Semiconductor Diode Gates

By LUTHER W. HUSSEY

(Manuscript received May 22, 1953)

Diode gates, which are the body of modern pulse communication and computing devices, are discussed. Methods of analysis, by which practical first order design is possible, are given with experimental verification. The general properties of the gates, both virtues and defects, are noted and methods shown for minimizing the defects.

INTRODUCTION

Semiconductor diodes are old from the viewpoint of communication engineering. The crystal detector was a fundamental part of the early radio receiver. It was also a troublesome part. The crystals were highly variable and unreliable and there was little theoretical understanding of their physics which could be used as a basis for successfully exploring and controlling their characteristics. While the semiconductor continued to be useful in the field of power rectifiers and the copper oxide modulator later found extensive use in carrier telephone systems, the development of the vacuum tube into a relatively reliable device, with controlled characteristics, tended to eliminate the semiconductor from most of the communication field.

The copper oxide modulator, in a carrier telephone system, is an early example of the type of application which could bring the semiconductor back into competition with the tube rectifier. In a radio receiver, the difference in unit cost, power consumption, space requirements, and maintenance expense, between a tube detector and a crystal, may be small compared with the engineering advantages; but in a telephone plant, with its multiplicity of units, each of these small increments of cost may be integrated up to a major item, and may determine the economic feasibility of the whole system.

The need for better semiconductors than copper-oxide, when carrier frequencies moved up into the megacycle range, was a major stimulus to continue research in the semiconductor field. Within the last few years, physicists, using their recently acquired understanding of the structure

and mechanisms involved, have returned to other semiconductors — particularly to silicon and germanium to obtain great improvements.

The interest in diodes was undoubtedly a factor in the chain of investigations which culminated in the invention of the transistor and the development of the transistor has in turn accelerated the improvement of the diode. The development of the transistor also has greatly increased the demand for diodes. Replacing vacuum tubes by transistors, in modern pulse communication and computer system design can expand the potentialities of such systems since such systems, even more than the telephone carrier, use large numbers of elements and the increments of cost (initial unit cost, power consumption, space requirements, maintenance) all become vital. In a typical system there may be a dozen diodes to every transistor or vacuum tube. In a vacuum tube system the use of tube diodes is generally undesirable. In a transistor system is is absurd.

FIELD OF APPLICATION

These two new and rapidly expanding applications of diodes (pulse communication and computing) are amazingly simple in their basic ideas and circuit building blocks. Practically all they do is:

1. Generate pulses or accept them from another source and regenerate them.
2. Store pulses.
3. Route pulses to a desired output.

The complexity of the modern computer with its thousands of tubes or transistors and diodes lies in the number of operations, not in the individual operation itself. The modern computer is capable of solving compex mathematical problems involving any of the normal mathematical operations and may do it primarily with diode networks — aided by amplifiers to compensate for losses, and delay networks, or timing devices, to insure that processes take place in the proper sequence[1]. The diode networks, which are the body of a modern computer or pulse communication system, are "gates" and routing circuits which, by controlling and guiding the passage of pulses, are capable of performing all the logical processes required.

TYPES OF DIODE GATES

A "gate", for our purposes, is an electrical device with an input, an output, and one or more control inputs. The control inputs and the other

[1] Felker, Regenerative Amplifier for Digital Computer Applications, Proc. I.R.E., **40**, Nov., 1952. Chen, Diode Coincidence and Mixing circuits, Proc., **38**, May, 1950.

input may be indistinguishable. When a certain combination of potentials is impressed on the controls an output appears at the output terminals. In the case of a "linear" gate the input signal has simply passed from input to output, so the output signal is approximately a replica of the input signal. In the case of a "switching" gate the output is a pulse which may have no resemblance, except in duration, to the pulses and potentials which are impressed on the controls.

There is essentially only one kind of linear gate. When a signal appears at its input, the potentials then present on the controls determine whether it shall pass to the output or be blocked. Switching gates, in spite of their apparent complexity in some particular applications, are constructed of two basic types — OR circuits and AND gates. A typical OR circuit is shown in Fig. 1. When a positive pulse is impressed on either of the input terminals (terminal 1 in the illustration) it drives the corresponding diode conducting and the pulse appears at the output (3).

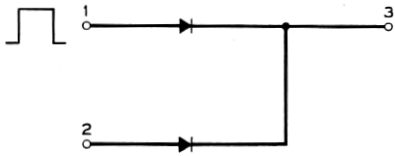

Fig. 1 — Diode OR circuit.

The signal pulse drives the other diode non-conducting, blocking that path.

An AND gate is shown in Fig. 2. The biases are so adjusted that, in the rest condition, the output voltage is zero, the diodes are conducting and the bias current from the bias current source (B) flows through the diodes and the control terminal resistors. In this case, when a positive pulse is impressed at one of the control inputs the corresponding diode is cut off and bias current ceases to flow in it. However, there is very little resulting change at the output because the bias can still flow in the second, low impedance control diode and resistance. If a second pulse is simultaneously impressed on the second control input, the second diode is also cut off and the bias current is forced to flow in the load, producing an output pulse of magnitude determined by the magnitude of the bias current and the magnitude of the load resistance.

There is a third fundamental gating concept which must also be introduced. That is the INHIBITING control. Its purpose is to prohibit an output, whatever may be done to the other controls. It is illustrated in Fig. 3. As far as control 1 is concerned, this is like the AND gate. The output, in the rest condition, is zero and the diode is conducting. There

might, in fact be two or more such controls, forming a standard AND gate. The second control shown is the INHIBIT control. It is so biased that the diode is in the high impedance, cut off condition. If let alone, it has negligible influence on the behavior of the remainder of the system. If a large negative pulse is impressed on it, overcoming the bias (C_2) the diode becomes conducting and the load remains shunted by the low control circuit impedance. Whatever the condition of the other control (or controls) very little current can get to the load and an output is inhibited.

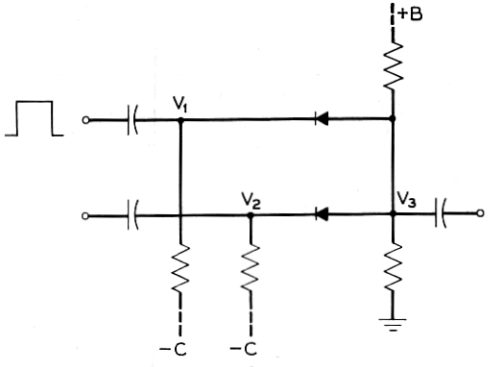

Fig. 2 — Diode AND gate.

There are several possible demands that might be put on a two control gate which can be shown in the following tabulation. In this tabulation a (1) means the presence of a pulse; a (0) means no pulse.

Control 1	*Control 2*	*Output*
0	0	0
0	1	0
0	1	1
1	1	1
1	1	0

These, and the corresponding relations when the two controls are interchanged, are all the available combinations of input and output. The first two are trivial — satisfied by a lack of any connection between input and output. The third demands an OR circuit. The fourth demands an AND gate. The final case requires a new configuration but can be satisfied by a combination of gates by adding the inhibiting controls. Fig. 4 shows such a gate. The boxes show gates of the type of Fig. 3, with the functional symbolism which is commonly used. A line with an

arrow aimed at the box represents a control; an arrow on a line, aimed away, represents an output connection; a control input with a small semicircle represents an inhibiting control. This combination satisfies what is called the AND NOT requirements. A pulse appearing at input (1) will pass gate (1) if there is no pulse simultaneously on gate (2). Similarly, a pulse on input (2) will pass gate (2). If pulses appear simultaneously on both inputs, each will inhibit the passage of the other and there will be no output.

In practice there are many complications encountered in using these

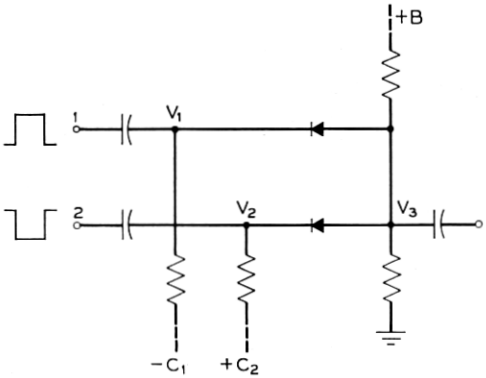

Fig. 3 — Gate with inhibiting control.

gates. In particular, networks using the INHIBITING control can, in the simple form show, get into difficulty due to the fact that pulses may not be exactly simultaneous. These difficulties are overcome in practical circuits by making part of the control voltages essentially dc potentials which set up the control conditions so that a final control pulse may produce an output, or be blocked, as the case requires. In a great many systems, this final pulse is an additional "clock pulse" from a master control or clock which synchronizes the sequence of operations of the entire computer. The details of how this is done and how the three types of gate may be combined is outside the scope of this paper. During the last few years there has been an extensive literature build up in the technical journals and a few books[2] published in this field.

TRANSMISSION GATE

Rather than analyze in detail the many forms a gate may take, only two forms will be discussed here. The methods (and in many cases the

[2] Keister, Ritchie and Washburn, *The Design of Switching Circuits*, Van Nostrand. Hartrie, *Calculating Instruments and Machines*, Univ. of Illinois Press.

results) can readily be applied to the numerous variants which may be encountered. The first is a form of "linear" or "transmission" gate. As previously stated, this gate has one or more control inputs, a signal input, and an output. When the control input enables the gate, a reasonably accurate replica of input signal should appear at the output. When the control input disables the gate, transmission of a signal should be effectively suppressed. Fig. 5 shows a form of this gate with a single control. This differs from the gates previously shown in that there is a diode in series with the output. Since this gate has superior discrimination and impedance characteristics, compared with the simpler forms it seemed desirable to analyze it as the typical transmission gate.

In this circuit, I_o, G_o represent the signal generator as a current generator with internal conductance G_o. Since the transmission proper-

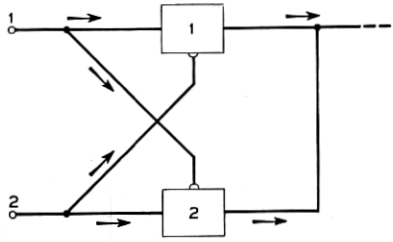

Fig. 4 — Combination of inhibited gates into AND NOT gate.

ties are of major interest, the load has been assumed to have the same conductance as the signal generator. A constant current bias I_b is impressed[3] at the midpoint of the gate. The control input voltage is represented by E_b. The internal control generator conductance, which should be large, is assumed to be included in the corresponding diode conductance, for computing purposes. The diodes are assumed to have a large conductance G, or a small conductance g depending on whether they are forward biased or reverse biased. Fig. 5 shows the gate enabled, with the series diodes in the conducting state and the control diode in the reverse bias, or non-conducting condition. If the G and g are interchanged the gate is in the disabled condition.

The relative magnitudes of V_o, V_1, V_2 evidently determine whether the diodes are biased forward or backward. Their magnitudes are immediately obtainable. The equations for the case shown in Fig. 5 are:

[3] In case the internal conductance of an actual bias source is not sufficiently small to be neglected, its main effect will be on transmission loss. In computing the loss, the bias conductance may be added to the conductance of the control connection.

$$(G_o + G)V - GV_1 = I_o \tag{1}$$

$$-GV_o + (g + 2G)V_1 - GV_2 = I_b + gE_b \tag{2}$$

$$-GV_1 + (G_o + G)V_2 = 0 \tag{3}$$

Their solutions are:

$$V_0 = \frac{I_0\left(g + G + \frac{G_0G}{G_0 + G}\right) + (I_b + gE_b)G}{gG_0 + gG + 2G_0G} \tag{4}$$

$$V_1 = \frac{I_0G + (I_b + gE_b)(G_0 + G)}{gG_0 + gG + 2G_0G} \tag{5}$$

$$V_2 = \frac{G}{G_0 + G}V_1 \tag{6}$$

The corresponding equations for the disabled gate are obtained by interchanging g and G in the above.

ENABLED GATE

Considering the enabled gate first, it should be evident from the figure that there are four requirements if the gate is to be properly enabled:

$$V_1 > V_o \tag{7}$$

$$V_1 > V_2 \tag{8}$$

$$V_1 < E_b \tag{9}$$

$$V_1 > 0 \tag{10}$$

V_1 is a positive voltage, so equation (6) insures that (8) will always be satisfied.

Putting the values of V_1 and V_o in (7) gives

$$(I_b + gE_b) > \left(\frac{g}{G_0} + \frac{G}{G + G_0}\right)I_0 \tag{11}$$

In the case that g is much smaller than G_o (which is normally true) gE_b is effectively the current from a constant current control generator and the above inequality has a simple interpretation:

In determining whether (7) is satisfied, and the input diode held conducting, the above inequality compares the total current which the bias and control generators put into midpoint (V_1) of the gate with the

maximum current which the signal generator could put into the same point when that point is grounded. If the control and bias current sum is larger, the current in the input diode cannot reverse.

The inequality (10) is to insure that the output diode remains conducting. Substituting the value of V_1 in it gives:

$$I_b + GE_b > \frac{-G}{G_0 + G} I_0 \tag{12}$$

The only way that the output diode could be cut off (with positive bias and control) is by a large negative signal current. The above inequality requires:

To hold the output diode conducting, the sum of bias and control generator currents must be greater in magnitude than the maximum negative signal current that the signal generator could put into the midpoint when the midpoint is grounded.

A zero potential on the midpoint is the boundary condition between the diodes being conducting or non-conducting. The two inequalities together compare the currents that the generators can put into the grounded midpoint. They require: The sum of bias and control generator currents should exceed in magnitude the maximum current of either polarity, that the signal generator can put into the grounded midpoint.

There remains the inequality (9) which is necessary if the control diode is to remain non-conducting. This gives:

$$\frac{GI_0}{G_0 + G} + I_b < 2\frac{G_0 G}{G_0 + G} E_b \tag{13}$$

This compares the same bias and signal generator currents with the current which would flow in the input and the output circuit if V_1 were replaced by E_b. If the inequality is satisfied V_1 can never get as large as E_b and the control input diode remains cut off.

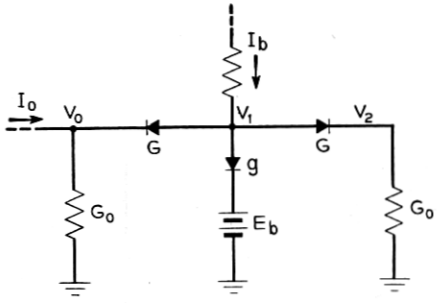

Fig. 5 — Transmission type diode gate.

DISABLED GATE

The condition under which the gate is held disabled is much simpler than the enabling conditions. All that is necessary is that V_1 be held more negative than the most negative signal generator voltage:

$$V_1 < V_0 < 0 \tag{14}$$

If equations (4), (5), and (6) are used in the disabled case (i.e., with the g's and G's interchanged), the condition can be obtained:

$$\frac{1}{G} I_b + E_b < \frac{1}{G_0} I_0 \left(1 + \frac{gG_0}{(G_0 + g)G}\right) \tag{15}$$

When g is much smaller than G_0, so that the signal generator is effectively open circuited, this is a direct comparison of voltage components on the two sides of the input diode. The first term on the left, is the positive voltage at the midpoint (V_1) due to the bias current I_b. The term on the right of the inequality is approximately the open circuit signal generator voltage. The inequality says that E_b must be sufficiently negative to overcome the positive bias and hold the midpoint more negative than the most negative signal voltage.

OPTIMUM CONDITIONS

What constitutes optimum conditions depends on the particular requirements of the associated circuit. It is, however, always true that the diode conductance ratio should be as large as possible:

$$G/g \text{ as large as possible.}$$

For *minimum power loss* the gate, which is a resistive T network, to the approximation considered here, should be terminated in its characteristic conductance:

$$G_0 = \sqrt{\frac{gG}{2 + g/G}} \tag{16}$$

For *small volltage loss* the terminating conductance should be small:

$$G_0 \ll G$$

For good discrimination (large loss in the disabled state) the terminating conductance should be large:

$$G_0 \gg g$$

These all add up to the general requirement:

$$g \ll G_0 \ll G \tag{17}$$

and the particular choice of G_0 is generally a compromise between good discrimination and low transmission loss.

PEDESTAL

A major defect of diode gates is an output voltage variation produced by control voltage changes. The output voltage is very small and negative when the gate is disabled, but when the control voltage switches the gate to the enabled condition a positive voltage, called the pedestal, appears at the output. From (5) and (6), with $I_0 = 0$, it is

$$V_2 = \frac{1}{G_0} \frac{I_b + gE_b}{2 + \frac{g}{G_0} + \frac{g}{G}} \tag{18}$$

Since g is small, this approximates the voltage due to half the bias and control currents flowing in the output. The signal output is superposed on this pedestal and may swing from zero to twice V_2. If the frequencies involved in the signal and the control voltages are widely different, this pedestal is not important since it can be filtered out from the transmitted signal but it is a serious output distortion in other cases where signal frequency components from the control pulse may be transmitted as a spurious signal.

EXPERIMENTAL CHECK

In the preceding discussion it was tacitly assumed that a diode switches between two constant conductance values. On a dc basis, the equations are still valid with a variable conductance, except for the signal loss relations. In computing the small signal loss, the conductances chosen should be the dynamic or differential conductances at the particular bias currents chosen.

The following is an example of the kind of experimental checks obtained in which, using 400B diodes, the actual behavior of a gate was tested on a dc basis. The particular units had approximately an impedance ratio (at about 5 volts) of 20,000/200. The load conductance was chosen (somewhat arbitrarily) as the geometric mean of the diode conductances, giving:

$$G = 5\ 10^{-3}$$

$$g = 5\ 10^{-5}$$

$$G_0 = 5\ 10^{-4}$$

If the maximum signal generator current is chosen as

$$I_0 = 5\ 10^{-3}$$

inequalities (11) and (13) take the form:

$$I_b + 5\ 10^{-5}\ E_b > 5.05\ 10^{-3}$$

$$-1.1\ I_b + 10^{-3}\ E_b > 5.0\ 10^{-3}$$

No negative signal voltages were used, so inequality (12) is not involved. The above inequalities limit I_b and E_b as shown in Fig. 6. I_b and E_b must be chosen from the shaded area. The values chosen were $I_b = 5ma$ and $E_b = 15$ volts.

Comparing experimental results with analytic, gives:

Voltage loss 1.0 db (computed 1.6 db)

Pedestal 5.05 volts (computed 5.45 volts)

Further experimental results, which are all in reasonable agreement with expectations are given on Figs. 7, 8 and 9. Fig. 7 shows how the pedestal voltage varies with gate voltage (E_b). This is also a measure of the load capacity, since the signal output can swing from zero to twice the pedestal. The useful range is above the point where the pedestal ceases to increase with gate voltage. In this range the control diode is cut off. Figure 8 shows the pedestal against bias current I_b. The limiting is not as sharp here because the forward conductance of the input and output

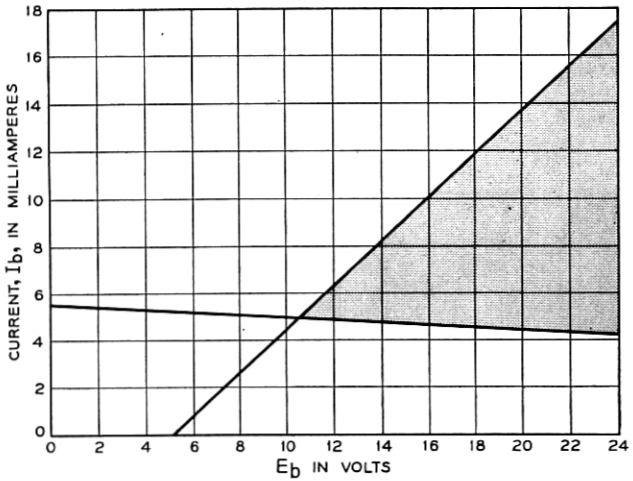

Fig. 6 — Bias restrictions on transmission gate.

diodes are involved and they do not change as rapidly with current in milliamperes as the reverse conductance does with volts. Fig. 9 shows the actual input output relation for the signal. It is linear over 80 to 90 per cent of the 5-volt range and then limits as expected.

The discrimination was not measured. It computes to better than 60-db voltage loss and discrimination of that order of magnitude has been measured in gates of this type.

SWITCHING GATE

A form of gate which is useful for pulse systems, since it lacks the pedestal, is shown on Fig. 10. This is, of course, basically the same configuration as that shown on Fig. 5, but it is operated quite differently, with pulses or dc potentials applied to the two control inputs and an output obtained by switching the bias current from flowing in a control path to flowing in the load. More specifically, if both E_1 and E_2 are sufficiently negative, diodes D_1 and D_2 are both conducting, V_b is negative, and D_3 is non-conducting. Thus practically all the bias current I_b flows in D_1 and D_2, and V_2 is zero or slightly negative. If one of the control voltages is increased until its diode cuts off, the bias current can still flow in the other control path and the change in the output voltage is extremely small. If both the control voltages are increased until the two control diodes are cut off, then V_b becomes positive, D_3 conducts and practically the entire bias current flows in the load, producing an output voltage

$$V_L = I_b R_L.$$

The above operation gives a two control AND gate with no pedestal

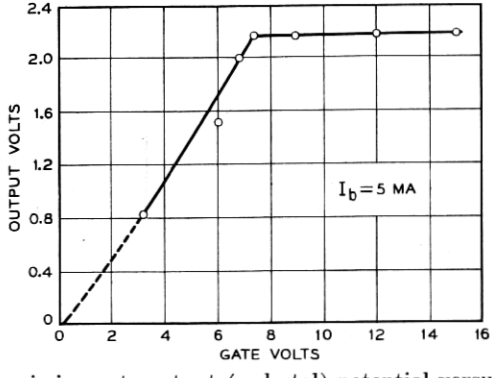

Fig. 7 — Transmission gate output (pedestal) potential versus gating control potential.

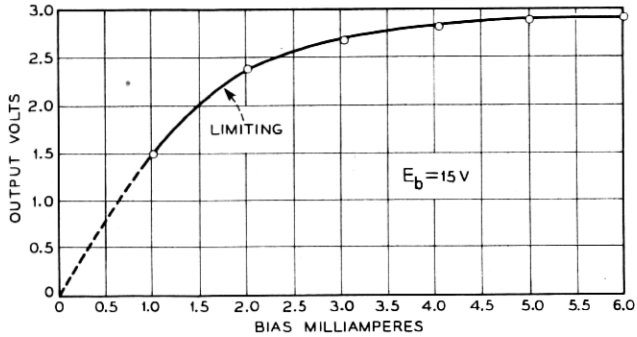

Fig. 8 — Transmission gate output (pedestal) potential versus bias current magnitude.

difficulty. One of the controls could instead be used as an inhibitor. For example, if E_2 were normally biased sufficiently positive it would have no effect on the output, which would be controlled by E_1 alone. However, a negative, or INHIBITING pulse, sufficient to make D_2 conducting, would permit the bias current to flow in that path and prevent an output, whatever the state of D_1.

This circuit could be analyzed in exactly the same manner as was done with the transmission gate. However, after a value of R_L has been chosen, a simple first approximation to a design may be carried out by assuming that the diodes are ideal, switching between zero and infinite resistance. In choosing R_L there are three major considerations:

1. R_L must be small compared with the reverse resistance of the diode, D_3, or there may be an appreciable negative output when the gate is disabled.

2. R_L must be large compared with the forward resistance of D_3 for efficient operation of the enabled gate,— preventing an appreciable voltage loss due to the voltage drop in D_3.

3. The peak amplitude of the output pulse is I_bR_L.

The value of R_L, which is chosen from the wide range of possibilities, is a matter of practical compromise, depending on the impedance levels in the system and the constant current generators which are available.

Having chosen R_L and I_b there remain only the control voltages, E_1 and E_2. The voltages which are necessary to hold the gate enabled can be obtained by noting that (in the ideal diode case)

$$V_b = V_3 = R_L I_b \tag{19}$$

To hold the control diodes non-conducting the voltages E_1 and E_2 must be greater than V_b. This gives:

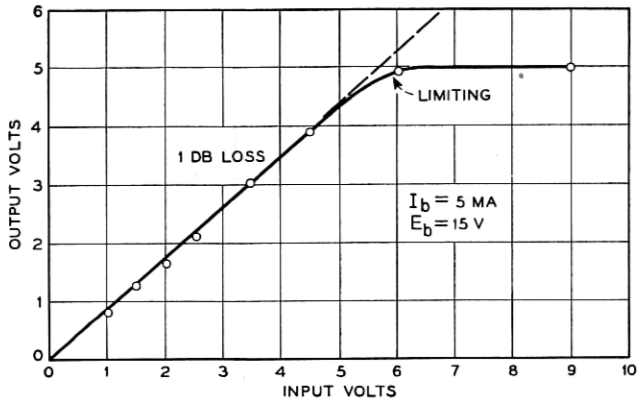

Fig. 9 — Transmission gate signal output potential versus input signal potential.

To *enable* the gate, voltages must be impressed such that

$$\begin{aligned} E_1 &> R_L I_b \\ E_2 &> R_L I_b \end{aligned} \tag{20}$$

In disabling the gate, either one of the control paths may have to carry the full bias current. If, for example, the bias current were flowing in D_2 it would bias that control path positive by an amount $R_2 I_b$. To overcome this, and keep V_b negative, the requirement is:

To *disable* the gate, voltages must be impressed such that

$$\begin{aligned} E_1 &< -R_1 I_b \\ E_2 &< -R_2 I_b \end{aligned} \tag{21}$$

These sets of conditions give the magnitudes of the biases which are necesssary to hold the gate either enabled or disabled, and the difference between them is the minimum magnitude of the necessary switching pulse.

EXPERIMENTAL GATE

An example may be given, using 400B diodes. The bias was chosen as 5 *ma* and the load resistance, 2,000 ohms. The control resistances were made small, as is desirable for reasons which will be discussed later. For this gate the output voltage is 10 volts. From (20) and (21) the gate could be enabled by 10 volt positive control pulses and disabled by very small negative control values. A larger than necessary control voltage was put on control 2,

$$E_2 = 15 \text{ volts}$$

and the output measured as a function of E_1. The results are shown on Fig. 11. Since the diodes are not ideal, there is a transition region, but, as predicted the output is very small at small negative control voltage and the output is the full 10 volts when E_1 is 10 volts.

The curve also shows what happens if one of the enabling biases are too small. A case is shown in which E_2 was only 5 volts. There is no significant difference until the output gets up to 5 volts. Above that voltage the diode, D_2, becomes conducting and the output is clamped at that voltage.

GATE CHARACTERISTICS

The main virtues of this type of gate is that there is no pedestal and a constant amplitude pulse is produced. It is also simple and has good discrimination. There are limitations:

1. Unless a very low control path resistance is used, there is a large loss — that is the output pulse is much smaller than the control pulse. For example, if the control resistance is equal to the output resistance there is a two to one loss.

2. A rather large load is put on the control generator, partly because it must produce the enabling voltage across a small resistance and also because, in some cases the total bias current flows in the control generator output.

3. A phenomenon called "hole storage", which is present to some extent in all semiconductor diodes can make trouble. When a diode has been resting in the conducting state with a current flowing in it and the voltage is reversed, the diode does not immediately change to high impedance. A reverse current flow for a short time — up to a few microseconds. This can result in very inconvenient, spurious, output pulses being produced by a gate which is supposed to be disabled.

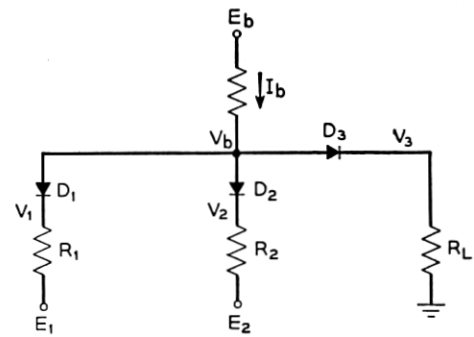

Fig. 10 — Switching type diode gate with two controls.

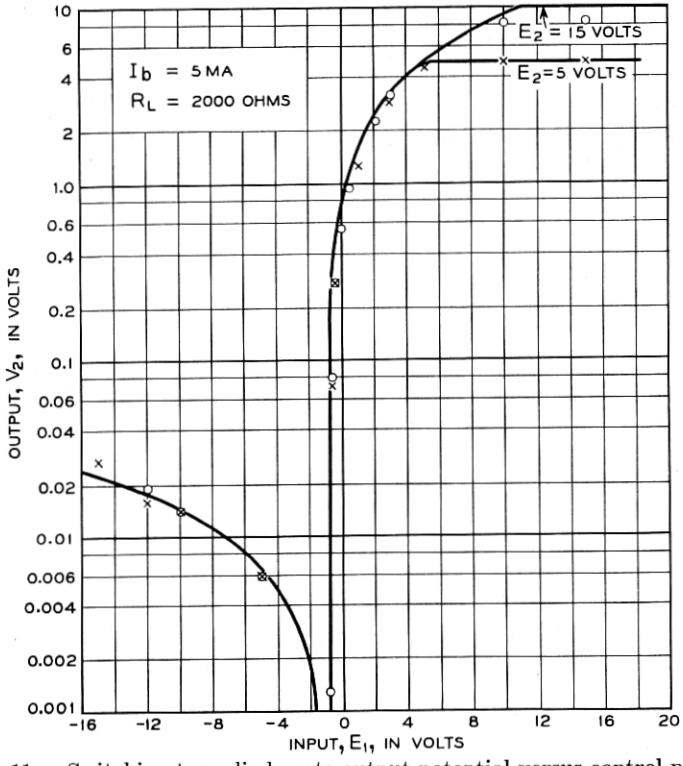

Fig. 11 — Switching type diode gate output potential versus control potential with a second control in the "enable" state.

SPECIAL CIRCUIT

It is not intended, in this paper, to attempt to list the numerous modifications which have been made of diode gates to overcome limitations and satisfy particular requirements, but it seems desirable to note an example of means to minimize the limitations.

One simple way of minimizing the difficulty, is to use point contact diodes in places where a spurious pulse could make trouble and use junction diodes elsewhere, since junction diodes have better impedance ratios but worse hole storage. For example, the output diodes in the switching gate could be a point contact unit, while the better discrimination of the junction unit made use of for control diodes.

A second means of avoiding hole storage effects is to avoid leaving diodes with large currents flowing in them, when they must be switched rapidly to the non-conducting state.

Fig. 12 illustrates both these design ideas. One of the control inputs has the control pulse impressed by means of a transformer and so has a very low dc impedance without making excessive demands on the control pulse generator. The biases are so adjusted that, in the disabled state, all the control diodes are just on the edge of conducting except the one in the transformer path. Because of the low DC impedance, practically all the bias current flows in this path. Thus there is no possibility of hole storage except in this one diode. If a positive control pulse is impressed while the potentials on the other controls are at their more negative value, the bias current just switches into those control paths; the output diode remains non-conducting and any spurious hole storage

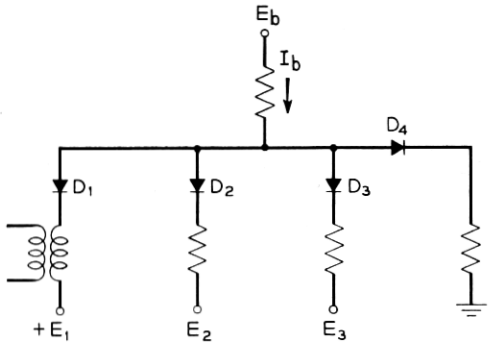

Fig. 12 — Switching type gate which minimizes "hole storage" effects.

current from the diode in the transformer control also goes into the other control paths. Since there is no storage in the other control paths, positive pulses may be simultaneously impressed on all the controls, the diodes in all but the transformer control path will immediately become high impedance and any hole storage current from the one diode will harmlessly add to the bias current flowing into the output.

Junction diodes are used in all the control inputs except the one with a transformer. A point contact unit is used here to minimize the hole storage. A point contact unit is also used in the output position. This is a critical location where good diode action is more important than a very high discrimination.

There is an additional advantage, in this configuration, that the relatively large bias current flows in the control generators only very briefly — while the disabled gate is being pulsed by the transformer control.

ACKNOWLEDGMENTS

It is impossible to give due credit to all the people who contributed to the development of the gates discussed here. Credit, however, should be given to W. D. Lewis, L. A. Meacham, and A. J. Rack who developed and explored the basic transmission gate analyzed here, and to J. R. Harris who made the modifications in the gate to avoid hole storage difficulties.

Acknowledgment is also due that part of the work reported on here was done in connection with a Joint Services contract.