Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

SUBJECT: THE FERROELECTRIC SWITCH

To: Norman H. Taylor

From: Dudley A. Buck

Date: April 16, 1952

Abstract: A multi-position ferroelectric switch is proposed which can accomplish many of the switching tasks in an information handling system; in particular, it can select among the rows and columns of a ferroelectric memory. The logical circuitry of the ferroelectric switch can be painted directly onto the two sides of a thin ferroelectric sheet.

The non-linear electric displacement-versus-field characteristics of a ferroelectric dielectric can be utilized to construct a condenser whose capacitance is a function of the applied voltage. This phenomenon, which makes possible dielectric amplifier operation,[1,2] is the basis for operation of the switch to be here described. Figure 1 illustrates the operation of the basic switch element--a simple R-C filter which uses a non-linear condenser as its series branch. With no direct voltage across the condenser (Fig. 1A), the circuit behaves like any ordinary π-section R-C filter with the exception that distortion will result if the input voltage is large enough to drive the dielectric out of its linear region. Transfer characteristics are shown for sinusoidal excitation. If a bias voltage, V_c, is inserted in the circuit as shown (Fig. 1B), the operating point for the transfer characteristics is shifted to a new point on the charge-versus-voltage characteristics of the non-linear condenser. At this new point, the condenser has a much lower capacity and, therefore, the filter characteristics are changed in such a way that the output diminishes in amplitude. With a fixed amplitude input voltage, then, the output voltage can be changed by varying the bias voltage, V_c. For ferroelectric switch operation, we need but two values for V_c: when $V_c = 0$ the switch is ON and when V_c is equal to some fixed value high enough to bias the dielectric well into its saturation region, the switch is OFF.

Figure 2A illustrates a two-position ferroelectric switch. The two non-linear condensers are made as a single unit by firing a large electrode on one side of a ferroelectric sheet and two smaller electrodes on the opposite side. With S in the position shown, output 2 is biased OFF and output 1 is ON.

---

1. Shepard Roberts, Barium Titanate and Barium-Strontium Titanate as Non-Linear Dielectrics, M. I. T. Sc. D. Thesis, Department of Electrical Engineering (1946).

2. Development and Application of Barium Titanate Ceramics as Non-Linear Circuit Elements, Final Report Contract No. W36-039 sc-44606 File No. 19028-PH-49-5(4060) Glenco Corporation (August 15, 1950).

In the opposite position, output 2 is ON.

Figure 2B illustrates an eight-position ferroelectric switch. Operation of the first stage, controlled by S_1, is the same as the two-position switch. Subsequent stages, however, have the lower ends of their resistors connected so that the even resistors are connected to ground when the odd resistors are connected to V_c and the even resistors are connected to V_c when the odd resistors are connected to ground. There are eight possible paths through the switch (Fig. 3) only one of which will have all of its condensers ON. With S_1, S_2 and S_3 of the eight-position switch set as shown, output zero is ON. Outputs 1, 3 and 7 have one condenser OFF, outputs 2, 4 and 6 have two condensers OFF, and output 5 has all three condensers OFF. The number of OFF condensers among the outputs follows a binomial distribution:

	All On	One OFF	Two OFF	Three OFF	Four OFF	Five OFF
4-position switch	1	2	1			
8-position switch	1	3	3	1		
16-position switch	1	4	6	4	1	
32-position switch	1	5	10	10	5	1

Successful operation of the switch postulates that a single OFF condenser leading to an output will cause that output to be OFF. To test this, an eight-position switch was constructed (Fig. 4) using a thin (.025") sheet of barium tinanate ceramic (Glenco body "X-18"). All of the non-linear condensers are placed on the same sheet by firing electrodes on the two sides as shown. The signal enters the sheet via a large fired electrode (back view). Two electrodes match this input electrode on the opposite side (front view). Among the two condensers thus formed, one will always be OFF and one will always be ON. Each of these two electrodes is enlarged to match up with two electrodes on the opposite side which are alongside the input electrode. One of each pair of this third set of electrodes will be OFF. Finally the signal goes through the dielectric a third time coming out on one of the eight small electrodes (front view).

The operation of the switch is illustrated graphically by Fig. 5. With a constant-amplitude, sine-wave input of variable frequency, the RMS output at terminal 7 was measured as a function of frequency for each of the eight possible combinations of S_1, S_2 and S_3. At 800 cps, the best operating frequency for this particular design, the ratio of ON voltage to the highest OFF voltage is greater than three to one. This operating frequency can be shifted higher or lower by changing the size of the condensers and resistors. Both steady-state and pulse tests on this dielectric indicate that the operating frequency can be shifted up to several megacycles per second. If the resistors are replaced by inductors, the output-versus-frequency characteristics can be improved and losses are lowered.

For pulsed operation of this switch, a non-linear condenser is used in both the series and shunt arms of the filter. Fig. 6A illustrates such a switch which is so arranged that when the series condenser is ON, the shunt condenser is OFF (Fig. 6B); and when the series condenser is OFF, the shunt condenser is ON (Fig. 6C). The filter looks like a condenser voltage-divider

to the rising edge of a pulse. The divider has either a large condenser in its upper leg and a small condenser in its lower leg or vice-versa, depending on whether the switch is ON or OFF.

The ferroelectric switch is proposed as a means for driving the rows and columns of a ferroelectric memory and for switching within an information-handling system. Its unique packaging makes it promising in applications where size, weight and cost are important considerations.

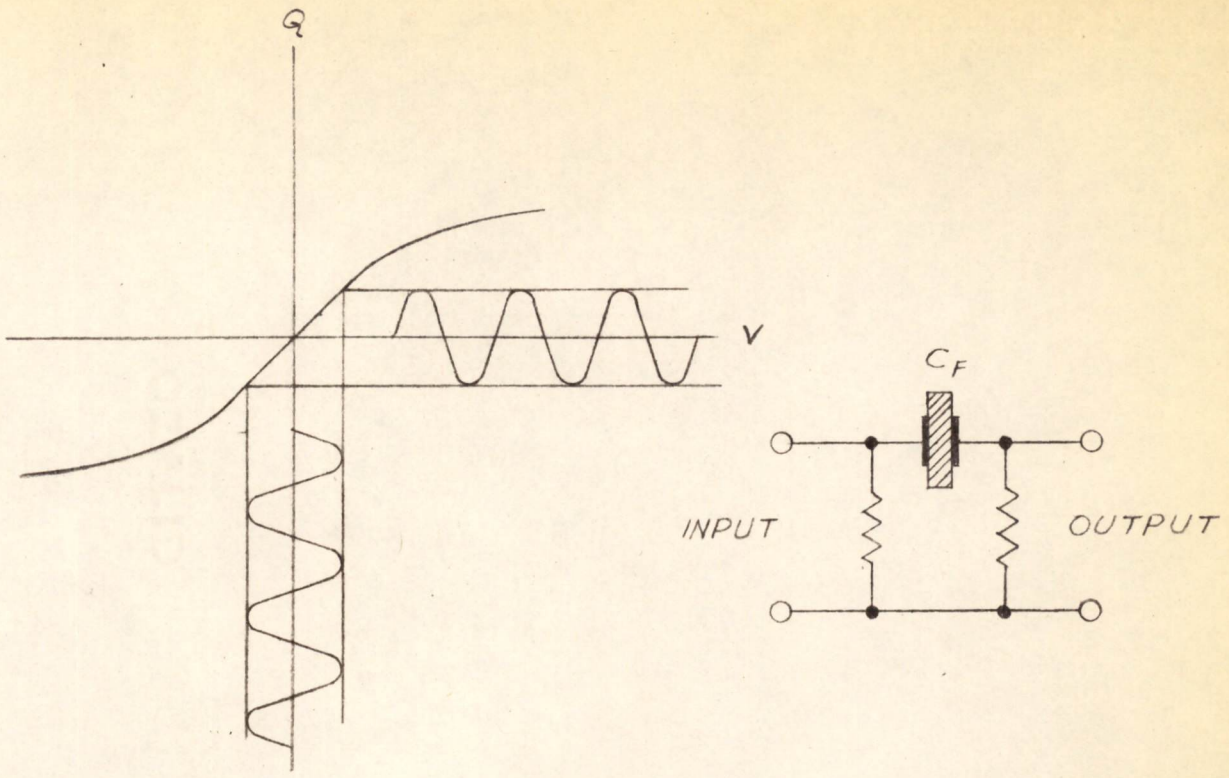
Signed Dudley A. Buck
Dudley A. Buck

Approved William N. Papian
William N. Papian

DAB/jk

Drawings attached:

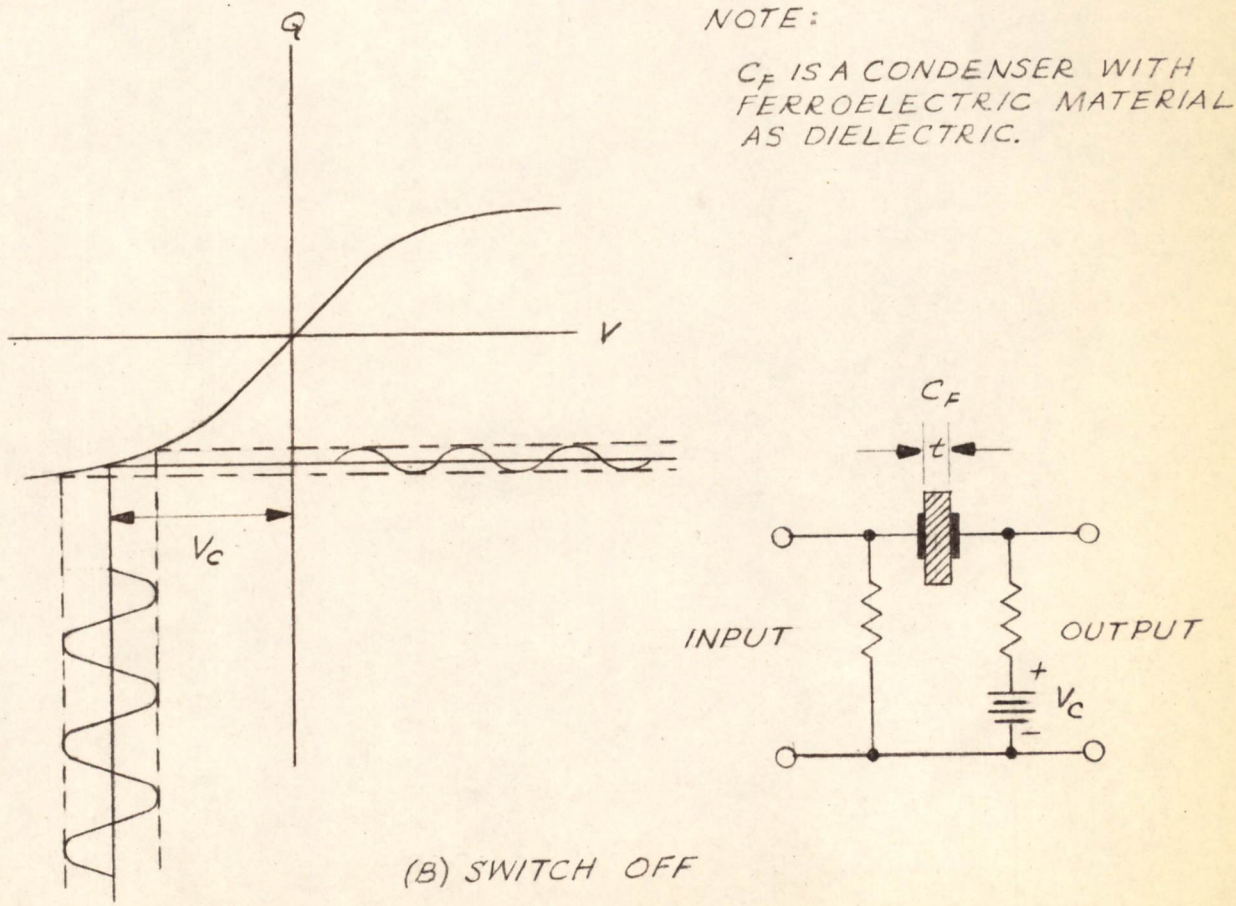
Fig. 1 - A-51155
Fig. 2 - A-51144
Fig. 3 - A-51151
Fig. 4 - A-50906
Fig. 5 - A-51148
Fig. 6 - A-51152



(A) SWITCH ON

NOTE:

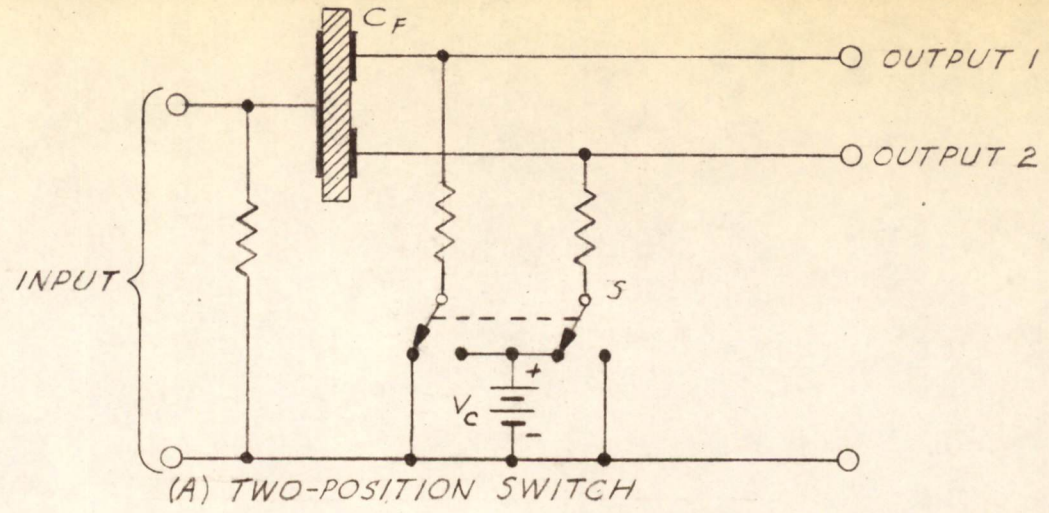
C_F IS A CONDENSER WITH FERROELECTRIC MATERIAL AS DIELECTRIC.



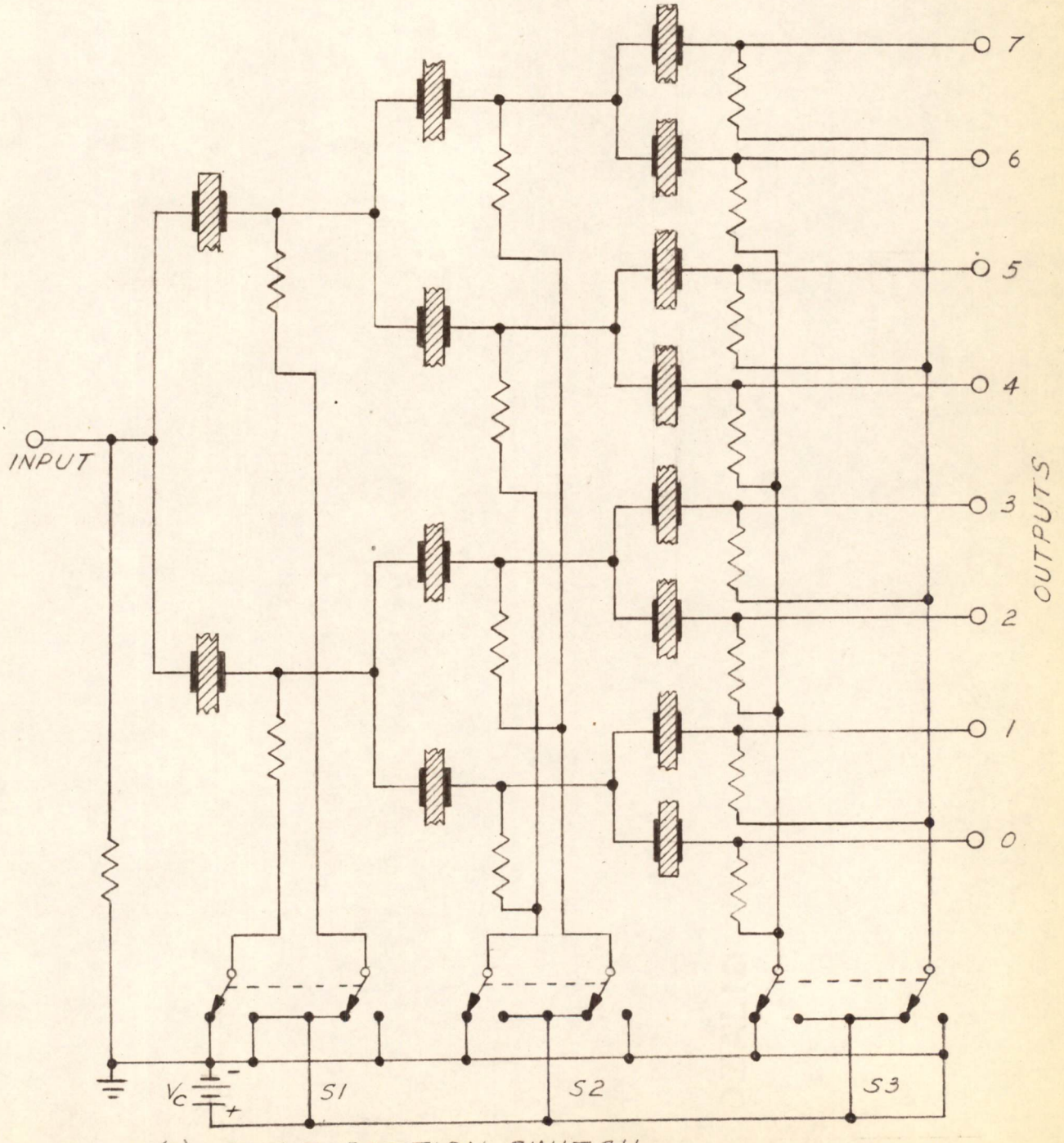
(B) SWITCH OFF

FERROELECTRIC SWITCH OPERATION

A-51155



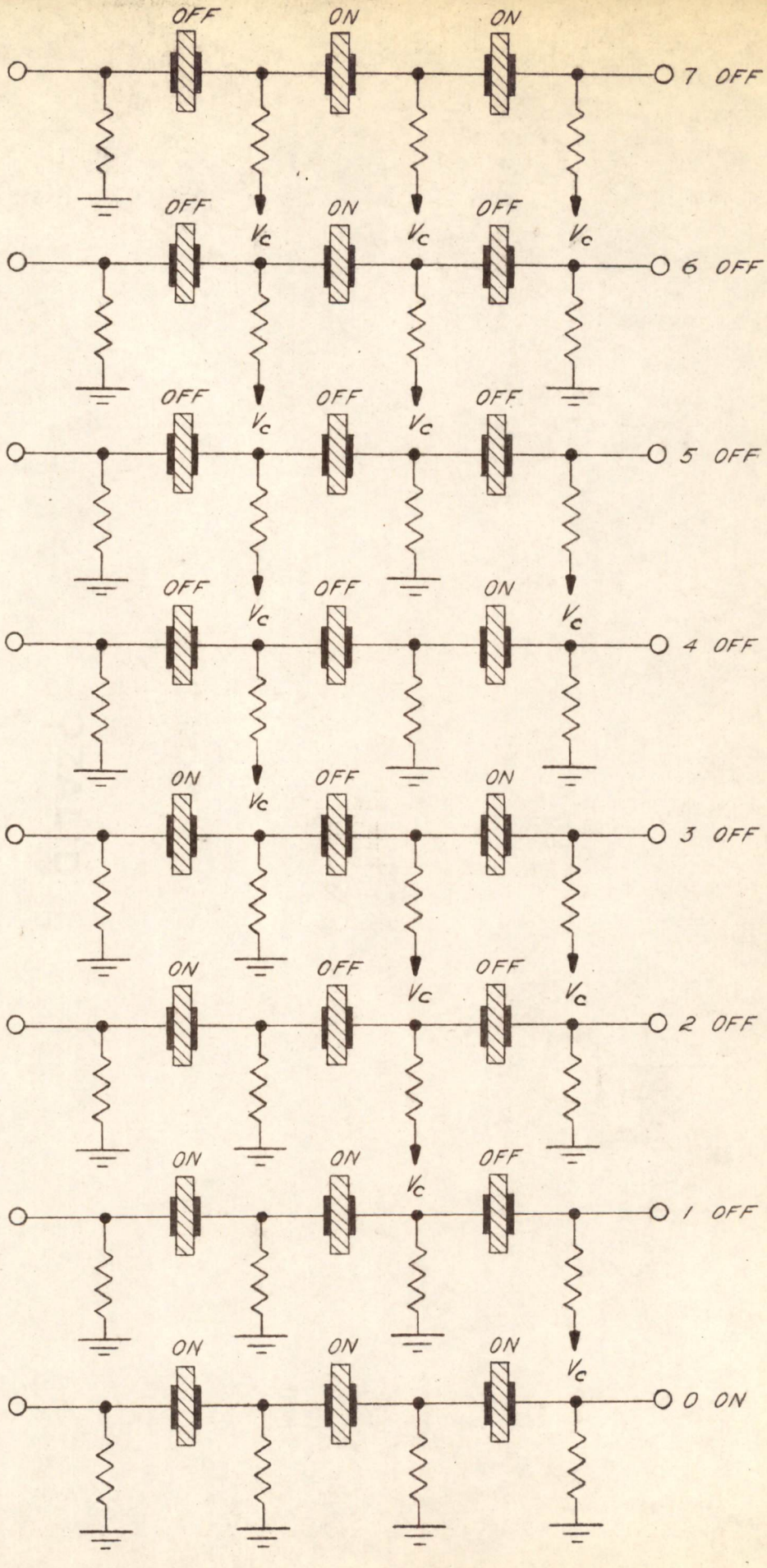
(A) TWO-POSITION SWITCH



(B) EIGHT-POSITION SWITCH

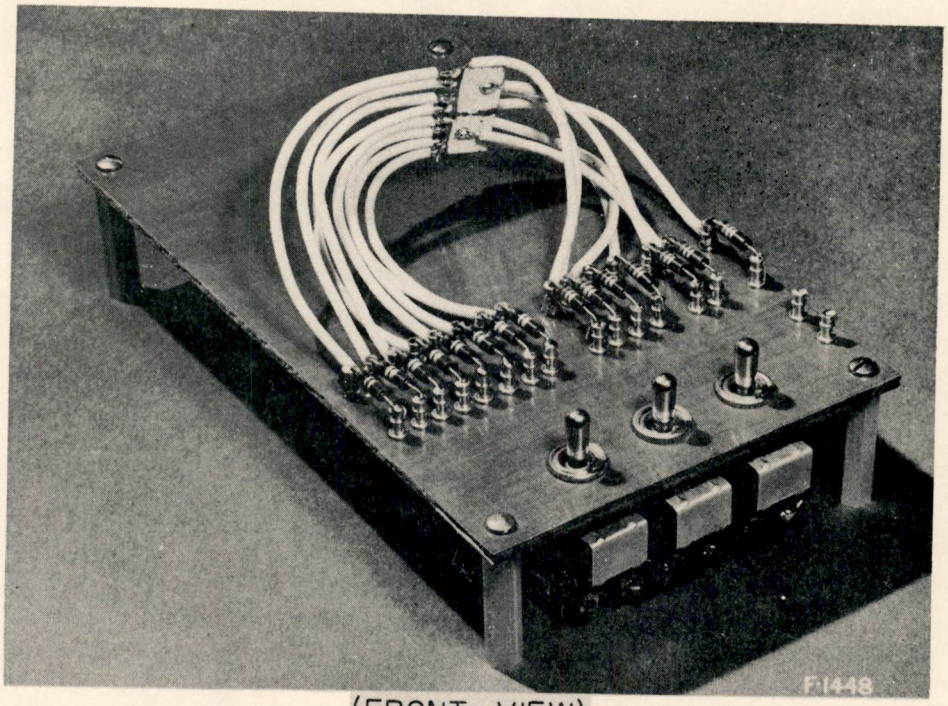
FERROELECTRIC SWITCH SCHEMATIC DIAGRAM

A-51144

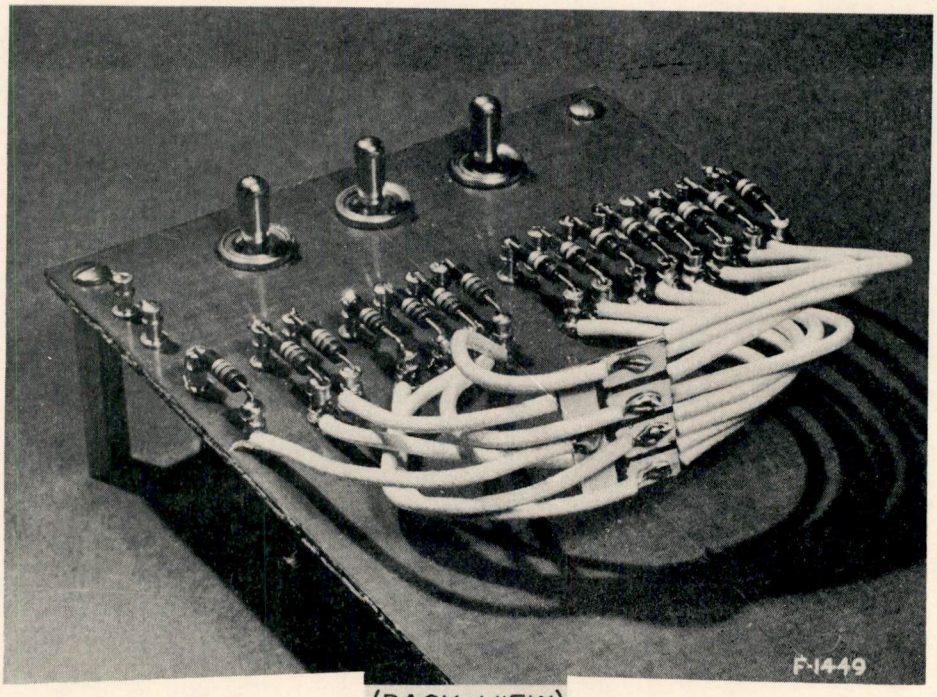


EIGHT-POSITION SWITCH ANALYSIS

A-51151



(FRONT VIEW)

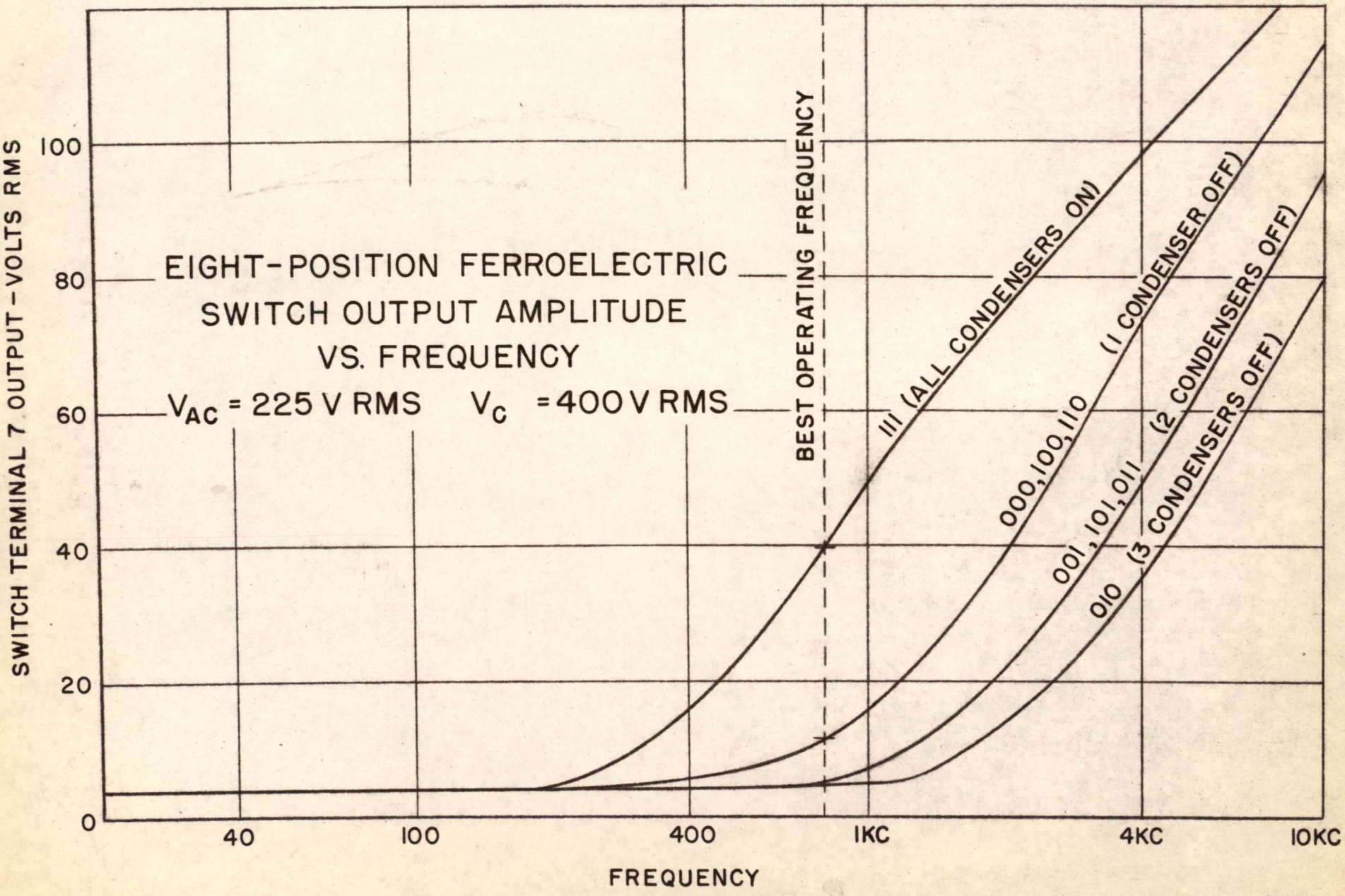


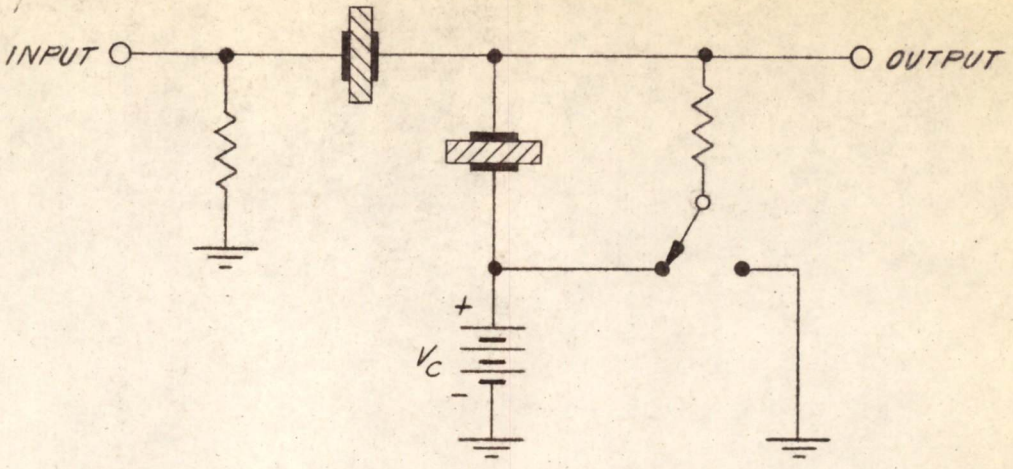
(BACK VIEW)

FERROELECTRIC MULTI—POSITION SWITCH

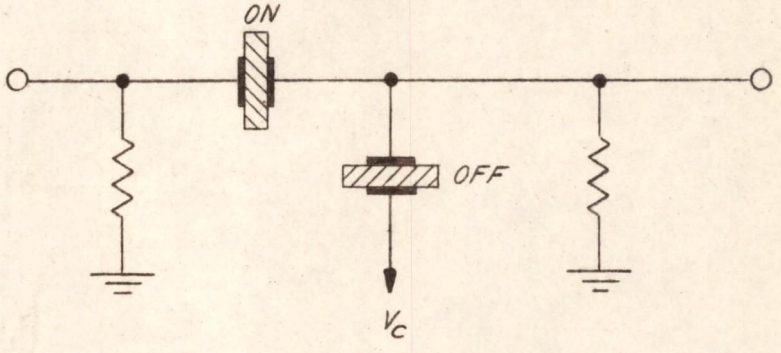
A-50906

A-51148

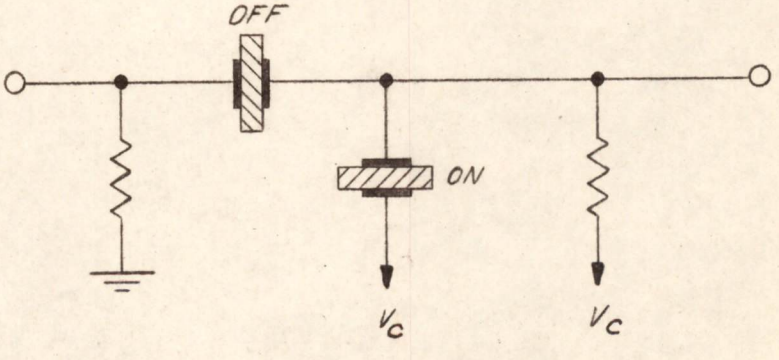

EIGHT-POSITION FERROELECTRIC
SWITCH OUTPUT AMPLITUDE
VS. FREQUENCY
V_{AC} = 225 V RMS V_C = 400 V RMS
SWITCH TERMINAL 7 OUTPUT - VOLTS RMS
0
20
40
60
80
100
40
100
400
1KC
4KC
10KC
FREQUENCY
BEST OPERATING FREQUENCY
111 (ALL CONDENSERS ON)
000, 100, 110 (1 CONDENSER OFF)
001, 101, 011 (2 CONDENSERS OFF)
010 (3 CONDENSERS OFF)



A. CIRCUIT SCHEMATIC OF PULSE-OPERATED TWO-POSITION FERROELECTRIC SWITCH



B. SWITCH ON



C. SWITCH OFF

FERROELECTRIC SWITCH FOR PULSES

A-51152

Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

SUBJECT: DESIGN OF A DIGITAL COMPUTER BY BOOLEAN ALGEBRA

To: N. H. Taylor

From: R. C. Jeffrey, I. S. Reed

Date: May 20, 1952

Abstract: The techniques described in E-458-1 are illustrated in a real-life situation: the design of a 4-order machine employing an unusual method of central control. For definiteness the memory size is taken to be 8 words, but no logical complexity is added when this is increased to a realistic figure.

1.0 SPECIFICATIONS FOR THE MACHINE: NOMENCLATURE

1.1 Words: 5 bits, interpreted as in WWI, with negative numbers represented in nine complement form.

Interpreted as a number: $\overbrace{Sg(X)}$

Interpreted as an instruction:

X_0	X_1	X_2	X_3	X_4
$\overbrace{Op\ (X)}$		$\overbrace{Ad(X)}$		

1.2 Memory: 8 words, stored in flip-flops. An actual memory would use some other device, such as iron cores; but the analysis for the flip-flop case can easily be applied to whatever device is actually used.

1.3 Registers

Arithmetic Element

A-Register

Receives number from storage

A_0	A_1	A_2	A_3	A_4
$\overbrace{Op\ (A)}$		$\overbrace{Ad(A)}$		

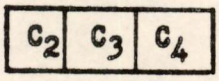
B-Register

Accumulator (A subtracter)

B_0	B_1	B_2	B_3	B_4
$\overbrace{Sg(B)}$				

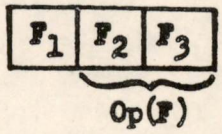
C-Register

Holds address of next memory register to be used.



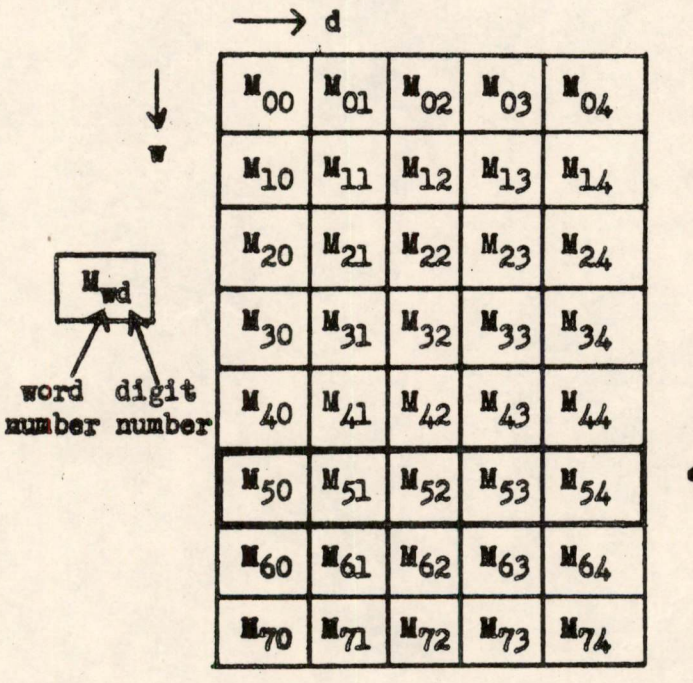
F-Register

Operation counter; central control.

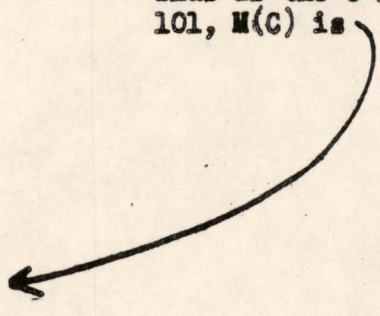


Start Flip-Flop

Memory



"M(C)" denotes the address in memory corresponding to the number in the C-register. Thus if the C-register holds 101, M(C) is



1.4 Operations

The decisions as to word length, number of registers, etc., summarized above come under the heading of machine planning, and were arrived at by cut-and-try. The last and most delicate part of the planning concerns the operation of central control.

Code No.	Name	Description
00xyz	Halt	The machine stops, i.e., all flip-flops remain in the states they were in when the operation was executed, until the start button is depressed, at which time program timing begins.
01xyz	Conditional Subprogram	If the sign digit of the B-register is 1, take next instruction from memory location xyz. If the sign of the B-register is 0, take next instruction in sequence.
10xyz	Subtract	Take number from memory location xyz and put it in the A-register; subtract it from the contents of the B-register and leave the result in B.
11xyz	Shift right	Shift the contents of the B-register right, depositing the original contents of B in M(C).

2.0 TIMING DIAGRAM

2.1 Notation

The above four operations are to be performed as sequences of commands to perform the elementary independent functions of which the machine is capable. In order to describe these functions we adopt a compact terminology:

M(C): The memory location whose number is stored in C.

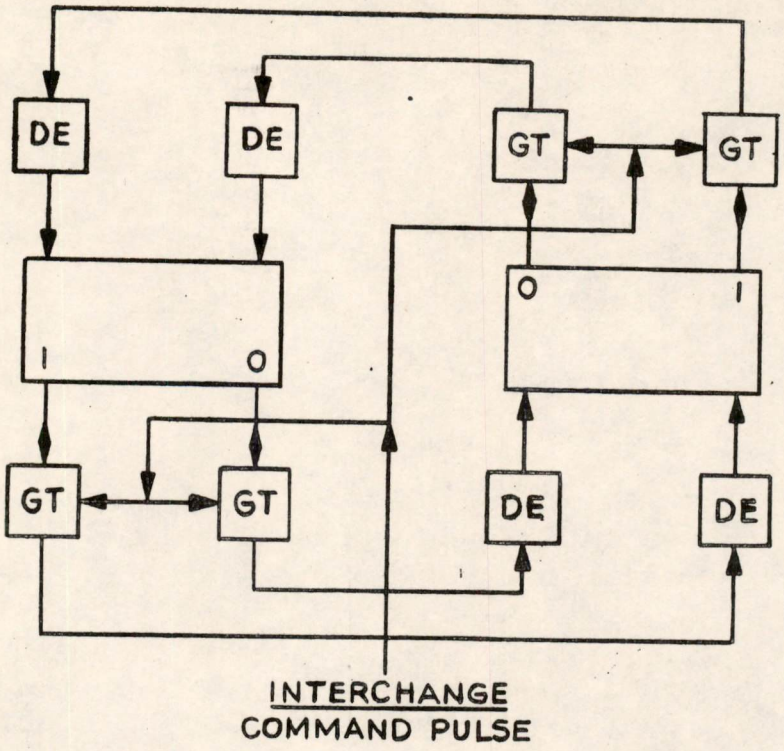
X: The X register.

(X): The contents of the X register.

(X) $\Rightarrow$ Y: The contents of register X at time t appear in register Y at time t + ϵ.

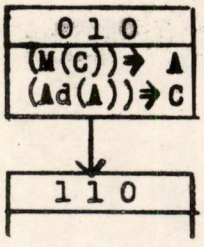
Then it is possible to have "simultaneous" interchanges:

"(X) $\Rightarrow$ Y and (Y) $\Rightarrow$ X" means that the contents of X at t appear in Y at t + ϵ, and that the contents of Y at t appear in X at t + ϵ. This can be implemented by delay elements (which we shall henceforth assume to be built into the FF's):



We now draw a flow diagram for central control (next page). The numbers at the tops of the boxes will represent states of the F-counter, and the notations inside the boxes represent the commands which are performed by the first time pulse during which the F-counter is in the indicated state.

For example, in the upper left corner on page 5

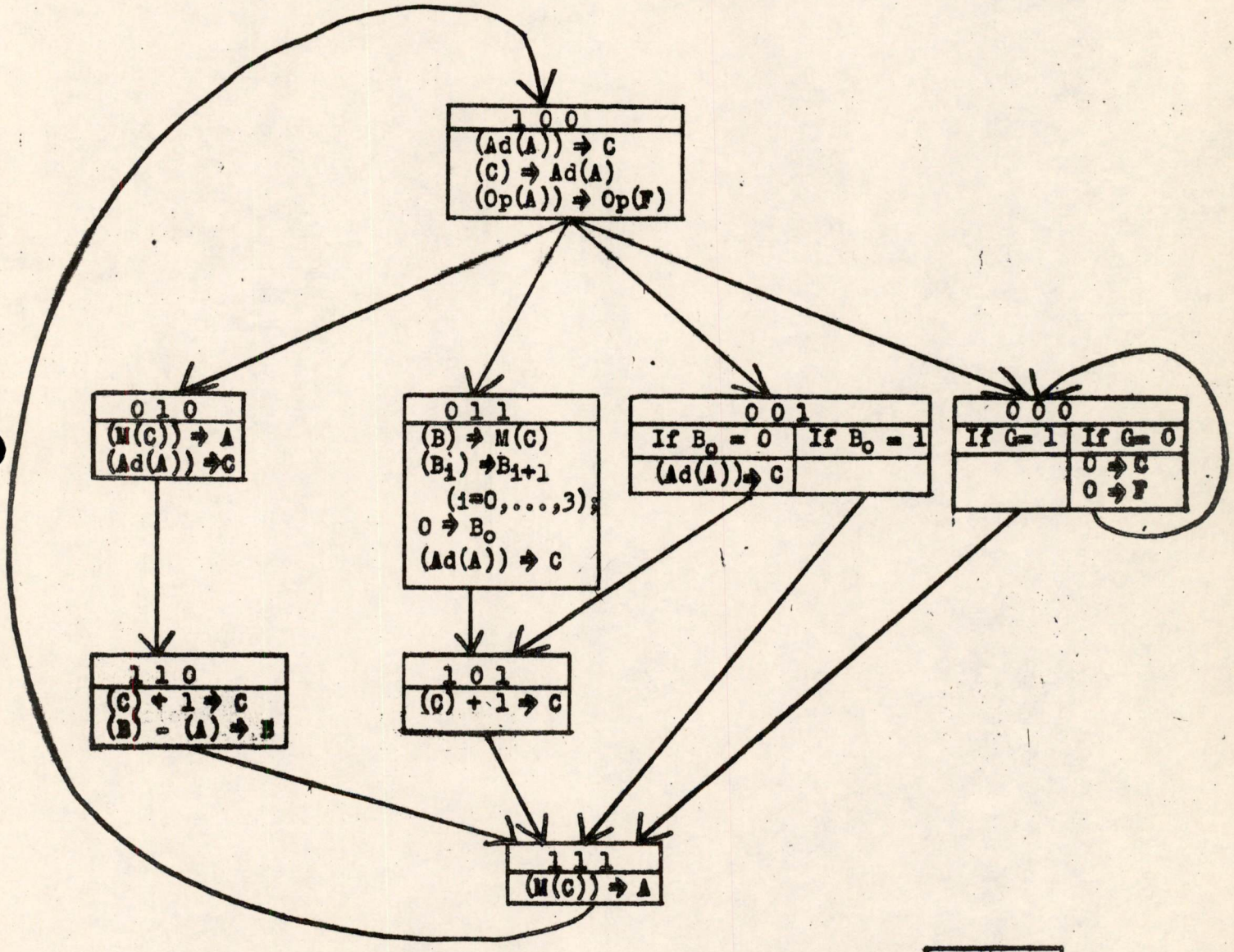


means that when the F-register is in the configuration $F_1 = 0$, $F_2 = 1$, $F_3 = 0$ the commands read out of memory location (C) into the A-register and read the last three (address) digits of the A-register into the C-register are performed and the F-counter changes to the configuration 110 (i.e., F_1 is complemented). This change in (F) is strictly speaking a command, but is indicated by the arrow between the boxes instead of by a third notation inside the upper box.

For further explanation of the flow diagram, see page 6.

2.2 Flow Diagram

(The numbering of the boxes should be thought of as arbitrary, for the present. See Section 4.0 below.)



Note: If G = 0, then regardless of the state of F, both the F and C registers are cleared:

If G = 0
0 ⇒ C
0 ⇒ F

Program timing occurs at 111 and 100. Note that the C-register is used both as a storage selection register and a program counter: in 110 and 101 we add 1 to (C) . Each of the four paths after 100 corresponds to an operation:

010 begins the operation timing for subtract (10xyz)
011 begins the operation timing for shift (11xyz)
001 begins the operation timing for cp (01xyz)
000 begins the operation timing for halt (00xyz)

The four way decision is determined by the result of the command $(Op(A)) \rightarrow Op(F)$, which reads (A_0) into F_2 and (A_1) into F_3. At the beginning of operation timing for operation number v w the F-register will be in configuration 0 v w, since previously we had (A_0) = v and (A_1) = w for Op(A). It is also necessary to arrange that F_1 be complemented when the F-counter leaves configuration 100.

The reader should now verify that the sequences of commands specified in the flow diagram really do add up to the four operations listed above. The logical details are discussed in section 3.

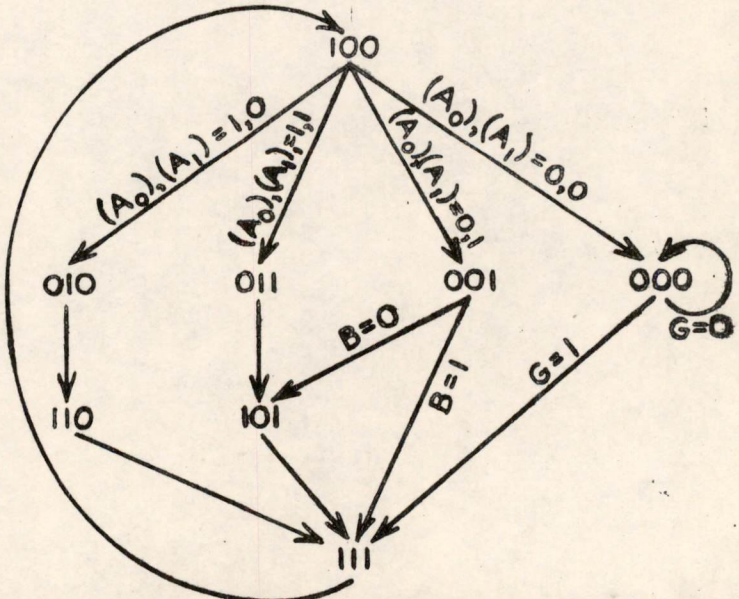
3.0 DESIGN OF THE MACHINE: INPUT EQUATIONS

3.1 Design of the F-Register

Apparently the F-register is to be a 3 stage binary counter with the following cycle:

We name the configurations of the F-register:

$P_0 = F_1' F_2' F_3'$ --- 000
$P_1 = F_1' F_2' F_3$ --- 001
$P_2 = F_1' F_2 F_3'$ --- 010
$P_3 = F_1' F_2 F_3$ --- 011
$P_4 = F_1 F_2' F_3'$ --- 100
$P_5 = F_1 F_2' F_3$ --- 101
$P_6 = F_1 F_2 F_3'$ --- 110
$P_7 = F_1 F_2 F_3$ --- 111

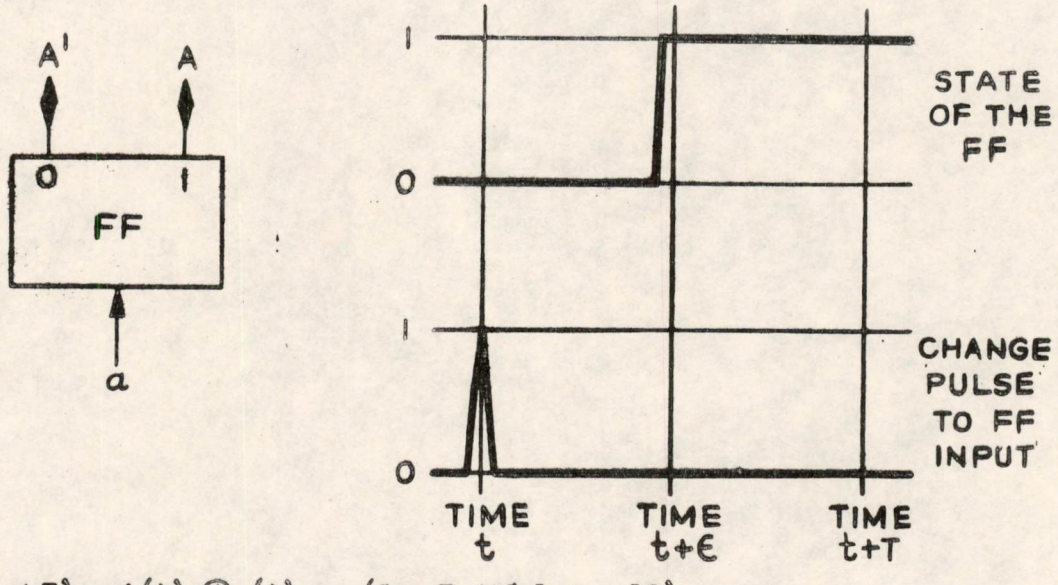


The block diagram on page 9 shows how voltages corresponding to the P_i are generated.

In the table above, the binary representation of the subscript of P indicates, by the distribution of zeros, the location of the primes in the product of the F's.

Now the successive states of the counter are determined by information external to the counter, as well as from the configuration of the counter itself. For example, from P_3 the counter must go to P_5, but from P_1 it may go to P_5 or to P_7. It is necessary to indicate the basis on which this selection is made; this is done in the cases where a choice is necessary.

Assume that the FF's used have one input, such that if that input is pulsed at time t, the FF is complemented at time $t+\epsilon$, ϵ being a delay less than T, the period of the clock.



$A(t+T) = A(t) \oplus a(t)$ (See E-458-1, p. 13)

Now to design the F-counter it is sufficient to write three equations, one for each stage, specifying when the change inputs are to be pulsed. We shall work through the lattice on page 6 (or page 5) level by level for each input. The equations we are about to build up are illustrated by a block diagram on page 9.

For the first digit (F_1) we see that in going from P_4 (=100) on the first level to the second, F_1 is to be complemented regardless of which of the four configurations on the second level is selected:

$$f_1 = (P_4 + \ldots\ldots\ldots)E$$

Here E denotes the clock pulse: the changes are to occur at each clock pulse, so that an operation will require at most 4 clock pulses for its completion.

To get the second term: if P_2 (=010) is the second level configuration selected, the first digit must change again to get the configuration 110:

$$f_1 = (P_4 + P_2 + \ldots\ldots)E$$

Ditto in case 011 was the second level configuration selected: to get from 011 to 101, F_1 must be complemented:

$$f_1 = (P_4 + P_2 + P_3 + \ldots\ldots)E$$

Similarly, whichever value B_0 may have, the F-counter will go from the configuration 001 to a configuration (101 or 111) in which the first digit is changed:

$$f_1 = (P_4 + P_2 + P_3 + P_1 + \ldots\ldots)E$$

Finally, the counter goes from 000 to 111 in case $G = 1$. Thus F_1 will change in case P_0G. ("P_0G" means $P_0 = 1$ and $G = 1$.)

$$\boxed{f_1 = (P_4 + P_2 + P_3 + P_1 + P_0G)E}$$

Since F_1 does not change in going from 111 to 100, the above formula represents all the conditions under which F_1 must be complemented.

Similarly, for F_2 we have

$$\boxed{f_2 = (A_0P_4 + P_3 + B_0P_1 + GP_0 + P_5 + P_7)E}$$

where the first term derives from the fact that in the two cases where $A_0 = 1$ $\boxed{(O_p(A)) = 10 \text{ and } (O_p(A)) = 11}$ the second digit of "100" must be changed, but in the other two cases (when "100" goes to "001" or to "000") this is not necessary. Finally

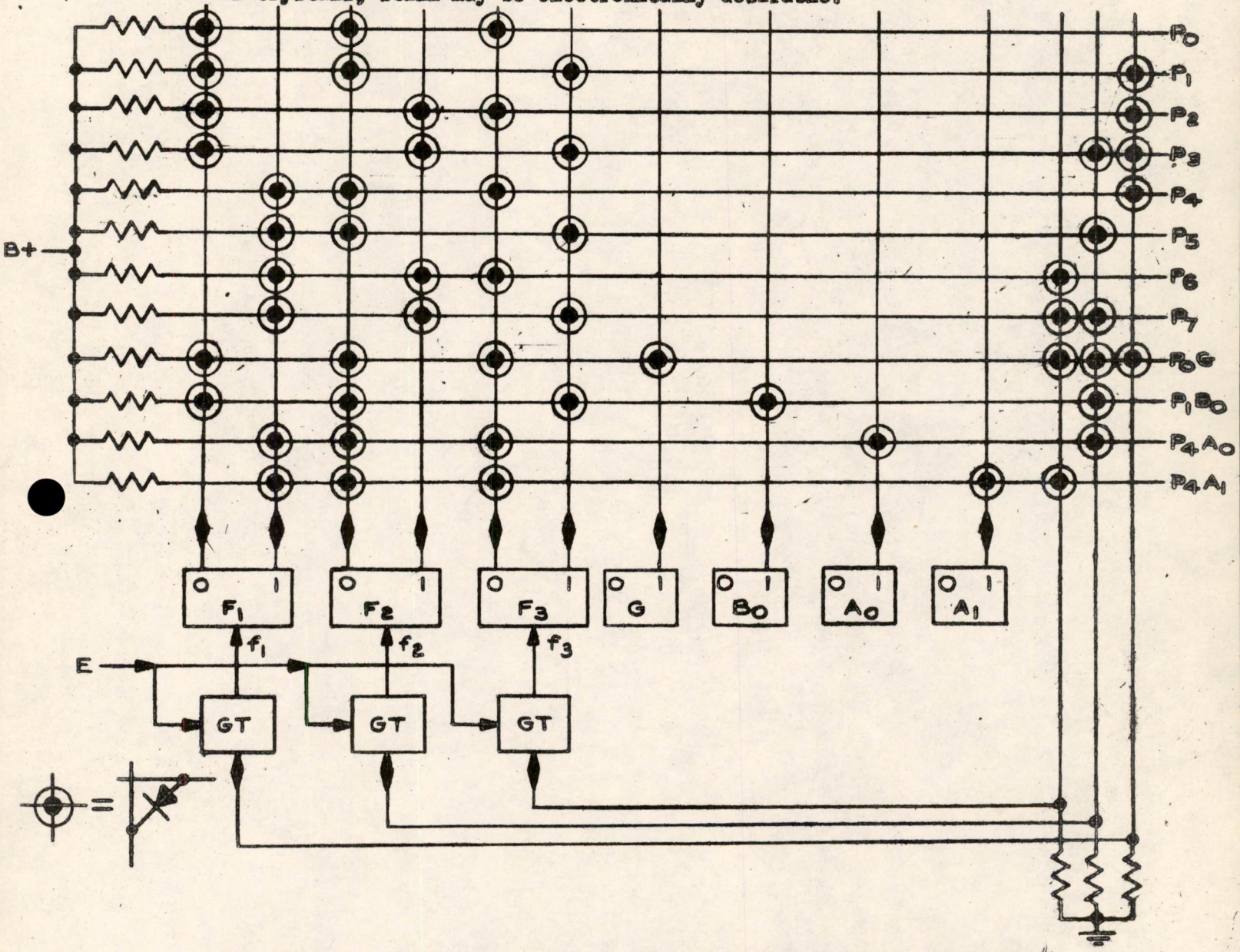
$$\boxed{f_3 = (A_1P_4 + GP_0 + P_6 + P_7)E}$$

It remains to implement the condition that if $G = 0$, $0 = F$. Note that if a halt is programmed, the F-counter steps from 100 to 000 and in so doing clears G. But as will be seen in Section 3.35 below, G can also be cleared by a pushbutton. We wish to provide that if $G = 0$ from any cause, the F-counter will assume the configuration 000 on the next time pulse.

This is most easily accomplished by writing three equations for the clear inputs to the F flip-flops:

$$\boxed{{}_0f_i = G'E \qquad i = 1,2,3}$$

Now these equations may be realized immediately by a two-level diode matrix which, although it does not represent a minimization of number of crystals, still may be electronically desirable.

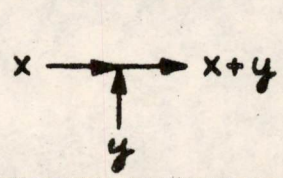
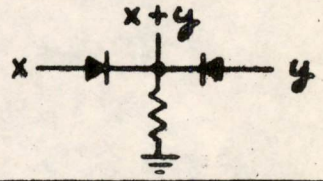
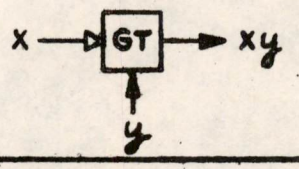
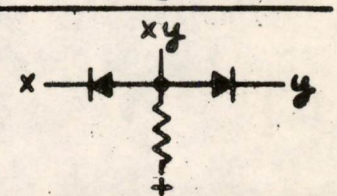
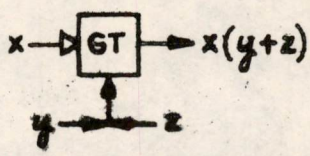
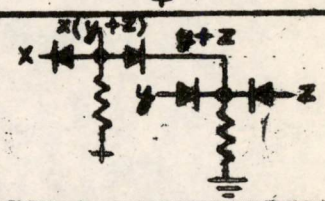
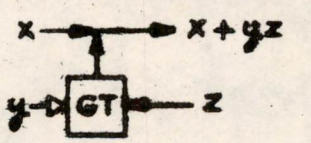
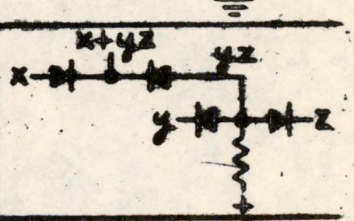
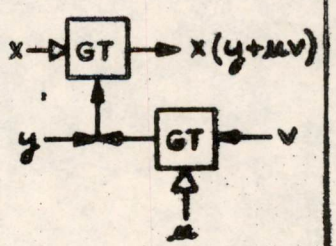
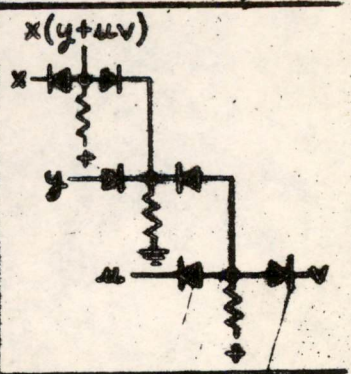


Here we have indicated only those 3 outputs of the diode matrix which feed back into the F-register. Actually we require other outputs: command pulses to the rest of the machine. They will be considered below in connection with the FF registers which they control.

3.2 Digression on "Levels"

Note that the three equations for the F-register can be factored, resulting in a reduction in the number of gates and mixers required for the realization of central control. But this numerical reduction is at the expense of an increase in the number of levels of gating and mixing. We illustrate the meaning of "levels" by examples.

EXAMPLE

No. of Levels	Equations	Block Diagrams	One Electronic Realization (Diode)
1	$x + y$		
	xy		
2	$x(y + z)$		
	$x + yz$		
3	$x[y + uv]$		

Apparently in terms of the equations, the number of levels is the same as the functional complexity of the expression for the output. That is, $x + y$ and xy are functions whose arguments are variables; $x(y + z)$ and $x + (yz)$ are functions (the first a product and the second a sum) in which one of the arguments is itself a function of variables.

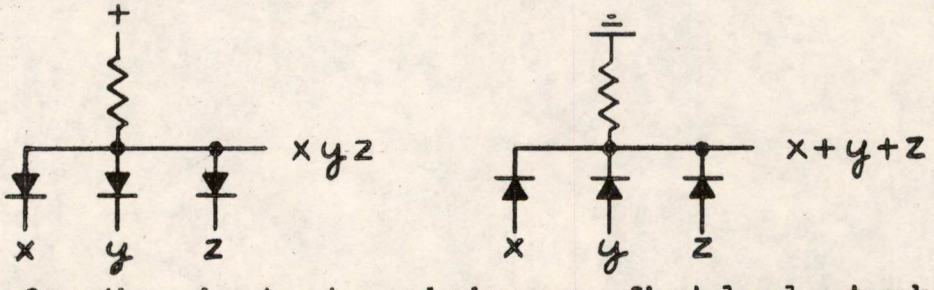
Finally, $x[y + uv]$ is a product one of whose factors is a sum one of whose terms is a product: "$\pi\{x, \sigma[y, \pi(u,v)]\}$" in functional notation where $\pi(a,b) = ab$ and $\sigma(a,b) = a + b$. Briefly, the number of levels in a Boolean function is identical with the number of nestings of parentheses within each other.

In terms of block diagrams, the number of levels in a device is determined by drawing lines from each input to the output; the largest number of gates and mixers through which such a line passes in the level of the system. Finally, the reason we are interested in levels is that if a function of level n is realized with diodes, n is the largest number of diodes through which any input current must pass in order to reach the output. In a high-level diode network considerable attention must be paid to the values of voltage and resistance associated with the diodes. Other difficulties arise with high-level vacuum tube and transistor networks. In all cases the difficulties simply require engineering attention: high-level networks are realizable.

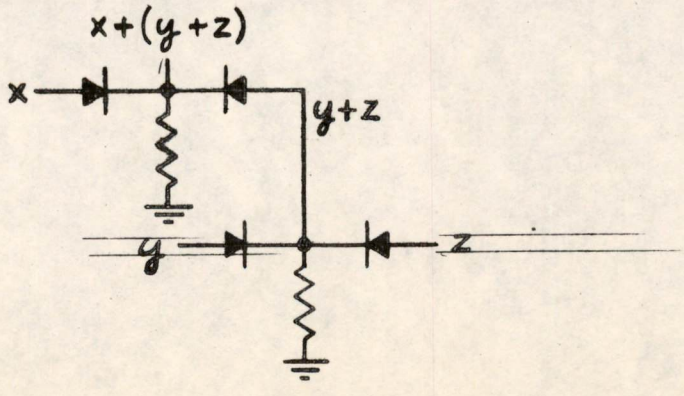
One last comment: in such expressions as "xyz" and "x + y + z", the product and sum are regarded as functions of three variables:

$$\pi_3(x,y,z) = xyz \text{ and } \sigma_3(x,y,z) = x + y + z$$

Electronically this corresponds to the use of three-input diode devices for mixing and gating:



and therefore three-input gates and mixers are first level networks. However, expressions like "$x + (y + z)$", which might be written in functional notation as "$\sigma_2(x, \sigma_2(y,z))$ have as their direct realizations two-level mixers like:



in which the longest current path passes through two diodes. The two sorts of mixers are logically equivalent, and the one-level type is used where possible.

To illustrate these points, let us factor the equations of the F-register:

$$f_1 = P_1 + P_2 + P_3 + P_4 + P_0G \quad \text{(Ignore the "E" for the present.)}$$

$$= \underline{F_1'}F_2'F_3 + \underline{F_1'}F_2F_3' + \underline{F_1'}F_2F_3 + F_1F_2'F_3' + \underline{F_1'}F_2'F_3'G$$

$$= F_1'(F_2'\underline{F_3} + F_2F_3' + F_2\underline{F_3} + F_2'F_3'G) + F_1F_2'F_3'$$

$$= F_1'\left[F_3(F_2' + F_2) + F_3'(F_2 + F_2'G)\right] + F_1F_2'F_3'$$

$$= F_1'\left[F_3 + (F_2 + G)\right] + F_1F_2'F_3' \quad \text{(by 2 applications of the rule: } x + x'y = x + y\text{)}$$

$$\boxed{f_1 = \left[F_1'(F_2 + F_3 + G) + F_1F_2'F_3'\right]E}$$

$$f_2 = A_0P_4 + P_3 + B_0P_1 + GP_0 + P_5 + P_7$$

$$= A_0\underline{F_1}F_2'F_3' + F_1'F_2F_3 + B_0F_1'F_2'F_3 + GF_1'F_2'F_3' + \underline{F_1}F_2'F_3 + \underline{F_1}F_2F_3$$

$$= F_1(A_0F_2'F_3' + F_2'\underline{F_3} + F_2\underline{F_3}) + F_1'(F_2F_3 + B_0F_2'F_3 + GF_2'F_3')$$

$$= F_1(A_0F_2'F_3' + F_3) + F_1'\left[F_3(F_2 + B_0F_2') + GF_2'F_3'\right]$$

$$= F_1(F_3 + A_0F_2') + F_1'\left[F_3(F_2 + B_0) + GF_2'F_3'\right]$$

$$\boxed{f_2 = \left\{F_1(F_3 + A_0F_2') + F_1'\left[F_3(F_2 + B_0) + GF_2'F_3'\right]\right\}E}$$

$$f_3 = A_1P_4 + GP_0 + P_6 + P_7$$

$$= A_1\underline{F_1}F_2'F_3' + GF_1'F_2'F_3' + \underline{F_1}F_2F_3' + \underline{F_1}F_2F_3$$

$$= F_1(A_1F_2'F_3' + \underline{F_2}F_3' + \underline{F_2}F_3) + GF_1'F_2'F_3'$$

$$= F_1(A_1\underline{F_2'}F_3' + \underline{F_2}) + GF_1'F_2'F_3'$$

$$\boxed{f_3 = \left[F_1(F_2 + A_1F_3') + GF_1'F_2'F_3'\right]E}$$

We have now reduced the total number of diodes (assuming a realization as a diode net) from 55 in the two-level matrix to 41 [10 for f_1, 19 for f_2, 12 for f_3: this can be determined directly from the equation by counting n diodes for each n input gate or mixer. E.g., for f_3..... ignoring E..... we have a 2 input mixer (the plus in front of "G") of which the inputs are a 4 input gate and a 2 input gate of which the inputs are F_1 and a 2 input mixer of which the inputs are F_2 and a 2 input gate. Adding up the numbers of inputs underlined above gives the 12 diodes required for f_3.]

We shall not draw the block diagrams for the factored equations. It would be very messy, and an engineer who understands the algebraic notation could design the network from the equations as easily as he could from the block diagrams (the equations have the added advantage of compactness: three rows of symbols, instead of a page crammed with boxes, lines and arrows).

3.3 The Arithmetic Element

3.31 The Subtractor (See diagram on p. 26)

The most complicated single command the machine must execute is subtract [(B) - (A) ⇒ B: from the contents of the B-register, subtract the contents of the A-register and deposit the result back in B.]

Now representing the binary number stored in the A-register by 'A', and similarly for B, and denoting the nines complement (result of complementing each digit) of A by '$\bar{A}$', we have

$$B = B_0 2^0 + B_1 2^{-1} + B_2 2^{-2} + B_3 2^{-3} + B_4 2^{-4}$$

$$A = A_0 2^0 + \ldots\ldots + A_4 2^{-4}$$

$$\bar{A} = (1 - A_0) 2^0 + \ldots\ldots + (1 - A_4) 2^{-4}$$

$$A + \bar{A} = 2^0 + \ldots\ldots + 2^{-4} = 2 - 2^{-4}$$

$$-A = \bar{A} + 2^{-4} - 2$$

$$\boxed{B - A = B + \bar{A} + 2^{-4} - 2}$$

We can now construct a table which describes the operation of the i^{th} stage of our subtractor; it is the ordinary table for binary addition where the second term is A_i' instead of its complement, A_i. Note that we require an "instantaneous" carry output, K_{i-1}.

$B_i(t)$	$A_i'(t)$	$K_i(t)$	$K_{i-1}(t)$	$B_i(t+T)$
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1
all possible inputs			$B + \bar{A}$	

We want in the B-register at time t + T not $B + \bar{A}$, but rather $B + \bar{A} + 2^{-4} - 2$, which we have shown above to be equal to B - A. (cf. end-around carry)

To get from $B + \bar{A}$ to $B + \bar{A} + 2^{-4} - 2$: we add the 2^{-4} by requiring that K_4, the carry input to the first stage, be 1.

We subtract the 2 by failing to provide a B_{-1} flip-flop.

Then a realization of the table which has the two additional properties just listed will subtract the contents of A from the contents of B and leave the result in the B-register T seconds after the subtract command pulse, P_6E (note that according to the table on p. 5 the (B) - (A) $\Rightarrow$ B command goes out when the F-counter is in the configuration 110 (P_6) and a clock pulse (E) occurs).

The input to B_i is to be pulsed whenever we wish $B_i(t + T)$ to be the complement of $B_i(t)$ and a subtract command pulse occurs. Comparing the first and last columns of the table we get

$$b = (B_i'A_iK_i + B_i'A_i'K_i' + B_iA_iK_i + B_iA_i'K_i')P_6E$$

If this seems wrong, notice that the table lists values of A', not values of A, and that $A_i' = 0$ means: $A_i = 1$.

This expression can be greatly simplified by factoring A_iK_i from the first and third terms, and $A_i'K_i'$ from the other two:

$$\boxed{b = (A_iK_i + A_i'K_i')P_6E = (A_i' \oplus K_i)P_6E} \qquad (i = 0, \ldots, 4)$$

which fact might have been seen directly from the table.

The carry output from this stage is

$$K_{i-1} = (B_i' A_i' K_i + B_i A_i K_i + B_i A_i' K_i' + B_i A_i' K_i) P_6 E$$

$$K_{i-1} = \left[A_i' K_i + B_i (A_i' \oplus K) \right] P_6 E \qquad i = 1,2,3,4;\ K_4 = 1$$

$$= A_i' K_i P_6 E + B_i b$$

3.32 Shifting

The B-register is used in shifting as well as subtracting. Then before we adopt the above equation for the input to the B-register we must add terms which take account of the additional inputs required for shifting.

Referring to the diagram on p. 5 we see that the function (B) - (A) $\Rightarrow$B, which occurs in configuration 110 of the F-counter, is not the only one that involves B.

In configuration 011 (P_3) of the F-counter two other functions are performed: (B) $\Rightarrow$M[C] and (B_i) $\Rightarrow$ $B_{i\ 1}$, it being understood that the number which appears in M[C] is (B) before the shift. (See remarks on "simultaneous" interchanges between registers in Section 2.1.) In the shift a zero is introduced in the left-most digit position, B_0.

The transfer to memory, (B) $\Rightarrow$ M[C], does not concern the input to the B-register: it will be taken account of in the input equations to memory. Then the input equations for b will be complete when we add expressions to realize (B_i) $\Rightarrow$ B_{i+1} ($i = 0,1,2,3$); $0 \Rightarrow B_0$. This is most easily done using the clear and set inputs to the B flip-flops, although it might have been accomplished with only the complement input, with suitable gating. Then we consider that we have the WWI type of FF with three inputs, labeled as shown; the complete equations for the B-register are then

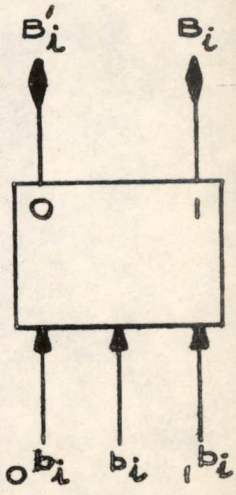
$$b_i = (A_i' \oplus K_i) P_6 E \qquad (i = 0,1,2,3,4) \quad \text{(difference digit)}$$

$$K_{i-1} = A_i' K_i P_6 E + B_i b \qquad (i = 1,2,3,4) \quad \text{(carry)}$$

$$K_4 = 1 \quad \text{("end-around carry")}$$

$$\left.\begin{aligned} {}_1b_i &= B_{i-1} P_3 E \\ {}_0b_i &= B_{i-1}' P_3 E \end{aligned}\right\} \qquad (i = 1,2,3,4) \quad \text{(shift right)}$$

$${}_0b_0 = P_3 E \qquad \text{(destructive shift)}$$



That is, when the F-counter is in configuration P_3, the clock pulse E acts as a command for the contents of register no. i-1 to appear in register no. i, and for a zero to appear in register B_0. It is assumed, once again, that there is a delay inherent in the inputs to our FF's such that this shifting will occur <u>after</u> the transfer to memory has taken place.

3.33 <u>The A-Register</u> (See diagram, p. 25)

To complete the discussion of the arithmetic element we write the equations of the inputs to the A-register. Transfers into A occur on configurations 100, 010, and 111 (P_4, P_2 and P_7) of the operation counter (see p. 5). Again it will be simplest to use the <u>set</u> and <u>clear</u> inputs rather than the <u>complement</u> input:

$$\left.\begin{aligned} 1^{a}_{i} &= \left[C_i P_4 + M_{hi}(P_2 + P_7)\right] E \\ 0^{a}_{i} &= \left[C'_i P_4 + M'_{hi}(P_2 + P_7)\right] E \end{aligned}\right\} \quad i = 2,3,4$$

$$\left.\begin{aligned} 1^{a}_{i} &= M_{hi}(P_2 + P_7) E \\ 0^{a}_{i} &= M'_{hi}(P_2 + P_7) E \end{aligned}\right\} \quad i = 0,1$$

$$h = (C)$$

It is necessary to treat i = 0,1 separately since there are no C_0 or C_1 flip-flops. Now we must explicitly indicate how h = (C) is to be implemented.

Let us assign names to the configurations of the C-register in analogy to the P_i as names for the configurations of the F-register:

$$\Gamma_0 = C'_2 C'_3 C'_4$$
$$\Gamma_1 = C'_2 C'_3 C_4$$
$$\Gamma_2 = C'_2 C_3 C'_4$$
$$\Gamma_3 = C'_2 C_3 C_4$$
$$\Gamma_4 = C_2 C'_3 C'_4$$
$$\Gamma_5 = C_2 C'_3 C_4$$
$$\Gamma_6 = C_2 C_3 C'_4$$
$$\Gamma_7 = C_2 C_3 C_4$$

Now the C-register will be in one and only one of these configurations at any time and hence for any fixed i

$$\sum_{h=0}^{7} M_{hi} \Gamma_h = M_{0i} \Gamma_0 + \ldots + M_{7i} \Gamma_7$$

will be 0 or 1 accordingly as the i^{th} digit of the word in memory which corresponds to (C) is 0 or 1.

This implements the condition h = (C) which we added verbally to the equation above.

Then the final equations for the A-register are

$$\left.\begin{aligned} {}_1{}^{a}{}_i &= \left[C_i P_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi}\,\Gamma_h\right)\right] E \\ {}_0{}^{a}{}_i &= \left[C_i' P_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M'_{hi}\,\Gamma_h\right)\right] E \end{aligned}\right\} \; i = 2,3,4$$

$$\left.\begin{aligned} {}_1{}^{a}{}_i &= (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi}\,\Gamma_h\right) E \\ {}_0{}^{a}{}_i &= (P_2 + P_7)\left(\sum_{h=0}^{7} M'_{hi}\,\Gamma_h\right) E \end{aligned}\right\} \; i = 0,1$$

$$\Big\} \; h = 0,\ldots,7$$

3.34 The C-Register

The inputs to the C-register are affected in configurations 100, 010, 011, 001 (in case $B_0 = 0$) and 101 of the operation counter, and in case G = 0, regardless of the state of F (see p. 5). Only three different things go on:

(1) (Ad A) = C (on P_4, P_2, P_3 and $P_1 B_0'$)

(2) 0 = C (on G')

(3) (C) + 1 = C (on P_5)

The first two functions are easily handled as above, using the set and clear inputs:

$$\left.\begin{aligned} {}_1{}^{c}{}_i &= (P_1 B_0' + P_2 + P_3 + P_4) A_i E \\ {}_0{}^{c}{}_i &= \left[(P_1 B_0' + P_2 + P_3 + P_4) A_i' + G'\right] E \end{aligned}\right\} \; i = 2,3,4$$

The "G'" term in the equation for ${}_0{}^{c}{}_i$ provides that when the computer is started up, the first memory address to be referred to will be 000.

The add one function can be implemented by a simple counter using the complement inputs. The cycle for the counter is:

C_2	C_3	C_4
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1
0	0	0

Apparently the stages should be complemented as follows:

$$c_2 = (C_2' C_3 C_4 + C_2 C_3 C_4) P_5 E$$

$$= C_3 C_4 P_5 E$$

$$c_3 = C_4 P_5 E$$

$$c_4 = P_5 E$$

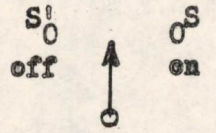
$$\boxed{\begin{aligned} c_2 &= C_3 c_3 \\ c_3 &= C_4 c_4 \\ c_4 &= P_5 E \end{aligned}}$$

3.35 The Start Flip-Flop, G

On the halt command, on the time pulse before the operation (F) counter reaches configuration 000 the G flip-flop is cleared and therefore the F-counter sticks on 000 (see p. 5). The machine is to be started again by use of a switch, S, which sets G and allows the next clock pulse to step the operation counter into configuration 111. The machine can be stopped not only on the halt command, but at any time by setting S to off.

$$_0g = (P_4 A_0' A_1' + S')E$$

$$_1g = SE$$



Shown in rest position. Can be locked in off. Contact in on state is momentary.

3.36 The Memory

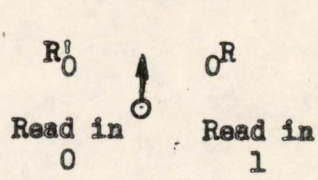
Read-in to the memory is to be accomplished in one of two ways

(1) Program: in the 011 configuration of the operation counter (B) $\Rightarrow$ M[C].

(2) Fill Buttons: the individual flip-flops can be filled "by hand"; this method is used for reading in programs.

The "fill buttons" are switches which in their rest positions excite neither $_0m$ nor $_1m$.

$$\left.\begin{aligned} 1^{m}_{wd} &= (R_{wd} + B_d \lceil_w P_3)E \\ 0^{m}_{wd} &= (R'_{wd} + B'_d \lceil_w P_3)E \end{aligned}\right\} \quad \begin{aligned} w &= 0,\ldots,7 \\ d &= 0,\ldots,5 \end{aligned}$$



Shown in rest position. Contact is momentary in both positions.

We now have a set of equations which completely describe the logic of our machine. They are summarized on p. 21 and p. 22.

4.0 GENERAL REMARKS

There are three general stages in the development of the logic of any particular computing machine. In order of decreasing generality of the decisions involved, they are:

(1) Planning. (Given the purpose of the machine, what operations shall be included? What word-length? What speed per operation?.....)

(2) Combinatorial Decisions. (What general arrangement of devices will best implement the plan: what shall be the cycle of the operation counter? What order code? What basic functions (commands)?.....)

(3) Design. (What configuration of memory-elements, gates, mixers, etc. best realizes the results of the combinatorial decisions?)

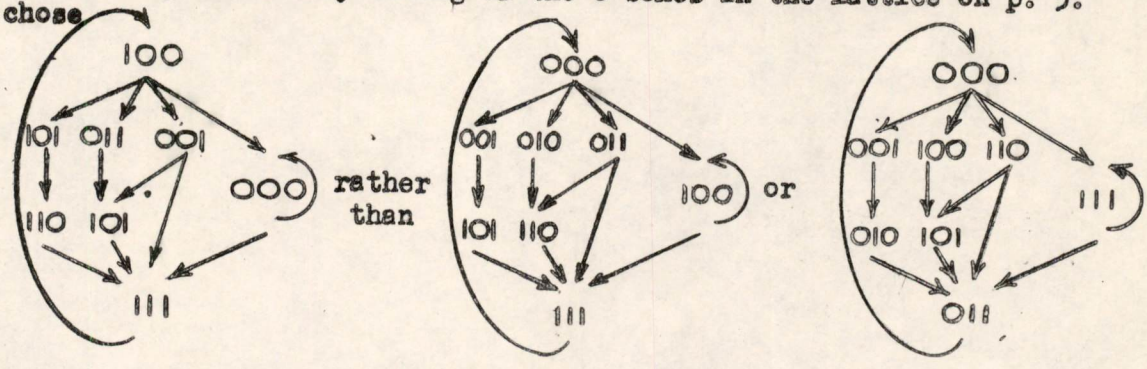
In the case of the present "sample" computer these steps went somewhat as follows:

(1) Planning: The purpose of the machine was to serve as an example of a method of design. Therefore it should have a small memory, short word length and few operations; but none of these should be so short, small or few as to make the design problem trivial. The results of the planning were presented, without discussion, on the first three pages.

(2) Combinatorial Decisions: These are summarized by the table on p. 5. There the cycle of the operation-counter is shown and the groups of basic commands indicated. Since the operations are few and simple, no thought at all was devoted to the coding of them (00 =halt, 01 = cp, etc.)

(3) Design: Throughout the design we referred to the table on p. 5 which summarized the combinatorial decisions. The design is almost mechanical, once the table has been drawn.

The sort of problem involved in the combinatorial decisions is illustrated by the binary coding of the 8 boxes in the lattice on p. 5. We chose



because we thought that of all the 8! ways of assigning numbers from 0 through 7 to the 8 points in the lattice, the one selected resulted in the simplest design. Of course we did not draw up all 8! lattices and from them write 8! sets of equations for the machine, examine them and choose the one that contained the fewest components. Such a program might be carried out by a machine like WWI, but not by human beings. Rather we considered e.g. that it would be desirable to have as little difference as possible between the numbers assigned to the points in the second level of the diagram, the outcomes of the four-way choice after 100. Accordingly, in a simple-minded way, we decided to use the four numbers beginning with "0" for those points. Similarly in the binary decision after 001 we chose "101" and "111" as the numbers of the next configurations because they differ from each other only in a single digit.

To systematize such combinatorial decisions, three developments would be helpful:

(1) A mathematical theory of the desirability of such choices as the two mentioned above (which were made on an intuitive basis).

(2) Assignment of numerical values to different components and their configurations (the unit might be dollars, or speed, reliability or some combination of those). Any such assignment should be in general terms so that the parameters might be changed as the characteristics of available components change.

(3) Programs which would allow a high-speed computer to survey large numbers of possibilities such as the 8! assignments of numbers to the points in the lattice above. It is possible to write down a fully mechanical procedure for simplifying equations and even for going from a diagram like the one on p. 5 to a set of equations, so that this development is <u>possible</u> - and perhaps practical.

5.0 SUMMARY

5.1 Complete Set of Equations for the Machine

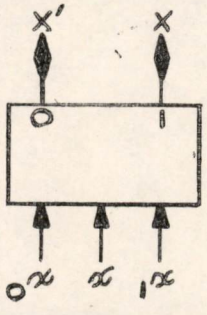
5.11 Abbreviations

$P_i = F_1^{*}F_2^{*}F_3^{*}$ where the n^{th} "*" is either a prime or a blank, depending on whether the n^{th} digit in the binary representation of i is a "0" or a "1". e.g., $P_5 = P_{101} = F_1F_2'F_3$.

$\Gamma_i = C^{*}C^{*}C^{*}$. E.g., $\Gamma_1 = \Gamma_{001} = C_1'C_2'C_3$

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n$$

5.12 Notation for Flip-Flops



5.13 The Operation Counter (F-Register)

$$f_1 = (P_0G + P_1 + P_2 + P_3 + P_4)E$$

$$f_2 = (P_0G + P_1B_0 + P_3 + P_4A_0 + P_5 + P_7)E$$

$$f_3 = (P_0G + P_4A_1 + P_6 + P_7)E$$

$$_0f_i = G'E \quad (i = 1,2,3)$$

5.14 The Accumulator (B-Register)

$$b_i = (A_i' \oplus K_i)P_6E \quad (i = 0,1,2,3,4)$$

$$\left.\begin{aligned} {}_1b_i &= B_{i-1}P_3E \\ {}_0b_i &= B_{i-1}'P_3E \end{aligned}\right\} \quad (i = 1,2,3,4)$$

$$_0b_0 = P_3E$$

$$K_{i-1} = A_i'K_iP_6E + B_ib \quad (i = 1,2,3,4); \quad K_4 = 1$$

5.15 The A-Register

$$\left.\begin{aligned} {}_1a_i &= \left[C_iP_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi}\left\lceil_h\right)\right]E \\ {}_0a_i &= \left[C_i'P_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi}'\left\lceil_h\right)\right]E \end{aligned}\right\} \quad \begin{aligned} &i = 2,3,4; \\ &h = 0,\ldots,7 \end{aligned}$$

Same for i = 0,1 except that last term is missing.

5.16 The Memory-Selection Register (C)

$$\left.\begin{aligned} {}_1c_i &= XA_iE \\ {}_0c_i &= (XA_i' + G')E \end{aligned}\right\} \quad i = 2,3,4$$

where $X = P_1B_0' + P_2 + P_3 + P_4$.

$c_i = C_{i+1} \quad (i = 2,3)$

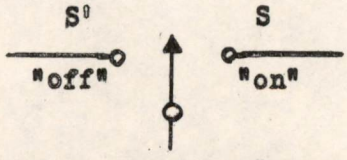
$c_4 = P_5E$

5.17 The Start-Stop Flip-Flop (G)

${}_0g = (P_4A_0'A_1' + S')E$

${}_1g = SE$

S is a double-throw switch which can be locked in the off position, makes momentary contact in the on position, and has a rest position in which neither contact is energized.

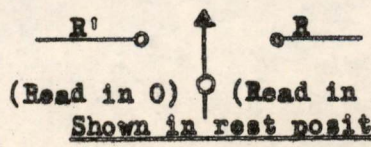


Shown in "rest" position.

5.18 The Memory Registers (M)

$$\left.\begin{aligned} {}_1m_{wd} &= (R_{wd} + B_d\lceil_w P_3)E \\ {}_0m_{wd} &= (R_{wd}' + B_d'\lceil_w P_3)E \end{aligned}\right\} \quad \begin{aligned} &w = 0,\ldots,7; \\ &d = 0,\ldots,5 \end{aligned}$$

The R's are double-throw switches ("fill buttons") which make momentary contact on either side and have a rest position in which both sides are open.



Shown in rest position

5.2 The Flip-Flops

These are assumed to have an inherent delay between receipt of an input pulse and the resulting change in the output. Speaking loosely, this delay permits reading new information into a register "at the same time" that the old information is being read out, and permits "simultaneous" transfers of information between two registers. (See Section 2.1.)

5.3 Operating Instructions

1. Turn on power and allow tubes to warm up. During this time the machine will be cycling through some meaningless program if the G flip-flop happens to start out holding a "1".

2. Lock switch S in the off position. This sets G to 0, sets the operation counter to 000, and reads 000 into C. Read program and data into memory with "push-buttons" (R).

3. Release S (nothing happens while S is in neutral position), push it to the on position and release. This sets G to 1 and the operation counter steps into 111, the beginning of program timing. Since C holds 000, the first word to be taken out of memory and transferred into the A-register will come from the first memory register (at location 000). The program should be stored with this fact in mind.

4. The program should end with a halt command, 00xyz, where x y z may be any number. This freezes the contents of all FF's except for G and the F and C registers, all of which are cleared. A new program may now be read into storage via the R switches and the machine may be re-started by pushing the S switch over to on.

5. Read-out is via neon bulbs attached to the memory FF's (not shown above).

5.4 Block Diagrams

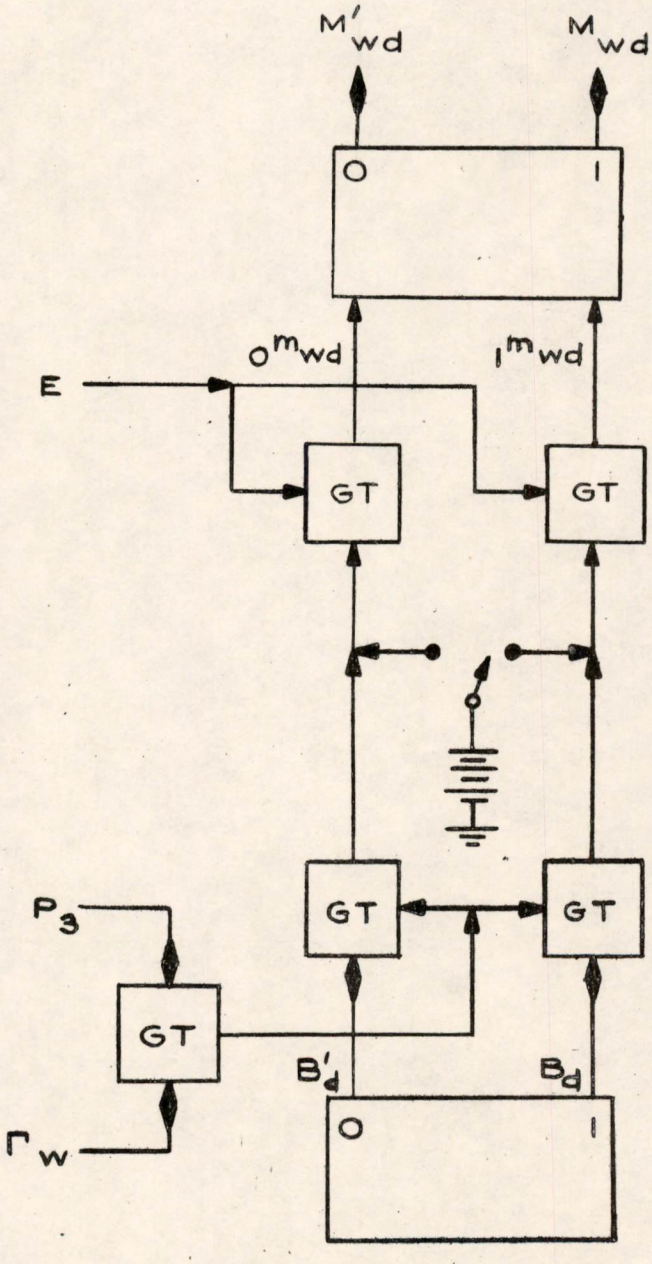
People who are unfamiliar with the algebraic notation may find it helpful to translate some of the equations into block diagrams. A little bit of this will go a long way toward promoting familiarity with the algebraic symbolism; with a little practice it will be found that as much information can be read directly from a few lines of equations as from a dense page of block diagrams.

Block diagrams for a few of the registers follow.

Note that the equations for a register show the inputs to that register. To find outputs of register X, look through all the equations for occurrences of the letter "X". (In the case of the F and C registers, look for "P" and "Γ" as well, since these abbreviate products of "F" s and "C"s.)

Memory

$$\left.\begin{aligned} {}_1m_{wd} &= (R_{wd} + B_d \lceil_w P_3) E \\ {}_0m_{wd} &= (R'_{wd} + B'_d \lceil_w P_3) E \end{aligned}\right\} \quad \begin{aligned} w &= 0,\ldots,7 \\ d &= 0,\ldots,5 \end{aligned}$$



Identical picture

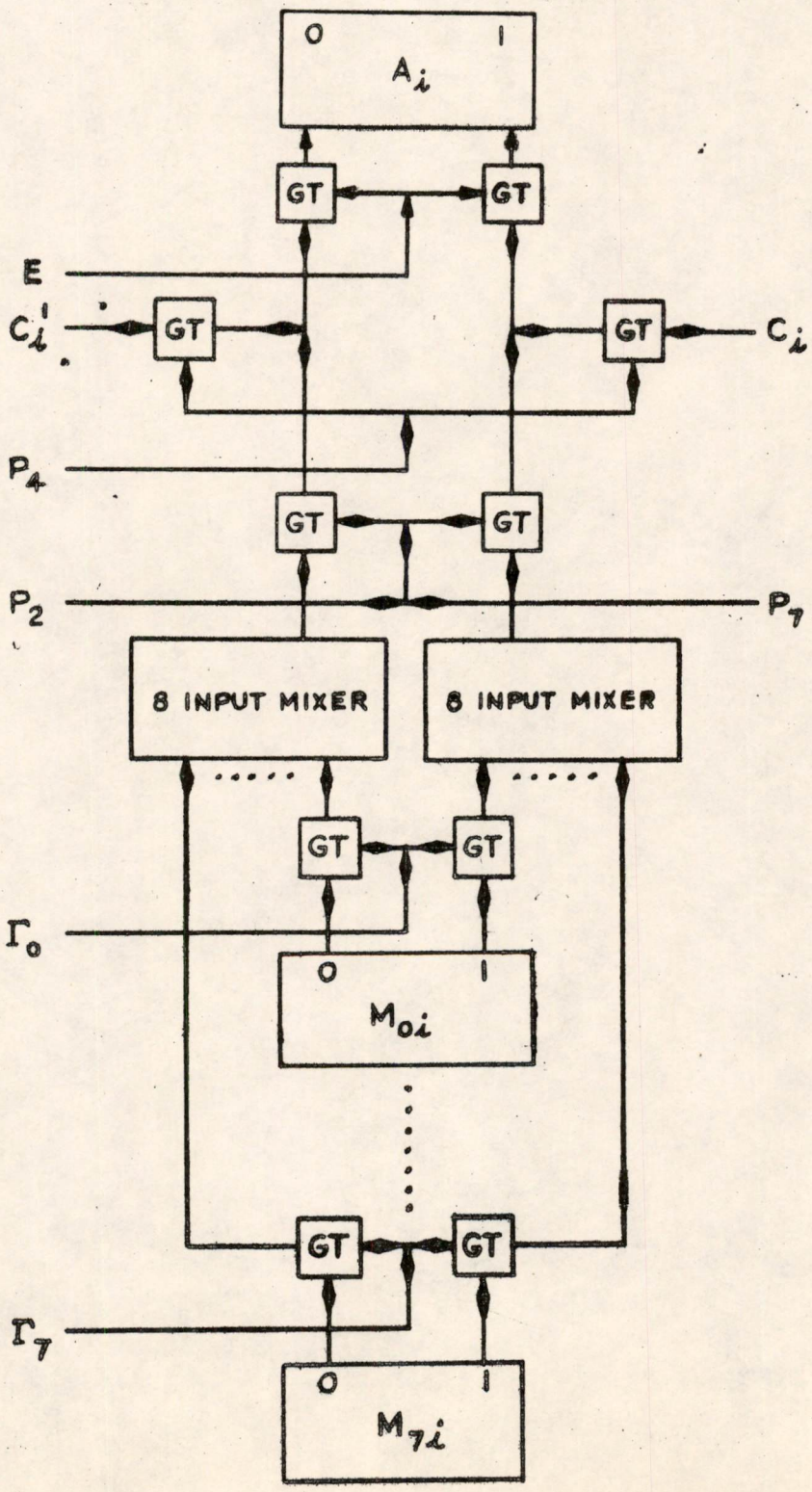
for each memory cell.

In case a magnetic core memory is used with an I/2 selection system, the read-in gates are simply points where pairs of insulated wires cross inside the cores.

The A-Register (last three digits)

$$1^{a}_{i} = \left[C_i P_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi} \lceil_h\right) \right] E \qquad i = 2,3,4;$$

$$0^{a}_{i} = \left[C'_i P_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M'_{hi} \lceil_h\right) \right] E \qquad h = 0,\ldots,7$$



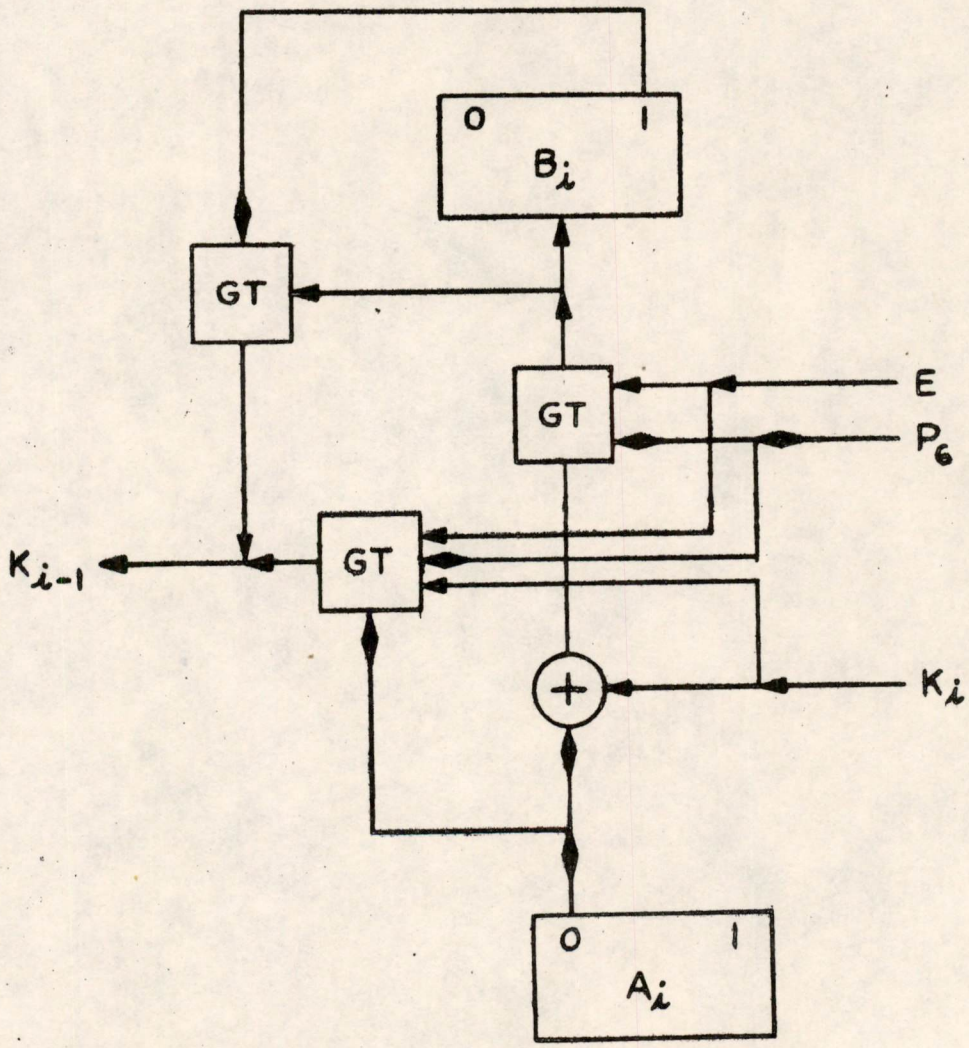
This memory system is not intended as a practical proposal. The following two modifications are desirable: (1) In reading out of memory, clear A on the preceding clock pulse and then read into only the set sides of the A FF's. This eliminates the left hand half of the memory read-out gates and the left hand 8 input mixer. (2) Use magnetic cores as memory cells. Then the remaining read-out gates are inherent in the cores, and the 8 input mixer is the sensing winding.

The Accumulator (complement inputs)

The set and clear inputs are used only for shifting. The reader can complete this diagram by adding them.

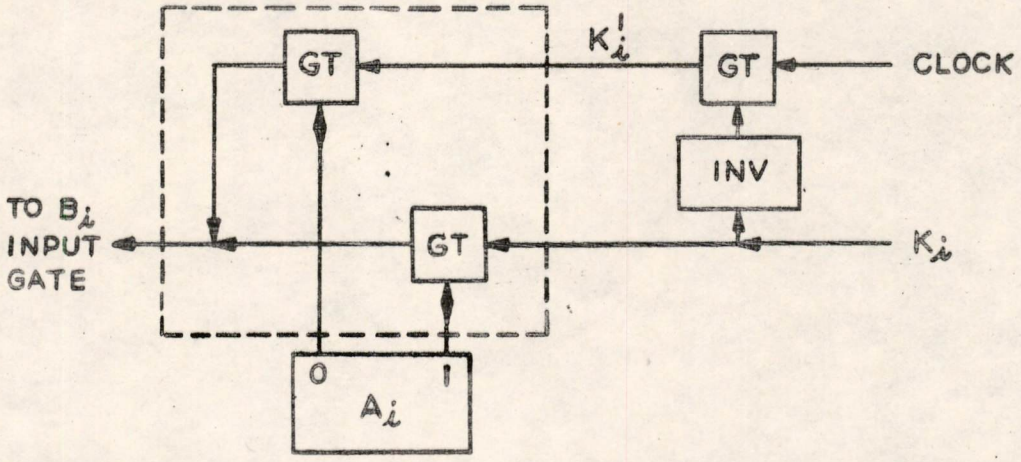
$$b_i = (A_i' \oplus K_i)P_6E \quad (i = 0,\ldots,4)$$

$$K_{i-1} = A_i'K_iP_6E + B_ib .$$



⊕ represents a black box which gives a pulse output whenever one but not both of the inputs are present. See next page.

The black box for partial sum ($\oplus$) is an oversimplification. Probably the best way to implement it with customary electronic components is to use not only A_i and K_i but also their complements:



This requires either providing an inverter to get K_i' out of K_i, or building up K_{i-1}' as an output from stage #i independent of K_i. (A carry zero line as well as a carry one line.)

SIGNED Richard C. Jeffrey
Richard C. Jeffrey

Irving S. Reed
Irving S. Reed

APPROVED
N. H. Taylor

RCJ/ISR/cp

cc: G. R. Briggs
D. R. Brown
D. A. Buck
H.R.J. Grosch
W. A. Hosier
J. Jacobs
W. Linvill
R. P. Mayer
J. A. O'Brien
W. Ogden
K. H. Olsen
W. N. Papian

BIBLIOGRAPHY

1. Jeffrey, R. C., Reed, I. S., "The Use of Boolean Algebra in Logical Design", Engineering Note E-458-1, Digital Computer Laboratory, M.I.T., (April, 1952).

2. Reed, I. S., Division II Status Report, Project Lincoln, (April, 1952).

3. Reed, I. S., "Some Mathematical Remarks on the Boolean Machine", Project Lincoln Technical Report No. 2, (December 19, 1951).

HARLAN E. ANDERSON
2.20

Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

SUBJECT: DESIGN OF A DIGITAL COMPUTER BY BOOLEAN ALGEBRA

To: N. H. Taylor

From: R. C. Jeffrey, I. S. Reed

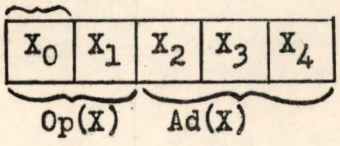
Date: May 20, 1952

Abstract: The techniques described in E-458-1 are illustrated in a real-life situation: the design of a 4-order machine employing an unusual method of central control. For definiteness the memory size is taken to be 8 words, but no logical complexity is added when this is increased to a realistic figure.

1.0 SPECIFICATIONS FOR THE MACHINE; NOMENCLATURE

1.1 Words: 5 bits, interpreted as in WWI, with negative numbers represented in nines complement form.

Interpreted as a number:

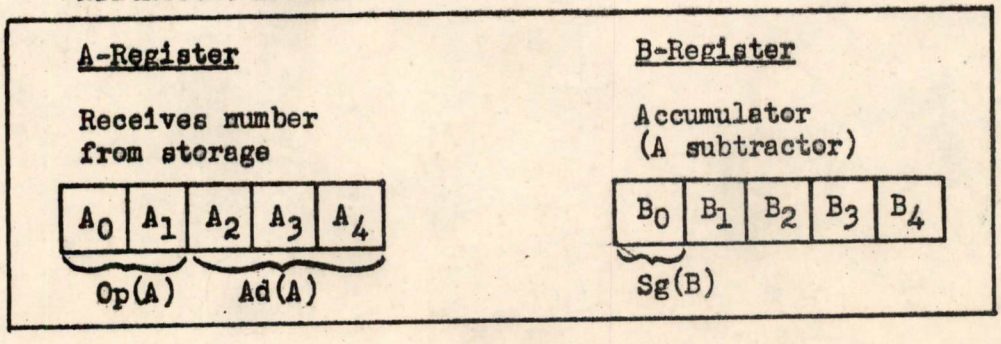


Interpreted as an instruction:

1.2 Memory: 8 words, stored in flip-flops. An actual memory would use some other device, such as iron cores; but the analysis for the flip-flop case can easily be applied to whatever device is actually used.

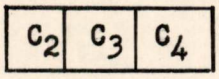
1.3 Registers

Arithmetic Element



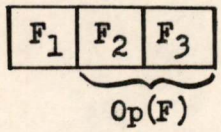
C-Register

Holds address of next memory register to be used.



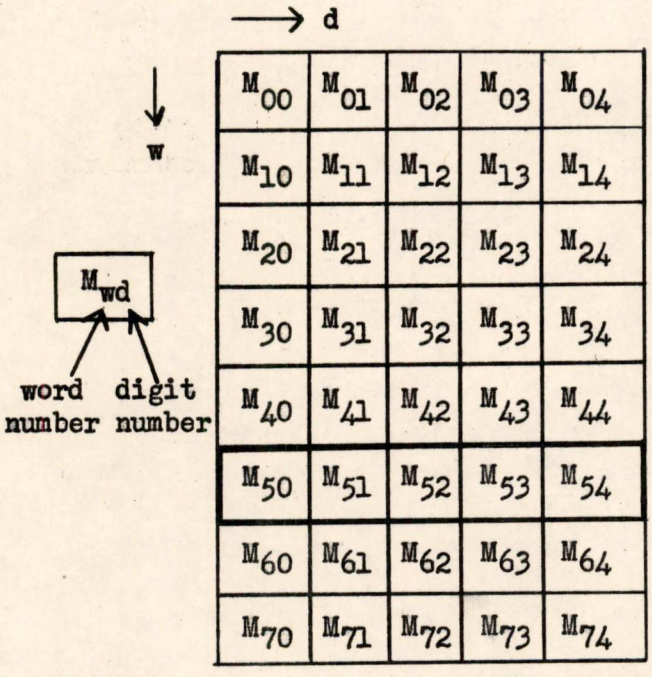
F-Register

Operation counter; central control.



Start Flip-Flop

Memory



"M(C)" denotes the address in memory corresponding to the number in the C-register. Thus if the C-register holds 101, M(C) is

1.4 Operations

The decisions as to word length, number of registers, etc., summarized above come under the heading of machine planning, and were arrived at by cut-and-try. The last and most delicate part of the planning concerns the operation of central control.

Code No.	Name	Description
00xyz	Halt	The machine stops, i.e., all flip-flops remain in the states they were in when the operation was executed, until the start button is depressed, at which time program timing begins.
01xyz	Conditional Subprogram	If the sign digit of the B-register is 1, take next instruction from memory location xyz. If the sign of the B-register is 0, take next instruction in sequence.
10xyz	Subtract	Take number from memory location xyz and put it in the A-register; subtract it from the contents of the B-register and leave the result in B.
11xyz	Shift right	Shift the contents of the B-register right, depositing the original contents of B in M(C).

2.0 TIMING DIAGRAM

2.1 Notation

The above four operations are to be performed as sequences of commands to perform the elementary independent functions of which the machine is capable. In order to describe these functions we adopt a compact terminology:

M(C): The memory location whose number is stored in C.

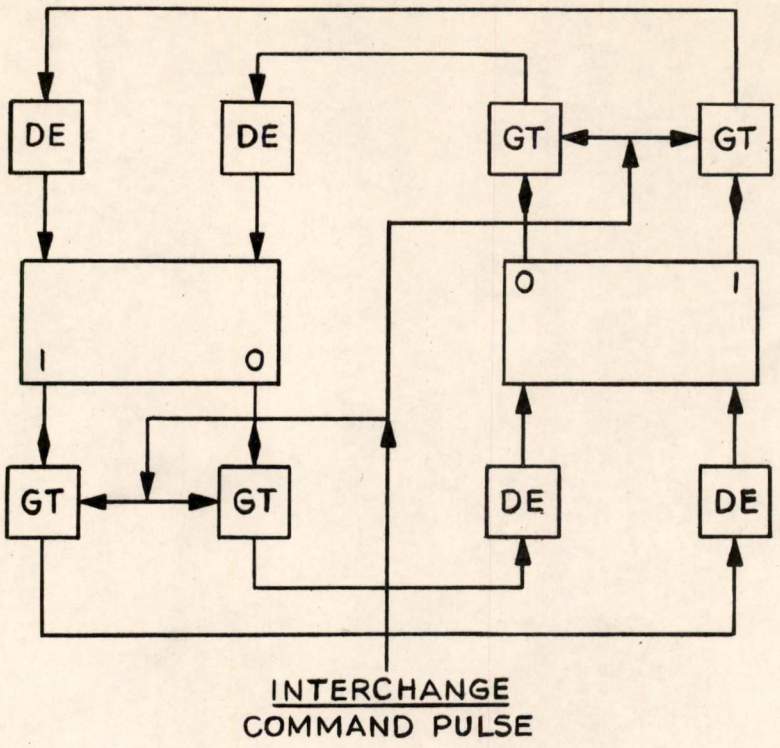
X: The X register.

(X): The contents of the X register.

(X) $\Rightarrow$ Y: The contents of register X at time t appear in register Y at time t + ϵ.

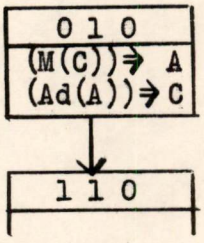
Then it is possible to have "simultaneous" interchanges:

"(X) $\Rightarrow$ Y and (Y) $\Rightarrow$ X" means that the contents of X at t appear in Y at t + ϵ, and that the contents of Y at t appear in X at t + ϵ. This can be implemented by delay elements (which we shall henceforth assume to be built into the FF's):



We now draw a flow diagram for central control (next page). The numbers at the tops of the boxes will represent states of the F-counter, and the notations inside the boxes represent the commands which are performed by the first time pulse during which the F-counter is in the indicated state.

For example, in the upper left corner on page 5

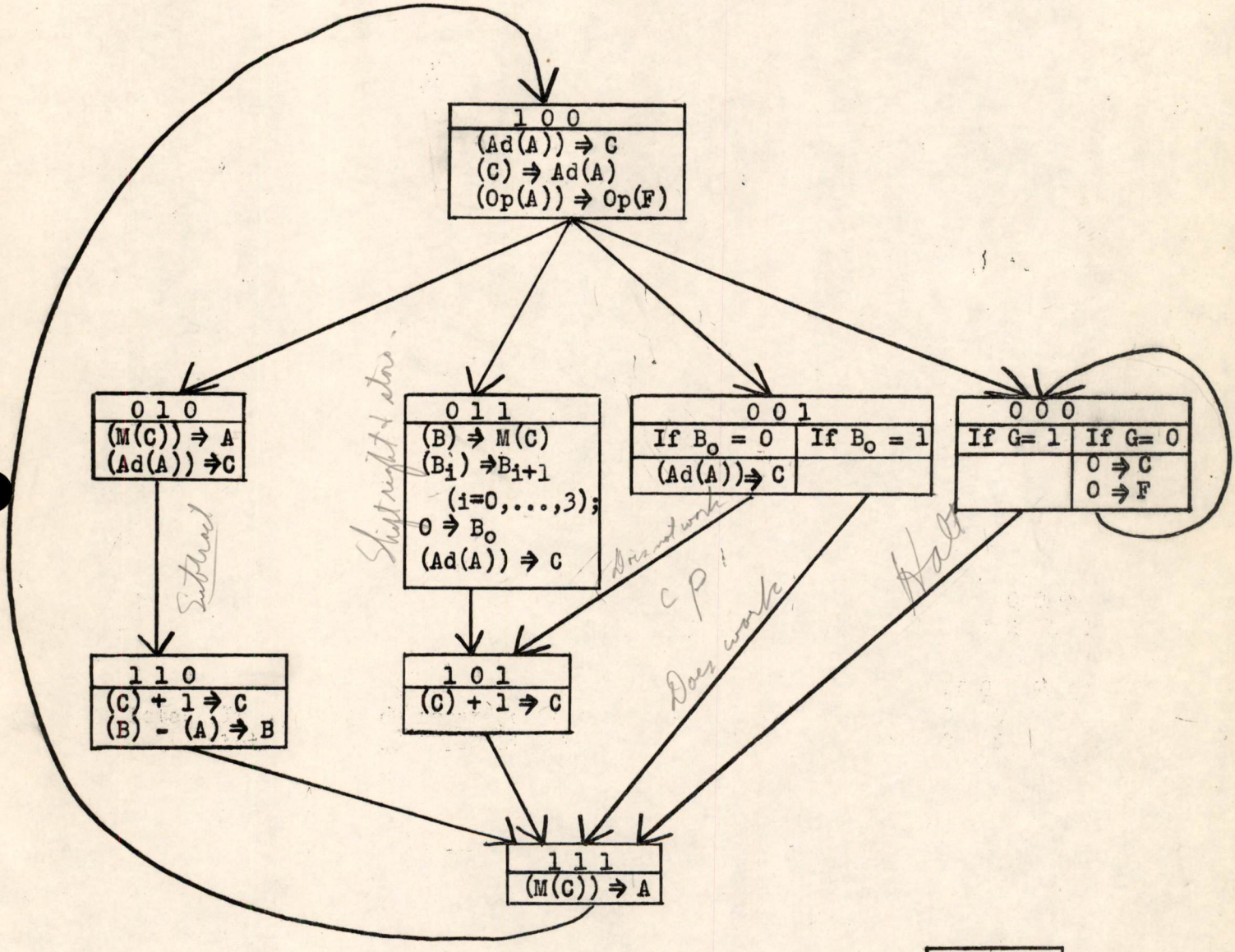


means that when the F-register is in the configuration $F_1 = 0$, $F_2 = 1$, $F_3 = 0$ the commands read out of memory location (C) into the A-register and read the last three (address) digits of the A-register into the C-register are performed and the F-counter changes to the configuration 110 (i.e., F_1 is complemented). This change in (F) is strictly speaking a command, but is indicated by the arrow between the boxes instead of by a third notation inside the upper box.

For further explanation of the flow diagram, see page 6.

2.2 Flow Diagram

(The numbering of the boxes should be thought of as arbitrary, for the present. See Section 4.0 below.)



Note: If G = 0, then regardless of the state of F, both the F and C registers are cleared:

If G = 0
0 ⇒ C
0 ⇒ F

Program timing occurs at 111 and 100. Note that the C-register is used both as a storage selection register and a program counter: in 110 and 101 we add 1 to (C) . Each of the four paths after 100 corresponds to an operation:

010 begins the operation timing for subtract (10xyz)
011 begins the operation timing for shift (11xyz)
001 begins the operation timing for cp (01xyz)
000 begins the operation timing for halt (00xyz)

The four way decision is determined by the result of the command $(Op(A)) \Rightarrow Op(F)$, which reads (A_0) into F_2 and (A_1) into F_3. At the beginning of operation timing for operation number v w the F-register will be in configuration 0 v w, since previously we had (A_0) = v and (A_1) = w for Op(A). It is also necessary to arrange that F_1 be complemented when the F-counter leaves configuration 100.

The reader should now verify that the sequences of commands specified in the flow diagram really do add up to the four operations listed above. The logical details are discussed in section 3.

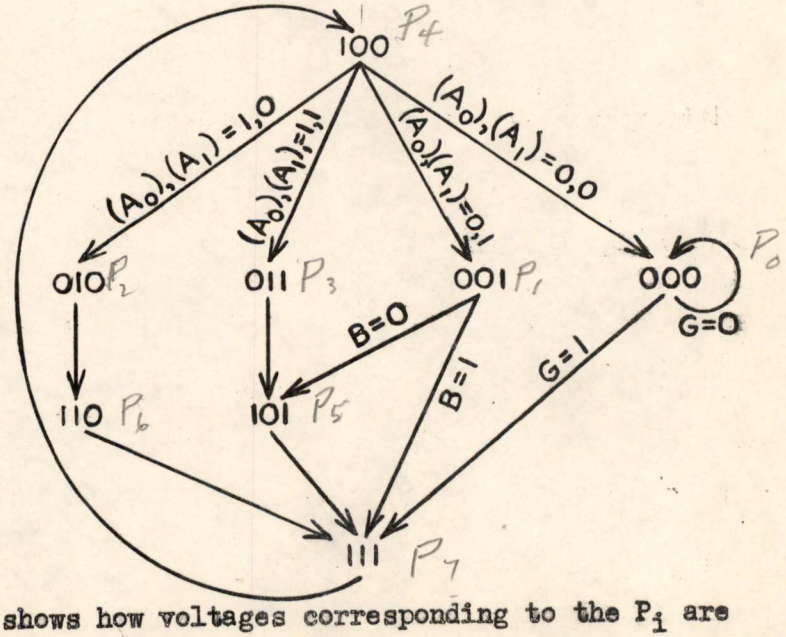
3.0 DESIGN OF THE MACHINE: INPUT EQUATIONS

3.1 Design of the F-Register

Apparently the F-register is to be a 3 stage binary counter with the following cycle:

We name the configurations of the F-register:

$P_0 = F_1'F_2'F_3'$ --- 000
$P_1 = F_1'F_2'F_3$ --- 001
$P_2 = F_1'F_2F_3'$ --- 010
$P_3 = F_1'F_2F_3$ --- 011
$P_4 = F_1F_2'F_3'$ --- 100
$P_5 = F_1F_2'F_3$ --- 101
$P_6 = F_1F_2F_3'$ --- 110
$P_7 = F_1F_2F_3$ --- 111

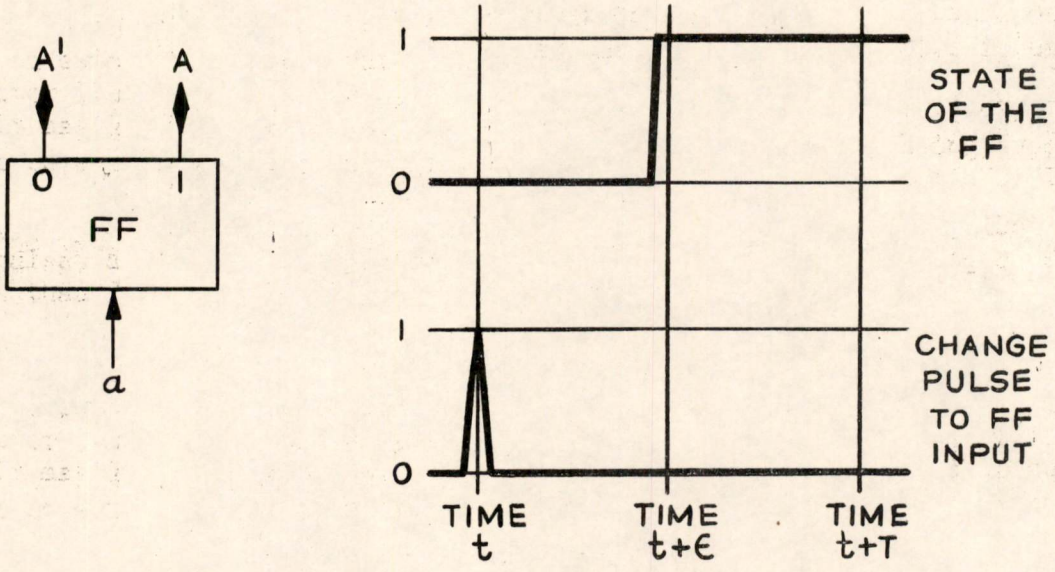


The block diagram on page 9 shows how voltages corresponding to the P_i are generated.

In the table above, the binary representation of the subscript of P indicates, by the distribution of zeros, the location of the primes in the product of the F's.

Now the successive states of the counter are determined by information external to the counter, as well as from the configuration of the counter itself. For example, from P_3 the counter must go to P_5, but from P_1 it may go to P_5 or to P_7. It is necessary to indicate the basis on which this selection is made; this is done in the cases where a choice is necessary.

Assume that the FF's used have one input, such that if that input is pulsed at time t, the FF is complemented at time $t+\epsilon$, ϵ being a delay less than T, the period of the clock.



$A(t+T) = A(t) \oplus a(t)$ (See E-458-1, p. 13)

Now to design the F-counter it is sufficient to write three equations, one for each stage, specifying when the change inputs are to be pulsed. We shall work through the lattice on page 6 (or page 5) level by level for each input. The equations we are about to build up are illustrated by a block diagram on page 9.

For the first digit (F_1) we see that in going from P_4 (=100) on the first level to the second, F_1 is to be complemented regardless of which of the four configurations on the second level is selected:

$$f_1 = (P_4 + \ldots\ldots\ldots)E$$

Here E denotes the clock pulse: the changes are to occur at each clock pulse, so that an operation will require at most 4 clock pulses for its completion.

To get the second term: if P_2 (=010) is the second level configuration selected, the first digit must change again to get the configuration 110:

$$f_1 = (P_4 + P_2 + ...)E$$

Ditto in case 011 was the second level configuration selected: to get from 011 to 101, F_1 must be complemented:

$$f_1 = (P_4 + P_2 + P_3 + ...)E$$

Similarly, whichever value B_0 may have, the F-counter will go from the configuration 001 to a configuration (101 or 111) in which the first digit is changed:

$$f_1 = (P_4 + P_2 + P_3 + P_1 + ...)E$$

Finally, the counter goes from 000 to 111 in case G = 1. Thus F_1 will change in case P_0G. ("P_0G" means $P_0 = 1$ and G = 1.)

$$\boxed{f_1 = (P_4 + P_2 + P_3 + P_1 + P_0G)E}$$

Since F_1 does not change in going from 111 to 100, the above formula represents all the conditions under which F_1 must be complemented.

Similarly, for F_2 we have

$$\boxed{f_2 = (A_0P_4 + P_3 + B_0P_1 + GP_0 + P_5 + P_7)E}$$

where the first term derives from the fact that in the two cases where $A_0 = 1$ $\boxed{(O_p(A)) = 10}$ and $\boxed{(O_p(A)) = 11}$ the second digit of "100" must be changed, but in the other two cases (when "100" goes to "001" or to "000") this is not necessary. Finally

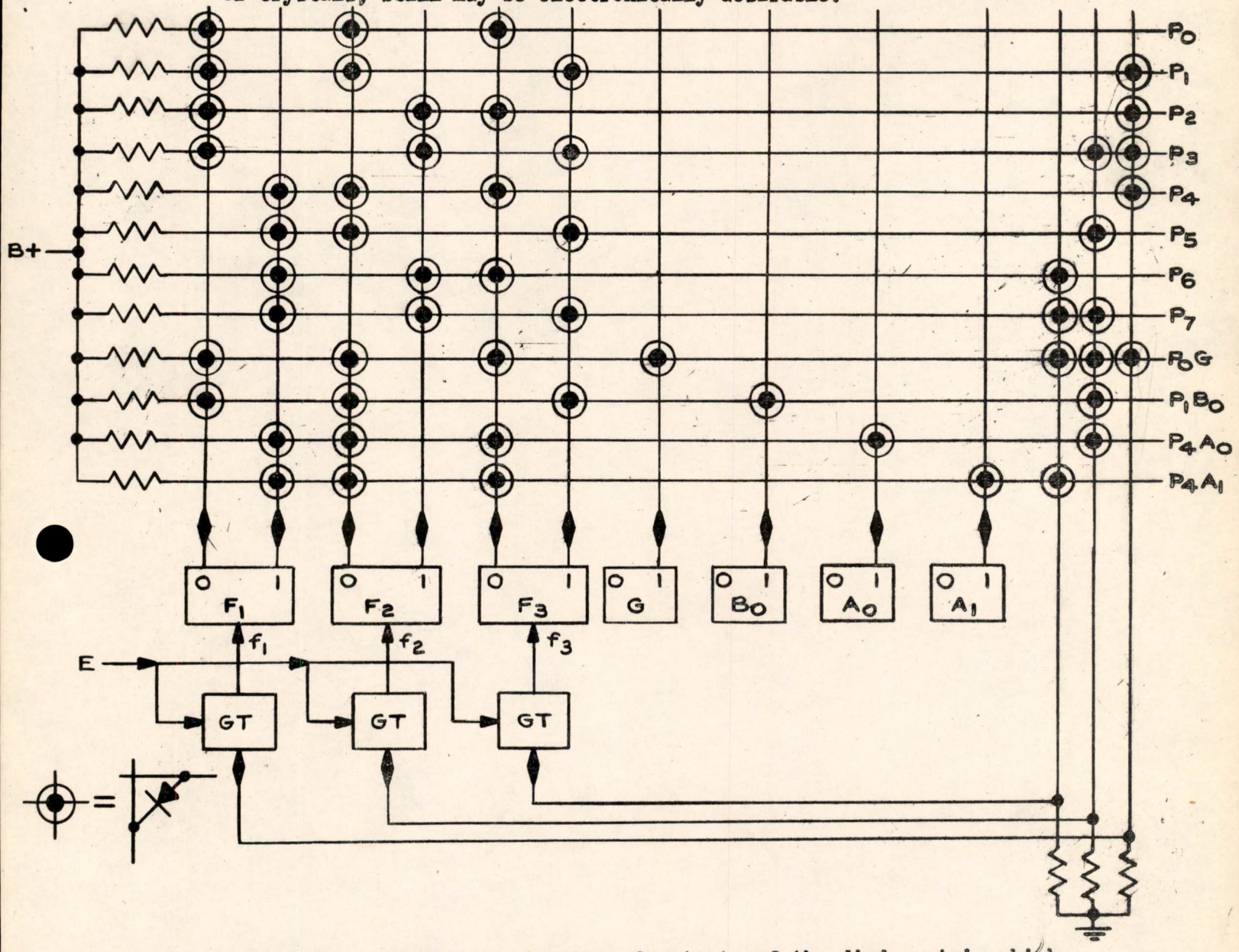
$$\boxed{f_3 = (A_1P_4 + GP_0 + P_6 + P_7)E}$$

It remains to implement the condition that if G = 0, 0 = F. Note that if a halt is programmed, the F-counter steps from 100 to 000 and in so doing clears G. But as will be seen in Section 3.35 below, G can also be cleared by a pushbutton. We wish to provide that if G = 0 from any cause, the F-counter will assume the configuration 000 on the next time pulse.

This is most easily accomplished by writing three equations for the clear inputs to the F flip-flops:

$$\boxed{{}_of_i = G'E \qquad i = 1,2,3}$$

Now these equations may be realized immediately by a two-level diode matrix which, although it does not represent a minimization of number of crystals, still may be electronically desirable.

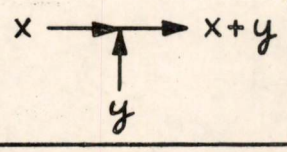
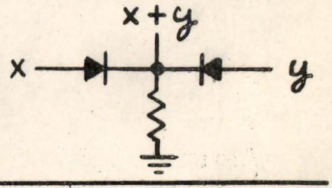
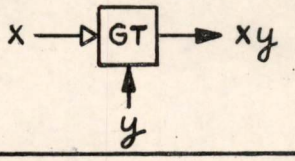
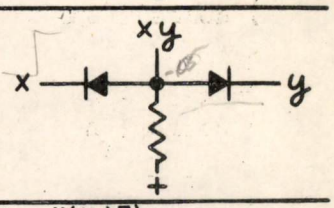
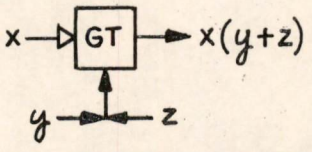
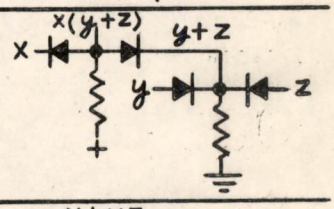
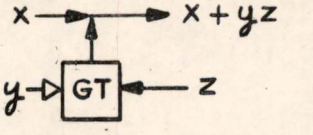
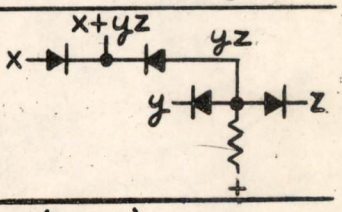
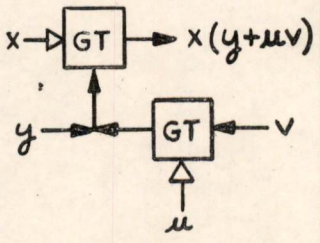
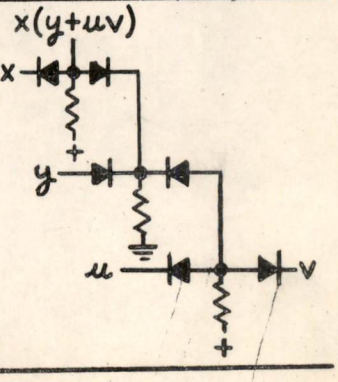


Here we have indicated only those 3 outputs of the diode matrix which feed back into the F-register. Actually we require other outputs: command pulses to the rest of the machine. They will be considered below in connection with the FF registers which they control.

3.2 Digression on "Levels"

Note that the three equations for the F-register can be factored, resulting in a reduction in the number of gates and mixers required for the realization of central control. But this numerical reduction is at the expense of an increase in the number of levels of gating and mixing. We illustrate the meaning of "levels" by examples.

EXAMPLE

No. of Levels	Equations	Block Diagrams	One Electronic Realization (Diode)
1	x + y		
	xy		
2	x(y + z)		
	x + yz		
3	x[y + uv]		

Apparently in terms of the equations, the number of levels is the same as the functional complexity of the expression for the output. That is, $x + y$ and xy are functions whose arguments are variables; $x(y + z)$ and $x + (yz)$ are functions (the first a product and the second a sum) in which one of the arguments is itself a function of variables.

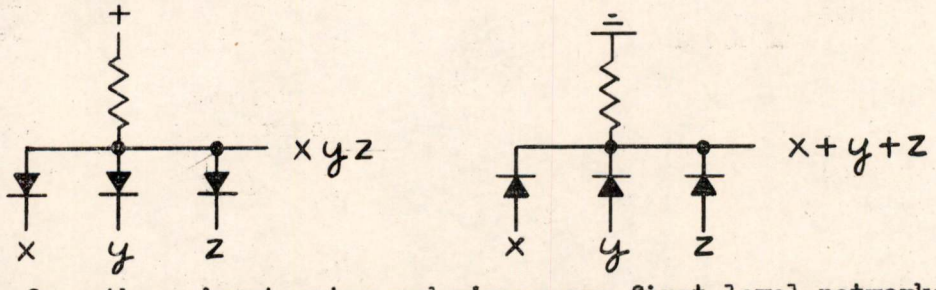
Finally, $x[y + uv]$ is a product one of whose factors is a sum one of whose terms is a product: "$\pi\{x, \sigma[y, \pi(u,v)]\}$" in functional notation where $\pi(a,b) = ab$ and $\sigma(a,b) = a + b$. Briefly, the number of levels in a Boolean function is identical with the number of nestings of parentheses within each other.

In terms of block diagrams, the number of levels in a device is determined by drawing lines from each input to the output; the largest number of gates and mixers through which such a line passes in the level of the system. Finally, the reason we are interested in levels is that if a function of level n is realized with diodes, n is the largest number of diodes through which any input current must pass in order to reach the output. In a high-level diode network considerable attention must be paid to the values of voltage and resistance associated with the diodes. Other difficulties arise with high-level vacuum tube and transistor networks. In all cases the difficulties simply require engineering attention: high-level networks are realizable.

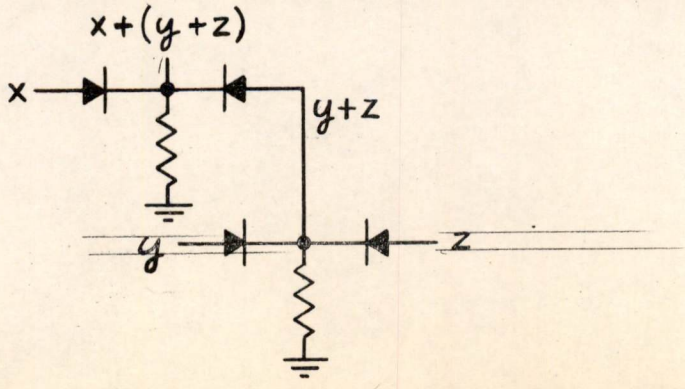
One last comment: in such expressions as "xyz" and "x + y + z", the product and sum are regarded as functions of three variables:

$$\pi_3(x,y,z) = xyz \text{ and } \sigma_3(x,y,z) = x + y + z$$

Electronically this corresponds to the use of three-input diode devices for mixing and gating:



and therefore three-input gates and mixers are first level networks. However, expressions like "$x + (y + z)$", which might be written in functional notation as "$\sigma_2(x, \sigma_2(y,z))$" have as their direct realizations two-level mixers like:



in which the longest current path passes through two diodes. The two sorts of mixers are logically equivalent, and the one-level type is used where possible.

To illustrate these points, let us factor the equations of the F-register:

$$f_1 = P_1 + P_2 + P_3 + P_4 + P_0G \quad \text{(Ignore the "E" for the present.)}$$

$$= \underline{F_1}F_2'F_3 + \underline{F_1}F_2F_3' + \underline{F_1}F_2F_3 + F_1F_2'F_3' + \underline{F_1}F_2'F_3'G$$

$$= F_1'(F_2'\underline{F_3} + F_2F_3' + F_2\underline{F_3} + F_2'F_3'G) + F_1F_2'F_3'$$

$$= F_1'\left[F_3(F_2' + F_2) + F_3'(F_2 + F_2'G)\right] + F_1F_2'F_3'$$

$$= F_1'\left[F_3 + (F_2 + G)\right] + F_1F_2'F_3' \quad \text{(by 2 applications of the rule: } x + x'y = x + y\text{)}$$

$$\boxed{f_1 = \left[F_1'(F_2 + F_3 + G) + F_1F_2'F_3'\right]E}$$

$$f_2 = A_0P_4 + P_3 + B_0P_1 + GP_0 + P_5 + P_7$$

$$= A_0\underline{F_1}F_2'F_3' + F_1'F_2F_3 + B_0F_1'F_2'F_3 + GF_1'F_2'F_3' + \underline{F_1}F_2'F_3 + \underline{F_1}F_2F_3$$

$$= F_1(A_0F_2'F_3' + F_2'\underline{F_3} + F_2\underline{F_3}) + F_1'(F_2F_3 + B_0F_2'F_3 + GF_2'F_3')$$

$$= F_1(A_0F_2'F_3' + F_3) + F_1'\left[F_3(F_2 + B_0F_2') + GF_2'\underline{F_3}\right]$$

$$= F_1(F_3 + A_0F_2') + F_1'\left[F_3(F_2 + B_0) + GF_2'\underline{F_3}\right]$$

$$\boxed{f_2 = \left\{F_1(F_3 + A_0F_2') + F_1'\left[F_3(F_2 + B_0) + GF_2'\underline{F_3}\right]\right\}E}$$

$$f_3 = A_1P_4 + GP_0 + P_6 + P_7$$

$$= A_1\underline{F_1}F_2'F_3' + GF_1'F_2'F_3' + \underline{F_1}F_2F_3' + \underline{F_1}F_2F_3$$

$$= F_1(A_1F_2'F_3' + \underline{F_2}F_3' + \underline{F_2}F_3) + GF_1'F_2'F_3'$$

$$= F_1(A_1F_2'F_3' + F_2) + GF_1'F_2'F_3'$$

$$\boxed{f_3 = \left[F_1(F_2 + A_1F_3') + GF_1'F_2'F_3'\right]E}$$

We have now reduced the total number of diodes (assuming a realization as a diode net) from 55 in the two-level matrix to 41 [10 for f_1, 19 for f_2, 12 for f_3: this can be determined directly from the equation by counting n diodes for each n input gate or mixer. E.g., for f_3..... ignoring E..... we have a 2 input mixer (the plus in front of "G") of which the inputs are a 4 input gate and a 2 input gate of which the inputs are F_1 and a 2 input mixer of which the inputs are F_2 and a 2 input gate. Adding up the numbers of inputs underlined above gives the 12 diodes required for f_3.]

We shall not draw the block diagrams for the factored equations. It would be very messy, and an engineer who understands the algebraic notation could design the network from the equations as easily as he could from the block diagrams (the equations have the added advantage of compactness: three rows of symbols, instead of a page crammed with boxes, lines and arrows).

3.3 The Arithmetic Element

3.31 The Subtractor (See diagram on p. 26)

The most complicated single command the machine must execute is subtract [(B) - (A) → B: from the contents of the B-register, subtract the contents of the A-register and deposit the result back in B.]

Now representing the binary number stored in the A-register by 'A', and similarly for B, and denoting the nines complement (result of complementing each digit) of A by '$\bar{A}$', we have

$$B = B_0 2^0 + B_1 2^{-1} + B_2 2^{-2} + B_3 2^{-3} + B_4 2^{-4}$$

$$A = A_0 2^0 + \ldots\ldots + A_4 2^{-4}$$

$$\bar{A} = (1 - A_0)2^0 + \ldots\ldots + (1 - A_4)2^{-4}$$

$$A + \bar{A} = 2^0 + \ldots\ldots + 2^{-4} = 2 - 2^{-4}$$

$$-A = \bar{A} + 2^{-4} - 2$$

$$\boxed{B - A = B + \bar{A} + 2^{-4} - 2}$$

We can now construct a table which describes the operation of the i^{th} stage of our subtractor; it is the ordinary table for binary addition where the second term is A_i' instead of its complement, A_i. Note that we require an "instantaneous" carry output, K_{i-1}.

$B_i(t)$	$A_i'(t)$	$K_i(t)$	$K_{i-1}(t)$	$B_i(t+T)$
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1
all possible inputs			$B + \bar{A}$	

We want in the B-register at time t + T not $B + \bar{A}$, but rather $B + \bar{A} + 2^{-4} - 2$, which we have shown above to be equal to B - A. (cf. end-around carry)

To get from $B + \bar{A}$ to $B + \bar{A} + 2^{-4} - 2$: we add the 2^{-4} by requiring that K_4, the carry input to the first stage, be 1.

We subtract the 2 by failing to provide a B_{-1} flip-flop.

Then a realization of the table which has the two additional properties just listed will subtract the contents of A from the contents of B and leave the result in the B-register T seconds after the subtract command pulse, P_6E (note that according to the table on p. 5 the (B) - (A) → B command goes out when the F-counter is in the configuration 110 (P_6) and a clock pulse (E) occurs).

The input to B_i is to be pulsed whenever we wish $B_i(t + T)$ to be the complement of $B_i(t)$ and a subtract command pulse occurs. Comparing the first and last columns of the table we get

$$b = (B_i'A_iK_i + B_i'A_i'K_i' + B_iA_iK_i + B_iA_i'K_i')P_6E$$

If this seems wrong, notice that the table lists values of A', not values of A, and that $A_i' = 0$ means: $A_i = 1$.

This expression can be greatly simplified by factoring A_iK_i from the first and third terms, and $A_i'K_i'$ from the other two:

$$\boxed{b = (A_iK_i + A_i'K_i')P_6E = (A_i' \oplus K_i)P_6E} \qquad (i = 0,\ldots,4)$$

which fact might have been seen directly from the table.

The carry output from this stage is

$$K_{i-1} = (B_i'A_i'K_i + B_iA_iK_i + B_iA_i'K_i' + B_iA_i'K_i)P_6E$$

$$K_{i-1} = \left[A_i'K_i + B_i(A_i' \oplus K)\right] P_6E \qquad i = 1,2,3,4;\ K_4 = 1$$

$$= A_i'K_iP_6E + B_ib$$

3.32 Shifting

The B-register is used in shifting as well as subtracting. Then before we adopt the above equation for the input to the B-register we must add terms which take account of the additional inputs required for shifting.

Referring to the diagram on p. 5 we see that the function (B) - (A) $\Rightarrow$B, which occurs in configuration 110 of the F-counter, is not the only one that involves B.

In configuration 011 (P_3) of the F-counter two other functions are performed: (B) $\Rightarrow$M[C] and (B_i) $\Rightarrow$ $B_{i\ 1}$, it being understood that the number which appears in M[C] is (B) before the shift. (See remarks on "simultaneous" interchanges between registers in Section 2.1.) In the shift a zero is introduced in the left-most digit position, B_0.

The transfer to memory, (B) $\Rightarrow$ M[C], does not concern the input to the B-register: it will be taken account of in the input equations to memory. Then the input equations for b will be complete when we add expressions to realize (B_i) $\Rightarrow$ B_{i+1} (i = 0,1,2,3); 0 $\Rightarrow$ B_0. This is most easily done using the clear and set inputs to the B flip-flops, although it might have been accomplished with only the complement input, with suitable gating. Then we consider that we have the WWI type of FF with three inputs, labeled as shown; the complete equations for the B-register are then

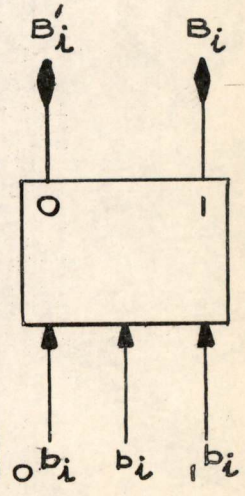
$$b_i = (A_i' \oplus K_i)P_6E \qquad (i = 0,1,2,3,4) \quad \text{(difference digit)}$$

$$K_{i-1} = A_i'K_iP_6E + B_ib \qquad (i = 1,2,3,4) \quad \text{(carry)}$$

$$K_4 = 1 \quad \text{("end-around carry")}$$

$$\left.\begin{aligned} {}_1b_i &= B_{i-1}P_3E \\ {}_0b_i &= B_{i-1}'P_3E \end{aligned}\right\} \qquad (i = 1,2,3,4) \quad \text{(shift right)}$$

$${}_0b_0 = P_3E \qquad \text{(destructive shift)}$$



That is, when the F-counter is in configuration P_3, the clock pulse E acts as a command for the contents of register no. i-1 to appear in register no. i, and for a zero to appear in register B_0. It is assumed, once again, that there is a delay inherent in the inputs to our FF's such that this shifting will occur <u>after</u> the transfer to memory has taken place.

3.33 <u>The A-Register</u> (See diagram, p. 25)

To complete the discussion of the arithmetic element we write the equations of the inputs to the A-register. Transfers into A occur on configurations 100, 010, and 111 (P_4, P_2 and P_7) of the operation counter (see p. 5). Again it will be simplest to use the <u>set</u> and <u>clear</u> inputs rather than the <u>complement</u> input:

$$\left.\begin{aligned} {}_1a_i &= \left[C_i P_4 + M_{hi}(P_2 + P_7)\right] E \\ {}_0a_i &= \left[C_i' P_4 + M_{hi}'(P_2 + P_7)\right] E \end{aligned}\right\} \quad i = 2,3,4$$

$$\left.\begin{aligned} {}_1a_i &= M_{hi}(P_2 + P_7) E \\ {}_0a_i &= M_{hi}'(P_2 + P_7) E \end{aligned}\right\} \quad i = 0,1$$

$$\Big\} \quad h = (C)$$

It is necessary to treat i = 0,1 separately since there are no C_0 or C_1 flip-flops. Now we must explicitly indicate how h = (C) is to be implemented.

Let us assign names to the configurations of the C-register in analogy to the P_i as names for the configurations of the F-register:

$$\Gamma_0 = C_2' C_3' C_4'$$
$$\Gamma_1 = C_2' C_3' C_4$$
$$\Gamma_2 = C_2' C_3 C_4'$$
$$\Gamma_3 = C_2' C_3 C_4$$
$$\Gamma_4 = C_2 C_3' C_4'$$
$$\Gamma_5 = C_2 C_3' C_4$$
$$\Gamma_6 = C_2 C_3 C_4'$$
$$\Gamma_7 = C_2 C_3 C_4$$

Now the C-register will be in one and only one of these configurations at any time and hence for any fixed i

$$\sum_{h=0}^{7} M_{hi} \Gamma_h = M_{0i} \Gamma_0 + \ldots + M_{7i} \Gamma_7$$

will be 0 or 1 accordingly as the i^{th} digit of the word in memory which corresponds to (C) is 0 or 1.

This implements the condition h = (C) which we added verbally to the equation above.

Then the final equations for the A-register are

$$\left.\begin{aligned} {}_1a_i &= \left[C_i P_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi}\ \Gamma_h\right)\right]E \\ {}_0a_i &= \left[C_i' P_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi}'\ \Gamma_h\right)\right]E \end{aligned}\right\} \quad i = 2,3,4$$

$$\left.\begin{aligned} {}_1a_i &= (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi}\ \Gamma_h\right)E \\ {}_0a_i &= (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi}'\ \Gamma_h\right)E \end{aligned}\right\} \quad i = 0,1$$

$$h = 0,\ldots,7$$

3.34 The C-Register

The inputs to the C-register are affected in configurations 100, 010, 011, 001 (in case $B_0 = 0$) and 101 of the operation counter, and in case G = 0, regardless of the state of F (see p. 5). Only three different things go on:

(1) (Ad A) = C (on P_4, P_2, P_3 and P_1B_0')

(2) 0 = C (on G')

(3) (C) + 1 = C (on P_5)

The first two functions are easily handled as above, using the set and clear inputs:

$$\left.\begin{aligned} {}_1c_i &= (P_1B_0' + P_2 + P_3 + P_4)A_i E \\ {}_0c_i &= \left[(P_1B_0' + P_2 + P_3 + P_4)A_i' + G'\right]E \end{aligned}\right\} \quad i = 2,3,4$$

The "G'" term in the equation for ${}_0c_i$ provides that when the computer is started up, the first memory address to be referred to will be 000.

The add one function can be implemented by a simple counter using the complement inputs. The cycle for the counter is:

C_2	C_3	C_4
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1
0	0	0

Apparently the stages should be complemented as follows:

$$c_2 = (C_2'C_3C_4 + C_2C_3C_4)P_5E$$

$$= C_3C_4P_5E$$

$$c_3 = C_4P_5E$$

$$c_4 = P_5E$$

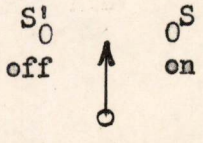
$$\boxed{\begin{aligned} c_2 &= C_3c_3 \\ c_3 &= C_4c_4 \\ c_4 &= P_5E \end{aligned}}$$

3.35 The Start Flip-Flop, G

On the halt command, on the time pulse before the operation (F) counter reaches configuration 000 the G flip-flop is cleared and therefore the F-counter sticks on 000 (see p. 5). The machine is to be started again by use of a switch, S, which sets G and allows the next clock pulse to step the operation counter into configuration 111. The machine can be stopped not only on the halt command, but at any time by setting S to off.

$${}_0g = (P_4A_0'A_1' + S')E$$

$${}_1g = SE$$



Shown in rest position. Can be locked in off. Contact in on state is momentary.

3.36 The Memory

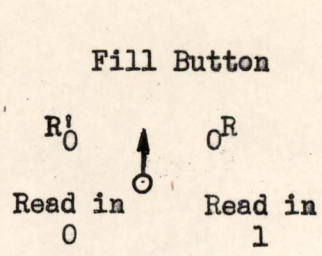
Read-in to the memory is to be accomplished in one of two ways

(1) Program: in the 011 configuration of the operation counter (B) $\Rightarrow$ M[C].

(2) Fill Buttons: the individual flip-flops can be filled "by hand"; this method is used for reading in programs.

The "fill buttons" are switches which in their rest positions excite neither ${}_0m$ nor ${}_1m$.

$$\left.\begin{aligned} {}_1m_{wd} &= (R_{wd} + B_d \lceil_w P_3)E \\ {}_0m_{wd} &= (R'_{wd} + B'_d \lceil_w P_3)E \end{aligned}\right\} \quad \begin{aligned} w &= 0,\ldots,7 \\ d &= 0,\ldots,5 \end{aligned}$$



Shown in rest position. Contact is momentary in both positions.

We now have a set of equations which completely describe the logic of our machine. They are summarized on p. 21 and p. 22.

4.0 GENERAL REMARKS

There are three general stages in the development of the logic of any particular computing machine. In order of decreasing generality of the decisions involved, they are:

(1) Planning. (Given the purpose of the machine, what operations shall be included? What word-length? What speed per operation?.....)

(2) Combinatorial Decisions. (What general arrangement of devices will best implement the plan: what shall be the cycle of the operation counter? What order code? What basic functions (commands)?.....)

(3) Design. (What configuration of memory-elements, gates, mixers, etc. best realizes the results of the combinatorial decisions?)

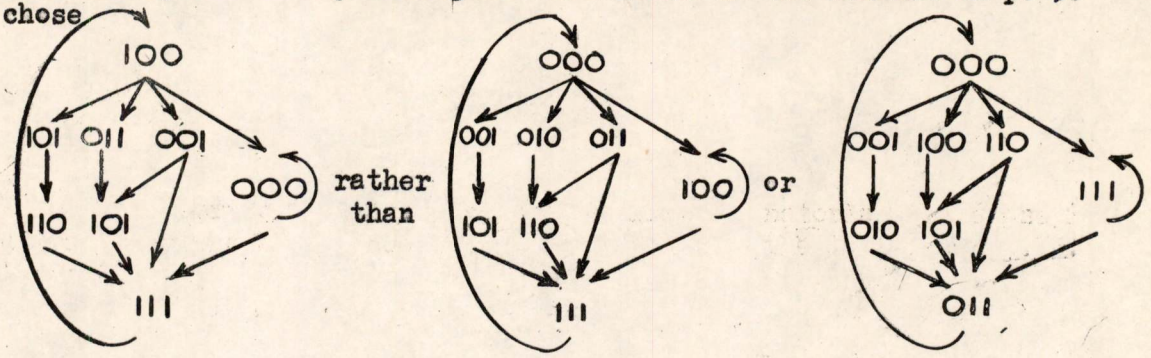
In the case of the present "sample" computer these steps went somewhat as follows:

(1) Planning: The purpose of the machine was to serve as an example of a method of design. Therefore it should have a small memory, short word length and few operations; but none of these should be so short, small or few as to make the design problem trivial. The results of the planning were presented, without discussion, on the first three pages.

(2) Combinatorial Decisions: These are summarized by the table on p. 5. There the cycle of the operation-counter is shown and the groups of basic commands indicated. Since the operations are few and simple, no thought at all was devoted to the coding of them (00 =halt, 01 = cp, etc.)

(3) Design: Throughout the design we referred to the table on p. 5 which summarized the combinatorial decisions. The design is almost mechanical, once the table has been drawn.

The sort of problem involved in the combinatorial decisions is illustrated by the binary coding of the 8 boxes in the lattice on p. 5. We chose



because we thought that of all the 8! ways of assigning numbers from 0 through 7 to the 8 points in the lattice, the one selected resulted in the simplest design. Of course we did not draw up all 8! lattices and from them write 8! sets of equations for the machine, examine them and choose the one that contained the fewest components. Such a program might be carried out by a machine like WWI, but not by human beings. Rather we considered e.g. that it would be desirable to have as little difference as possible between the numbers assigned to the points in the second level of the diagram, the outcomes of the four-way choice after 100. Accordingly, in a simple-minded way, we decided to use the four numbers beginning with "0" for those points. Similarly in the binary decision after 001 we chose "101" and "111" as the numbers of the next configurations because they differ from each other only in a single digit.

To systematize such combinatorial decisions, three developments would be helpful:

(1) A mathematical theory of the desirability of such choices as the two mentioned above (which were made on an intuitive basis).

(2) Assignment of numerical values to different components and their configurations (the unit might be dollars, or speed, reliability or some combination of those). Any such assignment should be in general terms so that the parameters might be changed as the characteristics of available components change.

(3) Programs which would allow a high-speed computer to survey large numbers of possibilities such as the 8! assignments of numbers to the points in the lattice above. It is possible to write down a fully mechanical procedure for simplifying equations and even for going from a diagram like the one on p. 5 to a set of equations, so that this development is <u>possible</u> - and perhaps practical.

5.0 SUMMARY

5.1 Complete Set of Equations for the Machine

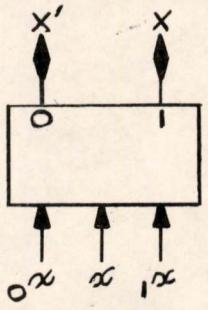
5.11 Abbreviations

$P_i = F_1^{*}F_2^{*}F_3^{*}$ where the n^{th} "*" is either a prime or a blank, depending on whether the n^{th} digit in the binary representation of i is a "0" or a "1". e.g., $P_5 = P_{101} = F_1F_2'F_3$.

$\Gamma_i = C^{*}C^{*}C^{*}$. E.g., $\Gamma_1 = \Gamma_{001} = C_1'C_2'C_3$

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n$$

5.12 Notation for Flip-Flops



5.13 The Operation Counter (F-Register)

$$f_1 = (P_0G + P_1 + P_2 + P_3 + P_4)E$$

$$f_2 = (P_0G + P_1B_0 + P_3 + P_4A_0 + P_5 + P_7)E$$

$$f_3 = (P_0G + P_4A_1 + P_6 + P_7)E$$

$${}_0f_i = G'E \quad (i = 1,2,3)$$

5.14 The Accumulator (B-Register)

$$b_i = (A_i' \oplus K_i)P_6E \quad (i = 0,1,2,3,4)$$

$$\left.\begin{aligned} {}_1b_i &= B_{i-1}P_3E \\ {}_0b_i &= B_{i-1}'P_3E \end{aligned}\right\} \quad (i = 1,2,3,4)$$

$${}_0b_0 = P_3E$$

$$K_{i-1} = A_i'K_iP_6E + B_ib \quad (i = 1,2,3,4); \quad K_4 = 1$$

5.15 The A-Register

$$\left.\begin{aligned} {}_1a_i &= \left[C_iP_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi}\lceil_h\right)\right]E \\ {}_0a_i &= \left[C_i'P_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi}'\lceil_h\right)\right]E \end{aligned}\right\} \quad \begin{aligned} &i = 2,3,4; \\ &h = 0,\ldots,7 \end{aligned}$$

Same for i = 0,1 except that 1st term is missing.

5.16 The Memory-Selection Register (C)

$$\left.\begin{aligned} {}_1c_i &= XA_iE \\ {}_0c_i &= (XA_i' + G')E \end{aligned}\right\} \quad i = 2,3,4$$

where $X = P_1B_0' + P_2 + P_3 + P_4$.

$$c_i = C_{i+1} \quad (i = 2,3)$$

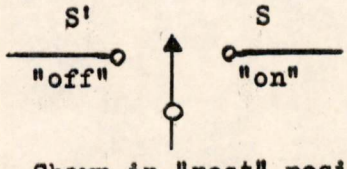
$$c_4 = P_5E$$

5.17 The Start-Stop Flip-Flop (G)

$${}_0g = (P_4A_0'A_1' + S')E$$

$${}_1g = SE$$

S is a double-throw switch which can be locked in the off position, makes momentary contact in the on position, and has a rest position in which neither contact is energized.

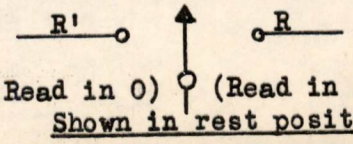


Shown in "rest" position.

5.18 The Memory Registers (M)

$$\left.\begin{aligned} {}_1m_{wd} &= (R_{wd} + B_d\lceil_w P_3)E \\ {}_0m_{wd} &= (R_{wd}' + B_d'\lceil_w P_3)E \end{aligned}\right\} \quad \begin{aligned} &w = 0,\ldots,7; \\ &d = 0,\ldots,5 \end{aligned}$$

The R's are double-throw switches ("fill buttons") which make momentary contact on either side and have a rest position in which both sides are open.



Shown in rest position

5.2 The Flip-Flops

These are assumed to have an inherent delay between receipt of an input pulse and the resulting change in the output. Speaking loosely, this delay permits reading new information into a register "at the same time" that the old information is being read out, and permits "simultaneous" transfers of information between two registers. (See Section 2.1.)

5.3 Operating Instructions

1. Turn on power and allow tubes to warm up. During this time the machine will be cycling through some meaningless program if the G flip-flop happens to start out holding a "1".

2. Lock switch S in the off position. This sets G to 0, sets the operation counter to 000, and reads 000 into C. Read program and data into memory with "push-buttons" (R).

3. Release S (nothing happens while S is in neutral position), push it to the on position and release. This sets G to 1 and the operation counter steps into 111, the beginning of program timing. Since C holds 000, the first word to be taken out of memory and transferred into the A-register will come from the first memory register (at location 000). The program should be stored with this fact in mind.

4. The program should end with a halt command, 00xyz, where x y z may be any number. This freezes the contents of all FF's except for G and the F and C registers, all of which are cleared. A new program may now be read into storage via the R switches and the machine may be re-started by pushing the S switch over to on.

5. Read-out is via neon bulbs attached to the memory FF's (not shown above).

5.4 Block Diagrams

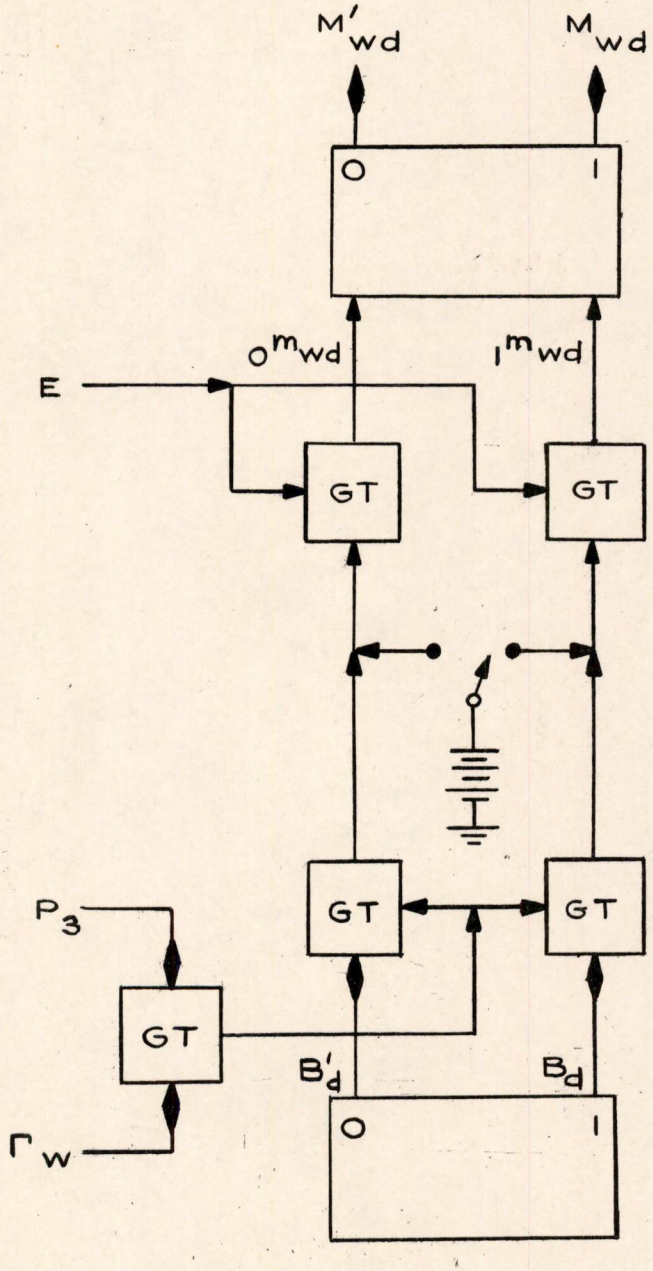
People who are unfamiliar with the algebraic notation may find it helpful to translate some of the equations into block diagrams. A little bit of this will go a long way toward promoting familiarity with the algebraic symbolism; with a little practice it will be found that as much information can be read directly from a few lines of equations as from a dense page of block diagrams.

Block diagrams for a few of the registers follow.

Note that the equations for a register show the inputs to that register. To find outputs of register X, look through all the equations for occurrences of the letter "X". (In the case of the F and C registers, look for "P" and "⌈" as well, since these abbreviate products of "F" s and "C"s.)

Memory

$$\left.\begin{aligned} {}_1m_{wd} &= (R_{wd} + B_d \lceil_w P_3)E \\ {}_0m_{wd} &= (R'_{wd} + B'_d \lceil_w P_3)E \end{aligned}\right\} \quad \begin{aligned} w &= 0,\ldots,7 \\ d &= 0,\ldots,5 \end{aligned}$$



Identical picture

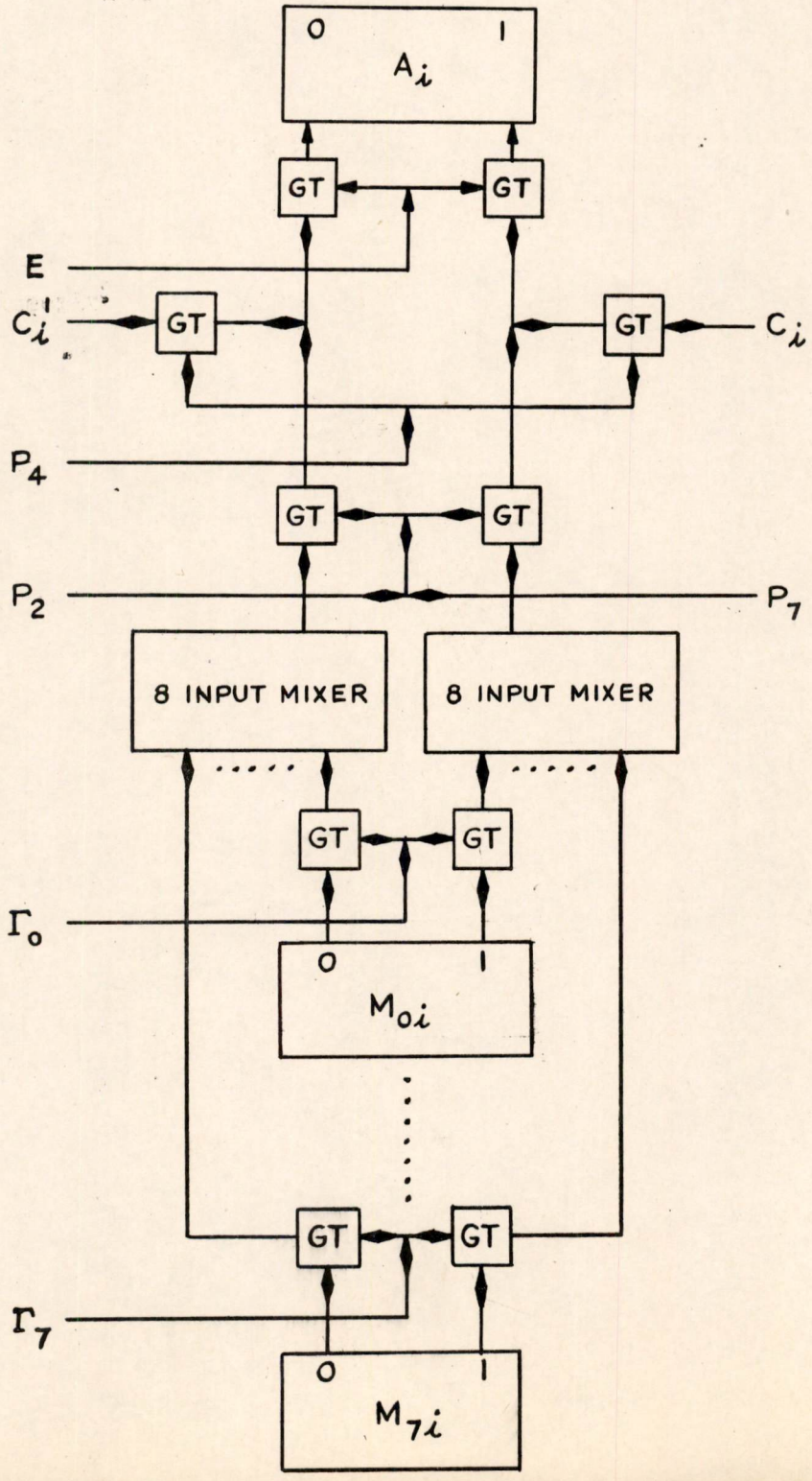
for each memory cell.

In case a magnetic core memory is used with an I/2 selection system, the read-in gates are simply points where pairs of insulated wires cross inside the cores.

The A-Register (last three digits)

$$1^{a_i} = \left[C_i P_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M_{hi} \lceil_h \right) \right] E \qquad i = 2,3,4;$$

$$0^{a_i} = \left[C_i' P_4 + (P_2 + P_7)\left(\sum_{h=0}^{7} M'_{hi} \lceil_h \right) \right] E \qquad h = 0,\ldots,7$$



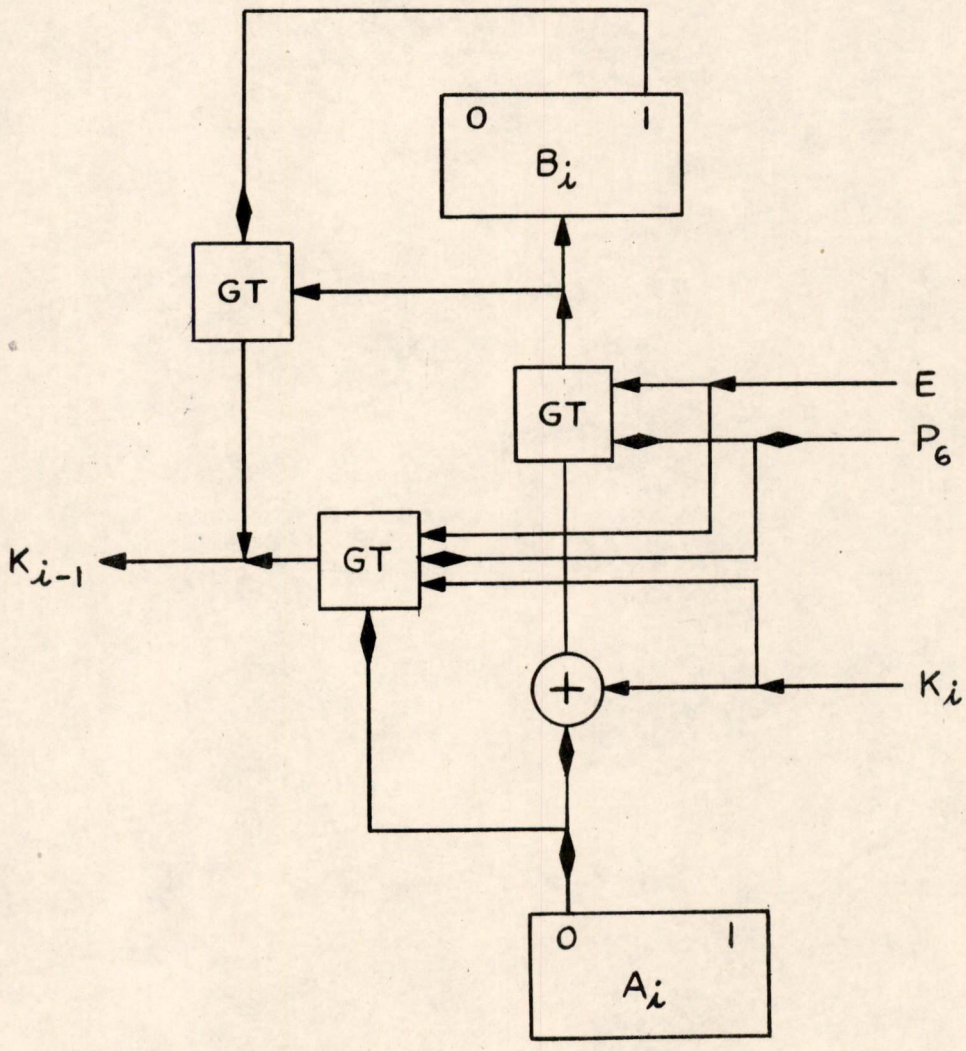
This memory system is not intended as a practical proposal. The following two modifications are desirable: (1) In reading out of memory, clear A on the preceding clock pulse and then read into only the set sides of the A FF's. This eliminates the left hand half of the memory read-out gates and the left hand 8 input mixer. (2) Use magnetic cores as memory cells. Then the remaining read-out gates are inherent in the cores, and the 8 input mixer is the sensing winding.

The Accumulator (complement inputs)

The set and clear inputs are used only for shifting. The reader can complete this diagram by adding them.

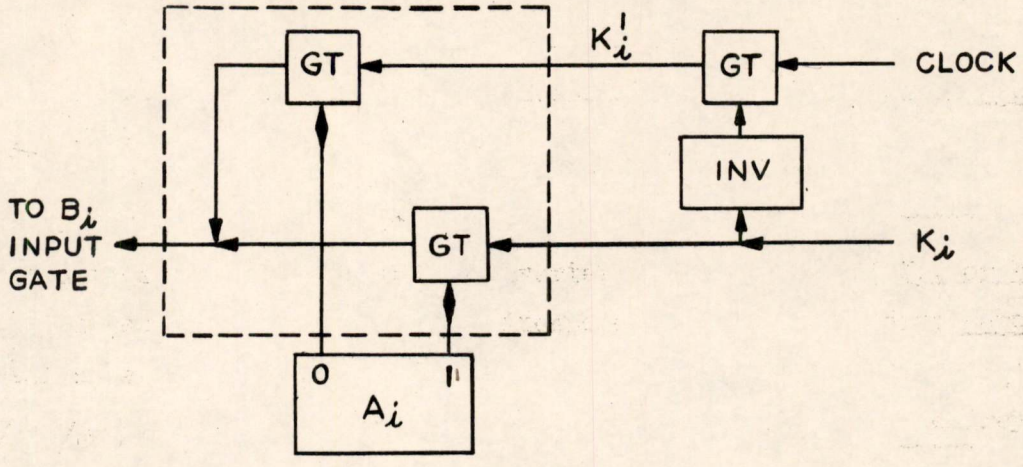
$$b_i = (A_i' \oplus K_i)P_6E \quad (i = 0,\ldots,4)$$

$$K_{i-1} = A_i'K_iP_6E + B_ib$$



+ represents a black box which gives a pulse output whenever one but not both of the inputs are present. See next page.

The black box for partial sum (⊕) is an oversimplification. Probably the best way to implement it with customary electronic components is to use not only A_i and K_i but also their complements:



This requires either providing an inverter to get K_i' out of K_i, or building up K_{i-1}' as an output from stage #i independent of K_i. (A carry zero line as well as a carry one line.)

SIGNED Richard C. Jeffrey
Richard C. Jeffrey

Irving S. Reed
Irving S. Reed

APPROVED
N. H. Taylor

RCJ/ISR/cp

cc: G. R. Briggs
D. R. Brown
D. A. Buck
H.R.J. Grosch
W. A. Hosier
J. Jacobs
W. Linvill
R. P. Mayer
J. A. O'Brien
W. Ogden
K. H. Olsen
W. N. Papian

BIBLIOGRAPHY

1. Jeffrey, R. C., Reed, I. S., "The Use of Boolean Algebra in Logical Design", Engineering Note E-458-1, Digital Computer Laboratory, M.I.T., (April, 1952).

2. Reed, I. S., Division II Status Report, Project Lincoln, (April, 1952).

3. Reed, I. S., "Some Mathematical Remarks on the Boolean Machine", Project Lincoln Technical Report No. 2, (December 19, 1951).

R. Best

Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

SUBJECT: INTRODUCTION TO THE THEORY OF SEMICONDUCTORS I,
SOME REMARKS ON QUANTUM MECHANICS

To: Transistor Group

From: Donald J. Eckl

Date: June 24, 1952

Abstract: This is the first of several papers dealing with the present theory of metals and semiconductors. The purpose of these papers is to familiarize the members of the transistor group with the basic physical theory behind the operation of semiconductor diodes and transistors. In so far as possible, the discussions will be kept on an intermediate and descriptive level. The results and implications of the more complicated mathematical theories will be emphasized rather than their step-by-step development.

The satisfactory discussion of modern solid state theory requires a certain background knowledge of quantum and statistical mechanics. These subjects are highly mathematical so only certain essential ideas and results, which will be required later, will be mentioned here. The present paper presents a brief discussion of some of the non-classical results of quantum or wave mechanics together with a short history of the development of the theory.

1.0 The Nuclear Atom

The original postulate of the nuclear atom was made by Rutherford in 1911. His atomic model, developed to explain the scattering of α-particles (helium nuclei) by heavier atoms, consisted of a positive nucleus surrounded by external electrons, making the complete system electrically neutral. It was necessary to assume that the electrons were travelling in orbits about the nucleus since otherwise the electrical forces of attraction would pull them into the positive nucleus.

However, on the basis of classical electromagnetic theory such a system should radiate electromagnetic energy with a resultant reduction of mechanical energy. As the mechanical energy decreases the orbital electrons should spiral into the nucleus, and the frequency of the emitter radiation should increase. Unfortunately these predictions do not agree with the observed facts. Atoms do not continuously emit electromagnetic radiation. They radiate only after having received excess energy and the radiation consists only of certain discrete frequencies or spectral lines. This enigma could not be solved by classical physics.

1.1 Planck's Quantum Hypothesis

The actual beginning of the quantum theory goes back to 1901 and Planck's attempt to derive theoretically his empirical expression for the spectral energy distribution in blackbody radiation. A blackbody radiator is one which will completely absorb incident radiation of every wave length. Planck's revolutionary assumption was that an oscillator (physical system) has a discrete set of energy values, $E = nh\nu$, where ν is the fundamental frequency of the oscillator, h is a constant (Planck's constant), and n is an integer: 1,2,3,4,---. If we define E/ν (energy x time) as a quantity called "action," then h becomes the "quantum" of action. Whenever a quantity may have only certain discrete values it is said to be quantized and the smallest permitted value is called a quantum. The integer n is referred to as a quantum number.

The quantum theory was extended to light in 1905 when Einstein made use of "photons" of energy $h\nu$ to explain photoelectric emission of electrons from metals.

1.2 The Bohr Theory

In 1913 Bohr combined the Rutherford nuclear atom with Planck's quantum hypothesis to develop an atomic model which explained the experimental results obtained with the spectrum of the hydrogen atom. This atom is the simplest of all, containing but a single orbital electron and a positive nucleus called a proton. Bohr contended that the electron rotates about the nucleus only in orbits fixed by certain quantum conditions, but that, contrary to classical electrodynamics, it does not radiate energy while in such "stationary" states. Moreover, each of these quantum orbits corresponds to a definite energy state E of the atom, the larger the orbit the greater the energy. The allowed orbits are governed by the quantum condition that the product of the circumference of the orbit and the momentum of the electron (dimensions of action) is equal to an integral multiple of Planck's constant h.

$$2\pi rmv = nh \qquad n = 1,2,3\text{---}. \tag{1.0}$$

Jumps by the electron from higher energy orbits to those of lower energy are accompanied by radiation of photons of light of energy $E_1 - E_2 = h\nu_{12}$, where the E_i are the orbit energies and ν_{12} is the frequency of the radiation produced by the transition from orbit 1 to orbit 2.

To deal with atoms of a more complicated nature than hydrogen, it is necessary to introduce quantum numbers other than n. These are j, the total angular momentum quantum number, and ℓ, the orbital angular momentum quantum number. In the Bohr theory, enlarged by Sommerfeld, ℓ is associated with the eccentricity of the elliptical electron orbit. In addition to orbital angular momentum the electron also has a spin giving a resultant total angular momentum associated with the quantum number j. By specifying the three quantum numbers, n, j, ℓ, it is possible to specify the energy state of the atom.

The foregoing is a crude, but useful, picture of the atom, and while it does give quantitative and qualitative results in agreement with experiment in certain cases, it breaks down under more complicated problems, such as the

helium atom for example. Moreover, it employs a classical mechanical model and yet rejects certain features of classical theory which it finds contrary to experiment. Finally, various physical quantities are quantized without any satisfactory reason why this should be done. The theory is at best a hybrid, yet its success shows that it must contain considerable truth.

1.3 The Schroedinger Theory

It has already been mentioned that Einstein showed how light, which had been supposed to be transmitted by electromagnetic waves, acted as if composed of particles (photons) or quanta when liberating electrons from a photo surface. In 1927, Davisson and Germer were able to show that electrons could exhibit wave characteristics and that the "wave length" was that predicted by de Broglie in 1925.

$$\lambda = \frac{h}{mv} \qquad (1.1)$$

Extending this idea of waves associated with particles, Schroedinger developed wave mechanics in 1926. This is a precise mathematical method of dealing with atomic systems. Another method called matrix mechanics was independently developed by Heisenberg in 1925. However, since the two methods may be shown to be formally equivalent, and since the Schroedinger scheme deals with the more familiar differential equations rather than matrices, we will discuss quantum mechanics from the Schroedinger view point.

Although it could be demonstrated that waves may be associated with particles, the physical significance of these oscillations was not immediately apparent. In particular, therewere the unsolved problems of the nature of the oscillation producing these waves. Schroedinger assumed that the oscillating quantity in these "matter waves" was a function Ψ, which he made no attempt to define, and showed that in general Ψ must satisfy the equation,

$$\frac{h^2}{8\pi^2 m}\nabla^2\Psi - V\Psi = \frac{-ih}{2\pi}\frac{\partial\Psi}{\partial t} \qquad (1.2)$$

where $$\nabla^2 = \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}, \quad i = \sqrt{-1}$$

and V = potential function for the system.

If the time factor is removed, we obtain Schroedinger's amplitude equation which is used in dealing with stationary states of atoms.

$$\nabla^2\psi + \frac{8\pi^2 m}{h^2}(V - E)\psi = 0. \qquad (1.3)$$

Useful solutions to this equation require that ψ remain finite even at infinity. This condition can be met only for certain discrete values of E, the proper values or eigin values corresponding to the proper functions ψ. These certain values of E are the allowed values of the energy corresponding to the stationary or quantum states of the particle in the given force field represented by the potential V.

Previously it was mentioned that the nature of the quantity ψ was not specified. It is possible to give a physical significance only to the quantity

$$\psi_E \psi_E^* d\tau \qquad (1.4)$$

where ψ^* is the complex conjugate of ψ. This quantity represents the probability for the state with energy E of finding the particle under question in the volume element $d\tau$. This indicates one of the fundamental differences between quantum and classical mechanics. Classical mechanics deals with known positions and known velocities. Quantum mechanics deals with most probable positions.

This attitude is expressed more exactly in the Heisenberg Uncertainty Principle which briefly states (in one form) that the more accurately the position of a particle is defined the less accurate is our knowledge of its momentum or velocity. Mathematically,

$$\Delta x \, \Delta p \geq \frac{h}{4\pi} = 5.3 \times 10^{-28} \text{ erg-sec.} \qquad (1.5)$$

also

$$\Delta E \, \Delta t \geq \frac{h}{4\pi} \qquad (1.6)$$

Note that both of these quantities have the dimensions of "action" and that therefore these expressions are equivalent to stating that the minimum quantum of action is $h/4\pi$. In other words, this inability to simultaneously determine momentum and position with perfect accuracy is not a limitation in our experimental apparatus but rather a fundamental property of nature. In the same way, more accurate equipment would not make it possible to measure any smaller quantity of charge than e, the electronic charge.

It is not too surprising that classical mechanics fails when applied to atomic systems. Classical mechanics explains the motion of dynamical systems in the light of our ordinary human experience. However, in the two extremes which are beyond the scope and imagination of the human observer classical mechanics breaks down. One of these extremes is, of course, the atomic system where distances are too small for human observation and measurement. The other is the universe where distances are extremely large (light years) and relativistic mechanics becomes necessary. It might be noted that some of the predictions of relativity (e.g. change of mass with velocity) have been shown to occur in the case of atomic particles in cyclotrons. Both of these extreme systems, however, must pass over into classical mechanics when applied to ordinary classical systems. Their unusual predictions will appear only in the special types of systems to which they are adapted. Where quantum and classical mechanics are shown to predict different results for the same type of system in the next section it should be understood that each result is correct for a system of the proper physical dimensions.

1.4 Applications of Quantum Mechanics to Dynamical Systems

At this time it will be helpful to consider the effects of quantum mechanics when applied to some rather common dynamical systems. First, take

the case of the simple harmonic oscillator--e.g., a pendulum or weighted spring. The equation of motion is

$$m\ddot{x} = -kx$$

and the solution

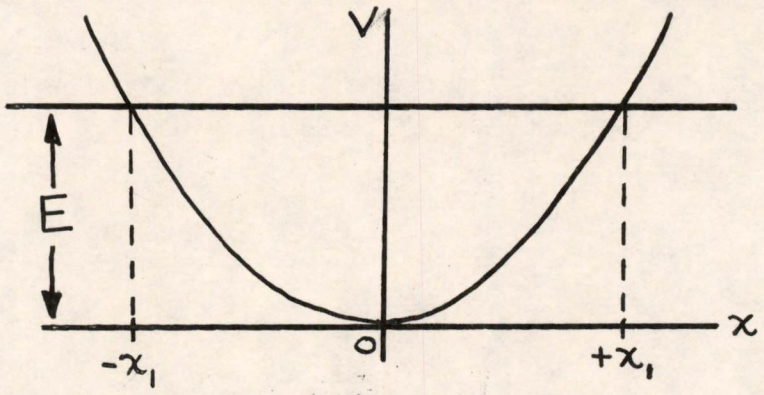
$$x = A \begin{matrix}\sin \\ \cos\end{matrix} \omega t$$

The force -kx is the negative gradient of the potential function.

$$\therefore V(x) = 1/2\, kx^2$$

A graph of this potential function is given in Figure 1.

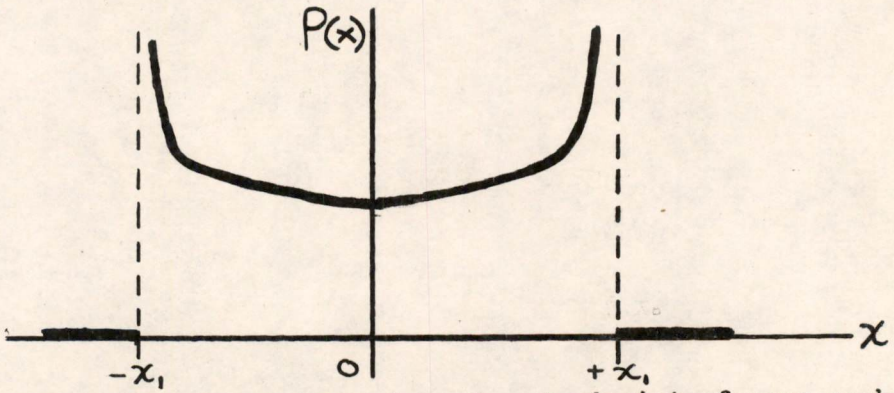
Fig. 1



Suppose E is the total energy of the system. Then at $\pm x_1$ all the energy is in the form of P.E. and the velocity v = 0. At x = 0, V = 0 and $E = 1/2\, mv^2$. $\pm x_1$ are called the classical limits of x of the oscillator since a precise measurement of position cannot yield a value of $|x| > |x_1|$.

A distribution of x-values for this system is shown in Figure 2. P(x) is the probability of occurrence of a given value of x.

Fig. 2



Now consider the same problem from the standpoint of wave mechanics. The Schroedinger equation must be solved with $V = 1/2\, kx^2$. Therefore we have, (c.f. Equation 4)

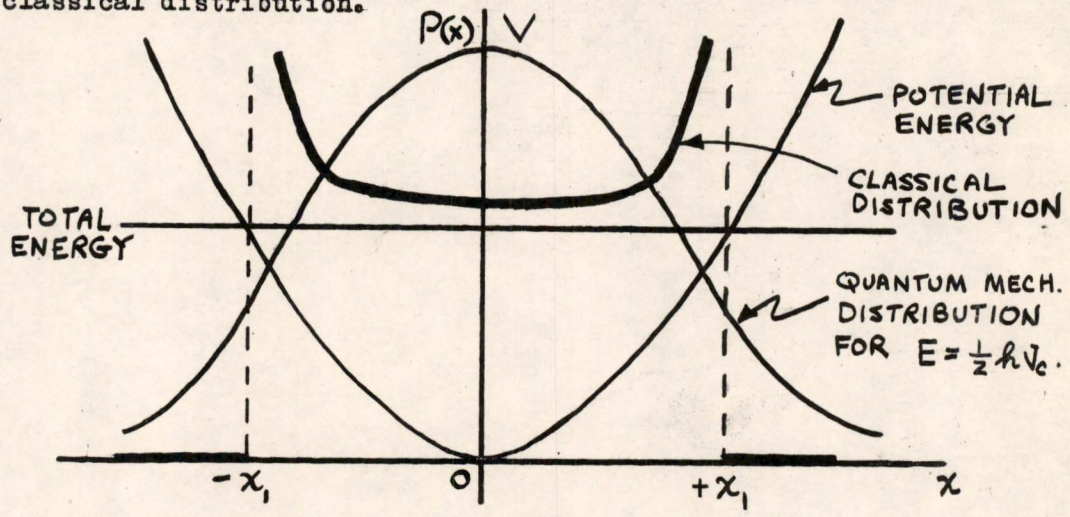
$$\frac{d^2\psi}{dx^2} + \frac{8\pi^2 m}{h^2}(E - 1/2\, kx^2)\psi = 0$$

The first result which differs from classical theory is the requirement that E have only certain discrete values given by

$$E = (n + 1/2)\, h\, \nu_c \qquad (1.7)$$

where $n = 0,1,2,3\ldots$ and ν_c = the classical frequency. Thus the energy is quantized. However, if the x distribution is calculated, a really striking result is obtained, as shown below in Figure 3, in comparison with the classical distribution.

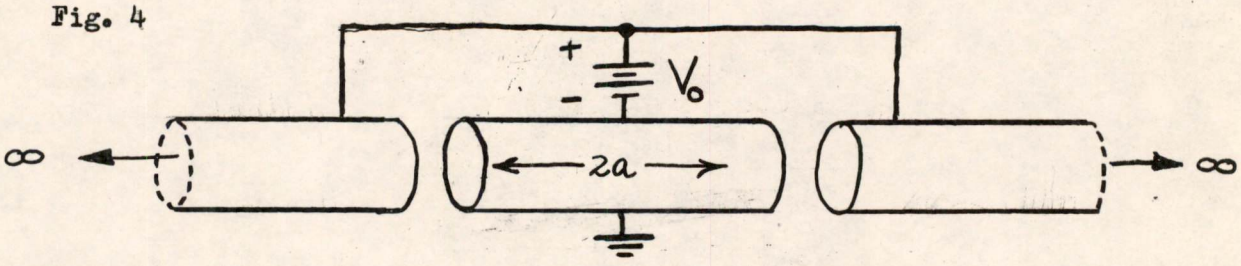
Fig. 3



Note that the quantum mechanical distribution in x does not vanish outside the classical limits of x. Therefore according to the quantum theory a particle may be found in regions that are not accessible to it on a classical basis.

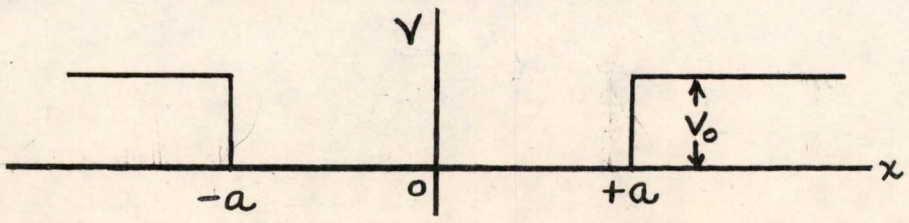
Next consider the case of the potential well in the classical system. This might be represented by the following diagram:

Fig. 4



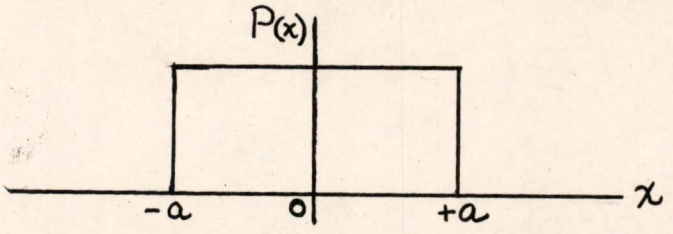
The potential field acting on a unit positive charge moving along the axis of the cylinders has the form of a well as shown below:

Fig. 5



The distribution in x for this system is as follows for $E < V_o$.

Fig. 6

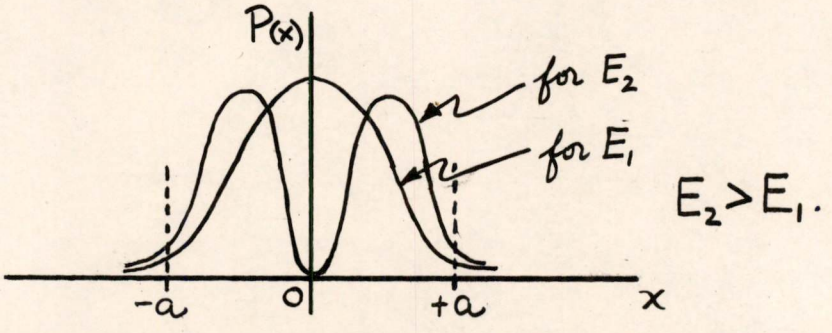


The positive charge has equal probability of being found at any point inside the center cylinder but cannot get outside.

When quantum mechanics is applied to a system of this type, it is again found that only certain values of E are allowed. Those who are interested in the mathematics of the solution are referred to Rojansky (see list of references at end). We will simply state results which are of particular interest.

The quantum mechanical distribution in x for two of the quantized values of E is shown below in Figure 7:

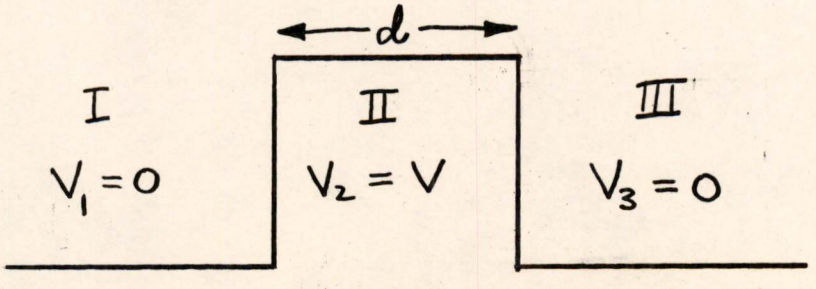
Fig. 7



Thus there is a small but finite possibility of finding the particle outside of the well which increases with the value of E.

An important problem which will arise in dealing with semiconductors is the case of the potential barrier. A classical particle cannot pass such a barrier unless it has an energy greater than required to surmount the barrier. Figure 8 shows a typical case.

Fig. 8



From the quantum mechanical viewpoint, however, the results are quite different. In this case the Schroedinger ψ function must satisfy the equations:

$$\frac{d^2\psi_1}{dx^2} + \frac{8\pi^2 m}{h^2} E\psi_1 = 0$$

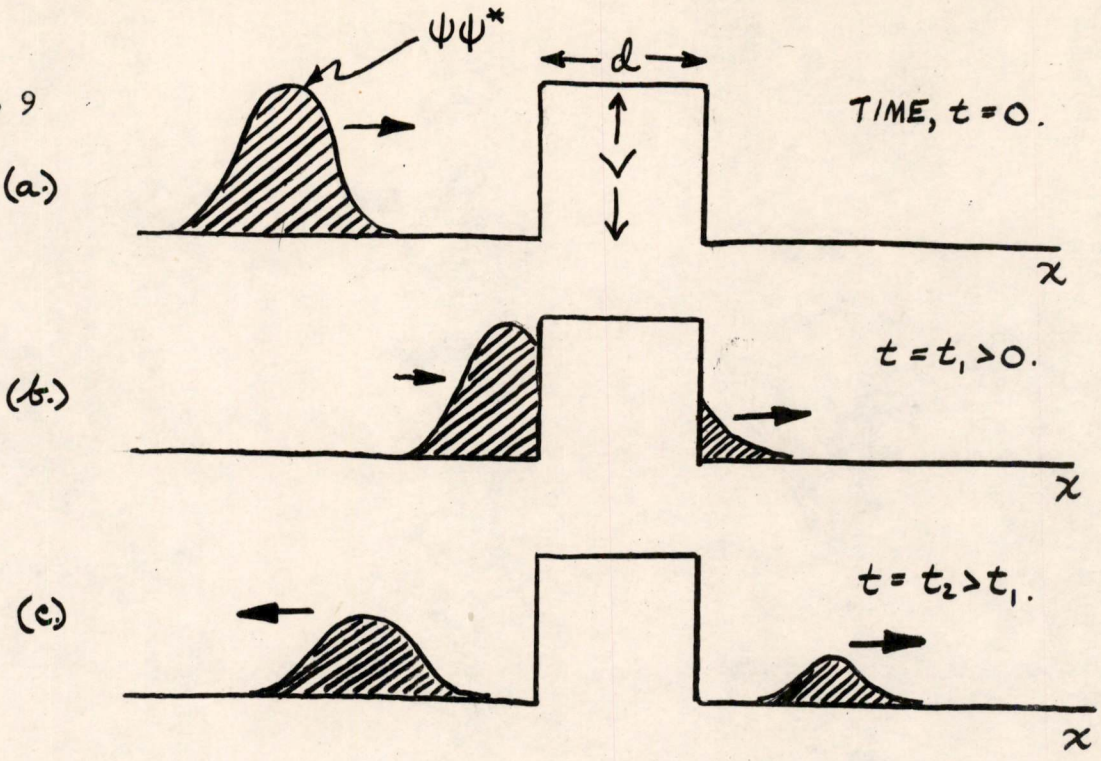
$$\frac{d^2\psi_2}{dx^2} + \frac{8\pi^2 m}{h^2} (E - V)\psi_2 = 0$$

$$\frac{d^2\psi_3}{dx^2} + \frac{8\pi^2 m}{h^2} E\psi_3 = 0$$

Also ψ and $d\psi/dx$ must be continuous at the two surfaces. In this case again we will discuss the important results and omit the rigorous mathematical treatment which may be found in several references. Now if we insist that E is certainly less than V, we find that this is quantum-mechanically incompatible with the requirement that the particle is certainly in region I because of the mathematical nature of the ψ functions. However, we can construct states for which E is certainly less than V and for which at t = 0 the particle is as likely to be in the left region as we please.

Therefore, at t = 0 the x-distribution, $\psi\psi^*$, may be represented by a "probability packet" approaching the barrier as shown in Figure 9 (a). This packet represents the probability of finding the particle at a given x position. Since this probability is essentially zero outside the packet, the particle must be within the packet and moving toward the barrier. As time passes, the situation at (b) occurs. Calculations of $\psi\psi^*$ to the left and right of the barrier show the incident packet as before, but also a transmitted packet so that there is now a possibility of finding the particle in the right-hand region. If $\psi\psi^*$ is calculated at a still later time, both transmitted and reflected packets appear. The relative sizes of these packets show that it is most probable that the particle has struck the barrier and rebounded.

Fig. 9



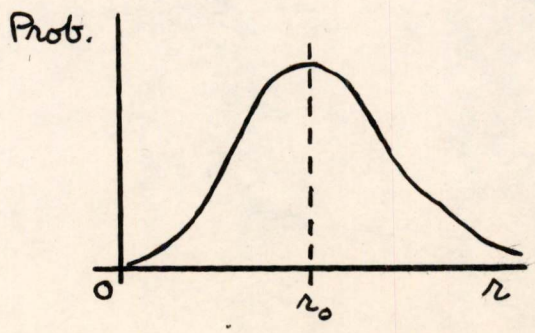
But the appearance of a transmitted packet implies that the probability of finding the particle in the right-hand region has grown from a negiligible value to a sizable one. A calculation shows that

$$\psi_3\psi_3^* = \frac{E\,(V - E)}{V^2}\, e^{-\frac{4\pi d}{h}\sqrt{2m\,(V - E)}} \qquad (1.8)$$

Thus the probability of finding the particle beyond the barrier (region III) is increased as d is made smaller and E approaches V. This penetration of a potential barrier is often referred to as the "tunnel effect."

Finally, a few remarks are in order on the effect of quantum mechanics on the Bohr model of the atom. For the hydrogen atom an evaluation of $\psi\psi^* d\tau$ for the electron in the ground (lowest energy) state gives a curve as shown below in Figure 10. The probability distributions for higher energy states have a different form and will not be discussed here.

Fig. 10



The electron is no longer in a fixed orbit for a given energy as required by classical mechanics but may be found at a wide range of distances. However, its most probable distance from the nucleus is r_o, the radius of the Bohr orbit in the ground state. The stationary energy states, postulated by Bohr to agree with observations, arise naturally from the Schroedinger equation as the proper values E_n, when the required potential function is used. The quantum numbers n, ℓ, m^* also arise naturally in the solution of the equation. The orbital angular momentum is now $\sqrt{\ell(\ell+1)}\, h/2\pi$ instead of $\ell h/2\pi$. The ψ functions, and, therefore, the state of the atom can be specified by the quantum numbers n, ℓ, m.

The important fact is that, despite (or rather because of) the rather startling features of wave mechanics mentioned above, it is able to predict results for complicated atomic systems which are in agreement with observation, a task at which both classical mechanics and early (Bohr) quantum theory were unsuccessful. Moreover, the undesirable features of the Bohr theory--the arbitrary introduction of quantum numbers and the semi-classical semi-quantum approach--have been eliminated. Wave mechanics is a completely self-consistent mathematical theory and perhaps some of our difficulties in understanding it arise from insisting on having "models" to explain its meaning.

1.5 The Pauli Exclusion Principle

There is one further requirement on the electrons in an atomic system which must be mentioned here because of its importance in the theory of metals. This is the Pauli Exclusion Principle which states that in a given system there can never be two or more electrons in the same state, i.e., with the same quantum numbers: n, j, ℓ, m. This is a fact derived from experience.

Signed Donald J. Eckl
Donald J. Eckl

Approved John F Jacobs
John F. Jacobs

Approved NHT
Norman H. Taylor

DJE/jk

* m is a magnetic quantum number.

References*

1. "Introduction to Modern Physics," F. K. Richtmyer and E. H. Kennard (McGraw-Hill)

2. "Elementary Wave Mechanics," W. Heitler (Oxford)

3. "Atomic Physics," W. Finkelnburg (McGraw-Hill)

4. "Quantum Theory of Matter," J. C. Slater (McGraw-Hill)

5. "Introductory Quantum Mechanics," V. B. Rojansky (Prentice-Hall)

* The first three of these references are mainly descriptive. Slater gives the necessary mathematical treatment as well as an extensive discussion of the results from the physical standpoint. Rojansky's book is more rigorous and mathematical and deals with the problems from the operator point of view.

Olsen

Engineering Note E-464 Page 1 of 1

Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

SUBJECT: A SQUARENESS RATIO FOR COINCIDENT-CURRENT MEMORY CORES

To: Group 63 Staff

From: David R. Brown

Date: July 17, 1952

Abstract: A quantitative index of hysteresis-loop squareness is defined. It can be determined directly from the hysteresis loop or by pulse test.

This engineering note defines a squareness ratio for coincident-current memory cores. The following discussion considers only a two-to-one selection ratio; however, the definition can easily be extended to other selection ratios.

Figure 1 illustrates the measurement of the squareness ratio from the hysteresis loop. The squareness ratio, R_s, has upper and lower bounds of plus and minus one; in the ideal case, it would be plus one. The larger the squareness ratio, the better the core. Note that R_s will be a function of I_m, the peak ampere turns. If R_s is plotted as a function of I_m, a single maximum will be found. The maximum squareness ratio, R_{sm}, and the I_m at which R_{sm} occurs, are usually figures of greatest significance.

Figure 2 illustrates the measurement of squareness ratio by a pulse test. Here again, R_s will be a function of i_m. The ratio R_{sm}, and its i_m, should be determined. The rise time of the flux, τ_ϕ, can also be determined.

Squareness ratios for ferrites have been determined from 60-cycle hysteresis loops. The best specimen of MF-1118 has an R_{sm} of 0.80. The optimum I_m as determined by measurement of R_{sm}, agrees with the optimum operating point as determined by the pulse measurements for ferrites. The validity of measurements of squareness ratios from 60-cycle hysteresis loops for metallic cores has not yet been established.

The measurement of squareness ratio by pulse test may offer some advantage over the presently used pulse tests. This advantage would be due to the fact that the determination would be dependent upon fewer variables. The measurement is not intended as a substitute for the present pulse test, but should be investigated and the results correlated with the results of our present pulse tests.

DRB/jk
Attached:
SA-51967 SA-51968

Signed David R. Brown
David R. Brown

cc: Jay Forrester, H. Fahnestock, R. Everett, N. Taylor, W. Papian, K. Olsen, G. Briggs, W. Ogden, R. von Buelow

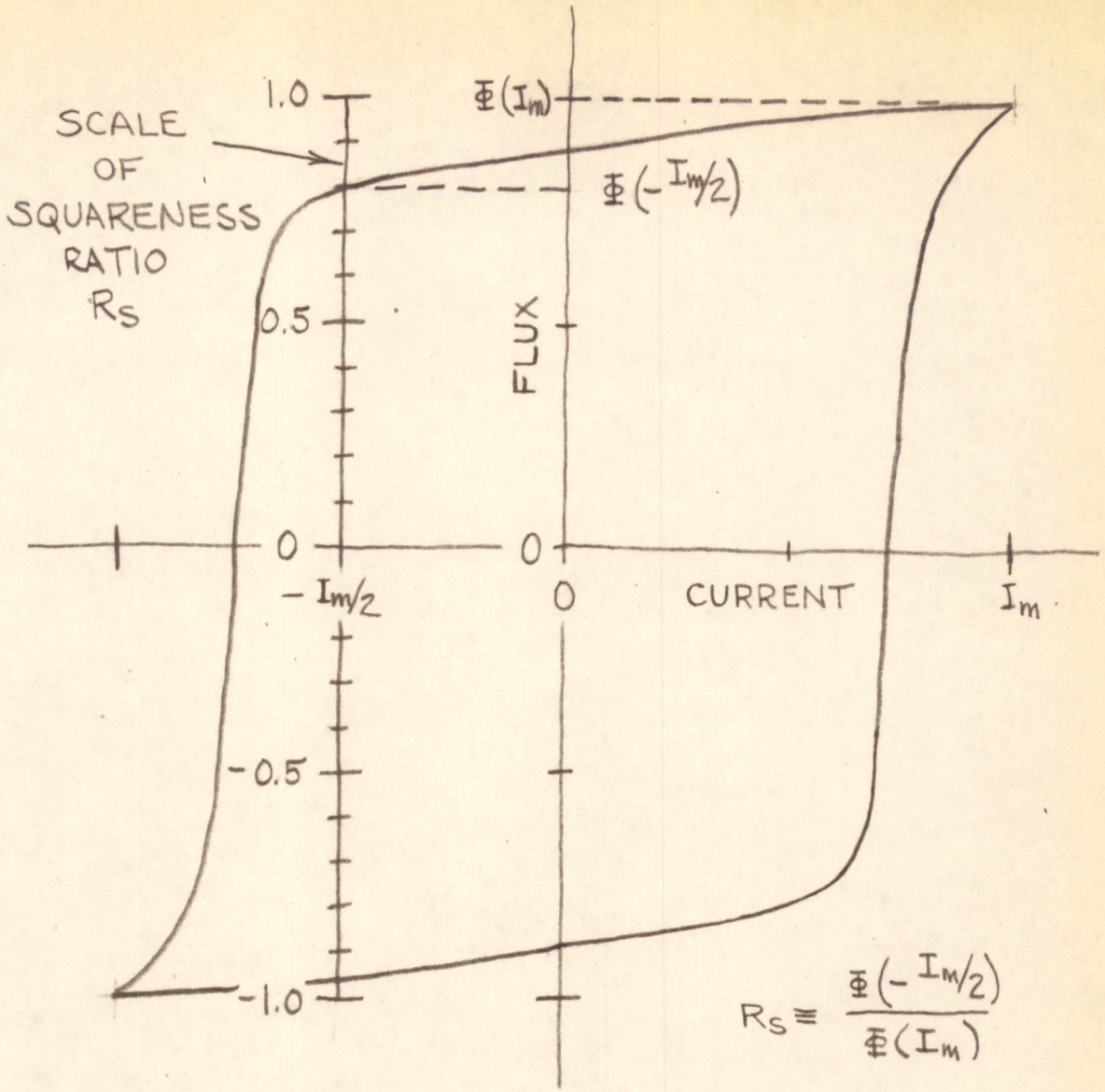


FIG. 1

MEASUREMENT OF SQUARENESS RATIO FROM HYSTERESIS LOOP

SA-51967

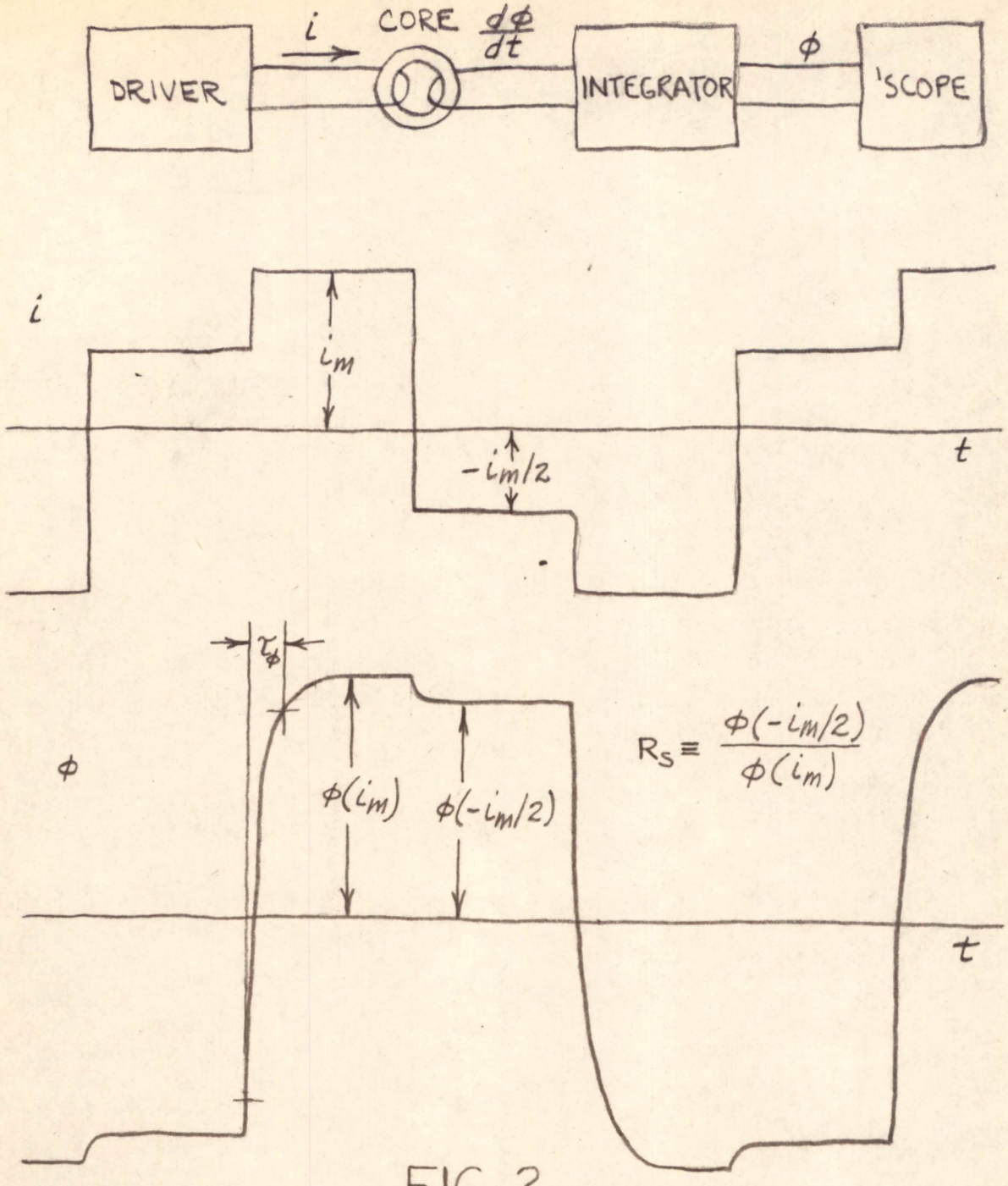


FIG. 2

MEASUREMENT OF SQUARENESS RATIO FROM PULSE TEST

SA-51968

R. Best

Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

SUBJECT: INTRODUCTION TO THE THEORY OF SEMICONDUCTORS II, A BRIEF DISCUSSION OF STATISTICAL MECHANICS.

To: Transistor Group

From: Donald J. Eckl

Date: July 28, 1952

Abstract: This is the second note dealing with the physics of semiconductors. The previous discussion, E-463, dealt with the microscopic viewpoint of the system under consideration -- i.e. - with the individual particle. The detailed solution of this problem requires the use of quantum mechanics. It is now necessary to look at the problem from the macroscopic viewpoint which involves statistical mechanics. The detailed mathematical solution of the problem has not been carried through, but sufficient mathematics has been included to indicate the method.

2.0 Introduction

The problem before us is one of interpreting and, as far as possible, predicting the properties of macroscopic (large scale) physical systems in terms of the properties of the microscopic systems of which they are composed. The classical method of dealing with a large system of, say, 10^{23} molecules in a gas would be to solve the 3×10^{23} equations describing the motion of the system. Aside from the practical considerations involved in obtaining such a solution, there remains the difficulty that the results would be expressed in terms of distances and velocities of individual particles, rather than physical quantities such as temperature, pressure, etc., which can readily be measured for the system as a whole. Obviously, a more useful approach must be employed and this is the role of statistical mechanics. In this method all considerations are based on the average behavior of particles and individual deviations from the mean are neglected. A considerable amount of information about the system can be obtained once the most probable distribution of energies for the particles in the system has been determined. Such a distribution would be

$$n_1 (E_1) ; \; n_2 (E_2) ; \; n_3 (E_3) ; \; \text{-----}$$

where there are n_1 particles with energy E_1, n_2 with energy E_2, etc. We are primarily interested in methods of determining this distribution.

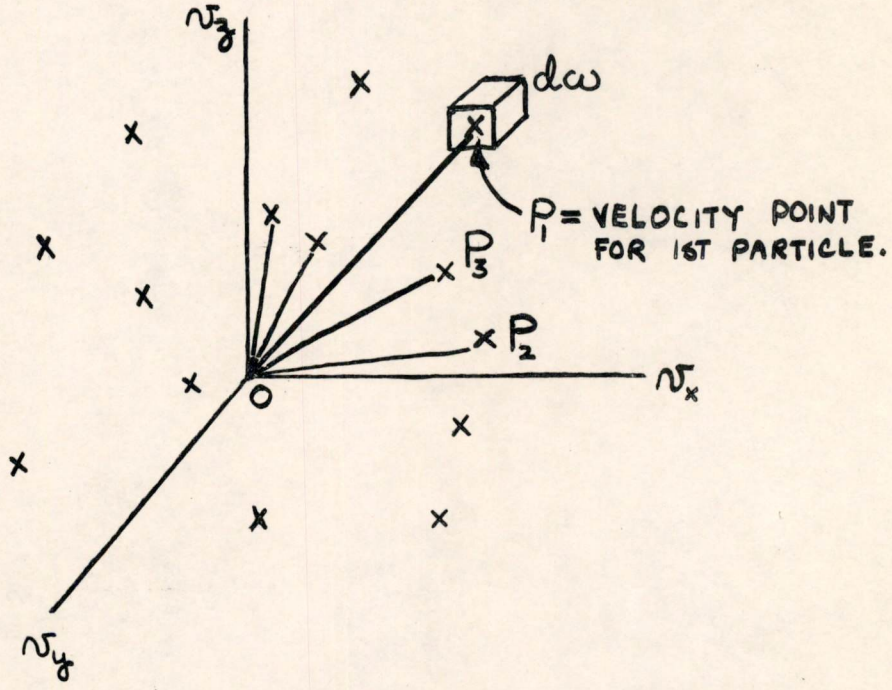
The degree to which it is possible to distinguish one particle from another (postulated) has a marked effect on the result and is the principal difference in the types of statistics described here. Sufficient introductory mathematics will be given in each case to show the general method of dealing with the problem and the main differences between the various systems.

2.1 Maxwell-Boltzmann Classical Statistics.

The statistical method is largely independent of all detailed mechanics. The particles under consideration are considered to be "point molecules". It is assumed that there are no interactions between molecules except in collisions -- i.e. the molecules form a perfect gas. The kinetic energy of the point molecules is supposed to make up the entire energy of the system.

The state of the system is represented by means of a velocity diagram as shown in Figure 1. Each particle in the system is represented

Fig. 1

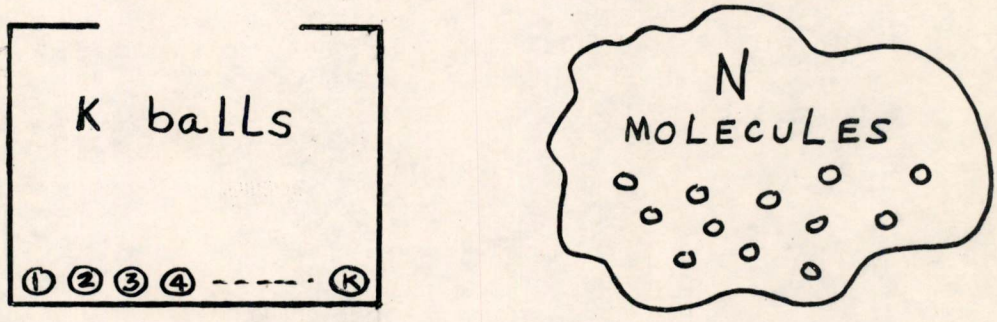


by a point whose coordinates are the components of the velocity of the particle in question. The velocity space is divided up into equal volume elements $d\omega = dv_x \cdot dv_y \cdot dv_z$. The problem then is to find the number of velocity points contained in each volume element -- i.e., the number of velocity components in the ranges v_x, $v_x + dv_x$; v_y, $v_y + dv_y$; v_z, $v_z + dv_z$. The volume elements are numbered 1,2,3,---,K. Assume that the system has N molecules.

We wish to devise a scheme whereby the molecules taken indi-

vidually have the same probability of being assigned to any one of the volume elements $d\omega_i$. One way of doing this is to set up a so-called Boltzmann lottery.[1] Suppose we have a box with k balls numbered 1,2,3, ---,k, one for each velocity volume element $d\omega$ as shown in Figure 2.

Fig. 2



A ball is removed from the box, its number inspected, and a molecule put into a cell of that number. The ball is returned and the process repeated. This is done until all N molecules are assigned to cells. A distribution is obtained with n_1 molecules in $d\omega_1$, n_2 in $d\omega_2$, etc.

$$\text{i.e. } (n_1, n_2, n_3, \text{-------}, n_k).$$

The whole process is repeated and we obtain other distributions:

$$(n_1', n_2', n_3', \text{-------}, n_k').$$

$$(n_1'', n_2'', n_3'', \text{-------}, n_k'').$$

$$\text{- - - - - - - - - - - - - - - - - - etc.}$$

In each case, of course, $\sum n_i = N$. Now the distributions are examined and that which occurs most frequently is selected. This is called the most probable distribution and we postulate that this is the actual distribution in the gas.

It is now necessary to find the probability of the occurrence of any given distribution $(n_1, n_2, n_3, \text{---}, n_k)$. To do this we will consider first a more simple problem, that of determining the probability of getting n_1 "heads" in N tosses of a coin.

Suppose N = 100. Then,

Number of cells: k = 2 i.e. -- heads, tails.

Possible Distributions: n_1 (heads) 49; 45; 50; etc.
n_2 (tails) 51; 55; 50; etc.

For two tosses of the coin there are four equally probable cases, i.e. --

H,T T,H

H,H T,T

For three tosses there are $2^3 = 8$ equally probable cases and for N tosses there are 2^N equally probable cases. Now the number of times n_1 heads may be expected is the number of combinations of N things taken n_1 at a time, or $C^N_{n_1} = \frac{N!}{n_1!(N-n_1)!} = \frac{N!}{n_1!\, n_2!}$. Therefore the probability of a distribution (n_1, n_2) is

$$P(n_1, n_2) = \frac{1}{K^N} \cdot \frac{N!}{n_1!\, n_2!}. \qquad (2.0)$$

i.e. -- the probability of a single distribution times the number of ways the desired distribution may occur.

This may be expanded to the more general case of the distribution $(n_1, n_2, n_3, \text{---}, n_k)$ and the probability of such a distribution is

$$P(n_1, n_2, \text{---}, n_K) = \frac{1}{K^N} \cdot \frac{N!}{n_1!\, n_2! \text{----} n_K!}. \qquad (2.1)$$

Here N! is the total number of possible permutations of the N particles in the system. Of these $n_1!$ represent rearrangements of the n_1 particles in ω_1 and so do not give a new distribution. This is also true for $n_2!$ and so on. Such a probability is a maximum for $n_1 = n_2 = \text{---} = n_k = N/K$, a uniform distribution of velocities. However, so far we have used only the auxiliary requirement,

$$\sum n_i = N. \qquad i = 1, 2, \text{----}, K. \qquad (2.2)$$

It is also necessary to assume that the gas has a constant total energy E. We then assign an energy ε_1 to cell ω_1, ε_2 to ω_2, etc., and require that

$$\sum n_i \varepsilon_i = E. \qquad i = 1, 2, \text{----}, K. \qquad (2.3)$$

All distributions incompatible with this second condition must be rejected.

Suppose we take an example:

Let k = 2 cells,

Let the number of particles N = 3. Label these a,b,c.

There are $k^N = 2^3 = 8$ equally probable cases shown in Figure 3:

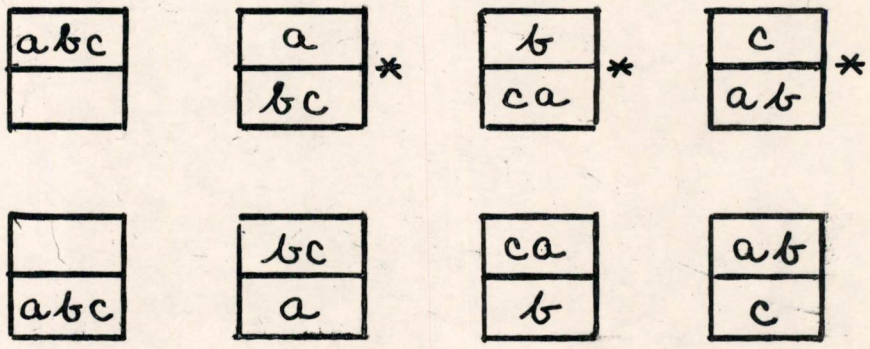


Fig. 3

Now the probability of the distribution $(n_1, n_2) = (2,1)$ is,

$$P(2,1) = \frac{1}{K^N} \frac{N!}{n_1! n_2!} = \frac{1}{8} \times \frac{3!}{2!1!} = \frac{3}{8}.$$

For the case $(n_1, n_2) = (3,0)$ the probability is,

$$P(3,0) = \frac{1}{8} \times \frac{3!}{3!0!} = \frac{1}{8}.$$

But if in addition we require that $\varepsilon_{TOP} = 0$, $\varepsilon_{BOTTOM} = 1$, and E = 2, only the starred distributions in Figure 3 are permitted. At this point, it should be observed that the molecules must be distinguishable for the enumeration of the distributions.

Returning now to the main problem, we must find the distribu-

tion for which

$$P = \text{CONST.} \times \frac{N!}{n_1!\, n_2! \cdots n_K!}.$$

is a maximum subject to the two auxiliary conditions (2.2) and (2.3). This problem is solved by the application to the above set of equations of Stirling's Rule and Lagrange's Method of Undetermined Multipliers. The result is

$$n_i = A e^{-\varepsilon_i/kT} \qquad (2.4)$$

where A = constant, k = Boltzmann's constant, and T = absolute temp. This is assumed to be the distribution in the gas and is called the Maxwell-Boltzmann distribution function.

With this distribution it is possible to calculate the number of molecules having an energy between E and E + dE. The result is

$$dN = \frac{2\pi N}{(\pi kT)^{3/2}} \sqrt{E}\, e^{-E/kT}\, dE. \qquad (2.5)$$

This gives the familiar curve shown in Figure 4 below for the Maxwell-Boltzmann distribution.

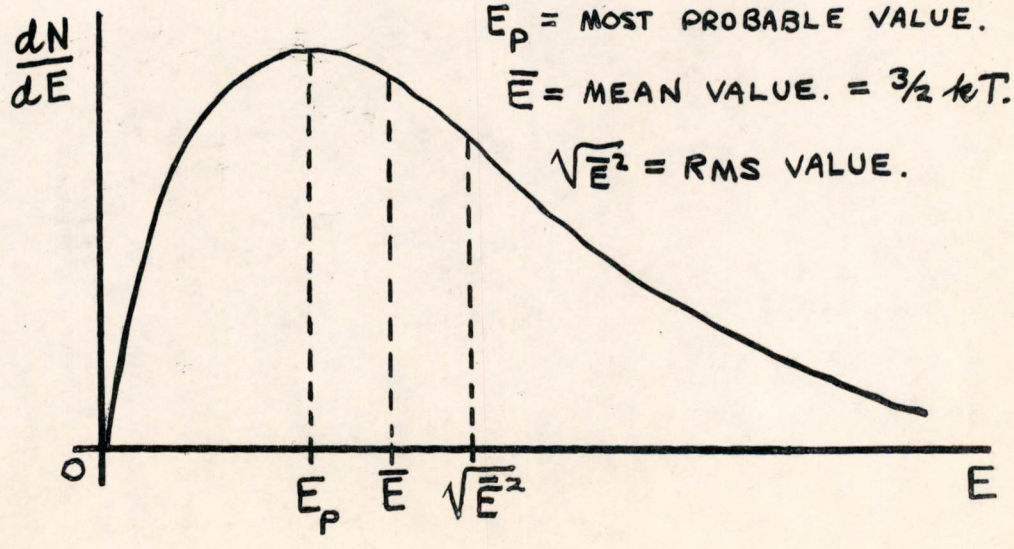


Fig. 4

The average energy $E = 3/2\ kT$. If this is entirely kinetic energy, then for a hydrogen molecule at 0°C the corresponding average velocity is 18×10^4 cm/sec or about 4000 miles per hour.

2.2 Bose-Einstein Statistics.

When classical statistics was applied to the problem of determining the spectral energy distribution in black-body radiation, the predicted result was not in agreement with experiment. As a result, a new quantum statistics was developed by Bose and Einstein based on complete indistinguishability of particles.

In classical statistics the velocity or momentum space is divided into cells $d\omega_i$. The new statistics assigns a volume h^3 (where h is Planck's constant) to the smallest volume element in accordance with the Heisenberg Uncertainty Principle. Therefore each of the previous volume elements is now subdivided into $d\omega_i/h^3 = g_i$ cells. $d\omega_i$ is usually referred to as a "sheet" and is taken as the spherical shell between p and p+ dp where $p = mv$, the momentum.

Suppose the individual cells of the sheet are numbered $z_1, z_2, z_3, \cdots, z_{g_i}$. In each sheet there are as before n_1 molecules. These must be distributed among the g_i cells of the sheet and the number of distinguishable arrangements determined. A typical arrangement would be, for example:

$$z_1\ (a_1a_2a_3),\ z_2\ (a_4),\ z_3\ (\ -\ ),\ z_4\ (a_5a_6),\ z_5\ (a_7a_8a_9a_{10})$$

All the different arrangements can be obtained as follows: Set down a "z" at the left side which can be done in g_i different ways. Then write down the $(g_i - 1 + n_i)$ remaining terms in any order. The total number of such arrangements is

$$g_i\ (g_i + n_i - 1)! \qquad (2.6)$$

Distributions obtained by permuting entire cells -- e.g., $z_1\ (a_5a_6)$ and $z_2\ (a_4)$ -- or by permuting particles within a given cell are not distinguishable. There are $g_i!\ n_i!$ such permutations. This requirement that $z_2\ (a_5a_6);\ z_4\ (a_4)$ and $z_2\ (a_4);\ z_4\ (a_5a_6)$ are not distinguishable is the essential departure from the previous case. The number of distinguishable arrangements in a particular sheet is

$$\frac{g_i\ (g_i + n_i - 1)!}{g_i!\ n_i!} \qquad (2.7)$$

But there are n_1 particles in the first sheet, n_2 in the second, etc., so that the total number of distinguishable arrangements for this distribution is

$$\prod_i \frac{(g_i + n_i - 1)!}{(g_i - 1)!\, n_i!} \qquad (2.8)$$

where the symbol $\prod_i$ indicates the product of all i terms. This is to be compared with

$$\left(\sum_i n_i\right)\left(\prod_i \frac{1}{n_i!}\right)$$

for the classical case. The probability of occurrence of a given distribution using this type of statistics is then

$$P = \text{Const.} \prod_i \frac{(g_i + n_i - 1)!}{(g_i - 1)!\, n_i!} \qquad (2.9)$$

As before the maximum value subject to the two auxiliary conditions (2.2) and (2.3) is required. Stirling's Equation and Lagrange's method lead to the Bose-Einstein Distribution Law:

$$n_i = \frac{g_i}{e^{\alpha + \frac{\varepsilon_i}{kT}} - 1} \qquad \alpha = \text{CONST.} \qquad (2.10)$$

For equal-sized sheets this may be written

$$n_i = \frac{B}{\frac{1}{A_0} e^{\varepsilon_i/kT} - 1} \qquad A_0 = e^{-\alpha} \qquad B = \text{CONSTANT.} \qquad (2.11)$$

This expression differs from the classical distribution (2.4) by the presence of the minus one in the denominator. The number of molecules with energies between E and E + dE is

$$dN = \frac{4\pi V (2m)^{3/2}}{h^3} \frac{\sqrt{E}\, dE}{e^{\alpha + (E/kT)} - 1} \qquad (2.12)$$

The quantity $A_o = e^{-\alpha}$ is called the degeneracy parameter. When α is large, and therefore A_o small, the one in the denominator can be neglected and the Bose-Einstein distribution reduces to the classical one. When A_o approaches unity deviations from the classical results occur and the gas is said to be degenerate.

It can be shown that for ordinary gasses degeneracy occurs only for extremely high pressures or low temperatures where the gas will no longer behave as an ideal gas in any case. Therefore, for ordinary gasses the quantum Bose-Einstein statistics shows no difference from classical Maxwell-Boltzmann statistics. However, when applied to black-body radiation the quantum statistics leads to Planck's law which agrees with experiment, whereas the classical method fails.

2.3 Fermi-Dirac Statistics.

This is a quantum statistics based on the Pauli Exclusion Principle. As a result of this principle there can be no more than one particle per cell in this case.

Suppose there are n_i particles in the i^{th} sheet distributed over the g_i cells of the sheet. Then n_i cells will be singly occupied and $g_i - n_i$ empty. This gives the following distribution:

	0 particles	1 particle
cells:	$z_2\ z_3\ z_5$ - - -	$z_1\ z_4\ z_6$ - - -

There are $g_i!$ such distributions corresponding to permutations of the g_i cells in this scheme. But there is no difference in the states obtained by permuting the g_i occupied or the $g_i - n_i$ empty cells. Therefore the probability of a given distribution with $n_1, n_2, \ldots, n_i$ cells occupied in the first, second, ----, i^{th} sheets is

$$P = \text{Const.} \prod_i \frac{g_i!}{n_i!\,(g_i - n_i)!} \qquad (2.13)$$

Again the same method of solution used in the two previous cases gives the Fermi-Dirac distribution law:

$$n_i = \frac{g_i}{e^{\alpha + (\varepsilon_i / kT)} + 1} \qquad (2.14)$$

or for equal numbers of cells per sheet

$$n_i = \frac{C}{\frac{1}{A_o} e^{\varepsilon_i/kT} + 1} \cdot \qquad C = \text{CONST.} \tag{2.15}$$

The plus one in the denominator distinguishes this distribution from the others. The number of particles with energies between E and E+ dE is

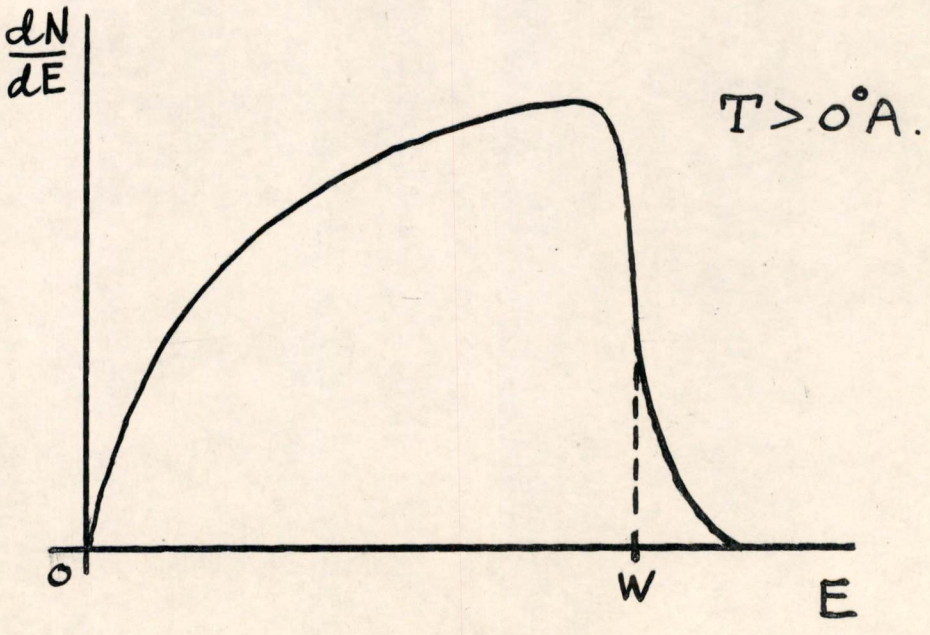
$$dN = \frac{4\pi V}{h^3}(2m)^{3/2} \frac{\sqrt{E}\, dE}{e^{\alpha + \frac{E}{kT}} + 1} \cdot \tag{2.16}$$

Here again when $A_o = e^{-\alpha}$ is extremely small the distribution passes into the classical case. When $A_o >> 1$ the gas is degenerate. It can be shown that for large A_o, $\alpha \approx -W/kT$, where W is a constant energy value, so that for the degenerate gas (2.16) becomes

$$dN = \frac{4\pi V}{h^3} \frac{(2m)^{3/2} \sqrt{E}\, dE}{e^{(E-W)/kT} + 1} \cdot \tag{2.17}$$

The form of this curve in general is shown below in Figure 5.

Fig. 5



A further discussion of the properties of the Fermi-Dirac distribution law will be given in the next section.

As a conclusion the following diagram will serve to illustrate the allowed distributions for two particles in two energy states in the three systems.

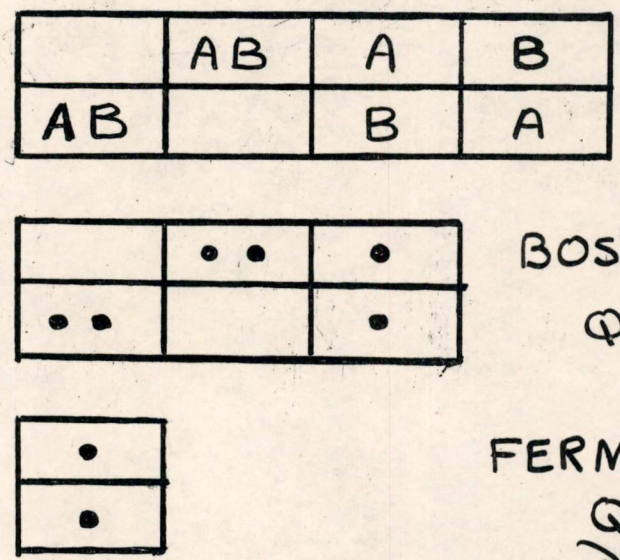


Fig. 6

Signed D. J. Eckl
Donald J. Eckl

References: "Atomic Physics", M.Born, (Hafner)

Approved John F Jacobs
John F. Jacobs

DJE/cs

Approved N H Taylor
Norman H. Taylor

Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

TITLE: THE MIRROR: A PROPOSED SIMPLIFIED SYMBOL FOR MAGNETIC CIRCUITS.

To: 6889 Engineers

From: R. P. Mayer

Date: August 14, 1952

Abstract: A coil can be wound on a magnetic core in one of only two directions. There are several dot-conventions which can be used to indicate which direction is desired. A simpler symbol, which is quicker to draw and easier to follow, is proposed: The "mirror symbol", which "reflects" current to find resultant flux.

The Mirror Symbol

Figure 1 shows a conventional coil symbol and the corresponding proposed mirror symbol. Note that the core is a single line with blocked ends. The coil is indicated by the short diagonal "mirror", which "reflects" applied current to show the direction of the flux it tends to produce. There are only two directions in which a coil can be wound, and they are both shown in Figure 1.

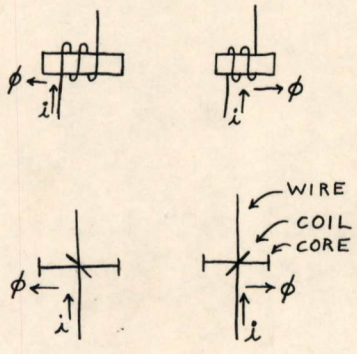
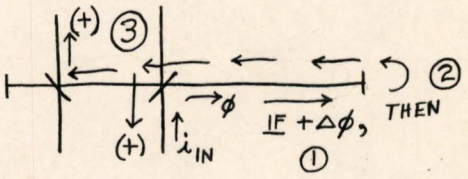


FIG. 1

Induced Voltages

Three simple steps (Figure 2) find the induced voltages: (1) Find the direction of positive flux change resulting from i_{IN} or other causes. (2) Go to the end of the core and bounce off. (3) Return back down core and reflect off all mirrors to find positive induced voltages.



"(+)" = INDUCED VOLTAGE, OR DIRECTION OF RESULTANT CURRENT.

FIG. 2

Applications

Figure 3 shows that a mirror in one place establishes coil direction, and that number of turns, etc., can be indicated (if desired) by any of several ways.

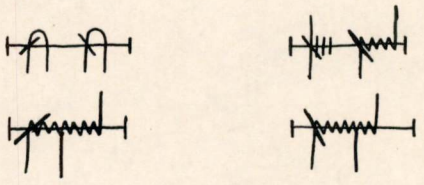


FIG. 3

Straight core symbols are strongly encouraged. But Figure 4 shows that a circular core symbol works almost as well. (Remember that you can find outputs only by reversing direction, even though the circle has no "end" to bounce off as in Figure 2).

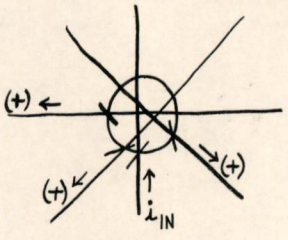


FIG. 4

Quadrature field windings can be shown as in Figure 5. The sensing pulse tends to decrease the main field, so the flux change is in the other direction.

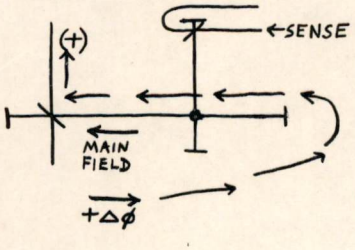


FIG. 5

Examples

The examples shown in Figures 6 and 7 show how a flip-flop and a stepping register work. These circuits were originally explained in M-1570 (Meeting of July 25, 1952). The "Method of Operation" shows the flow of information from input current, to ϕ, to reflected $\Delta\phi$, to output current, etc.

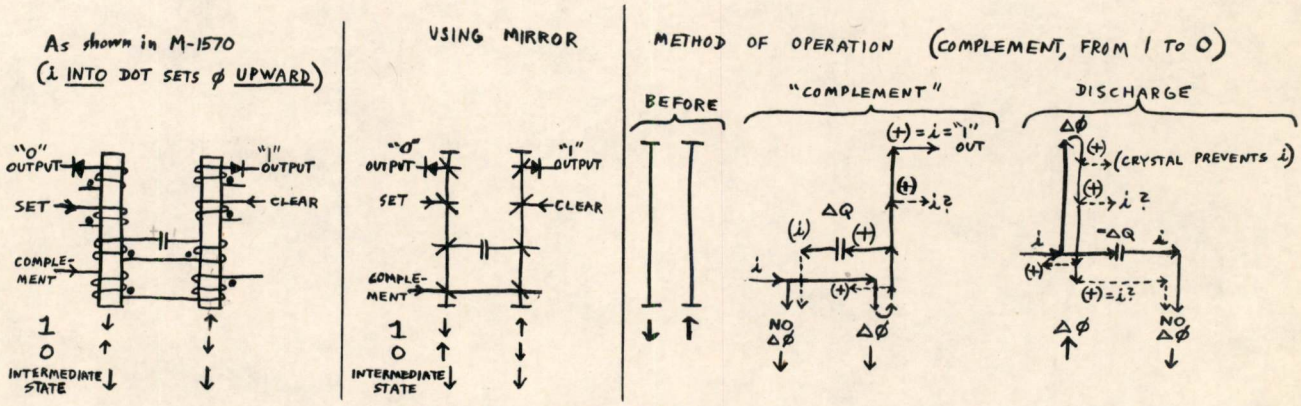


FIG. 6 - FLIP-FLOP

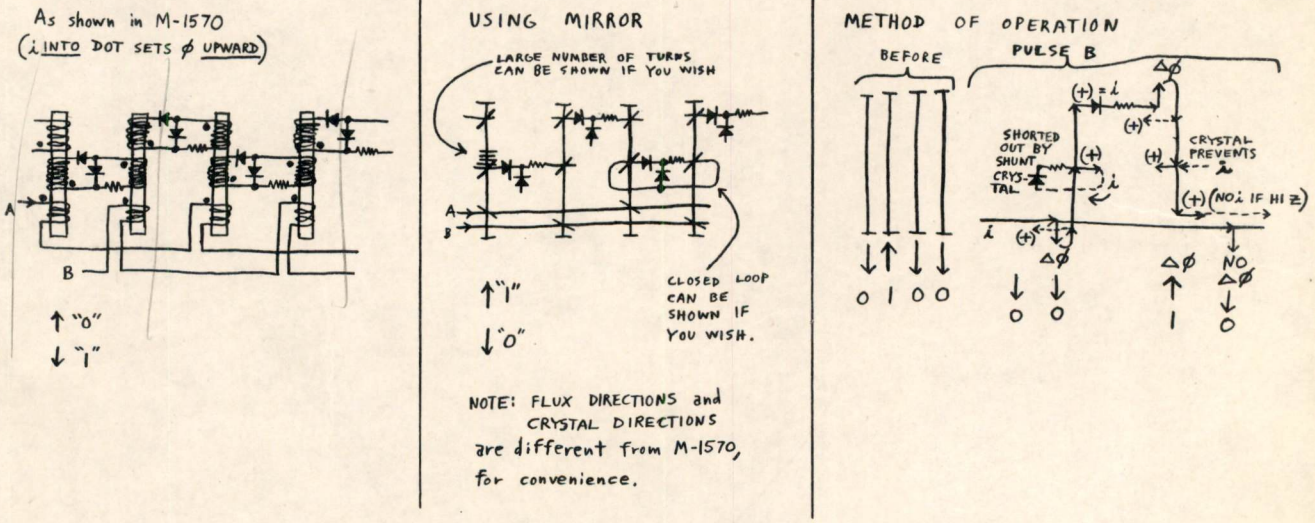


FIG. 7 - STEPPING REGISTER

Signed: R. P. Mayer
R. P. Mayer

Approved: William K. Linvill
W. K. Linvill

RPM/bs

R. Best

Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

SUBJECT: INTRODUCTION TO THE THEORY OF SEMICONDUCTORS III, CONDUCTION IN METALS; THE FIELD-FREE CASE

To: Transistor Group

From: D. J. Eckl

Date: August 28, 1952

Abstract: The two previous notes, E-463 and E-468 were written to supply a certain amount of essential background material. The present discussion serves as an introduction to the theory of conduction in metals. The method used gives an oversimplified picture of the conduction process but is useful in that it explains several physical phenomena. Also this mechanism is often employed and it is important for the reader to understand the assumptions made in its formulation and its basic limitations.

3.0 The Classical Theory of Metallic Conduction

Metals are composed of that class of elements in the periodic table which have one or two extra valence electrons outside of a completed shell. A completed shell is a tightly bound structure which is difficult to break, but valence electrons may be removed from the atom with little difficulty. It was suggested, therefore, that electrical conductivity in metals may be due to the tendency of these valence electrons to act as "free" electrons and thus appear as an electron gas in the interatomic spaces of the metal. The electrons in such a gas could drift under the influence of an electric field and thus transport charge.

This idea was developed into the classical theory of metallic conduction by Drude and Lorentz. They assumed that the electrons had the distribution of velocities proper to a gas at the temperature of the metal on the basis of the Maxwell-Boltzmann statistics. They were able to show that such a theory would account for thermal as well as electrical conductivity. However, a difficulty arose in the case of the specific heat. The specific heat is the amount of heat energy required to raise the temperature of 1 cc of the metal 1° Centegrade at a particular temperature. Its value depends on the change in the internal energy of the molecules of the metal with temperature. If the electrons form a gas which exhibits a change in energy with temperature as predicted by the Maxwell-Boltzmann distribution, they should contribute a considerable amount to the value of the specific heat. Experimental

values of the specific heat, however, can be entirely accounted for by the change in the molecular energies. Thus the expected large contribution of the electron gas fails to materialize.

The classical Maxwell-Boltzmann distribution requires zero translational energy for the electrons at absolute zero. As the temperature is increased, the average energy (being equal to 3/2kT) also increases. The energy distribution at approximately 300° absolute and 1500° absolute are shown below in Figure 1. At room temperature, 300°K,

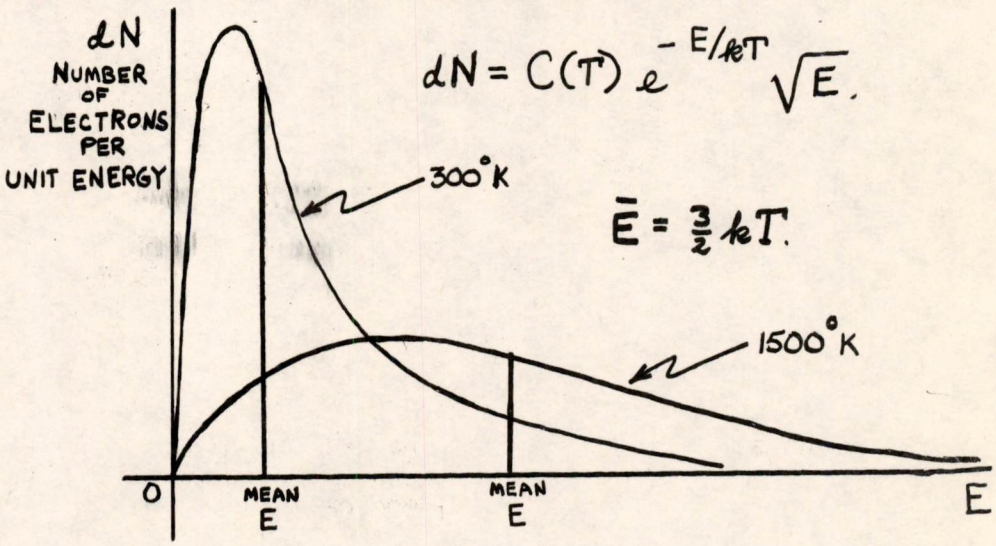


Figure 1

the average electron according to this theory would have an energy of .04 electron-volt. This corresponds to an electron velocity of 2.3×10^6 cm per second or 53,000 miles per hour. At 1500° K which is slightly above the melting point of copper the electrons would have an average energy of 0.2 electron-volt, corresponding to an electron velocity of 11×10^6 cm per second or 260,000 miles per hour.

3.1 The Quantum Viewpoint

Now consider the situation from the quantum point of view. In this case the electrons are restricted to quantized energy levels. In fact, for a metal with N identical atoms each energy state of the atom splits up into N states as a result of the interactions of the atoms. According to the Pauli Exclusion Principle each state may be occupied by only two electrons with opposite spins. The fact that there is only a discrete number of allowed energy levels, each of which may contain only two electrons, means that it is not possible for all electrons to have zero energy at absolute zero. The electrons seek to have minimum energy and so fill all energy levels, two per level, until a maximum energy is reached at which all electrons are taken care of. This maximum energy is called the zero point energy and is of the order of 8-10 electron-volts for metals at absolute zero. Its value may be found from the expression,

$$W = \frac{h^2}{2m}\left(\frac{3N}{8\pi}\right)^{2/3} = 3.6 \times 10^{-15} \times N^{2/3} \text{ e.v.} \qquad (3.0)$$

where N is the number of electrons per unit volume. Such an energy in the classical picture would require a temperature of 100,000° K. Under these conditions the degeneracy parameter (see E-468) $A_o = e^{-\alpha}$ is $>> 1$ so the electron gas is highly degenerate. Therefore, Fermi-Dirac statistics must be employed. For large A_o it can be shown that

$$\alpha = \frac{W}{kT} \qquad (3.1)$$

Therefore, the number of electrons per unit energy with energies between E and E + dE is given by the expression

$$dN = \frac{8\pi(2m)^{3/2}}{h^3} \frac{\sqrt{E}\, dE}{e^{\frac{E-W}{kT}} + 1} \qquad (3.2)$$

The distribution curve is drawn below in Figure 2.

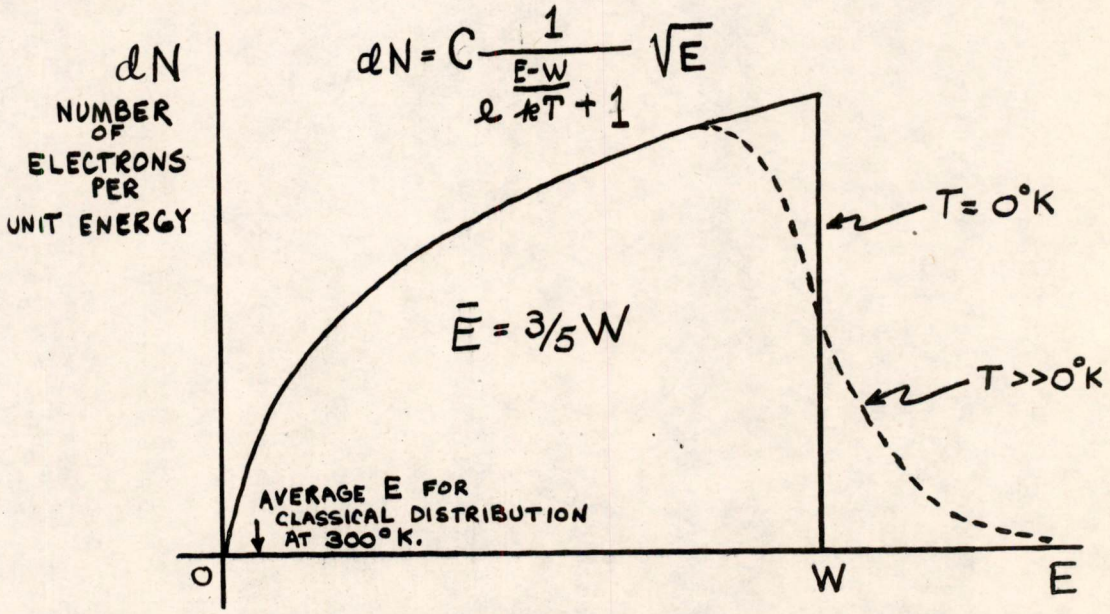


Figure 2

A comparison with the classical distribution in Figure 1 shows immediately that electron energies are much higher. Yet the problem of specific heats is taken care of by the fact that the total energy of the electrons changes very little with temperature. Thus the

electrons should make no noticeable contribution to the specific heat. Note also that as the thermal energy becomes greater than W the degeneracy disappears and the distribution approaches the classical one.

3.2 The Field-Free Assumption

The atoms in a metal form a uniform lattice—that is, they occupy fixed positions with a regular spacing. Suppose an electron is moved

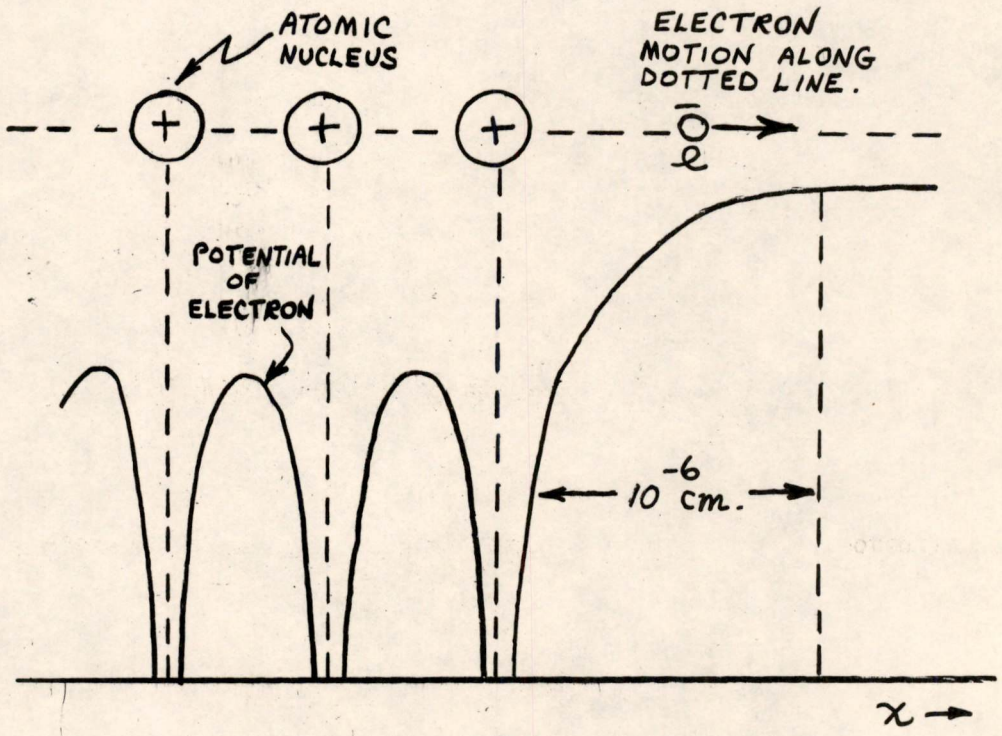


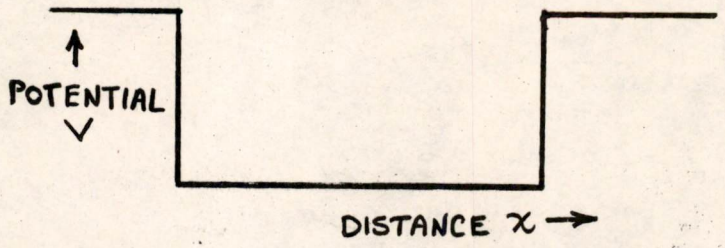
Figure 3

along the axis of such a row of atoms as shown in Figure 3. The potential energy of the electron as it moves from left to right is also shown in Figure 3. Inside of the metal where the electron is always close to an atom the attraction is due to the positive nucleus. When the electron leaves the metal and moves far enough away for the atoms to appear as a surface, the attraction is the result of the image force. An electron at a distance d from a conducting surface is attracted by the same force as would exist if there were an equal and opposite charge at a distance d inside the surface. The potential at large distances is given by the relation:

$$-e^2/16\pi\varepsilon_0 d \qquad (3.3)$$

The potential curve drawn in Figure 3 may be roughly approximated by the well drawn in Figure 4. This neglects all force fields inside the metal resulting

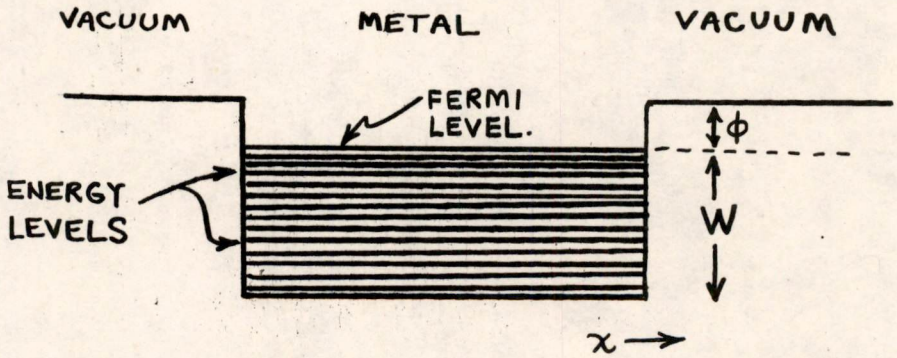
Figure 4



from the presence of the metallic atoms and is thus called the field-free case. Since this form of potential function is frequently drawn for metals, it is important to recognize that it is only an approximation. As might be expected, when the correct picture is used some considerably different results are obtained. However, the field-free approximation is close enough to give a useful insight to the nature of some physical phenomena.

The energy levels in such a system are quantized as shown in Figure 5. The height of the top filled band W depends solely on the number of free

Figure 5



electrons available in the metal. There is a potential barrier of ϕ electron-volts which prevents electrons from leaving the metal. The top level W is also referred to as the Fermi level.

3.3 Electrical Conductivity

The classical picture of the free electron gas in which the electrons are accelerated by the application of an electric field and hence produce a transport of charge must be modified. With the Fermi-Dirac distribution only the electrons in the upper levels may be excited and take part in conduction, since an acceleration requires a change of energy. The electrons in a given level have a fixed energy and since all lower levels must be filled, there is no possibility of a jump from a lower level to one slightly higher. The movement through the metal of those electrons which can be accelerated by the field

is retarded by "collisions" with the atoms in the metallic crystal lattice. A collision in this case must not be thought of as a billiard-ball type of impact. A collision occurs whenever the momentum mv of the electron is altered even though no physical contact occurs. These retarding collisions in which electrons lose energy can only occur for those in the upper levels, and they result in the electrical resistance of the metal.

Suppose the mean free time between collisions is defined as τ seconds. If a field E volts/cm is applied, the force on an electron is -eE. Let $P = p_1+p_2+\cdots+p_n$, the sum of the values of momentum for all n electrons. This sum is normally zero for a random distribution with no external field. It is now necessary to calculate the change in P produced in a time dt by both the electric field and the collisions. We need only consider the x-direction. The increase in total momentum in time dt caused by the presence of the field is given by

$$dP_x = - n e E_x dt. \tag{3.4}$$

Since τ is the mean free time between collisions, in time dt a fraction dt/τ of the electrons will lose momentum by collisions. We will assume that they eliminate exactly their proportional share of the total momentum. Therefore, the decrease in the total momentum in time dt as a result of collisions is:

$$dP_x = -P_x \frac{dt}{\tau} \tag{3.5}$$

The steady state condition is reached when the loss by collisions is equal to the increase due to the field.

$$-P_x \frac{dt}{\tau} = n e E_x dt. \tag{3.6}$$

Therefore the steady-state total momentum in the x direction is:

$$P_x = - n e E_x \tau. \tag{3.7}$$

For a general field, the total momentum is then:

$$P = - n e E \tau. \tag{3.8}$$

This gives the average value of momentum for a single electron as:

$$\bar{p} = \frac{P}{n} = - e E \tau. \tag{3.9}$$

The average velocity of an electron is:

$$\bar{v} = \frac{\bar{p}}{m} = - \frac{e \tau}{m} E \tag{3.10}$$

The current carried by n electrons per cc is:

$$I = - ne\bar{v} = ne \frac{e \tau}{m} E \tag{3.11}$$

is retarded by "collisions" with the atoms in the metallic crystal lattice. A collision in this case must not be thought of as a billiard-ball type of impact. A collision occurs whenever the momentum mv of the electron is altered even though no physical contact occurs. These retarding collisions in which electrons lose energy can only occur for those in the upper levels, and they result in the electrical resistance of the metal.

Suppose the mean free time between collisions is defined as τ seconds. If a field E volts/cm is applied, the force on an electron is -eE. Let $P = p_1+p_2+\cdots+p_n$, the sum of the values of momentum for all n electrons. This sum is normally zero for a random distribution with no external field. It is now necessary to calculate the change in P produced in a time dt by both the electric field and the collisions. We need only consider the x-direction. The increase in total momentum in time dt caused by the presence of the field is given by

$$dP_x = - n e E_x dt. \tag{3.4}$$

Since τ is the mean free time between collisions, in time dt a fraction dt/τ of the electrons will lose momentum by collisions. We will assume that they eliminate exactly their proportional share of the total momentum. Therefore, the decrease in the total momentum in time dt as a result of collisions is:

$$dP_x = -P_x \frac{dt}{\tau} \tag{3.5}$$

The steady state condition is reached when the loss by collisions is equal to the increase due to the field.

$$-P_x \frac{dt}{\tau} = n e E_x dt. \tag{3.6}$$

Therefore the steady-state total momentum in the x direction is:

$$P_x = - n e E_x \tau. \tag{3.7}$$

For a general field, the total momentum is then:

$$P = - n e E \tau. \tag{3.8}$$

This gives the average value of momentum for a single electron as:

$$\bar{p} = \frac{P}{n} = - e E \tau. \tag{3.9}$$

The average velocity of an electron is:

$$\bar{v} = \frac{\bar{p}}{m} = - \frac{e \tau}{m} E \tag{3.10}$$

The current carried by n electrons per cc is:

$$I = - ne\bar{v} = ne \frac{e \tau}{m} E \tag{3.11}$$

The conductivity is the ratio of this current to the voltage and is:

$$\sigma = \frac{I}{E} = \frac{ne^2\tau}{m} \qquad (3.12)$$

The dependence of τ on temperature determines how the conductivity will depend on temperature. The magnitude of the ratio of the mean velocity to the electric field is called the mobility of the electron and is:

$$\mu = \frac{e\tau}{m} \qquad (3.13)$$

Thus the conductivity may also be expressed

$$\sigma = ne\mu \qquad (3.14)$$

3.4 Thermionic Emission

At zero degrees absolute the maximum energy possessed by an electron is W as shown in Figure 6. The energy distribution is shown at the left of the diagram. An electron must have an energy in excess of W by an amount ϕ in order

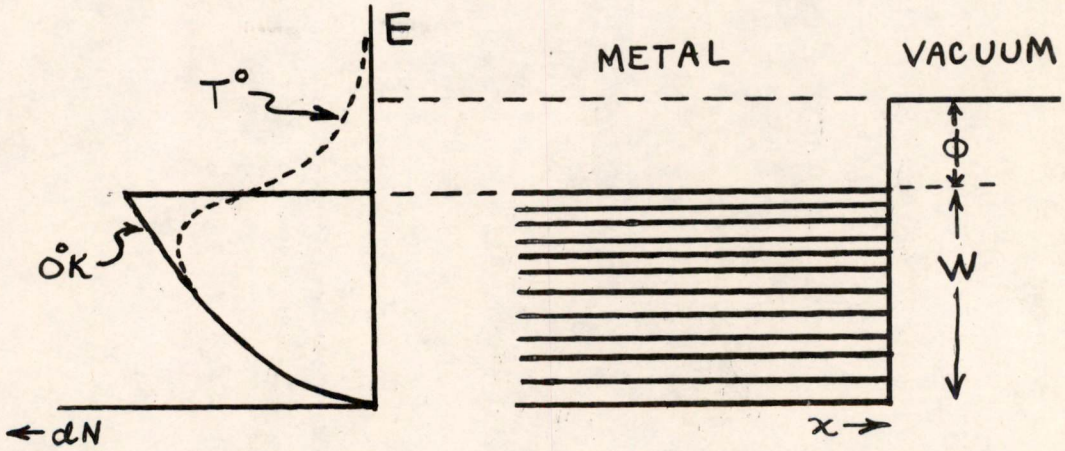


Figure 6

to escape from the metal. At zero degrees there are no electrons with this much energy but as the temperature is increased the energy distribution obtains a tail which exceeds $W + \phi$. This tail becomes higher as the temperature increases, indicating the presence of more and more electrons with an energy greater than $W + \phi$. Therefore, at higher temperatures, there will be electrons which can escape, and their number will increase with temperature. The quantity ϕ is called the work function of the metal. It represents the work which must be done by an electron in escaping from the surface of the metal. The expression for current against temperature is Richardson's equation:

$$I = AT^2 e^{-\phi/kT} \qquad (3.15)$$

Where I = current per unit area.
A = constant
T = temperature in degrees Kelvin.

3.5 Contact Potential

Consider the case where two different metals, A and B, which have potential diagrams as shown in Figure 7a, are brought into contact. As the two metals approach, thermionic emission will allow electrons to escape from one metal and reach the other. There will be a greater transfer of electrons from A to B since $\phi_B > \phi_A$. The result is that B acquires a negative potential

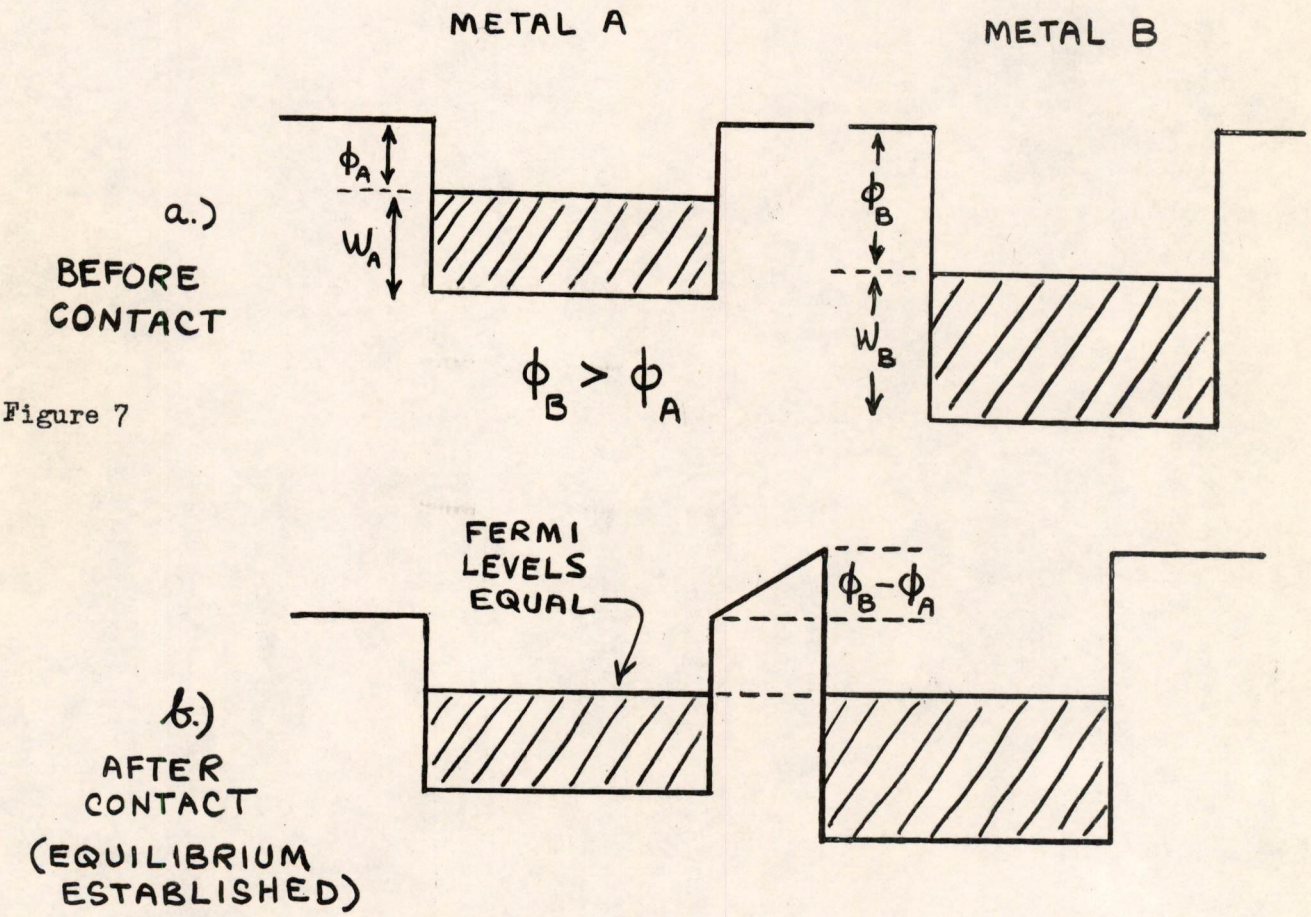


Figure 7

and A a positive potential as indicated by the slope between them in Figure 7b. Moreover as the barrier begins to form there will continue to be a net electron flow from A to B due to the tunnel effect since the barrier is smaller for A. This continues until an equilibrium is reached and it can be shown that the Fermi levels for the two metals in contact must then be at the same height. The difference in potential across the boundary is called CONTACT POTENTIAL.

3.6 High Field Emission

The question of retarding and accelerating fields may also be treated briefly at this time. The retarding field is of little interest. This case is shown in Figure 8a. The height of the barrier is effectively increased and

emission current is given by the equation

$$I_r = I_o e^{-\frac{eV}{kT}} \qquad (3.16)$$

When an accelerating field is present, however, the barrier is modified as

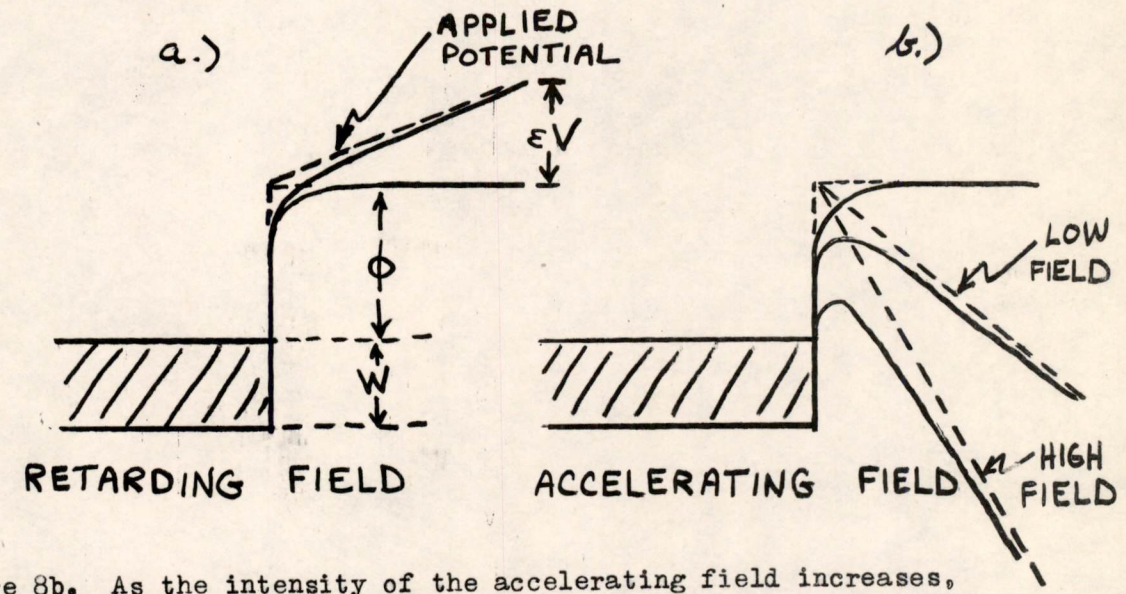


Figure 8

shown in Figure 8b. As the intensity of the accelerating field increases, the height of the barrier decreases and the thermal emission increases. The work function never becomes zero--i.e., the barrier does not vanish--but it becomes small and narrow enough for appreciable tunnel effect to take place. This requires a field of from 20 to 50 million volts per cm. The emission is given by the Schottky equation:

$$I_a = I_o e^{\frac{\varepsilon^{3/2} V^{1/2}}{kT}} \qquad (3.17)$$

Signed Donald J. Eckl
Donald J. Eckl

Approved John F Jacobs
John F. Jacobs

Norman H. Taylor

DJE/jk

References

"Semiconductors," D. A. Wright (Wiley)

"Applied Electronics," EE Staff, M. I. T. (Wiley)

E-468, "A Brief Discussion of Statistical Mechanics," D. J. Eckl

Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

Subject: A MAGNETIC-CORE GATE AND ITS APPLICATION IN A STEPPING REGISTER

To: 6889 Engineers

From: G. R. Briggs

Date: October 30, 1952

Abstract: A magnetic-core gate to replace diodes is a necessity if an all-core computer is to be made feasible. A device is described composed of two magnetic cores and a resistance which will operate as a gate between the cores of a stepping register, replacing the diodes conventionally used. The theory of operation is presented, indicating an upper limit to the operating frequency of about 50 kc at present. A possible method of increasing the operating frequency is discussed.

1.0 Introduction

A magnetic-core device to prevent undesirable core interaction is needed in an all-core computer to insure that information pass only in a forward direction through the machine. Diodes have been used so far, but they have two serious disadvantages:

1) Preventing current flow in one direction is not sufficient in itself to eliminate unwanted core interaction.

2) Diodes in core circuits operate with large back-voltage forward-current products at the frequencies desired. Diodes commercially available, operating with the back-voltage forward-current products required with even the smallest cores now obtainable, are not too reliable. This situation is expected to improve considerably with the availability of smaller cores, such as the F-291 ferramic series. Diode heating is also a serious problem at high speeds.

A better device than a diode for preventing undesirable core interaction would be a switch in series with core-coupling windings. This switch could be closed only when useful information transfer is desired and left open at all other times. Applied to a stepping register,[1] (see Figure 1), the switch would be closed when transferring a "one" from the first to the second core, but open during transfer of the "one" from the second core to the third and while setting up the first core to a "one". This switch would operate properly if it were closed by the advance time pulse applied to the first core. The switch could be thought of as a time pulse operated gate circuit. Such a device, approximating a time pulse operated switch, using magnetic cores, without diodes, is the subject of this note.

2.0 A Magnetic-Core Gate Applied to a Stepping Register

A saturable reactor could be used to couple stepping-register cores, as shown in Figure 2. If it is desired that small coupling exist between core 1 and core 2, then the saturable reactor should be capable of absorbing the voltage output on core 1, producing no voltage drop across winding N_2 of core 2. This means that

$$e_1 = \frac{N_1 d\phi_1}{dt} = e_3 = N_3 \frac{d\phi_3}{dt}, \qquad (1)$$

or, if $\Delta\phi_1$ and $\Delta\phi_3$ are the flux changes that occur in a given time, by integrating equation (1);

$$N_1\Delta\phi_1 = N_3\Delta\phi_3. \qquad (2)$$

For the flux change $\Delta\phi_1$, the smallest possible current flow in the coupling circuit is desired to keep the loading of core 1 as small as possible. This means that $\frac{\Delta\phi_3}{N_3 I}$ unsaturated should be as large as possible, where $N_3 I$ is the mmf required by core 3 to produce the flux change $\Delta\phi_3$. On the other hand, if it is desirable to be able to pass information from core 1 to core 2 with as little loss as possible, core 3 should be capable of being saturated by an externally applied current pulse. The flux change in core 3 should then be as small as possible,

---

1. For a complete discussion of diode-coupled stepping registers, see Sims, Robert C., An Investigation of Magnetic-Core Stepping Registers for Digital Computers, Thesis, Digital Computer Laboratory, Massachusetts Institute of Technology, 1952.

whereas the information current should be as large as possible;

$\left(\frac{\Delta\Phi_3}{N_3 I}\right)_{\text{saturated}}$ should be as small as possible. This is best accomplished by a core exhibiting a rectangular hysteresis loop, as shown in Figure 4. If the core is d-c biased by a current source to point b, and if c is the saturation operating point, it is obvious that the ratio

$$\frac{\left(\frac{\Delta\Phi_3}{N_3 I}\right)_{\text{unsaturated}}}{\left(\frac{\Delta\Phi_3}{N_3 I}\right)_{\text{saturated}}}$$

will have its greatest value. If the core were not biased to point b, but left unbiased at a, the mmf pulse required would have to be larger by an amount equal to the magnetizing mmf. This reduces $\frac{\Delta\Phi_3}{N_3 I}$ considerably for the materials available and is undesirable. This effect is illustrated in Figure 3, a plot of switching time of a stepping-register information core versus applied mmf. The data were taken for an arrangement which operated satisfactorily as a gate core. The switching time of the stepping core was measured as a function of the mmf drive applied to the stepping core for:

(1) an unloaded stepping core,

(2) a stepping core loaded with an unbiased gate core, and

(3) a stepping core loaded with a gate core biased almost to point b in Figure 4.

Major loop operation of the gate core was insured by a resetting saturation-current pulse applied to the gate core simultaneously with the resetting pulse applied to the stepping core. From Figure 3 it is plain that the biased gate core loads the stepping core much less than the unbiased gate core at the moderate switching times used in practice (half a microsecond and up).

Let us now follow the action of this gate during a stepping-register cycle of operation. Starting initially with "zero's" in both core 1 and core 2 and with the gate core biased to point b, as shown in Figure 5, repeated application of advance pulses 1 and 2 and

of the gate-core saturation pulse will have no effect because none of the cores are switched. Suppose core 1 is switched to a "one" by an information pulse from the previous stepping core; then in order to reduce the loading of core 1 as much as possible as it is switching, the gate core should keep the coupling-circuit current flow as small as possible. The gate core flux changes to some point, say c, under the influence of the small current I_1. This current loads core 1 negligibly and has no effect on core 2 since it is not in the proper direction to switch this core. At the end of this information pulse core 1 is in a "one" state, the gate core is at point d, and core 2 is still in a "zero" state. On the next advance pulse (1) it is necessary for the core to exhibit small $\frac{\Delta\phi_3}{N_3 1}$ in order that the "one" in core 1 may be passed on to core 2. Application of the gate-core saturation pulse will switch the gate core to point f, where $\frac{\Delta\phi_3}{N_3 1}$ will now be small. This means that the gate-core flux change that occurred in setting up core 1 to a "one" state must be removed before the gate core works like a closed switch. During this process current flows in the coupling circuit in a direction opposed to the gate-core saturation pulse (Lenz' law) or in the I_1 direction in Figure 5. This current tends to switch core 1 back to a "zero" state before core 1 has time to pass the "one" on to core 2. For a coupling circuit containing no resistance, from equation (2) for setting up core 1 to a "one" state,

$$\Delta\phi_3 = \frac{N_1}{N_3}\Delta\phi_1, \qquad (3)$$

and if the flux change $\Delta\phi_3$ is removed, the corresponding flux change in core 1

$$\Delta\phi'_1 = \frac{N_3}{N_1}\Delta\phi'_3 = \frac{N_3}{N_1}\Delta\phi_3 = \frac{N_1}{N_3}\,\frac{N_3}{N_1}\Delta\phi_1 = \Delta\phi_1. \qquad (4)$$

This means that core 1 is switched completely back to a "zero" state before use can be made of its stored information. If some series resistance is introduced into the coupling circuit, however, (R in Figure 2), then it becomes possible to switch the gate core from point d to point e so slowly that the output emf of the gate core can never become large enough to drive sufficient current through the resistance to produce a large enough $N_1 I$ to reach the "knee" of the hysteresis loop of core 1 and switch this core. The best procedure is to apply a gate-core current pulse corresponding to approximately twice the magnetizing mmf of the core, slowly switching the core to a point in the neighborhood of e in Figure 5, then applying

a relatively fast, large, saturation-current pulse, moving on the loop rapidly to point f. Since the total flux change from e to f is small, much more rapid variation of mmf is possible without generating too large an output emf. After the switch-back of the gate core is complete, advance pulse 1 can occur, driving core 2 to a "one" state and resetting core 1 to a "zero" state. This is shown in Figure 6. I_2 is the coupling-circuit current flowing in this case. At the end of this process, the gate-core saturation-current pulse is removed, allowing the gate core to return to point h, which is the same as point b in Figure 5.

Next, one must consider the effect of advance pulse 2, which will reset core 2 to "zero", driving the "one" into the next stepping register core through a similar gate circuit. Information will tend to pass backward into core 1 now, as shown in Figure 7. The current I_3 tends to flow in the coupling circuit, and since the gate core still looks like a short circuit to this current, core 1 could be switched again. This situation can be remedied by adding a second gate core, as shown in Figure 8.

The windings on the two gate cores are poled as shown. Operation can be followed by reference to Figure 9. With "zero's" in both core 1 and core 2, application of advance pulses 1 and 2 and of the saturation-current pulse to the gate cores produces no switching of any core. On setting up core 1 to a "one" state, current I_1 flows in the coupling circuit, switching gate core A to point b, but not switching gate core B because of the reversed winding polarity. Gate core A is then slowly switched to d and then more rapidly to e by the gate-core saturation pulse, as discussed above. Gate core B is also saturated by the same gate-core current pulse, with negligibly small change in flux. On advance pulse 1, the "one" is driven from core 1 to core 2, as shown in Figure 10. I_2 in Figure 10 is the coupling-circuit current flow in this case, and if the gate-core saturation mmf is equal to or greater than the mmf applied to gate core B by the current I_2, then the switching of gate core B can be inhibited; the combination of gate cores looks like a closed switch, and core 2 will be switched to a "one" state. At the end of advance pulse 1, the gate-core saturation pulse can be removed, leaving the cores in the state represented by point g in Figure 10. So far in this discussion, core B has served no purpose. On advance pulse 2 (Figure 11), the "one" is switched out of core 2 into the following gate core, setting up a current I_3 (Lenz' law) in the coupling circuit. I_3 tends to switch core B (not affecting gate core A), and this switching absorbs the voltage output of core 2, keeping the current I_3 small. I_3 can be kept small enough to produce no switching of core 1 and also small enough to load core 2 negligibly.

Before the next advance pulse 1 can occur, gate core B must be switched back again to point j at least, or spurious information would be introduced into the coupling circuit by the action of the gate-core saturation pulse on core B. This is again done by applying a long current pulse of approximately twice the magnetizing mmf of core B, (the same pulse used to reset core A when driving core 1 to a "one"), slowly switching gate core B to point j in Figure 11. The output voltage of the gate core can be so small during this process that the current it can force through R is insufficient to switch core 2 to a significant extent. Thus core A acts as an isolation element upon setting up core 1 to a "one" state, whereas core B acts as an isolation element upon reading out the transferred "one" from core 2.

3.0 Analysis of the Performance of the Magnetic-Gate Stepping Register

For purposes of analysis, it is sufficient to consider only one gate core at a time; see Figure 12. Divide the problem into two cases:

(a) Setting up core 1 to a "one" state and switching back the gate core prior to applying advance pulse 1.

(b) Transferring of the "one" from core 1 to core 2 and then driving the "one" out of core 2 into the following stepping core and finally switching back the gate core prior to applying advance pulse 1 again.

Let N_1, N_2, N_3 be the number of turns on the cores, as shown in Figure 12. Also let:

$\Delta\phi_1$ = flux change produced by switching core 1.

$\Delta\phi_2$ = flux change produced by switching core 2. Major loop operation is assumed.

$\Delta\phi_3$ = flux change produced by partially or completely switching the gate core.

I_{m1} = maximum current flow in the coupling circuit produced by switching back core 3 for case (a).

F_{m1} = corresponding mmf in core 1.

I_{m2} = maximum current flow in the coupling circuit produced by switching back core 3 for case (b).

F_{m2} = corresponding mmf in core 2.

t_1 = switching time of core 1.

t_2 = switching time of core 2.

t_3 = switching time of core 3 while blocking current flow.

T = switching time of core 3 during resetting of the core.

F_1 = mmf required to switch unloaded core 1 in time t_1.

F_2 = mmf required to switch unloaded core 2 in time t_2.

F_{L2} = mmf reflected into core 2 by the output load on core 2.

F_{L1} = mmf reflected into core 1 by the input load on core 1.

In analyzing the operation, the approximation made by Sims[1] will be used. He assumed that the output voltage of any core is a rectangular pulse of constant voltage amplitude and of fixed duration, t. This implies $\frac{d\phi}{dt}$ is a constant during switching of a core. This approximation is very crude and can be expected to yield only qualitatively correct results. The true output voltage shapes, in practice, are more nearly triangular. But use of any more accurate approximation complicates the theory to a tremendous extent.

Case a: On switching core 1 to a "one" state, a relatively small current flows in the coupling circuit, causing a flux change in the gate core to occur. The drop across R will be so small during this process that it can be neglected and since the current flow is not in the proper direction to switch core 2, we have approximately,

$$N_1\Delta\phi_1 = N_3\Delta\phi_3, \qquad (5)$$

Where $\Delta\phi_3$ is the flux change that has occurred in the gate core. In order to reduce the loading effect on core 1 as much as possible, the current flow in the coupling circuit should be as small as possible. This is best done by decreasing $\Delta\phi_3$ as much as possible, reducing the mmf required by the gate core to produce the flux change to a minimum. By equation (5), N_3 must be increased ($N_1\Delta\phi_1$ assumed constant, of course).

---

1. Sims, R. C., loc. cit.

This means that the minimum mmf requirement is brought about by the maximum number of turns N_3 and that, accordingly, the coupling-circuit current is a minimum. This is the same as requiring the energy loss in the gate core to be as small as possible. $\Delta\phi_3$ can be generated by completely switching a small gate core or only partially switching a larger gate core. The former method is advantageous, because the largest possible ratio of unsaturated to saturated $\frac{\Delta\phi}{NI}$ is obtained by allowing the gate core to switch as far as possible in the unsaturated state. This is true because in practice the gate core cannot be biased quite to the "knee" of the hysteresis loop for reasons of stability, so that $\frac{\Delta\phi}{NI}$ will first be small until the "knee" is reached, then will increase greatly during the rest of the total flux change excursion. The average $\frac{\Delta\phi}{NI}$ that occurs during the process is thus reduced as the $\Delta\phi$ excursion is reduced. Another advantage gained with a small core of the same magnetic material is that usually a reduction in switching mmf can be obtained for a given switching time, t , as the core size is reduced. One serious disadvantage in practice is that N_3 must increase (or $\frac{N_3}{N_1}$, at least) as the gate core size is reduced. More turns must be wound on a core with a smaller central hole. The wire size must thus be very rapidly reduced as the core size is reduced. An optimum gate core size must therefore exist in the practical case.

As explained in section 2.0, the flux change $\Delta\phi_3$ must be gotten rid of before the occurrence of the next time pulse. The maximum $\frac{d\phi_3}{dt}$ that can be tolerated is one that will force a current of magnitude less than that which will tend to switch core 1 appreciably in a time T, the length of time it takes to get rid of the flux change in the gate core. This current corresponds to an mmf slightly larger than the magnetizing mmf of the stepping core. If this current is called I_{m1},

$$N_3\left(\frac{d\phi_3}{dt}\right)_{max.} = I_{m1}R; \qquad (6)$$

or, since $F_{m1} = N_1 I_{m1}$,

$$\frac{d\phi_3}{dt}_{max.} = \frac{F_{m1}R}{N_1 N_3} . \qquad (7)$$

If T_A is the switch-back time of the gate core A,

$$\left(\frac{d\phi_3}{dt}\right)_{max.} = \frac{\Delta\phi_3}{T_A} \quad \text{by Sims' approximation.} \tag{8}$$

This leads finally to:

$$T_{A\ min.} = \frac{N_1 N_3}{F_{m1} R}\Delta\phi_3 = \frac{N_1^2 \Delta\phi_1}{F_{m1} R}\ , \tag{9}$$

by substitution of $\Delta\phi_3$ by $\Delta\phi_1$ by equation (5). $T_{A\ min.}$ is the minimum switch-back time that can be tolerated; longer times are, of course, allowed. $T_{A\ min.}$ depends upon the value of R and varies inversely with R. R has an upper limit, as explained below. Use of the Sims approximation here gives too low a value of T_a, because in reality the maximum $\frac{d\phi_3}{dt}$ will exceed the average $\frac{d\phi_3}{dt}$.

$$\text{Or, } \left(\frac{d\phi_3}{dt}\right)_{max.} > \left(\frac{d\phi_3}{dt}\right)_{ave.} = \frac{\Delta\phi_3}{T}, \text{ or } T > \frac{\Delta\phi_3}{\left(\frac{d\phi_3}{dt}\right)_{max.}} = T_{min.} \tag{10}$$

Case B will next be considered. Here gate core B blocks coupling current flow on reading out of core 2. On resetting gate core B, a current I_{m2} flows as shown in Figure 12, tending to switch core 2 again. Carrying out the same analysis for this case, an expression for $T_{B\ min.}$ can be found which is the same as equation (9) with N_2^2, replacing N_1^2, if core 1 and core 2 are identical, as they are in all practical stepping registers. N_1 greater than N_2 is always required (see p. 15). This means that $T_{B\ min.}$ is less than $T_{A\ min.}$, but the value of T which must be used is the larger of the two values or $T_{A\ min.}$, because resetting of core A and core B must be done on every advance pulse for one or more coupling circuits in a register composed of many stepping cores. The net stepping period T must then be greater than the larger of T_A or T_B.

Upon driving the "one" from core 1 into core 2, the coupling-circuit current flow must supply the mmf F_2 to switch core 2 in time t_2, plus any additional mmf (F_{L2}) reflected back into core 2 by current flow in the load circuit (the next gate core). The voltage loop equation of the coupling circuit becomes, again using Sims' approximation:

$$\frac{R(F_2+F_{L2})}{N_2} + \frac{N_2\Delta\phi_2}{t_2} = \frac{N_1\Delta\phi_{1p}}{t_2}\ . \tag{11}$$

$\Delta\phi_{1p}$ is the flux change that occurs in core 1 during the switching of core 2, and $\frac{\Delta\phi_{1p}}{t_2} = \frac{d\phi_1}{dt}$ during this process, $\frac{d\phi_1}{dt}$ being considered constant. $\Delta\phi_{1p}$, rather than $\Delta\phi_1$, is used because core 1 is in general not switched completely by the time core 2 is switched completely. This is necessary to prevent possible attenuation of the information as it passes down a line of cores.[2] After the core has completely switched, all the output voltage of core 1 must appear across R; core 1 is therefore loaded much more heavily and requires a fairly long time to finish switching. In fact, it may not have time to switch completely by the end of advance pulse 1. If the core is practically completely switched, say greater than nine-tenths of a complete switch, by the time core 2 is completely switched, whether or not it is able to complete the switching process by the end of advance pulse 1 is not very important in practice; the introduction of spurious information is negligible on the next advance pulse 1. The ratio $\frac{\Delta\phi_{1p}}{\Delta\phi_1}$ depends upon the turns ratio $\frac{N_1}{N_2}$ as well as upon R, from equation (11).

Solve equation (11) for $\Delta\phi_{1p}$:

$$\Delta\phi_{1p} = \frac{R}{N_1 N_2} t_2 (F_2 + F_{L2}) + \frac{N_2}{N_1} \Delta\phi_2,$$

and dividing by $\Delta\phi_1$,

$$\frac{\Delta\phi_{1p}}{\Delta\phi_1} = \frac{R\, t_2}{N_1 N_2 \Delta\phi_1} (F_2 + F_{L2}) + \frac{N_2}{N_1} \frac{\Delta\phi_2}{\Delta\phi_1}. \qquad (12)$$

Since usually core 1 and core 2 are the same size, $\Delta\phi_1 = \Delta\phi_2$, and

$$\frac{\Delta\phi_{1p}}{\Delta\phi} = \frac{R\, t_2}{N_1 N_2 \Delta\phi} (F_2 + F_{L2}) + \frac{N_2}{N_1} \qquad (13)$$

where $\Delta\phi \equiv \Delta\phi_1 = \Delta\phi_2$.

This shows if $\frac{\Delta\phi_{1p}}{\Delta\phi_1}$ is less than 1, as it must be for proper operation, $\frac{N_2}{N_1}$ less than 1 is always required. If $\frac{\Delta\phi_{1p}}{\Delta\phi}$ is equal to a maximum of 0.9, say, then $\frac{N_2}{N_1}$ is 0.9 maximum, and then only if R = 0. As R is increased to allow the recovery of the gate core in a reasonable

2. Buck, D. A., Binary Counting with Magnetic Cores, E-438, Digital Computer Laboratory, Massachusetts Institute of Technology, December 2, 1952.

time, $\frac{N_2}{N_1}$ must be correspondingly decreased. Substituting for R its value from equation (9) and solving for $\frac{T_{min.}}{t_2}$,

$$\frac{T_{min.}}{t_2} = \frac{N_1 (F_2 + F_{L2})}{N_2 F_{m1}} \quad \frac{1}{\frac{\Delta\phi_{1p}}{\Delta\phi} - \frac{N_2}{N_1}} .$$

Since $\frac{\Delta\phi_{1p}}{\Delta\phi} = 0.9$, we get finally,

$$\frac{T_{min.}}{t_2} = \frac{F_2 + F_{L2}}{F_{m1}} \quad \frac{1}{\frac{N_2}{N_1}\left(0.9 - \frac{N_2}{N_1}\right)} . \qquad (14)$$

It is, of course, desirable for maximum operating speed to keep $\frac{T_{min.}}{t_2}$ as small as possible. $T_{min.}$ has its minimum value for $\frac{N_1}{N_2} = \frac{2}{0.9} = 2.2$. Although this turns ratio gives the fastest operation, it also requires a rather large core-driving mmf.

To calculate the core 1 driving mmf required, the mmf reflected into core 1 from the coupling circuit to core 2 is:

$$F_{\text{reflected into core 1}} = \frac{N_1}{N_2} (F_2 + F_{L2}). \qquad (15)$$

Also, the mmf required to switch core 1 is very nearly F_2, since t_1 for $\frac{\Delta\phi_{1p}}{\Delta\phi} = 0.9$ is equal to t_2. This means that the mmf required to switch core 1 only 90% in the same time as a 100% switch of core 2 is less than F_2. No data are available to give an idea of how much less F_1 is, but for the purposes of this report, F_1 can be assumed equal to $F_2 \equiv F$. Also, another mmf is reflected into the core 1 from the previous gate core, which looks almost like an open circuit to core 1 on advance pulse 1, F_{L1}. F_{L2} is the load of the gate core following core 2. This gate core also looks almost like an open circuit to core 2, but F_{L2} is not equal to F_{L1} because the turns ratios are different in the two cases. Finally, the mmf drive which must be supplied to core 1

by advance pulse 1 is:

$$F_{drive} \cong F_{L1} + F + \frac{N_1}{N_2}(F + F_{L2}). \qquad (16)$$

If the gate core material is rectangular enough that voltage drop across the gate cores can be neglected when they are saturated, then only the loading effect of the unsaturated gate cores need be considered in finding the driving mmf. This gate core loading effect is best determined experimentally; Figure 3 gives data on the best ferrite gate cores tried. These cores were identical to the stepping cores and were considerably larger than is desirable; smaller cores are now becoming available and will soon be tested as gate cores. Molybdenum-permalloy cores also show great promise as gate cores, but the data for them are incomplete as yet.

T for the case of MF-1118 stepping cores of the F-259 size will now be calculated for a typical circuit, using the data of Figure 3 and also the data of Figure 13. Figure 13 contains curves of applied mmf versus switching time for the case of 5%, 10%, and 20% of complete core switching. This information is of importance in determining F_{m1}, for F_{m1} is the maximum mmf that can be applied to core 1 by the slow switching of the gate core without causing more than a negligible percentage of switching of this core.

The first step is to pick a reasonable value for $t_1 \cong t_2$, the switching time of the core. Let us pick $t_1 = 0.7$ of a microsecond (say), then $F + F_{L2} = 8.3$ ampere turns, from Figure 3. From Figure 13, assuming a 10% allowable core switching by the gate core, F_{m1} is about 2 ampere turns. For a turns ratio $\frac{N_1}{N_2} = 1.5$, for which the data of Figure 3 were taken, $T_{min.} = 19$ microseconds. This means an upper frequency limit of about 50 kc for the circuit tested. This experimental value of T can be used to find a better value of F_{m1} from Figure 13. A better value would be $F_{m1} = 1.9$ ampere turns. $T_{min.}$ can now be corrected slightly with this better value. Unfortunately this result cannot as yet be verified because the minimum prf of the test equipment is 150 kc at present. The mmf necessary to drive the register can also be calculated from equation (16) for this case. $F_{L1} + F = 7.4$ ampere turns from Figure 3, and the driving mmf = $7.4 + 1.5\ (8.3) = 19.9$ ampere turns. Switching an unloaded core in 0.7 microsecond takes $F = 6.2$ ampere turns. For 100% efficiency, it should take 12.4 ampere turns to switch both core 1 and core 2, or 7.5 ampere turns are necessary to handle gate core and resistive losses in this circuit. This efficiency is better than that of the diode-coupled stepping register[1] for this low turns ratio of $\frac{N_1}{N_2} = 1.5$.

---

1. Sims, R. C., loc. cit.

To indicate future possibilities, consider the driving mmf for the case $F_{L1} = F_{L2} = 0$, or the gate cores acting as perfect open circuits. Then for $\frac{N_1}{N_2} = 2.2$ (optimum) the driving mmf = 3.2F. This indicates a maximum efficiency of 62% for the fastest possible circuit. Also, $\frac{T_{min.}}{t_2} = \frac{4.9\ F_2}{F_{m1}}$ from equation (14) for the ideal case. $\frac{F_2}{F_{m1}}$ depends mostly upon the rectangularity of the core; it is 1 for a perfectly rectangular loop. It is about 3 for the ferrite cores presently in use. Thus $T_{min.}$ lies ideally between 5 and 15 times the switching time of the stepping cores.

4.0 Summary

The feasible speed of the magnetic-core plus resistance gate with ferrite cores of the F-259 size appears to be about 50 kc, which is 20 times slower than the corresponding diode-coupled circuits. Gate core driving requirements are very moderate, as the back emf's generated by the gate core during the saturation and slow switch-back pulses are small. The emf's generated across the gate-core external-pulse windings during the gate-core blocking operations are, however, quite large. These emf's always tend to increase the pulser plate voltage and require a pulser tube with sharp cut-off characteristics, biased considerably below cut-off. This can be quite serious if a long stepping register contains a fairly large number of "one's". Another disadvantage is that the individual cores must be switched in a time much less than the period of operation, producing larger back emf's upon driving than would be necessary with a diode-coupled register of the same transfer speed.

The logical place to use this circuit and other low-speed circuits based upon this gate would be for in-out problems in a computer of WWI speed. Even here the improvement in reliability over comparable diode-coupled circuits would be decreased because of the greater reliability of the diode circuits at low speeds, where the diode forward currents and back emf's can be kept small. Reducing the speed of diode circuits is limited by the forward resistance of the diodes, however, and this requires use of rather bulky selenium diodes at low speeds.

A possible method of increasing the frequency of operation is under investigation, and results will probably be incorporated into a future E-note. A capacitor is used instead of the resistor in the coupling circuit. The "ringing" current flow in this capacitor can be used to switch the gate core back after a blocking operation, requiring no external

gate-core current pulse. The reversed polarity of current flow in the coupling circuit does not tend to switch core 1 back, and the partial switching of core 2 that occurs instead is not serious. The stepping rate of the circuit is quite critical; it can only be varied a few percent from optimum for given circuit parameters. A "one" has been successfully cycled around a closed loop of four stepping cores at 300 kc by this circuit. In common with other capacitor-coupled core circuits tested it will require a great deal of engineering to make the circuit less critical to variations in parameters and driving currents.

Signed George R. Briggs
George R. Briggs

Approved NHT
Norman H. Taylor

Approved William K. Linvill
William K. Linvill

GRB/bs

Drawings Attached:

A-52750 Fig. 1 and 2
A-52792 Fig. 3
A-52757 Fig. 4 and 5
A-52751 Fig. 6 and 7
A-52752 Fig. 8 and 9
A-52758 Fig. 10 and 11
A-52753 Fig. 12 and 13

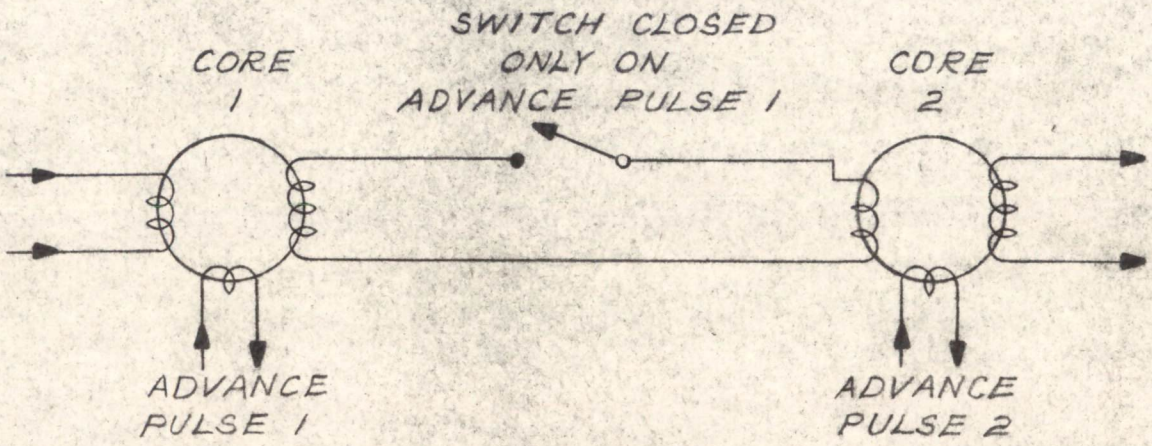


FIG. 1

STEPPING REGISTER WITH SWITCH ISOLATION

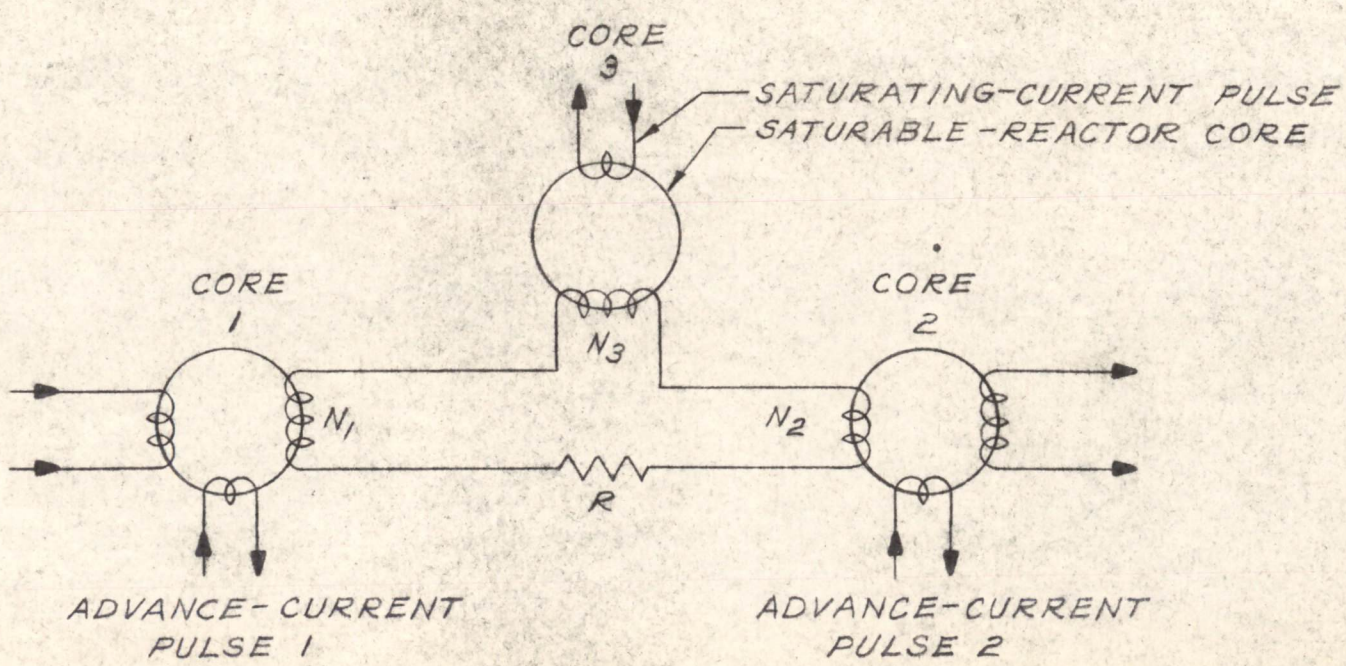


FIG. 2

A MAGNETIC-CORE GATE STEPPING REGISTER

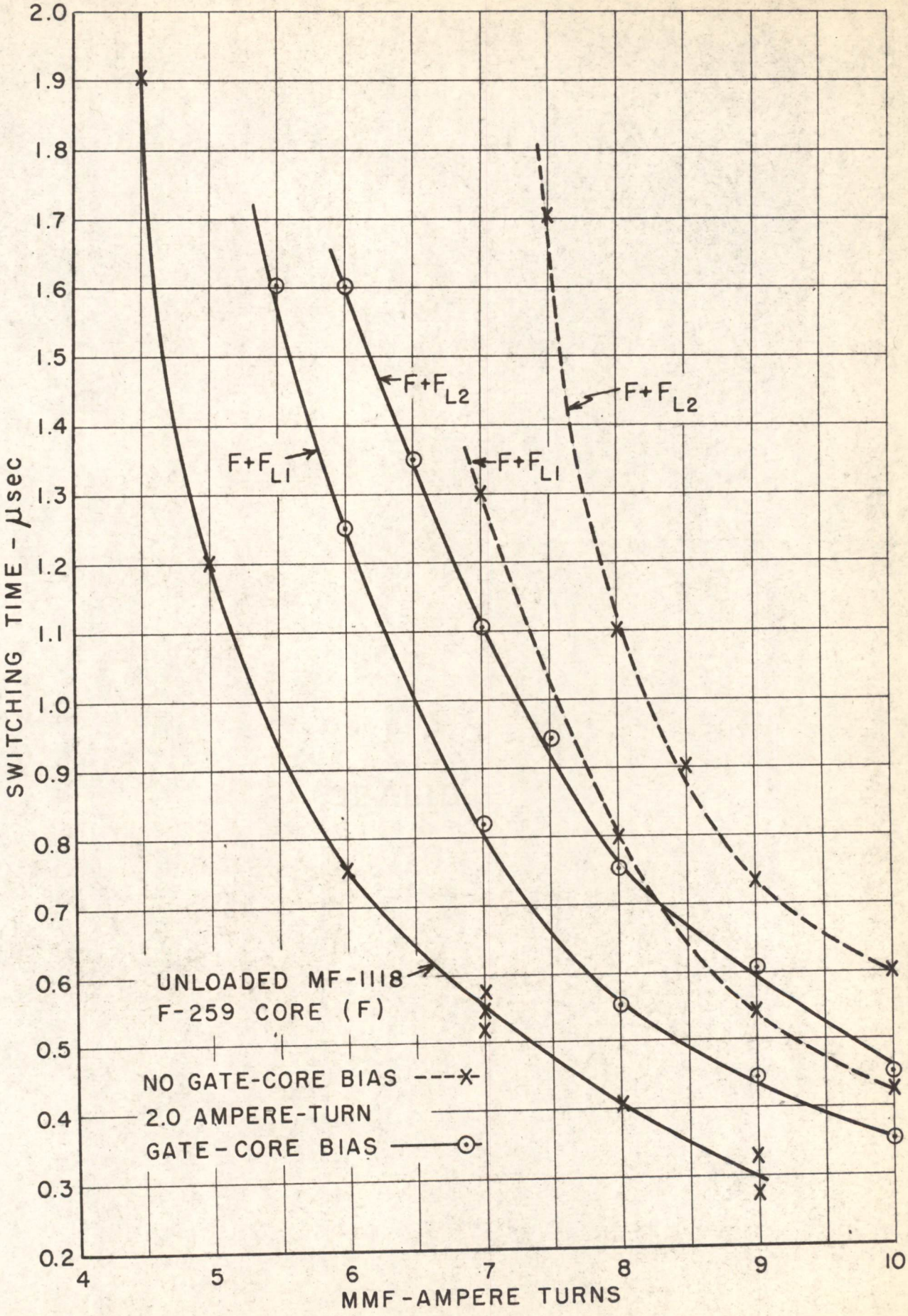


FIG. 3

SWITCHING TIME OF STEPPING CORE
WITH AND WITHOUT GATE-CORE LOAD

A-52792

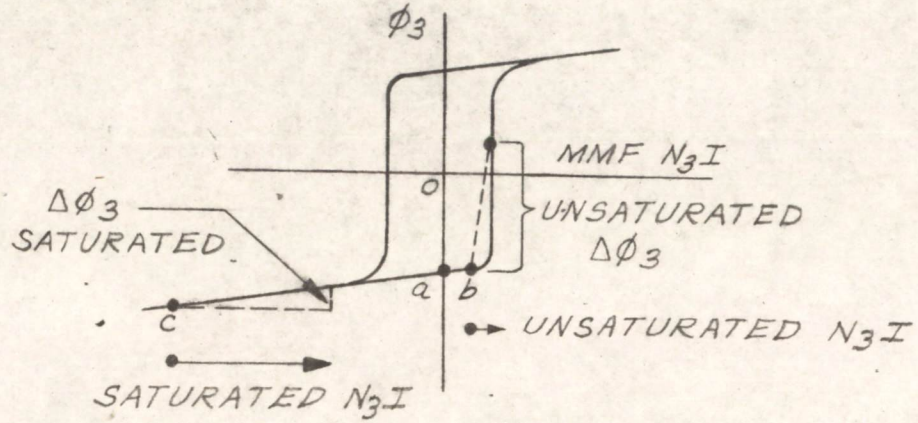


FIG. 4

GATE-CORE OPERATION

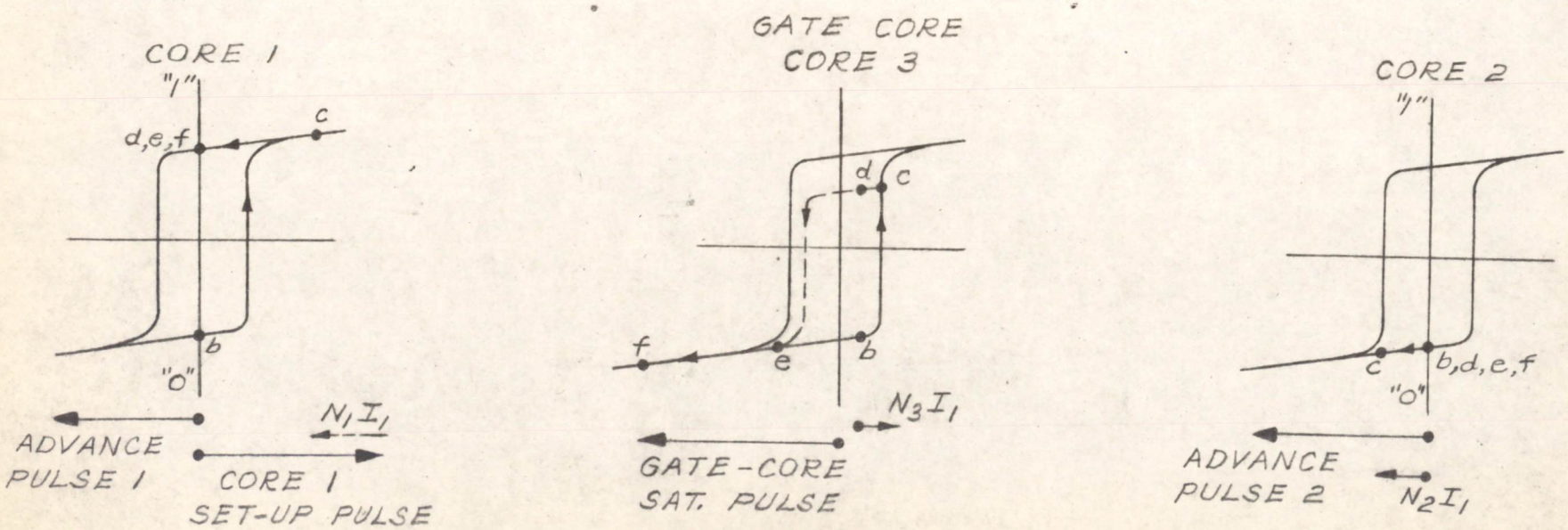


FIG. 5

SET-UP OF CORE 1 AND RESET CORE 3

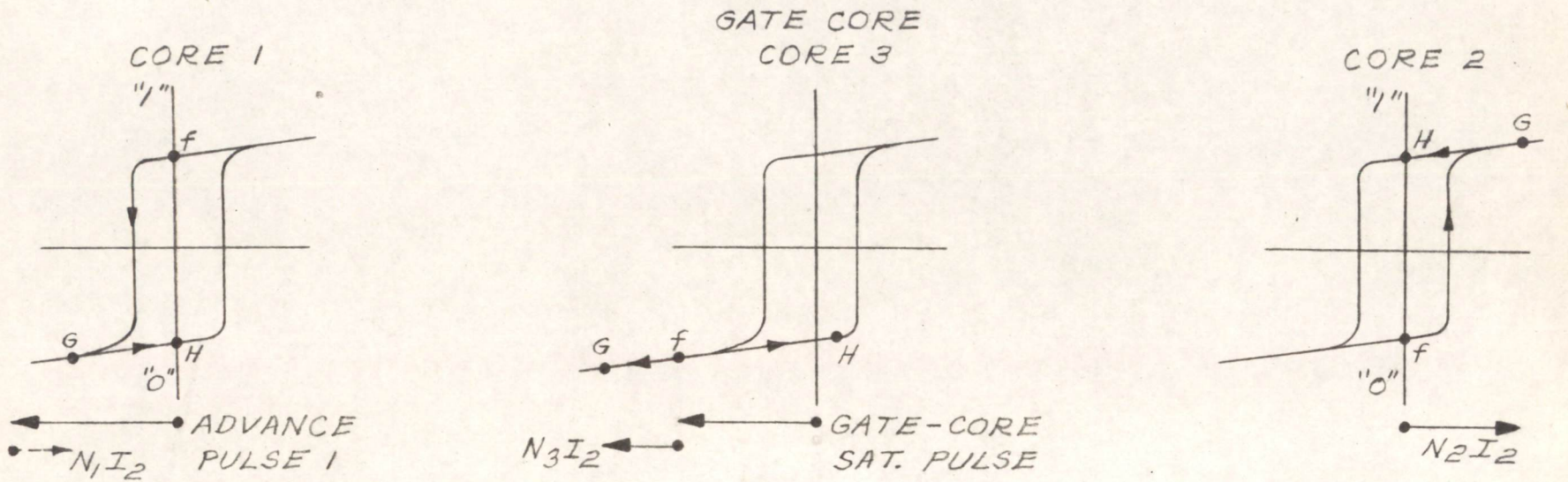


FIG. 6

TRANSFER OF "1" FROM CORE 1 TO CORE 2

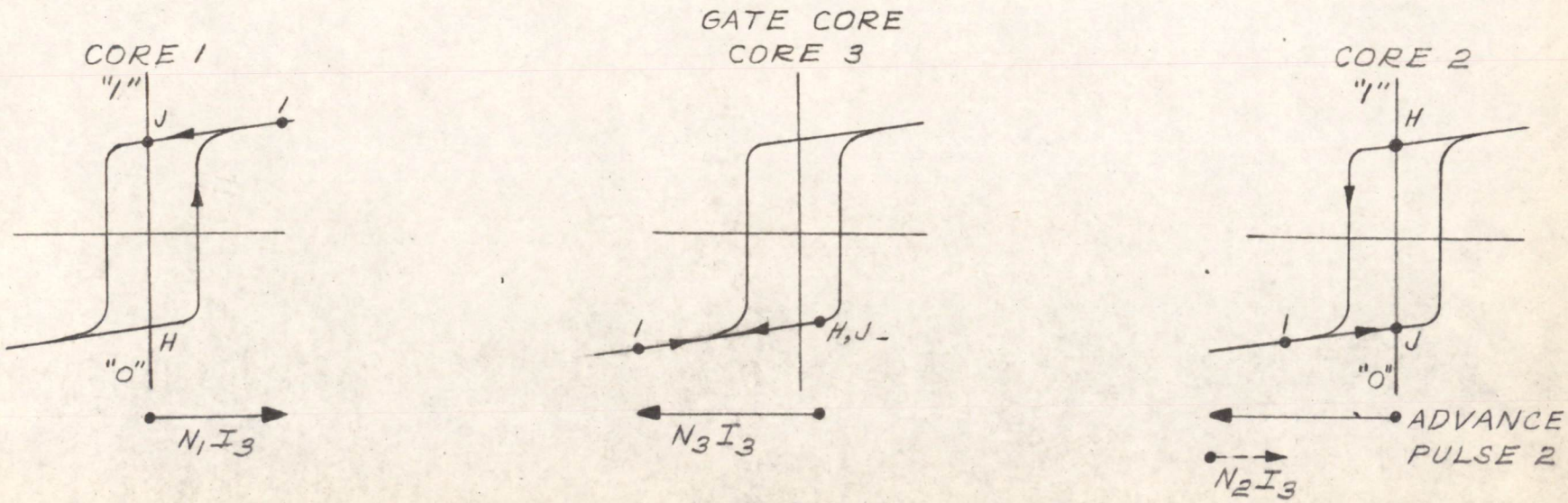


FIG. 7

READING OUT "1" FROM CORE 2

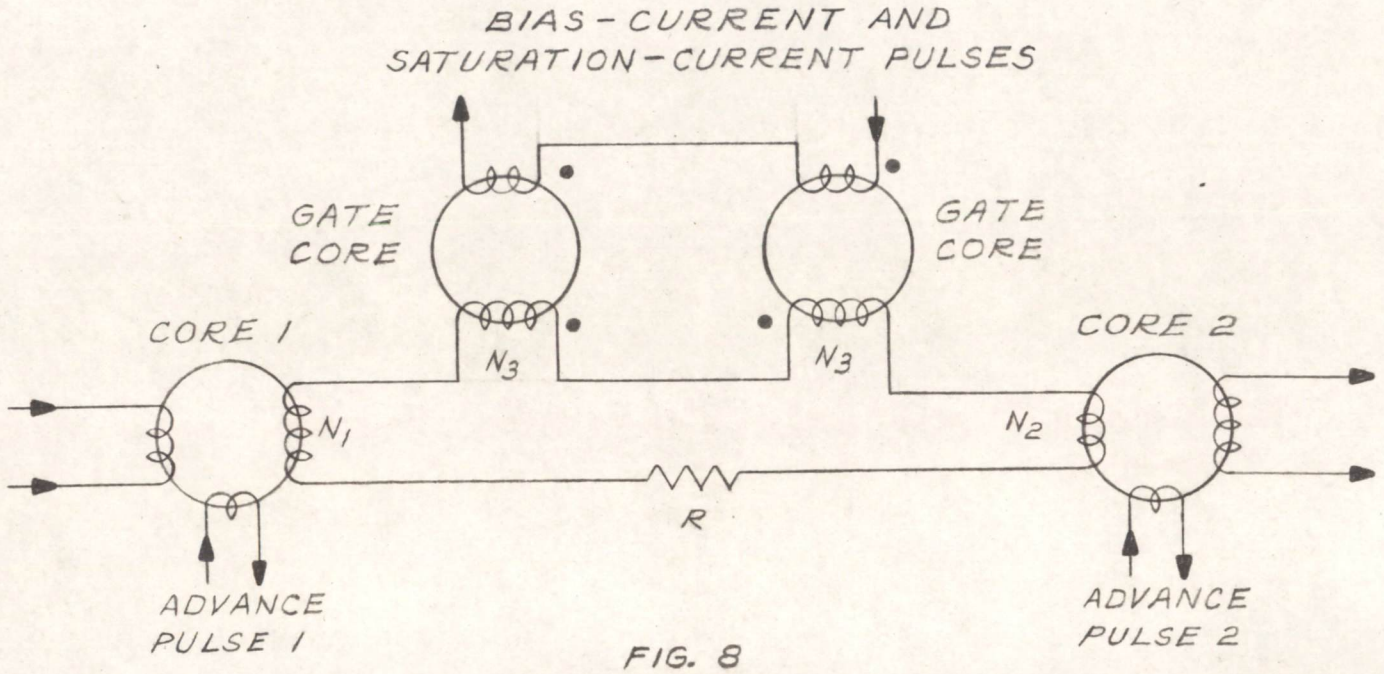


FIG. 8

FINAL CIRCUIT OF MAGNETIC-CORE GATE IN A STEPPING REGISTER

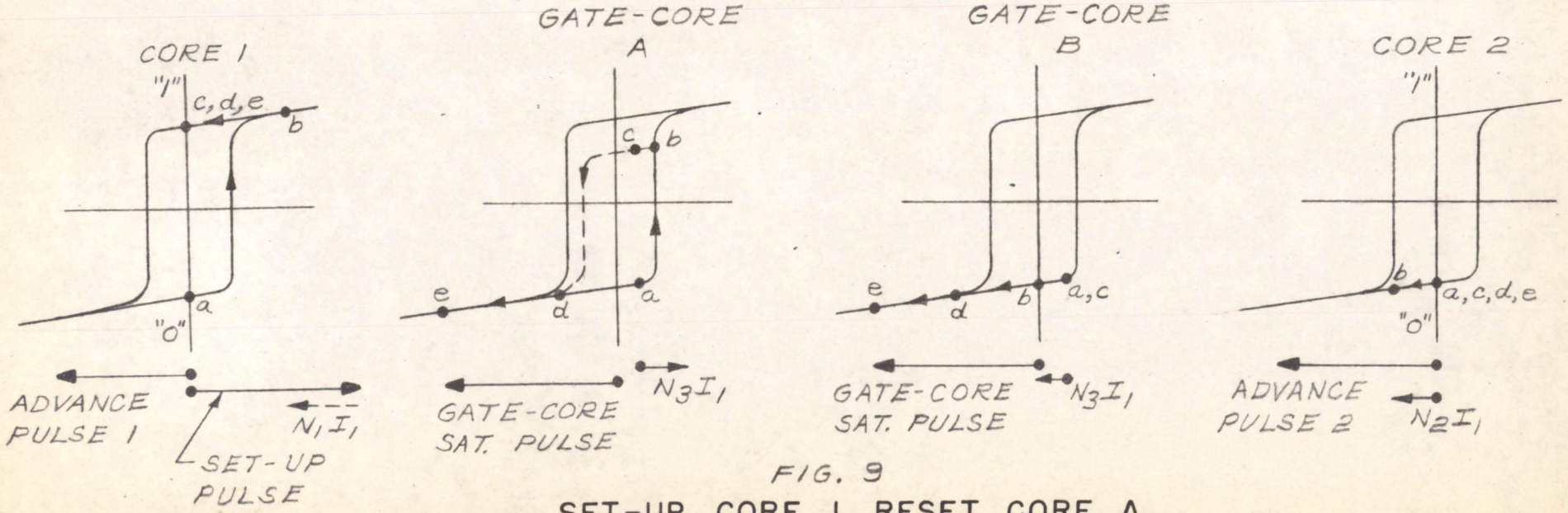


FIG. 9

SET-UP CORE 1, RESET CORE A

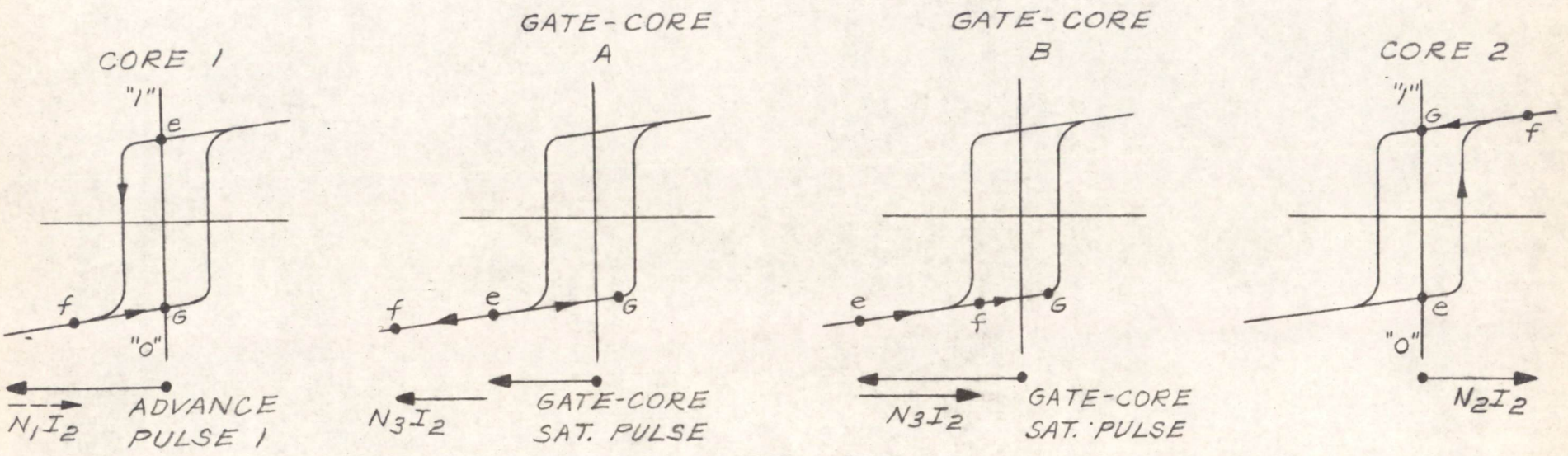


FIG. 10

TRANSFER OF "1" FROM CORE 1 TO CORE 2

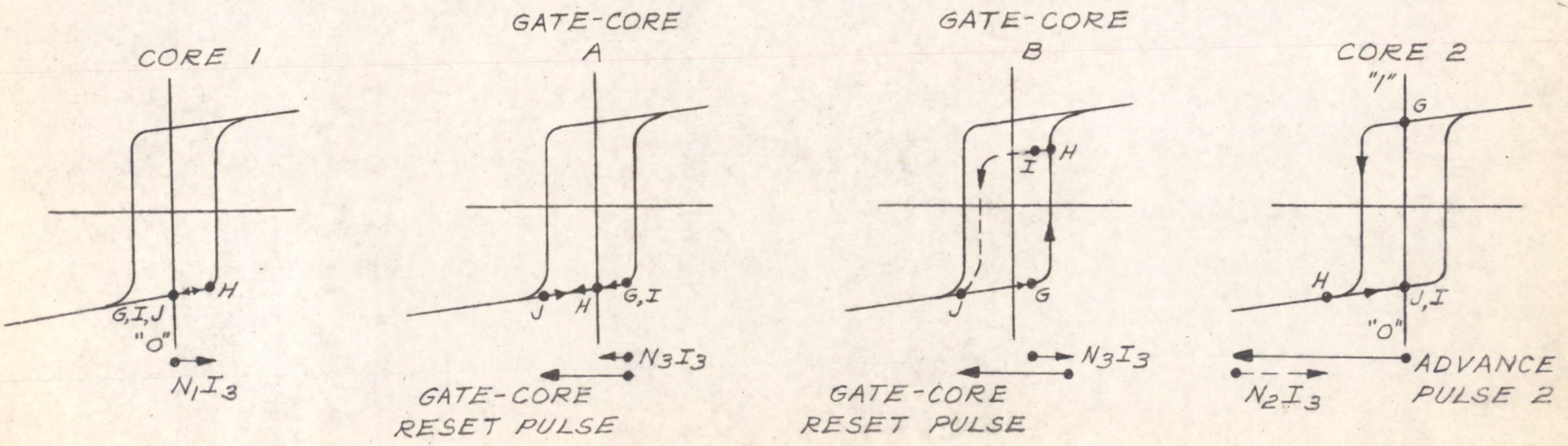


FIG. 11

READ-OUT "1" FROM CORE 2 AND RESET CORE B

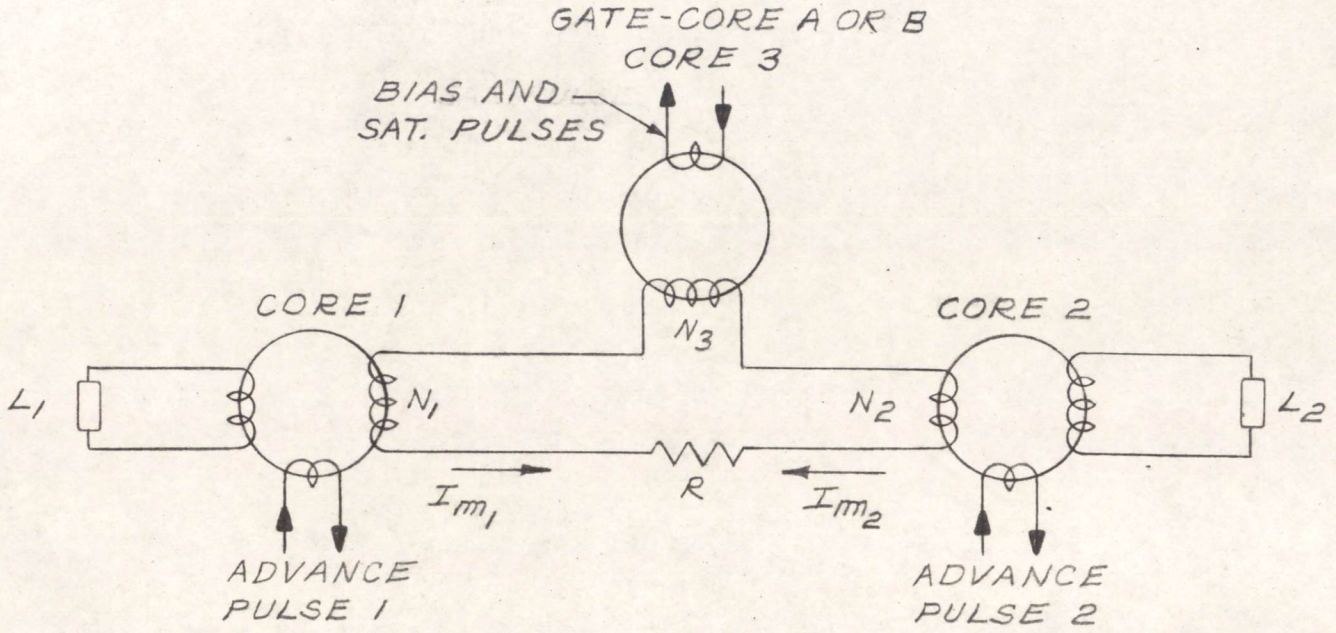


FIG. 12

GENERALIZED DIAGRAM OF CIRCUIT

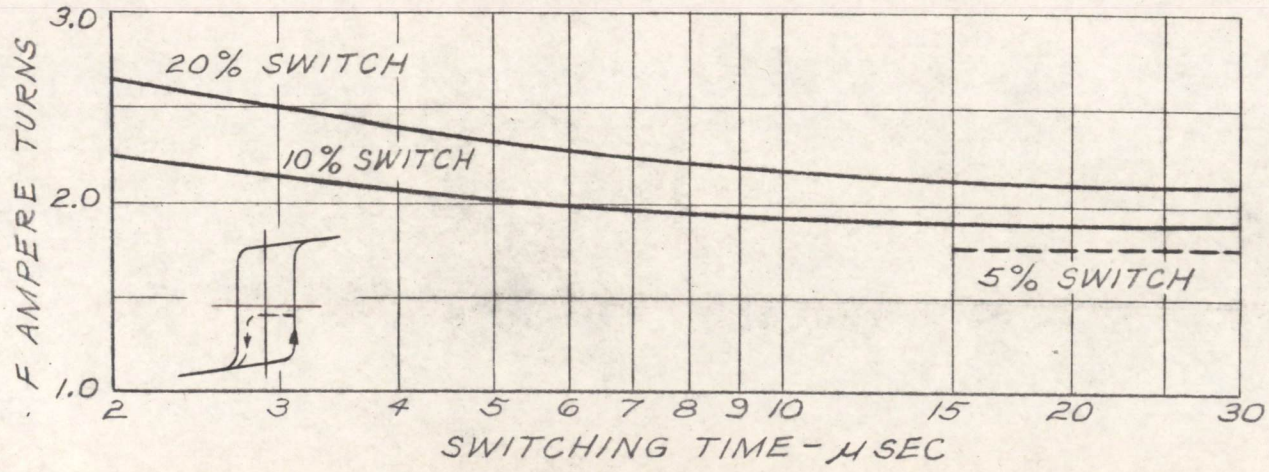


FIG. 13

F VS. SWITCHING TIME FOR PARTIAL SWITCHING
OF A FERRITE CORE MF-1118 (F-259)

Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

SUBJECT: MAGNETIC AND DIELECTRIC AMPLIFIERS

To: Jay W. Forrester

From: Dudley A. Buck

Date: August 28, 1952

Abstract: This paper was prepared at the request of Professor Arthur R. von Hippel for presentation at a summer symposium on the Theory and Applications of Dielectric Materials, September 3-12, 1952, at the Massachusetts Institute of Technology.

A. INTRODUCTION

The prime motivation behind recent magnetic and dielectric amplifier research is the quest for amplifiers which are more reliable and more rugged than vacuum-tube amplifiers. Amplifiers of increased reliability and ruggedness are needed in military and industrial control circuits, in communication circuits, and in information-handling machines.

Before describing how a piece of iron alloy or a thin sheet of barium titanate can act as an amplifier, let us briefly consider amplifiers in general. The term amplifier here refers to a power amplifier.

Power amplifiers are three terminal-pair devices (Fig. 1). One terminal pair is associated with the input, or signal to be amplified; one terminal pair is associated with the output, or load; and the third terminal pair is associated with the power supply.

The power supply contains no information, that is, it contains no meaningful fluctuating component that is to be amplified. The power supply can be a direct voltage, a sinusoidal voltage, a sequence of pulses, or any other waveform for which the amplifier is designed to operate. Vacuum-tube and transistor amplifiers normally use direct voltage power supply, while magnetic and dielectric amplifiers, by their very nature, must use alternating current or pulsed power supplies.

Power amplification, or power gain, may be defined as the average output power divided by the average input power, without reference to the average power delivered by the power supply.

The non-linearity in the electrical properties of ferromagnetic and ferroelectric materials makes possible their use in three terminal-pair devices which are capable of supplying more average power at the output terminals than is supplied to them at the input terminals. These devices are called magnetic and dielectric amplifiers. We shall consider them in this order.

B. MAGNETIC AMPLIFIERS

1. OPERATING PRINCIPLES OF MAGNETIC AMPLIFIERS

Magnetic amplifiers can be divided into two categories—those in which magnetic flux changes produce a voltage which subtracts from the output voltage, and those in which magnetic flux changes produce a voltage which adds to the output voltage. Amplifiers of the first category are called saturable inductors while those of the second category are called saturable transformers.

A large variety of circuits have been worked out for both categories, a few of which will be described here.

a. Saturable Inductors

Figure 2A illustrates a simple one-core saturable inductor. The reactance winding has been placed in series with the alternating current power supply and the output. Voltage across the reactance winding, caused by flux changes in the core, subtracts from the voltage available to the output. As long as flux in the core is changing, very little voltage is applied to the load. If, during the cycle, the core should reach saturation so that the flux can no longer change, then the voltage across the reactance winding of the core drops to zero and the full supply voltage is applied to the output.

The curve traced by the core material in the hysteresis-loop plane varies with the input (control) voltage (Fig. 2B). With zero input voltage, the hysteresis loop is asymmetrical due to the bias mmf. Input voltages which add to the effect of the bias voltage make the flux change smaller, that is, the core material is made to traverse a smaller hysteresis loop. Conversely, input voltages which subtract from the effect of the bias voltage cause the core material to go through a larger change in flux and thereby traverse a larger loop. In the two extremes, the input voltage may become so large in one direction that the core is always saturated, giving maximum output (load) voltage, or in the other direction so large that the major (largest possible) hysteresis loop is traversed each cycle, giving minimum output (load) voltage.

If the input winding and the bias winding were to be removed from the single-core saturable inductor of Figure 2, one might recognize the device as the ballast used with fluorescent fixtures to drop line voltage. One could wind an input winding on a fluorescent light ballast, making it a saturable inductor, and cause the light to shine brighter at will by applying a voltage to the input winding. This is indeed the device in common use at movie houses for turning the lights on during intermission, the d.c. sometimes being supplied by a small motor-generator set.

One cannot dim the lights by applying a voltage of opposite polarity to the input winding, however, for the core material is already traversing its largest possible hysteresis loop at zero input voltage. An input voltage of either polarity will cause the lights to shine brighter. To make possible

a lowering of the output voltage with input voltages of one polarity and an increase of the output voltage with input voltages of the opposite polarity, the bias voltage is added. Let us say that the bias voltage turns the lights half-way on. Then input voltages which assist the bias voltage will brighten the lights, while input voltages which oppose the bias voltage will dim the lights.

Bias voltage is here applied to a separate bias winding. Often a separate bias winding is not required, the bias voltage being simply added in series with the input voltage.

As a practical device, this particular circuit has the disadvantage of relatively large power losses in the input circuit. An improvement can immediately be made by connecting two such saturable inductors (Fig. 3) with their reactance windings in series aiding and their control windings in series opposition. The voltage induced in the control winding of one inductor will be opposite in sign to that induced in the control winding of the other, and the two voltages will therefore tend to cancel whenever flux is changing in both cores. The reactance windings, whose voltages add, will operate as before.

Using a straight-line approximation for the hysteresis loop of a core material having a near-rectangular loop, the voltage, current, and flux waveforms of a saturable inductor can be derived. Figure 4A shows the loaded saturable inductor at one input voltage. Figure 4B shows the same waveforms, actually observed, for a resistance-loaded saturable inductor using Deltamax cores.[1] Load current curves are given for three different input voltages. It can be seen that as the input voltage is increased, the load current increases and lasts for a greater portion of each cycle, thus increasing the power delivered to the load. The position in the a.c. cycle at which saturation occurs and load current starts to flow, measured from the start of the cycle in electrical degrees, is called the firing angle.

b. Saturable Transformers

Figure 5A illustrates a simple one-core saturable transformer. The operation is similar, but inverse in logic, to the operation of the saturable inductor just described. Here, flux changes in the core produce the output voltage rather than a voltage which subtracts from the output voltage. The asymmetrical hysteresis loops of Figure 2B can also be used to describe the operation of the saturable transformer, remembering that traversal of a large loop now means a large output voltage, while traversal of a small loop now means a small output voltage.

Neglecting the effect of the bias voltage for the moment, it will be noted that with zero input voltage, this device has maximum output voltage--in contrast to the previous device which has minimum output voltage for zero input voltage. If a movie house found that its lights were on more than they were off (e.g.--a burlesque show) then this device might well be the more practical of the two for control of the lights.

---

1. T. G. Wilson, "Series-Connected Magnetic Amplifier with Inductive Loading," NRL Report 3923, Naval Research Laboratory, Washington, D.C., Jan. 9, 1952.

A two-core scheme (Fig. 5B) can be used once again to minimize input power losses.

c. Introduction of Feedback

A portion of the output voltage is often added to or subtracted from the input voltage so as to produce positive or negative feedback. Figure 6 illustrates the effect of feedback on the operating curve of a magnetic amplifier. Negative feedback, wherein a portion of the output voltage is subtracted from the input voltage, lowers the amplification but improves the linearity of the amplifier. Linearity, while it would probably be unimportant in the case of the movie house light dimmer, is of great importance in the case of magnetic amplifiers for voice amplification and for certain control systems. Positive feedback, on the other hand, increases amplification, but reduces linearity. In many circuits, excessive positive feedback will create instability, resulting in a "snap-action" (dotted lines). Under this condition, the output voltage will actually jump between two values as the input voltage increases or decreases through a certain range.

Figure 7 illustrates one of the many possible circuit arrangements which provide feedback. A portion of the a.c. output voltage is rectified and the resulting d.c. voltage is applied to a second input winding. As the a.c. output voltage increases, so does the rectified d.c. voltage. In the case of the saturable inductor (Fig. 7A), the feedback voltage will produce positive feedback if it assists the bias voltage and negative feedback if it opposes the bias voltage. The converse is true in the case of the saturable transformer (Fig. 7B).

The feedback schemes of Figure 7 are representative of a class which provide external feedback. Whenever a portion of the output is fed back to the input via an external loop, the scheme is known as external feedback.

A somewhat simpler means for introducing feedback (Fig. 8) is called internal feedback. Here, a rectifier is used to create a uni-directional current flow in the output circuit thereby causing feedback. As the output voltage increases, the magnitude of the uni-directional current in the output circuit increases, and the feedback increases.

Internal feedback, as thus far described, would be suitable only for driving d.c. loads. Using two cores operated in push-pull (Fig. 9), either a.c. loads (Fig. 9A) or d.c. loads (Fig. 9B) can be driven. The latter circuit gives a full-wave rectification.

d. Core and Circuit Variations

More complicated magnetic cores are often encountered in practice. The magnetic amplifiers which have been described involve one or two simple magnetic cores having but a single flux path. The two-core combinations can often be combined into a single three-legged core to facilitate fabrication. In addition, three-legged and four-legged cores can be built wherein magnetic fluxes are added and subtracted in the various legs just as voltages are added.

and subtracted in windings.

Many circuit variations have also been made, some involving other types of feedback and methods for controlling feedback, some involving saturable inductors in bridges, and some involving condensers which allow certain windings to be operated near resonance. For a description of these schemes, the interested reader is referred to a survey of magnetic amplifier types made by J. G. Miles[2] and to his very excellent bibliography of magnetic amplifier devices and the saturable reactor art.[3]

Very little has been said about the nature of the input voltage. Clearly, it can be variable direct voltage shich adds to or subtracts from the effect of the bias. In the case of internal feedback, it is seen that the feedback voltage is not a direct voltage, but rather a pulsating direct voltage. In a similar manner, the input voltage can also be pulsating direct voltage.

e. Response Time of Magnetic Amplifiers

Response time of magnetic amplifiers, defined as the time for the output voltage to reach 63% of its final value following a change in input voltage, is usually a direct function of the power supply frequency, and as such is usually measured in cycles at the frequency. Storm[4] and Ramey [5, 6] have shown that response time of many saturable inductor amplifiers is a function of the input circuit resistance, the output circuit resistance, and the square of the turns ratio between the two windings. Lower resistance in the output circuit or higher resistance in the input circuit make for shorter response time. Typical response times are from 1-1/2 to 3 cycles.

f. Half-cycle Response Time Magnetic Amplifiers

Circuits have been developed which have a response time of one-half cycle. Two such circuits will be described. The first, a saturable inductor circuit, was developed by R. A. Ramey at the Naval Research Laboratory;

---

2. J. G. Miles, Types of Magnetic Amplifiers--A Survey, Engineering Research Associates, St. Paul, Minnesota, January 24, 1951

3. J. G. Miles, "Bibliography of Magnetic Amplifier Devices and the Saturable Reactor Art," A.I.E.E. Transactions, Vol. 70, pp. 2104-2123 (1951); also A.I.E.E. Technical Paper 51-388, September, 1951.

4. H. F. Storm, "Series-connected Saturable Reactor with Control Source of Comparatively Low Impedance," A.I.E.E. Technical Paper 50-123 (1950).

5. R. A. Ramey, "On the Mechanics of Magnetic Amplifier Operation," A.I.E.E. Transactions, Vol. 70, pp. 1214-23 (1951); also NRL Report 3799, Naval Research Laboratory, Washington, D.C.

6. R. A. Ramey, "On the Control of Magnetic Amplifiers," A.I.E.E. Transactions, Vol. 70, pp. 2124-28 (1951); also NRL Report 3869, Naval Research Lab., Washington, D.C.

while the second, a saturable transformer circuit, is becoming important in information-handling machines. The half-cycle response time saturable transformer is pulse-operated.

1.) Half-cycle Response Time Saturable Inductor

Ramey[5] has developed a saturable inductor circuit (Fig. 10A) which uses rectifier switches to alternately connect the magnetic core to the input circuit and the output circuit. The rectifier in the input circuit has an inverse voltage across it during one half of the power supply cycle, thereby effectively removing the input circuit from the core during that half-cycle. During the next half-cycle, the rectifier in the output circuit has an inverse voltage across it thereby effectively removing the load circuit from the core. (The power supply voltage in the input circuit must be higher than in the output circuit.)

During the half-cycle that the output circuit is removed from the core, the input circuit changes the flux in one direction. During the next half-cycle, the input circuit is removed, and the output circuit changes the flux back in the opposite direction. Because the core reaches saturation during the output half-cycle, the flux can change only as much as it changed during the input half-cycle. This amount is determined by the power supply voltage in the input circuit minus the input voltage. If there is zero input voltage, the input half-cycle will change the flux by the maximum amount, and the output terminals will receive the lowest possible voltage. In the other extreme, if the input voltage equals the power supply voltage, no flux change will occur on either half-cycle and the output terminals will receive the full supply voltage. Because of this isolation between input and output, this amplifier will always have a half-cycle response time.

2.) Half-cycle Response Time Saturable Transformer--A Pulse-Operated Device

Figure 10B illustrates a half-cycle response time saturable transformer derived from the preceding circuit. During the output half-cycle, the flux change in the core has been left unchanged in direction. The output current must therefore flow in the opposite direction. Therefore, the series rectifier must be reversed. This change brings about the very interesting fact that the rectifier in the output circuit need no longer be cut off by the half-cycle of supply voltage. The input voltage, as it causes flux changes in the core, induces a reverse-polarity voltage in the output circuit. The input circuit, acting along, thus effectively removes the output circuit during the input half cycle. We can use a power supply which consists of a sequency of uni-directional pulses to drive the output.

Since the change from inductor to transformer inverts the control logic, the power supply voltage and input voltage need no longer be differenced, so that the half-cycle from which the input was formerly subtracted can be eliminated. The power supply in charge of cutting off the input circuit

---

5. Ramey, loc. cit.

can also consist of a sequence of uni-directional pulses.

An amplifier of this type with a modification in the input circuit (Fig. 10C), which eliminates the input power supply at the expense of somewhat higher losses, is in wide use in information-handling machines. One large digital computer is being built almost entirely with magnetic amplifiers of this type.

2. MATERIALS AND PERFORMANCE OF MAGNETIC AMPLIFIERS

Magnetic amplifiers have been used most extensively with 60 cycles-per-second and 400 cycles-per-second alternating current power supplies. They are performing such tasks as the control of d.c. motor fields and armatures, control of a.c. motors, control of rectifier power supplies, etc.

Recently, in connection with voice amplification and computer service, power supply frequencies have been increasing. Some magnetic amplifiers are in operation with power supply frequencies in the region between 10^4 and 10^5 cycles per second.

In the low frequency applications, magnetic amplifiers are commercially available which deliver anywhere from a few watts to a few hundred watts of output power with amplification as high as 10,000. They have proven themselves rugged and reliable in mobile service. In this field, they are being used in increasing numbers both in conjunction with and as substitutes for vacuum-tube amplifiers.

It was implicit in the foregoing description of magnetic amplifier circuits that the core material should have a saturable region--one in which the flux changes but little for large changes in mmf. It may also have been evident that it is highly desirable to have a sharp break between the saturated region and the unsaturated region as mmf is increased. These desirable properties are summarized by stating that the hysteresis loop should be as nearly rectangular as possible. A rectangular hysteresis loop material, in addition, will give the greatest possible linearity in most of the circuits described.

Chiefly responsible for the recent widespread use of magnetic amplifiers are the recently available core materials with rectangular hysteresis loops. Once a laboratory curiosity, rectangular hysteresis loops are now built into commercially available core materials by processes of grain-orientation and domain-orientation. The following widely used alloys are listed in Table I with certain of their trade names and approximate compositions.

---

7. F. Benjamin, "Improvements Extended Magnetic Amplifier Applications," Electronics, Vol. 25, No. 6, June, 1952.

TABLE I

MAGNETIC MATERIALS OF CURRENT INTEREST FOR MAGNETIC AMPLIFIERS AND SATURABLE INDUCTORS
Hipersil, Trancor XXX, Silectron Type C (Grain-oriented Silicon steels — 3% Si, 97% Fe)
Mu-Metal (75% Ni, 2% Cr, 5% cu, 18% Fe)
Supermalloy (79% Ni, 5% Mo, 15% Fe)
4-79 Mo Permalloy (79% Ni, 4% Mo, 16% Fe)
Deltamax, Hipernik V, Orthonik, Permeron, Orthonol (50% Ni, 50% Fe)

These core materials are generally available in the form of thin (.005" or less) ribbons of any desired width. Rectangular hysteresis loop properties are exhibited if the mmf is applied in the direction for which the grains or domains have been oriented. This is usually in the plane of the ribbon parallel to the edges, so that fabrication of a core involves simply winding the ribbon as on a spool. Thin insulating layers separate the wraps of tape so as to minimize eddy currents.

As one leaves the low frequency applications and increases the power supply frequency, one finds that the rectangular hysteresis loop with its steep sides causes eddy current effects to show up at relatively low frequencies. Eddy currents produce a "magnetic skin effect" which limits the depth of flux penetration in a core driven with alternating current, thereby effectively lowering the cross-sectional area of the magnetic flux path. The remedy for this ill effect, with a given material, is to use thinner and thinner ribbon as the frequency is raised. Ribbons as thin as .000125" (1/8 mil) are available (1/4 mil seems to be the thinnest ribbon in common use).

An alternate solution is to use a different material--preferably one with a higher volume resistivity. Unfortunately, the materials already mentioned all have about the same volume resistivity. A radically new core material whose volume resistivity is many powers of ten higher than the materials mentioned has recently become available with a rectangular hysteresis loop. This new material is a ferrite, composed of oxides of iron and other metals,

which is manufactured by a ceramics process. Figure 2 was made from a photograph of one of these rectangular hysteresis loop ferrites. One year ago, it was not thought possible to make a ferrite--a polycrystalline ceramic--with such a rectangular hysteresis loop. It is not yet as rectangular, however, as the metallic alloys, and it has the disadvantages of relatively low saturation induction and relatively high coercive force (short, fat hysteresis loop). This material does promise, however, to extend the frequency range of magnetic amplifiers.

As one considers the computer applications of magnetic amplifiers, pulse operated, one finds it more natural to think of the eddy-current phenomena in the time domain rather than in the frequency domain. Switching time is then defined as the time required for the material to traverse one-half of its hysteresis loop. Switching time is a function of the applied mmf. With relatively high mmf's (25 times the coercive force) switching times as short as 10^{-7} sec have been observed for 1/4-mil 4-79 Mo Permalloy ribbon-wound cores.

Heating of the core material due to hysteresis loss is a second factor which limits the frequency at which magnetic amplifiers can be operated. Heating very often results in a change in the magnetic properties of the core material. With the metallic cores, this change is often irreversible upon cooling. With ferrite cores, changes in the magnetic properties due to heating seem to be reversible, but the magnetic properties often vary over a somewhat narrower temperature range.

C. DIELECTRIC AMPLIFIERS

A power amplifier can, as we have seen, be built which utilizes for its operation the non-linear electrical properties of a ferromagnetic material. In a similar manner, the non-linear electrical properties of a ferroelectric material can be used to build a power amplifier. This latter type, called a dielectric amplifier, is somewhat the electrical dual of the magnetic amplifier.

Before describing dielectric amplifier circuits and their operation, let us consider briefly the electrical properties of ferroelectric materials.

1. FERROELECTRIC MATERIALS

Ferromagnetic materials owe their non-linearity primarily to the existence within the material of domains of permanent magnetic dipoles. Recently, materials have been discovered within which domains of permanent electric dipoles exist.[8,9,10] These materials, named ferroelectrics, exhibit

---

8. A. von Hippel and co-workers, NDRC Reports 14-300 (August, 1944) and 14-540 (1945).

9. A. von Hippel, "Ferroelectricity, Domain Structure, and Phase Transitions in Barium Titanate," Laboratory for Insulation Research Technical Report IIVII, Massachusetts Institute of Technology, March, 1950; Reviews of Modern Physics, Vol. 22, No. 3, pp. 221-237, (July, 1950).

10. D. A. Buck, "Ferroelectrics for Digital Information Storage and Switching," Digital Computer Laboratory Report R-212, M.I.T., June 5, 1952.

hysteresis loops in the D-E plane similar to those of the ferromagnetic materials in the B-H plane (Fig. 11).

Ferroelectricity is exhibited by three groups of materials whose representatives are the tetrahydrate of potassium-sodium tartrate (rochelle salt), dihydrogen potassium phosphate and arsenate, and barium titanate. Materials of the latter group, represented by barium titanate, have been considered for dielectric amplifier use. Unlike the others, barium titanate can be prepared in the form of a rugged ceramic which exhibits ferroelectricity over a wide temperature range, and which, when compounded with other titanates, can be tailored to a wide variety of electrical porperties. This third group has recently been augmented by the addition of the ferroelectric tantalates, niobates, and tungstates, most of which have yet to be tested for dielectric amplifier use.

The hysteresis loops of Figure 12 are those of ceramic barium titanate (Glenco body "X-18"). They are not nearly as rectangular as the hysteresis loops of the magnetic cores used for magnetic amplifiers. If one desires hysteresis loops of greater rectangularity, the only available method at present is to abandon ceramics and grow single crystals of barium titanate.

As with the ferromagnetic materials, the hysteresis loop disappears as the material is heated through the Curie temperature, and the electric dipole domains within the material are no longer able to align themselves spontaneously against the randomizing action of thermal vibrations (Fig. 12).

2. DIELECTRIC AMPLIFIER CIRCUITS

Considering dielectric amplifiers as the dual of magnetic amplifiers, interchanging everywhere the words voltage and current, we immediately encounter the rather troublesome fact that ferroelectric condensers are one terminal-pair devices while magnetic cores with their windings can be multi-terminal-pair devices. The engineer recognizes this as a distinct handicap, for he must often incorporate additional circuit elements--linear condensers and linear transformers--to obtain multiple connections to the ferroelectric, d-c isolation between terminal pairs, and impedance matching.

In principle, there are two basic dielectric amplifier circuits just as there are two basic magnetic amplifier circuits--one in which the current through the dielectric adds to the output current, and one in which current through the dielectric subtracts from (shunts) the output current.

A simple dielectric amplifier of the first type (Fig. 13A) can be made by simply replacing the reactance winding of the saturable inductor of Figure 2A with a saturable condenser--a ferroelectric condenser--and then describe the device somewhat as we did the saturable transformer of Figure 5A. The saturable condenser is in series with the power supply and the load, as was the saturable inductor. It passes a current which contributes to the load current whenever its charge changes. Therefore, the output power is greatest when the ferroelectric condenser traverses the largest possible hysteresis loop and smallest when the ferroelectric condenser is biased into a saturated region.

An improvement with regards to lowering of input power can be made by using more than one ferroelectric condenser. One such circuit (Fig. 13B) locates the ferroelectric condensers in two legs of a bridge and linear condensers in the other two legs. Unbalance in the bridge due to input voltage causes an output current to flow.

Quite commonly, dielectric amplifiers incorporate linear inductances which allow the circuit to operate near resonance (Fig. 14). One such mode of operation is obtained when the resonant circuit has too large a capacitance to resonate at the power supply frequency with zero input voltage. An increase of the input voltage drives the ferroelectric condensers towards saturation, effectively lowering their capacity and bringing the circuit into tune. This amplifier is of the second type; the large capacity effectively shunting the output during quiescence.

Dielectric amplifiers have been built which operate with power gains greater than 50 using power supply frequencies in the 2 megacycle region.[8] Power output levels, however, are in the milliwatts.

A dearth of suitable materials seems to be the one thing which is holding back the full scale development of dielectric amplifiers. The ceramics, inexpensive and rugged, do not exhibit saturation regions as sharply defined as is desired. Grown single crystals, while they have most of the desirable characteristics, are quite expensive.

In spite of this fact, the future of dielectric amplifiers is indeed exciting. The packaging possibilities alone offer something that is not possible with magnetic circuits, for nature made magnetic fields divergenceless so that closed paths--rings or shells--are needed for the lines of flux, whereas electric fields can start and stop on any material within which we can mobilize charge carriers. We can therefore think of many small ferroelectric condensers being fabricated side by side on a single thin sheet of ceramic or grown crystal barium titanate where fabrication involves such methods as silk-screening, evaporation, or photo-engraving. A multi-position switch[9] has been built using saturable ferroelectric condensers which illustrates one such possibility.

D. POWER SUPPLY REQUIREMENTS

Some general statements can be made regarding the various magnetic and dielectric amplifier circuits and their relationship to the relative impedances of the power supply and the load on the output terminals.

We can define a "quasi-voltage" source as a power supply whose internal impedance is lower than that of the device to which it supplies

---

8. H. Urkowitz, "A Ferroelectric Amplifier," Philco Report No. 199-M, Philco Research Division, Philadelphia, Pennsylvania.

9. D. A. Buck, "A Ferroelectric Switch", M.I.T. Digital Computer Laboratory Report E-460, April 16, 1952.

power, and a "quasi-current" source as one whose internal impedance exceeds that of the device to which it supplies power.

Maximum power is delivered to a load from a source when the respective impedances are equal. On either side of this matched condition, the power transferred to the load drops off. One can see immediately that if a load is driven by a quasi-voltage source, additional power is requested from the source by lowering the load impedance. If, on the other hand, the load is driven by a quasi-current source, one must increase the load impedance to request more power from the source.

One notices that in the saturable inductor circuit of Figure 2A more power is requested from the power supply by lowering the impedance of the circuit. One should therefore use a quasi-voltage source to drive such a circuit. Indeed, one sees that a true current source could not be used to drive the circuit because the current in the load would not be controlled by the saturable inductor; it would always be the same regardless of input voltage.

In a like manner, one sees that the saturable transformer circuit of Figure 5A needs a quasi-current source for a power supply. If driven by a true voltage source, the driven winding would demand infinite current when the core material saturates.

The dielectric amplifiers described in Figure 13 require quasi-voltage sources since their impedance lowers in order to demand more power. If driven by a true current source, one sees that at saturation an infinite voltage would appear across the terminals of the ferroelectric condenser.

In the dielectric amplifier circuit of Figure 14, the ferroelectric condenser shunts the load. With a quasi-current source power supply, the circuit impedance increases at saturation thereby demanding more power.

These four basic circuits are summarized in shorthand form in Figure 15. The saturable transformer is drawn with only its magnetizing inductance showing; the leakage inductance and the usual ideal transformer are omitted.

In the literature, most emphasis has been given to the two schemes which require quasi-voltage sources. This seems quite reasonable since most power supplies are quasi-voltage sources; the commercial 60-cycle power distribution is a quasi-voltage source.

More often than not, one has a number of magnetic or dielectric amplifiers all operating from a common power supply. Minimum interaction between the amplifiers is usually desired. This dictates that amplifiers which require quasi-voltage sources be connected to parallel, while amplifiers which require quasi-current sources be connected in series.

E. CONCLUSION

This paper has been written as an introduction to magnetic and dielectric amplifiers, both of which are being investigated in the quest for new and more suitable components for communication, computation and control.

Signed Dudley A. Buck
Dudley A. Buck

Approved DRB
David R. Brown

DAB/jk

Drawings attached:

A-51945	Figure	1
A-51948	"	2
A-51962	"	3
A-51970	"	4
A-51983	"	5
A-51985	"	6
A-51972	"	7
A-51971	"	8
A-51980	"	9
A-51987	"	10
A-50545	"	11
A-51351	"	12
A-52312	"	13
A-52393	"	14
A-52313	"	15

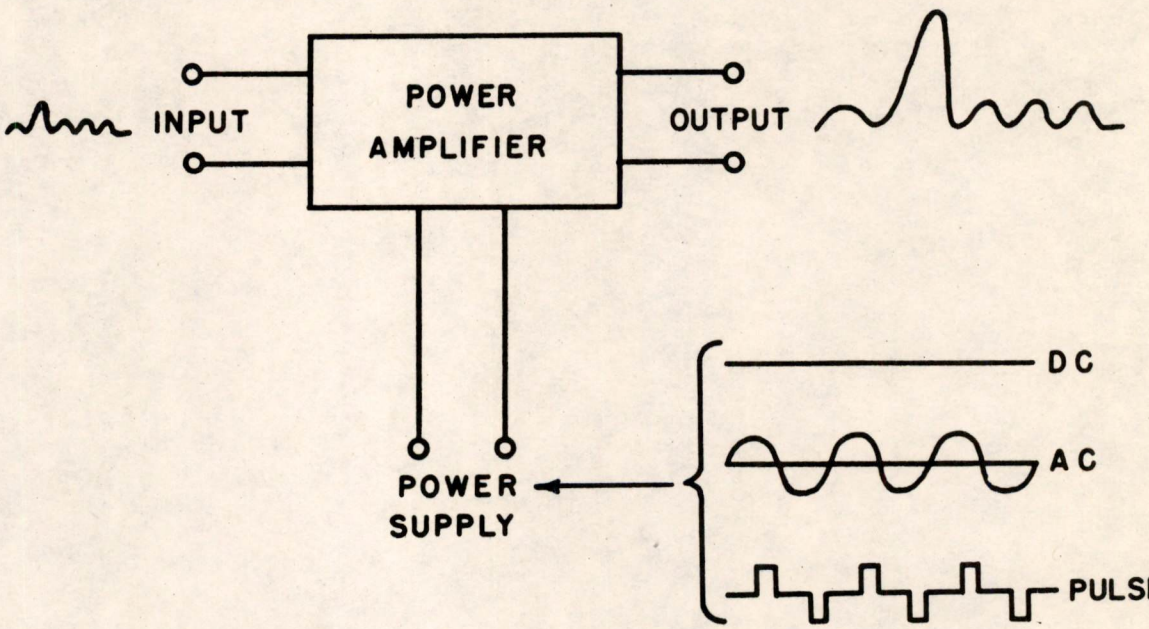
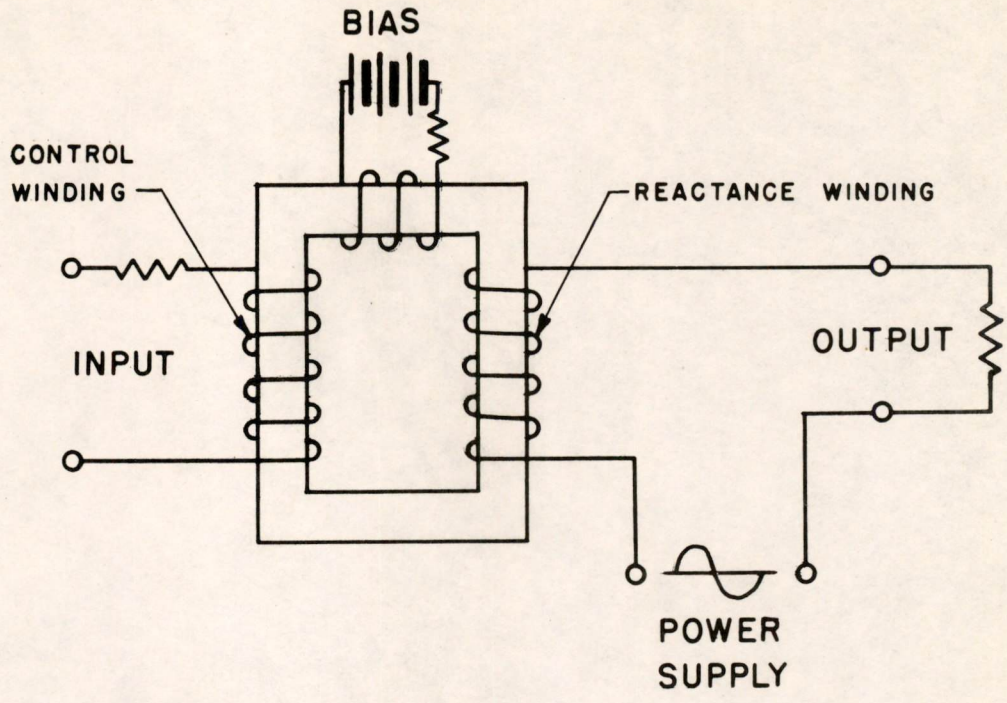


FIG. 1

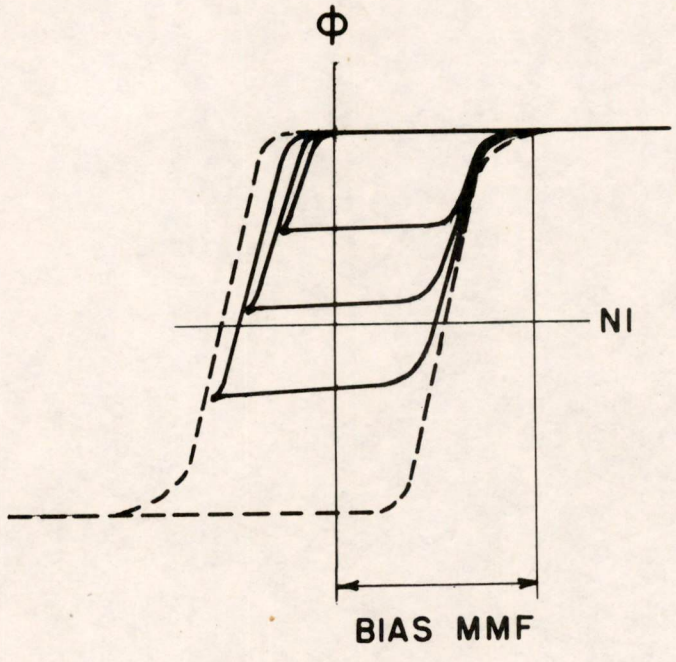
POWER AMPLIFIER

A THREE TERMINAL-PAIR DEVICE

A-51945



A. SINGLE-CORE SATURABLE INDUCTOR



B. ASYMMETRICAL HYSTERESIS LOOPS OF A RECTANGULAR-LOOP FERRITE CORE MATERIAL MF-1118(259)

FIG. 2

SATURABLE INDUCTOR

A-51948

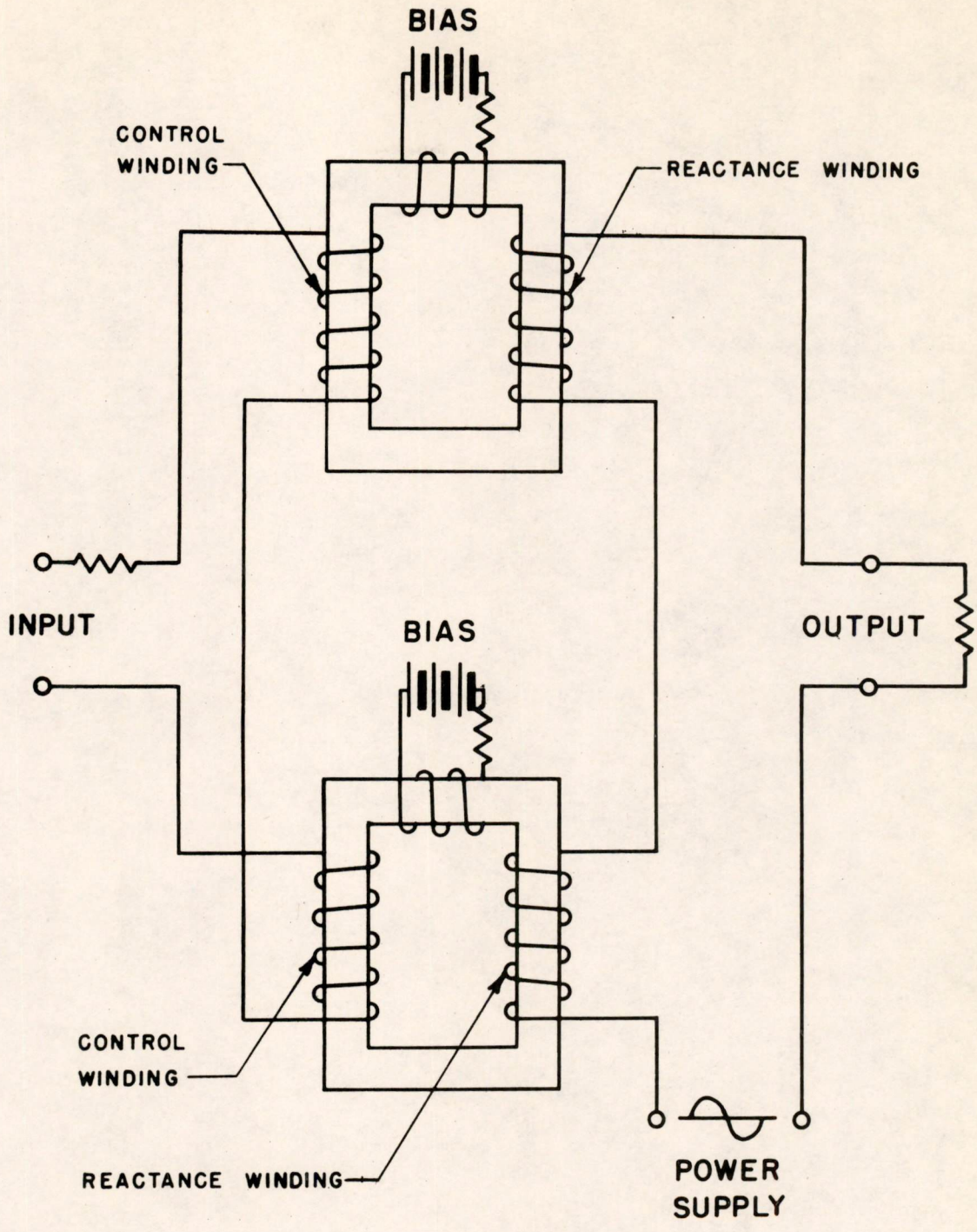
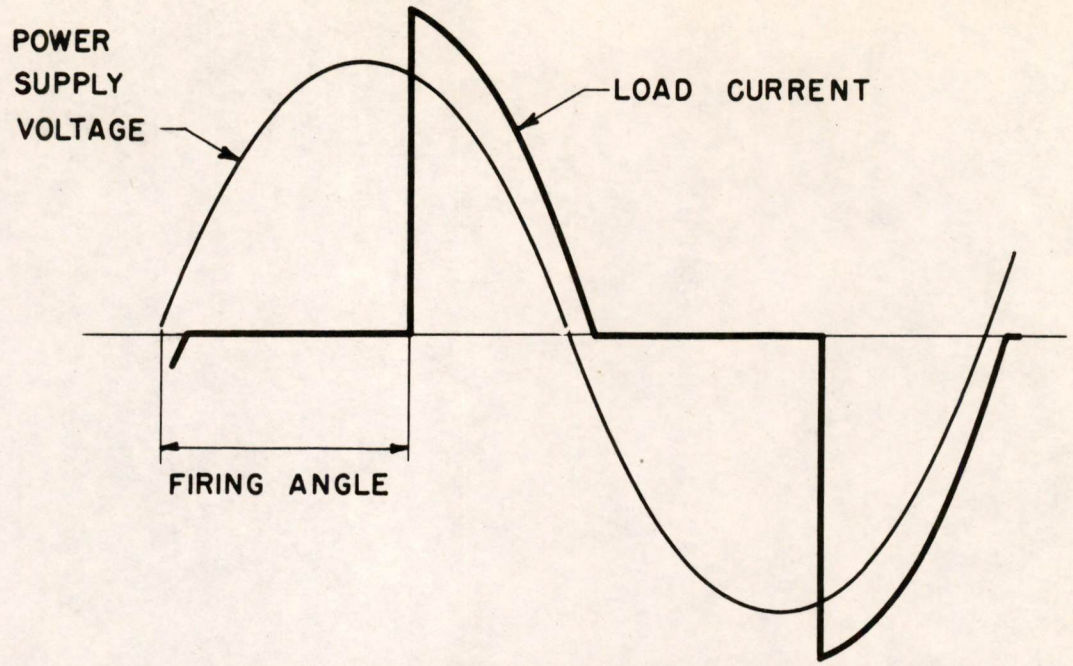


FIG. 3

TWO-CORE SATURABLE INDUCTOR

A-51962



A. DERIVED WAVEFORMS OF RESISTANCE—LOADED SATURABLE INDUCTOR USING IDEALIZED CORES

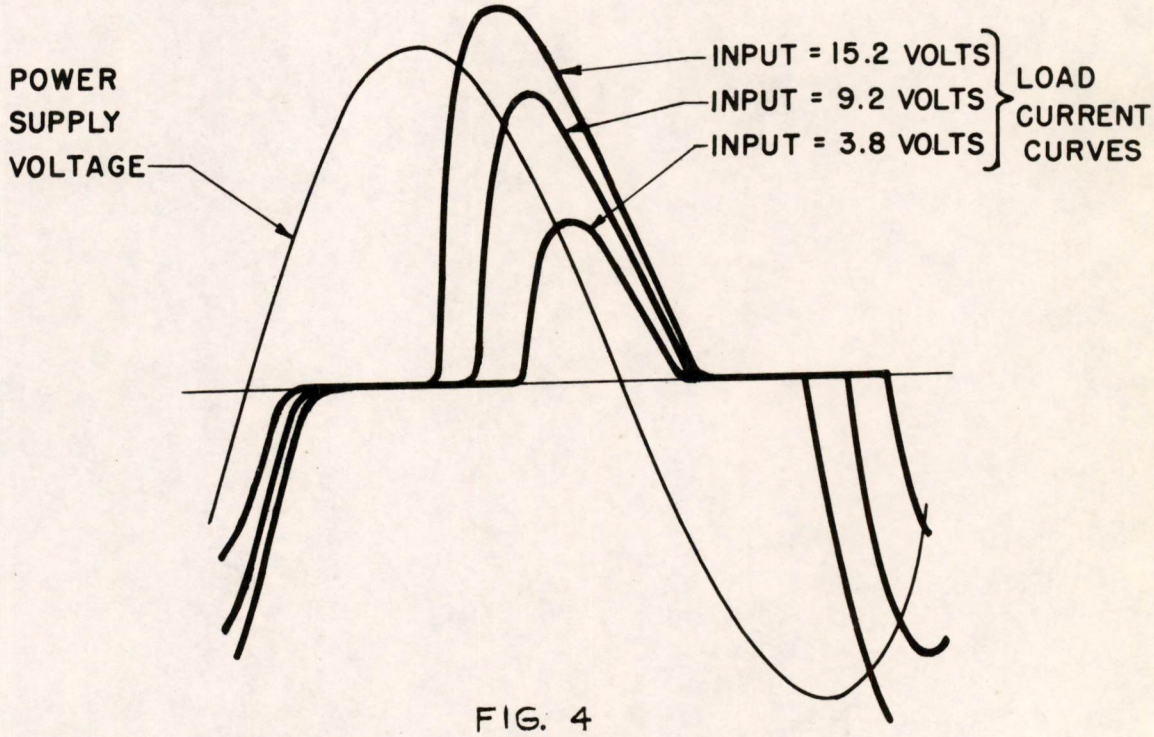
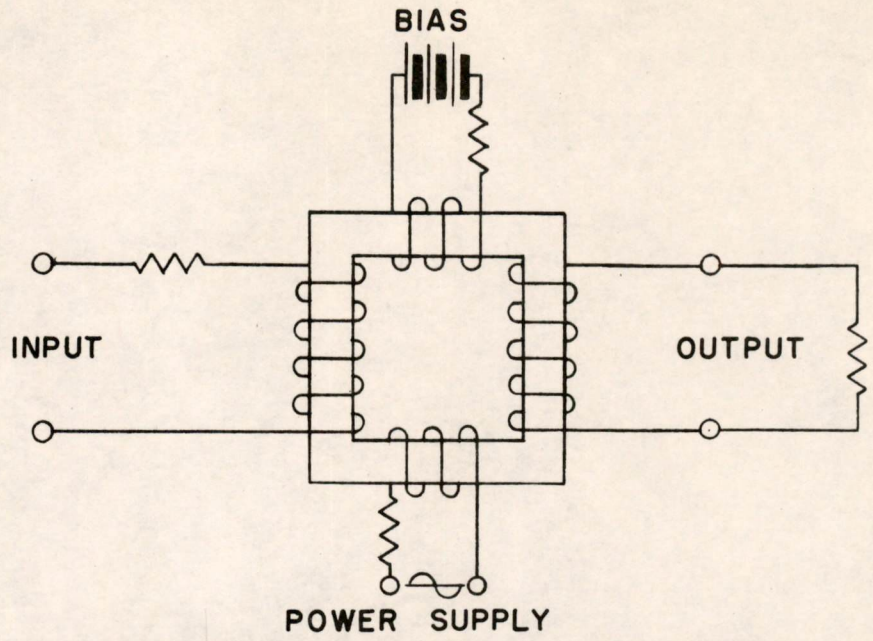


FIG. 4

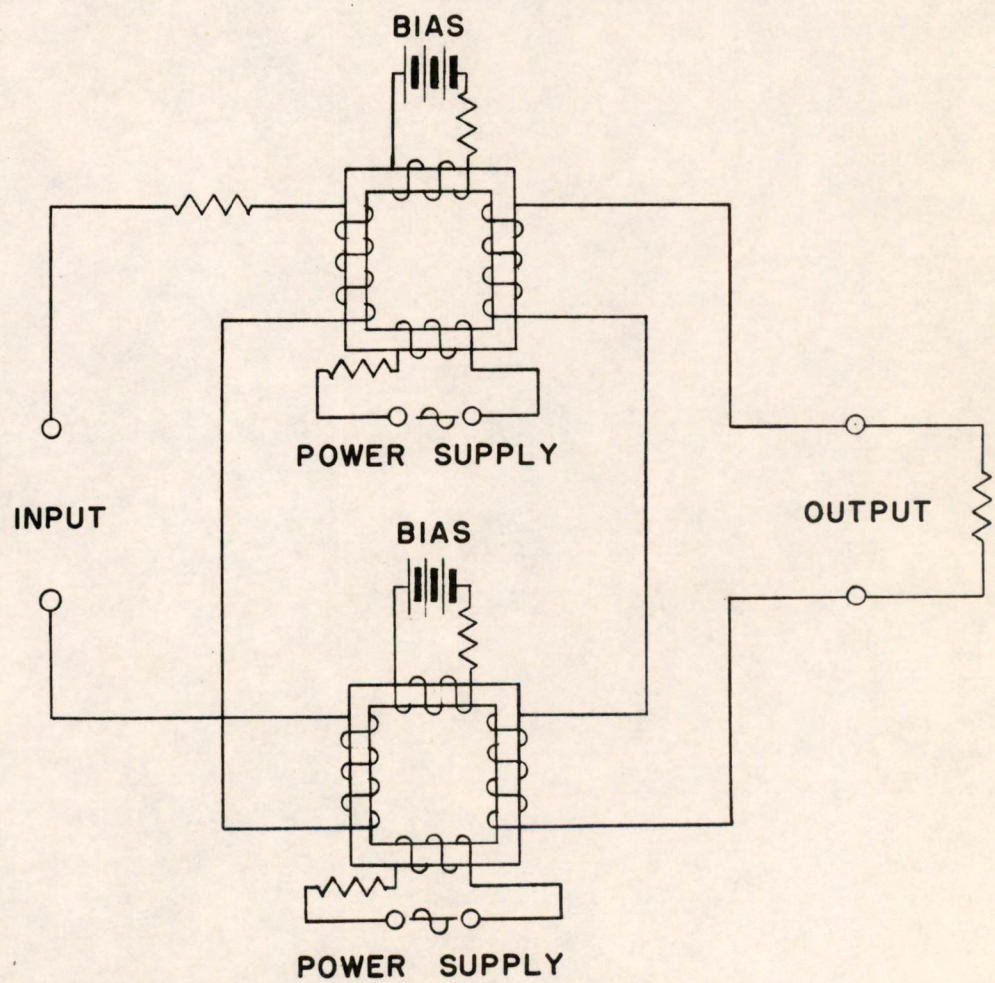
B. OBSERVED WAVEFORMS OF RESISTANCE—LOADED SATURABLE INDUCTOR USING DELTAMAX CORES (AFTER T.G. WILSON)

SATURABLE INDUCTOR WAVEFORMS

A-51970



A. SINGLE-CORE SATURABLE TRANSFORMER



B. TWO-CORE SATURABLE TRANSFORMER

FIG. 5

SATURABLE TRANSFORMER

A-51983

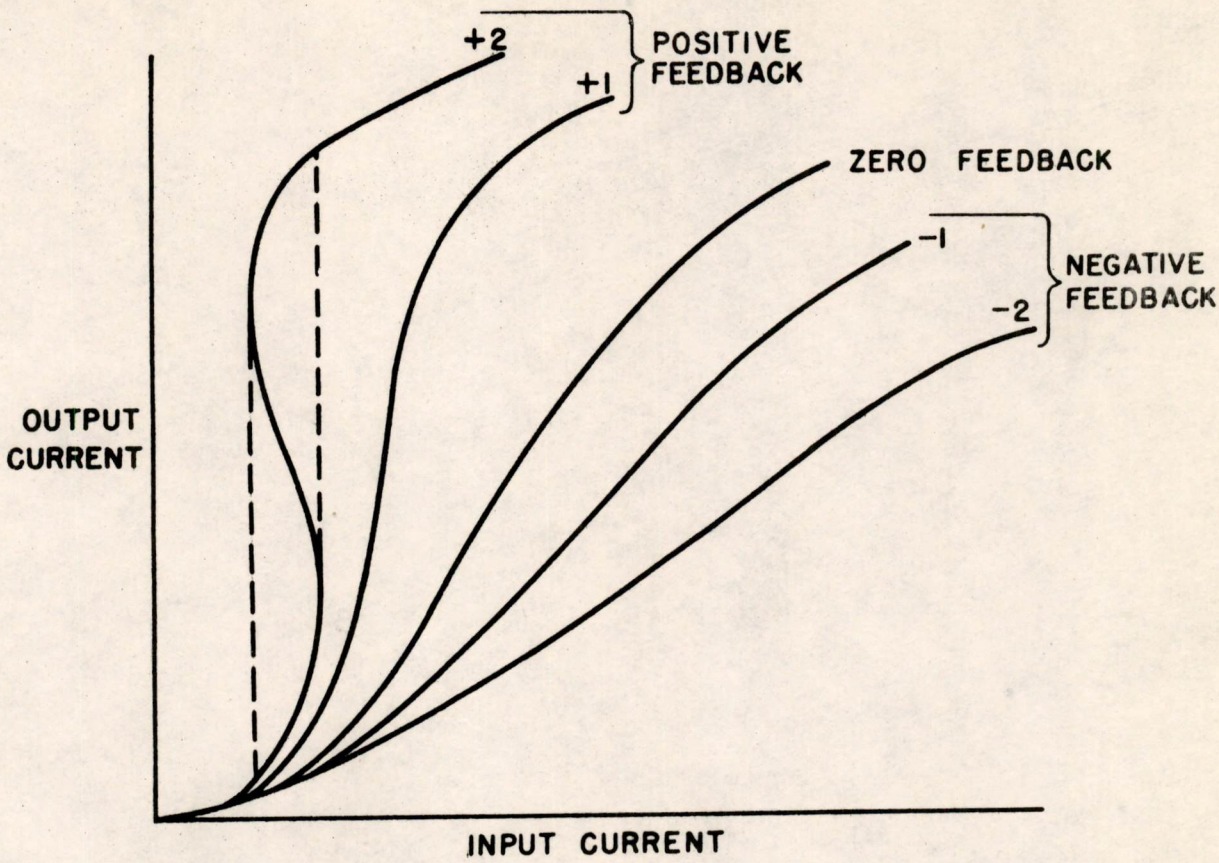
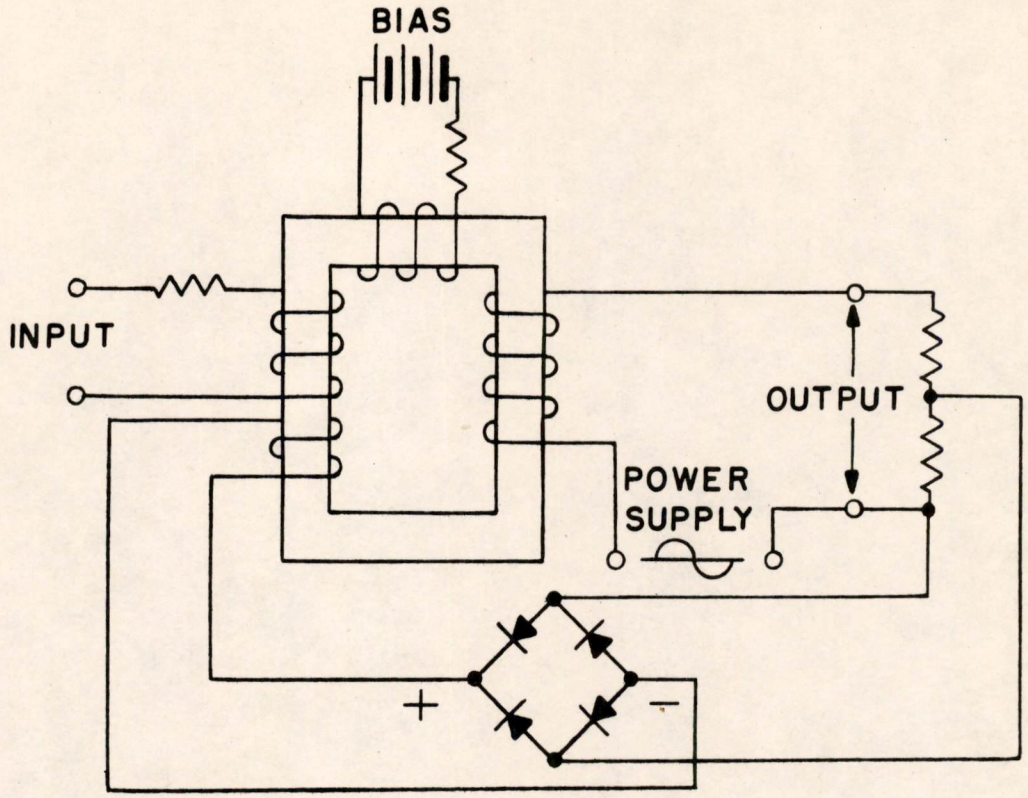


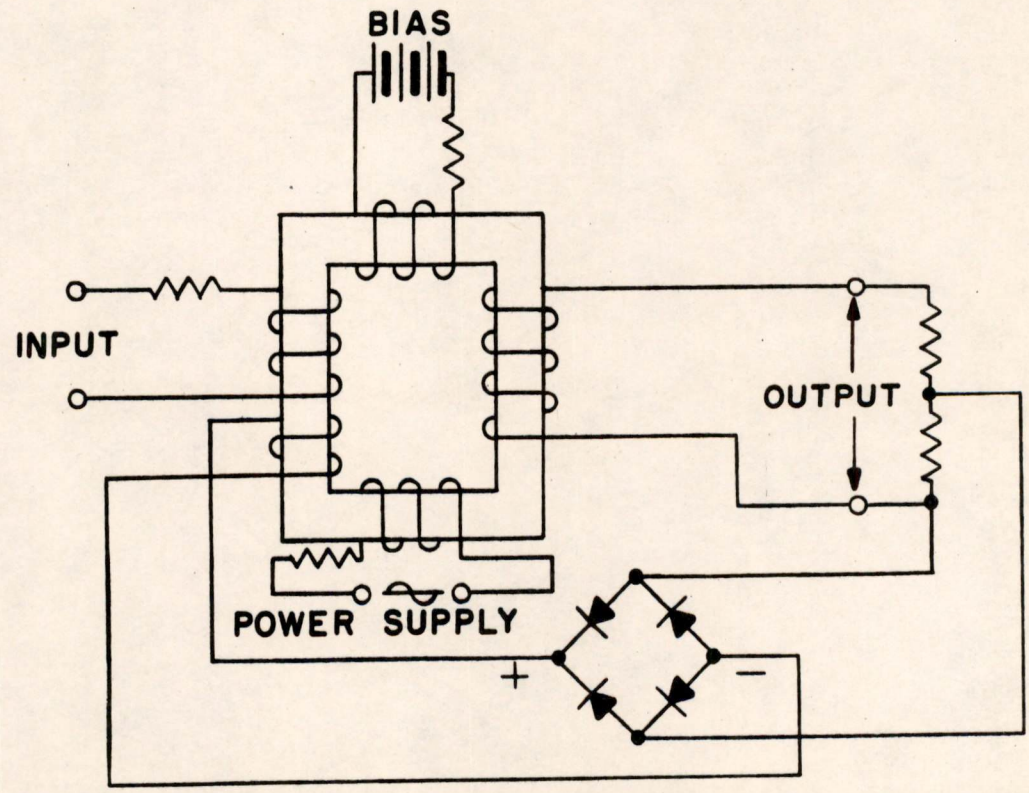
FIG. 6

EFFECT OF FEEDBACK ON MAGNETIC AMPLIFIER OPERATION

A-51985



A. SATURABLE INDUCTOR WITH EXTERNAL FEEDBACK

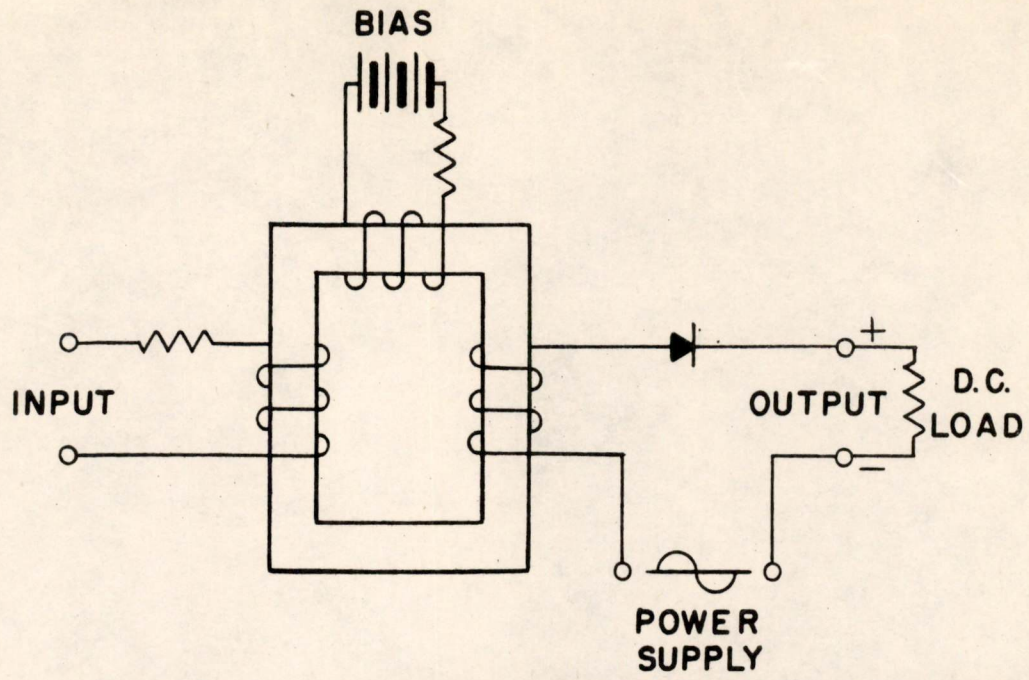


B. SATURABLE TRANSFORMER WITH EXTERNAL FEEDBACK

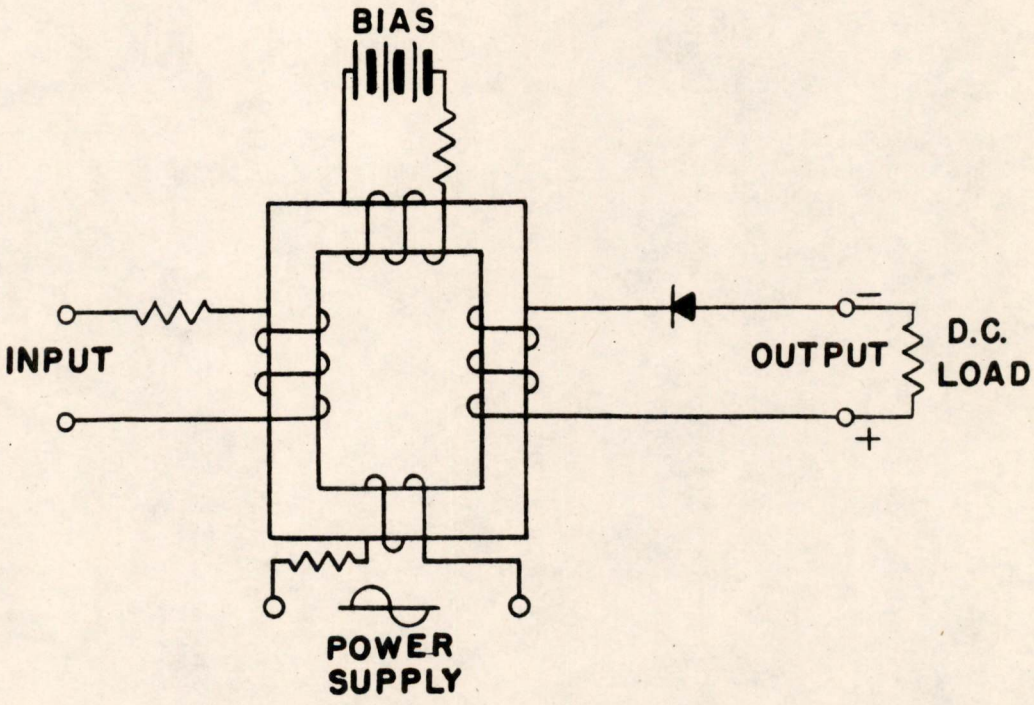
FIG. 7

EXTERNAL FEEDBACK

A-51972



A. SATURABLE INDUCTOR WITH INTERNAL FEEDBACK

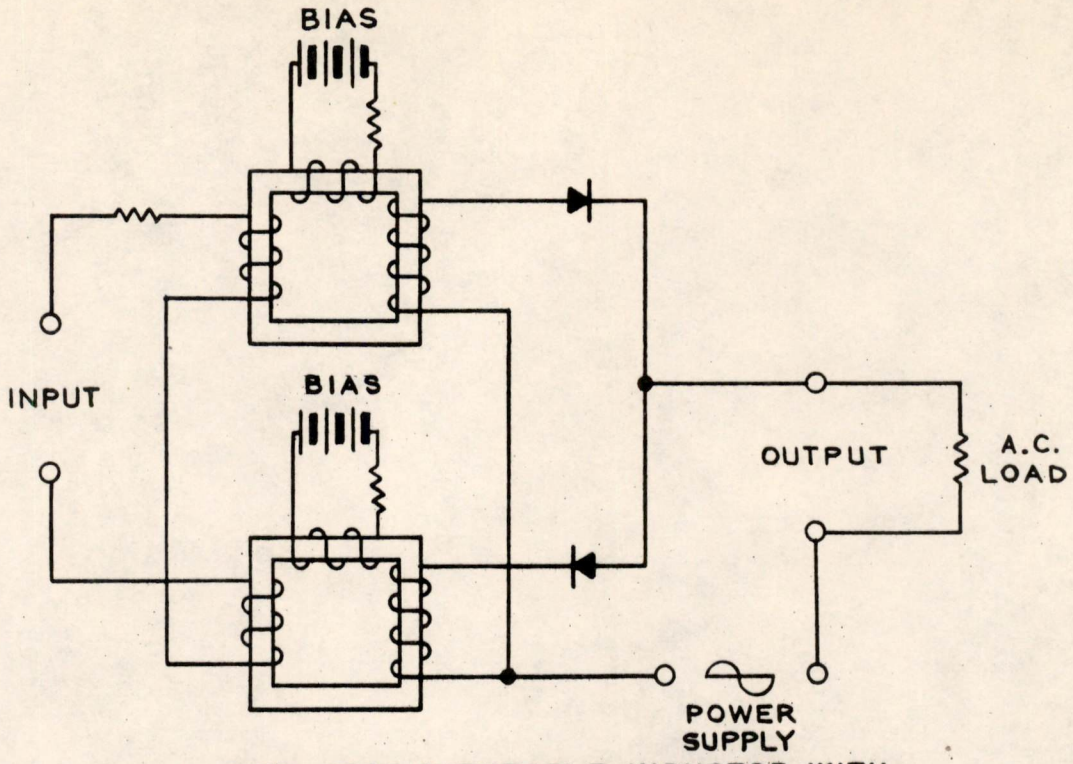


B. SATURABLE TRANSFORMER WITH INTERNAL FEEDBACK

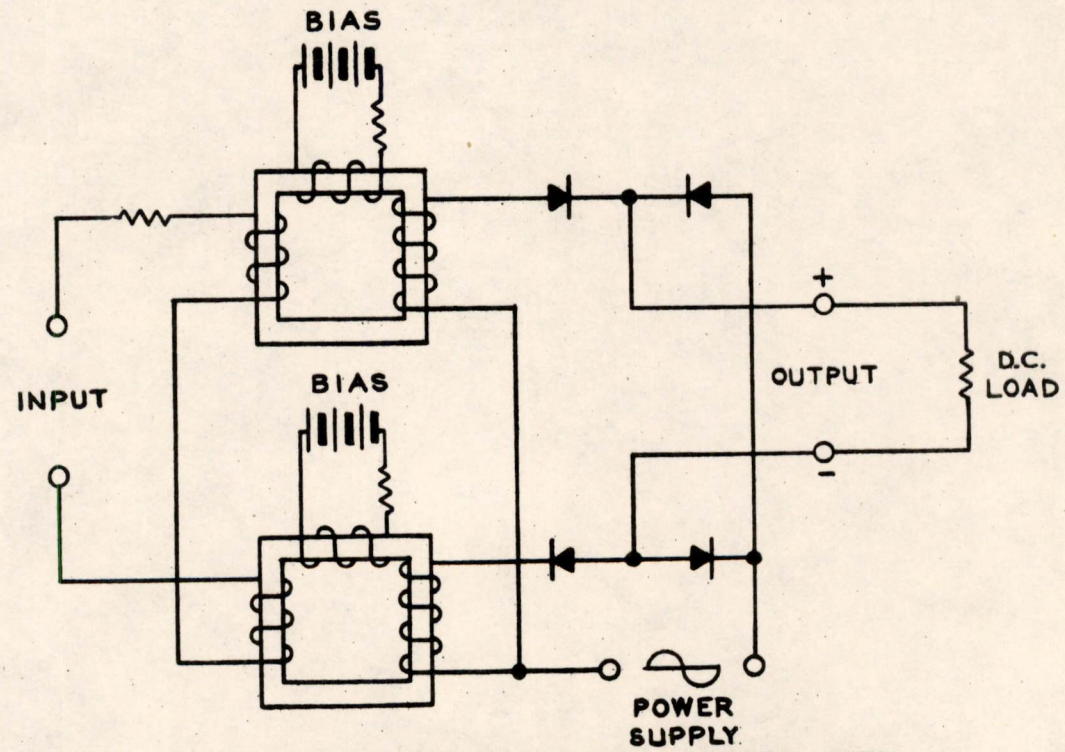
FIG. 8

INTERNAL FEEDBACK

A-51971



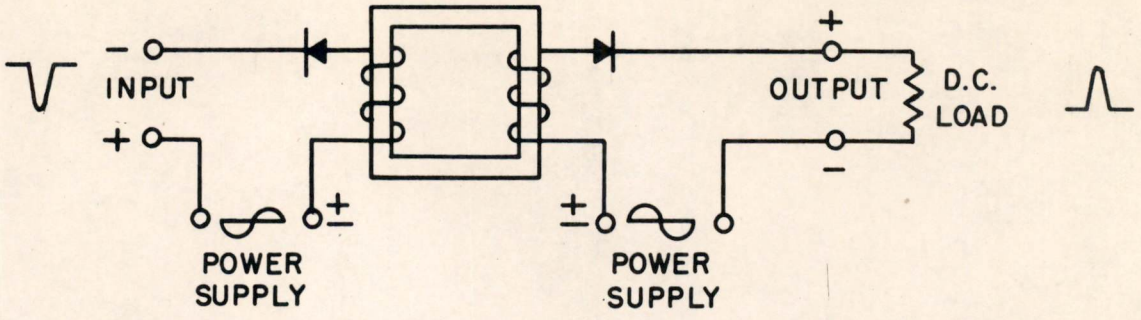
A. TWO-CORE SATURABLE INDUCTOR WITH INTERNAL FEEDBACK FOR A.C. LOAD



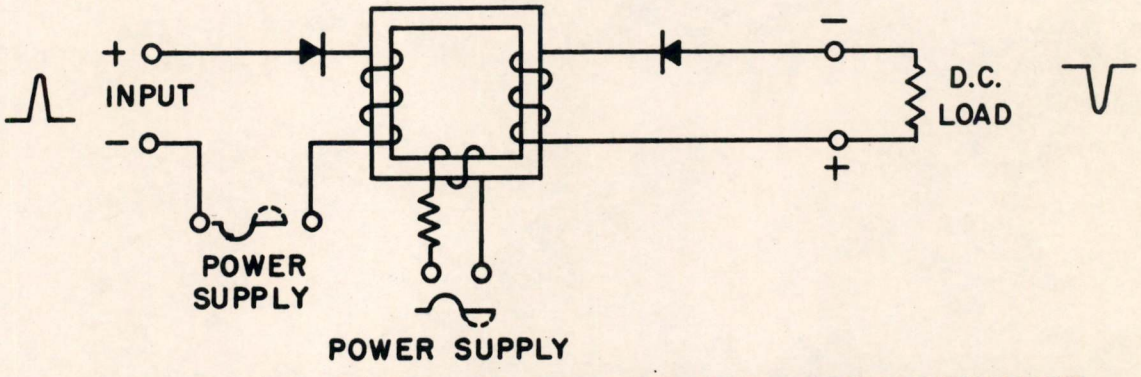
B. TWO-CORE SATURABLE INDUCTOR WITH INTERNAL FEEDBACK FOR D.C. LOAD

FIG. 9 TWO-CORE SATURABLE INDUCTORS WITH INTERNAL FEEDBACK

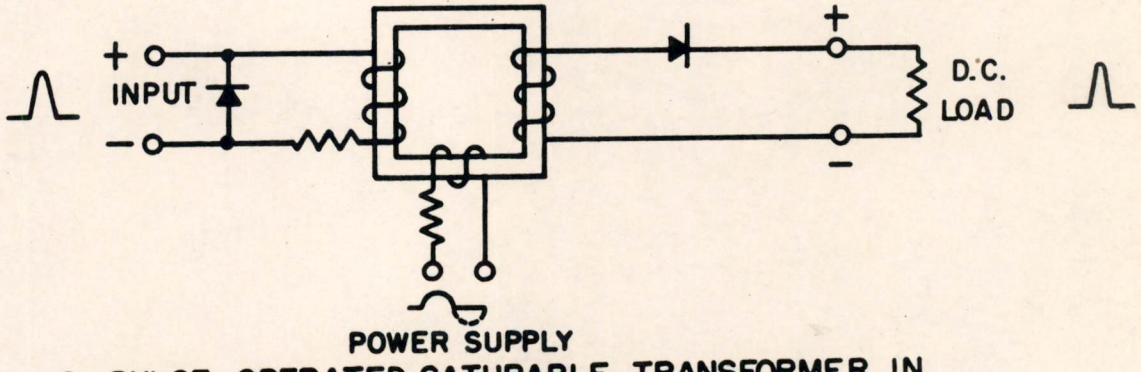
A-51990



A. HALF-CYCLE RESPONSE TIME SATURABLE INDUCTOR
(AFTER RAMEY)



B. HALF-CYCLE RESPONSE TIME SATURABLE TRANSFORMER
(PULSE-OPERATED)



C. PULSE-OPERATED SATURABLE TRANSFORMER IN
COMPUTER SERVICE

FIG. 10
HALF-CYCLE RESPONSE TIME CIRCUITS

A-51987

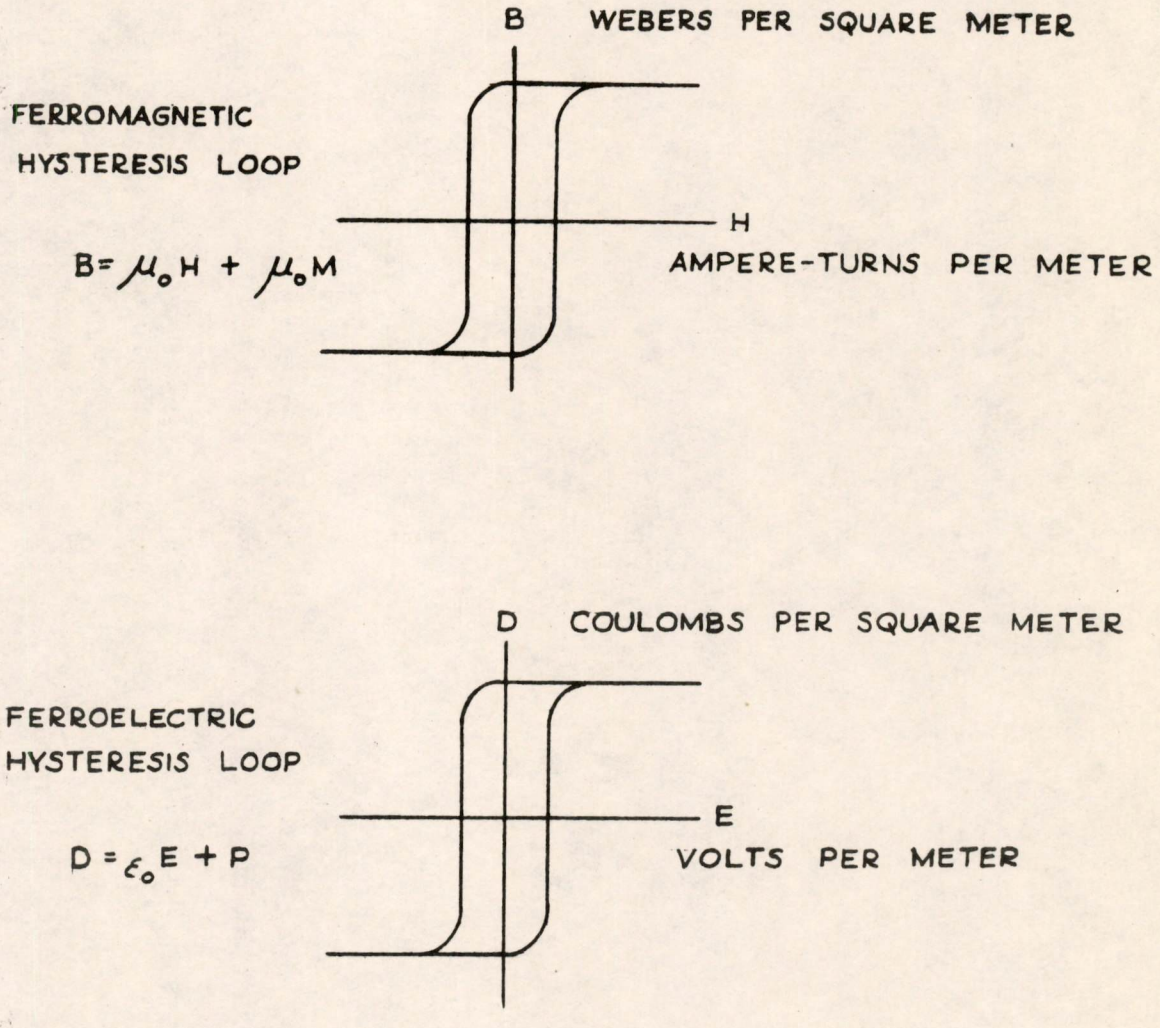


FIG. 11

COMPARISON OF FERROMAGNETIC AND FERROELECTRIC HYSTERESIS LOOPS

A-50545

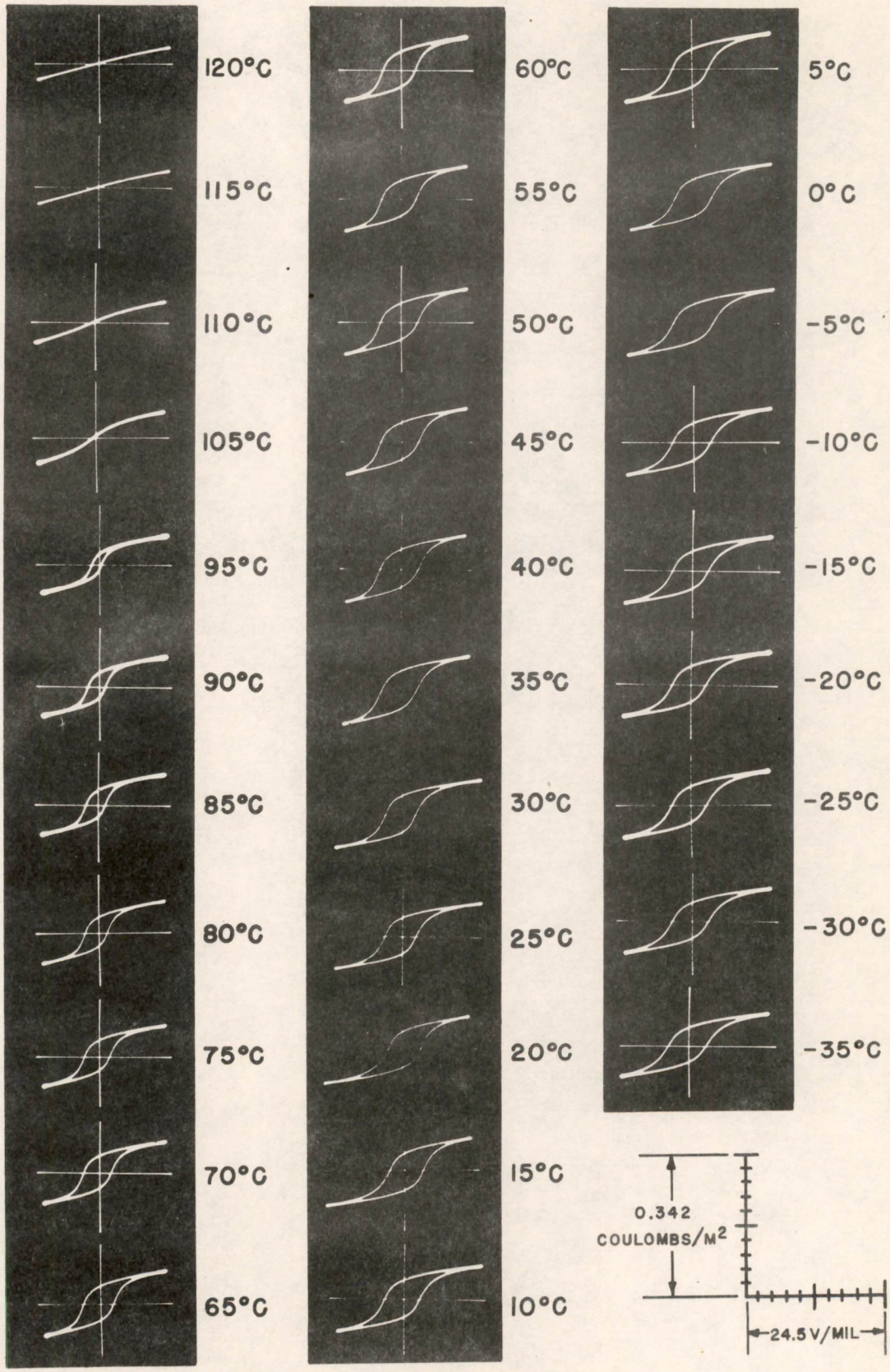
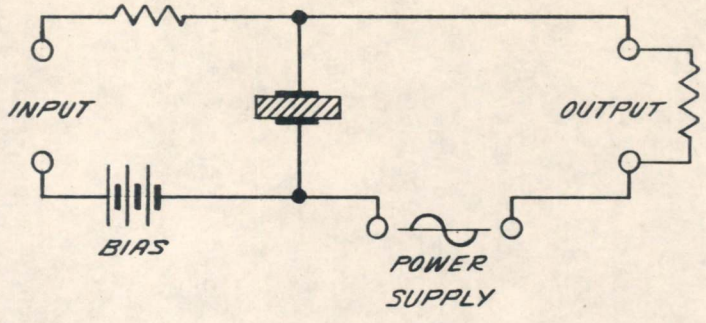


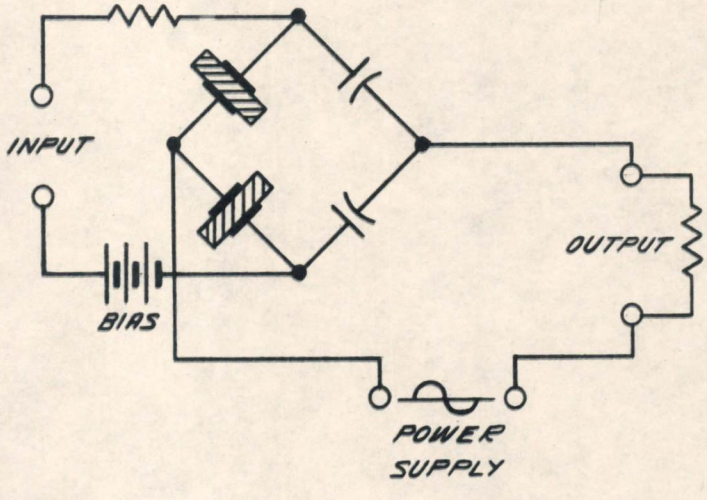
FIG. 12

HYSTERESIS LOOPS OF BARIUM TITANATE CERAMIC AS A FUNCTION OF TEMPERATURE

A-51351



A. SIMPLE ONE-ELEMENT SATURABLE CONDENSER



B. BRIDGE-CONNECTED SATURABLE CONDENSERS

FIG. 13

DIELECTRIC AMPLIFIER CIRCUITS

A-52312

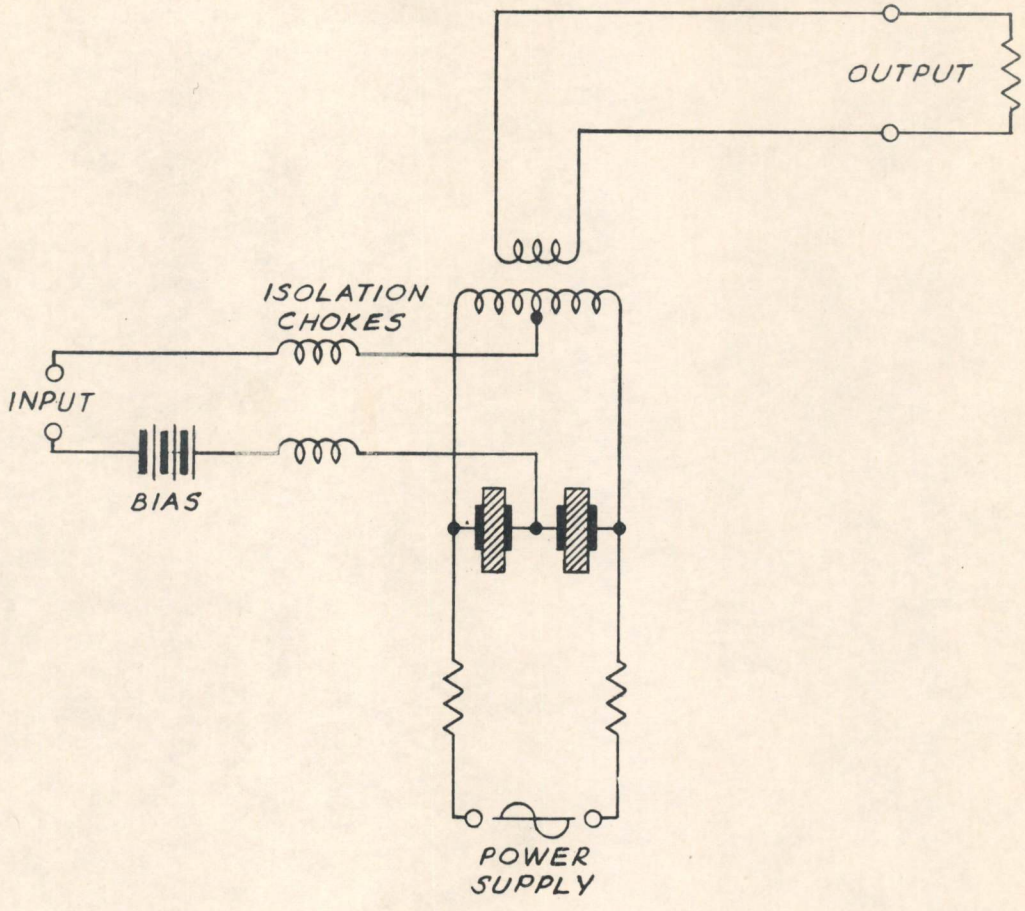
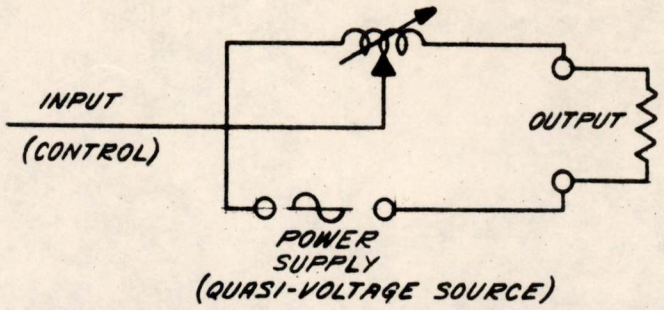


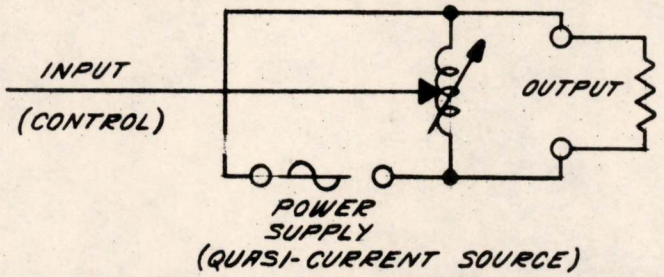
FIG. 14

DIELECTRIC AMPLIFIER
OPERATED NEAR RESONANCE

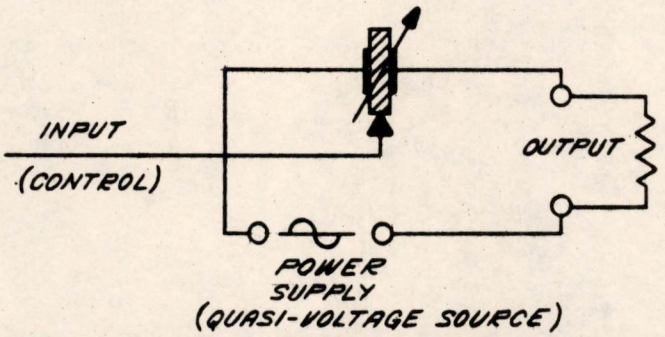
A-52393



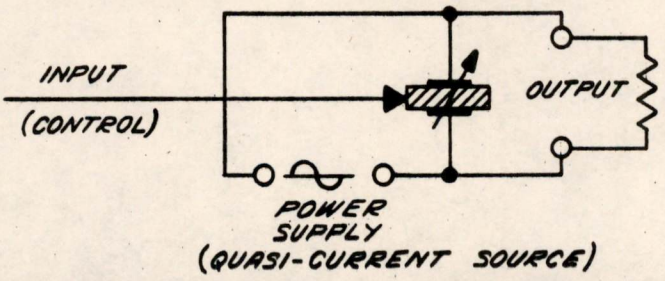
A. SATURABLE INDUCTOR EQUIVALENT CIRCUIT



B. SATURABLE TRANSFORMER EQUIVALENT CIRCUIT



C. SERIES SATURABLE CONDENSER EQUIVALENT CIRCUIT



D. SHUNT SATURABLE CONDENSER EQUIVALENT CIRCUIT

FIG. 15

MAGNETIC AND DIELECTRIC AMPLIFIER
EQUIVALENT CIRCUITS

A-52313

Digital Computer Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

SUBJECT: TOGGLE SWITCH INPUTS AND INDICATOR LIGHT OUTPUTS AS EXTERNAL UNITS

To: 6889 Engineers

From: Guy A. Young

Date: September 17, 1952

Abstract: Several methods of Toggle Switch Input to the computer and Indicator Light Output from the computer are discussed. The possibility of using magnetic cores in Toggle Switch Input is also considered.

1.0

Several possible modes of Toggle Switch Input (TSI) operation are listed below. With each mode is a list of equipment needed, and the sequence of operation. All of these modes operate on an "si (TSI)" and a "rd", as conventional input units. All operate through the IOE.

For matching purposes, some type of driving system will have to be added for each Toggle Switch Register. Three methods have been suggested. One method uses a Buffer Amplifier for each digit. One set of 16 amplifiers would serve all registers that are physically close together. Another method requires one gas tube for each digit of each register. These gas tubes have a duty cycle limitation of about one millisecond. A third method would use D.C. levels from the IOS. An amplifier would be required for each register. A gate for each digit would be placed physically close to IOR. (See Fig. 4.)

All of the modes use approximately the same amount of equipment. It is felt that either "mode 1" or "mode 4" is preferable. Both of these modes allow the computer to skip an order.

1.10

Mode 1 - TSI reads in information only if the word has been changed. If no information has been read in, the computer ships an order. (See Fig. 2.)

1.11 Operation

si - selects TSI. Reads word into IOR if the word has been changed.

rd - if word has been changed, reads contents of IOR to AC.

1.12 Equipment (for each TSI register)

1. One 16-digit toggle switch register

2. Three gates

3. One flip-flop

4. One push button synchronizing gate

It may be possible to use one synchronizing gate for several registers. The register could be selected with a toggle switch.

No changes would be required in the IOE except those contemplated for magnetic drums.

It may be desirable to place a toggle switch between the synchronizing gate and the flip-flop. (See Fig. 2.) This would allow the operator to eliminate a word if new information were available before the old information had been accepted by the computer.

1.20

Mode 2 - TSI reads in only if the word has been changed. If the word has not been changed, "+0" is placed in AC on a "rd" order.

1.21 Operation

si - selects TSI. Reads word into IOR if word has been changed.

rd - reads contents of IOR into AC. If word has not been changed, AC contains "+0".

1.22 Equipment - same as listed in 1.12.

1.23 No changes would be required in the IOE.

1.30

Mode 3 - TSI reads in only if word has been changed. Computer waits until word has been changed.

1.31 Operation

si - selects TSI. Reads word into IOR.

rd - reads word into AC. Computer must wait until si has been completed.

1.32 Equipment - same as listed in 1.12.

1.33 This mode is undesirable since the operator could stop the computer for an indefinite length of time.

1.40

Mode 4 - TSI reads in unless word is being changed. If word is being changed, the computer ships an order. (See Fig. 1.)

1.41 Operation - similar to mode 1.

1.42 Equipment (for each TSI register).

1. One 16-digit toggle switch register

2. Three gates

3. One flip-flop

4. One toggle switch

1.43 No changes would be required in the IOE except those contemplated for magnetic drums.

1.50

Mode 5 - TSI reads in unless word is being changed. If word is being changed, a "rd" will put "+0" in AC.

1.51 Operation - similar to mode 2.

1.52 Equipment - same as listed in 1.42.

1.53 No changes would be required in IOE.

1.60

Mode 6 - TSI reads in unless word is being changed. Computer waits while word is being changed.

1.61 Operation - similar to mode 3.

1.62 Equipment - same as listed in 1.42.

1.63 Undesirable, since computer could be made to wait an indefinite length of time.

1.70

Mode 7 - Computer always reads in. Two toggle switch registers could be used. The operator sets one while the computer reads in the other.

1.71 Operation

si - selects TSI. Reads word into IOR.

rd - reads word from IOR to AC.

1.72 Equipment (for each TSI register)

1. Two 16-digit toggle switch registers

2. Three gates

3. One flip-flop

4. One toggle switch

1.73 No changes would be required in IOE.

1.80

Mode 8 - Computer always reads in. The operator indicates, with a light gun for example, that TSI is ready to be read in. (See Fig. 3.)

1.81 Operation - same as mode 7.

1.82 Equipment (for each TSI register)

1. One 16-digit toggle switch register

2. One gate

1.83 No changes would be required in IOE.

1.90

Several of these modes could be combined using the equipment listed in 1.12 plus a few switches.

2.0

The possibility of using cores has also been considered. It is felt that a core could not be used as a synchronizer. It could be used as a memory, however. It may be possible to replace the flip-flop of "mode 1" by a magnetic core. The feasibility of doing this depends on the "jitter" of the synchronizing generator and the set-up time of the core. If two flip-flops were used to obtain a synchronous "set" pulse, the core would serve no purpose.

Assuming that the circuit would operate, the use of a flip-flop still has two advantages over the use of a core. An indication that a word has been read into the computer is much more easily done with a flip-flop than with a magnetic core. Also, once a process has been started, it cannot be stopped if a magnetic core is used. This is not true if a flip-flop is used. For these reasons, the use of a flip-flop seems to be the best method.

3.0

A method has been suggested for recording information from the recorder into a neon light indicator. A gas tube would be required for each digit of each register. Two possible modes of operation are:

mode 1 - The computer clears the indicator when it is ready to read in a new word. (See Fig. 5.)

mode 2 - The operator clears the indicator when he has read the word.

3.10 "Mode 1" Operation

si - the indicator is released, and a delay for the relay (about 12ms delay) started. Completion clears interlock and allows "rc" to proceed.

rc - place word in IOR. IOC reset pulse sets indicator to the contents of IOR.

3.20 "Mode 2" Operation

si - if the operator has read the word, he presses a push button. The next "si" releases the indicator light and starts a delay. If the operator has not read the word, IOC will be set up to skip an order on the next "rc".

rc - either sets indicator to new word or skips an order in the program.

3.21 "Mode 2" requires two flip-flops and several gates.

3.30

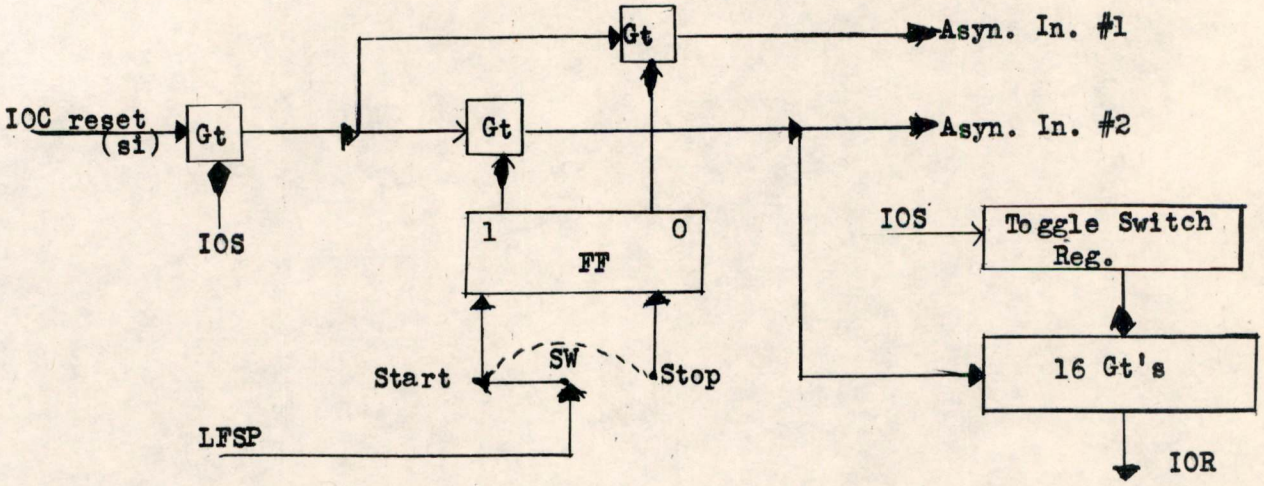
For either indicator system, a gate in IOC is required to set IODC. "Mode 1" requires no additional changes in IOC. "Mode 2" requires no additional changes in IOC other than those contemplated for the magnetic drum.

The above indicator systems display information in octal form. It may be desirable to display information in decimal form. Available methods seem to be either too slow or too complex.

Signed Guy A. Young (BEM)
Guy A. Young

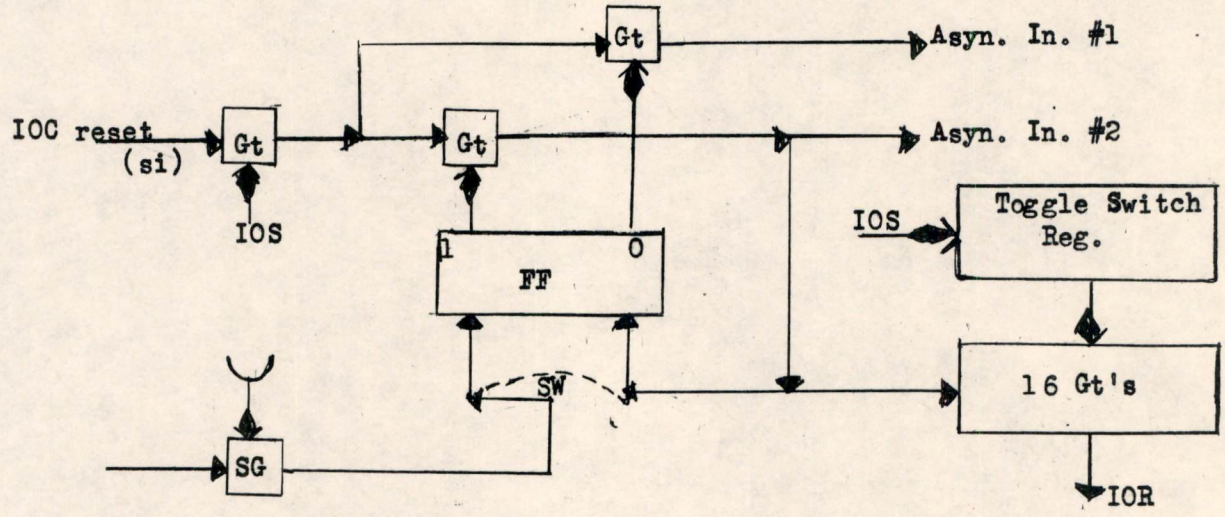
Approved E.S. Rich
E.S. Rich

GAY/mrs
cc: C.R. Wieser
C.A. Adams
J.A. Arnow
S. Dodd
N. Daggett
D. Israel
J.A. O'Brien
E.S. Rich
R.L. Walquist



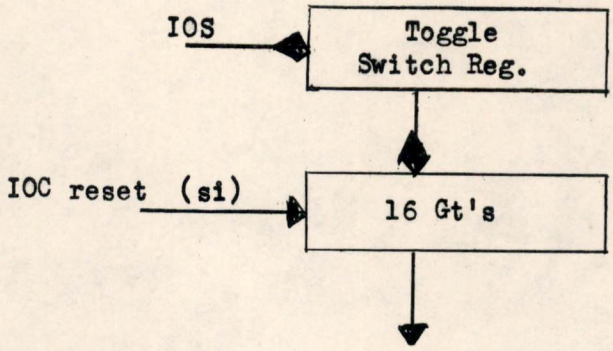
Toggle Switch Input - mode 4

Fig. 1



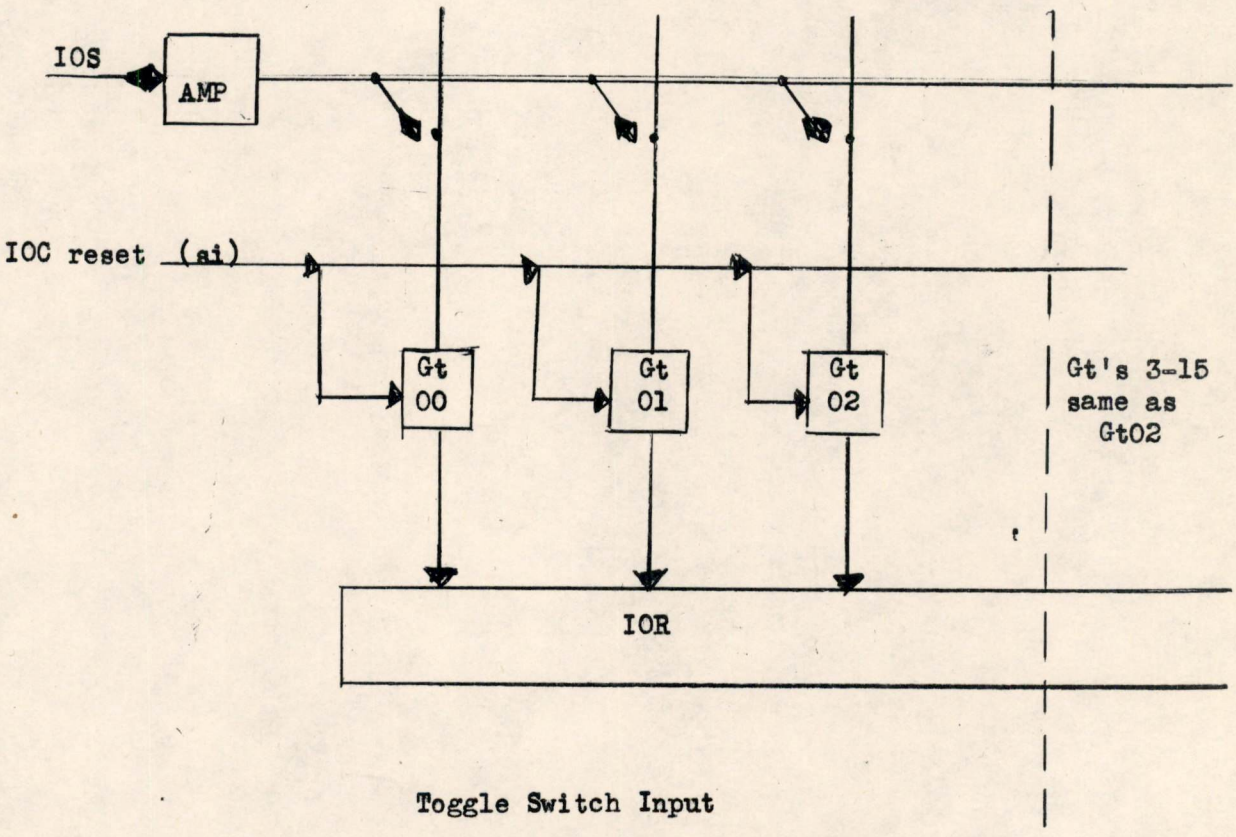
Toggle Switch Input - mode 1

Fig. 2



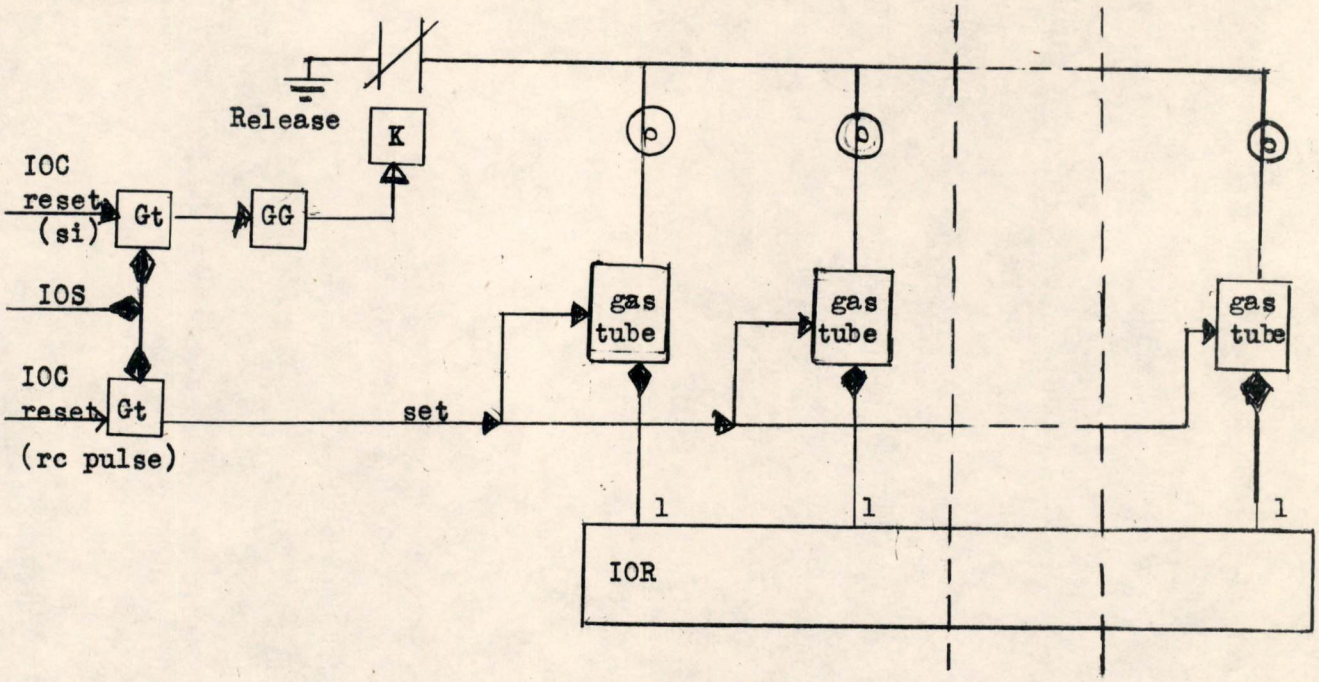
Toggle Switch Input - mode 8

Fig. 3



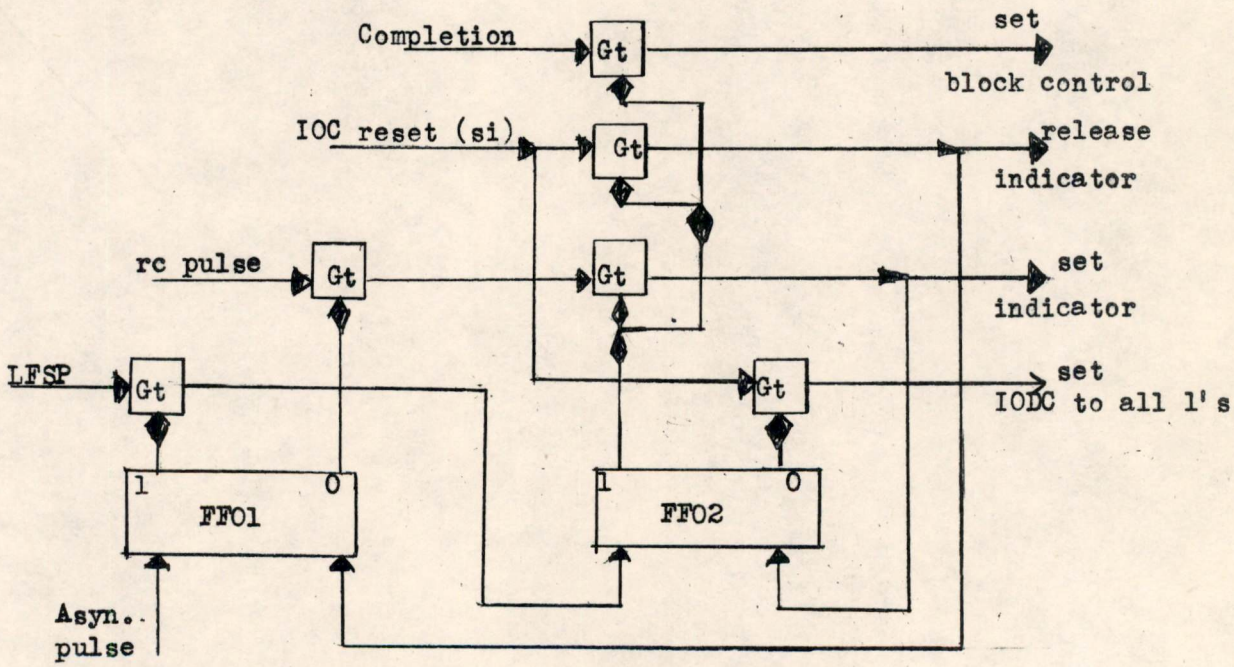
Toggle Switch Input

Fig. 4



Indicator Light Unit - modes 1 and 2

Fig. 5



Indicator Light Control - mode 2

Fig. 6