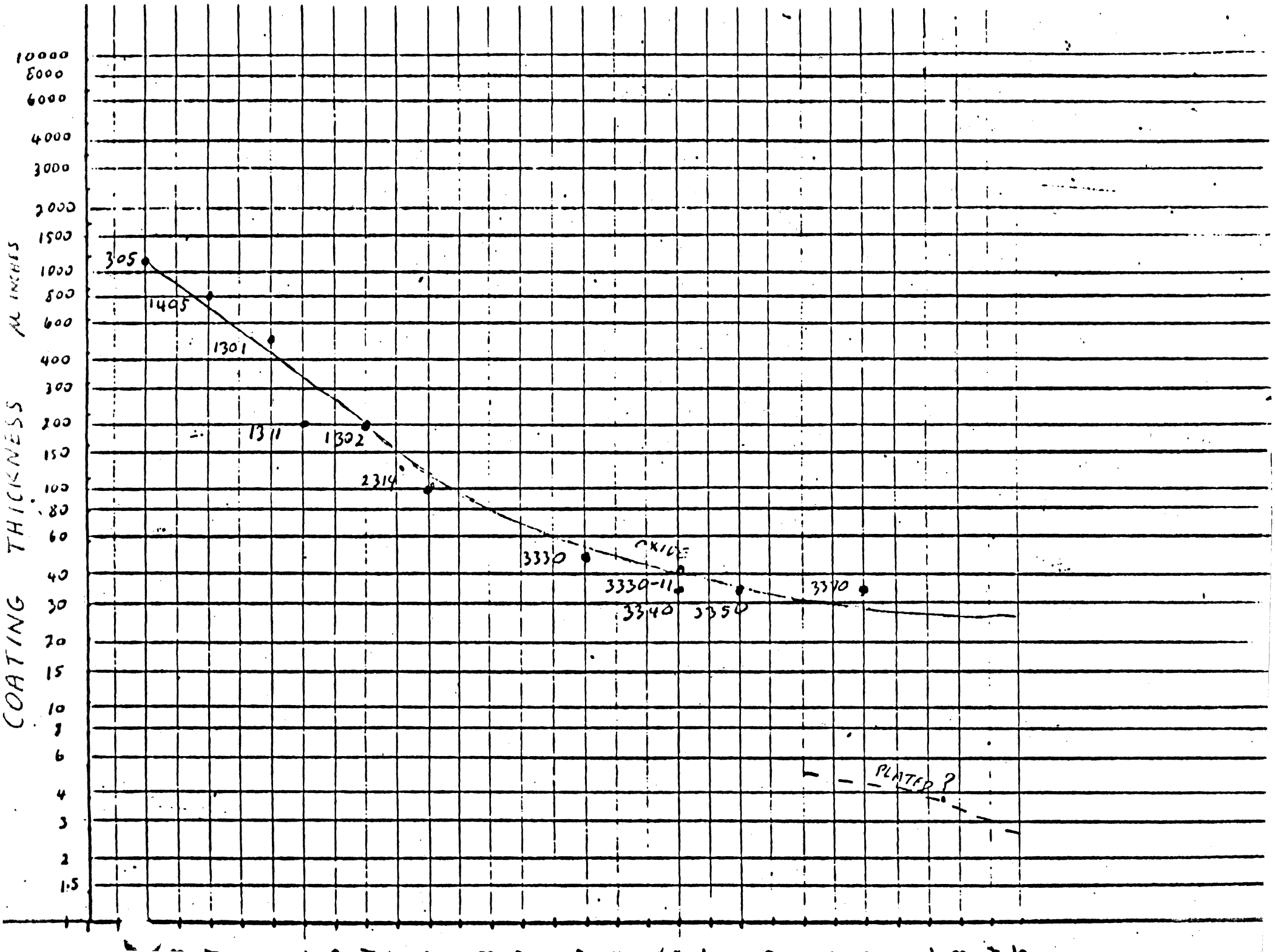

000
00
000
000
000
1000
2000
1500
1000
800
600
400
300
200
150
100
80
60
40
30
20
15
10
8
6
4
3
2
1.5
3380
3370
3682
3350
1302
3330-11
3330
3340
-20
1301
2314
3340
-35
1405
1311
2811
350

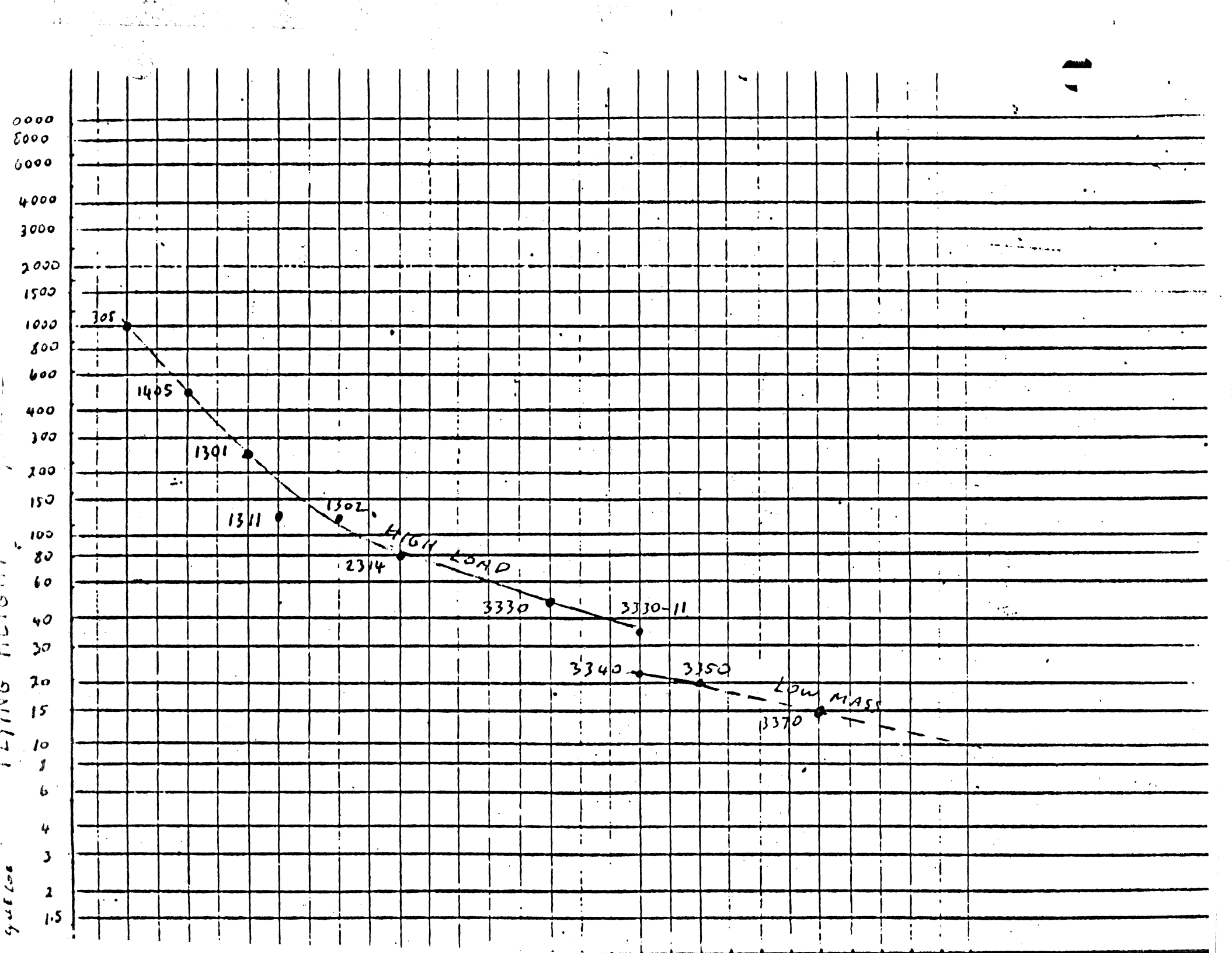

AREAL DENSITY
3000
2000
1500
1000
800
600
400
300
200
150
100
80
60
40
30
20
15
10
9
6
4
3
2
1.5
3305
1405
1301
2311
1302
2314
3330
3340
3350
3602

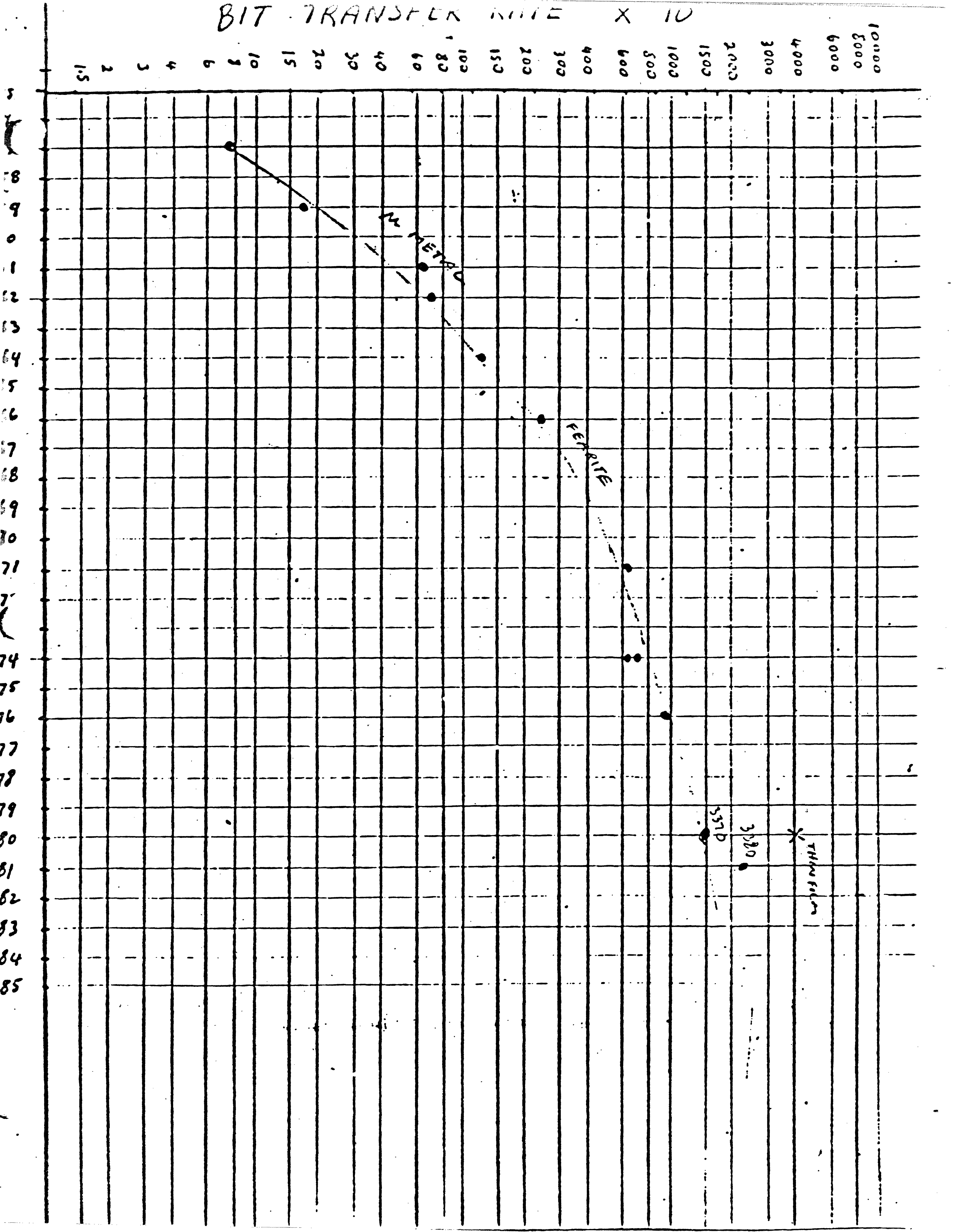

TRACK DENSITY
10000
8000
6000
4000
3000
2000
1500
1000
800
600
400
300
200
150
100
80
60
40
30
20
15
10
9
6
4
3
2
1.5
3652
3380
3370
3340
3330-11
3330
2311
2314
1405
1301
395


10000
8000
6000
4000
3000
2000
1500
1000
800
600
400
300
200
150
100
80
60
40
30
20
15
10
9
8
7
6
5
4
3
2
1.5
LINEAR
4 Cycle Log
305
1405
1301
1311
1302
2311
2314
3330
3330-11
3340
3352
3370


GAP LENGTH μ INCHES
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
μ METAL
2311
2314
FERRITE
2319
3330
FERRITE
3350
3330-11
3340
3370
PLATED?


COATING THICKNESS IN INCHES
10000
8000
6000
4000
3000
2000
1500
1000
800
600
400
300
200
150
100
80
60
40
30
20
15
10
8
6
4
3
2
1.5
305
1495
1301
1311
1302
2314
3330
3339-11
3340
3350
3340
PLATED


0000
5000
6000
4000
3000
2000
1500
1000
800
600
400
300
200
150
100
80
60
40
30
20
15
10
1
6
4
3
2
1.5
305
1405
1301
1311
1302
2314
LOW
LOW
3330
3330-11
3340
3350
LOW MASS
3370



BIT TRANSFER RATE X 10
1.5
2
3
4
6
8
10
15
20
30
40
60
80
100
150
200
300
400
600
800
1000
1500
2000
3000
4000
6000
8000
10000
M METAL
FERRITE
3330
3380
THIN FILM
78
79
80
81
82
83
84
85
74
75
76
77
78
79
80
81
82
83
84
85

the above model. Our conclusions are that the composition $Cu^{1+}_{0.8}Fe^{3+}_{2.2}O_4$ can exist as a stable compound when quenched from temperatures between 1210° and 1350°C in accordance with the published phase diagram.[6]

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. B. A. Calhoun for suggesting this problem and for his continued interest in the work. We thank Dr. E. L. Boyd for many helpful discussions. Our special thanks are to Professor F. D. Barros, Physics Department of Carnegie Mellon University, Pittsburgh, for making the helium runs. We also express our appreciation to Messrs. T. R. Klein, J. Angelelo, H. Lilienthal, and J. Jacobs for their help during the course of this work.

[1] F. Bertaut and C. Delorme, Compt. Rend. 236, 74 (1953).
[2] A. Bergstein, Int. Symp. Reactivity Solids, 6th, Schenectady, 1968 (in preprints).
[3] K. Ohbayashi and S. Iida, J. Phys. Soc. Japan 23, 776 (1967).
[4] E. Kordes and E. Röttig, Z. Anorg. Allgem. Chem. 264, 34 (1951).
[5] J. Théry and R. Collongues, Compt. Rend. 254, 685 (1962).
[6] S. Miyahara and Y. Kino, Jap. J. Appl. Phys. 4, 310 (1965).
[7] C. F. Jefferson, J. Appl. Phys. 36, 1165 (1965).
[8] E. W. Gorter, Advan. Phys. 6, 336 (1957).
[9] R. L. Cohen, P. G. McMullin, and G. K. Wertheim, Rev. Sci. Instrum. 34, 67 (1963).
[10] J. K. Lees, Ph.D. thesis, Carnegie Institute of Technology, Pittsburgh, 1966).
[11] B. J. Evans and S. S. Hafner, J. Phys. Chem. Solids 29, 1573 (1968).

JOURNAL OF APPLIED PHYSICS VOLUME 41, NUMBER 4 15 MARCH 1970

Analysis of Saturation Magnetic Recording Based on Arctangent Magnetization Transitions

ROBERT I. POTTER
International Business Machines Corporation, San Jose, California 95114

(Received 13 August 1969; in final form 13 October 1969)

The minimum transition length for the arctangent magnetization transition is calculated. Readback voltage for the arctangent transition, where the magnetization contains components both in and normal to the coating plane, is calculated via the reciprocity theorem and Karlquist's fringe field equations. Several misconceptions currently existing in the literature are discussed.

I. INTRODUCTION

A theoretical upper limit to the longitudinal data density attainable in conventional saturation magnetic recording exists, since current magnetic coatings are incapable of supporting abrupt transitions, and since the read process introduces an additional apparent broadening of the transition region. The term "saturation" is used here to describe the case where peak magnetization in the coating remains approximately equal to the remanent magnetization of the major hysteresis loop. The limitation on transition length arises in conventional coatings primarily[1] because the magnetization $\mathbf{M}(\mathbf{r})$ in the coating produces a demagnetizing field

$$\mathbf{H}(\mathbf{r}) = -\int[(\mathbf{r}-\mathbf{r}')\nabla\cdot\mathbf{M}(\mathbf{r}')/|\mathbf{r}-\mathbf{r}'|^3]dv', \quad (1)$$

which, being opposed to $\mathbf{M}$, cannot exceed the coercivity H_c of the hysteresis loop associated with each point in the medium. Since possible magnetization states must simultaneously satisfy Eq. (1) and the M-H relationships dictated by the hysteresis properties of the material, the recorded transition length must be calculated self-consistently.

Certain simplifications are necessary if the minimum transition length which a given coating will support is to be estimated without performing this self-consistent calculation. It is assumed that the hysteresis loops are square, and that the final state of magnetization $\mathbf{M}(\mathbf{r})$ after the external write field has vanished can be adequately described by an *assumed* functional form with adjustable parameters. These parameters are then chosen, after Chapman,[2] such that the demagnetizing field resulting from $\mathbf{M}(\mathbf{r})$ never exceeds the coercivity H_c of the major hysteresis loop. Two possible choices for $\mathbf{M}(\mathbf{r})$ are the ramp transition,

$$\begin{aligned} M_x(x) &= M_r && (-\infty < x \leq -\pi a/2), \\ &= -(2M_r/\pi)(x/a) && (-\pi a/2 < x < \pi a/2), \\ &= -M_r && (\pi a/2 \leq x < \infty), \end{aligned} \quad (2)$$

and the arctangent transition,

$$M_x(x) = -(2M_r/\pi)\tan^{-1}(x/a) \qquad (-\infty < x < \infty,\ a \geq 0), \quad (3)$$

where $M_y = M_z = 0$. x is measured longitudinally along the track, M_r is the remanent magnetization of the major hysteresis loop, and the parameter a describes the sharpness of the transition. These functions are normalized such that their derivatives at $x=0$ and limiting values at $x=\pm\infty$ are equal. The transition

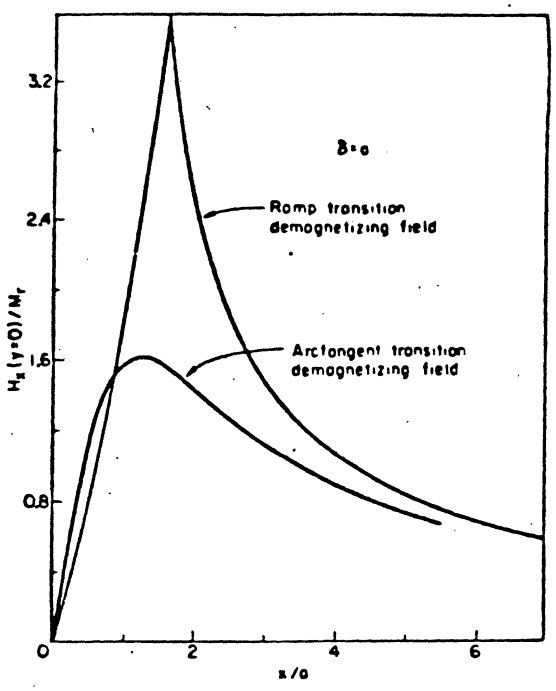


FIG. 1. A comparison of the demagnetizing fields for arctangent and ramp transitions of equal length.

length of the arctangent function as defined by the slope at $x=0$ is πa. The restriction $a \geq 0$, imposed for the sake of convenience in the calculations, does not result in any loss of generality since the sign of a can be absorbed in M_r. The step transition can be treated as a special case by letting a approach zero. The arctangent function is of special interest, because it is a simple function that has a continuous derivative, is antisymmetric, and asymptotically approaches a finite nonzero value. Our purpose here is to investigate further the consequences of assuming that **M** has this functional form, and to correct several misconceptions currently existing in the literature.

The external field due to an isolated arctangent transition was first correctly calculated by Miyata and Hartel.[3] This same[4] result was later obtained by Speliotis and Morrison.[5] Chapman,[2] in his analysis of the minimum transition length, used a modification of Miyata and Hartel's expression as the internal demagnetizing field, imposing the restriction $|a| > \delta/2$, where δ is the coating thickness. Bonyhard *et al.*,[6] give an asymptotic expression for the transition length, valid for $\delta/a \ll 1$ and based on the arctangent model. Treatments based on single and multiple ramp transitions have been given by Speliotis and Morrison[5] and Aharoni,[7] respectively.

We point out that Miyata and Hartel's expression for the field intensity (even as modified by Chapman) is not valid inside the coating, and hence Chapman's calculation of the minimum transition length is incorrect. The difficulty with Miyata and Hartel's expression inside the coating ($y<\delta/2$) is evidenced by lack of the correct symmetry. We give an expression for the field intensity that is valid both inside and outside the coating, and use this result to obtain the minimum transition length in terms of M_r, H_c, and δ.

An expression for the readback voltage resulting from an arctangent transition with the magnetization in the coating plane has been obtained by Speliotis and Morrison[5] and Speliotis.[8] Speliotis and Morrison argue that the reciprocity theorem is not applicable to this problem, and proceed by calculating the flux passing through the surface of one pole piece of the transducer. We argue below that the reciprocity theorem is applicable and, summarizing their method here for comparison, show their assumptions to be equivalent to those made by Karlquist[9] in his analysis of the transducer fringe field. We give a more general expression for the readback voltage, treating the case where the magnetization contains components both in and normal to the coating plane. This result is obtained using the reciprocity theorem and Karlquist's fringe field. The normal component of magnetization is expected to exist (primarily in particulate coatings) since, near the trailing edge of the gap where recording occurs, the fringe field is canted out of the coating plane.

II. THE MINIMUM TRANSITION LENGTH

The demagnetizing field (in oersteds) for a single arctangent transition is calculated[10] in two dimensions from Eqs. (1) and (3). The result, for the special case where **M** is in the coating plane, is

$$H_x(x,y) = 4M_r\left[\tan^{-1}\left(\frac{(\delta/2+y)x}{x^2+a^2+|\delta/2+y|a}\right) + \tan^{-1}\left(\frac{(\delta/2-y)x}{x^2+a^2+|\delta/2-y|a}\right)\right], \quad (4a)$$

and

$$H_y(x,y) = 2M_r \ln\left(\frac{x^2+(|\delta/2+y|+a)^2}{x^2+(|\delta/2-y|+a)^2}\right), \quad (4b)$$

where δ is the coating thickness, x is measured parallel to the track from the center of the transition, y is measured normal to the coating from the center plane, and a and M_r (in emu/cc) are as previously defined. These equations are valid for all x, y, and $a \geq 0$. Equation (4a) can be reduced to the previously obtained result[3,5] via trigonometric identities when $y \geq \delta/2$.

According to Eq. (4) the maximum field magnitude occurs on the coating surface, and H_y exceeds H_x over a considerable portion of the transition region. H_y tends to create a y component of the magnetization m_y, where m_y is an even function of x and an odd function of y. This is in addition to a possible M_y resulting from the y component of the record transducer fringe field, M_y being an odd function of x. The existence of the total y component m_y+M_y is countered by the additional demagnetizing field resulting from $\partial m_y/\partial y + \partial M_y/\partial y$

near $y=\pm\delta/2$ and, in some coatings, by magnetic anisotropy induced by one or more of the following: strain, preferential orientation of crystallites possessing nonzero magnetocrystalline anisotropy constant K_1, and preferential orientation of acicular particles (shape anisotropy). The field as calculated on the assumption M is fixed in the coating plane is in error by approximately this additional demagnetizing field. The error is greatest near the coating surfaces and results primarily in an overestimation of H_y. Therefore, when estimating the minimum transition length which a coating will support according to the method of Chapman and in conjunction with Eqs. (4), it is more meaningful to use the field in the center plane, Eq. (4a) with $y=0$, than to use the magnitude of H elsewhere.

In Fig. 1 both the arctangent transition demagnetizing field and the ramp transition demagnetizing field[5] are plotted versus x in the center plane. Peak fields in this figure differ by more than a factor of two, indicating that the arctangent transition results in a considerable decrease in the minimum transition length which a coating will support as estimated via the method described above. The maximum value of $H_x(x,0)$ as obtained from Eq. (4a) is

$$H_x(\max)=8M_r\tan^{-1}[\delta/4a(1+\delta/2a)^{1/2}]. \quad (5)$$

Upon inverting this equation and setting $H_x(\max)=H_c$, the minimum transition length (as determined by considering demagnetization only) is found to be

$$\begin{aligned} l_{\min}=\pi a_{\min}&=(\pi\delta/4)[\csc H_c/8M_r-1], && H_c/M_r<4\pi \\ &=0, && H_c/M_r\geq 4\pi. \end{aligned} \quad (6)$$

Equation (5) reduces to the previously obtained[6] result, $l_{\min}\simeq 2\pi\delta M_r/H_c$, when $\delta/a\ll 1$. Estimates for $l_{\min}$ based on Eq. (6) may be misleading when $l_{\min}/\delta<$ about 5 (or equivalently $M_r/H_c<$ about 1), since a transition with no y dependence, e.g., Eq. (3), is not a realistic approximation when coating thickness is large compared to transition length. Square hysteresis loops are also assumed, and in this respect[1] together with the condition $M_r/H_c>1$, this estimation of $l_{\min}$ may be more appropriate to thin metallic coatings.

Several observations can be made concerning the microscopic details of the magnetic transition. First, Eq. (3) [and Eq. (9) below] results in all three components of M vanishing at $x=0$. This does not contradict the Stoner-Wohlfarth coherent rotation model[11] or the more appropriate[12,13] single domain incoherent rotation models,[14,15] since the magnetization in reality is free to form a z component. We consider only two dimensions of the three-dimensional problem, a justifiable simplification since the read transducer is insensitive to M_z. Moreover, the magnetization may be considered in the macroscopic or averaged sense whenever particle dimensions are negligible compared to dimensions of interest, thus removing the restriction that at each point $|\mathbf{M}|=M_s(T)$. When dimensions in the x-y plane (i.e., coating thickness, transducer gap width, and transducer-media spacing) are not large compared to domain or particle dimensions, a macroscopic M may still be appropriate in the statistical sense, in that it represents an average of M_x and M_y over the track width or z axis. Second, the relative importance of exchange in limiting the transition length has been investigated by Aharoni[7] and found to be insignificant for current values of coating thickness.

III. THE READBACK VOLTAGE

Readback voltage $e(\bar{x})$, where $\bar{x}\equiv vt$ and v is the relative coating-transducer velocity, may be obtained either directly by calculating the flux through the transducer core, or indirectly by use of the reciprocity theorem. The direct calculation of Speliotis and Morrison[5] first assumes that the read transducer may be replaced by one continuous semi-infinite slab of permeability μ, spaced a distance d above the coating. The method of images is then used to obtain the magnetic induction B inside this slab,

$$\mathbf{B}=[2\mu/(\mu+1)]\mathbf{H}, \quad (7)$$

where H, given by the equivalent (for $y>\delta/2$) of Eq. (4) for an arctangent transition with M in the plane of the coating, is the field intensity that would exist if the slab were absent. Total flux Φ per unit track width through one pole face of a transducer with finite gap length g is then assumed to be

$$\Phi(\bar{x})=\int_{-\infty}^{\bar{x}-g/2}B_y(x,\delta/2+d)\,dx + \int_{\bar{x}-g/2}^{\bar{x}+g/2}[(\bar{x}+g/2-x)/g]B_y(x,\delta/2+d)\,dx, \quad (8)$$

where the second integral apportions the flux entering the gap region between the two pole pieces according to the linear weighting factor $[(\bar{x}+g/2-x)/g]$. The lower limit of integration may be changed from minus infinity to zero without affecting $e(\bar{x})$, which is proportional to $d\Phi(\bar{x})/d\bar{x}$.

In the indirect method employed here, the readback voltage is calculated for the transition

$$\mathbf{M}(x,y)=(-2/\pi)(\hat{\imath}M_x-\hat{\jmath}M_y)\tan^{-1}(x/a), \quad (9)$$

using the reciprocity theorem[16]

$$e(\bar{x})=4\pi NvW\alpha\times 10^{-8}\int_{-\infty}^{\infty}dx \times\int_{d}^{d+\delta}dy\,\frac{\partial\mathbf{M}(x-\bar{x},y)}{\partial\bar{x}}\cdot\mathbf{H}_f(x,y), \quad (10)$$

where N is the number of turns, v is the velocity, W is the track width, and $\mathbf{H}_f$ is the gap fringe field resulting

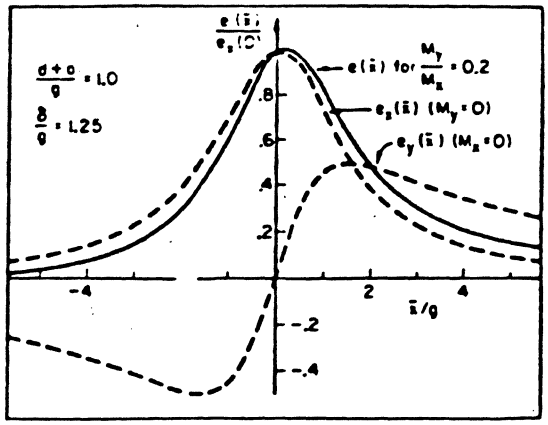


FIG. 2. Readback pulse distortion caused by a nonzero y component of magnetization. The separate contributions of each magnetization component are indicated.

from unit total current flowing around the transducer core. The head efficiency factor α is given by

$$\alpha=\mathcal{R}_g/(\mathcal{R}_g+\mathcal{R}_c), \tag{11}$$

where $\mathcal{R}_g$ and $\mathcal{R}_c$ are the reluctances of the transducer gap and core, respectively. This approach relegates the major approximations to the fringe field calculation. The fringe field, due to Karlquist,[9] is given by

$$H_{fx}(x,y)=\frac{1}{\pi g}\left[\tan^{-1}\left(\frac{g/2+x}{y}\right)+\tan^{-1}\left(\frac{g/2-x}{y}\right)\right], \tag{12a}$$

and

$$H_{fy}(x,y)=\frac{-1}{2\pi g}\ln\left(\frac{(g/2+x)^2+y^2}{(g/2-x)^2+y^2}\right), \tag{12b}$$

assuming that the transducer permeability is infinite, the medium permeability is unity, the pole pieces extend to infinity, and the magnetic scalar potential varies linearly across the gap at $y=0$. This last assumption is equivalent to the linear weighting factor employed by Speliotis and Morrison. Thus, that portion of the readback signal due to the x component of magnetization will be identical to their expression, except for a factor of $\mu/(\mu+1)$, which does not appear here. However, Speliotis and Morrison assume that the flux linking the read coil is α times the flux through one pole piece at $y=\delta/2+d$, and this is valid only if $\mu\gg1$. Upon inserting Eqs. (9) and (12) into Eq. (10), the readback voltage $e(\bar{x})$ in volts is determined[10] to be

$$e(\bar{x})=8.\text{Val}\,\alpha\times10^{-8}\{M_x[f(\bar{x})+f(-\bar{x})]+M_y[h(\bar{x})-h(-\bar{x})]\}, \tag{13a}$$

where

$$f(\bar{x})\equiv\frac{a}{\pi g}\int_{-\infty}^{\infty}dx\int_{d}^{d+\delta}dy\,\frac{\tan^{-1}[(g/2+x)/y]}{(x-\bar{x})^2+a^2}$$

$$=\frac{d+a+\delta}{g}\tan^{-1}\left(\frac{g/2+\bar{x}}{d+a+\delta}\right)-\frac{d+a}{g}\times\tan^{-1}\left(\frac{g/2+\bar{x}}{d+a}\right)+\frac{1}{2g}(g/2+\bar{x})\times\ln\left(\frac{(g/2+\bar{x})^2+(d+a+\delta)^2}{(g/2+\bar{x})^2+(d+a)^2}\right), \tag{13b}$$

where

$$h(\bar{x})\equiv\frac{a}{2\pi g}\int_{-\infty}^{\infty}dx\int_{d}^{d+\delta}dy\,\frac{\ln[(g/2+x)^2+y^2]}{(x-\bar{x})^2+a^2}$$

$$=[(d+a+\delta)/2g]\ln[(g/2+\bar{x})^2+(d+a+\delta)^2]-[(d+a)/2g]\ln[(g/2+\bar{x})^2+(d+a)^2]-(1/g)(g/2+\bar{x})\times\left[\tan^{-1}\left(\frac{g/2+\bar{x}}{d+a+\delta}\right)-\tan^{-1}\left(\frac{g/2+\bar{x}}{d+a}\right)\right], \tag{13c}$$

and where velocity v is measured in cm/sec, track width W in cm, and M_x, M_y in emu/cc. The coating thickness is δ and the transducer-coating spacing is d.

The justification for considering M_y but not m_y in the readback voltage calculation, where M_y is due to H_{fy} and m_y is due to the demagnetizing field, is as follows: M_y contributes to pulse asymmetry, whereas m_y produces only a symmetric first-order correction to the dominant and symmetrical pulse resulting from M_x. A comparison of $e_x(\bar{x})$ and $e_y(\bar{x})$, the contributions to $e(\bar{x})$ due to the parallel and normal magnetization

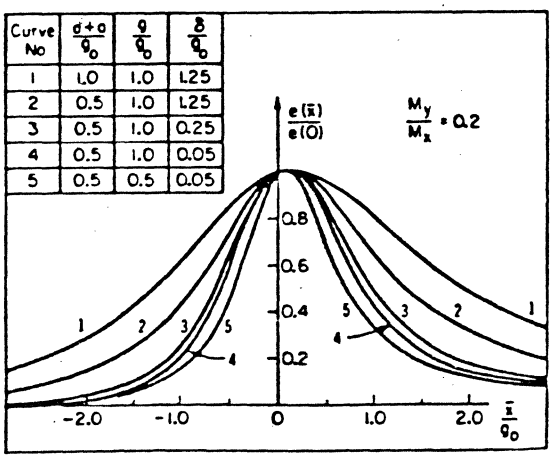


FIG. 3. Readback voltage for several values of $d+a$, δ, and g; and for $M_y/M_x=0.2$. All dimensions are normalized to an arbitrary gap length g_0.

such uV_1 and rV_1. The rounding of the S curve near the threshold indicates that stress renders the core "disturb sensitive" with a corresponding decrease in rV_1 and increase in wV_2. These effects are illustrated in Figs. 4 and 5 for composition 27, which is more stress sensitive than 9.

Effects of Substitutions

We have substituted small amounts of Zn for Mn in compositions 8 and 9 (see Table II). Pulse test data show that the substitution of 0.1 atom Zn in 8 (26) gives optimum results. It increases the voltage outputs uV_1, rV_1, and the rV_1/wV_2. The substitution of 0.1 atom of Cu for Mn in compositions 3 and 19 resulted in a substantial reduction of rV_1/wV_2. The substitution of 0.01 atom of Ca in 8 and 9 resulted in a reduction of the output and an increase in the switching constant without any improvement in other properties.

Conclusions

In the $Li_{0.5}Fe_{2.5}O_4$–$MnFe_2O_4$–Mn_3O_4 system, two compositions (8 and 9) have been found from which reasonably fast-switching memory cores can be made. Such cores have [illegible] temperature or stress than are the square-loop Cu–Mn cores. The incorporation of a small amount of Zn in these two Li–Mn ferrite compositions (26 and 29) increases their output signals without any appreciable deleterious effects on their rV_1/wV_2, switching speed, temperature, or stress sensitivity.

Acknowledgment

The author would like to thank Dr. B. A. Calhoun for suggesting this investigation and for his continued interest. Thanks are also due Dr. E. L. Boyd and R. W. Dam for many helpful discussions, F. J. Hoey for preparing and measuring the samples, J. W. Mitchell for the low-temperature magnetization measurements, and W. W. Yeomans for the pulse measurements.

References

[1] E. A. Schwabe and D. A. Campbell, "Influence of grain size on square-loop properties of lithium ferrite," *J. Appl. Phys.*, vol. 34, p. 1251, 1963.

[2] R. S. Weisz and D. L. Brown, "Square-loop properties of copper-manganese ferrites," *J. Appl. Phys.*, vol. 31, p. 2095, 1960.

[3] E. Albers-Schoenberg, "Square-loop ferrites in the system MgO–MnO–Fe_2O_3," *Bull. Am. Ceram. Soc.*, vol. 35, p. 276, 1956.

The Effect of Finite Flux Rise Time on Recording Performance

JOHN E. LEE AND NORMAN N. TRUMAN

Abstract—An experimental and theoretical investigation has been made of the effect of finite flux rise time on transition length and delay time for thin recording media. This was done by tracing the field history of particles as they pass the write head and by assuming that the remanent magnetization of each particle depends on the maximum field it experienced. For linear rise times and for separation greater than half a gap width, analytic expressions have been derived showing the dependence of transition length and delay time on rise time and the other recording parameters. Numerical solutions valid for separations down to one quarter of a gap width and including exponential rise times are also shown.

Introduction

PRESENT-day computing requires fast-access high-density storage. One such device is the disk store. By operating at high packing densities and high surface speeds a disk store can offer bit transfer rates of the order 10^6 per second together with access times of better than 100 ms. In order to achieve a high packing density the magnetic storage medium must be of high coercivity [1]. Writing with conventional ring-type heads requires therefore that a high current be switched into an inductive load with only a finite voltage available. This must constitute a finite lower limit to the flux rise time. Also, a further limitation may be introduced by relaxation phenomena in the head core material itself. During the finite period taken for the flux to rise the recording medium is moving past the write gap at high speed. This could cause the written transition to become broader than would be expected under ideal, i.e., zero rise-time conditions [2], [3].

The purpose of the present work has been to investigate in more detail the effect of finite flux rise time and to compare its effect on the packing density limitations with that of self-demagnetization which takes place in the recording medium after the write process has been completed. Kostyshyn [4] and others have analyzed the

Manuscript received March 18, 1969; revised September 18, 1969. Paper 2.6, presented at the 1969 INTERMAG Conference, Amsterdam, The Netherlands, April 15–18.

The authors are with Computer Research and Advanced Development Division, International Computers Ltd., Stevenage, Herts., England.

ording process in considerable detail and a similar proach could have been used. However, elaborate calations are not required in order to demonstrate the mparative limitations imposed by finite flux rise time d demagnetization. Therefore, a less rigorous approach s been adopted.

Model for Write Process

Before any attempt can be made to determine a written nsition length, three basic assumptions must be made. ese assumptions are concerned with 1) the properties the medium, 2) the spatial distribution of the head-kage field, and 3) the nature of the flux rise time.

operties of the Medium

The following analysis will be concerned only with thin etallic films, and the magnetic behavior will be om the $M_r - H$ characteristic. No attempt will be take into account either the microscopic nature of the versal process or the detailed effects of self-demagnetization [5]. The remanent magnetization of each particle in e medium will be assumed to depend solely on the aximum longitudinal field it experiences as it approaches, sses, and leaves the write-head gap. The effect of the rtical field component [6] will not be considered.

ead-Field Distribution

The most accurate determinations of the spatial field stribution have been made by Fay [7] using conformal nsformation and by Daniker [8] using matrix analysis. rlquist's expression for the longitudinal field $H_x(x,y)$ a less exact representation of the distribution but has e advantage of being analytic. According to Karlquist

$$H_x(x,y) = (H_g/\pi)[\tan^{-1}(g+x)/y + \tan^{-1}(g-x)/y]. \quad (1)$$

quation (1) becomes a better fit the greater the y coordinate, and even when $y = g/4$ the error is less than percent.

In the following analysis the longitudinal field will be etermined using (1). The field distribution will be assumed to be independent of the presence of the medium nd y will be taken as the separation since the medium is thin film, i.e., $D \ll y$.

ux Rise Time

Relaxation phenomena within the head core material ill be ignored and the flux rise time will be taken as the urrent rise time, since experiments were carried out using urrent rise times greater than the relaxation time constant. For simplicity we shall deal only with saturation RZ recording, i.e., $H_x(0,y) > H_c$ for all y.

Theoretical

We shall assume that the medium is initially remanent $-M_r$ and that the gap field switches from negative to positive. This means that only the positive fields determine the final remanent state of the medium. Similar considerations may be carried out for fields switching from positive to negative. However, in this latter case the final remanence may be determined by both the positive and the negative fields, hence the calculation may be more complicated.

From (1) the longitudinal field seen at time t by a particle positioned at x_s with respect to the head gap center line as origin is given by

$$H_x(x_s,y,t) = (H_g/\pi)f(t) \cdot [\tan^{-1}(g + x_s)/y + \tan^{-1}(g - x_s)/y] \quad (2)$$

where $f(t)$ represents the form of the head current rise time. Simplifying the Karlquist expression

$$H_x(x_s,y,t) = (H_g/\pi)f(t)\tan^{-1}[2gy/(y^2 - g^2 + x_s^2)] \quad (3)$$

and provided $(y^2 - g^2 + x_s^2) \gg 2gy$, (3) may be simplified further to give

$$H_x(x_s,y,t) = \frac{2H_g ygf(t)}{\pi(y^2 - g^2 + x_s^2)}. \quad (4)$$

Linear Rise Times

For linear rise times, $f(t)$ is given by

$$f(t) = t/T, \quad 0 < t \leq T$$
$$f(t) = 1, \qquad t > T \quad (5)$$

where T is the time taken for the current to rise from zero to its maximum value and shall hitherto be referred to as the current rise time. Therefore, substituting into (4) for $f(t)$, the field seen at time t by a particle positioned at x_s becomes

$$H_x(x_s,y,t) = \frac{2H_g ygt}{\pi T(y^2 - g^2 + x_s^2)} \quad (6)$$

where

$$x_s = x_m + vt. \quad (7)$$

Differentiating $H_x(x_s,y,t)$ with respect to t and equating to zero for a stationary value yields

$$0 = x_s^2 - 2vtx_s + y^2 - g^2 \quad (8)$$

for a given separation y, gap width $2g$, and velocity v. Substituting for x_s in (6) yields

$$t = \frac{H_g yg}{2H_x\pi Tv^2} + \frac{(y^2 - g^2)H_x\pi T}{2H_g yg} \quad (9)$$

which together with (7) gives

$$x_m = \frac{H_g yg}{2\pi H_x vT} - \frac{(y^2 - g^2)\pi H_x vT}{2H_g yg} \quad (10)$$

where t represents the time taken for the particle labelled x_m to see a maximum field H_x and hence to acquire its remanent magnetization.

Delay Time

When x_m is chosen such that $H_x = H_c$, that particle will leave the write head with zero remanence and x_m will

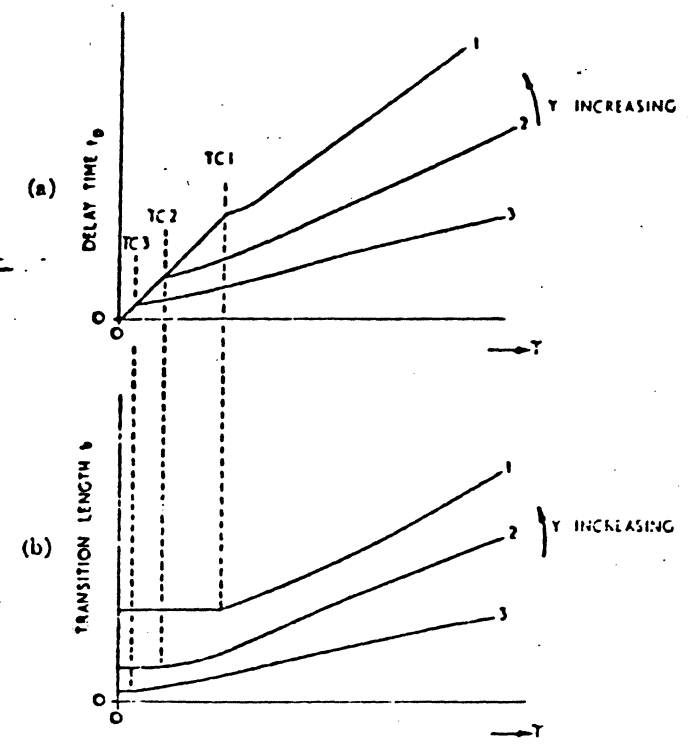


Fig. 1. (a) Delay time versus rise time (schematic). (b) Transition length versus rise time (schematic).

represent the center of the transition region. The corresponding time $t = t_D$ given by

$$t_D = \frac{H_g yg}{2H_c \pi v^2} + \frac{(y^2 - g^2)H_c \pi T}{2H_g yg} \qquad (11)$$

is the delay time and represents the time taken, measured from the instant when the write current is zero, to write the center of the transition on the medium. The form of this relationship is shown in Fig. 1(a) where delay time has been plotted versus rise time for arbitrary known values of H_c/H_c, g, y, and v. Since (11) holds only in the region $t_D \leq T$ there will be a discontinuity in the curve, as shown. This will occur at some critical rise time T_c obtained from (11) by substituting $t_D = T = T_c$, i.e.,

$$T_c^2 = \frac{H_g yg/2H_c \pi v^2}{1 - [(y^2 - g^2)H_c \pi/2H_g yg]}. \qquad (12)$$

In this region, i.e., $T < T_c$, the appropriate particles see a peak (though not stationary) field equal to the coercivity H_c after a time $t = T$ after which time they see ever decreasing fields as they move further from the head gap. When

$$0 = 1 - [(y^2 - g^2)H_c \pi/2H_g yg] \qquad (13)$$

the critical rise time appears to be infinite. This never happens in practice, however, since comparison of (13) and (6) shows that this condition corresponds to

$$H_x(0,y,T) = H_c,$$

and the medium would never be correctly switched (in the usual sense).

For convenience H_g may be chosen such that the field gradient $\partial H_x(x_m,y)/\partial x_m$ is a maximum at $H_x(x_m,y) = H_c$, where $H_x(x_m,y)$ is the field distribution round the head when $t > T$ and given by

$$H_x(x_m,y) = 2gH_g y/\pi(y^2 - g^2 + x_m^2). \qquad (14)$$

This condition can be shown to occur at

$$x_m^2 = y^2 - g^2 \qquad (15)$$

where

$$H_c = 2gH_g y/\pi(y^2 - g^2 + x_m^2) \qquad (16)$$

that is,

$$H_c/H_g = gy/\pi(y^2 - g^2). \qquad (17)$$

Therefore, from (12)

$$T_c = (y^2 - g^2)^{1/2}/v. \qquad (18)$$

Calculations have shown that this value of T_c is accurate to within 10 percent for $y \geq 1.2g$.

Transition Length

Although $M_r(x_m)$ may be determined using (10), together with the $M_r(H)$ characteristic, the procedure is a lengthy one. Therefore, the transition length was defined and determined using a procedure similar to that due to Davies [9]. According to Davies, since

$$\frac{dM_r}{dt} = \frac{dM_r}{dH}\frac{dH}{dx_m}\frac{dx_m}{dt}$$

then for a given medium velocity, since $dM_r/dH \gg dH/dx_m$, the peak amplitude of the output pulse on read corresponds to the region where dM_r/dH has its maximum value, i.e., at the center of the transition where $M_r = 0$, $H = H_c$. Therefore, the transition length may be conveniently defined with respect to the fields H_1 and H_2 where dM_r/dH has its half-peak value and the output pulse half its peak amplitude. Using this definition the transition length $2b$ will be given by

$$2b = (dx_m/dH_x)_{H_x = H_c}(H_2 - H_1) \qquad (19a)$$

that is, by differentiating (10)

$$2b = \left[\frac{H_g yg}{2\pi H_c^2 vT} + \frac{(y^2 - g^2)\pi vT}{2H_g yg}\right](H_2 - H_1). \qquad (19b)$$

The form of this relationship is shown in Fig. 1(b) where $2b$ has been plotted versus rise time for arbitrary known values of H_g, H_c, $(H_2 - H_1)$, g, y, and v. Again, since (11) holds only for $t_D \leq T$ there will be a discontinuity as shown. In the region $T \leq T_c$ the appropriate particles see peak fields equal to H_1 and H_2 after a time $t = T$, and from (7)

$$x_{s1} = x_{m1} + vT$$
$$x_{s2} = x_{m2} + vT \qquad (20)$$

that is,

$$x_{s1} - x_{s2} = x_{m1} - x_{m2}.$$

That is, the transition length becomes independent of rise time and equal to the zero rise-time value [10]. When

H_g is chosen such that

$$H_c/H_g = y g^{-1}(y^2 - g^2)^{1/2} \quad (17')$$

the transition length becomes

$$2b = v(H_2 - H_1)(T_c^2 + T^2)/2H_cT, \quad (0 < T \leq T_c). \quad (21)$$

Numerical Solution

When the more accurate expression for the field given by (3) is used the expression for delay time, critical rise time, and transition length are no longer analytic. Because of the large number of variables involved the problem is best solved numerically using a computer program, and all theoretical results quoted below have been calculated in this way.

Exponential Rise Times

For exponential rise times $f(t)$ is given by

$$f(t) = 1 - e^{-t/T} \quad (22)$$

where T is the rise time and is the time taken for the current to rise from zero to 63 percent of its maximum value. Substituting this into (2) gives

$$H_x(x_s,y,t) = (H_g/\pi)(1 - e^{-t/T}) \cdot \left[\tan^{-1}\frac{g + x_s}{y} + \tan^{-1}\frac{g - x_s}{y}\right] \quad (23)$$

and the analyses for calculating the delay time and transition length now follow in a manner similar to that for linear rise times described in the previous sections.

Experimental

All experimental work was carried out with reference to the analysis presented above relating to linear rise times. The current waveform used is shown in Fig. 2(a) and was generated using a circuit which provided rise times T_2 from 10 μs to 10 ms.

The delay time t_D was determined by measuring the time interval t_1 between the crossover points of the current waveform and the time interval t_2 between the peaks of the output pulses Fig. 2(b). From Fig. 2(b),

$$t_D = t_2 - (t_1 + t_3)$$

and by arranging for the rise times T_1 and T_2 to be such that

$$T_2 \geq 100T_1$$

then to a very good approximation t_3 may be ignored and

$$t_D = t_2 - t_1.$$

The transition length $2b$ was determined by measuring the 50 percent pulsewidth. According to Bonyhard *et al.* [1], who assume no write losses, the 50-percent pulsewidth is

$$p_{50} = 2[(a + y_s)^2 + g^2]^{1/2} \quad (24)$$

where $2a$ defines the minimum transition length in terms of the physical properties of the medium and is given by

$$a = 2M_rD/H_c.$$

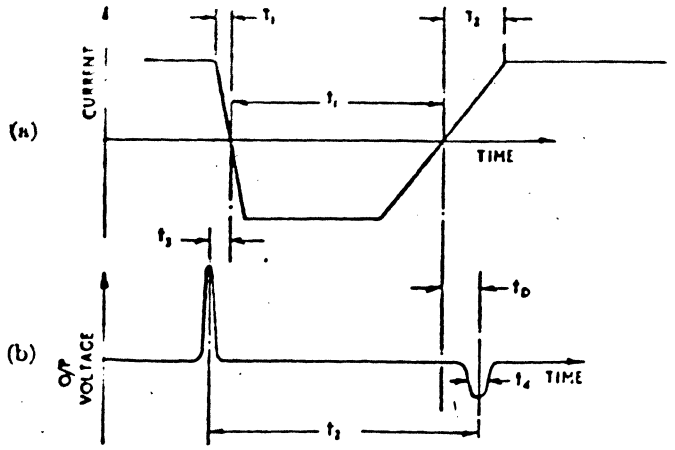


Fig. 2. (a) Write-head current versus time. (b) Read-head output versus time.

For the present work a is replaced by b, as defined in (19a), and with reference to Fig. 2(b), (24) becomes

$$t_4 = (2/v)[(b + y)^2 + g^2]^{1/2}. \quad (25)$$

The parameters used in these experiments were

v = 75 or 37.5 in·s^{-1}
y = 100–500 μin
$2g$ = 580 μin
H_g = 3300 Oe
H_c = 630 Oe
D = 12 μin
M_r = 600 gauss

The head field within the gap H_y ($t > T$) was determined by writing static transitions [6] (i.e., at zero tape velocity) and measuring the distance between the peaks of the output dipulse.

Results and Discussion

Experimental

The results for delay time are shown in Fig. 3. The solid lines represent the calculated values and the points experimental values. Although some inconsistent results occurred as a result of the speed instability in the tape transport used, the agreement between theory and experiment was fair.

The agreement between theory and experiment for the transition lengths, however, was poor except at long rise times where it was good. A possible reason for this is that in contrast to the delay time measurements the transition lengths were determined indirectly from the pulsewidths using (25). The validity of this equation depends on the assumption made in deriving it that the transition is of the form

$$M_r(x) = (2M_0/\pi)\tan^{-1}(x/b).$$

This assumption is not strictly true because the form of the written transition clearly depends on the $M_r - H$ characteristic of the medium and on the write-field distribution. When the transition is long compared with the read gap however, i.e., at long rise times, the pulsewidth

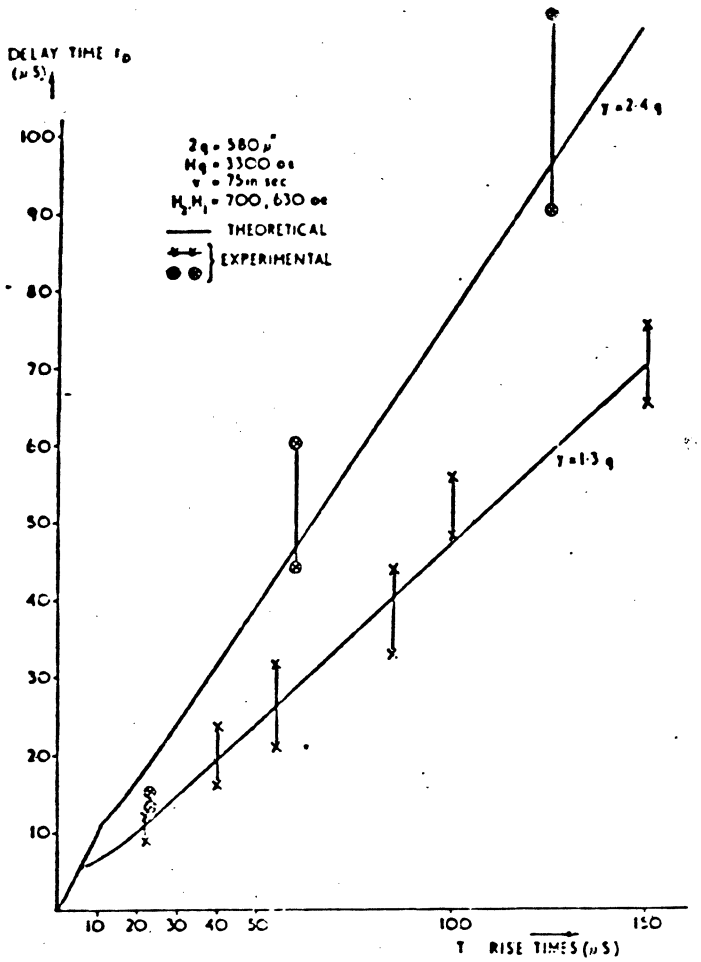


Fig. 3. Comparison of experimental with theoretical results for delay time versus rise time.

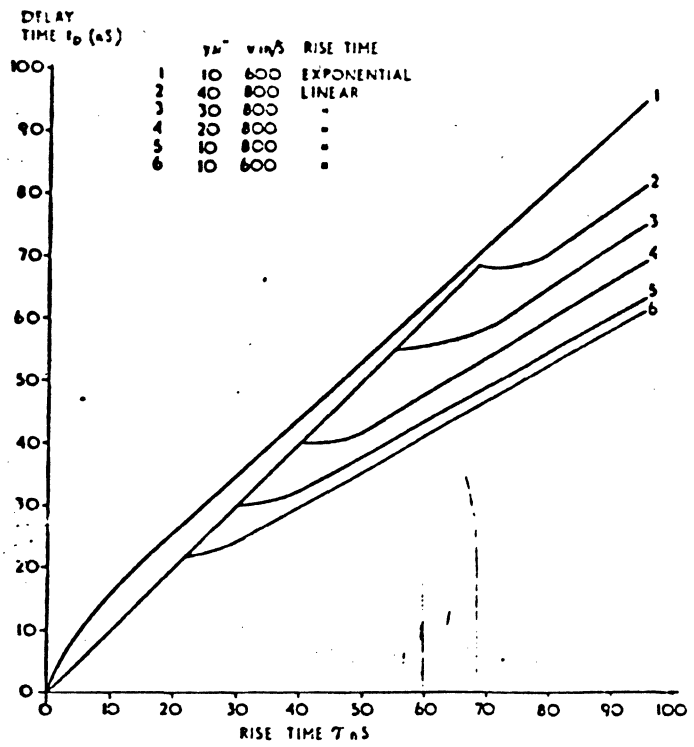


Fig. 4. Delay time (theoretical) versus rise time for various velocities and separations $H_2 = 860$ Oe, $H_1 = 540$ Oe.

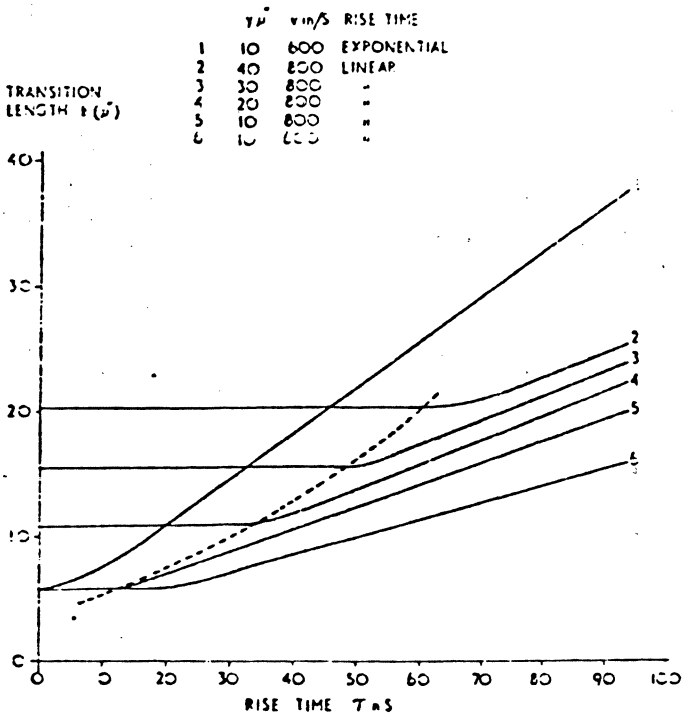


Fig. 5. Transition length (theoretical) versus rise time for various velocities and separations $H_2 = 860$ Oe, $H_1 = 540$ Oe. The dotted line is the locus of T_c.

becomes the same as the transition width and less dependent on the details of the replay process. This makes the assumptions made previously more nearly correct.

Multiplying (11) by v we have

$$vt_D = \frac{yg}{2\pi}\left(\frac{H_g}{H_c vT}\right) + \frac{(y^2 - g^2)}{2yg}\left(\frac{H_c vT}{H_g}\right).$$

That is, for a given head gap $2g$ and separation y then provided $H_g/H_c vT$ be held constant then vt_D should remain constant. Similarly, from (19b) the transition length should remain constant for a given $H_c/H_c vT$. This was investigated experimentally by varying H_g, H_c, v, and T.

The agreement here was better than for the previous results. This is to be expected if the theory is correct, since the experiments were designed to demonstrate a dependence rather than to determine absolute values.

Theoretical

Figs. 4 and 5 show delay time and transition length versus rise time for parameters typical of a high density disk recording system using thin metallic films. H_g was chosen in each case such that, for zero rise time, the transitions were written in the maximum head field gradient. The gradients of the curves of Fig. 5 decrease with decreasing separations, i.e., b is insensitive to changes in rise time even for $T > T_c$. Also for exponential rise times there is no discontinuity in the curves, i.e., there is no critical rise time.

From Fig. 5 the pulsewidth and hence the output versus packing density curve may be obtained using (25). For linear rise times there is no change in b and hence no change in p_{50} for T less than T_c. Even if the flux rise time is exponential which is more likely in practice, there is an effective time T_c below which no appreciable change in p_{50}

a length $2a$ [1] on leaving the write head. Thus
ne medium is large, the value of the written tran-
ngth may still be such that the pulsewidth is
$y^2 + g^2]^{1/2}$ and is hence independent of rise time
ide range. The above shows that a rise time equal
ving a transition length $2b$ approximately twice
lue of the medium would be an optimum.

Fig. 4 the delay time is only zero when the rise
zero. Since in a real system, timing difficulties
y decide the maximum usable packing density,
ay-time considerations may often mean that a
rise time than is determined on grounds of pulse-
s necessary.

Conclusion

ave shown that the theory gives reasonable agree-
ith experiment for delay time, although the agree-
or transition lengths was poor. This was due to
ficulties in maintaining and measuring the head-
n separation, and to the inaccuracy of some of the
imations. This approach has, however, enabled
pendence of transition length and delay time on
ne to be determined in analytic form and an ex-
n for critical rise time to be derived which is ac-
even for low separations.

authors have shown that the results given may be
o compare the effects of self-demagnetization with
f finite flux rise time.

Nomenclature

ongitudinal coordinate fixed in medium
ongitudinal coordinate fixed relative to gap center
ead to medium separation
ap width
demagnetization transition length
transition length
time
T rise time
t_d delay time
T_c critical rise time
H_x longitudinal component of head field
H_x maximum longitudinal field seen by particle
H_g head field within gap ($t > T$)
H_c coercivity of medium
M_0 retentivity of medium
M_r remanent magnetization.

Acknowledgment

The authors wish to thank M. Aspland for writing the computer programs and the directors of International Computers Limited for their permission to publish.

References

[1] P. I. Bonyhard, A. V. Davies, and B. K. Middleton, "A theory of digital magnetic recording on metallic films," *IEEE Trans. Magnetics*, vol. MAG-2, pp. 1–5, March 1966.

[2] I. Stein, "Generalized pulse recording," *IEEE Trans. Electronic Computers*, vol. EC-12, pp. 77–92, April 1963.

[3] J. R. Herbert, "A computer simulation of the magnetic recording process," *Proc. 1965 INTERMAG Conf.* (Washington, D.C.), pp. 15-1-1–15-1-8.

[4] B. Kostyshyn, "A theoretical model for a quantitative evaluation of magnetic recording systems," *IEEE Trans. Magnetics*, vol. MAG-2, pp. 236–242, September 1966.

[5] J. J. Miyata and R. R. Hartel, "The recording and reproduction of signals on magnetic medium using saturation-type recording," *IRE Trans. Electronic Computers*, vol. EC-8, pp. 159–169, June 1959.

[6] J. E. Lee and N. N. Truman, "The effect of vertical head field component on the switching of thin metallic films," *IEEE Trans. Magnetics*, vol. MAG-4, pp. 277–280, September 1968.

[7] L. E. Fay, "Two dimensional theory of the recording head," *Internatl. Conf. on Magnetic Recording*, pp. 71–77, July 1969.

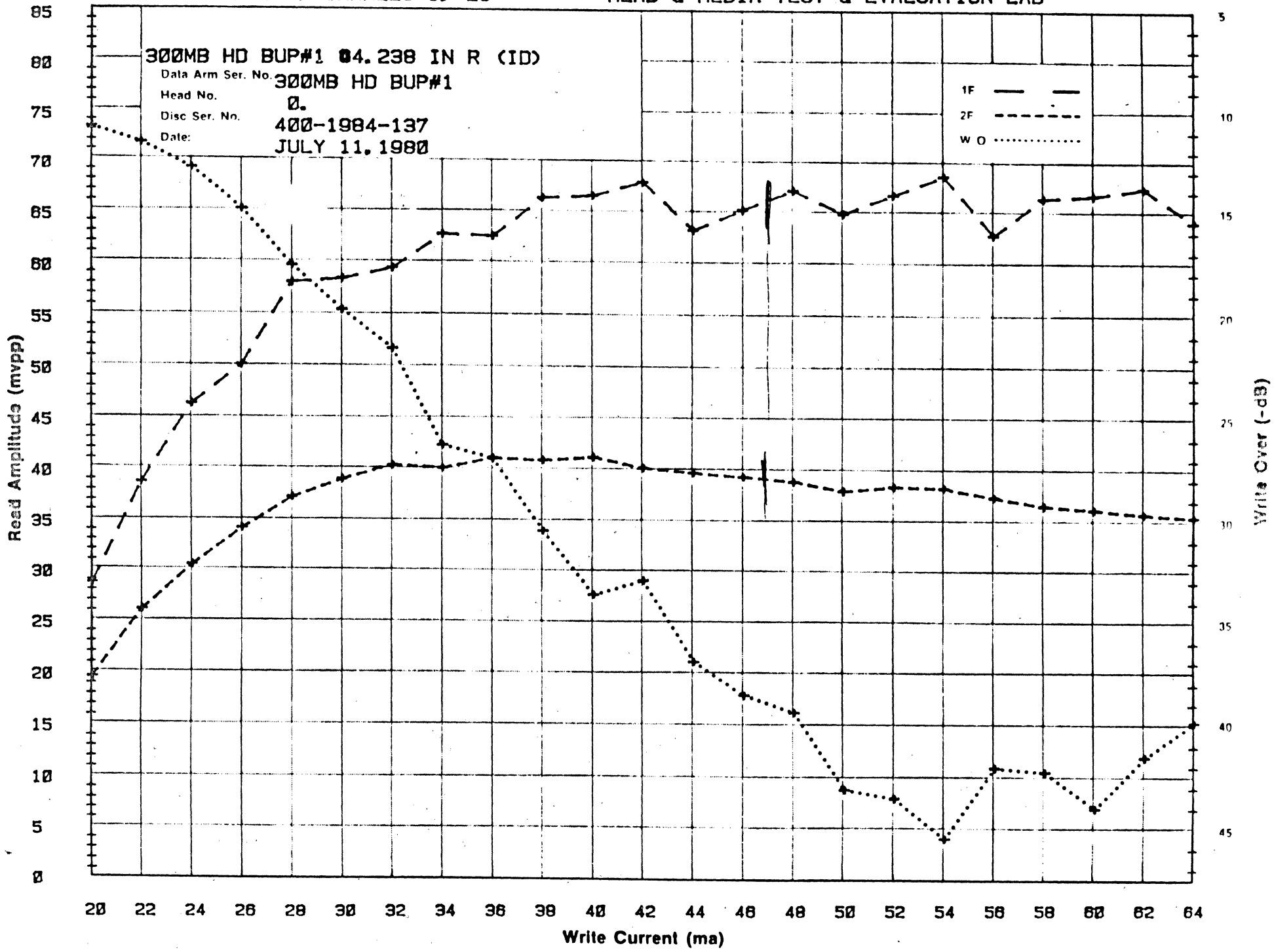
[8] S. Duinker, "On the resolving power in the process of magnetic recording," *Tijdschr. Ned. Radiogen.*, vol. 22, pp. 29–48, January 1957.

[9] A. V. Davies, "Effects of the writing process and cross-talk on the timing accuracy of pulses in NRZ digital recording," *IEEE Trans. Magnetics*, vol. MAG-3, pp. 217–222, September 1967.

[10] A. S. Hoagland, *Digital Magnetic Recording*, 1st ed. New York: Wiley, pp. 120–122.

PREPARED BY : CHARLES C. LU

HEAD & MEDIA TEST & EVALUATION LAB



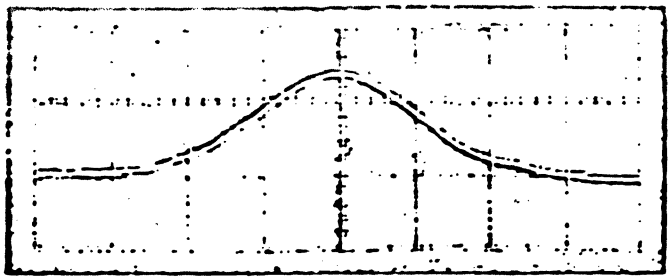


Fig. 6—Oscillogram of an isolated read pulse.

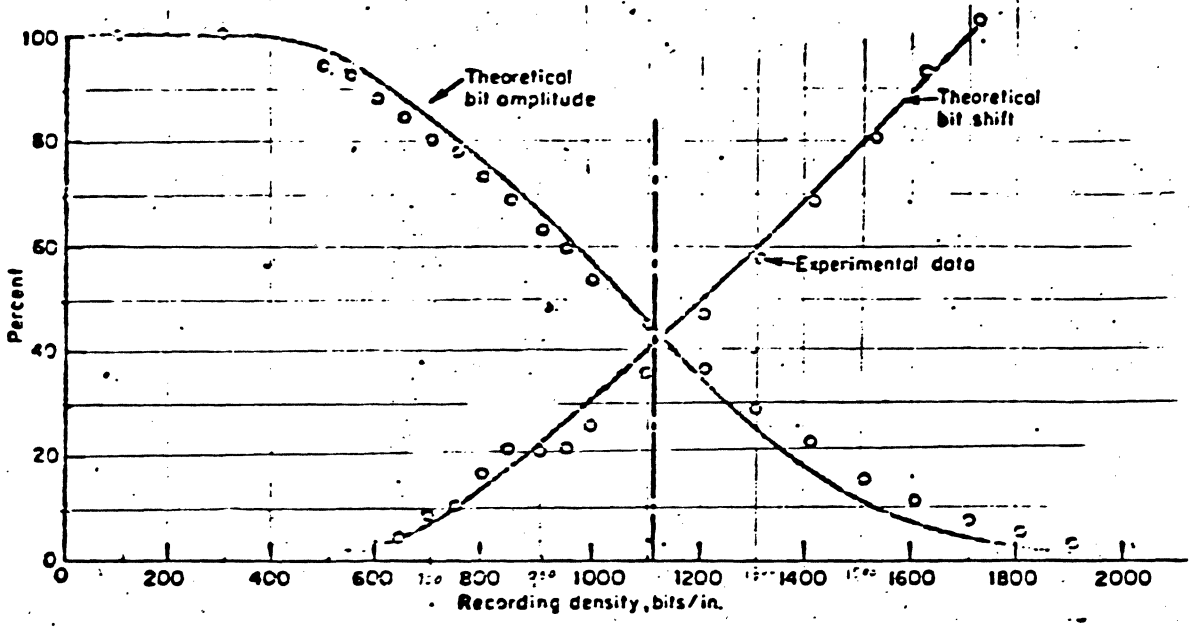


Fig. 7—Comparison of experimental and theoretical results for bit shift and amplitude. The dots represent experimental data taken with a digital-readout oscilloscope.

where $n = 0, \pm1, \pm2, \pm3, \ldots$. At $t = 0$ there is a maximum of

$$f(0) = 2\sum_{n=0}^{\infty}(-1)^n \exp[-(n\tau)^2] - 1$$

Since the maximum of an isolated normalized pulse is 1, the bit amplitude in percent is given by

$$BA = \left\{2\sum_{n=0}^{\infty}(-1)^n \exp\left[-(n\tau)^2\right] - 1\right\} \times 100 \quad (8)$$

Equation (8) has also been programmed on the 7090. The resulting plot of bit amplitude versus T/τ is given in Fig. 4.

Experimental Verification

The theoretical results described above were verified by running laboratory tests on an actual read head. A non-contact head with a mechanical loading of 350 grams was operated from data recorded on a magnetic disk of aluminum coated with iron oxide. The track diameter was 8 in., disk coating thickness was 200 μin., linear velocity was 1200 in./sec, and write current was 100 ma. A continuous sequence of 1's was used to test for bit amplitude. To test for bit shift, the format 0000000110000000 was used.

An oscillogram of an isolated read pulse is shown in Fig. 6. The experimental and theoretical curves for bit shift and bit amplitude are compared in Fig. 7. They show good agreement between predicted values and test results.

Acknowledgments

The author wishes to acknowledge the assistance of Donald J. Kastella, who wrote the programs referred to in the text, and John D. Bartlett, who assisted during the laboratory evaluation.

Bibliography

"A Logical Reading System for Non-Return-to-Zero Magnetic Recording," A. S. Hoagland, *IRE Trans. on Electronic Computers*, September 1955, pp 93-95.

"High Density Digital Magnetic Recording Techniques," G. C. Bacon and A. S. Hoagland, *Proc. IRE*, January 1961, p 265.

"Transfer Function and Error Probability of a Digital Magnetic Tape Recording System," J. W. Hung, *J. Appl. Phys.*, Vol 31, No. 5, May 1960, pp 39oS, 397S.

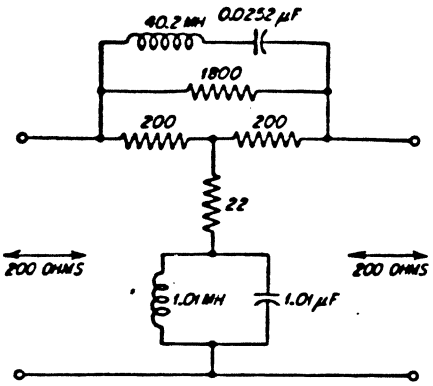


FIG. 17.17. Bridged-T equalizer for Example 17.4.

$$L_1 = 7.96 \times \frac{10 - 1}{\sqrt{10}} \times \frac{1}{1.25^2 - 1} = 40.2 \text{ mh}$$

$$L_2 = 7.96 \times \frac{\sqrt{10}}{10 - 1} \times \frac{1.25^2 - 1}{1.25^2} = 1.01 \text{ mh}$$

$$C_1 = 0.199 \times \frac{\sqrt{10}}{10 - 1} \times \frac{1.25^2 - 1}{1.25^2} = 0.0252 \ \mu f$$

$$C_2 = 0.199 \times \frac{10 - 1}{\sqrt{10}} \times \frac{1}{1.25^2 - 1} = 1.01 \ \mu f$$

$$R_1 = 200(10 - 1) = 1{,}800 \text{ ohms}$$

$$R_2 = 200 \times \frac{1}{10 - 1} = 22.2 \text{ ohms}$$

(Refer to Figs. 17.17 and 17.18

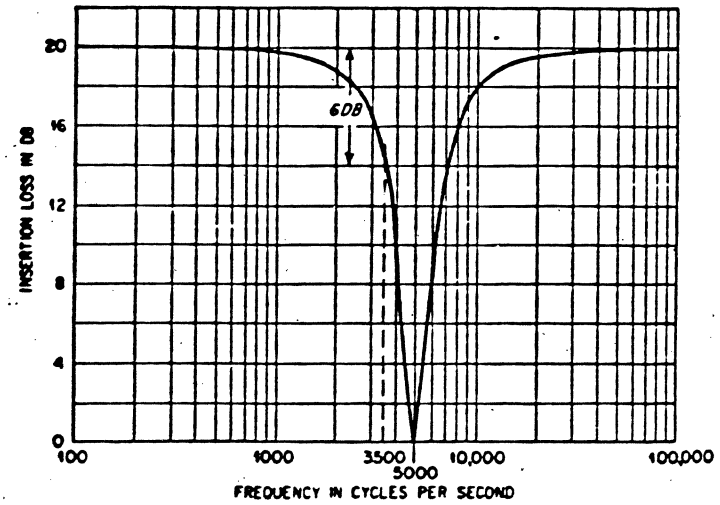


FIG. 17.18. Insertion-loss characteristics of network shown in Fig. 17.17.

17.3. Phase Equalizers. The types of phase equalizers treated are those w theoretically introduce either zero or a fixed amount of attenuation at all frequen They can therefore be added to existing circuits for phase correction without torting the gain characteristics.

The shape of electrical impulses which contain many frequency components ca distorted in passing through an electrical circuit even though the circuit has the gain for the different frequency components. If such is the case, the distort due to unequal transmission delays for the different frequency components. type of distortion is called *phase distortion* and can be corrected by adding a net which will cause the total transmission period for all frequencies to be identical. added network must, therefore, be a network in which the phase character can be controlled.

Equal transmission periods for different frequency components through a c stipulates that the circuit must introduce either no phase shift or an amount of shift which is directly proportional to frequency. This is identical to stating the transmission period must be either zero or a constant amount at all frequ

$$7.96 \times \frac{10-1}{\sqrt{10}} \times \frac{1}{1.25^2-1}$$

$$= 40.2 \text{ mh}$$

$$7.96 \times \frac{\sqrt{10}}{10-1} \times \frac{1.25^2-1}{1.25^2}$$

$$= 1.01 \text{ mh}$$

$$.199 \times \frac{\sqrt{10}}{10-1} \times \frac{1.25^2-1}{1.25^2}$$

$$= 0.0252 \ \mu f$$

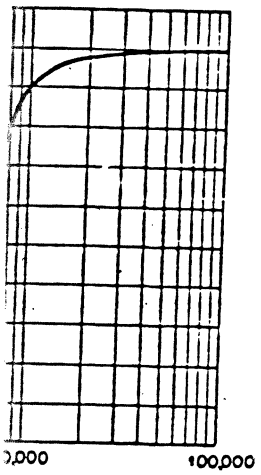
$$.199 \times \frac{10-1}{\sqrt{10}} \times \frac{1}{1.25^2-1}$$

$$= 1.01 \ \mu f$$

$$00(10-1) = 1{,}800 \text{ ohms}$$

$$00 \times \frac{1}{10-1} = 22.2 \text{ ohms}$$

(Refer to Figs. 17.17 and 18.)



,000 100,000

ECOND

twork shown in Fig. 17.17.

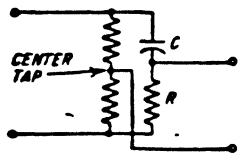
qualizers treated are those which
of attenuation at all frequencies
for phase correction without dis-

ny frequency components can be
n though the circuit has the same
uch is the case, the distortion is
nt frequency components. This
be corrected by adding a network
frequencies to be identical. The
which the phase characteristics

y components through a circuit
hase shift or an amount of phase
This is identical to stating that
nstant amount at all frequencies

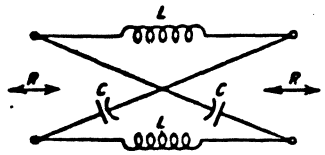
Four different configurations of phase equalizers[1] are shown in Figs. 17.19 to 17.21. It should be noted that the four-terminal networks can be used in only those applications in which the input and output circuits are either both balanced or are in no way connected to each other.

The circuit in Fig. 17.19 introduces an insertion loss of 6 db and does not have constant input and output impedances as a function of frequency. In addition, the



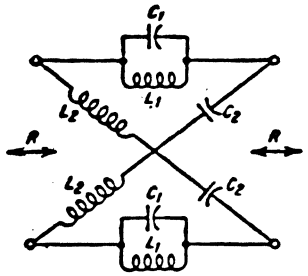
$f_0 = \frac{1}{2\pi RC}$
WHERE f_0 IS THE FREQUENCY AT WHICH THE PHASE SHIFT β IS EQUAL TO $-90°$

FIG. 17.19. Phase equalizer with a fixed insertion loss of 6 db. The phase characteristics are exactly the same as for the lattice network shown in Fig. 17.20, provided the output is not loaded.

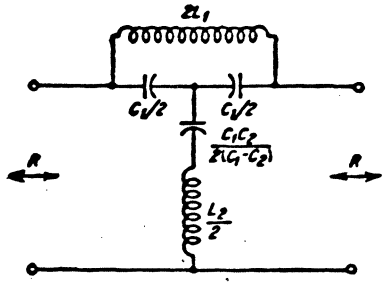


$L = \frac{2R}{2\pi}$ $\quad a = 1/f_0$
WHERE f_0 IS THE FREQUENCY AT WHICH NETWORK PHASE SHIFT IS $-90°$
$C = \frac{L}{R^2}$
$\tan \frac{\beta}{2} = af = \left(\frac{f}{f_0}\right)$

FIG. 17.20. Phase-shift network with zero attenuation. Refer to Fig. 17.22 for phase characteristics.



(a) LATTICE NETWORK



(b) BRIDGED-T EQUIVALENT TO LATTICE IF $C_1 > C_2$

$L_1 = \frac{2R}{2\pi}$ $\quad C_1 = \frac{b}{2\pi aR}$
$L_2 = R^2C_1$ $\quad C_2 = \frac{L_1}{R^2}$
$b = V_{f_0}$ WHERE f_0 IS THE FREQUENCY AT WHICH THE NETWORK PHASE SHIFT IS $-180°$

$$\tan \frac{\beta}{2} = \frac{af}{1-bf^2} = \frac{af_0\left(\frac{f}{f_0}\right)}{1-\left(\frac{f}{f_0}\right)^2}$$

FIG. 17.21. Phase-shift network with zero attenuation. Refer to Fig. 17.23 for phase characteristics.

phase-shift curve for the circuit, Fig. 17.22, is based on there being no terminating impedance.

A center-tapped transformer secondary winding could be substituted for the resistor in Fig. 17.19, provided the amplitude and phase characteristics of the transformer were acceptable.

The networks shown in Figs. 17.20 and 17.21 have constant input and output characteristic impedances as a function of frequency and provide phase shift without attenuation. Figures 17.22 and 17.23 indicate the phase characteristics which can be obtained.

The phase equalizers shown in Figs. 17.19 to 17.21 introduce a lagging phase shift.

[1] These networks are also referred to as *all-pass* filters.

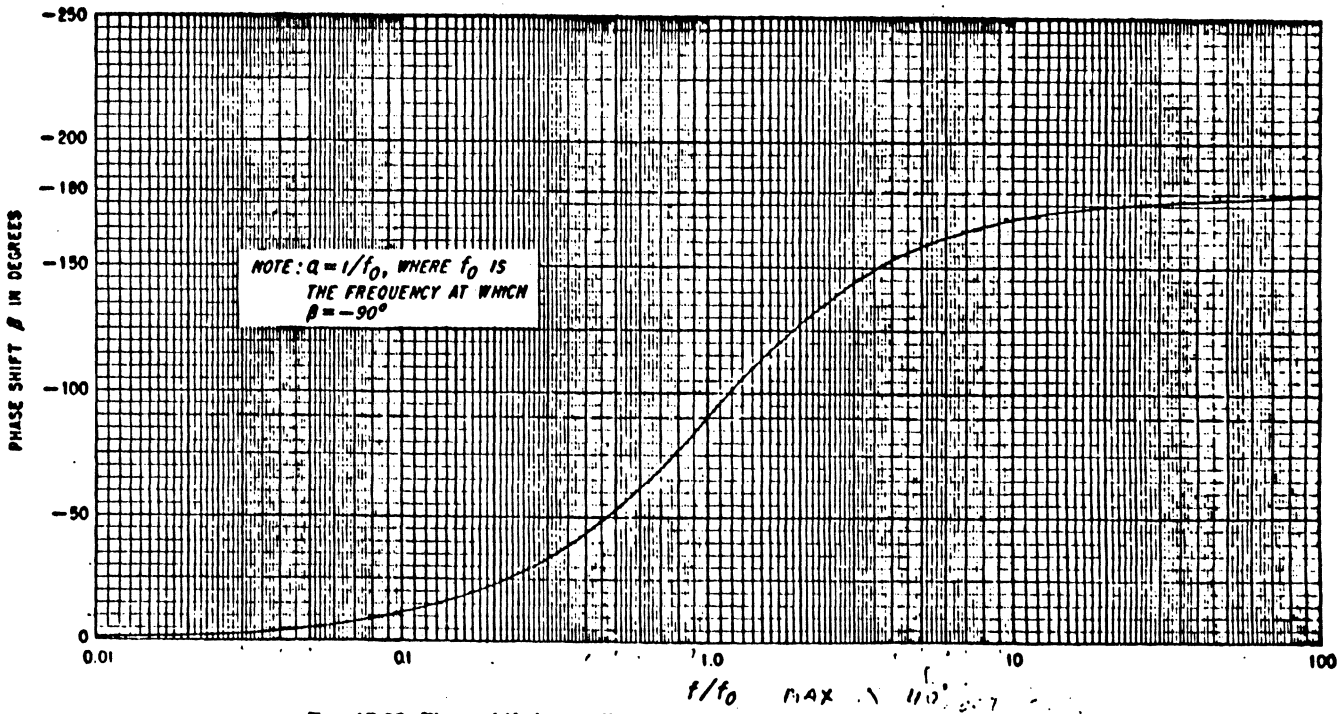


FIG. 17.22. Phase shift in an all-pass filter of the type shown in Fig. 17.20.

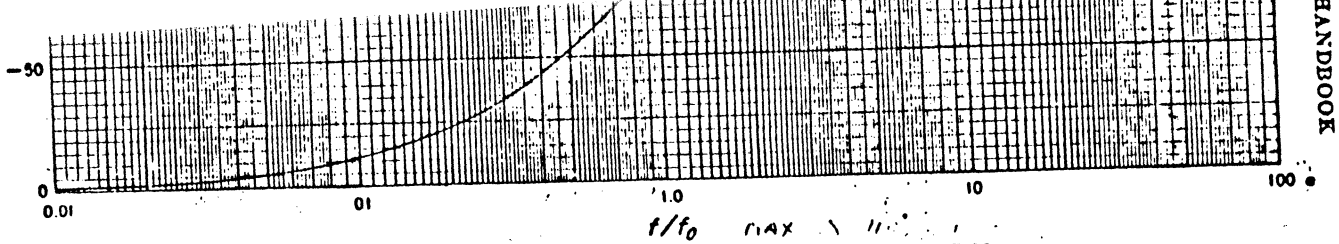


FIG. 17.22. Phase shift in an all-pass filter of the type shown in Fig. 17.20.

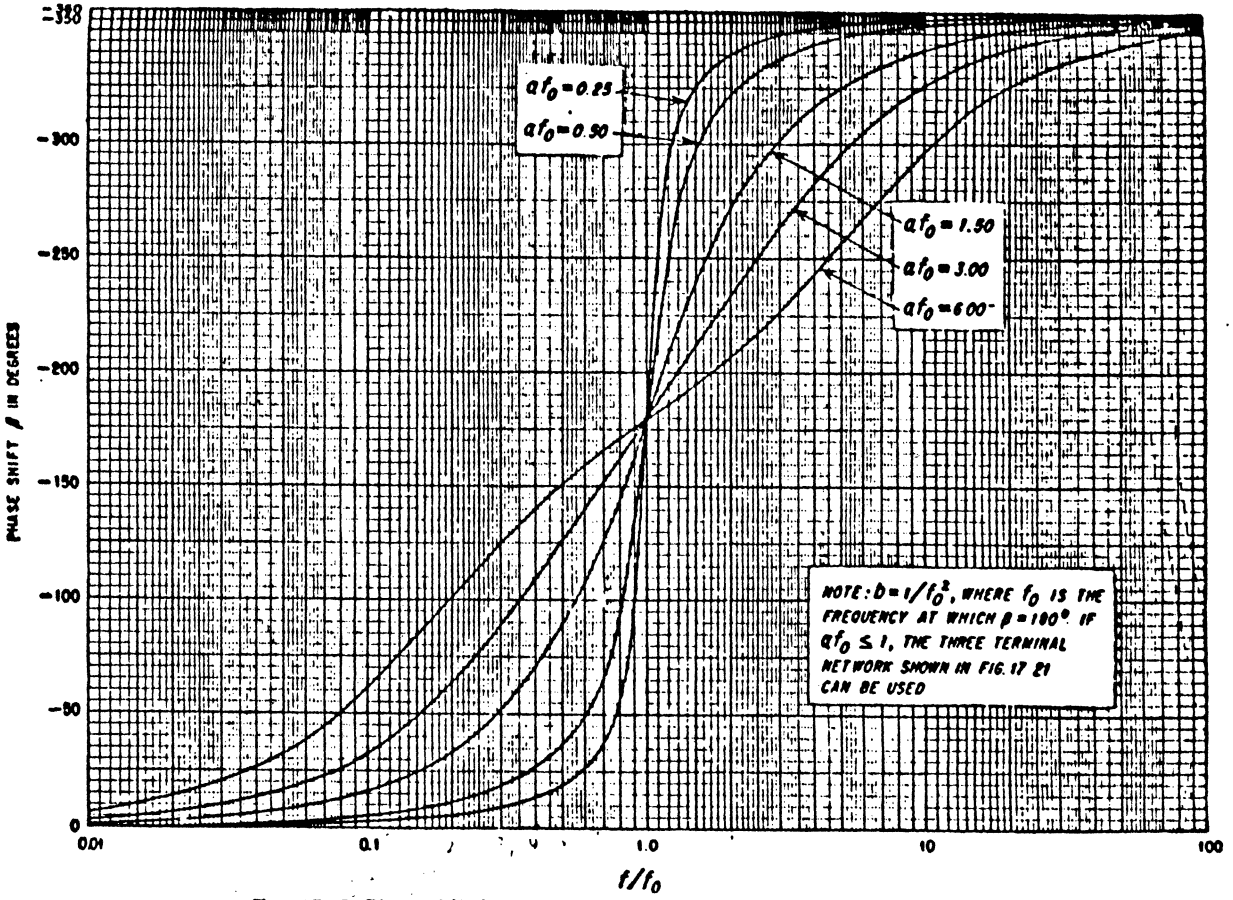


FIG. 17.23. Phase shift in an all-pass filter of the type shown in Fig. 17.21.

Example 17.5

Assume that intelligence must be transmitted in the 10- to 20-kc frequency band and that the circuit employed introduces phase shift in accordance with the following tabulation.

Phase Shift	*Frequency*
−27°	10 kc
−43.5°	15 kc
−63°	20 kc

Design a phase equalizer of the lattice type with a characteristic impedance of 1,000 ohms for use with this circuit.

Solution

1. Determine the required phase characteristics of the phase equalizer.

The departure from linear phase shift as a function of frequency for the existing circuit must first be determined. Since the phase shift at 20 kc is −63°, the phase shift at 10 kc should be ½ × −63, or −31.5°, and the phase shift at 15 kc should be ¾ × −63, or −47.25°. The existing network therefore introduces a phase error of +4.5° at 10 kc and +3.75° at 15 kc.

Phase Error	*Frequency*
+4.5°	10 kc
+3.75°	15 kc
0°	20 kc

The phase equalizer must therefore exhibit the inverse characteristics, i.e.,

Phase Error	*Frequency*
−4.5°	10 kc
−3.75°	15 kc
0°	20 kc

2. Determine from Figs. 17.22 and 17.23 if the required conditions tabulated in step 1 can be satisfied with either of the networks shown in Figs. 17.20 or 17.21.

Since the network shown in Fig. 17.20 is simpler, the curve shown in Fig. 17.22 should first be examined.

The procedure is to determine if the phase shift in the equalizer at any three values of f/f_e, which are related in the same proportions as are 10, 15, and 20 kc, will depart from linear phase shift as a function of frequency by the desired amount. A few experimental groups of values of f/f_e reveal that the phase shifts for f/f_e equal to 0.4, 0.6, and 0.8 are equal to −43, −61.5, and −77°, respectively, and satisfy the specified requirements. This is true since a phase shift of −77° at $f/f_e = 0.8$ requires that the phase shift be −38.5 and −57.75° at f/f_e equal to 0.4 and 0.6, respectively, for linear phase characteristics. The phase equalizer therefore introduces phase errors of −4.5 and −3.75° when f/f_e is equal to 0.4 and 0.6, respectively. It should be apparent that the three values of f/f_e, that is, 0.4, 0.6, and 0.8, correspond to f being equal to 10, 15, and 20 kc, respectively.

3. Determine f_e and the values for the lattice elements.

$$\frac{f}{f_e} = 0.8 \text{ (at } f = 20 \text{ kc)}$$

$$f_e = 25 \text{ kc}$$

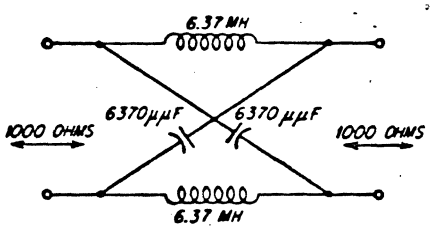


FIG. 17.24. Lattice network for Example 17.5.

From Fig. 17.22

$$a = \frac{1}{25{,}000} = 4 \times 10^{-5}$$

From Fig. 17.20

$$L = \frac{4 \times 10^{-5} \times 10^{3}}{2 \times 3.14}$$

$$= 6.37 \times 10^{-3} \text{ henry, or } 6.37 \text{ mh}$$

$$C = \frac{6.37 \times 10^{-3}}{10^{6}}$$

$$= 6{,}370 \times 10^{-12} \text{ farad, or } 6{,}370\ \mu\mu\text{f}$$

The lattice network is shown in Fig. 17.24.

C. P. EGGEN, Lockheed Missiles and Space Co...

MODERN LOW-PASS FILTER CHARACTERISTICS

Side-by-side comparison of filter characteristics can help the designer to choose wisely between the available types of networks. This article displays phase and group delay as well as magnitude characteristics.

A NUMBER of different types of networks are used to implement low-pass filters, and designers often need to compare the various types for possible use in a given application. Magnitude-versus-frequency characteristics are usually well known, but less information has been published on phase response and group delay; yet the latter characteristics are often equally important. This article describes the phase and delay characteristics of a number of important modern filter types so that they can be easily evaluated and compared.

The network types considered here include Butterworth or maximally flat filters and Chebyshev or equi-ripple filters [1–5]; Bessel, Thompson, or maximally flat delay filters [6–8]; transitional Butterworth-Thompson filters [9]; and Legendre, optimum monotonic, or L-type filters [10–12].

The first three types are widely known. [13] The transitional Butterworth-Thompson filter attempts to combine the best features of the Butterworth and Bessel filters. The design is based on a map that smoothly transforms the filter poles from the Butterworth locations to the normalized Bessel locations as a parameter m changes from 0 to 1. The transitional Butterworth-Thompson poles are in various compromise locations defined by values of m between 0 and 1.

The Legendre characteristic attempts to combine the best characteristics of the Butterworth and Chebyshev characteristics. Here, the stopband-magnitude characteristic is made as steep as possible with the restriction that the characteristic be monotonic.

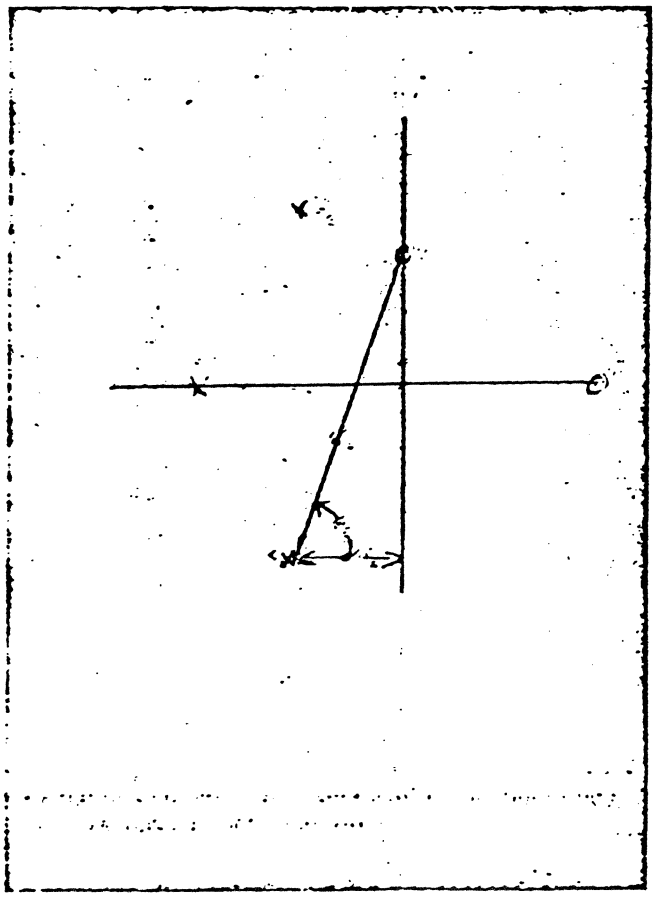
Filter designers face many difficult choices in matching available design techniques to specific applications. Graphical comparisons of techniques emphasize their limitations and potentialities.

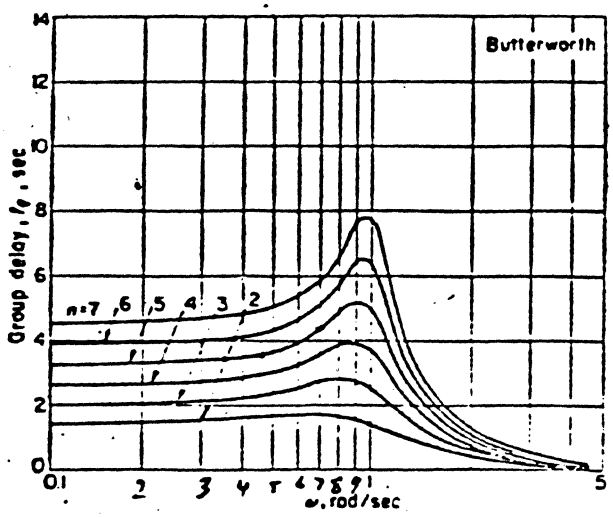
Lumped-Parameter Low-Pass Filters

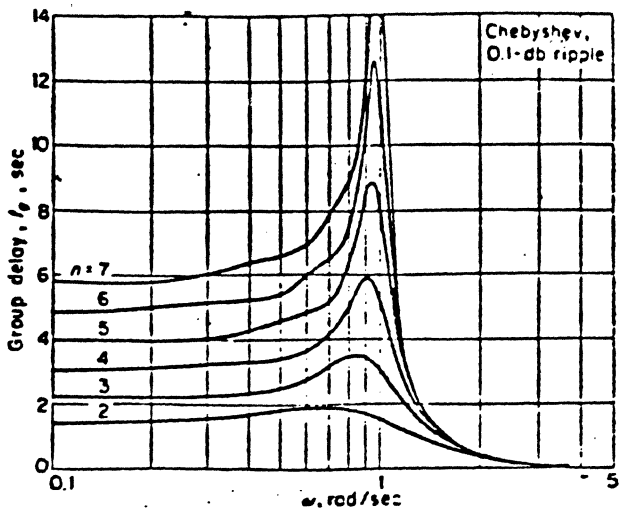
The filters discussed here are all low-pass all-pole networks. The transfer function for such a network is of the form:

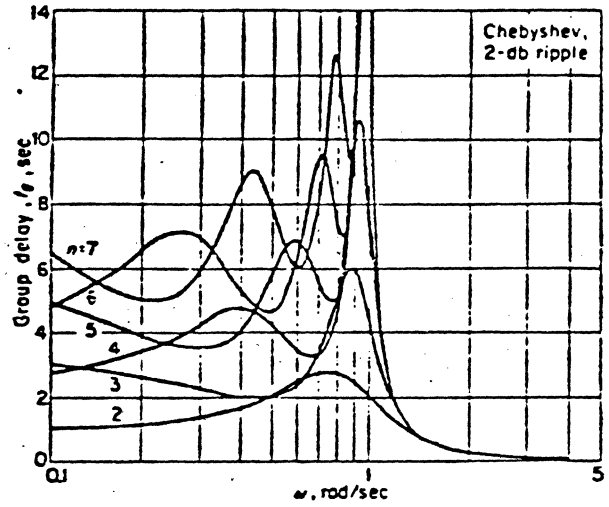
$$H(s) = \frac{K}{Q(s)} = \frac{K}{(s - s_1)(s - s_2)\cdots(s - s_n)} \qquad (1)$$

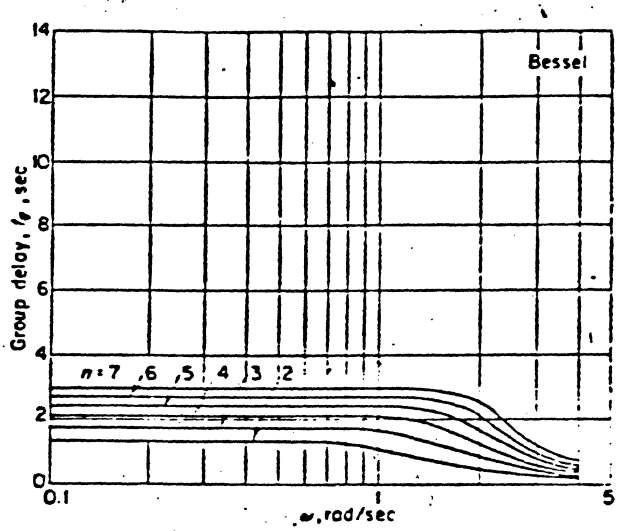
where $Q(s)$ is a polynomial in s of order n (n is also the order of the filter), and s_i are the poles, whose locations determine the characteristics of the filter. When subjected to sinusoidal input, the magnitude, phase, and group delay of the network are found by setting $s = j\omega$. Then

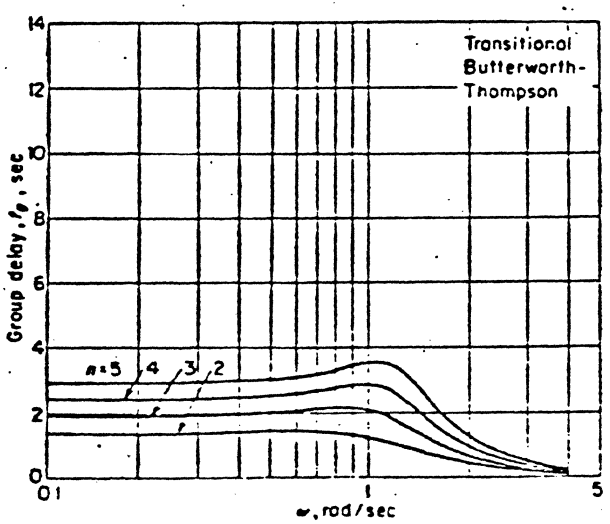


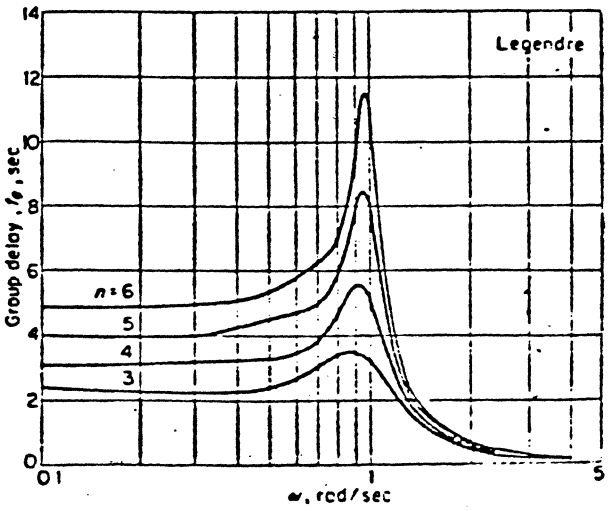

Butterworth
Group delay, t_g, sec
n=7 ,6 ,5 ,4 ,3 ,2
0.1
2 3 4 5 6 7 8 9 1
5
ω, rad/sec

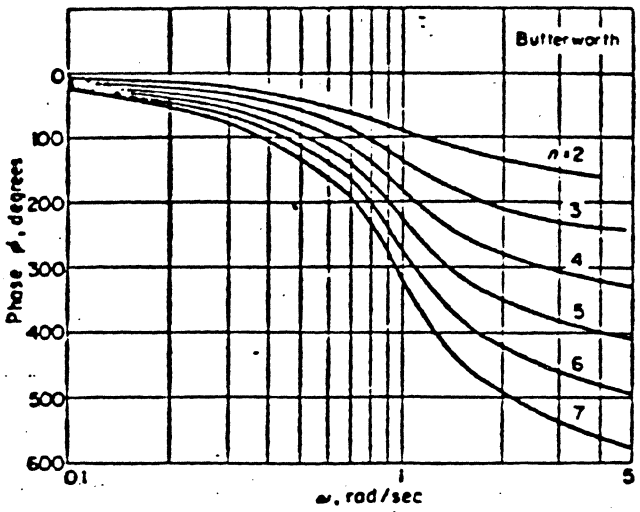

Chebyshev, 0.1-db ripple
Group delay, t_g, sec
n=7
6
5
4
3
2
0.1
1
5
ω, rad/sec

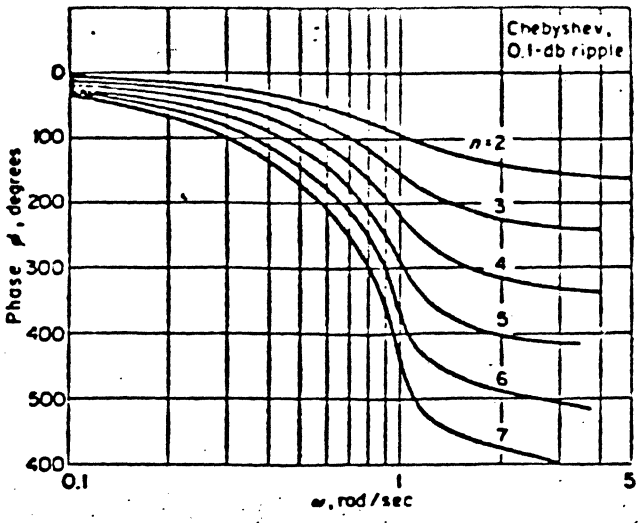

Chebyshev, 2-db ripple
Group delay, t_g, sec
n=7
6
5
4
3
2
0.1
1
5
ω, rad/sec

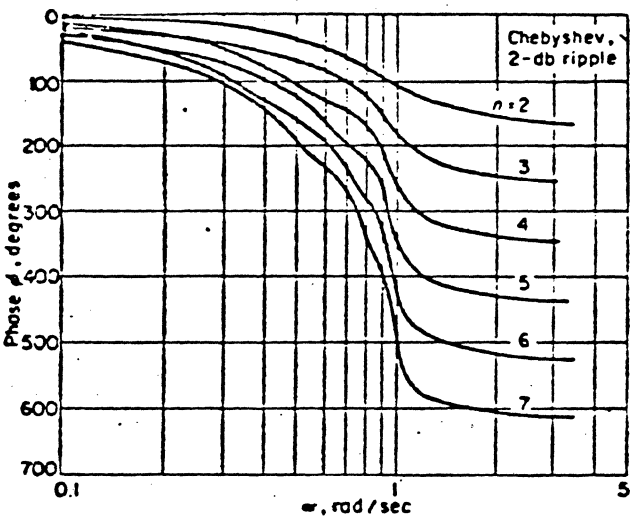

Bessel
Group delay, t_g, sec
n=7 ,6 ,5 ,4 ,3 ,2
0.1
1
5
ω, rad/sec

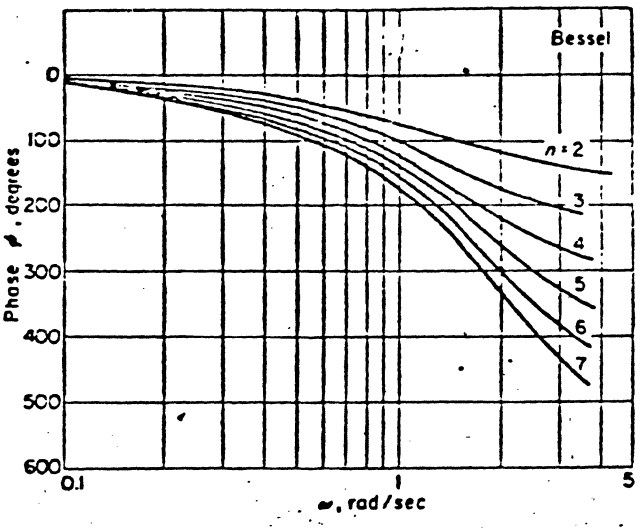

Transitional Butterworth-Thompson
Group delay, t_g, sec
n=5 ,4 ,3 ,2
0.1
1
5
ω, rad/sec

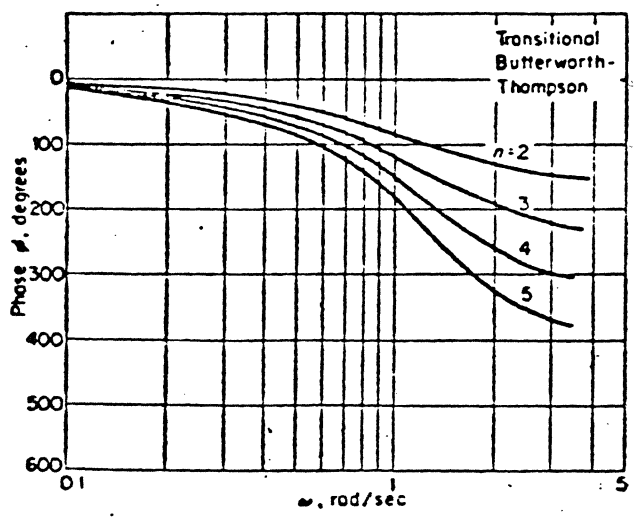

Legendre
Group delay, t_g, sec
n=6
5
4
3
0.1
1
5
ω, rad/sec

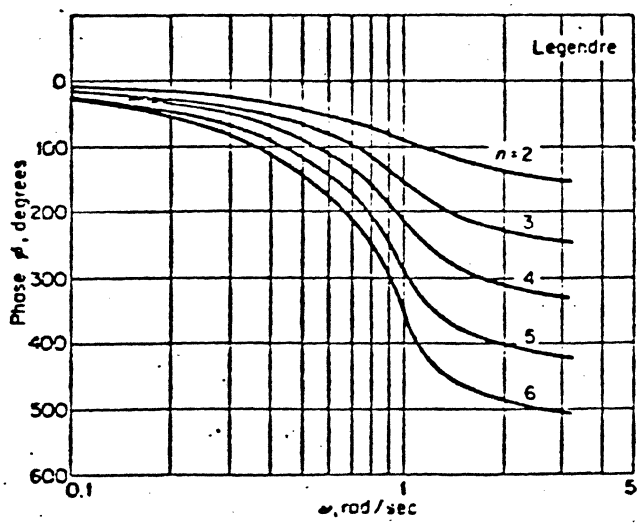

Butterworth
n=2
3
4
5
6
7
Phase φ, degrees
0
100
200
300
400
500
600
0.1
1
5
ω, rad/sec

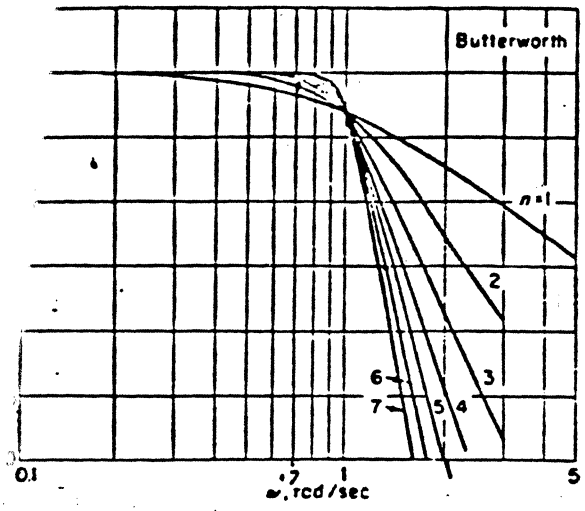

Chebyshev, 0.1-db ripple
n=2
3
4
5
6
7
Phase φ, degrees
0
100
200
300
400
500
0.1
1
5
ω, rad/sec

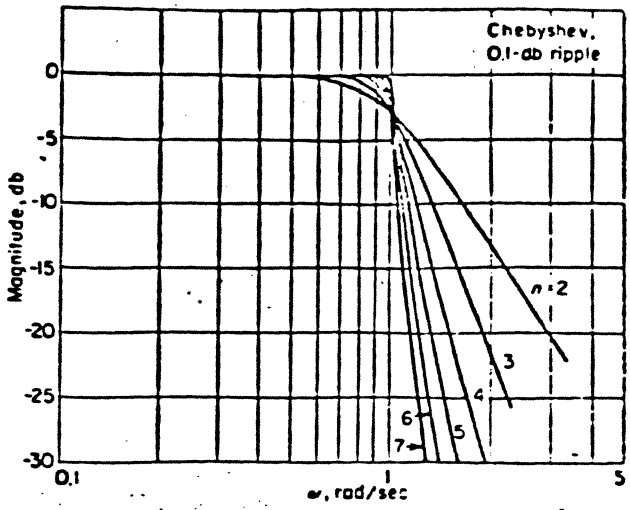

Chebyshev, 2-db ripple
n=2
3
4
5
6
7
Phase φ, degrees
0
100
200
300
400
500
600
700
0.1
1
5
ω, rad/sec

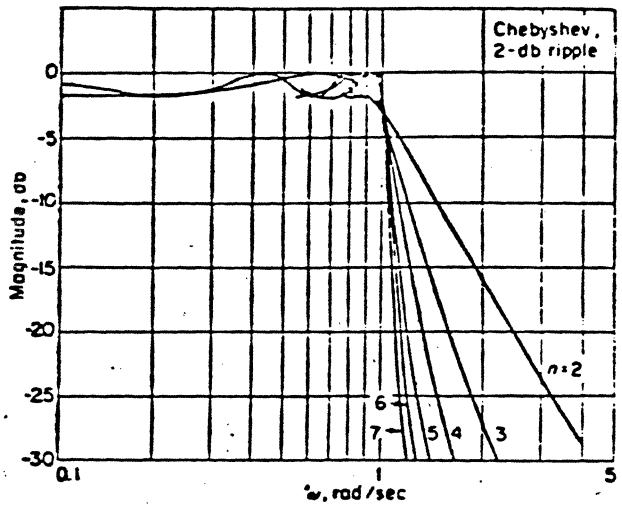

Bessel
n=2
3
4
5
6
7
Phase φ, degrees
0
100
200
300
400
500
600
0.1
1
5
ω, rad/sec

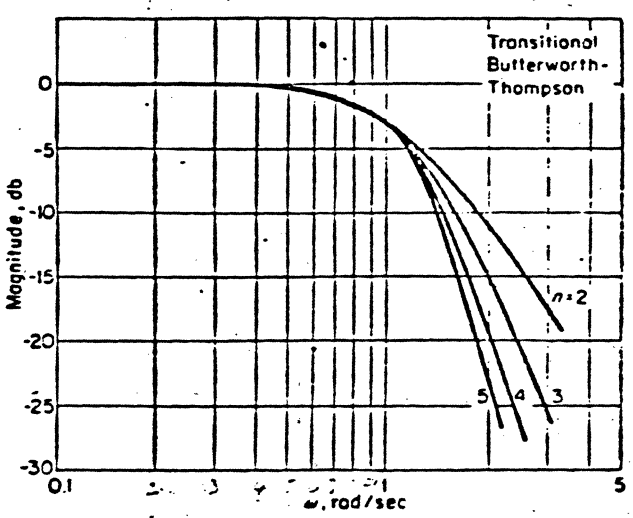

Transitional Butterworth-Thompson
n=2
3
4
5
Phase φ, degrees
0
100
200
300
400
500
600
0.1
1
5
ω, rad/sec

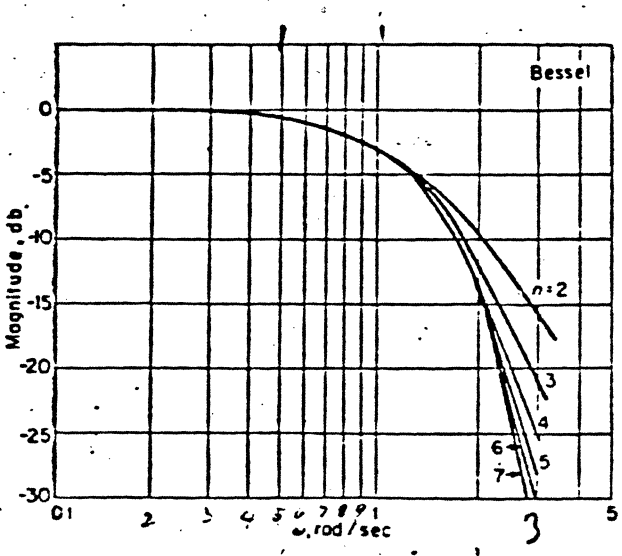

Legendre
n=2
3
4
5
6
Phase φ, degrees
0
100
200
300
400
500
600
0.1
1
5
ω, rad/sec

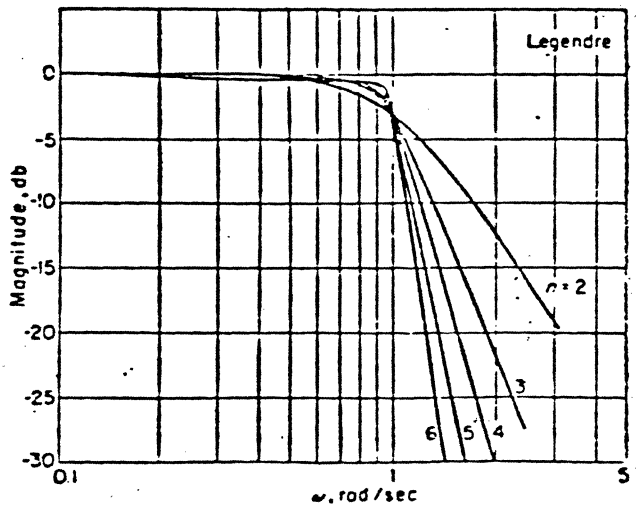

Butterworth
n=1
2
3
4
5
6
7
0.1
1
5
ω, rad/sec


Chebyshev,
0.1-db ripple
Magnitude, db
0
-5
-10
-15
-20
-25
-30
n=2
3
4
5
6
7
0.1
1
5
ω, rad/sec


Chebyshev,
2-db ripple
Magnitude, db
0
-5
-10
-15
-20
-25
-30
n=2
3
4
5
6
7
0.1
1
5
ω, rad/sec


Transitional
Butterworth-
Thompson
Magnitude, db
0
-5
-10
-15
-20
-25
-30
n=2
3
4
5
0.1
1
5
ω, rad/sec


Bessel
Magnitude, db
0
-5
-10
-15
-20
-25
-30
n=2
3
4
5
6
7
0.1
1
5
ω, rad/sec


Legendre
Magnitude, db
0
-5
-10
-15
-20
-25
-30
n=2
3
4
5
6
0.1
1
5
ω, rad/sec

Normalized Solutions

	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$
Butterworth	$s_{1,2}=-0.7071 \pm j0.7071$	$s_1=-1.0000+j0$ $s_{2,3}=-0.5000 \pm j0.8660$	$s_{1,2}=-0.9239 \pm j0.3827$ $s_{3,4}=-0.3827 \pm j0.9239$	$s_1=-1.0000+j0$ $s_{2,3}=-0.8090 \pm j0.5878$ $s_{3,4}=-0.3090 \pm j0.9511$	$s_{1,2}=-0.9659 \pm j0.2588$ $s_{3,4}=-0.7071 \pm j0.7071$ $s_{5,6}=-0.2588 \pm j0.9659$	$s_1=-1.0000+j0$ $s_{2,3}=-0.9010 \pm j0.4339$ $s_{4,5}=-0.6235 \pm j0.7818$ $s_{6,7}=-0.2225 \pm j0.9749$
Chebyshev, 0.1-db ripple	$s_{1,2}=-0.6104 \pm j0.7106$	$s_1=-0.6979+j0$ $s_{2,3}=-0.3489 \pm j0.8683$	$s_{1,2}=-0.5257 \pm j0.3833$ $s_{3,4}=-0.2177 \pm j0.9254$	$s_1=-0.4749+j0$ $s_{2,3}=-0.3842 \pm j0.5884$ $s_{5,6}=-0.1467 \pm j0.9521$	$s_{1,2}=-0.3916 \pm j0.2590$ $s_{3,4}=-0.2867 \pm j0.7076$ $s_{5,6}=-0.1049 \pm j0.9666$	$s_1=-0.3527+j0$ $s_{2,3}=-0.3178 \pm j0.4341$ $s_{4,5}=-0.2199 \pm j0.7022$ $s_{6,7}=-0.0785 \pm j0.9754$
Chebyshev, 2-db ripple	$s_{1,2}=-0.3741 \pm j0.7572$	$s_1=-0.3572+j0$ $s_{2,3}=-0.1786 \pm j0.8938$	$s_{1,2}=-0.2486 \pm j0.3896$ $s_{3,4}=-0.1029 \pm j0.9406$	$s_1=-0.2157+j0$ $s_{2,3}=-0.1745 \pm j0.5946$ $s_{4,5}=-0.0666 \pm j0.9621$	$s_{1,2}=-0.1738 \pm j0.2609$ $s_{3,4}=-0.1272 \pm j0.7128$ $s_{5,6}=-0.0465 \pm j0.9737$	$s_1=-0.1544+j0$ $s_{2,3}=-0.1391 \pm j0.4364$ $s_{4,5}=-0.0962 \pm j0.7065$ $s_{6,7}=-0.0343 \pm j0.9807$
Transitional Butterworth-Thompson	$s_{1,2}=-0.8615 \pm j0.6977$	$s_1=-1.1249+j0$ $s_{2,3}=-0.6942 \pm j0.9368$	$s_{1,2}=-1.0858 \pm j0.3987$ $s_{3,4}=-0.5843 \pm j1.0605$	$s_1=-1.1771+j0$ $s_{2,3}=-1.0059 \pm j0.6428$ $s_{4,5}=-0.5103 \pm j1.1442$		
Bessel	$s_{1,2}=-1.1016 \pm j0.6364$	$s_1=-1.3226+j0$ $s_{2,3}=-1.0474 \pm j0.9992$	$s_{1,2}=-1.3700 \pm j0.4102$ $s_{3,4}=-0.9952 \pm j1.2571$	$s_1=-1.5023+j0$ $s_{2,3}=-1.3808 \pm j0.7179$ $s_{4,5}=-0.9576 \pm j1.4711$	$s_{1,2}=-1.5716 \pm j0.3209$ $s_{3,4}=-1.3819 \pm j0.9715$ $s_{5,6}=-0.9307 \pm j1.6620$	$s_1=-1.6827+j0$ $s_{2,3}=-1.6104 \pm j0.5886$ $s_{4,5}=-1.3775 \pm j1.1904$ $s_{6,7}=-0.9089 \pm j1.8346$
Legendre		$s_1=-0.6200+j0$ $s_{2,3}=-0.3450 \pm j0.9010$	$s_{1,2}=-0.5500 \pm j0.3590$ $s_{3,4}=-0.2320 \pm j0.9460$	$s_1=-0.4680+j0$ $s_{2,3}=-0.3880 \pm j0.5890$ $s_{4,5}=-0.1540 \pm j0.9680$	$s_{1,2}=-0.4390 \pm j0.2400$ $s_{3,4}=-0.3090 \pm j0.6980$ $s_{5,6}=-0.1152 \pm j0.9780$	

$$H(j\omega) = \frac{K}{(j\omega - s_1)(j\omega - s_2)\cdots(j\omega - s_n)}$$

$$= \frac{K}{(M_1/\phi_1)(M_2/\phi_2)\cdots(M_n/\phi_n)}$$

$$= \frac{K}{M(\omega)/\phi(\omega)} \qquad (2)$$

where k is a constant and

$$M(\omega) = M_1 M_2 \ldots M_n$$

$$\phi(\omega) = \phi_1 + \phi_2 + \cdots + \phi_n$$

$$M_i/\phi_i = j\omega - s_i = j\omega - (\sigma_i + j\omega_i)$$

$$= -\sigma_i + j(\omega - \omega_i)$$

so that

$$M_i(\omega) = \sqrt{\sigma_i^2 + (\omega - \omega_i)^2}$$

and

$$\phi(\omega) = \tan^{-1}\left(\frac{\omega - \omega_i}{-\sigma_i}\right)$$

The relationship between M_i, σ_i, s_i, and ϕ_i is shown in the figure. The magnitude and phase of the transfer function $H(j\omega)$ are:

$$|H(j\omega)| = \frac{K}{M(\omega)} = \frac{K}{\prod_{i=1}^{n} \sqrt{\sigma_i^2 + (\omega - \omega_i)^2}} \qquad (3)$$

and

$$\arg[H(j\omega)] = -\phi(\omega)$$

$$= -\sum_{i=1}^{n} \tan^{-1}\left(\frac{\omega - \omega_i}{-\sigma_i}\right) \qquad (4)$$

The group delay is defined as

$$t_g = \frac{d}{d\omega}\,[-\arg[H(j\omega)]] = \frac{d\phi}{d\omega} \qquad (5)$$

The figure and Eq (2) show that the total phase at frequency ω is

$$\phi(\omega) = \sum_{i=1}^{n} \phi_i = \sum_{i=1}^{n} \tan^{-1}\left(\frac{\omega - \omega_i}{-\sigma_i}\right) \qquad (6)$$

Differentiating, we find

$$t_g(\omega) = \sum_{i=1}^{n} \frac{\sigma_i}{\sigma_i^2 + (\omega - \omega_i)^2} \qquad (7)$$

Thus, when the pole locations are known, the group delay can be found by a simple summation.

Normalization

The characteristics given in this article have common normalizations in both magnitude and frequency so that the results are comparable for the various filter types. The magnitude is normalized so that its maximum value is 1 for all values of ω. This maximum usually occurs at $\omega = 0$ so that $H(0) = 1$: in fact, the one exception to this rule is the Chebyshev filter for n odd. The characteristics are normalized in frequency so that the 3-db loss point occurs at $\omega = 1$. It should be noted that, although this normalization is usual for Butterworth and Legendre filters, it is not the one customarily used for Chebyshev, Bessel, or transitional Butterworth-Thompson filters.

Magnitude characteristics, phase characteristics, group delays, and normalized pole locations for various filter types are shown in the accompanying panels. Information has been published elsewhere on the transient responses of the Butterworth, Chebyshev, and Bessel filters [14] and of the transitional Butterworth-Thompson filter. [9] However, these sources use different frequency normalizations from the one used in this article.

The actual time delay of any filter may be found easily by dividing the normalized group delay by the cutoff frequency. Thus, if the delay of a normalized filter is t_g seconds at a particular frequency, the delay of the filter with a cutoff of f_c cps is

$$t_g/\omega_c = t_g/2\pi f_c \qquad (8)$$

For example, the normalized fifth-order Legendre filter has a group delay of $t_g = 3.99$ sec at $\omega = 0.1$ rad/sec. If this filter is built with a bandwidth of $f_c = 2$ Mc, the delay at $f = 0.2$ Mc is

$$\frac{3.99}{(2\pi)(2 \times 10^6)} = 3.18 \times 10^{-7}$$

$$= 0.318\ \mu s$$

Cited References

[1]. *Synthesis of Passive Networks*, E. A. Guillemin, John Wiley & Sons, Inc., New York (1957).

[2]. *Network Synthesis*, N. Balabanian, Prentice-Hall, Inc., Englewood Cliffs, N. J. (1958).

[3]. *Network Synthesis*, D. F. Tuttle, Jr., Vol 1, John Wiley & Sons, Inc., New York (1958).

[4]. *Network Analysis and Synthesis*, L. Weinberg, McGraw-Hill Book Co., Inc., New York (1962).

[5]. "The Phase and Envelope Delay of Butterworth and Tchebycheff Filters," H. J. Orchard, *IRE Trans. on Circuits Theory*, Vol CT-7, June 1960, pp 180, 181.

[6]. "Delay Networks Having Maximally Flat Frequency Characteristics," W. E. Thompson, *Proc. IEE*, Vol 96, Pt. C, 1949, pp 487–490.

[7]. "Networks with Maximally Flat Delay," W. E. Thompson, *Wireless Engineer*, Vol 29, October 1952, pp 256-263.

[8]. "Synthesis of Constant-Time-Delay Ladder Networks Using Bessel Polynomials," L. Storch, *Proc. IRE*, Vol 42, November 1954, pp 1666-1675.

[9]. "Analysis and Synthesis of Transitional Butterworth-Thompson Filters and Bandpass Amplifiers," Y. Peless and T. Murakami, *RCA Review*, Vol 18, March 1957, pp 60-94.

[10]. "Optimum Filters with Monotonic Response," A. Papoulis, *Proc. IRE*, Vol 46, March 1958, pp 606-609; also, Vol 47, February 1959, pp 332, 333.

[11]. "Optimum Filters of Even Orders with Monotonic Response," M. Fukada, *IRE Trans. on Circuit Theory*, Vol CT-6, No. 3, September 1959, pp 277-281.

[12]. *Network Analysis and Synthesis*, F. F. Kuo, John Wiley & Sons, Inc., New York (1962).

[13]. "Introduction to Filters," A. I. Zverev, ELECTRO-TECHNOLOGY, June 1964, pp 61-70.

[14]. "Transient Responses of Conventional Filters," K. W. Henderson and W. H. Kautz, *IRE Trans. on Circuit Theory*, Vol CT-5, No. 4, December 1958, pp 333-347; further discussion in G. A. Beck, Vol CT-8, No. 2, June 1961, pp 166, 167.

Reprinted from the AUGUST 1966 issue of ELECTRO-TECHNOLOGY

DATA DETECTION METHODS VS HEAD RESOLUTION IN DIGITAL MAGNETIC RECORDING

I. H. Graham*

ABSTRACT

The detection methods for determining the presence of a magnetically recorded bit depend on the operating conditions of the magnetic recording head and media. The Bit Density curve which best describes the head and media operating conditions is divided into four regions of interest. Methods for data detection suitable for each of the regions are listed together with block diagrams of typical detection circuits. Circuit parameters that effect correct data bit placement in a serial data stream are discussed.

I. INTRODUCTION

Theoretical treatment of data reproducibility and performance in digital magnetic recording systems has largely been focused on the magnetics of the head and media. The electronics necessary in such systems has been neglected in the literature. For many years the thrust towards higher recording densities was achieved by making proportional changes in the various system dimensions. A high resolution head signal was deemed essential for best data reproducibility. The work of Karlquist[1], Hoagland[2] and many others has resulted in a better understanding of the recording and read-back process which has permitted the design of machines with increased data density. The recent higher density machines have focused on other than the ideal high resolution head signals. The design of the electronics necessary to resolve a bit is heavily influenced by the linear density or BPI curve of the magnetic system. This paper addresses itself to the data detection methods that best determine the presence of a recorded transition as a function of the resolution of the head signals.

II. THE BIT DENSITY CURVE

A particular magnetic recording head and media combination can best be characterized by a set of performance curves. A family of saturation curves can be made by plotting the normalized head output voltage as a function of the write current magnitude for each of several different recording densities. A second and equally valuable curve is taken from the saturation curves. It is made by plotting the normalized head output voltage at a particular value of write current against the recording density. This curve is the Bit Density Curve, or BPI curve, for that head and media. Sometimes the point of ideal write current for a particular recording density is plotted instead of at a fixed write current. This takes into effect the field spreading associated with higher write currents at higher recording densities and results in a more realistic view of the head's and media's capability. The head resolution is determined from the saturation curves and is the ratio of the head output voltage peak-to-peak at the highest bit density of interest to the head output voltage peak-to-peak at the lowest bit density of interest:

$$\text{Resolution} = \frac{\text{2F output voltage pp}}{\text{1F output voltage pp}} \text{ for FM Codes}$$

$$\text{Resolution} = \frac{\text{All ones output voltage pp}}{\text{Isolated pulse output volt pp}} \text{ for NRZI}$$

The Bit Density Curve may be divided into four regions as shown in Fig. 1. Region 1 covers the higher resolution area: Region 2, the corner area; Region 3, the poor resolution area; and Region 4, the very poor resolution area.

IV. CIRCUITRY

The type of circuit utilized for detecting the presence of a recorded bit changes with the region of operation. In Region 1 the read-back signals are essentially isolated pulses. The traditional detection scheme, which dates from the early 1950s, calls for an amplitude-sensitive circuit that removes the baseline area from the signal such that only the signal peaks above some threshold voltage,

Manuscript received January 15, 1978; revised April 14, 1978.

* Memorex Corporation, San Tomas at Central Expressway, Santa Clara, California 95052.

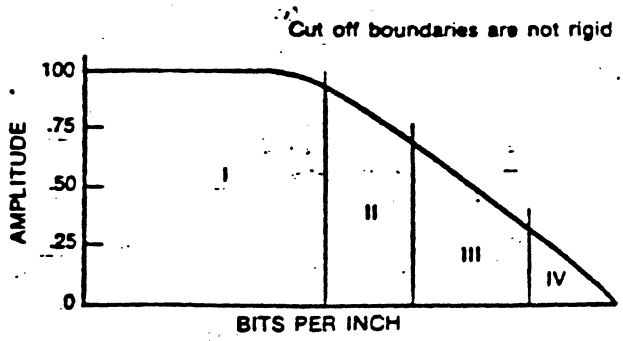


Fig. 1. Bit Density Curve

or clipping level, are passed onto a peak sensing and gating circuit. Figure 2 shows a block diagram of one such circuit. The signal peak is the actual center of the recorded bit, which is the instant of the maximum flux change recorded in the media plus some phase shift through the head circuit. In digital systems this bit position/time information is preserved in order to achieve accurate declocking of the recorded bit. The low pass filter bandwidth needs to include the 3/2 harmonic of the highest flux reversal pattern recorded. Spectrum analysis reveals the presence of this harmonic in any pattern where the read signal returns to the baseline. This bandwidth will permit minimum amplitude distortion as the 3/2 harmonic is essential to accurate peak timing. Also, the filter needs to have linear time delay characteristics within the pass band for the same reason. Butterworth, Butterworth-Thompson, and Bessel filters have been successfully used in this position. A differentiator is used to provide a baseline crossing for each signal peak. The differentiated signal is limited to provide a square wave edge corresponding to the input signal peak. The amplitude of the input signal to the differentiator should be sufficient to remove the ambiguity of the limiting circuits. The gain out of the limiter is usually made equal to that required to cause one electrical degree of the minimum expected input signal to result in a full amplitude change out of the limiter. The differentiated and limited signal contains not only the baseline crossing information resulting from signal peaks but also those resulting from all sources of noise.[5] The gate generator's function then is to block those peaks that are noise related but pass those resulting from the true signal peaks. This is done by utilizing a clipping level detector that generates a gate only when some predetermined clipping level is exceeded. The clipping level is usually expressed as a percentage of the signal amplitude; consequently automatic gain control circuits are used prior to the detection circuits. Sometimes the same purpose can be achieved by generating an automatic clipping level derived from the input amplitude. This approach is illustrated by the dashed line to the gate generator. The output pulses from the "and" of the gate and the differentiated and limited signal is processed by timing circuits that clock the bit into its correct position in the serial data stream. Phaselocked loops perform this function today. Before 1965 this function was provided by clock tracks.

In Region 3 the signal has very poor resolution. It was first used in the early 1960s with the introduction of self-clocking codes such as Phase Modulation and Frequency Modulation. These codes always provide a transition at least every bit cell time resulting in signals referred to as 1F for clock bits only and 2F for clock bits and data bits. Intersymbol interference is so great that the signal reduces essentially to sinusoids with very little third harmonic content. The concept of a baseline disappears; therefore, a clipping level approach is impractical since many head signals may not cross the baseline. The waveform is then essentially continuous. The signal can be processed by utilizing the circuit block diagram shown in Fig. 3. The circuit consists of an amplifier, low pass filter, differentiator and limiter, followed by a bidirectional single shot. No automatic gain control is required nor is any gating. The output pulse represents each signal peak. The filter bandwidth needs to be wide enough to include the 3rd harmonic of 1F as before, and linear phase is required.

an be improved by using write nmetrical declocking windows, which s despite high intersymbol interference. rately writes bits either early or late to resulting from intersymbol interference locking windows take advantage of the tion codes, where the clock bit is always aracteristics of minimum bit shift for the led on both sides by a clock bit, and ck bit if the one bit is missing. The floppy this type of detector circuit.

onsiderable 3rd harmonic content in the in a tendency for the signal to return to shouldering, and it occurs between pulses ition is not great enough to permit the in Region 1, due to some signal peaks not s too high to use the simple circuit used in drive the differentiated signal back to the crossing of the differentiated signal that ge of the recorded transition; therefore, any noise in the shoulder area can drive the signal across the baseline and generate a false bit.

One method of detecting a bit in this region requires the same block diagram as used in Region 1 (Fig. 2) with a major change to the gate generator. Here the reference relates to a percentage of the signal peak rather than a percentage of the signal amplitude around the baseline. Automatic gain control is required because amplitude is one of the criteria for detection. Another approach is to devise a variation to the circuit of Fig. 3 as shown in Fig. 4. A delay line and a "D" flip flop can be added that ignores the shoulder area noise by taking advantage of the alternate polarity signals of magnetic digital recording. The "D" input is clocked at a time less than the time to the noisy area but greater than the width of the noisy area. Thus a square wave is generated that does not contain the uncertain area resulting from the differentiated shoulder. This then may be processed as before by using a bidirectional single shot to produce an output pulse for each square wave edge. Again, write precompensation and/or unbalanced declocking windows are often necessary to maintain error performance. Pulse slimming filters such as described by Sierra[3] and Huber[4] may be used to improve the head signal resolution to that found in

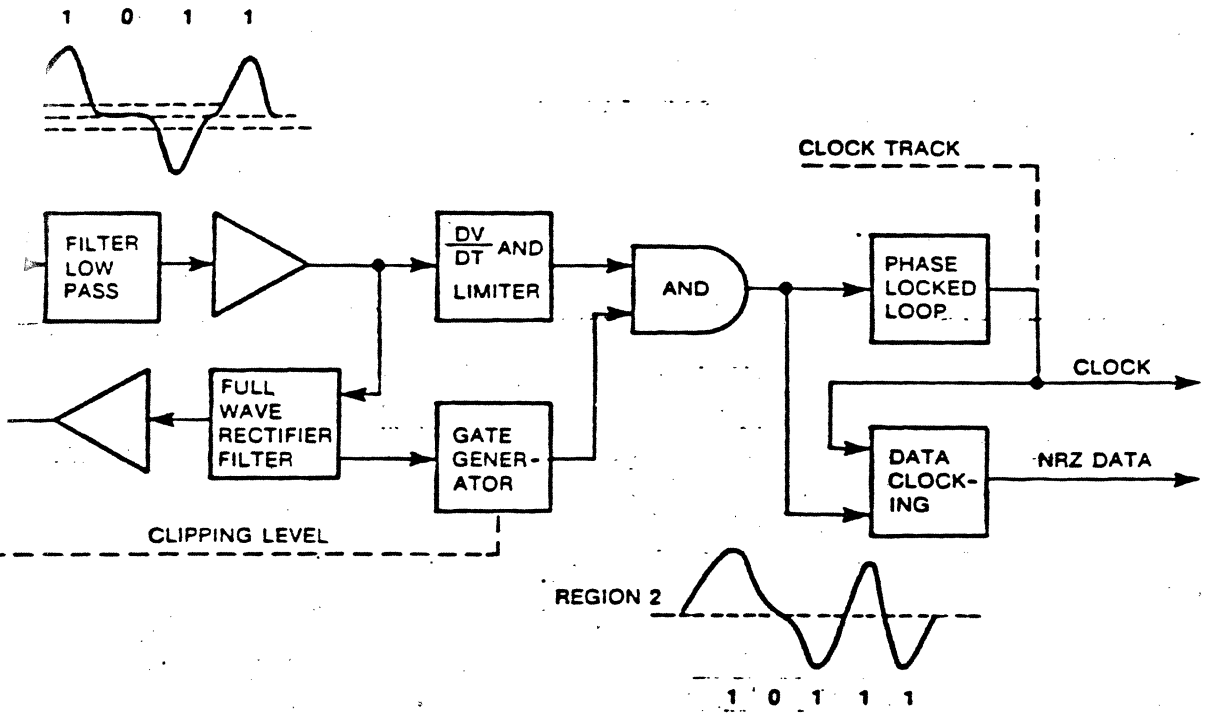


Fig. 2. Traditional Detector, Region 1 used only for high resolution recording

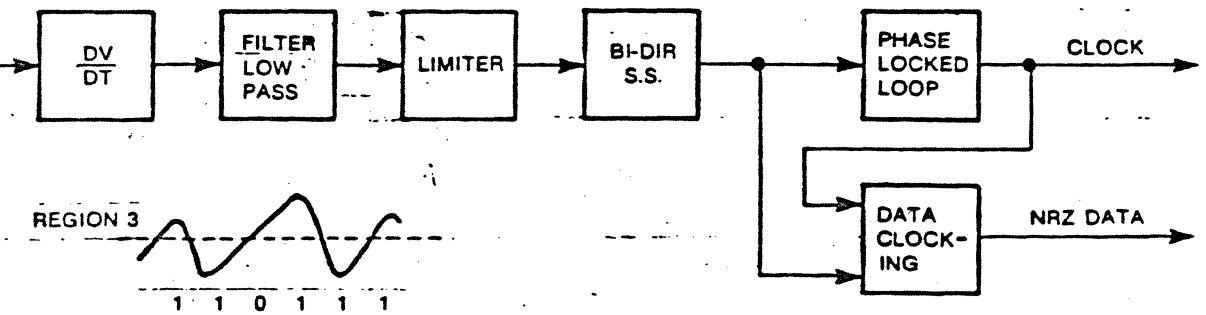


Fig. 3. Region 3 Detector

Fig. 4a. Is a block diagram of a region 2 detector.

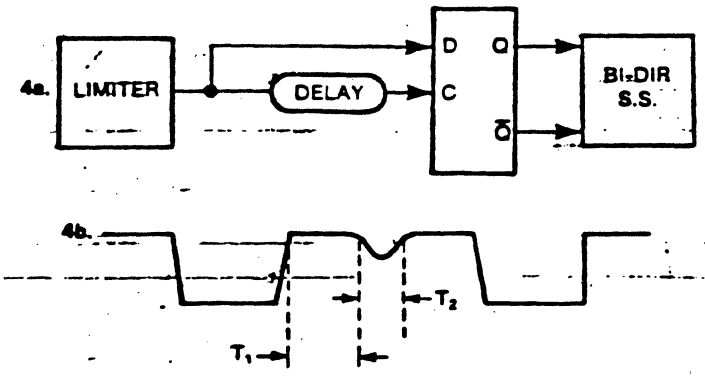


Fig. 4b. Is the differentiated and limited head signal. The area bounded within T_2 results from a differentiated shoulder in Region 2 where $T_2 < DELAY < T_1$.

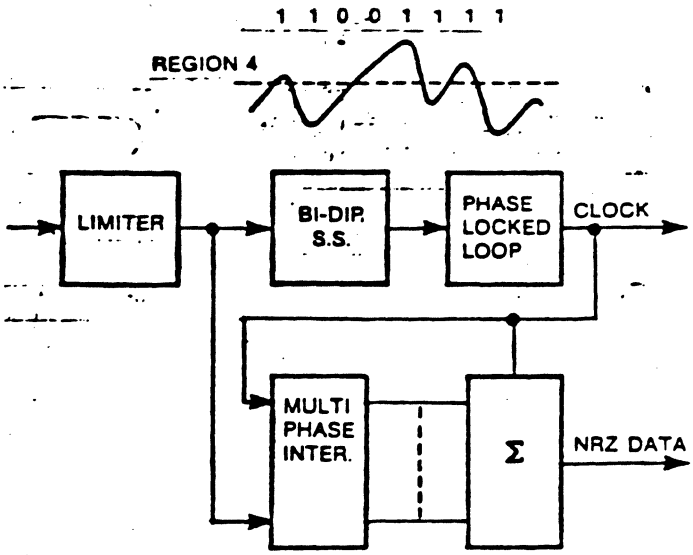


Fig. 5. Region 4 Detector used in Tape Machines Exclusively

Region 1 so that those circuits may be utilized. However, some types can increase the noise content of the signal or worsen the signal-to-noise ratio because they emphasize the higher frequency noise. Media noise often reduces the benefits of using slimming filters.

Beyond Region 3 lies an area utilized by the magnetic tape industry almost exclusively. Again intersymbol interference is so great and resolution so poor that the signal is essentially a series of sinusoids. Bit shift is so great that normal declocking circuits cannot properly position a bit into its own time slot. Voting techniques have been used, as illustrated in Fig. 5, that divide the limited differentiated signal into small time slots and then weigh the total count of each polarity against a criterion of the recording code used. Here the code becomes a part of the declocking scheme since the decision as to whether the bit was early or late depends on superposition.

V. CONCLUSIONS

From the foregoing it can be readily seen that a detection scheme may be designed for any portion of the Bit Density Curve. The skill of the recording circuit designer who knows the alternatives available to him in terms of bandwidth, signal-to-noise ratios,[5] bit shift,[5] declocking tolerances, data codes, and circuit types is just as important to a successful machine design as the head and media design. The history of data density increases is very impressive. Figures 6 and 7 show the trends over the past 20 years of digital magnetic disk recording. There have been breakthroughs in many areas of head design, air bearing design, disk surface finishes, coating formulations and processes, electronic components, positioning mechanisms and skills that have contributed to higher density data storage. The present use of a head mounted integrated circuit is the culmination of an idea that dates back to the late 1950s. It was technology before its time then, but it is a necessity today in terms of signal-to-noise ratios and bandwidths.

The future still appears to be in very high frequency, narrow track recording. Ferrite heads, which have permitted such high gains in recording density in the past, are reaching the limit of grain boundary dimensions in track width. Thin-film heads have the characteristics that will fill the need for higher track densities at substantially higher transfer rates. Again, the circuit block diagrams will need to be revised to accommodate the electrical signal characteristics from the thin-film heads.

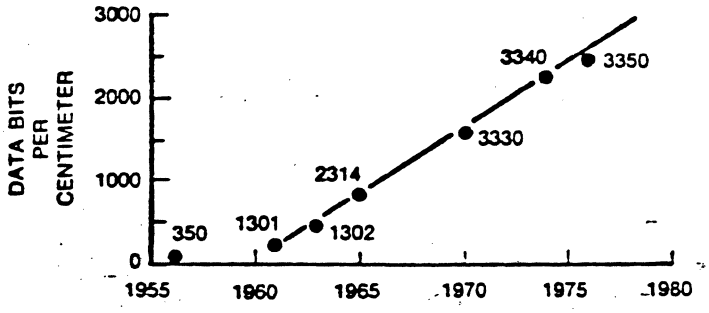


Fig. 6. Linear Bit Density

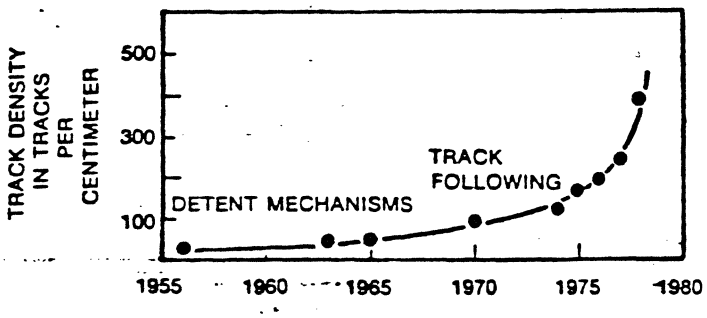


Fig. 7. Track Density

REFERENCES

1. O. Karlquist, "Calculation of the Magnetic Field in the Feromagnetic Layer of a Magnetic Drum," Transaction, Royal Institute of Technology, Stockholm, Sweden, No. 86-1954.
2. A. S. Hoagland, "Digital Magnetic Recording," Wiley, New York, 1963.
3. H. M. Sierra, "Increased Magnetic Recording Read Back Resolution by Means of a Linear Passive Network," IBM Journal of Research and Development 6, 22 (1962).
4. D. Huber, "Maximization of Lineal Recording Density," IEEE Transactions on Magnetics Vol. MAG-13, No. 5, 1208 (Sept. 1977)
5. E. R. Katz, T. G. Campbell, "Effect of Bitshift Distribution on Error Rate in Magnetic Recording," Intermag Conference 1978 (to be published).

EFFECT OF BITSHIFT DISTRIBUTION ON ERROR RATE IN MAGNETIC RECORDING

Eric R. Katz, Member, IEEE, and Thomas G. Campbell, Member, IEEE

ABSTRACT

A formalism is presented which relates the intrinsic error rate of a digital recording channel to measurable channel parameters: inter-pulse interactions and signal-to-noise ratio. Experimental data is provided to support the theoretical model. Implementation of the formalism allows clear identification of recording channel limitations and provides a method for determining the most reliable use of the channel.

INTRODUCTION

The bit error rate achieved by a digital magnetic recording system is a valuable criterion to use in evaluating its performance. The bit error rate is defined as the fraction of bits detected in a bit stream that differ from those bits actually recorded in the media. It is a raw error rate which occurs before error correction, and includes errors in clock pulses as well as data pulses. The observed bit error rate in a real system results from the combination of many contributing effects, which include noise transients, media defects, overwrite, adjacent track pick-up, and external disturbances.

This paper was presented at the International Magnetics Conference, Florence, Italy, May 1978.

The authors are with Memorex Corporation, Santa Clara, California 95052

However, even in the absence of these effects, the bit error rate can never be less than an intrinsic error rate, which is shown here to be dependent upon the bitshift distribution that results from the given implementation of the read/write head, media, and channel. This error rate is particularly significant because it identifies the best potential performance of the recording channel. This error rate has been discussed previously [1,2] in terms of the signal-to-noise ratio (SNR), but it is described here in terms of the inter-pulse interference as well as a more clearly defined SNR. By treating both of these contributions simultaneously, a more complete description of the error rate is obtained; and the importance of inter-pulse interference, which is determined by the readback pulse width, linear bit density, and the data encoding method, can be compared with the effects of SNR, which is dependent on the media and electronic noises and the recorded track width. This allows the channel performance to be optimized with respect to the various design parameters available.

THEORY

Consider the data recovery process in digital magnetic recording systems in which there is a series of bit intervals with one interval for each bit in the logic stream. These intervals are phase-locked to the average phase of the incoming signal to allow minor variations in the readback data rate. It is useful to think of each bit interval as a window in time; if a pulse is detected in a window, the bit is interpreted as a logic "one", otherwise, it is a logic "zero". In order to describe the relation between bit error rate and the system parameters, it is convenient to introduce the bitshift distribution. It is formed by taking all pulse peaks in a data stream and superposing their positions relative to their own data windows onto a single data

. The bitshift distribution is therefore a histogram of the positions he pulse peaks in the individual bit windows with the horizontal axis esenting the position in the time window; the vertical axis represents number of pulse peaks occurring at any given position with respect to the ter of the intended time window. In the ideal case of a noiseless channel no inter-pulse interactions, there is no bitshift; the bitshift distribu-n is simply a delta function located at the center of the time window. is is shown in Figure 1a. In real systems, where pulses are packed closely gether to maximize the data storage density, the effect of pulse interactions to shift the positions of various pulses from the centers of their data indows by unique discrete amounts [3]. Figure 1b shows the splitting of the itshift distribution for noninteracting pulses into several smaller peaks, ome shifted forward in time, while others are shifted backwards. This "interaction-induced" bitshift is dependent on both the pulse density (the ratio of the pulse width to the interpulse spacing) and the patterns in the logic stream. Pulse peaks with symmetric environments, as in an "all ones" pattern, will have no pattern-induced bitshift; whereas pulse peaks with asymmetric peak environments will be shifted away from the center of the data window.

In addition to "interaction-induced" bitshift, the pulse peaks will be subject to additional bitshift due to the effects of noise according to:

$$\frac{d}{dt}[S(t_o + \Delta t) + n(t)] = 0 \qquad (1)$$

where $S(t)$ is the signal waveform, t_o is the unperturbed position of the pulse peak, Δt is the bitshift, and $n(t)$ is the noise voltage as a function of time. If $S(t)$ is expanded in a Taylor series about t_o, then to first order the

bitshift is given by:

$$|\Delta t| = n'(t_0)/S_i'' \qquad (2)$$

where $n'(t_0)$ is the noise as measured after filtering and differentiation, and S_i'' is the second derivative of the signal waveform at the i^{th} pulse. Consequently, every pulse peak in the logic stream is shifted in a manner that can be described by a probability density function, P(t), which is related to $S_i''(t)$ and to P(n'), the probability distribution of the differentiated noise. Figure 1c shows the spreading of the bitshift distribution due to the "noise-induced" bitshift. The intrinsic error rate of the channel is simply the area of the tails of the bitshift distribution beyond $|T_w|$ normalized to the total area of the distribution, viz. the fraction of bits falling outside the intended window.

In cases where the noise has a Gaussian probability distribution and for times t small with respect to the pulse width, one can show that the bisthift distribution is given by:

$$P(t) = \tfrac{1}{2}\sum A_i\{\exp[-(t - \tau_{1i})^2/2\tau_{2i}^2] + \exp[-(t + \tau_{1i})^2/2\tau_{2i}^2]\} \qquad (3)$$

where the A_i represent the fraction of pulse peaks in the i^{th} local data pattern, the τ_{1i} are the bitshifts induced by inter-pulse interaction, and the τ_{2i} describe the broadening of the distribution due to the random noise according to (2). Equation (3) may also be considered to be the probability of finding any bit in the data stream displaced from the center of its data window by time t. The encoding method determines the number of terms in (3) as well as the size of the data window, and the structures of the data patterns allowed by the encoding method determine the magnitudes of the τ_{1i}. The τ_{1i} may be

d by using linear superposition of characteristic pulses, or by more
e methods [4,5]. The τ_{2i}, the RMS values of the noise-induced bit-
are found from (2), with $n'(t_0)$ replaced with the RMS value of the
Generally, the τ_{2i} will be slightly different for each local data
, as the sharpness of the pulse peaks, S_i", depends on the details of
tern.

miting value of the system bit error rate, the intrinsic error rate, is
by examining the bit stream comprised of that data pattern in which the
s undergo the greatest peak shift. In practice, this worst case pattern
ally consists of pulses whose inter-pulse spacing alternates between the
ium interval and the minimum interval allowed by the code. The bitshift
ibution for this data stream is:

$$P(t) = \tfrac{1}{2}\{\exp[-(t - \tau_1)^2/2\tau_2^2] + \exp[-(t + \tau_1)^2/2\tau_2^2]\} \qquad (4)$$

re τ_1 is the "interaction-induced" bitshift for the above worst case pattern,
τ_2 is the corresponding RMS value for the "noise induced" bitshift. The
rinsic error rate is then:

$$E = \tfrac{1}{2}[\mathrm{erfc}(x_-) + \mathrm{erfc}(x_+)] \qquad (5)$$

re $x_+ = (T_w + \tau_1)/\tau_2$, $x_- = (T_w - \tau_1)/\tau_2$, and $2T_w$ is the data window width.

the most general case the error rate may not have the erfc form, which
ulted from the assumptions of a Gaussian probability distribution due to

random noise and the retention of only three terms in the Taylor expansion of the pulse in the vicinity of the peak. In any case the error rate will be the value of the integral of the actual bitshift distribution for times larger than $|T_w|$:

$$E = \int_{-\infty}^{-T_w} P(t)dt + \int_{T_w}^{+\infty} P(t)dt. \qquad (6)$$

The bitshift distribution P(t) is found by solving (1) explicitly for τ_{2i} for each pattern allowed by the code, finding τ_{1i} for each data pattern, and by finally summing all terms as in (3) but with the proper functional dependence on time as dictated by the actual noise probability distribution. The noise probability distribution will have a Gaussian component from the random noise due to the media noise and broadband electronic noise, and may have other components from overwrite (partially erased earlier data) or adjacent track pick-up, which will generally have sinusoidal forms, as well as contributions from any other extraneous noise sources.

APPLICATION

A recording system may be evaluated directly by measuring the error rate as function of data window size. This information is obtained by using a data separator that runs at the data clocking rate and has an adjustable window width; the bits that fall outside of the chosen window are counted as errors. Fitting this data, taken at convenient window widths, to (5) determines τ_1 and τ_2; these parameters are then used with the actual data window in (5) to predict the operating error rate. The key point here is that the intrinsic error rate in most practical digital magnetic recording systems is too

e directly in a reasonable time; however, it may be determined
knowing the functional form of the error rate curve together
al form of the curve at small windows. A typical error rate
n in Figure 2 with the range of obtainable data points indicated
line and the intrinsic error rate identified as the value of the
half width (T_w) of the data window. Figure 3 shows actual
n a Memorex 3650-type R/W interface using MFM code and a data window
= 52 nsec. Two different patterns were measured and for each pattern
is made with theoretical curves that were generated from (5) with
τ_2 = 2.0 nsec being the parameters fitting the 101010...
τ_1 = 4.1 nsec and τ_2 = 1.7 nsec being the parameters fitting
.. pattern. The limitation of the recording system due to noise
en in the error rate curve for the 101010... pattern in which the
the same symmetric environment and τ_1 is consequently zero.
interaction effect is seen in the error rate curve for the 110110...
hich is the worst-case pattern for this particular code; the
uced" error rate curve tends to be displaced toward the window edge
nt equal to the "pattern-induced" bit shift. Thus, the relatively
on of the error rate curves at small window sizes is predominately
se-interaction effect, while the error rate curvature at large win-
ds on the SNR [1]. Figure 4 compares error rate curves for dif-
nal-to-noise ratios, which were obtained by displacing the head
rom the written track. The data in this figure, which was gener-
an IBM 3340-type interface having a track width of 65 μm, and a
w of $2T_w$ = 76 nsec, shows that the slope of the error rate curve
indows decreases as the SNR decreases, as expected. This is
monstrated in Figure 5, which gives the value of τ_2 describing the
curve as function of the displacement of the head from the written

track. The dotted line shows the expected variation of τ_2 with the assumption that the head has a side-reading width of 5 μm.

While the experimental error rate curve can be used to determine the channel resolution and SNR, it also identifies other error mechanisms. Figures 6 and 7 illustrate the types of error rate dependences typically found for hard media defects and intermittent errors, respectively. One type of intermittent error is the "soft" media defect, where the magnitude of the signal loss is not sufficient to cause a complete bit dropout. However, the local degradation of the SNR results in a local broad error distribution that contains only the small number of bits affected. This distribution is shown schematically as the dotted line in Figure 7. Thus the shape alone of the error rate curve often distinguishes a deficiency in channel design from media defects or other intermittent errors.

Since the error rate curve identifies the recording channel limitations, the optimum implementation of the channel is facilitated. For example, the relative utility of different digital codes is evaluated directly by comparing the experimental error rate curves for the worst case data patterns for each code. The design of the recording channel may be optimized by constructing theoretical error rate curves for different values of pulse density and SNR. From these curves the error rates expected for various digital codes are compared and trade-offs are then made between linear bit density (pattern-induced bitshift) and track density (noise-induced bitshift) in maximizing the areal bit density for a given error rate. Likewise, the benefit of pulse narrowing equalization is estimated by inserting the reduced pulse width into the calculation for pattern-induced bitshift, and

using the lower SNR in the determination of the noise-induced bitshift, and then computing the combined effect on the overall intrinsic error rate. Consequently, improvements required in a recording channel can be estimated quantitatively with (5) and the resultant effects of pulse narrowing or a SNR enhancement are then verified by measuring the error rate curve for the modified channel.

CONCLUSIONS

An experimental technique for analyzing the recording performance of a digital magnetic recording channel has been described; the technique can determine very low error rates by extrapolation and also can detect anomalies such as media defects and recurring noise transients. A theoretical model has been presented to support the experimental method and this model can be used to predict the most effective utilization of a channel. The basis for comparison has been the channel error rate which is due to the combined effects of inter-pulse interactions and signal-to-noise ratio; and because this error rate exists for all recording channels, it has been called the "intrinsic error rate."

ACKNOWLEDGEMENTS

The authors wish to recognize Mr. Michael Monett for his invaluable contributions in the design and fabrication of the original variable-window data separator. The authors are also grateful to Mr. Michael Hammer for assistance in obtaining data.

REFERENCES

1. G. V. Kelley and G. L. Blank, "Track Density Considerations in Magnetic Recording," IEEE Trans. Magnetics MAG-10, 910-912 (1974).

2. Y. Tahara, Y. Miura and Y. Ikeda, "Peak Shift Caused by Gaussian Noise in Digital Magnetic Recording," Electronics and Communications in Japan, 59-C (10), 77-86 (1976).

3. J. C. Mallinson and C. W. Steele, "Theory of Linear Superposition in Tape Recording," IEEE Trans. Magnetics MAG-5, 886-890 (1969).

4. S. I. Iwasaki and T. Suzuki, "Dynamical Interpretation of Magnetic Recording Process," IEEE Trans. Magnetics MAG-4, 269-276 (1968).

5. Chao Chi and D. E. Speliotis, "Dynamic Self-Consistent Iterative Simulation of High Bit Density Digital Magnetic Recording," IEEE Trans. Magnetics MAG-10, 765-768 (1974).

JURE CAPTIONS

g. 1a Bitshift Distribution for a data stream in a noiseless channel with no pulse interactions

ig. 1b Bitshift Distribution for a data stream in a noiseless channel with pulse interactions.

ig. 1c Bitshift Distribution with pulse interactions in a noisy channel. The intrinsic error rate is the area of the tails beyond $|T_w|$ normalized to the total area of the distribution.

Fig. 2 A typical error rate curve with the range of experimental points and intrinsic error rate illustrated.

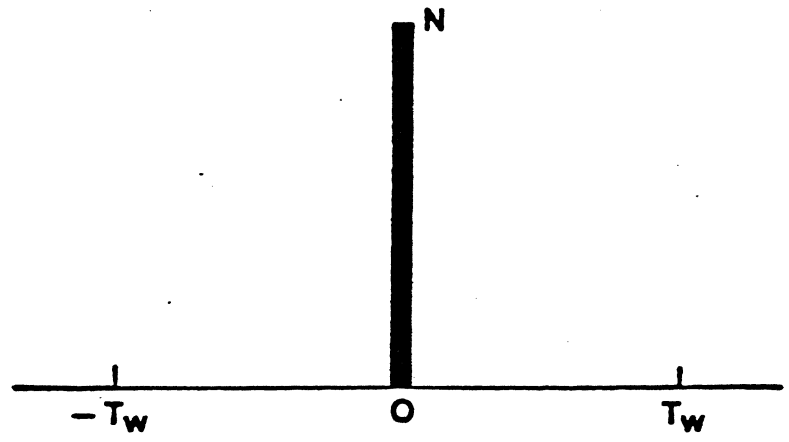
Fig. 3 Error Rate Curves for two patterns taken with a Memorex 3650-type R/W interface.

Fig. 4 Comparison of error rate curves with differing signal-to-noise ratio. Data was obtained with an IBM 3340-type R/W interface.

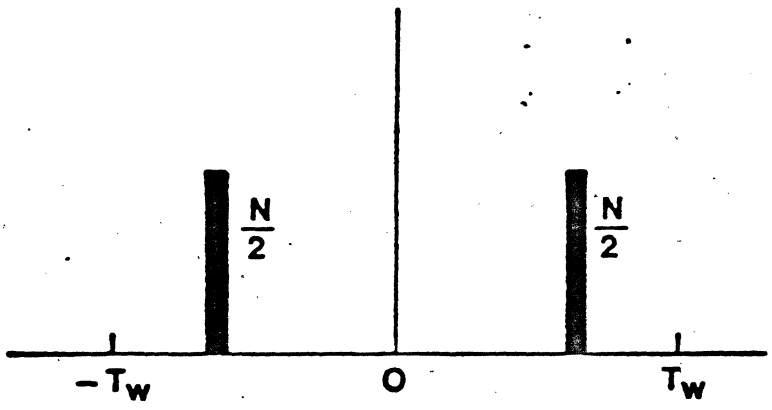
Fig. 5 Increase in τ_2 calculated from least-squares fit to data obtained with displacement of the head from the written track. The R/W interface is the same as in Fig. 4.

Fig. 6 Expected error rate curve when a hard media defect is present on the track. Solid curve is the observed error rate curve; dashed curves represent the two components comprising the overall error rate.

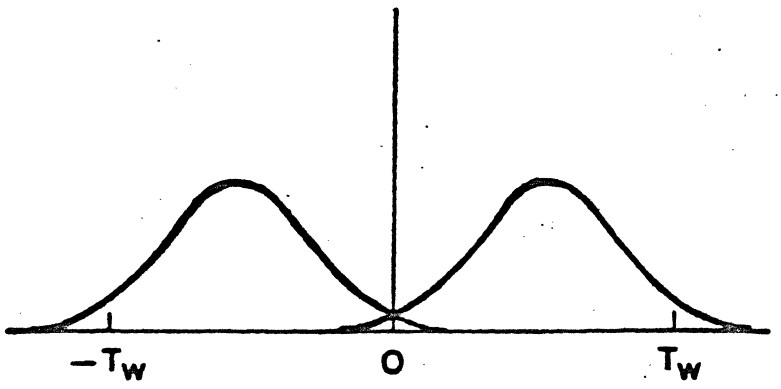
Fig. 7 Expected error rate curve for an intermittent error, such as a "soft" media defect.



a

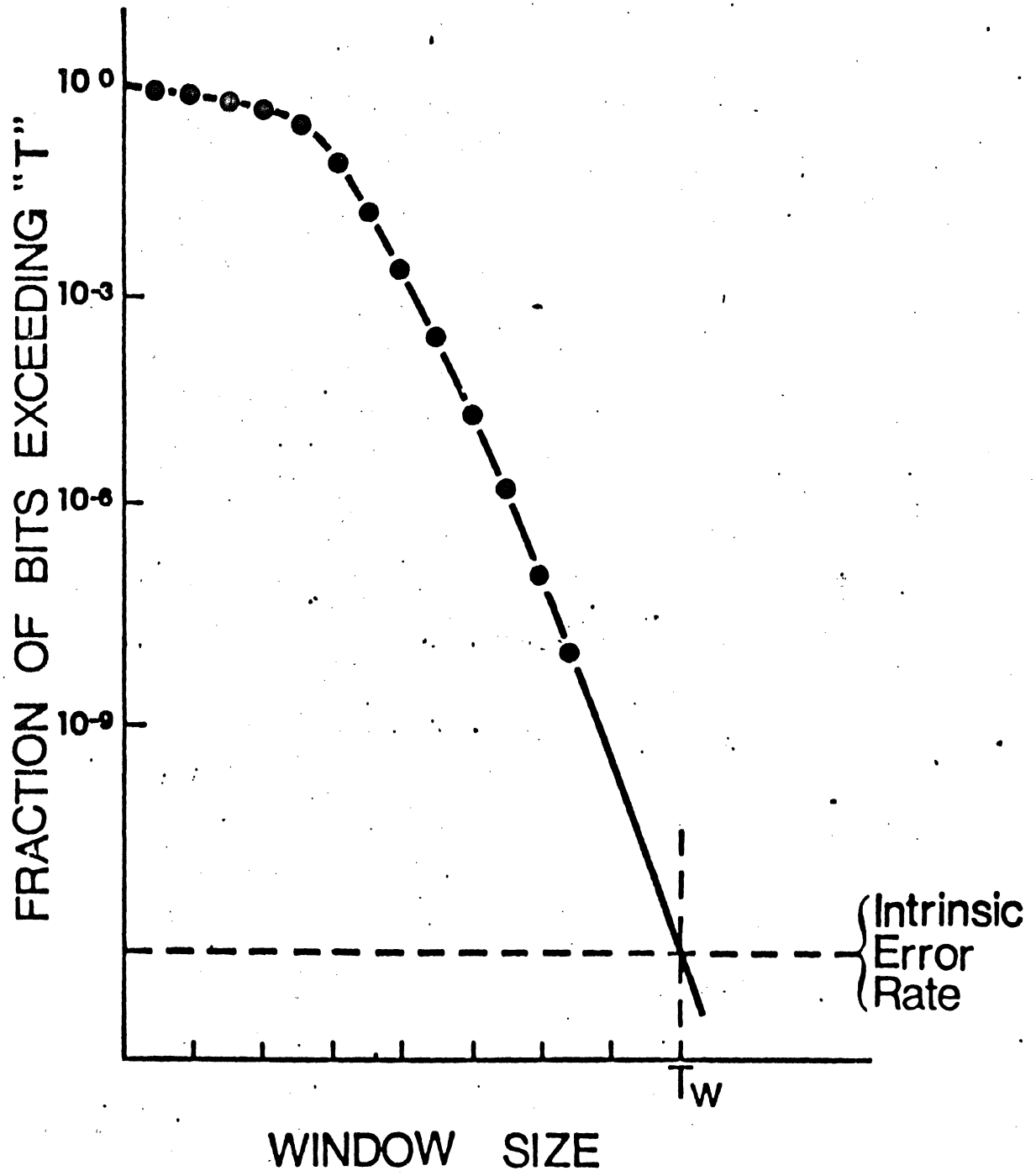


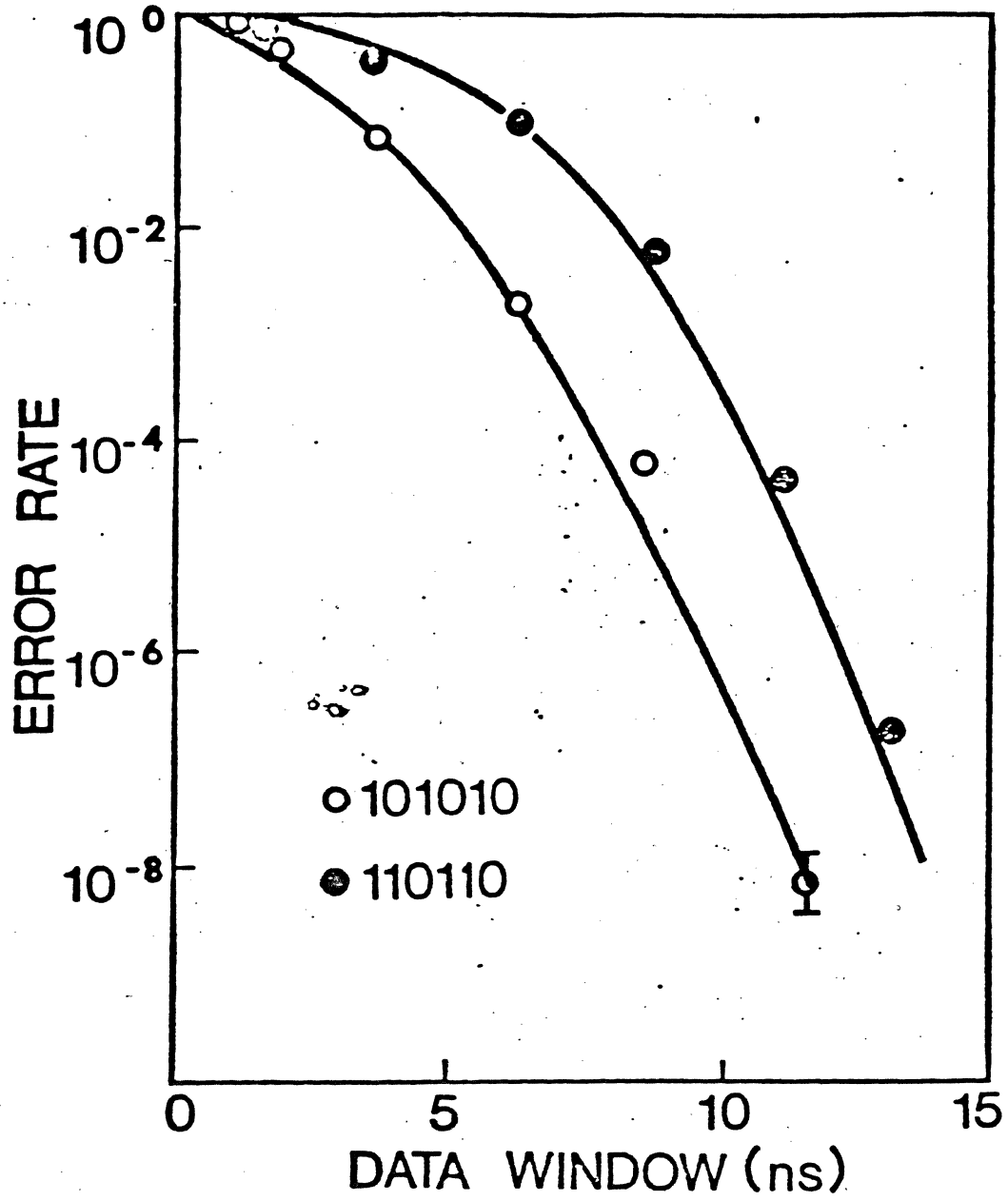
b

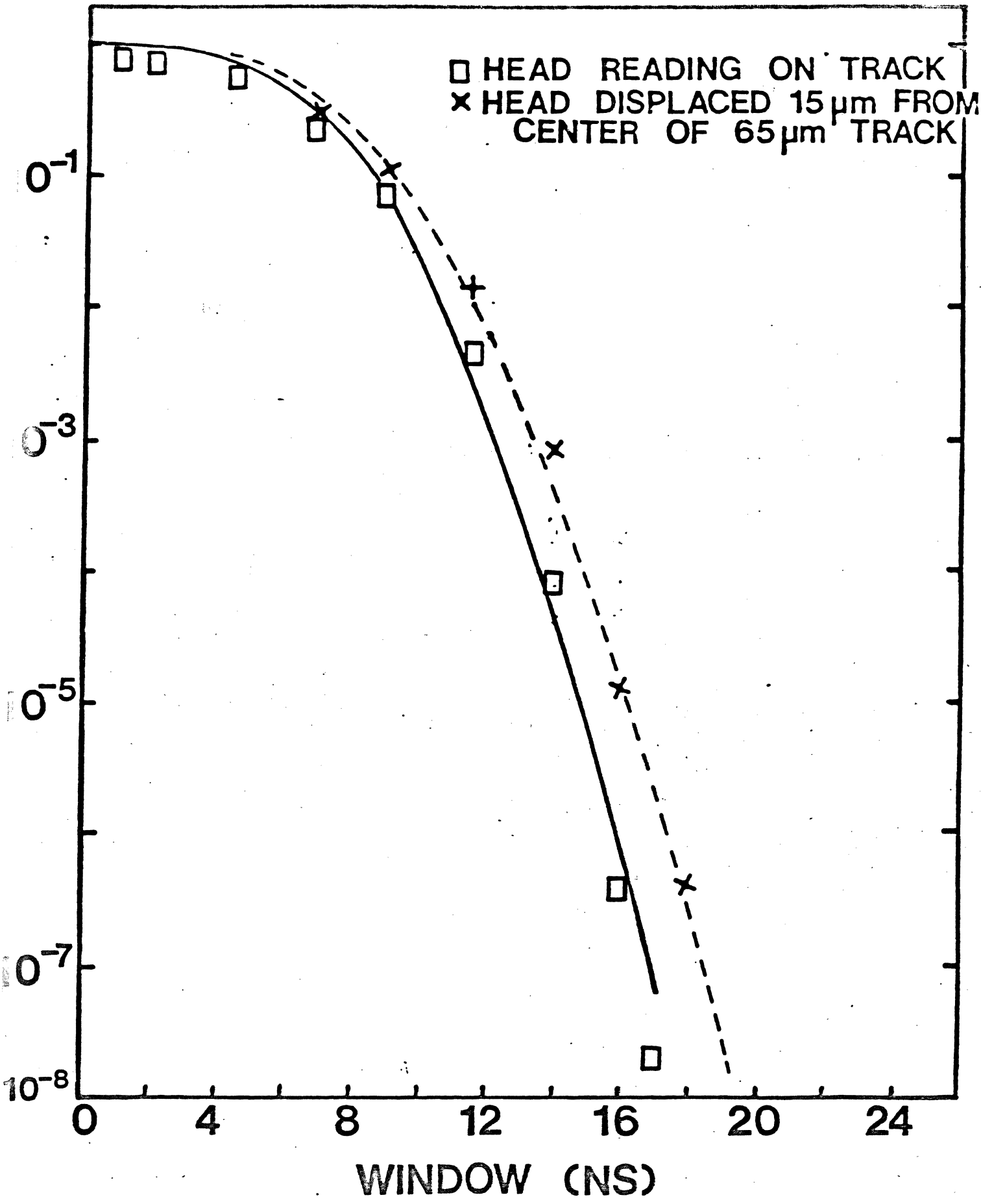


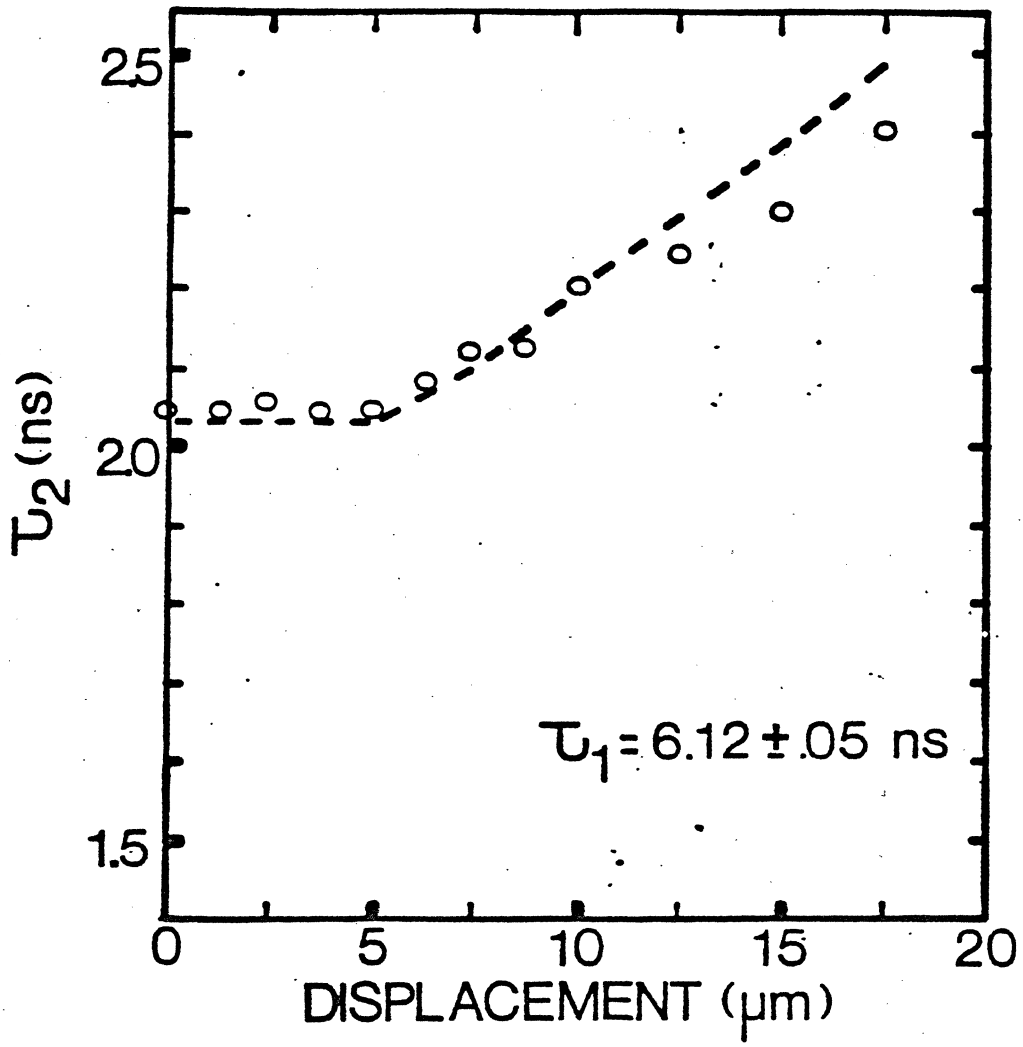
c

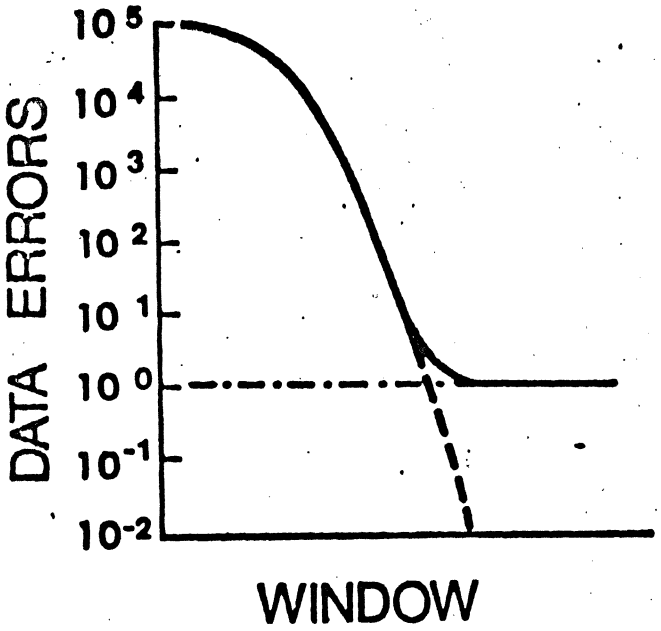
Fig 1

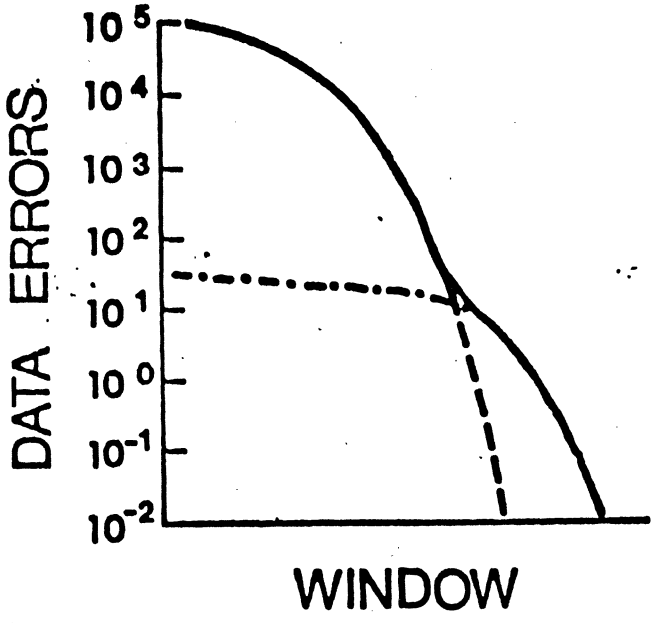

FRACTION OF BITS EXCEEDING "T"
10 0
10-3
10-6
10-9
Intrinsic
Error
Rate
Tw
WINDOW SIZE


10 0
10-2
10-4
10-6
10-8
ERROR RATE
101010
110110
0
5
10
15
DATA WINDOW (ns)


HEAD READING ON TRACK
HEAD DISPLACED 15 μm FROM CENTER OF 65 μm TRACK
10^{-1}
10^{-3}
10^{-5}
10^{-7}
10^{-8}
0
4
8
12
16
20
24
WINDOW (NS)


2.5
2.0
1.5
τ_2 (ns)
$\tau_1 = 6.12 \pm .05$ ns
0
5
10
15
20
DISPLACEMENT (μm)


DATA ERRORS
10 5
10 4
10 3
10 2
10 1
10 0
10-1
10-2
WINDOW


DATA ERRORS
10 5
10 4
10 3
10 2
10 1
10 0
10-1
10-2
WINDOW

DATA COMPRESSION AND BLOCK CODES

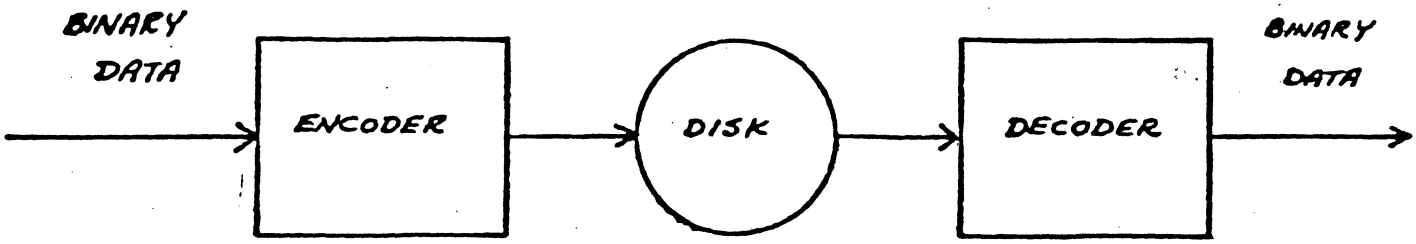
I. INTRODUCTION

II. EXISTENCE OF CODES

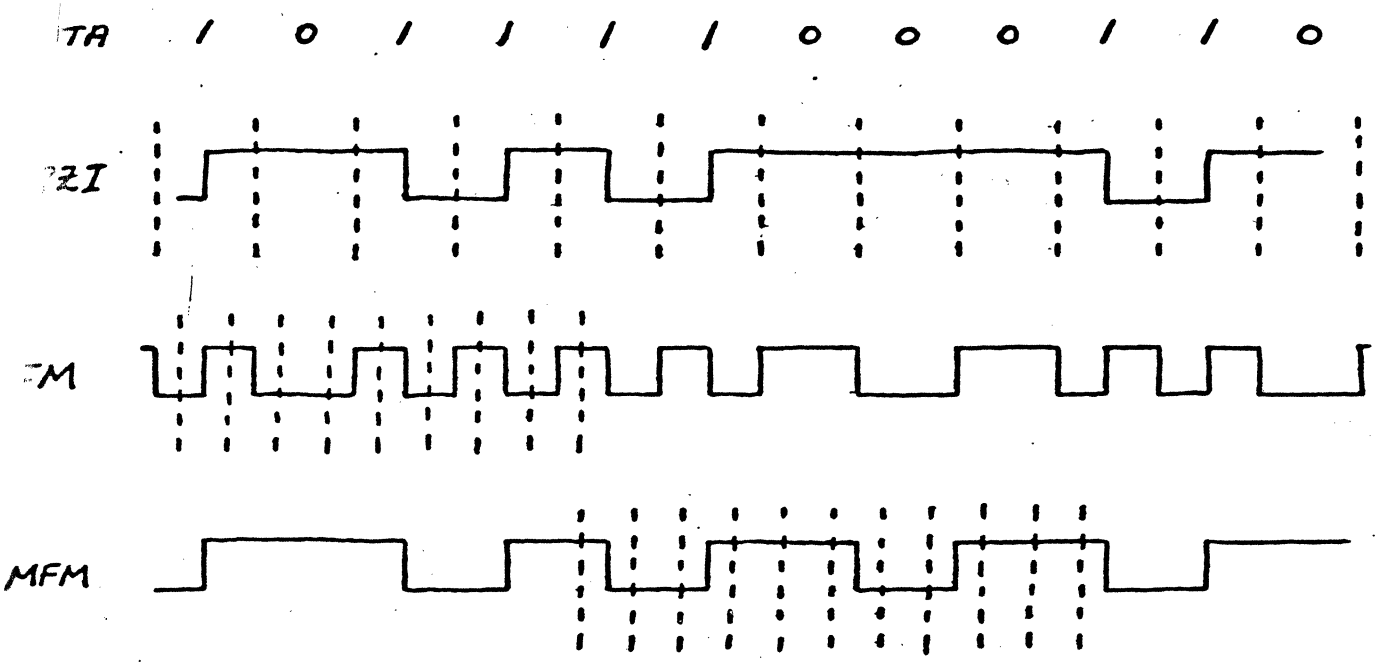
III. APPLICATION OF CODES

T. G. CAMPBELL

FEBRUARY 15, 1978


BINARY
DATA
ENCODER
DISK
DECODER
BINARY
DATA

DIGITAL ENCODING METHODS



EXISTENCE OF CODES

A. CONVENTIONAL BLOCK CODES

n	number of cells in code word
d	minimum number of zeros between adjacent transitions
m	length of binary data word
2^m	number of data words of length m

$CW(n,d)$ number of code words of length n with restriction d

$$CW(n,d) = \sum_{i=1}^{R_{max}} \frac{(n-di)!}{i!\,(n-(d+1)i)!}$$

$$R_{max} \leq \frac{n}{d+1}$$

DR density ratio

$$DR = \frac{m}{n}(d+1)$$

where $m \leq \log_2 [CW(n,d)]$

$n = 10$

$d = 2$

	1	2	3	4	5	6	7	8	9	10	
α_1	1	0	0	0	0	0	0	0	0	0	C_1
α_2	0	1	0	0	0	0	0	0	0	0	
α_3	0	0	1	0	0	0	0	0	0	0	
α_4	0	0	0	1	0	0	0	0	0	0	
α_5	0	0	0	0	1	0	0	0	0	0	
α_6	0	0	0	0	0	1	0	0	0	0	
α_7	0	0	0	0	0	0	1	0	0	0	
α_8	0	0	0	0	0	0	0	1	0	0	
α_9	1	0	0	1	0	0	0	0	0	0	C_2
α_{10}	1	0	0	0	1	0	0	0	0	0	
α_{11}	1	0	0	0	0	1	0	0	0	0	
α_{12}	1	0	0	0	0	0	1	0	0	0	
α_{13}	1	0	0	0	0	0	0	1	0	0	
α_{14}	0	1	0	0	1	0	0	0	0	0	
α_{15}	0	1	0	0	0	1	0	0	0	0	
α_{16}	0	1	0	0	0	0	1	0	0	0	
α_{17}	0	1	0	0	0	0	0	1	0	0	
α_{18}	0	0	1	0	0	1	0	0	0	0	
α_{19}	0	0	1	0	0	0	1	0	0	0	
α_{20}	0	0	1	0	0	0	0	1	0	0	
α_{21}	0	0	0	1	0	0	1	0	0	0	
α_{22}	0	0	0	1	0	0	0	1	0	0	
α_{23}	0	0	0	0	1	0	0	1	0	0	
α_{24}	1	0	0	1	0	0	1	0	0	0	C_3
α_{25}	1	0	0	1	0	0	0	1	0	0	
α_{26}	1	0	0	0	1	0	0	1	0	0	
α_{27}	0	1	0	0	1	0	0	1	0	0	

BLOCK CODE POSSIBILITIES FOR USE IN MAGNETIC RECORDING ENCODING CIRCUITRY

d	R_{MAX}	C_1	C_2	C_3	C_4	C_5	C_6	CW	m	DR_{MAX}
1	2	3	1	–	–	–	–	4	2	1.00
1	2	4	3	–	–	–	–	7	2	0.80
1	3	5	6	1	–	–	–	12	3	1.00
1	3	6	10	4	–	–	–	20	4	1.143
1	4	7	15	10	1	–	–	33	5	1.25
1	4	8	21	20	5	–	–	54	5	1.111
1	5	9	28	35	15	1	–	88	6	1.20
1	5	10	36	56	35	6	–	143	7	1.273
1	6	11	45	84	70	21	1	232	7	1.167
2	1	2	–	–	–	–	–	2	1	0.75
2	1	3	–	–	–	–	–	3	1	0.60
2	2	4	1	–	–	–	–	5	2	1.00
2	2	5	3	–	–	–	–	8	3	1.286
2	2	6	6	–	–	–	–	12	3	1.125
2	3	7	10	1	–	–	–	18	4	1.333
2	3	8	15	4	–	–	–	27	4	1.20
2	3	9	21	10	–	–	–	40	5	1.364
2	4	10	28	20	1	–	–	59	5	1.20
3	1	1	–	–	–	–	–	1	–	–
3	1	2	–	–	–	–	–	2	1	0.80
3	1	3	–	–	–	–	–	3	1	0.66
3	1	4	–	–	–	–	–	4	2	1.143
3	2	5	1	–	–	–	–	6	2	1.00
3	2	6	3	–	–	–	–	9	3	1.333
3	2	7	6	–	–	–	–	13	3	1.20
3	2	8	10	–	–	–	–	18	4	1.455
3	3	9	15	1	–	–	–	25	4	1.333
3	3	10	21	4	–	–	–	35	5	1.54

n	d	R_{max}	c_1	c_2	c_3	CW	m	DR_{max}
5	4	1	1	–	–	1	–	–
6	4	1	2	–	–	2	1	.833
7	4	1	3	–	–	3	1	.714
8	4	1	4	–	–	4	2	1.25
9	4	1	5	–	–	5	2	1.11
10	4	2	6	1	–	7	2	1.00
11	4	2	7	3	–	10	3	1.364
12	4	2	8	6	–	14	3	1.25
13	4	2	9	10	–	19	4	1.54
14	4	2	10	15	–	25	4	1.43
15	4	3	11	21	1	33	5	1.67
16	4	3	12	28	4	44	5	1.56
6	5	1	1	–	–	1	–	–
7	5	1	2	–	–	2	1	.86
8	5	1	3	–	–	3	1	.75
9	5	1	4	–	–	4	2	1.33
10	5	1	5	–	–	5	2	1.20
11	5	1	6	–	–	6	2	1.09
12	5	2	7	1	–	8	3	1.50
13	5	2	8	3	–	11	3	1.38
14	5	2	9	6	–	15	3	1.29
15	5	2	10	10	–	20	4	1.60
16	5	2	11	15	–	26	4	1.50

B. BLOCK CODES WITH MERGING

$CW'(n,d)$ additional code words that can be formed when merging is allowed

$d \geq 2$

	1	2	3	4	5	6	
β_1	1	0	0	0	0	0	
β_2	0	1	0	0	0	0	
β_3	0	0	1	0	0	0	$CW(6,2) = 4+1 = 5$
β_4	0	0	0	1	0	0	
β_5	1	0	0	1	0	0	
β_6	0	0	0	0	1	0	
β_7	0	1	0	0	1	0	$CW'(6,2) = 1+2 = 3$
β_8	1	0	0	0	1	0	

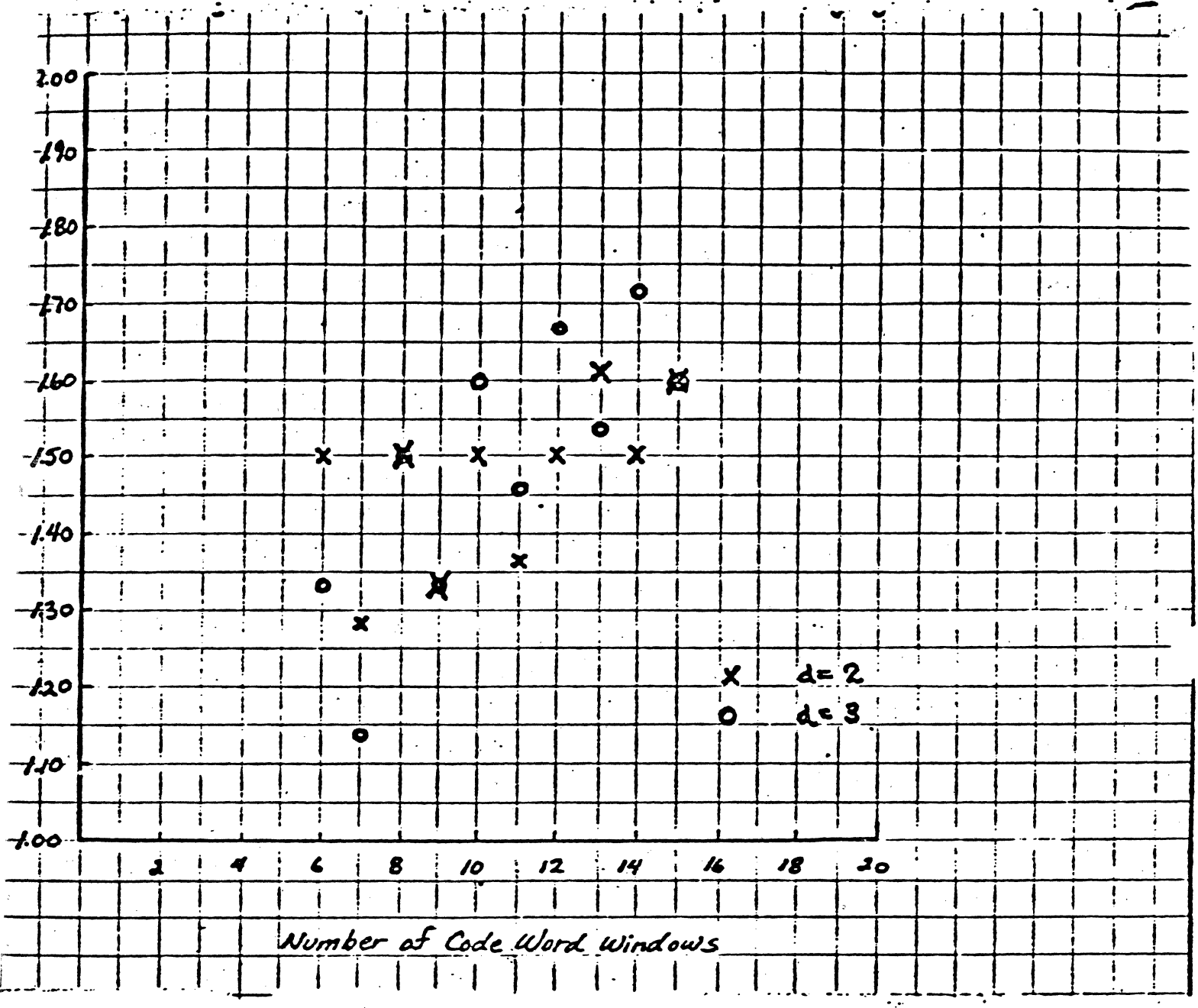
$$CW'(n,d) = \sum_{j=0}^{R'} \frac{(n-d-dj)!}{j!\,(n-d-(d+1)j)!}$$

$$R' \leq \frac{n-d}{d+1}$$

DATA WORD	INFLUENCE OF ADJACENT WORDS		TRANSITION POSITIONS						
	Preceding	Following	P_6'	P_1	P_2	P_3	P_4	P_5	P_6
β_1	1	X	1	0	0	0	0	0	0
β_1	0	X	0	1	0	0	0	0	0
β_2	X	X	0	0	1	0	0	0	0
β_3	X	X	0	0	0	1	0	0	0
β_4	1	X	0	0	0	0	1	0	0
β_5	0	X	1	0	0	0	1	0	0
β_5	X	0	0	1	0	0	1	0	0
β_6	X	1	0	0	0	0	0	1	0
β_6	X	0	0	0	0	0	0	0	1
β_7	X	1	0	0	1	0	0	1	0
β_7	0	0	0	0	1	0	0	0	1
β_8	0	1	0	1	0	0	0	1	0
β_8	1	0	0	1	0	0	0	0	1
β_8	1	1	1	0	0	0	0	1	0
β_8			1	0	0	0	0	0	1

n	d	R_{max}	c_1	c_2	c_3	c_4	c_5	CW	$R'+1$	c_1'	c_2'	c_3'	c_4'	c_5'	CW'	$CW+CW'$	m	DR_{max}
6	2	2	4	1	–	–	–	5	2	1	2	–	–	–	3	8	3	1.500
7	2	2	5	3	–	–	–	8	2	1	3	–	–	–	4	12	3	1.286
8	2	2	6	6	–	–	–	12	3	1	4	1	–	–	6	18	4	1.500
9	2	3	7	10	1	–	–	18	3	1	5	3	–	–	9	27	4	1.333
10	2	3	8	15	4	–	–	27	3	1	6	6	–	–	13	40	5	1.500
11	2	3	9	21	10	–	–	40	4	1	7	10	1	–	19	59	5	1.364
12	2	4	10	28	20	1	–	59	4	1	8	15	4	–	28	87	6	1.500
13	2	4	11	36	35	5	–	87	4	1	9	21	10	–	41	128	7	1.615
14	2	4	12	45	56	15	–	128	5	1	10	28	20	1	60	188	7	1.500
15	2	5	13	55	84	35	1	188	5	1	11	36	35	5	88	276	8	1.600
6	3	1	3	–	–	–	–	3	1	1	–	–	–	–	1	4	2	1.333
7	3	1	4	–	–	–	–	4	2	1	1	–	–	–	2	6	2	1.143
8	3	2	5	1	–	–	–	6	2	1	2	–	–	–	3	9	3	1.500
9	3	2	6	3	–	–	–	9	2	1	3	–	–	–	4	13	3	1.333
10	3	2	7	6	–	–	–	13	2	1	4	–	–	–	5	18	4	1.600
11	3	2	8	10	–	–	–	18	3	1	5	1	–	–	7	25	4	1.455
12	3	3	9	15	1	–	–	25	3	1	6	3	–	–	10	35	5	1.667
13	3	3	10	21	4	–	–	35	3	1	7	6	–	–	14	49	5	1.538
14	3	3	11	28	10	–	–	49	3	1	8	10	–	–	19	68	6	1.714
15	3	3	12	36	20	–	–	68	4	1	9	15	1	–	26	94	6	1.600

n	d	R_{MAX}	C_1	C_2	C_3	C_4	C_5	CW	R'_{+1}	C'_1	C'_2	C'_3	C'_4	C'_5	CW'	$CW+CW'$	m	DR_{MAX}
9	4	1	5	-	-	-	-	5	2	1	1	-	-	-	2	7	2	1.111
10	4	2	6	1	-	-	-	7	2	1	2	-	-	-	3	10	3	1.500
11	4	2	7	3	-	-	-	10	2	1	3	-	-	-	4	14	3	1.364
12	4	2	8	6	-	-	-	14	2	1	4	-	-	-	5	19	4	1.667
13	4	2	9	10	-	-	-	19	2	1	5	-	-	-	6	25	4	1.538
14	4	2	10	15	-	-	-	25	3	1	6	1	-	-	8	33	5	1.786
15	4	3	11	21	1	-	-	33	3	1	7	3	-	-	11	44	5	1.667
9	5	1	4	-	-	-	-	4	1	1	-	-	-	-	1	5	2	1.333
10	5	1	5	-	-	-	-	5	1	1	-	-	-	-	1	6	2	1.200
11	5	1	6	-	-	-	-	6	2	1	1	-	-	-	2	8	3	1.636
12	5	2	7	1	-	-	-	8	2	1	2	-	-	-	3	11	3	1.500
13	5	2	8	3	-	-	-	11	2	1	3	-	-	-	4	15	3	1.384
14	5	2	9	6	-	-	-	15	2	1	4	-	-	-	5	20	4	1.714
15	5	2	10	10	-	-	-	20	2	1	5	-	-	-	6	26	4	1.600
16	5	2	11	15	-	-	-	26	2	1	6	-	-	-	7	33	5	1.875


200
190
180
170
160
150
140
130
120
110
100
2
4
6
8
10
12
14
16
18
20
d = 2
d = 3
Number of Code Word Windows

n	d	T_{min}	T_{max}	Bit Rate	Window Size	Clock Rate	DR
2	1	T	2T	$1/T$	$T/2$	$2/T$	1.00
6	2	T	4T	$3/2T$	$T/3$	$3/T$	1.50
8	1	T	7T	$5/4T$	$T/2$	$2/T$	1.25
8	2	T	$\frac{16T}{3}$	$3/2T$	$T/3$	$3/T$	1.50
8	3	T	4T	$3/2T$	$T/4$	$4/T$	1.50
10	5	T	$\frac{11T}{3}$	$9/5T$	$T/6$	$6/T$	1.80
15	5	T	$\frac{16T}{3}$	$2/T$	$T/6$	$6/T$	2.00

APPLICATION OF CODES

Definitions

PW_{50}	Width at half maximum of isolated pulse
T	Bit to bit spacing of binary data (Bit rate $\equiv \frac{1}{T}$)
T_W	Window half width
T_P	Pattern-induced bitshift for "worst case" pattern
T_S	Noise-induced bitshift
T_{MI}	Window Margin to 10^{-10} error rate

$$T_{MI} \approx T_W - (6.5\,T_S + T_P)$$

For 3650:

PW_{50} (ns)	T (ns)	T_W (ns)	T_P (ns)	T_S (ns)	T_{MI} (ns)
100	104	26	2.9	1.94	9.5

$$\frac{PW_{50}}{T} \approx 1.00$$

More Definitions

$$\frac{PW_{50}\ (CODED)}{PW_{50}\ (MFM)} \equiv \lambda$$

$$\frac{PW_{50}\ (CODED)}{(d+1)\,T_c} \equiv \kappa = \text{actual pulse density on disk}$$

$$T_c = \text{code window size}$$

$$\text{Density Ratio} = \frac{\text{length of data word}}{\text{length of code word}} = \frac{mT}{nT_c}$$

$$= \frac{m}{n}(d+1)\frac{\kappa}{\lambda}$$

λ	κ	d	T_P (ns)	T_S (ns)	T_W (ns)	T_{MI} (ns)
1.0	1.0	1	3.7	1.94	25.0	8.69
1.0	1.0	2	3.7	1.94	16.66	0.36
1.0	1.0	3	3.7	1.94	12.50	-3.81
1.0	1.0	4	3.7	1.94	10.0	***
1.0	1.0	5	3.7	1.94	8.33	***
1.0	0.9	1	3.03	1.82	27.8	12.94
1.0	0.9	2	3.03	1.82	18.5	3.64
1.0	0.9	3	3.03	1.82	13.9	-0.96
1.0	0.9	4	3.03	1.82	11.11	***
1.0	0.9	5	3.03	1.82	9.26	***
1.0	0.8	1	2.41	1.71	31.25	17.72
1.0	0.8	2	2.41	1.71	20.8	7.27
1.0	0.8	3	2.41	1.71	15.6	2.10
1.0	0.8	4	2.41	1.71	12.5	-1.03
1.0	0.8	5	2.41	1.71	10.4	***
1.0	0.7	1	1.84	1.61	35.75	23.45
1.0	0.7	2	1.84	1.61	23.8	11.5
1.0	0.7	3	1.84	1.61	17.9	5.6
1.0	0.7	4	1.84	1.61	14.3	2.0
1.0	0.7	5	1.84	1.61	11.9	-0.4
1.0	0.5	1	1.0	1.45	50.0	39.6
1.0	0.5	2	1.0	1.45	33.3	22.9
1.0	0.5	3	1.0	1.45	25.0	14.6
1.0	0.5	4	1.0	1.45	20.0	9.6
1.0	0.5	5	1.0	1.45	16.6	6.2

λ	K	d	T_P (ns)	T_S (ns)	T_W (ns)	T_{MI} (ns)
0.9	0.9	1	2.73	1.82	25.0	10.44
0.9	0.9	2	2.73	1.82	16.66	2.10
0.9	0.9	3	2.73	1.82	12.5	-2.06
0.9	0.9	4	2.73	1.82	10.0	***
0.9	0.9	5	2.73	1.82	8.33	***
0.8	0.8	1	1.93	1.71	25.0	11.95
0.8	0.8	2	1.93	1.71	16.66	3.61
0.8	0.8	3	1.93	1.71	12.5	-0.55
0.8	0.8	4	1.93	1.71	10.0	***
0.8	0.8	5	1.93	1.71	8.33	***
0.7	0.7	1	1.29	1.61	25.0	13.24
0.7	0.7	2	1.29	1.61	16.66	4.90
0.7	0.7	3	1.29	1.61	12.5	0.74
0.7	0.7	4	1.29	1.61	10.0	-1.76
0.7	0.7	5	1.29	1.61	8.33	***
0.5	0.5	1	0.5	1.45	25.0	15.07
0.5	0.5	2	0.5	1.45	16.66	6.73
0.5	0.5	3	0.5	1.45	12.5	2.57
0.5	0.5	4	0.5	1.45	10.0	0.07
0.5	0.5	5	0.5	1.45	8.33	-1.6

CONCLUSIONS

1. Many codes exist with density ratio greater than, or equal to 1.50

2. Implementation of any data compression code with $d > 2$ would be very difficult

3. Implementation of a data compression code with $d = 2$ would require pulse equalization

4. Implementation of a data compression code with $d = 1$ would not provide attractive density enhancement

Code Performance and Head/Media Interface

R.D. FISHER, *Member IEEE*, and J.J. NEWMAN, *Member IEEE*

:T— The limitations of selected run length limited codes relative to ical recording density at given error rates are examined with a d/disc system interface. This paper presents a method for the of run length limited codes on the basis of error rate as a function m linear density. Experimental measurements of the intrinsic error ction of the maximum linear density were utilized with theoretical stics of the code to determine the practical data density limitations. upon the system configuration, the effectiveness of the code can either by the noise characteristics and/or pattern induced peak 3% increase in linear density at an error rate of 10^{-10} can be by the proper choice of code relative to MFM without changing the ia/system interface.

I. INTRODUCTION

operational performance of digital disc recording s may be optimized by implementing the proper encod- oding scheme. Disc storage systems generally utilize coding by peak detection within a phase locked loop ted timing window. Data pulse patterns are shifted in lative to the detection window due to pattern induced hift, media noise, etc., which may result in errors [1], [2]. the linear density or data transfer rate increases (given y), the available window for peak detection decreases. for a given recording system, the reduced window ns increase the error susceptibility. In high density ding, it is desirable to minimize pulse crowding effects by g the spacing between consecutive magnetic transitions.

vever, the error rate probability depends on both pulse rn induced peak shift and signal to noise ratio. This paper ribes a method for characterizing a given head/disc/ m interface relative to error rate and provides a method electing a code for the maximum linear density with num error rates without altering the system.

II. CODE CHARACTERISTICS

n length limited codes convert N data bits into C code bits the constraint that consecutive code bits or magnetic tran- ns are separated by at least d, but not more then K, empty tion intervals or detents. Codes of this type can reduce or mize peak shift by separating consecutive transitions. en the data period is T, the minimum and maximum time rvals between consecutive transitions are $T_{min}=(N/C)$ 1)T and $T_{max}=(N/C)(k+1)T$ which can be related to a n system through the data transfer rate (DTR = media city x data bit density = V x B = 1/T). Then we may write $_{max}=(C/N)B/(d+1)$ and $FCI_{min}=(C/N)B/(k+1)$. The ction window, T_w, and the clock rate are defined by $=NT/C$ and clock rate $=1/T_w=C/NT=(C/N)$ x DTR. Code iency or density ratio can be defined as the ratio of the timum number of data bits to the maximum number of tran- ons or filled detents, i.e., $E=B/FCI_{max}=(N/(d+1))=T_{min}/T$. A summary of parameters for selected run th limited codes are shown in Table I [4], [5], [6].

TABLE I
CODE PARAMETERS

C	d	k	T_{min}	T_{max}	T_w	E	Clock Rate
2	1	3	T	2T	0.5T	1	2/T
6	2	11	1.5T	6T	0.5T	1.5	2/T
2	2	7	1.5T	4T	0.5T	1.5	2/T
3	1	6	4T/3	14T/3	2T/3	4/3	3/2T

uthors are with the Recording Technology Center, Memorex Corpora- Santa Clara, California 95052

III. EXPERIMENTAL

The operational performance of digital recording systems is directly related to the transition error rate (BER for MFM). For peak detection systems utilizing a phase-locked loop generated timing window as in disc data systems, the transition error rate is defined as the number of transitions in a data stream which occur outside of the system detection window per transition. These errors are dependent on the signal to noise ratio, pattern induced peak shift, signal processing and detection window width. Therefore, in order to determine the system transition error rate, a marginalized VFO (MVFO) which is designed to count all flux transitions in a data stream occurring outside a preset but variable detection window was utilized. The results of this measurement for a system with the characteristics given in Table II are plotted as error rate versus detection window in Fig. 1. The MVFO plots were obtained using a worst case flux reversal pattern (110110---)for four disc radii. The minimum detection window required to assure error rates to 10^{-6} was determined directly. Error rates less than 10^{-6} can be obtained by extrapolation of the exponential portion of the MVFO plots on semilog paper as shown.

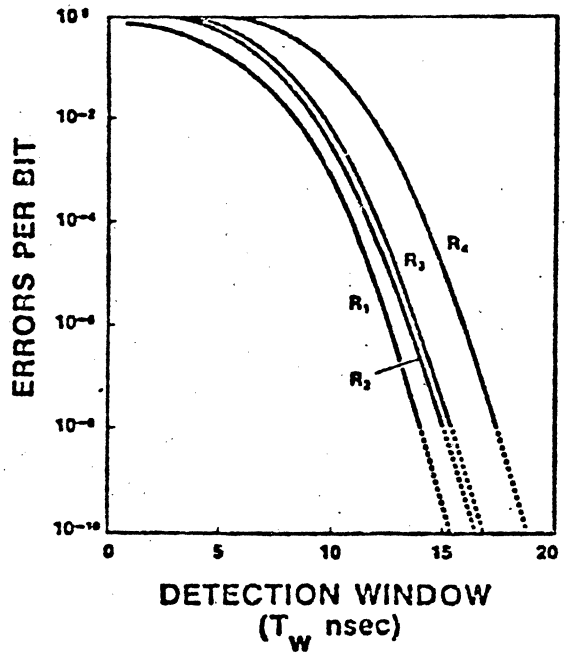


Fig. 1 MVFO Curves of Errors per Bit vs. 1/2 Detection Window at Four Radii

TABLE II
TEST SYSTEM CHARACTERISTICS

P_W50 = 112 nsec	Bandwidth = 2.4 MHz to 8.2 MHz
Vel. = 1508 ips @ 4 inch radius	SNR ≃ 23 dB Peak/rms (after differentiation)

The FCI content varies as a function of radii (6400 x 4/R). Since the MVFO plots are on an error/track basis, the window for a given error rate can be read directly from the MVFO plots and replotted versus their relative FCI, (Fig. 2.). The half window required as a function of FCI for error rates from 10^{-6} to

17 are shown. The MVFO plots used to form the error rate ves in Fig. 1 and Fig. 2 were obtained by averaging the ividual MVFO plots from 15 head/surface combinations in a ical HDA module. These plots were obtained on inside cks (R=4.000 inches and 5.397 inches) and outside tracks =5.17 inches and 6.568 inches) for both ID and OD head s. The two head sets at the two radial positions provided the r data points.

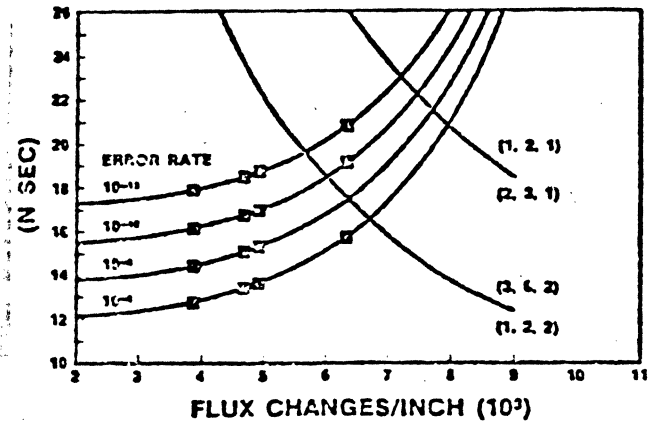


Fig. 2 Code Window and Error Window as a Function Linear Density

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The behavior of the required detection window versus FCI urves indicate that at low linear densities, the window is ssentially constant for each error rate, i.e., independent of CI. This constant window at each error rate can be attributed the subsystem SNR (media, head, electronic noise, etc.). At igh linear densities, the window versus FCI at various error ates rapidly increases from the additional affects of pattern nduced peak shift (adjacent pulse interaction). An analysis of he experimental detection window as a function of flux ransition density shows that the resultant curves may be fit ith a series of exponential functions. Over the limited range of ransition densities and error rates investigated, the xperimental data can be well approximated by an equation of he form:

$$T_w = A \exp(M \cdot FCI \times 10^{-3}) + b \quad (1)$$

At a given error rate, A, M, and b are constants. These constants for the experimental system were found to be 0.158, 0.512 and 15.2 respectively (error rate $=10^{-10}$) with a correlation coefficient of 0.99. Values of b can be obtained to fit the curves at each error rate (10^{-8} to 10^{-12}) and provide a means of extrapolating around the measured points. The value of A+b, which is a constant for each error rate, is equivalent to the extrapolation of the detection window to zero FCI. This value may be attributed to the system noise induced peak shift (T_n) which includes media noise, electronics noise, VFO jitter, etc., but obviously exludes pattern induced peakshift. The accuracy of the extrapolation to zero FCI can be verified by utilizing the expression by Tamura, et.al.[7]

$$T_n = \frac{1}{2\pi F_s} \sin^{-1} \frac{1}{S/N} \quad (2)$$

where T_n is the noise induced peakshift, F_s is the maximum signal frequency and S/N is the signal to noise ratio of the system (peak signal to RMS noise). Using this equation with the appropriate sigma value (Gaussian approximation), the theoretical and experimental values can be compared as shown in Table III.

TABLE III

Sigma Value	Error Rate	Theoretical T_n (nsec)	Experimental T_n (nsec)
5.61	10^{-8}	13.56	13.5
6.36	10^{-10}	15.50	15.2
7.03	10^{-12}	17.28	17.2

The window required for an error rate at a given linear density can be considered to consist of two parts due to pattern induced peak shift, T_p, and noise, T_n. Consequently, at a given error rate and linear density, the minimum detection window must meet the condition:

$$\tfrac{1}{2} T_{w\min} = |T_p| + |T_n| \quad (3)$$

Therefore, the peak shift can be obtained by subtracting T_n from the minimum detection window in Fig. 2 as

$$|T_p| = \tfrac{1}{2} T_{w\min} - |T_n| \quad (4)$$

The experimental peak shift (worst case pattern) may then be plotted as a function of linear density as shown in Fig. 3. Peak shift values from Fig. 3 compare within ±1 nsec up to 7500 FCI with values calculated for linear superposition of Lorentzian pulses.

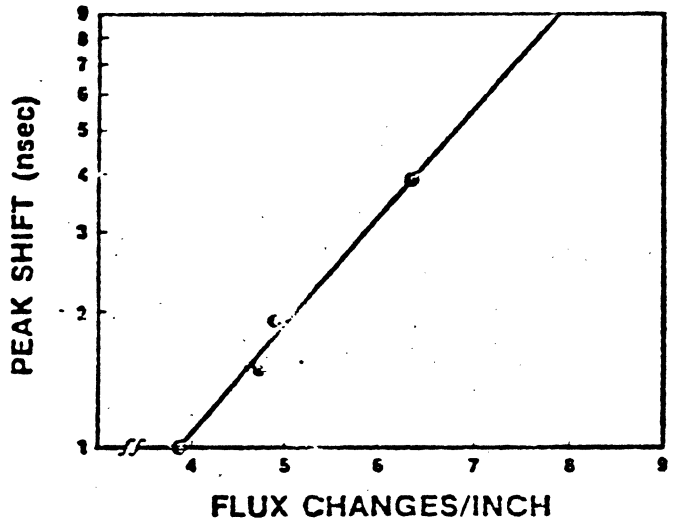


Fig. 3 Peak Shift as a Function of Linear Density for the Experimental System

The expression for the code half detection window value can be derived as

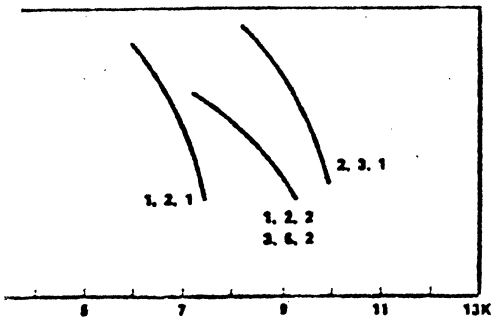
$$\tfrac{1}{2} T_w = \frac{1}{2V(d+1)FCI} \quad (5)$$

The code detection window curves are shown plotted on Fig. 2. The maximum linear transition densities defined by the intersection of the code window curves at each error rate may then be converted to data bit density through the code parameters using the relation

$$B = \frac{N}{C}(d+1)\,FCI \quad (6)$$

and plotted as shown in Fig. 4.

The following comparisons are made assuming no change in system characteristics or detection scheme. Examination of Fig. 4 indicates the MFM code, (1,2,1), exhibits a limiting density of 7450 bpi; the (3,6,2) and (1,2,2) codes exhibit a limiting



Rate vs. Linear Data Density for Selected Codes

) bpi; and the (2,3,1) code exhibits a value of error rate of 10^{-10}. An increase in linear density MFM of 33% is obtained with the (2,3,1) code obtained with the (3,6,2)/(1,2,2) codes. At error nd 10^{-6}, the linear density improvements for the codes increase to 24% and 26%. The greater observed with the (3,6,2) code in Ref. [4] was proved analog signal spectral shaping.

V. CONCLUSIONS

for determining code performance based on he detection window versus transition density at ates has been presented. The experimental data retical calculations of system noise induced peak ulations of pattern induced peak shift by simple osition. The method allows for determination as to recording performance is limited by the system shift characteristics.

Based simply on the code efficiency, the (3,6,2)/(1,2,2) codes should exhibit a 50% increase in linear density; and the (2,3,1) code a 33% increase in linear density relative to MFM. However, it can be seen from the experimental results that, although the (2,3,1) code does exhibit a 33% increase in density as suggested by the efficiency; the (3,6,2)/(1,2,2) codes only increase the linear density by 22 to 26%. This may be attributed to the fact that the experimental system was noise dominated, i.e., $T_n > T_p$ at the maximum operating FCI; and therefore, relative window width has a more significant effect on maximum linear density than efficiency.

VI. ACKNOWLEDGEMENT

The authors wish to thank W.D. Huber for many helpful discussions and F.J. Sordello and S. Puthuff for support.

VII. References

[1] E.R. Katz and T.G. Campbell, *IEEE Trans. MAG*, MAG-15, No. 3, pp. 1050-53, May, 1979.

[2] J.J. Newman and R.D. Fisher, *IEEE Trans. MAG*, MAG-16, No. 1, pp. 23-25, January, 1980.

[3] G.F. Hughes and R.K. Schmidt, *IEEE Trans. MAG*, MAG-12, No. 6, pp. 752-54, November, 1976.

[4] G.V. Jacoby, *IEEE Trans. MAG*, MAG-13, No. 5, pp. 1202-04, September 1977.

[5] T. Horiguchi and K. Morita, *IEEE Trans. MAG*, MAG-12, No. 6., pp. 740-42, November, 1976.

[6] P.A. Franaszek, *IBM Journal of Research & Development*, Vol. 14, pp. 3767-83, July, 1970.

[7] T. Tamura, M. Tsutsumi, H. Aoi, H. Matsuishi, K. Nakagoshi, S. Kawano, and M. Makita, "A Coding Method in Digital Magnetic Recording," *IEEE Trans. MAG*, MAG-8, No. 3., pp 612-14, September 1972.

High performance in magnetic media demands much from conventional encoding methods, but each method has its own disadvantages; yet a simple encoding technique for mass storage has a number of attractive characteristics distinctly its own

4/7

~~FAW GRAHAM~~

FOR YOUR INFO

New Method for Magnetic Encoding Combines Advantages of Older Techniques

Arvind M. Patel

General Products Division
IBM Corporation
San Jose, California

A unique method of magnetic recording combines two advantages not both found in previous methods—absence of a dc component in the signal read from tape, and maintenance of a high recording efficiency on the tape itself. In addition, the method retains principal advantages of previous methods in that it is self-clocking for any pattern of recorded data and is not seriously affected by baseline or peak shift in the readback signal.

Called zero modulation or ZM, the method is used in the IBM 3850 Mass Storage System (see box at right). This machine reads and writes on a magnetic surface with a rotating read/write head. Rotation requires transformer coupling, which cannot handle a dc component; therefore an encoding method that imposes neither a dc component nor the disadvantages of other techniques is required. ZM is a significant improvement over earlier encoding methods, some of which merely assigned transitions of magnetic polarity to bits in some more or less straightforward way and often achieved a new advantage at the expense of an old one. These methods included non-return-to-zero inverted (NRZI), phase encoding (PE), group-coded recording (GCR), frequency modulation (FM), and modified frequency modulation (MFM), sometimes called delay modulation. Zero modulation uses sophisticated coding of data to match idiosyncracies of the magnetic recording channel with waveform properties of the recorded signal.

Importance of efficiency and the absent dc component is brought out when ZM is compared with some earlier codes. NRZI, for example, is the simplest en-

Honeycomb Storage

The IBM 3850 is a mass storage system that serves as a virtual storage medium supporting magnetic disc files, in much the same way as the discs serve as virtual storage supporting main memory. It consists of an array of data cartridges about 2 in. in diameter and 4 in. long with a capacity of 50 million characters each. Although the cartridge contains a length of magnetic tape, stored data are organized in cylinders analogous to those of a disc file, and can be transferred to the disc file a cylinder at a time—that is, without moving the disc read/write heads during the transfer. Up to 4720 cartridges are stored in hexagonal compartments in a honeycomb-like apparatus that includes a mechanism for fetching cartridges from the compartments, reading or writing data on them, and replacing them.

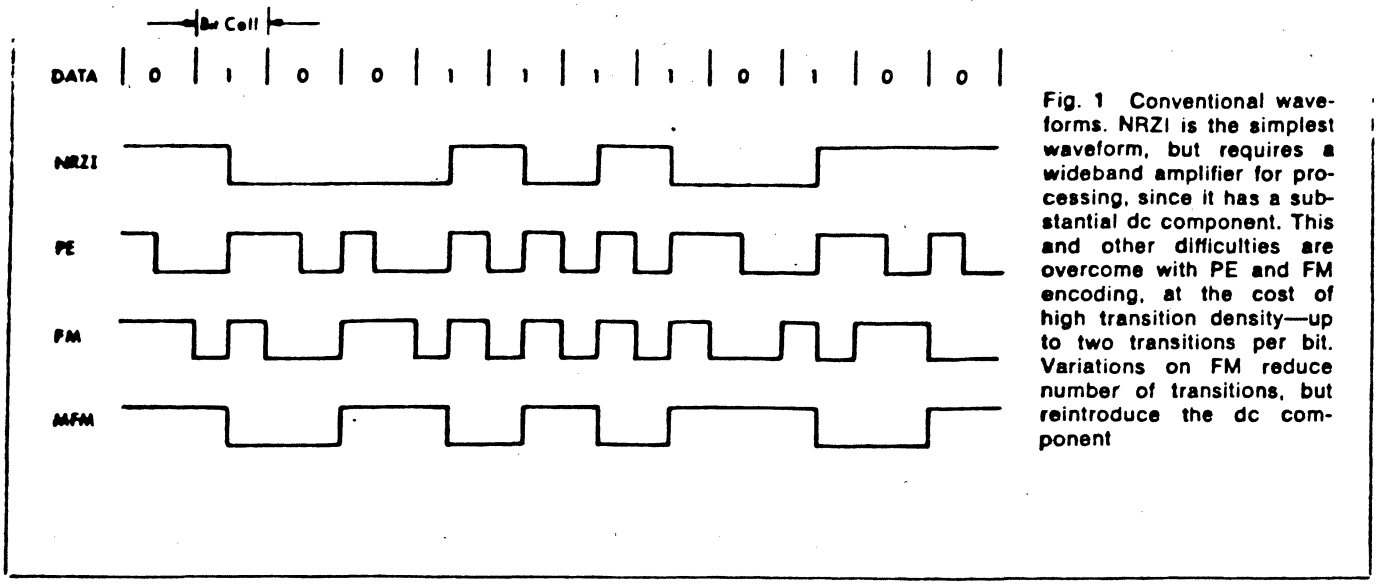


Fig. 1 Conventional waveforms. NRZI is the simplest waveform, but requires a wideband amplifier for processing, since it has a substantial dc component. This and other difficulties are overcome with PE and FM encoding, at the cost of high transition density—up to two transitions per bit. Variations on FM reduce number of transitions, but reintroduce the dc component

coding method, used on ½-in. wide magnetic tape at data recording densities up to 800 bits/in.[1] Presence or absence of a transition in the NRZI magnetic waveform corresponds to 1 or 0 respectively in the binary data stream (Fig. 1); successive magnetic transitions, which alternate in polarity, produce alternately positive and negative electric pulses in the readback signal, nominally symmetric about a base line. Principal drawback of NRZI is that long strings of 0s are recorded as correspondingly long periods with no magnetic transitions, and hence no pulses in the readback signal; the read clock can lose synchronization during those periods. Furthermore, wideband circuits with dc response are required for signal processing and data detection.

NRZI is also subject to a more subtle difficulty: baseline and peak shift, alluded to previously. At high densities, the readback pulses produced by a string of consecutive 1s tend to interfere with one another. When such a string of 1s is preceded or followed by a string of consecutive 0s, this interference is asymmetrical, causing the first or last few pulses to have larger amplitudes; thus the base line appears to drift. A similar shift can arise from a string of consecutive 0s if the dc response of the electronic circuitry is slightly off specification. Interference also causes the pulses to seem to slide into the signal-free zone occupied by the 0s, creating a displacement in time of the pulse peaks. This peak shift can be a significant fraction of the nominal time between bits.

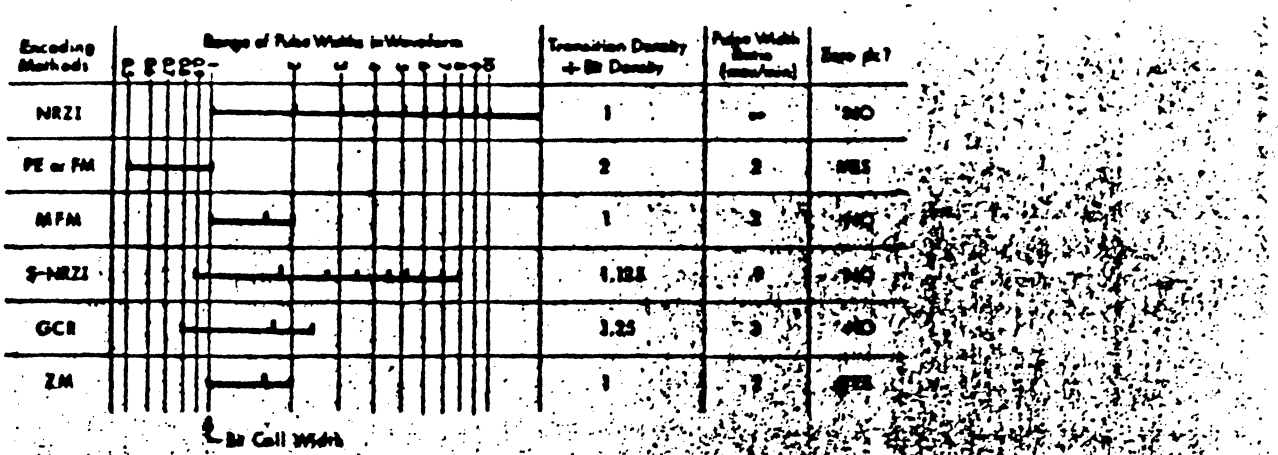
Phase encoding (PE)[2] was devised to alleviate these problems, and is used on ½-in. tapes recorded at 1600 bits/in. In PE a 1 corresponds to an up-going transition and a 0 to a down-going transition at the center of the bit cell. Where two or more 1s or 0s occur in succession, extra transitions are inserted at the bit-cell boundaries. Resulting waveform is self-clocking and has no dc component, since the waveform in each bit cell has up (positive) and down (negative) signal levels of equal duration. However, PE requires twice the transition density of the NRZI method for a random data pattern. Thus recording efficiency is poor.

Frequency modulation has transitions at every bit-cell boundary. However, although FM is similar to PE in all waveform properties, the 1 and 0 correspond to a presence or absence, respectively, of any transition at the center of the corresponding bit cell, rather than to an up or down transition. In the IBM 3330 disc file FM has been supplanted by MFM, or delay modulation, which provides enough clocking transitions without doubling the transition density.[3] In MFM (as in FM) a 1 and a 0 correspond to the presence or absence, respectively, of a transition in the center of the corresponding bit cell. However, additional transitions at the cell boundaries occur only between bit cells that contain consecutive 0s. This method retains an adequate minimum rate of transitions for clock synchronization without exceeding the maximum transition density of NRZI, but at the cost of a more complex data detection process. An additional disadvantage is the dc component, with indefinitely large accumulated dc charge for some data patterns such as 0110110110....

Modified forms of NRZI include synchronized NRZI (S-NRZI), and group-coded recording (GCR) used to record magnetic tape at 6250 bits/in.[4] In these methods, the data stream is precoded by adding bits to break up long strings of 0s. In S-NRZI, a 1 extends each 8-bit group to nine bits and establishes a synchronization transition. In GCR, each 4-bit group is mapped into five bits using a fixed assignment that guarantees that the coded data stream never contains more than two consecutive 0s. Detection process for both methods is the same as for NRZI, but maximum transition density is necessarily higher, and the dc component can be quite prominent.

TABLE 1

Encoding Methods: Parameters



Advantages of ZM

Zero modulation encoding produces waveforms that closely match characteristics of the magnetic recording channel. Among these characteristics (Table 1) are:

Minimum energy at low frequencies—achieved because the long-term dc component is 0, and maximum accumulated unbalance between positive and negative pulse durations of the waveform is small. This helps alleviate base-line shift of high density read signals.

Narrow bandwidth—conserved by constraining pulse width and ratio of maximum to minimum pulse width. This helps reduce the problem of peak shift. Resulting read signal amplitude is relatively uniform—that is, it has a small dynamic range for any kind of data stream.

Adequate transition rate—as a consequence of the constraint on the maximum pulse width, a clock signal can be derived from the transition rate.

High efficiency—ratio of highest transition density to bit density is close to 1. With NRZI and MFM the ratio is exactly 1, and with other codes it is fractionally higher, rising to 2 for PE and FM.

Reliable Error Checking—as a consequence of these characteristics and various other constraints of the method, the code structure of ZM is such that error checking is easy and reliable.

Zero Modulation Algorithm

The ZM algorithm can be readily implemented by a practicing engineer with the information that follows. Theoretical details and proofs have been published elsewhere.[5] The ZM algorithm maps every data bit sequentially into two binary digits in such a way that any two consecutive 1s are separated by at least one and at the most three 0s. This sequence is recorded in NRZI. Consequently, the narrowest pulse in the ZM waveform spans two digits in the coded sequence, and this is the width of the data-bit cell. Similarly the widest pulse spans four digits, which is twice the narrowest pulse width. The resulting waveform is similar to that of MFM.

This limited range of pulse widths corresponds to a narrow bandwidth for the waveform, and the narrow bandwidth has obvious electronic advantages. Short "pieces" of the waveform of one polarity are always balanced very soon by equally short pieces of the waveform with the opposite sign. Two such sections taken together make up a section with a zero dc component, and the waveform made up of such balanced sections contains minimal energy at low frequencies.

A quick way to estimate the amount of dc in a train of rectangular pulses representing data encoded in ZM or any other code is to count positive- and negative-going excursions, giving a value of ±1 to the narrowest pulses, ±2 to pulses twice as wide, and so on. This corresponds to integrating the waveform mathematically, or collecting charge on a capacitor fed by a current with the given waveform. If the total strays far from 0, the dc component is present. Under the zero modulation algorithm, the total changes sign frequently and never exceeds ±3, corresponding to the limit of three 0s between any pair of 1s.

Two-for-one mapping is a nonlinear function of the preceding and following data sequences and requires an encoder with memory. In a practical implementation, amount of memory can be limited as desired by adding a small amount of redundancy. In its functional form, however, the ZM algorithm, in general, requires unlimited memory.

The algorithm can be described in terms of a data bit to be encoded, one preceding and one following data bit, and the two coded digits corresponding to the preceding data bit; and in terms of two parity functions that look ahead and back relative to the

TABLE 2

Encoding and Decoding Algorithms

ENCODING ALGORITHM

d_0	a_0b_0	CONDITION
0	10	$d_{-1} = 0$
0	10	$d_{-1} = 1$ and $a_{-1}b_{-1} = 00$
0	00	$d_{-1} = 1$ and $a_{-1}b_{-1} \neq 00$
1	10	$d_{-1} = 0$ and $P(A) = 0$ and $P(B) = 1$
1	10	$d_{-1} = 1$ and $a_{-1}b_{-1} = 00$
1	00	$d_{-1} = 1$ and $a_{-1}b_{-1} = 10$
1	01	otherwise

$$a_0 = \bar{d}_0\bar{d}_{-1} + d_0\bar{d}_{-1}\,\overline{P(A)}\,P(B) + d_{-1}\bar{a}_{-1}\bar{b}_{-1}$$
$$b_0 = d_0(P(A)\,\bar{d}_{-1} + \overline{P(B)} + b_{-1})$$

DECODING ALGORITHM

a_0b_0	d_0	CONDITION
10	1	$a_{-1}b_{-1} = 00$
10	0	$a_{-1}b_{-1} \neq 00$
00	1	$a_{-1}b_{-1} \neq 10$
00	0	$a_{-1}b_{-1} = 10$
01	1	none

$$d_0 = b_0 + a_0\bar{a}_1\bar{b}_{-1} + \bar{a}_0a_{-1}\bar{b}_{-1}$$

bit being encoded. Look-ahead parity, P(A), is the count modulo 2 of 1s in the data stream beginning with the data bit being encoded and counting forward to the next 0 bit; look-back parity, P(B), is the count modulo 2 of all 0s in the data stream from its beginning up to the present bit. For example, in the data sequence 01011110, P(A) = 1 at the second, fifth, and seventh bits from the left, P(A) = 0 at all 0 bits and all the other 1 bits; while P(B) = 1 at the first, second, and eighth bits and 0 elsewhere.

These rules can be translated into encoding and decoding functions expressed either in the form of tables or as equations (Table 2). In both, the symbol d represents a data bit; a and b, coded digits; and subscripts −1, 0, and +1, preceding, current, and following bits, respectively. For convenience, the nonexistent bit preceding the first data bit is assumed to be 1 and its look-back parity is 0; the nonexistent bit following the last bit is 0.

The example (Fig. 2) illustrates the relationships among the various parameters of the data sequence and the corresponding ZM waveforms.

Although the ZM waveform looks similar to the MFM, or delay modulation, waveform, the important difference is that the MFM waveform contains a dc component, and the accumulated charge often increases indefinitely, whereas the maximum accumulated charge in the ZM waveform is ±3 units.

ZM Algorithm With Limited Memory

Look-back parity is accumulated from the number of 0s simply by updating a 1-bit storage cell as data bits are encoded. However, look-ahead parity depends on the length of a string of 1s in the following data sequence. Since the algorithm imposes no limit on the length of this string, memory requirement for computation of P(A), in general, is unlimited. However, a variant of the basic ZM algorithm limits the memory requirement to any specified number of bits.

When accumulated look-back parity P(B) = 0, encoding functions become independent of the look-ahead sequence:

$$\left.\begin{aligned} a_0 &= \bar{d}_0\bar{d}_{-1} + d_{-1}\bar{a}_{-1}\bar{b}_{-1} \\ b_0 &= d_0 \end{aligned}\right\} \text{if } P(B) = 0.$$

To force P(B) to 0, a digit, P, can be inserted in a continuous data stream at fixed intervals of f bits. This insertion implies that look-ahead parity, P(A), of the sequence of 1s at the end and beginning of any section of f + 1 bits has no effect on ZM mapping in the modified data sequence. In Fig. 3, some values of P(A) are denoted by ϕ indicating a "don't care" value when P(B) = 0. The only sequences of 1s affecting the mapping, then, are those between two 0s in the same section of f + 1 digits. The longest such sequence is f − 1 digits long. Thus, the memory required to compute P(A) is f − 1 bits.

Thus, to limit the amount of memory, the ZM algorithm is given two important modifications: First, an extra P-bit with the value of P(B) at position f is inserted at the end of every section of f data

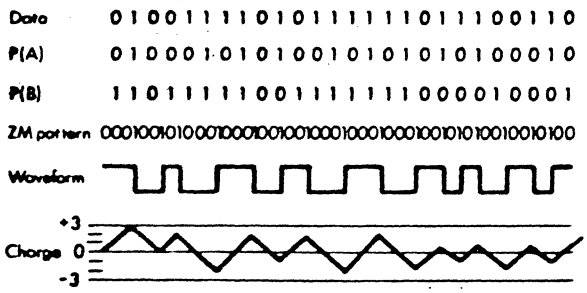


Fig. 2 Basic zero modulation. Substituting two encoded bits for each data bit on the basis of information in preceding and following bits, and recording the encoded bits in NRZI form, eliminates the dc component without other waveforms' disadvantages. However, since it places no constraints on the data pattern, it can require an infinitely large memory for implementation

bits. Second, computation of P(A) at any data bit extends only to the following f − 1 data bits, as a binary logic function of the data stored in f bits of memory:

$$P(A) = d_0\bar{d}_1 + d_0d_2\bar{d}_3 + d_0d_2d_4\bar{d}_5 + \ldots + d_0d_2d_4 \ldots d_{t-3}\bar{d}_{t-2} + d_0d_2d_4 \ldots d_{t-3}d_{t-2}$$

where t = f if f is even and t = f − 1 if f is odd. Look-back parity is merely a count of 0s as in the case for unlimited memory.

Encoding process is delayed by f bit periods in a continuous stream of data, while the memory is loaded for computing P(A), but the decoding process is delayed by only one bit period. Thus decoding errors in ZM do not propagate.

In the diagram, a value of f = 8 is assumed for illustrative purposes; but in fact, the value of f has no theoretical limits. As described later, a shift register delays each bit while look-ahead parity is generated. Small values of f require short shift registers and encode data quickly; however they add a larger proportion of redundancy in the form of the extra P-bit than do large values. Nevertheless, they may be convenient in some applications. For efficient utilization of the magnetic recording medium, large values of f—eg, 100 or more—are mandated. Although they impose substantial encoding delays, these are still negligible compared to the access time of a mass storage system, which is measured in seconds. Large values of f do not necessarily imply complex encoding logic; the look-ahead parity generator merely counts the bits as they pass, and a monolithic shift register is relatively inexpensive.

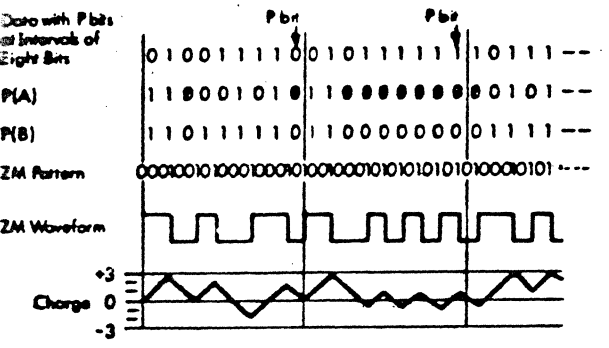


Fig. 3 Practical zero modulation. Adding an extra bit to the data stream at regular intervals in accordance with a prescribed pattern permits the two-for-one substitution of ZM to be made with a memory of realizable capacity. In the waveform, dc component is still 0, transition density is low, yet added redundancy is minor

Synchronization Signal

When reading, a ZM waveform is decoded into a data sequence with the help of a clock, which is usually derived from the waveform. A synchronizing signal of sufficient length and recognizable ending is useful in marking the beginning of data. A similar resynchronizing signal may also be inserted at predetermined intervals in the waveform, such as at ZM memory boundaries, for protection in case of temporary loss of synchronization in magnetic defects.

Several characteristics are required of the synchronization signal:

(1) It must be distinctive enough not to be confused with the normal data waveform in its original or shifted position.

(2) It must satisfy the ZM constraints of maximum and minimum pulse widths.

(3) It must have no net dc component over its length, although unlike the encoded data, the integrated total may exceed three units within itself. The accumulation may be four, five, or six units (or even more if the synchronization signal is short and infrequent).

(4) Basic synchronization signal should be reasonably short and the endings compatible with the ZM algorithm for insertion at the memory boundary without modification.

These specifications could be easily satisfied if some binary sequence were known to be inadmissible as data. However, no such sequence is possible for serial data. Alternatively, a synchronizing signal can be chosen from inadmissible sequences in ZM-coded patterns. These sequences must satisfy the ZM pulse-width constraint but may exceed the maximum charge constraint.

Among the sequences that satisfy the ZM run-length constraints, 0 0 1 0 1 0 0 0 1 0 1 0 0 0, either forward or backward, is the shortest that does not occur in any ZM pattern. Any pattern containing one of these sequences can be used as a synchronizing pattern. Two examples are:

$W_1 = 0100010010100010100010001010010$

$W_2 = 0100010100010100010$

The second of these does not satisfy specification 3. Both examples begin and end with 01, and can be padded by any number of 01 digit pairs if desired for clocking. These endings also allow the synchronization signal to be placed at the ZM memory boundary without modification.

In actual application, the synchronization pattern is placed at predetermined intervals at ZM memory boundaries. In case of synchronization loss, the clock is regenerated by the read waveform as soon as the defect or other cause has passed. With the clock running, the signal detector can produce the binary ZM pattern, but the pattern cannot be decoded until the synchronization pattern re-establishes the ZM pair relation with respect to the clock. If the clock is found to be out of synchronization, ie, one ZM digit out of step with the read signal, its complement may be used for decoding.

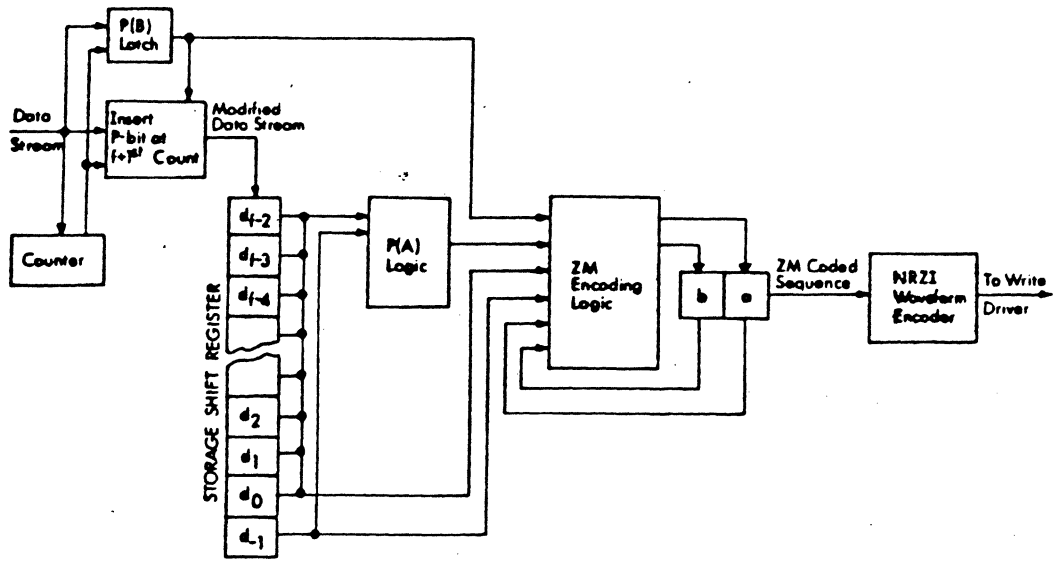


Fig. 4 ZM encoder. Principal feature is a shift register from which data bits are fed to encoding logic, which is straightforward implementation of equations in text

Error Check in ZM Patterns

Patterns of digits generated by the ZM algorithm satisfy various constraints, including parity in the case of ZM with limited memory. These constraints provide a powerful check capability for bit-detection errors and synchronization errors at the receiver.

Because error-free ZM patterns possess run lengths of one, two, or three 0s between two 1s, two consecutive 1s indicate a pick-up, or a 0 incorrectly read as 1, while four or more consecutive 0s signal a drop-out error. Acquisition of excessive dc charge can be detected with an up-down counter that increments for every code bit position recorded with a positive level, decrements likewise when the level is negative, and signals an error if the total exceeds ±3 at any time. An alternative method of implementing this check has been worked out.[5] Finally, value of P(B) and charge value must both be 0 at the memory boundary—a simple but effective check on both synchronization and random errors. These two checks at the memory boundary are equivalent in the sense that neither detects any error missed by the other as long as the checking circuits are working properly. Error checking of ZM patterns at the receiver, an additional benefit derived from the stringent ZM constraints, need be implemented only to enhance reliability even further.

Implementation of ZM Algorithm

In a ZM encoder (Fig. 4), the first step is to modify the binary data sequence by inserting the P-bit at fixed intervals of f bits. The P-bit is the value of P(B) at count f, computed by a simple latch triggered by each 0 in the data stream. At count f + 1, the P-bit is inserted in the data stream, and the P(B)-latch and counter are reset to 0. The modified data stream passes through a shift register f bits long, which stores the previous and current data bits and the following f − 2 data bits. From these stored bits the look-ahead parity function P(A) is generated in accordance with the binary logic function given previously. From these values of P(A) and P(B), the current and next previous bits, and the just-computed ZM bits, the ZM code sequence is generated. Two feedback latches store each pair of bits of the ZM pattern as they are computed, for use in the next bit cycle. These latches are updated continually as the data bits are sequentially encoded into ZM patterns.

Initially, look-back parity is set to 0 and feedback latches are set to 01. Encoding starts after a delay of f − 1 clock periods during which the first data bits are shifted into the storage register. This puts d_0 in the next to last cell of the shift register, from which it

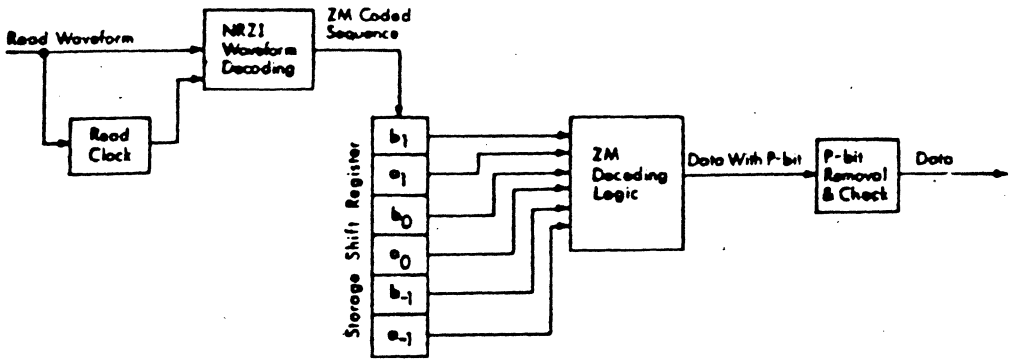


Fig. 5 ZM decoder. Data bits are defined in terms of pairs of code bits, taking into account preceding and following pair along with current pair. Hence the key element is a 6-bit shift register that shifts two bits at a time and feeds code information to decoding logic

is encoded into the first pair of code bits. The "previous" bit, assumed to be 1 and initialized accordingly, either has been inserted in the shift register just ahead of the first real data bit or can be gated to the ZM encoder during the first encoding cycle. Alternatively it can be initialized simply by placing all 1s in the shift register. This sequential encoding process is continued for the ZM coded sequence which is converted into a conventional NRZI waveform. ZM synchronization pattern, if necessary, may be inserted in the coded pattern at selected $(f + 1)$-bit boundaries.

ZM decoder (Fig. 5) converts the received waveform into a ZM pattern by means of conventional NRZI clocking and detection circuits. This ZM pattern passes sequentially through a 6-bit storage register which stores three pairs of pattern bits. As in encoding, for the first pair of bits, the "preceding" pair is set to 01. Decoding starts after the second pair arrives in the storage register. The decoder generates each data bit as a function of the preceding, current, and following bit pairs, as described previously.

Coded sequence received at the decoder satisfies all ZM constraints, and every $(f + 1)$st bit in the decoded data stream is the correct P-bit. Any departure from the constraints or an incorrect P-bit clearly indicates an error. Some or all of these checks can easily be implemented for error checking.

References

1. American National Standard, Recorded Magnetic Tape for Information Interchange (800 CPI, NRZI), American National Standards Institute, Document ANSI-X 3.22-1973
2. American National Standard, Recorded Magnetic Tape for Information Interchange (1600 CPI, P E), American National Standards Institute, Document ANSI-X 3.39-1973
3. M. Hecht and A. Guida, "Delay Modulation," *Proceedings of the IEEE*, July 1969, pp 1314-1316
4. Proposed American National Standard, Recorded Magnetic Tape for Information Interchange (6250 CPI, Group Coded Recording), American National Standards Institute
5. A. M. Patel, "Zero Modulation Encoding in Magnetic Recording," *IBM Journal of Research and Development*, July 1975, pp 366-378

Dr Arvind M. Patel is a senior engineer with the future storage technology group at IBM's General Products Div, where he is currently involved with development of advanced tape products. He holds a BE degree from Sardar Vallabh-bhai Vidyapeeth, India, an MS degree from the University of Illinois, and a PhD from the University of Colorado, all in electrical engineering.

This review of the basic theory of polynomial error detecting codes is expanded into a real-world application involving remote terminals

Increasing interest in computer stations and terminals located some distance from a large central computer has stimulated a great demand for data link and communication facilities. Inherent in any data link are the associated phenomena such as electrical noise and signal degradation that introduce errors into data transmitted over long distances. Consequently most data communication facilities include error detection capabilities which may allow limited error correction on a character basis and which may enable retransmission of complete messages or data blocks. One method of detecting errors within a data block is to send check bits after the data block as well as a simple parity bit with each character. However, large data blocks may accumulate independent errors which are mutually compensating and may not be encoded in the check bits. Polynomial codes have been used to resolve this problem by providing burst error detection as well as all single- and double-bit error detection.

This article describes specific hardware designed and developed for physical implementation of polynomial check code generation and detection. Hardware is implemented as an I/O device attached to a small computer for use in a remote job entry terminal being developed. The polynomial code generator/detector was designed as a generalized logic block to be used with most small general-purpose computers and to require minimal interface logic design.

Michael A. Liccardo is engaged in the design and development of advanced computer system interfaces, and has had minicomputer systems design as well as extensive general programming experience. He received BSEE and MSEE degrees from the University of California (Berkeley) and is studying for a PhD degree at Stanford University.

Basic Theory

Representation of Binary Information

Consider a message of k bits to be transmitted serially over a data link, and with $n - k$ bits to be appended as check bits unique to the message being transmitted. The binary digit positions of a message can be considered as the coefficients of a polynomial of arbitrary variable x. Each coefficient is restricted to one of two values (ie. 0, 1). If the message consists of k bits, it is convenient to consider the first bit transmitted as the most significant coefficient of the polynomial (ie, the kth bit corresponds to the coefficient b_{k-1}). For example, consider the 6-bit message 101101 written least significant bit to most significant bit from left to right. This message is represented by the general polynomial $f(x) = b_0 + b_1x + b_2x^2 + b_3x^3 + b_4x^4 + b_5x^5$, where the coefficients b_0, b_1, b_2, b_3, b_4, b_5 are 1, 0, 1, 1, 0, 1, respectively. Extending this representation, note that any arbitrary binary message of k bits can be represented by a polynomial of the form

$$f(x) = b_0 + b_1x^1 \ldots + b_{k-2}x^{k-2} + b_{k-1}x^{k-1} = \sum_{i=1}^{k-1} b_ix^i$$

Arithmetic operations may be performed on the polynomials according to the laws of ordinary algebra. How-

* Work performed under the auspices of the U. S. Atomic Energy Commission.

r, attention here is restricted to binary codes in which nomial coefficients are elements from the set (0, 1). as been shown[1] that polynomial codes satisfying error-ecting and correcting properties can be described by special algebraic system which essentially requires t arithmetic operations on polynomials whose coeffi-nts are elements of a field containing q elements be formed modulo q. For binary polynomials this re-ires addition and subtraction to be carried out odulo 2 (ie, the exclusive OR function).

efinitions

veral pertinent definitions and properties of poly-mial operations are: (1) The degree of a polynomial the greatest power of x in which the coefficient is nzero. For example, *n* binary digits is represented by n − 1 degree polynomial. (2) The degree of the poly-mial resulting from the product of two polynomials is e sum of their degrees. (3) If R(x), S(x), and T(x) re polynomials such that T(x) = R(x)S(x), then T(x) said to be divisible by R(x), or R(x) divides T(x). (x) and S(x) are also termed factors of T(x).

With these concepts in mind, the division of poly-omials can be defined according to the Euclidean di-ision algorithm. Given any two polynomials H(x) and (x), there is a unique pair of polynomials Q(x), R(x) uch that H(x) = Q(x)P(x) + R(x), where Q(x) is ermed the quotient and R(x) the remainder. Q(x) and R(x) can be obtained from the division H(x)/P(x) nder the corresponding algebraic system. The degree of R(x) is less than the degree of H(x) and P(x). R(x) may be interpreted as the remains after H(x) has been evenly divided by P(x); and since division is essentially a subtraction operation, R(x) is the remainder after P(x) has been subtracted from H(x) an integral number of times. The special case when R(x) = 0 is precisely the case when H(x) is divisible by P(x), as defined previously.

Data Encoding

The polynomial division process provides some insight into the encoding of *k* bits of information for transmission. If H(x) represents the message and P(x) defines a polynomial the previously defined division will produce a remainder, R(x), which is unique to a small subset of messages including H(x). P(x) is called a generator polynomial since its function is that of generating unique check bits R(x), given any message H(x). If errors occurred during transmission, they can be detected by dividing the received message bits by P(x) and comparing that remainder with the received check bits. If the comparison results in an equality condition, the assumption is that both the message, H(x), and the check, R(x), were transmitted correctly.

Mathematically, H(x) = Q(x)P(x) + R(x), where H(x) is a k − 1 degree polynomial representation of *k* message bits; P(x) is a generator polynomial of degree m − 1 where m − 1 < k − 1. H(x) − R(x) = Q(x)P(x), since addition and subtraction are defined modulo 2, ie, H(x) − R(x) = H(x) + R(x) = Q(x)P(x).

The addition of R(x) modulo 2 essentially modifies the message H(x) in the last m − 1 bits, since R(x), by definition, is of degree less than [illegible] bits). Physically it is more convenient to send R(x) as the last m − 1 bits of the encoded message. This is easily accomplished by considering H(x) to be k + (m − 1) bits in length with m − 1 low order 0 coefficients. Multiplication by x^{m-1} performs the transformation.

$$x^{m-1}H(x) + R(x) = F(x)$$

where F(x) is termed the code polynomial and represents the encoded message of k + (m − 1) bits to be transmitted. *Example*: Encode the message 1010010001, corresponding to the polynomial $H(x) = 1 + x^2 + x^5 + x^9$, using the generator $P(x) = 1 + x^2 + x^4 + x^5$. Multiplying H(x) by x^5 and dividing by P(x) results in:

$$\begin{array}{rl} x^5 H(x) = & 00000\ 1010010001 \\ R(x) = & 11000 \\ \hline F(x) = & \underbrace{11000}_{\text{check bits}}\ \underbrace{1010010001}_{\text{message bits}} \end{array}$$

The encoded message, F(x), consists of ten higher-order message bits, H(x), and five lower-order check bits, R(x).

The encoded message received after transmission over a data link can be represented by B(x) = F(x) + E(x), where F(x) is the correct message and E(x) is the error message. Since arithmetic has been defined modulo 2, E(x) will contain nonzero coefficients in each erroneous bit position. If B(x) is not divisible by P(x), an error has occurred. If the resulting division generates no remainder, B(x) is accepted as the true encoded message. It is possible, however, that enough errors in appropriate bit positions have been generated so that E(x) is divisible by P(x) and B(x) will be decoded as the true message. To ensure effective error detection, the generator, P(x), must be chosen such that no error pattern, E(x), is divisible by P(x). It can be shown that P(x) must have certain properties to enable this error detecting scheme to function. It can also be shown that the ability of a specific code to detect the erroneous data is related to its error correction capabilities.

A more rigorous mathematical treatment of error correcting codes and their properties is given in Ref. 1. A few basic results are given without proof to illustrate typical properties of generator polynomials: (1) A code generated by any polynomial P(x) with more than one term detects all single errors (ie, an error in exactly one position). (2) Any polynomial of the form $1 + x^c$ (c an integer) will detect any odd number of errors (ie, typical odd or even parity error detecting). (3) A polynomial P(x) of length n detects all single and double errors if n < e, where e is the least integer such that P(x) divides $x^e - 1$ ($= x^e + 1_{\text{mod } 2}$). (4) A polynomial P(x) of length n detects any burst-error of length n or less. A burst-error of length n is defined as the number of errors occurring between the first and last errors, inclusive.

Physical Realization

The theory of how to encode and decode messages for error detection has been shown, but the important operation to accomplish this is the division, under addition modulo 2, of messages by a fixed polynomial, P(x). Consider a long-hand calculation of the division of 1 +

```
                    1 1010
1101001 | 1010010 0001
          1101001        STEP 1
          1110110 001
          1101001        STEP 2
          0111110 1
          1101001        STEP 3
          0101011
```

Fig. 1 Manual calculation of $1 + x^3 + x^5 + x^{10}$ divided by $1 + x^3 + x^5 + x^6$

$x^5 + x^8 + x^{10}$ by $1 + x^3 + x^5 + x^6$. Since addition modulo 2 is simply an Exclusive OR function, drop the variable x and manipulate the binary coefficients of the polynomials (see Fig. 1). The quotient is $Q(x) = x^4 + x^3 + x$ and the remainder is $R(x) = x^5 + x^3 + 1$.

Referring to the manual division example (Fig. 1), the division algorithm is: (1) Align high-order coefficients of P(x) and the partial remainder. The first iteration aligns the divisor P(x) with the dividend. (The dotted underline references the partial remainder for each step). (2) Subtract (modulo 2) P(x) from the partial remainder. (3) Go back to step (1) if the degree of the partial remainder is greater than or equal to the degree of divisor P(x); otherwise the partial remainder is the remainder R(x). Notice that in step (3) the alignment of high-order coefficients required the partial remainder to be shifted left by one bit and the entry of the next two dividend bits into the low-order position. In general, at each step the next dividend bit is "brought down" or shifted into the low position of the partial remainder until the most significant bit is 1. This becomes clearer if the addition is thought of as being performed in a 6-bit register in which the most significant bit of the generator and partial remainder are ignored. These bits will always result in $1 + 1 = 0$.

Consider a 6-bit register and step (2) of the previous division, with reference to Fig. 2(A). Shifting the partial remainder left by one bit will align the higher-order coefficients as shown in Fig. 2(B). Since the subtraction is performed after the coefficients are aligned, the most significant bit shifted out of the register is used to enable the subtract logic; the result after subtraction is shown in Fig. 2(C). Each step may be implemented in a similar manner, except that the division is terminated after all data bits have been shifted into the register.

The hardware required to implement this algorithm (ie, shift and subtract modulo 2) is simply a feedback shift register with Exclusive OR gating. Subtraction and addition modulo 2 is implemented by the Exclusive OR function. The number of shift register bit positions is equal to the degree of the divisor P(x). The shift register shown in Fig. 3(A) is oriented with the low- to

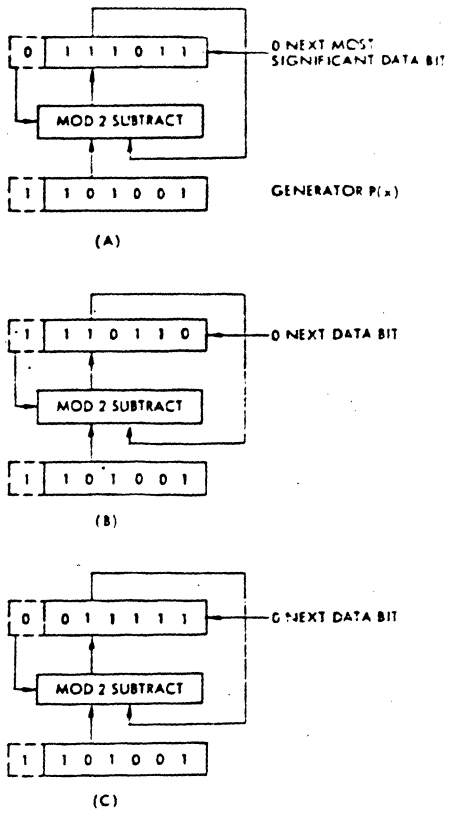


Fig. 2 Shift register alignment sequence

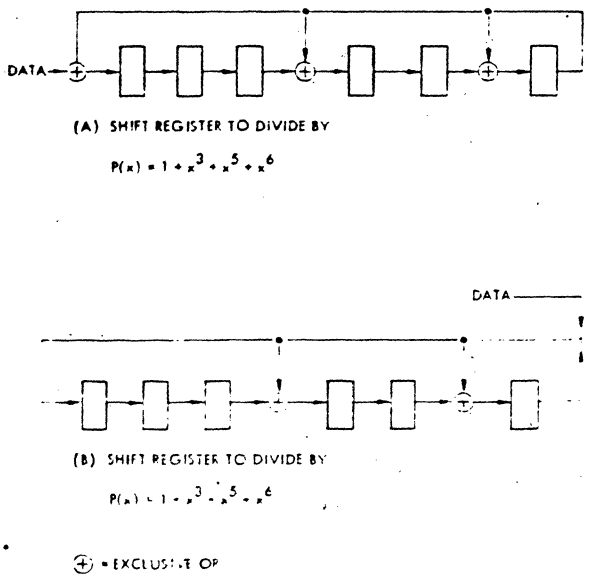


Fig. 3 Shift register divide implementation

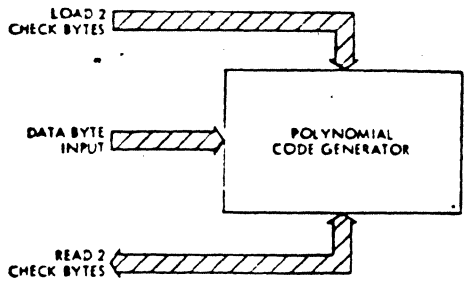


Fig. 4 Functional block diagram

high-order position from left to right. The most significant data bit enters the left most shift register bit position. A delay of 6-bit-shift times is associated with this shift register since six extra 0's must be shifted in after the data bits to complete the encoding process. This delay can be avoided by treating the data as if it were shifted out of the high-order end. The register can be easily modified for this more efficient scheme[2] as shown in Fig. 3(B). Since encoding and decoding of messages are precisely equivalent, this circuit can be used for both functions. This and the simple logic implementation are the attractive features of polynomial error detection. Error correction implementation is much more complex.[1]

Hardware Description

Design Philosophy

The initial design objective was to develop polynomial code generation electronic logic that could easily interface with a small 8-, 12-, or 16-bit computer. The specific application was to implement an I/O device for error detection at remote job entry stations and at communications concentrator stations. Both of these stations were to be provided with a small general-purpose computer to be used as a data communications controller. To allow for expansion and upgrading of the remote terminal system, the polynomial code device design objectives were incorporated with enough flexibility to allow interfacing to most commercial minicomputers.

Using the black box concept, the polynomial code generator accepts 8-bit data bytes and produces two 8-bit check bytes. These two check bytes may be read from the black box and subsequently loaded into a computer. Conversely, two check bytes can be loaded into the black box from the computer I/O bus, allowing previous data-error-code generation/detection to continue from a previously interrupted state. (See Fig. 4.) This facility is very convenient for time-multiplexing error-code generation/detection between several communication lines and eliminates unnecessary replication of error-detection hardware at a data concentration station.

The 8-bit byte was chosen as the basic data structure because most character code sets used in data communication are easily specified with 8-level coding. It also appears that most small and medium size computers have integral-number byte word sizes. In essence, there appears to be an evolving byte-oriented standardization of most large computer systems using communication facilities.

Operating Characteristics

A more detailed block diagram of the polynomial code device is shown in Fig. 5. The device consists of an 8-bit buffer register (through which all data bytes pass for encoding), a 16-bit polynomial register with associated feedback gating for a specific generator polynomial, and the control logic.

The polynomial code device performs four functions as specified by the four command bits. These bits, strobed into the control logic by a pulse from the computer, are:

LOAD LEFT — The left byte of the polynomial register is loaded with data from the I/O bus. The LOAD and EOR/LEFT bits are high for this command, and the device is set for EOR mode.

LOAD RIGHT — The right byte of the polynomial register is loaded with data from the I/O bus. The LOAD and CRC16/RIGHT bits are high for this command, and the device is set to CRC16 mode.

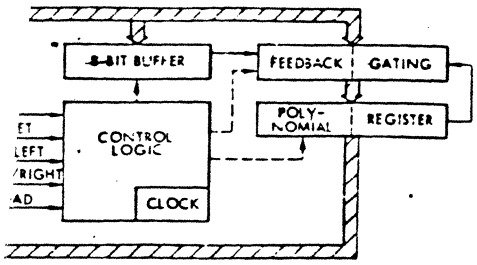
RESET — This command overrides all other command bits and clears the buffer and polynomial registers to all 0's. The RESET bit is high for this command.

GO — The buffer register is loaded with a data byte from the I/O bus. The data byte is then encoded with polynomial register contents as specified by EOR and CRC16 bits.

- EOR = High — Exclusive OR of data byte with right byte of polynomial register.
- CRC16 = High — Generate polynomial check code of data byte and polynomial register contents.

If both EOR/LEFT and CRC16/RIGHT bits are low, the encoding mode is determined by the last mode command given.

The polynomial register contents may be strobed onto the I/O bus by providing the gating appropriate to the desired computer. A separate I/O signal must be used for this function since no command bits have been assigned for this purpose. This provides more flexibility in interfacing the polynomial code device to computers with different word sizes.



Polynomial code device block diagram

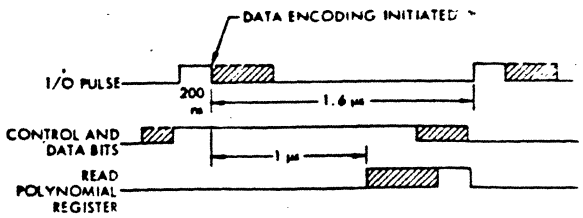


Fig. 6 Timing diagram

ce and Time Constraints

I/O timing constraints are shown in Fig. 6. and bits must be stable at the leading edge e I/O pulse. Encoding of a data byte is initie trailing edge of the I/O pulse and requires completion. A minimum I/O pulsewidth of its the device to a 1.8-μs repetition rate, which in the required I/O instruction execution time small computers. Consequently, this assures the of a data byte in one I/O instruction time. time consuming I/O device skip loop programne contents of the polynomial register may be onto the I/O bus 1 μs after the trailing edge of pulse.

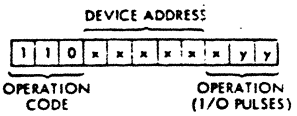


Fig. 7 IOT instruction format

uter Implementation

lynomial code device has been interfaced and th a Digital Equipment Corp PDP-8/L computer unctions in a remote terminal environment and ntrol several remote devices and a data communiine. Briefly, the computer's characteristics are ord, 1.6-μs memory cycle, and one programmable (ie, the AC register). The instruction set is very and contains one I/O instruction (ie, the IOT on). All programmed I/O data must pass through which provides 12 buffered data output lines data input lines. The IOT instruction format con-3-bit operation code, a 6-bit device address, and operation specification as shown in Fig. 7.

execution of the IOT instruction places 6-bit ddress on the device address lines monitored by O device. When a device recognizes its assigned address, it gates the last three bits of the IOT ion into the device control logic. Each operation ifies an action associated with the device. The peration bits correspond to three control pulses ccur in sequence and may be combined in the truction to affect one, two, or three sequential perations.

all I/O data must pass through the AC, the con- data bits must be loaded into the AC prior to

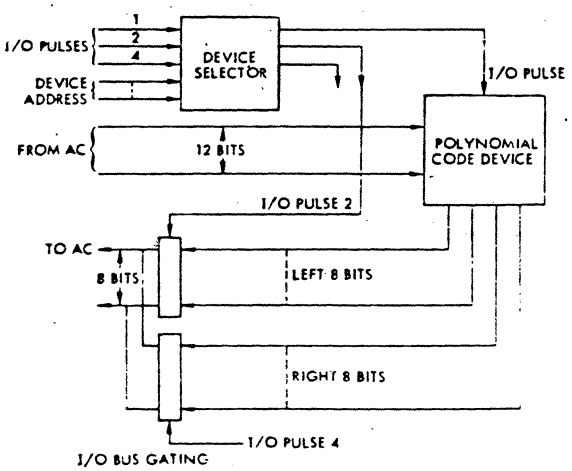


Fig. 8 PDP-8/L interface to device

an IOT instruction. The general interfacing of the polynomial code device to the computer is shown in Fig. 8.

Notice that only either the right or left byte from the polynomial register can be read at one time since the

PDP-8/L Data Encode Program

Instructions	Comment
CLA	Clear AC
TAD K14$\phi\phi$	
TAD RIGHT	Get old right check byte
IOT 1	Load right byte of polynominal register
TAD K24$\phi\phi$	
TAD LEFT	Get old left check byte
IOT 1	Load left byte of polynominal register. Sets device to CRC16
TAD DATA	Get data byte
IOT 5	Generate check bytes, clear AC, load AC with new right check byte
DCA RIGHT	Deposit new right check byte, clear AC
IOT 2	Load AC with new left check byte
DCA LEFT	Deposit new left check byte, clear AC

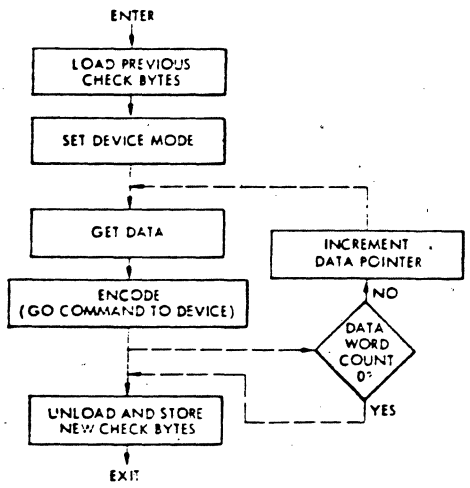


Fig. 9 Device data encode flowchart

data to the AC is only 12 bits wide. This is accomplished by assigning I/O pulses 2 and 4 to read left and right, respectively. For a computer with a 16-bit I/O data bus, only one control pulse would be needed to read both bytes of the polynomial register.

Software Considerations

Device Programming

The polynomial code device is easily programmed on any general-purpose minicomputer. A flowchart for generating check bytes given one data byte is shown in Fig. 9. The corresponding program can be implemented with six to ten instructions on most small computers. Notice that no decision branches are required and that no device flag test loop is required per data byte. This results from the assurance that data byte encoding is completed within one instruction time. The dotted line branches in Fig. 9 are included to illustrate the additional flowcharting required to generate check bytes for a block of data bytes. A corresponding error detection program can similarly be implemented with this flowchart with the additional test-for-zero instructions on the check bytes as the last step. A specific programming example for a 1-byte encode operation is shown in the table. The execution time on a PDP-8/L computer is 39.4 μs/data byte and includes seven memory references.

Hardware vs Software

The justification of the effort and expense of developing special-purpose hardware for a computer system is inherently influenced by the equivalent software required to perform the same task. Hardware/software tradeoffs consequently become central to the decision between hardware and software implementation. In remote computing applications where a small computer is used to control such items as peripherals, format communications data, and translate code sets, gross amounts of time can be consumed in communication line error-checking routines. This effectively lowers the data transmission rate which the computer can maintain while satisfying its other commitments. Cost and performance criteria may force the designer to accept lower communication line bandwidth and/or slower remote peripheral devices. This in turn enhances the I/O throughout limitations of the central computer facilities, particularly for a large number of remote stations.

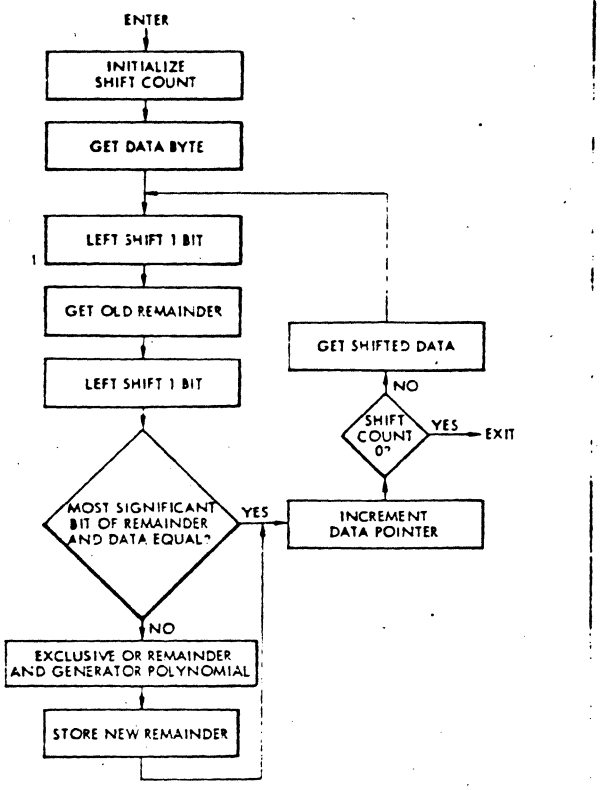


Fig. 10 Programmed data encode flowchart

Software implementation of polynomial code generation and detection is possible on any machine with shift and basic logical instructions. The algorithm consists of three basic operations: (1) shift data byte and remainder, (2) compare most significant bits, and (3) Exclusive OR remainder and generator polynomial (dependent on compare test).

A flowchart for encoding one data byte is shown in Fig. 10. Notice that eight iterations of the shift loop are required for every data byte. Worst-case data encoding requires eight iterations with an Exclusive OR for each iteration. A typical 16-bit computer with an Exclusive OR instruction and two or more programmable registers would require 80 to 120 instruction executions per data byte. Assuming a polynomial code I/O device connected to the same computer could be programmed with six to eight instructions, a performance increase factor of 10 to 20 is obtained.

The performance increase factor for a PDP-8/L is much higher, due to its 12-bit word size. For an integral number of byte generator polynomials, an integral number of 12-bit memory locations must be used for check-byte storage. In this case, only 8 of the 12 bits are used since packing and unpacking 8-bit bytes is time consuming on a PDP-8/L, which probably represents the worst-case programming effort for implementation of polynomial code generation. A PDP-8/L program that I wrote requires 432 instruction executions and 1.1 ms execution time per data byte in the worst case. Assuming 20% of the total computer time is allotted for error checking, about 5 ms/data byte of computing would be required. In this case, the PDP-8/L could support a 200-baud communication line, which hardly satisfies remote job entry requirements. Using the PDP-8/L execution times with and without the polynomial code device, a performance increase ratio is obtained of 1.1 ms/39.4 μs = 28. The larger factor for the PDP-8/L results from the lack of an Exclusive-OR instruction. The equivalent of an Exclusive OR requires 10 PDP-8/L instructions and 80 extra instructions per data byte in the worst case.

The polynomial code generator/detector described earlier has been built. The required logic was assembled on three 2½ x 5", 36-pin PC cards; 32 TTL IC packages were used. The PC cards were compatible with DEC M-series logic cards and could conveniently be placed within a PDP-8/L extended memory cabinet. In addition, two unique logic cards were required to satisfy PDP-8/L I/O bus conventions (ie, device selection card and an open collector bus driver card).

Conclusion

The attempt has been made here to provide the reader with an understanding of polynomial error code generation and detection with minimal mathematical rigor. Practical aspects and application of this error-coding technique have been considered in a real-world situation involving remote job entry computing terminals. It has been shown that significant performance improvements can be realized by "hardwired software" capable of implementing a specific generator polynomial for error detection. The significant advantages of the polynomial code device described are: (1) inexpensive ($200 for materials, parts, card assembly), (2) generalized I/O interfacing to small computers, (3) elimination of costly error encoding/detecting software, (4) time-shared capability between several communication lines, with minimum software, and (5) decreased number of undetected errors passed through remote job entry and concentrator subsystems.

The development of more general-purpose logic for implementing any arbitrary generator polynomial has been initiated. Each PC card will contain four bits of the polynomial register. A polynomial register of any length may be assembled and hardwired to select Exclusive OR bit positions or not and to select the last stage of the polynomial register.

References

1. W. W. Peterson, *Error Correcting Codes*, The MIT Press, Cambridge, Mass, 1961
2. W. W. Peterson, "Cyclic Codes for Error Detection," *Proceedings of the IRE*, Jan 1961, pp 228-235

A NEW LOOK-AHEAD CODE FOR INCREASED DATA DENSITY

George V. Jacoby
Senior Member IEEE
ISS, Div. of Sperry-Univac, Cupertino, CA 95014

Abstract

This paper describes a novel run-length limited code, termed 3PM. A group of three data bits is converted into six code bits which are represented by the presence or absence of signal transitions. At least two zeros are maintained between two consecutive ones, that is a minimum distance of three positions between transitions, resulting in great reduction of pulse crowding. The minimum distance is assured by a unique merging rule at the boundary of adjacent code words. This rule distinguishes the code from both fixed and variable length codes and results in very simple encoding and decoding algorithms. An actual 50% density increase has been accomplished in saturation recording by using the 3PM code in combination with other electronic techniques. The new code is used in a current ISS/Univac high density disk storage system, featuring 2500 bits/cm (6300 BPI) linear density, 10 Mbits/sec data rate, 338 MByte capacity and one bit in 10 billion raw error rate on conventional Mod-11 head/disk interface.

Desirable Code Properties

The basic problem of this paper is increasing the data density in digital magnetic recording without changing the physical parameters, that is flying height, head inductance, oxide thickness and magnetic properties, in an existing head medium interface, while maintaining the same reliability. The first task is the selection of a code that is capable of increasing the linear bit density with the least possible degradation in the analog signal waveform. This means that the new code should not aggravate the write conditions, and it should not increase pulse crowding at the higher data density.

At this point, it is desirable to survey the major parameters of some existing codes. The definitions follow those of Patel.[1] A data group, consisting of m data bits, is converted into a code group of n code bits. The ratio m/n is called the code rate. The code bits of a code group, ones and zeros, are recorded by the presence or absence of magnetization transitions in n uniformly spaced positions. Two consecutive ones are separated by at least d but not more than k zeros so that the minimum distance between two consecutive transitions is (d+1) positions. This property makes the code run-length limited.[2,3]

If the time interval of one data bit is T sec, the minimum and maximum time intervals between transitions are: Tmin = (m/n)(d+1)T sec and Tmax = (m/n)(k+1)T sec. The efficiency of the code is measured by the density ratio, defined as:

$$DR = \frac{\text{data density}}{\text{max. transition density}} = \frac{Tmin}{T} = Tminr = (\frac{m}{n})(d+1)$$

where T is the data bit period and Tminr is the ratio of minimum time interval to the data bit period. The density ratio expresses the number of data bits per flux reversals or transitions, spaced at minimum distance. Additional important parameters are the detection window, W = (m/n) · T and the clock rate, which is the reciprocal of the window. The window equals the time duration of one code bit or the length of one position. The code rate (m/n) is a measure of the relative window (Wr), i.e., the detection window related to the data bit period: Wr = W/T = m/n.

Many run-length-limited codes have been reported and analyzed in the past.[1-9] The write data waveforms and major code parameters of some codes are shown in Figure 1 and Figure 2 for the same data density.

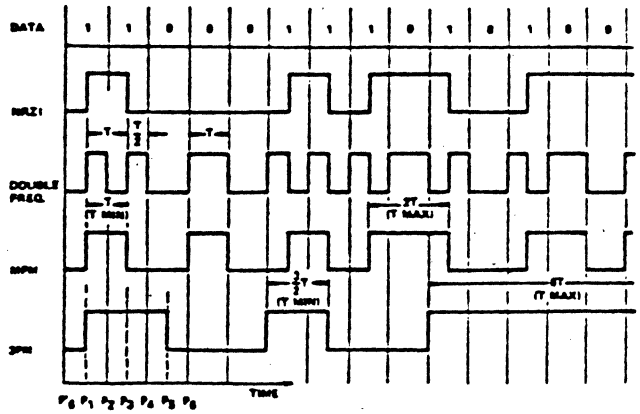


FIGURE 1. Write Data Waveforms

CODE	m	n	CODE RATE	d	k	T MIN	T MAX	DENSITY RATIO	WINDOW	CLOCK RATE
NRZI	1	1	1	0	∞	T	UNLIMITED	1	T	NO CLOCK
DOUBLE FREQ.	1	2	0.5	0	1	T/2	T	0.5	0.5T	2/T
MFM	1	2	0.5	1	3	T	3T	1	0.5T	2/T
GCR 4/5	4	5	0.5	0	2	0.5T	2.5T	0.5	0.5T	1.25/T
3PM	3	6	0.5	2	11	2T/3	8T	1.5	0.5T	3/T

FIGURE 2. Major Code Parameters. Bit Rate: $\frac{1}{T}$

In the write mode in saturation recording, the most critical parameter is the maximum flux reversal density that the head is capable of writing and the medium can support. This is limited by the inductance of the head and by pulse crowding, demagnetization and other nonlinear effects of the medium that increase with flux density. It is therefore, desirable to maintain the same flux density in a system, where the head and medium cannot be changed. The highest data density will be achieved, if a code with the largest Tmin or DR value is selected.

In the read mode, high density ratio is again important as will be seen later. It is equivalent to large Tminr, which expresses the fact that pulse crowding is removed or limited. The other desirable factor is a large window that results in relative insensitivity to noise. However, large window is less efficient in increasing the data density than large Tmin. This is illustrated by the GCR 4/5 code that has a 60% wider window, but reduced Tminr value, relative to MFM. It can increase the data density over MFM only by 8%, as shown in a study by Tamura et al.[5] A large value of Tminr is equivalent to a low flux density resulting in reduced pulse crowding. Tminr is effective in both writing and reading, while the relative window is effective only in reading.

The old codes without exception have density ratios less than, or at best, equal to one. The 3PM code accomplishes 50% greater DR than MFM, while maintaining the same window and the same clock rate. This is due to using three

positions (d+1) in the Tmin distance, while MFM uses only two positions (Figure 1). This basic feature of the code originated its name: Three-position modulation (3PM). In a general sense, there is a direct evolution from double frequency to MFM and then to 3PM. The number of code bit positions in the Tmin distance increased from one to two and then to three, while the window remained the same. The density ratio increased in proportion to Tmin. It is seen that if the minimum transition interval of an existing MFM recording system is maintained but the code is changed to 3PM, the potential exists of increasing the data density by 50%. The maximum transition interval (Tmax) in the 3PM code is considerably longer than in MFM. This requires a more tightly controlled phase-locked oscillator that is capable of maintaining accurate clocking over a maximum of 12 clock periods. This problem has been solved successfully in the window detection of the 3PM code.

The 3PM code write data waveform has considerable DC-content and digital sum variation, as most run-length-limited codes do. This is of no concern in writing since the write head is DC-coupled. In the read mode the question of digital sum variations enters into consideration only if we want to restore the write data waveform.[10] However, if we reconstruct the read pulses by proper spectral shaping to the point that they do not interfere, the location of the pulse peaks contain the data accurately and conventional peak detection can be used. In this process, there is no need to consider DC-content or DSV of the original write data waveform. The optimum spectral shaping is treated in a companion paper, submitted by W.D. Huber. Based on his work the following general equation has been derived for the code parameters, when optimum spectral shaping, or optimum equalization is applied to the analog waveshape. The relative importance of the code and analog waveshape parameters in the read mode can be observed from this relationship:

$$\text{Max. data density} = (\text{flux rev. density}) \cdot DR =$$

$$= \frac{1}{v \cdot T_{min}} \left(\frac{m}{n}\right)(d+1) = \frac{K}{v \cdot T_{50}} (SNR_i)^q \left(\frac{m}{n}\right)(d+1)^{1-q} =$$

$$= \frac{K}{v \cdot T_{50}} (SNR_i)^q \, (DR)^{1-q} \, W_R^q$$

Where: $K = 0.4$, $q = 0.37$ for $(d+1) = 2$
$q = 0.4$ for $(d+1) = 3$

v is the head to medium speed, T_{50} is PW_{50}, measured in time, of the isolated pulse. SNR_i is the input signal to noise ratio at the head, expressed as peak single pulse amplitude to RMS noise. The above relationship has been worked out for an error rate (without error correction) of one bit in 10 billion and with a generous 33% window margin to cover fixed bit shift, jitter and circuit tolerances. It is valid for a range of SNR_i between 30 and 34 dB. The second part of the equation focuses attention on the analog properties of the waveshape. It is seen that density ratio or minimum transition distance is more efficient in increasing the data density than the relative window. The 3PM code increases the density ratio by 50% over MFM, while leaving the relative window the same. An actual data density increase of 1.56-times was achieved by using this code, combined with optimum spectral shaping equalization at 31 dB SNR_i. Thus, 3PM combines the advantages in both write and read through its Tmin = 1.5T feature and appears to be the best code to increase data density without changing the head/medium interface.

3PM Code Algorithm

A code with properties described above can be implemented with a number of different algorithms. The most straightforward method is fixed length block coding. This, however, is not efficient; requiring large memory or complex logic. Variable length codes are much more efficient and can be implemented with less memory. However, the logical operations in decoding, due to the nature of variable length groups that must be handled, could be quite complex. Furthermore, special precaution must be used against error propagation, which can be limited in general to the number of bits in the longest word.

The unique novel algorithm of the 3PM code is similar to fixed length block encoding and decoding with one additional rule. The algorithm converts three bit data groups into six bit code groups. The encoding is explained by looking at Figure 1 and Figure 3. The space allowed for a word of three data bits is divided into six equidistant positions: P_1 to P_6. The basic encoding table for one word is shown in Figure 3.

BINARY DATA WORD			TRANSITION POSITIONS					
			P1	P2	P3	P4	P5	P6
0	0	0	0	0	0	0	1	0
0	0	1	0	0	0	1	0	0
0	1	0	0	1	0	0	0	0
0	1	1	0	1	0	0	1	0
1	0	0	0	0	1	0	0	0
1	0	1	1	0	0	0	0	0
1	1	0	1	0	0	0	1	0
1	1	1	1	0	0	1	0	0

FIGURE 3. Initial Encoding Table

BINARY DATA WORD			INFLUENCE OF ADJACENT WORDS		TRANSITION POSITIONS						
			PRECEDING	FOLLOWING	P0	P1	P2	P3	P4	P5	P6
0	0	0	X	0	0	0	0	0	0	1	0
0	0	0	X	1	0	0	0	0	0	0	1
0	0	1	X	X	0	0	0	0	1	0	0
0	1	0	X	X	0	0	1	0	0	0	0
0	1	1	X	0	0	0	1	0	0	1	0
0	1	1	X	1	0	0	1	0	0	0	1
1	0	0	X	X	0	0	0	1	0	0	0
1	0	1	0	X	0	1	0	0	0	0	0
1	0	1	1	X	1	0	0	0	0	0	0
1	1	0	0	0	0	1	0	0	0	1	0
1	1	0	1	0	1	0	0	0	0	1	0
1	1	0	0	1	0	1	0	0	0	0	1
1	1	0	1	1	1	0	0	0	0	0	1
1	1	1	0	X	0	1	0	0	1	0	0
1	1	1	1	X	1	0	0	0	1	0	0

1 : YES
0 : NO
X : DON'T CARE

FIGURE 4. Final Encoding with Merging

Minimum two zeros are maintained between adjacent ones. The boundary position (P_6) is occupied by zeros in all code words. In a sequence of words, where a one occurs at P_5 of the present word and also at P_1 of the following word the d=2 condition would be violated. The special rule of the 3PM code provides that in this case the P_5 transition of the present word and the P_1 transition of the following word will not be written in their original locations but will be replaced by a single transition at P_6. The two original transitions, P_5 and P_1, will be merged into one transition at P_6. The P_6 position in the initial encoding table is reserved for this merging operation.

The results of the final encoding, after the merging rule has been carried out, is shown in Figure 4. Here all combinations of binary data words of three bits each are shown together with the influence of adjacent data words. If the preceding word ends in P_5, a P_1 transition of the present word will be shifted to the sixth position of the previous word denoted by P_6'. Similarly, if the following word starts with P_1, a P_5 transition of the present word will be shifted to P_6. The result of this merging rule is that any number of octal data words can be catenated, while simultaneously maintaining the d=2 condition everywhere in the sequence. The encoding logic that implements this rule is very simple. It has to look back to the P_5 position of the previous word and look ahead to the P_1 position of the following word, thus dealing with 9 positions simultaneously. This way, the words are chained to one another in a natural way, preserving the fixed block length property of the code.

Decoding is done in a similar manner. The 7 transition positions are observed simultaneously. They uniquely identify the binary data word, as shown in Figure 4. Decoding, therefore, is state independent, maintaining the advantage of fixed length blocks, that greatly simplifies the logic. The decoded data words are identified by a word clock, derived from a general clock system that runs synchronously with the transitions. Error propagation is limited to a maximum of three data bits (one word). This may result from the drop-out or drop-in of a single transition, or from one transition shifting by one position in detection. This is better than the limit of error propagation in variable length codes with the same parameters.

It is interesting to compare the code with the smallest fixed-length block code that achieves the run lengths of 3PM. To make the comparison as stringent as possible consider state-dependent codes where past as well as present information may be used in forming each code word. By means of a program devised at the Sperry Research Center, it was found that the shortest block code with minimum spacing of 1.5 data-bit times and rate 0.5 converts each 8 bits of data into 16 bits of code.[11] Thus, this code requires 12,288 bits of storage for its encode/decode tables, or logic sufficient for a rather complex 8-input, 16-output switching circuit. With this comparison the simplicity of the 3PM logic can be properly appreciated.

This system has been implemented in a recently released ISS/Univac high density disk file, featuring 2500 bits/cm (6300 BPI) data density, 10 Mbits/sec data rate and 338 MByte capacity on conventional Mod-11 type head disk interface. In a companion paper, A. Geffon discusses the complete system implementation. Figure 5 shows the analog waveshape of a pattern sequence after equalization. It is seen that the pulses are completely separated, the interaction is completely removed by optimum spectral shaping. As a result, the pulse peak locations are restored to great accuracy. A number of different time intervals between adjacent pulse peaks can be observed which is characteristic of this code.

Conclusion

A new code has been described that uses the same code rate as the MFM code, but increases the minimum distance by 50%, thereby achieving a density ratio of 1.5 data bits per minimum transition distance. The actual achievable data bit density is related to the code parameters and signal to noise ratio in general terms. The new code has been implemented in a commercial digital recording disk file and reliably accomplished 56% increase in bit density with the same head/disk interface, used by an earlier model with MFM code.

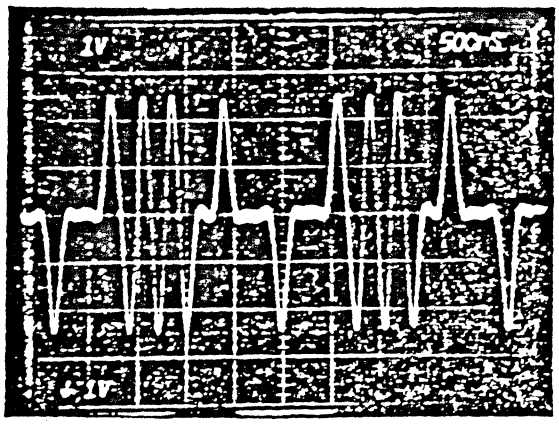


FIGURE 5. Equalized Analog Waveshape

Acknowledgment

The author wishes to express his thanks to Dr. M. Cohn of the Sperry Research Center for his advice in the field of code theory and to C.A. Bates, A.P. Geffon and W.D. Huber of ISS for the implementation of this code.

References

1. Patel, A.M., "Zero-Modulation Encoding in Magnetic Recording," IBM Journal of Res. & Dev., Vol. 19, July 1975, pp. 366-378.
2. Tang, D.T., "Run-Length Limited Codes," IEEE International Symposium on Information Theory, 1969.
3. Franaszek, P.A., "Sequence-state Methods for Run-Length-Limited Coding," IBM Journal of Res. & Dev., Vol. 14, July 1970, pp. 376-383.
4. Cullum, C.D., "Encoding and Signal Processing," Annals of the New York Academy of Sciences, Vol.189, January 3, 1972, pp. 52-62.
5. Tamura, T. et al., "A Coding Method in Digital Magnetic Recording," IEEE Transactions on Magnetics, September 1972, pp. 612-614.
6. Kiwimagi, R.G. et al., "Channel Coding for Digital Recording," IEEE Transactions on Magnetics, September 1974, pp. 515-518.
7. Spitzer, C.F., "Digital Magnetic Recording of Wideband Analog Signals," Computer Design, September 1973, pp. 83-90.
8. Mallinson, J.C. and Miller, J.W., "On Optimal Codes for Digital Magnetic Recording," Proceedings of the 1976 Birmingham Conference on Video and Digital Recording, pp. 161-169.
9. Horiguchi, T. and Morita, K., "On Optimization of Modulation Codes in Digital Recording," IEEE Transactions on Magnetics, November 1976, pp. 740-742.
10. Jacoby, G.V., "Signal Equalization in Digital Magnetic Recording," IEEE Transactions on Magnetics, September 1968, pp. 302-305.
11. Lempel, A. and Cohn, M., "Run-Bounding Binary Block Codes for Magnetic Recording," Internal Sperry Rand Research Report, RM-71-19, April 1971.

MEMOREX

Interoffice Correspondence

To: Dave Gordon

Date: May 30, 1980

From: Bob Beckenhauer

Copy to: E. Asato, W. Cheney, S. Dinsmore, D. Huber, H. Kwok, D. Moynahan, L. Raney, R. Singleton, T. Yung

Subject: 2,7 Encoder/Decoder Circuits

The purpose of this memo is to summarize the results obtained from an experimental breadboard of a 2,7 encoder/decoder circuit, for possible application to future Memorex disk files.

Schematics[1] for the breadboard were provided to Tony Yung and Hoover Kwok at three stages in the development process - original design level, original breadboard build level, and functional breadboard level (after debug). The functional breadboard schematics were provided on May 28, 1980, after the results described in this memo were obtained.

TEST CIRCUIT

The test circuit consisted of essentially those functions described in the block diagram on the following page.

The on-board test oscillator provided a 2F frequency of approximately 25MHZ. While this frequency is slightly lower than the IBM 3370 rate, it should be close enough for the purposes of this demonstration. Also, using the MECL 10K logic family, much higher switching rates should be attainable without major problems.

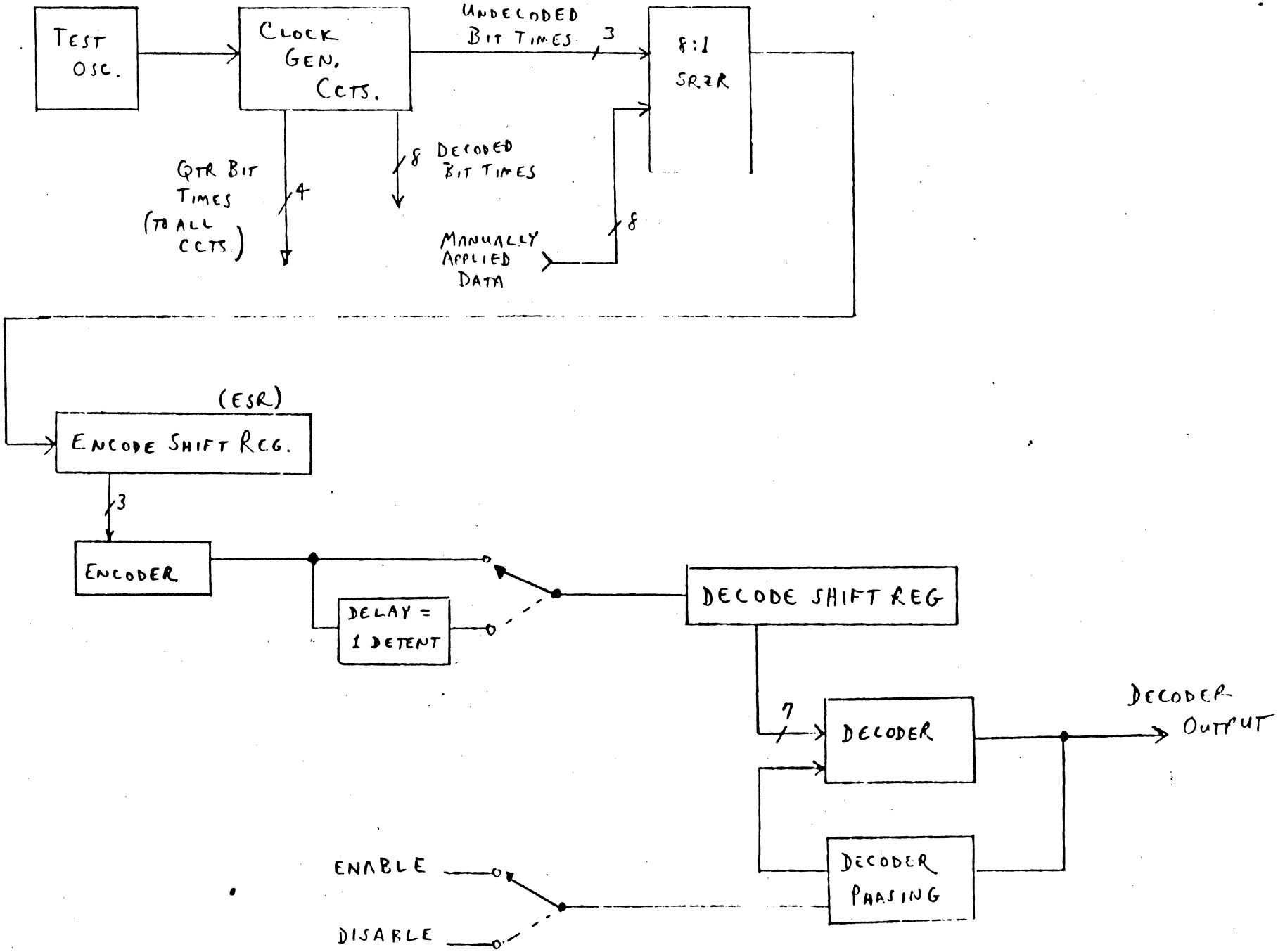
(Futhermore, it is expected that timing changes would be necessary anyway to integrate the circuits into a disk controller. This tends to reduce the value of detailed timing analyses on an isolated board.)

The main purposes, therefore, of the breadboard were to establish:

1) A working MECL 10K prototype which would encode and decode all words in the Franaszek 2,7 code dictionary.[2,3]

2) A simple, reliable decoder phasing technique for read-back.

3) A potential scheme for generating system timing pulses for the associated disk controller.



2,7 CODE TEST CIRCUIT FLOCK DIAGRAM

RESULTS

The following chart describes the 2,7 code dictionary:

DATA WORD		CODE WORD	CODE WORD BY DETENT ASSIGNMENT
1 0	Hi Rep Rate →	0 1 0 0	0 T_1 0 0
0 1 0		1 0 0 1 0 0	To 0 0 T_1 0 0
0 0 1 0		0 0 1 0 0 1 0 0	0 0 To 0 0 T_1 0 0
1 1		1 0 0 0	To 0 0 0
0 1 1	Lo Rep Rate →	0 0 1 0 0 0	0 0 To 0 0 0
0 0 1 1		0 0 0 0 1 0 0 0	0 0 0 0 To 0 0 0
0 0 0		0 0 0 1 0 0	0 0 0 T_1 0 0

A "1" in the code word represents a flux transition recorded on, or read back from, the disk surface. Since it is a 1/2 rate code, there are two possible detent assignments, which are arbitrarily labelled To and T_1.

In the photographs of figures 5 and 6, an encoded data pulse which aligns with Clock B represents a 1, or transition, in the To detent. An encoded data pulse which is in the space between Clock B pulses represents a 1, or transition, in the T_1 detent.

Since the breadboard circuits were capable of repeatedly serializing a byte of data (applied by manual switches), the test patterns selected were combinations of either two or three 2,7 words which had their word boundaries at the data byte boundary. The results are shown in figures 1 through 4. (Note that this by no means exhausts all the possible permutations of 2,7 words, but is a reliable indication that the hardware's encode/decode algorithm works. It may still be possible to have data-dependent timing problems.)

Figure 5 shows the generation of the highest rep rate pattern of 2,7 encoded data in relation to the B Clock. This is obtained during the time that the word 010 is being repeatedly encoded.

Figure 6 shows the lowest possible rep rate, obtainable by encoding the word 0011 repeatedly.

Figures 7 through 10 illustrate the results of the decoder phasing technique. The test sequence was as follows:

1) Circuits are repeatedly encoding and decoding a byte of 1's. (fig. 7)

2) A one-detent (1/2 bit time) delay is introduced into the undecoded data stream. Note new timing relationship between - BIT TIME Ø (scope trace 1) and - DECODER A*0 GATE OUT (trace 3), in fig. 8, as compared to figure 7. This results in erroneously

decoding a stream of ones as alternate ones and zeroes. This is in accordance with the 2,7 dictionary:

DATA WORD	CODE WORD
1 0	0 1 0 0
1 1	1 0 0 0

To Detent ↑ ↑ T_1 Detent

3) In figure 9, the decoder phasing has been momentarily enabled, and then disabled, which causes the decoder to decode a stream of one's again (even though the undecoded data stream is now arriving at the decoder one detent time later).

4) In figure 10, the one detent delay has been removed, and the decoder again erroneously decodes alternate ones and zeroes, (since the decoder phasing was disabled), even though the undecoded data stream now arrives at the time it originally did.

This decoder phasing technique is the subject of a Memorex invention disclosure which is now being written. I recommend it as being simpler, and more reliable then the method used in the IBM 3370[4], since it requires no adjustment to the system clocks.

Figures 11 through 15 illustrate the timing relationships of the bread-board clocking system. Note that for 2F = 29 to 30 MHZ, the ABCD Clock pulses derived will only be about 17 nanoseconds in width. At higher data rates, to obtain greater margin for setting latches, the signals ± 1F and ± 1FD could be used instead of the ABCD Clocks. The 1F and 1FD pulses will be twice as wide.

COMMENTS RELATED TO SCHEMATICS (PAGE TS30)

In figure 14, timing margins can be improved by inserting more delay between the 2F input and the clock input to the 1F and 1FD - generating flip flops. This, of course, will affect all the downstream system timing.

In figure 15, note that the equivalent of Clock D (pin 8D-13, page TS30) is used to drive the bit counter. The bit times are generated by means of a 3-bit Gray code counter and a 3-to-8 bit decoder.

The 3-to-8 decoder is disabled by the same pulse which triggers the counter. The "dead time" between bits may be eliminated by keeping the 3:8 decoder enabled at all times. Since it is a Gray code counter, the decoder output bits should be glitch-free.

ADDITIONAL COMMENTS

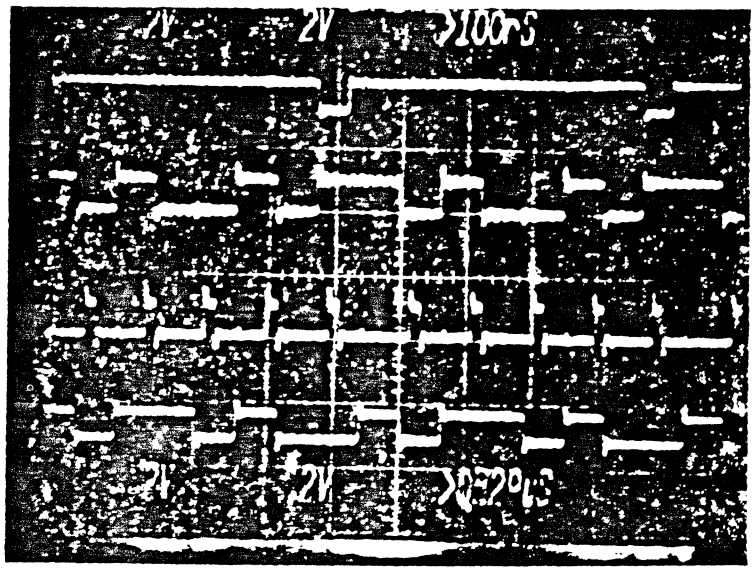
The 10141 shift register module type used in the breadboard is not a standard Memorex part. The 10141 seemed much more noise-prone than the other 10K modules. The worst noise was present on pin 1, particularly at the 2F shifting frequency. This was resolved by ensuring that the VCC pins (pins 1 and 16) were tied directly to the ground plane by a via-hole, or with heavy guage wire, rather than just connecting pins 1 and 16 together with light gauge wire.

The breadboard also contained hardware for generating the high-rep-rate 2,7 encoded pattern by means of a 3 bit shift register. This circuit was not tested, but should be very easy to implement. This method (or some equally trivial technique) is recommended instead of the six bit data loop used in the IBM 3370[5], since it would be simpler, more reliable, and avoid patent infringement.

Bob Beckhaney

BB:ps

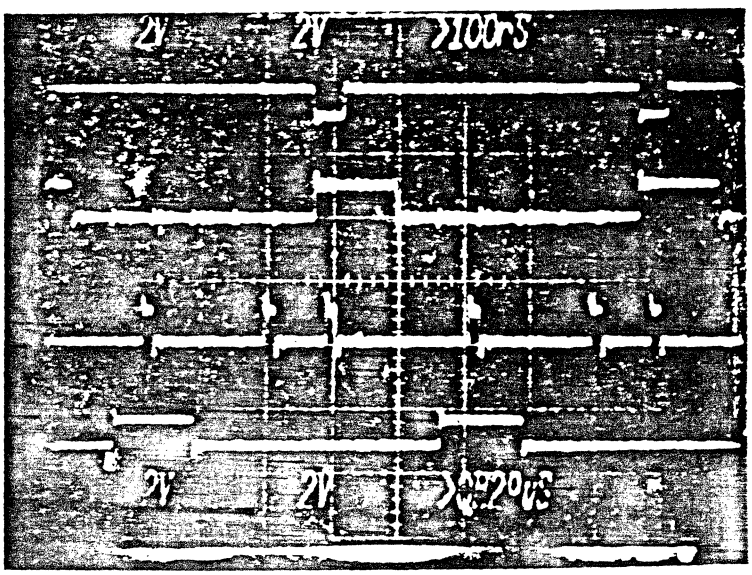
FIGURES 1 THROUGH 4 REPRESENT ENCODE & DECODE OF THE SEVEN WORDS OF THE 2,7 DICTIONARY.



C10.010.11

- BIT TIME Φ (PIN 10D-06, PG. TS 30)
+ 3RD BIT OF ESR (05B-02, TS10)
+ 2,7 ENCODED DATA (04C-07, TS20)
+ DECODED DATA (c2c-02, TS20)

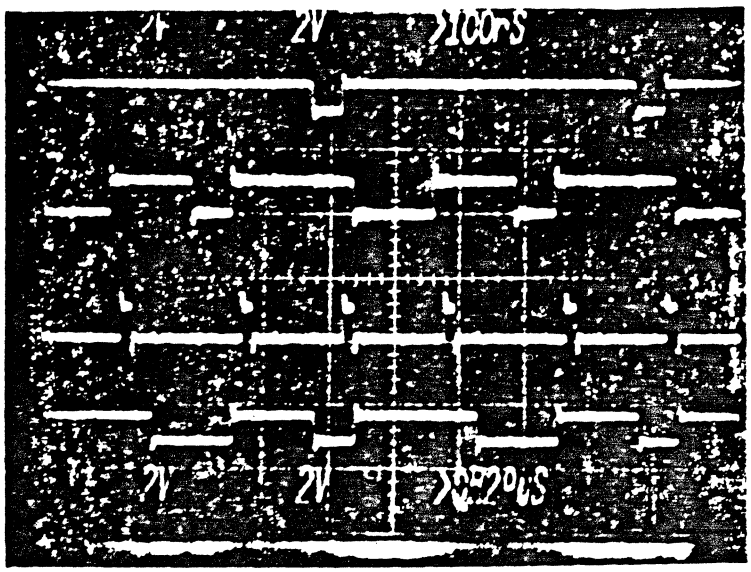
FIG. 1 — DATA BYTE = 010.010.11
ENCODES AS (T0) 00 (T1) 00 (T0) 00 (T1) 00 (T0) 000.



000.000.11

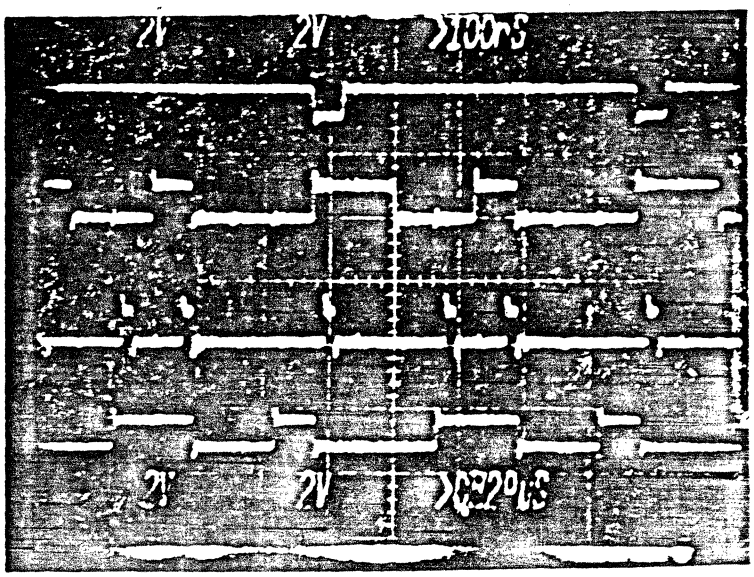
SAME PINS AS FIG. 1.

FIG. 2 — DATA BYTE = 000.000.11
ENCODES AS 000 (T1) 00 000 (T1) 00 (T0) 000



SAME PINS
AS FIG. 1

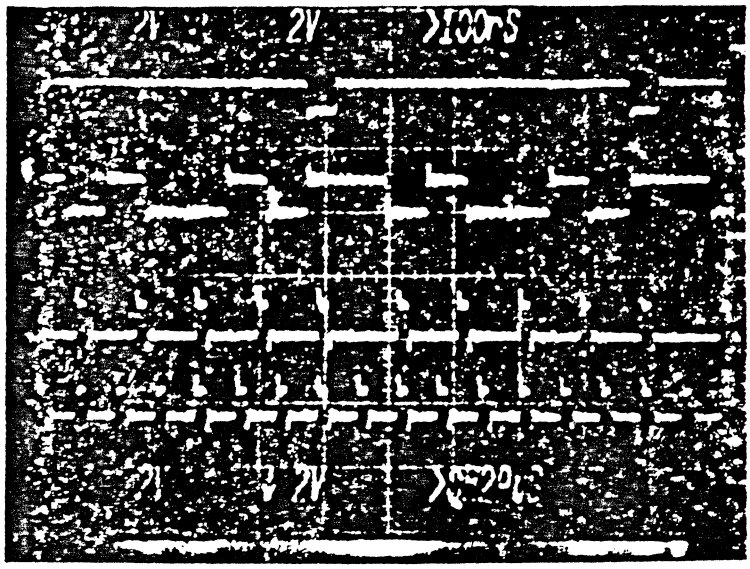
FIG. 3 — DATA BYTE = 011.011.10
ENCODES AS .CC(T$_0$)CCC. CC(T$_0$)CCC. C(T$_1$)CC



SAME PINS
AS FIG. 1.

FIG. 4 — DATA BYTE = CC1C.CC11
ENCODES AS CC(T$_0$)CC(T$_1$)CC. CCCC(T$_0$)CCC

FIGURES 5 & 6 SHOW GENERATION of HIGHEST & LOWEST 2,7 ENCODED DATA REP RATES.



HI REP RATE ENC OUT C1C C1C . 11

- BIT TIME ϕ (PIN 10D-06, PG. TS30)

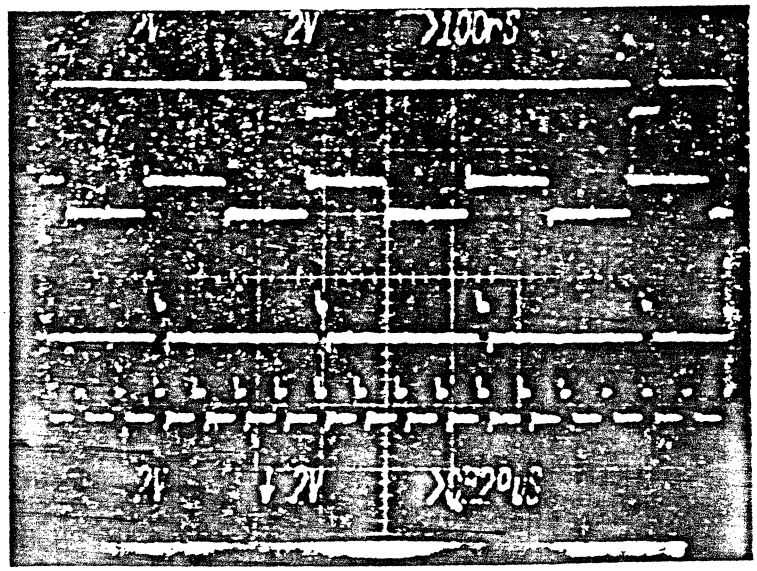
+ 3RD BIT OF ESR (053-02, TS1C)

+ 2,7 ENCODED DATA (04C-07, TS20)

- CLOCK B TO STROBE TΦ, INVERTED. (05B-05, TS1C)

FIG. 5 — DATA BYTE = C1C . C1C . 11

1ST 6 BITS ENCODE AS (T₀)CC(T₁)CO . (T₀)OO(T₁)CC , HIGHEST RATE



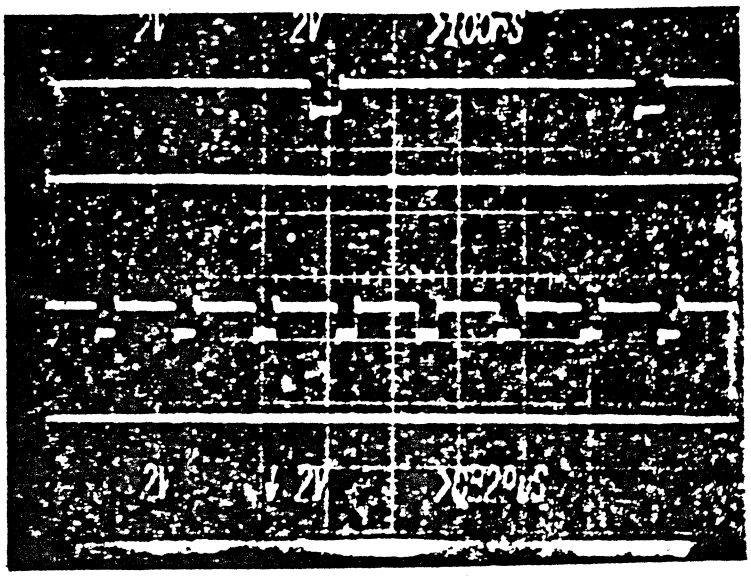
LO REP RATE ENC OUT CCII CCII

SAME PINS AS FIG. 5.

FIG. 6 — DATE BYTE = CCII . CCII

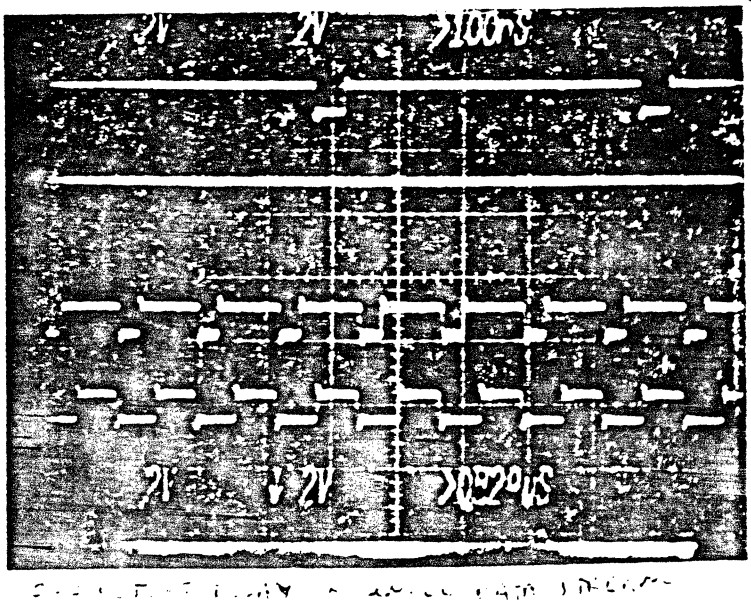
ENCODES AS OOOO(T₀)OOO . OOOO(T₀)OOO , LOWEST RATE

FIGURES 7 THROUGH 10 ILLUSTRATE THE DECODER PHASING (NOTE: 4TH TRACE IS INVERTED)



① - BIT TIME Φ (FIN 10D-06, PG. TS3C)
+ 3RD BIT OF ESR (08B-02, TS1C)
- DECODER A#C GATE OUT (03B-03, TS2C)
+ DECODER LATCHED DATA OUT, INVERTED (02C-02, TS2C)

FIG. 7 — SWITCH 4, MODULE 14C, TS4C, CLOSED.
SWITCH 5, MODULE 14C, TS4C, OPEN.



② SAME PINS AS FIG. 7.

FIG. 8 — SWITCH 4 OPEN.

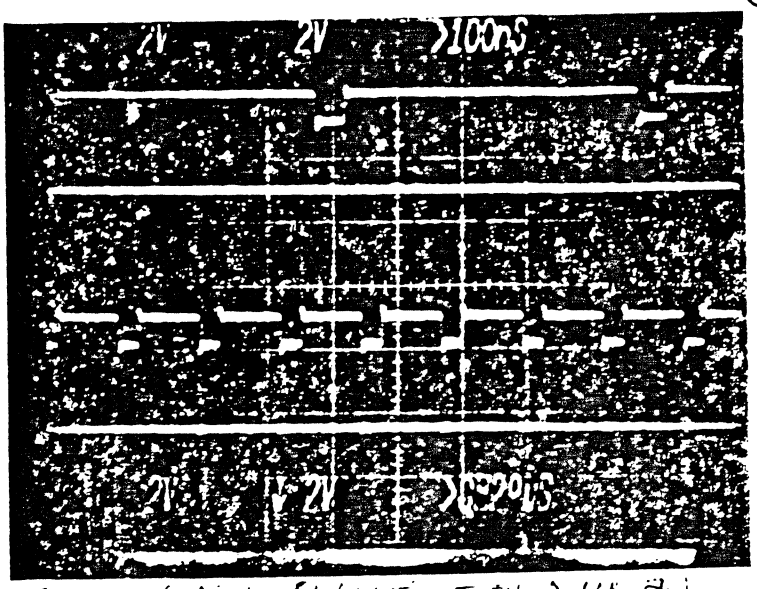


FIG. 9 — SWITCH 5 CLOSED, THEN OPENED AGAIN.

SAME PINS AS FIG. 7

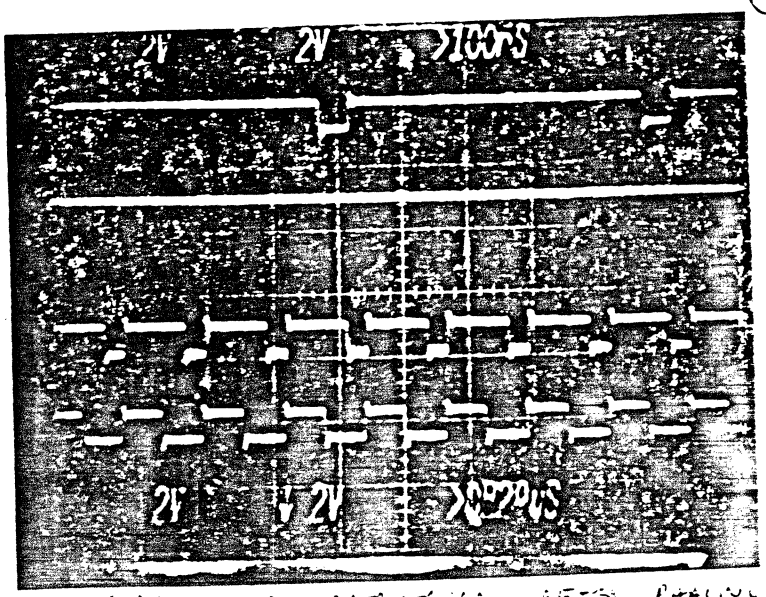
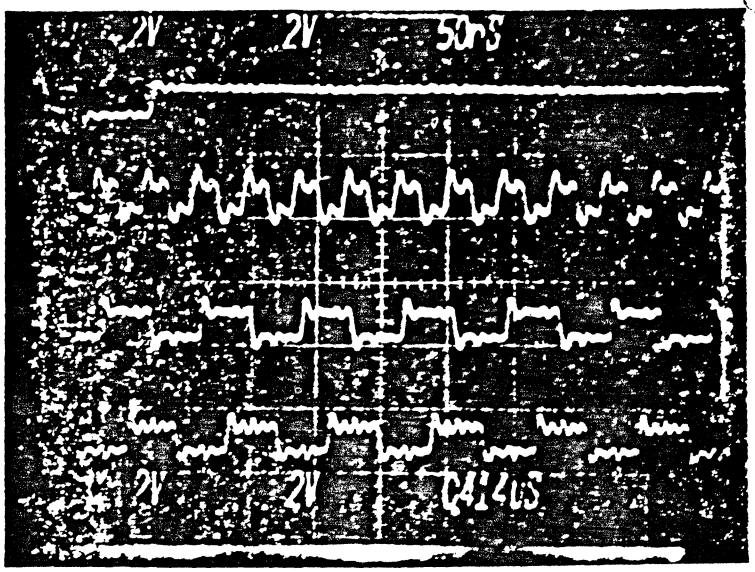


FIG 10 — SWITCH 4 CLOSED AGAIN.

SAME PINS AS FIG. 7.

FIGURES 11 THROUGH 13 SHOW CLOCK SYSTEM TIMING RELATIONSHIPS

(I)



- BIT TIME Φ
(PIN 10D-06, PG TS3C)

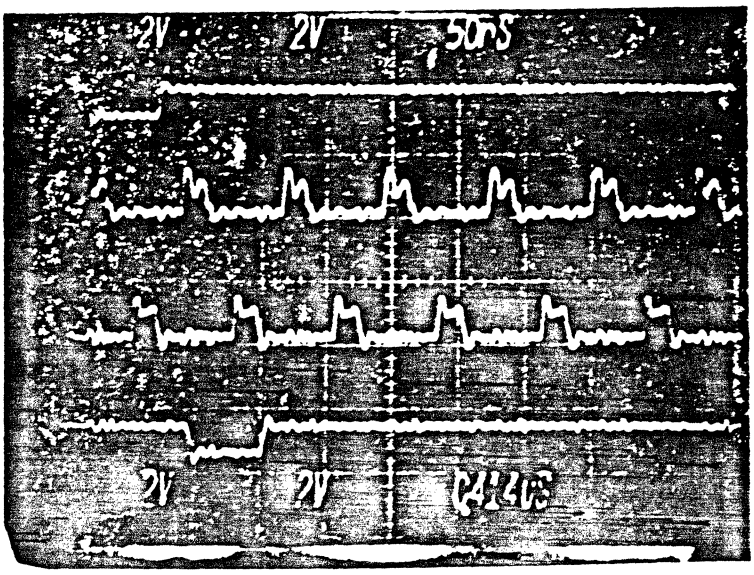
+ TEST OSC. IN
(07D-05, TS3C)

+ 1F
(08C-15, TS3C)

+ 1FD
(08C-02, TS3C)

FIG. 11 — RELATIONSHIP OF 2F INPUT TO 1F & 1FD.

(II)



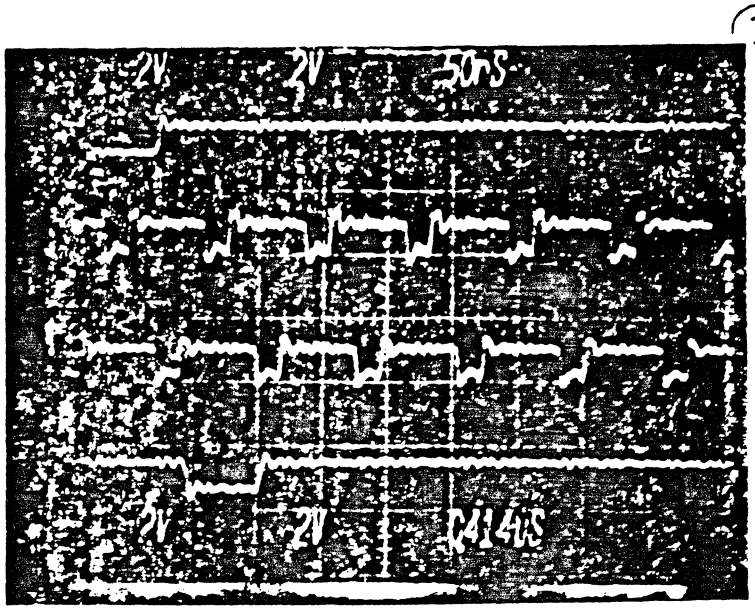
- BIT TIME Φ
(10D-06, TS3C)

+ CK A TO ESR
(06C-03, TS3C)

+ CK C TO PATH LATCHES
(06C-12, TS3C)

- BIT TIME 1
(10D-05, TS3C)

FIG. 12 — RELATIONSHIP OF A & C CLOCKS TO BIT TIMES.



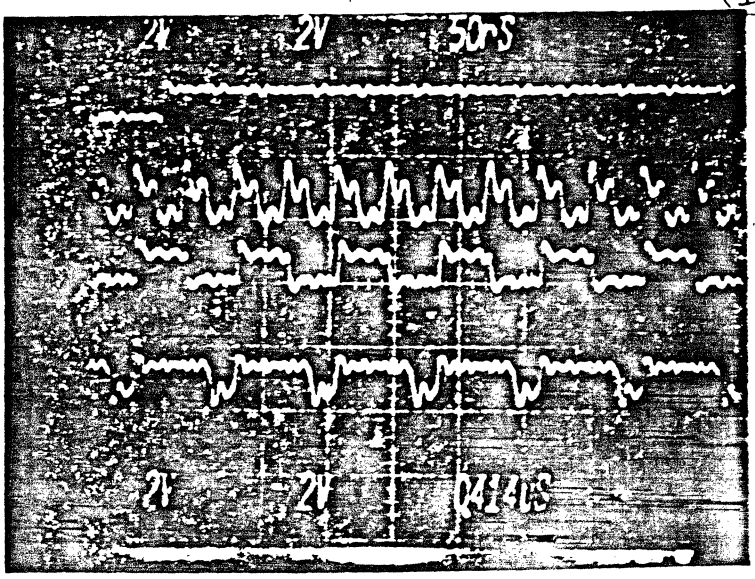
- BIT TIME ϕ
(PIN 10D-06, PG. TS3C)

- CK B TO STROBE Tϕ.
(05C-02, TS3C)

- CK D TO STROBE T1
(05C-14, TS3C)

- BIT TIME 1
(10D-05, TS3C)

FIG. 13 — RELATIONSHIP of B & D CLOCKS TO BIT TIMES.



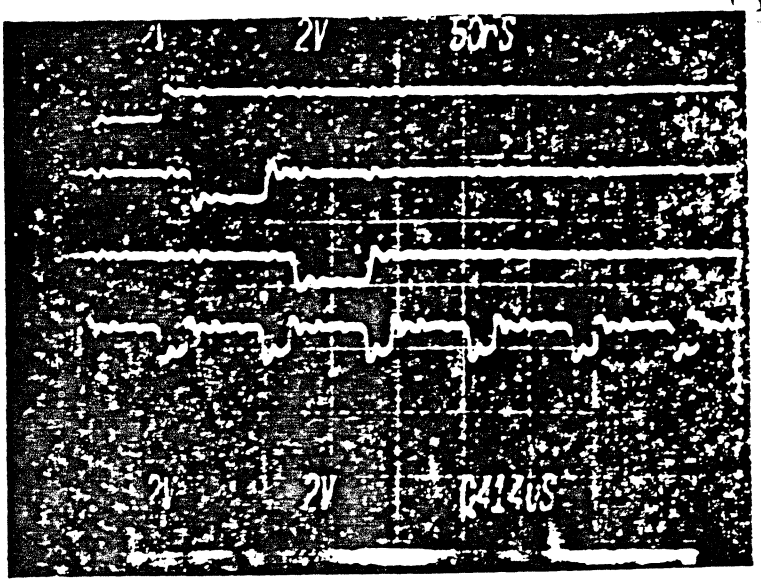
- BIT TIME ϕ
(PIN 10D-06, PG. TS3C)

+2F OSC IN AFTER ONE INVERSION
(05C-05, TS3C)

+1FD
(05C-06, TS3C)

- CLOCK B TO STROBE Tϕ
(05C-02, TS3C)

FIG. 14 — GENERATION OF CLOCK B.



(1)

- BIT TIME 4
(PIN 10D-C6, PC TS 3C)
- BIT TIME 1
(10D-C5, TS 3C)
- BIT TIME 2
(10D-C3, TS 3C)
- CLOCK D TO STROBE T1.
(05C-14, TS 3C)

FIG. 15 — RELATIONSHIP of CLOCK D TO "DEAD" TIME" BETWEEN BIT TIMES

REFERENCES

1. Schematics of 2,7 Breadboard Circuits, pages TS10, TS20, TS30, and TS40 (Beckenhauer)

2. U.S. Patent #3,689,899 (Franaszek)

3. U.S. Patent #4,155,760 (Eggenberger & Hodges)

4. IBM Technical Disclosure Bulletin, Vol. 21, No. 1, June 1978, Pages 326-327. "Phase Alignment of 1F and 2F Clocks" (Beckenhauer & Schauble)

5. U.S. Patent #4,146,909 (Beckenhauer & Schauble)