L. Esaki

Characterization of Tunnel Diode Performance in Terms of Device Figure of Merit and Circuit Time Constant

Abstract: Tunnel diode oscillation, flip-flop switching and Goto twin operation have been characterized on the basis of numerical integration of a nonlinear differential equation representing transient behavior in the simplified, lumped equivalent circuit. The results have provided some information for the maximum utility of a tunnel diode having a given figure of merit. A useful relationship is given for single diodes, in which switching time vs external-circuit time constant L/R is shown for various device figures of merit. Another relationship is given for twin diodes in plots of cut-off frequency vs L/R. The study is then extended to the case of oscillator diodes.

Introduction

Since negative resistance was originally observed in a tunnel junction,[1] considerable progress has been made toward the development of a high-figure-of-merit tunnel diode. From these attempts has arisen a considerable body of literature showing various fabrication methods and giving diode performance data. At this time, it might be worth while to analyze the capabilities of this simple two-terminal device.

The systems development engineer might be interested in the limitations of power and of switching time. Suppose a circuit engineer plans to use a given type of tunnel diode. He certainly has to know the maximum usage of a tunnel diode having a certain figure of merit. Would it be reasonable, for example, to expect a tunnel diode oscillator to deliver a considerable amount of power at 10 gigacycles? Or would it be physically possible to obtain a 0.1 nanosecond switching circuit with a diode of advanced design? The purpose of the present paper is to consider these kinds of problems on the basis of numerical calculation of the simplified, lumped equivalent circuit, in a way somewhat similar to Schuller-Gärtner's method.[2] The results of this analysis will clarify diode operation and show the necessary conditions, viz., requirements for both the diode and the circuit, in order to attain a certain goal of system capability. The analysis will also provide us with a useful rule of thumb for the relation between the device's figure of merit and the external circuit time constant. This relation is given in terms of the switching time, for the single diode, and in terms of the cut-off frequency, for the twin diode. The oscillation behavior, including output power of the fundamental, the second and third harmonics, is also investigated at the various bias-point load conditions for a diode having a given figure of merit. The result can easily be extended to the general case.

The figure of merit of the diode will be mostly expressed by the $c\langle r\rangle$ time constant, where c and $\langle r\rangle$ are the capacitance and the average negative resistance, respectively. This purely mathematical approach, without any physical consideration of the detailed mechanism of tunneling, may be entirely justified because of the far smaller time constant of the tunneling (probably much less than 10^{-13} sec) than that of the $c\langle r\rangle$ time constant of the diode (larger than 10^{-12} sec). In other words, the capacitance-resistance parallel equivalent circuit of the diode is valid over the wide frequency range of interest.

Formulation

Now let us consider such a greatly simplified circuit common to various applications, as shown in Fig. 1(a). The diode which is shown with a parallel circuit of nonlinear conductance g and capacitance c, could be an amplifier, oscillator or switching element, according to the given values of the dc source voltage V_s, the external resistance R, and the inductance L. When we define the currents i, i_0 and j as indicated in Fig. 1(a),

the behavior can be written by the following two fundamental loop equations,

$$\int \frac{i_0}{c}\,dt + v = 0$$

$$v + Ri + L\frac{di}{dt} = V_s\,,$$

where v is a terminal voltage of an ideal diode, excluding the voltage drop due to series resistance. Also we can write the relations

$$i_0 = j - i$$

$$j = j(v)$$

and

$$c = c_0\left(1 - \frac{v}{E_g}\right)^{-1/2},$$

where the energy gap E_g is taken as 0.6 v for germanium, and a step junction is assumed. Eliminating i from the above equations, we obtain a nonlinear second-order differential equation,[2]

$$Lc\frac{d^2v}{dt^2} + \left(L\frac{dj}{dv} + Rc\right)\frac{dv}{dt} + L\frac{dc}{dv}\left(\frac{dv}{dt}\right)^2 + v = V_s - Rj\,. \tag{1}$$

We can put any diode characteristics $j(v)$ in this equation. For the numerical integration, a current-voltage curve was chosen which closely approximates experimentally observed Ge tunnel diode characteristics. It is composed of three components: the tunneling, the excess current and the diffusion component. We employ Kane's theoretical expression for the tunneling (direct tunneling case),[3] one-tenth of the peak current I_p for the excess current and diffusion current of the form $K\exp(40\,v)$. This curve is always used in the present study, as illustrated in Fig. 2.

It should be mentioned that Eq. (1) is invariant in the following substitutions,

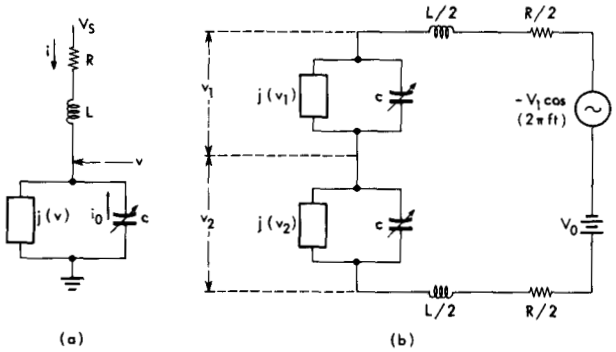


Figure 1 **Simplified equivalent single diode circuit (a), and twin diode circuit (b).**

$$c' = \alpha c,\ j' = \alpha j,\ L' = \frac{L}{\alpha} \text{ and } R' = \frac{R}{\alpha} \tag{2a}$$

or

$$c' = \beta c,\ L' = \beta L \text{ and } t' = \beta t\,, \tag{2b}$$

where α or β is a numerical constant. Therefore we may deal only with a particular diode, of which $c_0 = 10^{-12}$ farad and $I_p = 1$ ma, without losing generality.

The fundamental circuit of Goto's twin diode switching may be represented by the circuit in Fig. 1(b). The following nonlinear simultaneous equations describe the circuit:

$$Lc\frac{d^2v_1}{dt^2} + \left(L\frac{dj(v_1)}{dv^1} + Rc\right)\frac{dv_1}{dt} + L\frac{dc}{dv_1}\left(\frac{dv_1}{dt}\right)^2$$

$$+ Rj(v_1) + v_1 + v_2 = V_0 - V_1\cos(2\pi ft)\,,$$

$$Lc\frac{d^2v_2}{dt^2} + \left(L\frac{dj(v_2)}{dv_2} + Rc\right)\frac{dv_2}{dt} + L\frac{dc}{dv_2}\left(\frac{dv_2}{dt}\right)^2$$

$$+ Rj(v_2) + v_1 + v_2 = V_0 - V_1\cos(2\pi ft)\,, \tag{3}$$

where notations are indicated in the Figure. A sinusoidal power source $-V_1\cos(2\pi ft)$ for excitation is employed together with dc bias V_0 instead of pulse wave, where V_0 and V_1 are 0.22 and 0.14 v, respectively.

In the linear approximation that both c and g are constant, where g is a negative conductance, we obtain from Eq. (1),

$$\frac{d^2v}{d\tau^2} + (X + Y)\frac{dv}{d\tau} + (XY + 1)v = \text{const.}\,, \tag{4}$$

where

$$\tau = \omega t,\ \omega = \frac{1}{\sqrt{Lc}}$$

$$X = \frac{g}{\omega c}$$

and

$$Y = \frac{R}{\omega L}\,.$$

The same type of equation will also be written for the current i. The solutions are, obviously,

$$v \text{ or } i = A\exp(\alpha_1\tau) + B\exp(\alpha_2\tau) + \text{const.}\,, \tag{5}$$

where α_1 and $\alpha_2 = \frac{1}{2}[-(X + Y) \pm \sqrt{(X - Y)^2 - 4}]$. The dimensionless parameters X and Y, of which the inverses represent the device time constant and the external-circuit time constant, respectively, will determine all behaviors of $v(t)$ or $i(t)$. For example, the conditions for the standing sinusoidal oscillation are easily derived as

$$X + Y = 0,\quad XY + 1 > 0\,,$$

that is,

$$\frac{R}{L} = -\frac{g}{c}, \quad 1 + gR > 0,$$

and the frequency f is obviously given by the following well-known equation:

$$2\pi f = \omega\sqrt{1 - X^2}$$

$$= \frac{1}{\sqrt{Lc}}\sqrt{1 + gR}. \tag{6}$$

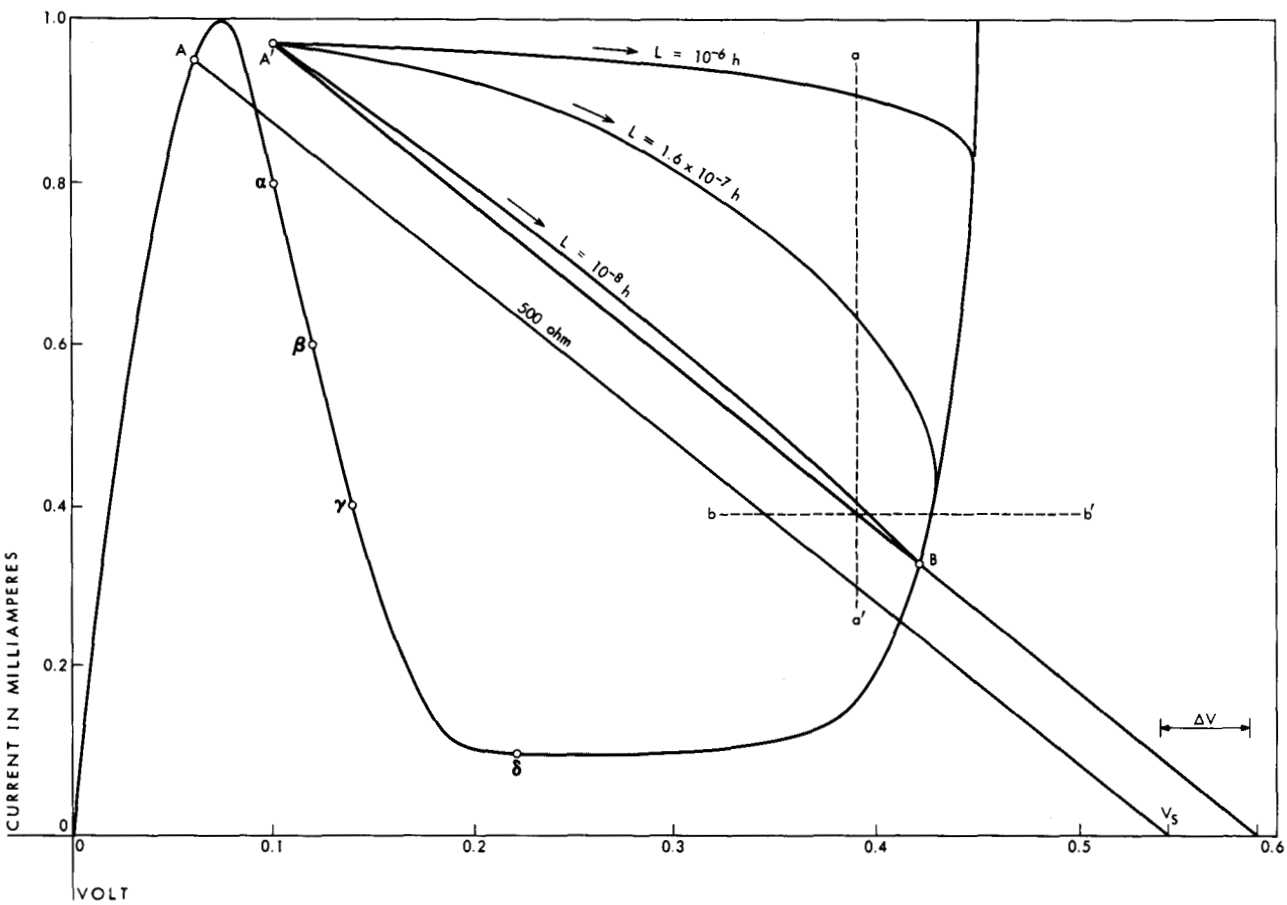
Results

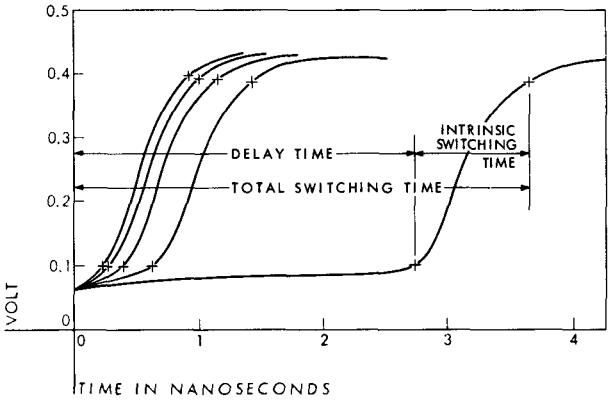
Flip-flop switching

The bistable switching operation is obtained with suitable choice of V_s and R for a given diode. The value L will have a large effect on the transient behavior. We obtained v and i as functions of time through the numerical integration of the Eq. (1). Figure 2 shows a conventional plot of v versus i of the switching transient at three values of L. In this Figure, we started from A' instead of A to eliminate the triggering. The switching time is customarily defined by the time when either the voltage change or the current change reaches 90% of the total change expected. Therefore we drew two lines a-a' and b-b' at the places of 90% change in the current and the voltage, respectively. The v-i trajectory starting from A', initially, hits the voltage line b-b' and then hits the current line a-a'. However, the current switching time may be more significant than the voltage one from a practical viewpoint, because of output power considerations, though the two switching times are practically the same in the case of small L or small L/R as in the case of $L = 10^{-8}$ h in Fig. 2.

The choice of triggering is very important in order to obtain the proper switching time. We chose rather arbitrarily a bias point A of 0.95 ma and 0.06 v to begin with, while the peak point is 1 ma and 0.075 v. The triggering voltage V_{in} or current I_{in} should be large enough to surmount the peak point; that is, $V_{in} > 0.038$ v or $I_{in} > 0.075$ ma for $R = 500$ ohms. The triggering voltage and current are defined here as step-function voltage and current sources connected in series and parallel, respectively, with the power supply. Figure 3 shows the transient behavior in the switching at triggering voltages 0.043, 0.063, 0.083,

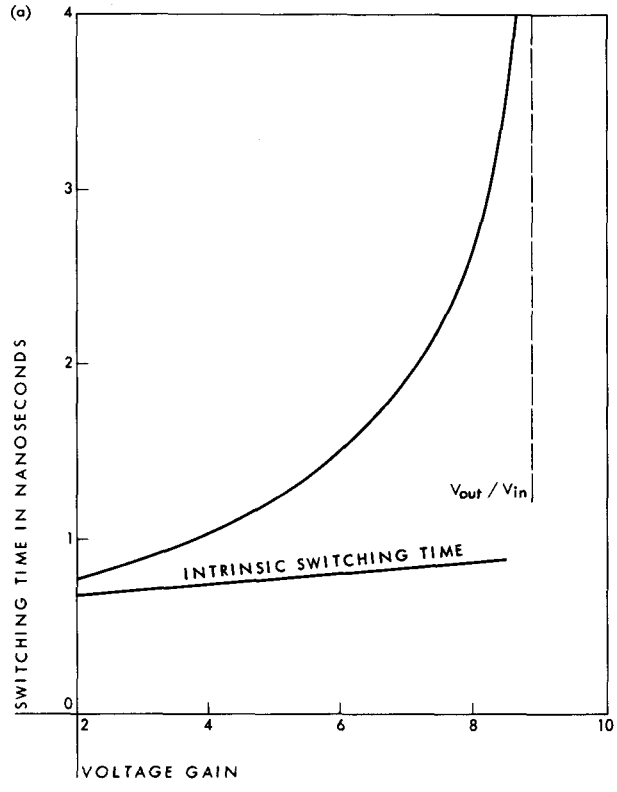
Figure 2 **Diode current-voltage characteristic v-j and switching transient v-i for $L = 10$, 160 and 1000 nh ($R = 500$ ohms, $c_0 = 1$ pf).**

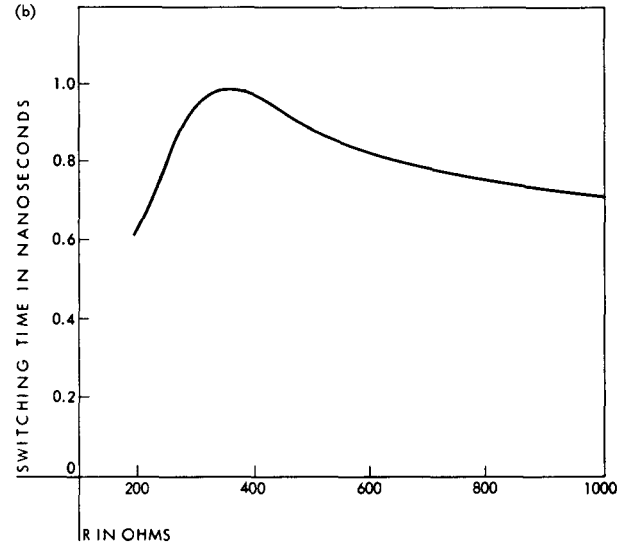




← *Figure 3* **Switching transient v(t) for various triggerings in the case of $L = 10$ nh and $R = 500$ ohms.**

Figure 4 **(a) Switching time (the upper curve) and intrinsic switching time (the lower curve) versus voltage gain V_{out}/V_{in}. (b) Switching time versus load resistance R.**





0.103 and 0.123 v from right to left respectively, given by step functions for the case of $L = 10^{-8}$ h and $R = 500$ ohms in Fig. 2. The excess voltage or current above the required minimum triggering input always plays a dominant role, whatever the bias point. The upper curve in Fig. 4(a) indicates how the switching time depends on the triggering voltage V_{in} in the terms of the ratio V_{out} to V_{in} (voltage gain). We would like here to introduce a concept which we call *the intrinsic switching time*, that is, the transition interval, which is the time required for from 10% to 90% change in output voltage or output current. The lower curve in Fig. 4(a) is this switching transition time, which does not seem to vary strongly with the triggering voltage and therefore is used in the following analysis. Figure 4(b) shows the intrinsic current switching time versus the load resistance R, which indicates a broad maximum around $R = 350$ ohms. The reverse switching behavior was analyzed in the same way. No significant difference was noticed.

We next calculate the intrinsic current switching time over the wide range of the inductance L at the fixed resistance R ($= 500$ ohms). The results are shown with curves indicated as "single" in Fig. 10, where the switching time was plotted against L/R on the logarithmic graph. Once we get one curve for the diode of $c\langle r\rangle = 10^{-10}$ sec, we can obtain a curve for a diode having any figure of merit simply with parallel transfer, using the substitutions (2b). The figure of merit is frequently indicated by the ratio of the capacitance c to the peak current I_p. The $c\langle r\rangle$ product 10^{-10} sec approximately corresponds to 1 pf/ma in the germanium tunnel diode. It is notable that the curves in Fig. 10 consist of flat regions and 45° slope regions. In the former region, the switching time is determined by the device time constant $c\langle r\rangle$, while in the latter region, that is subject to the external time constant L/R.

• *Oscillation*

One can expect a sine-wave oscillation when the diode is biased at the negative resistance region, and the following inequalities are satisfied:

Table 1 **Oscillation frequencies calculated for bias points α, β, and γ in Fig. 2; oscillation powers dissipated in resistance R for fundamental frequency and for 2nd and 3rd harmonics.**

	α-point r −80Ω	β-point r −90Ω	γ-point r −110Ω
R = 20Ω			
Osc. Frequency (Gc)	1.36	1.39	1.40
Osc. Power (μw)			
fundamental	13.52	13.76	19.88
2nd harmonic	0.123	0.063	0.023
3rd harmonic	0.0095	0.0019	0.0143
R = 40Ω			
Osc. Frequency (Gc)	1.28	1.30	1.31
Osc. Power (μw)			
fundamental	16.89	19.43	22.52
2nd harmonic	0.097	0.040	0.003
3rd harmonic	0.0135	0.0162	0.0152
R = 60Ω			
Osc. Frequency (Gc)	1.10	1.11	1.12
Osc. Power (μw)			
fundamental	14.89	16.62	18.55
2nd harmonic	0.062	0.027	0.006
3rd harmonic	0.0157	0.0172	0.0135

$$0 \geqq X + Y$$

and

$$2 > Y - X ,$$

according to the solution (5) of the linear approximation. We made a numerical integration of the nonlinear equation for the biased points α, β and γ of the negative resistance region in Fig. 2. The inductance, $L = 10^{-8}$ h, is fixed and the three resistances, R = 20, 40 and 60 ohms, are employed. The oscillation was sinusoidal, mixed with higher harmonics. The fundamental oscillation frequencies are listed in Table 1, together with the oscillation powers dissipated in the resistance R for the fundamental, the second and the third harmonic. One can see that the oscillation power in the fundamental frequency increases from the α-point toward the γ-point, while the power in the second harmonic decreases by such movement of the bias point, because of the decrease of the curvature radius in our diode characteristics. Also one can get the maximum power when the load resistance R is about half the negative resistance r. It is clear from Eq. (6) that R should be small in order to get really high-frequency oscillation. The actual diode, however, must have some series resistance R_s, which may depend on the structure and the substrate material. The ratio of the average negative resistance $\langle r \rangle$ to R_s may be 10 in an ordinary diode and may be much less in the high-figure-of-merit diode. Therefore R in Fig. 1 should be divided into two components, the load resistance R_1 and the series resistance R_s, and the really available output power may be only that dissipated in R_1. This sort of argument clearly informs us that the ratio $\langle r \rangle / R_s$ should increase in order to get high oscillation power at the high frequency. However, in most cases, we see the very small ratio $\langle r \rangle / R_s$ in the extremely high-figure-of-merit diode because of the small junction area. We carried out the output power calculation in the case of R_s = 10 ohms for the same diode (I_p = 1 ma, $c_0 = 10^{-12}$ farad and $\langle r \rangle \sim$ 100 ohms) at various inductance and load resistance values, which determine the frequency. The output power was plotted against the frequency in Fig. 5. A slotted envelope indicates the maximum power output attainable by choosing suitable L and R_1. As long as $\langle r \rangle / R_s$ remains constant, this diagram can be simply extended to diodes having other peak currents or figures of merit by shifting the power level or the frequency according to the rules (2a) and (2b).

We get interesting oscillation performance, unpredictable from linear theory, if the diode was biased at the δ-point in Fig. 2, a little beyond the negative-resistance region. Generally speaking, a standing-wave oscillation of the tunnel diode in such a circuit as as shown in Fig. 1 should satisfy the following equation:

$$\int_t^{t+T} (vj + Ri^2)dt = 0 ,$$

from the requirement of rf energy conservation, where T is the period of one cycle. One may get a set of the

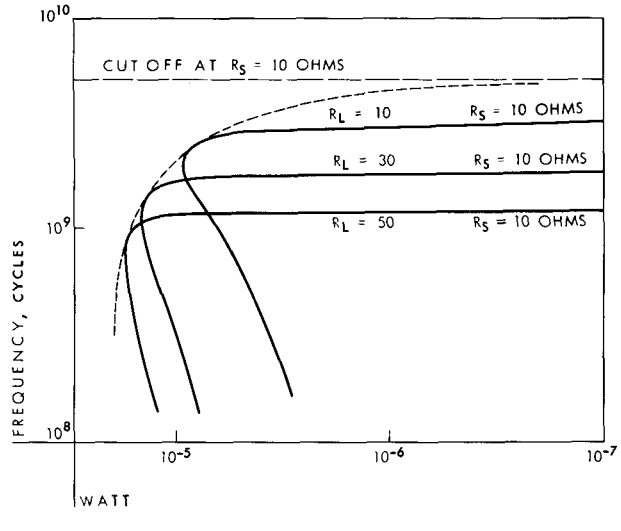


Figure 5 **Oscillation power versus frequency for I_p = 1 ma, c_0 = 1 pf and R_s = 10 ohms ($c\langle r \rangle \sim 10^{-10}$ sec).**

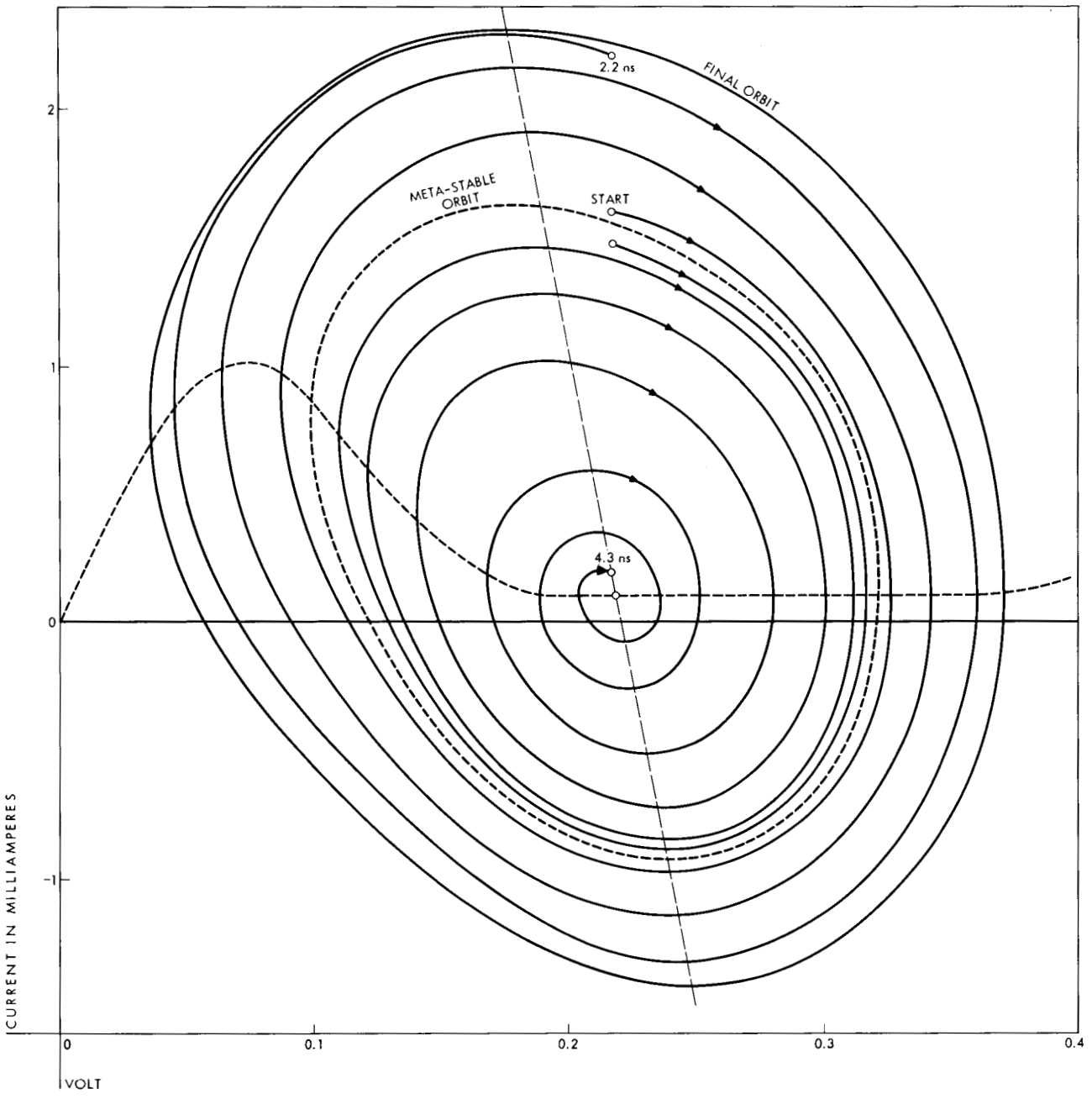


Figure 6 **Transient *v-i* of the "Class C oscillation" at large and small triggering kicks.**

unique solutions $i(t)$ and $v(t)$ satisfying the above equation if the diode is biased just inside the negative-resistance region, while one can obtain two sets of the solutions $i_m(t)$, $v_m(t)$ and $i_f(t)$, $v_f(t)$ if that is biased at the voltage range from 0.18 through 0.24 v at the end of, or beyond, the negative-resistance region. One set of the solutions $i_m(t)$ and $v_m(t)$, represents a metastable state, while the other set $i_f(t)$ and $v_f(t)$ is a stable solution. Figure 6 shows the *v-i* trajectories of the transient, the final state $[i_f(t), v_f(t)]$ and the metastable state $[i_m(t), v_m(t)]$, which were obtained from Eq. (1) for $V_s = 0.22$ v, $R = 20$ ohms, and $L = 10^{-8}$ h. If the trigger kick is inside the i_m-v_m orbit, it will "die out", while, if it extends to the outside of that orbit, a sustaining oscillation of 1.5×10^9 cycles will be obtained after a few turns, of which the rf power dissipated in the load R is about 35×10^{-6} w, a relatively large quantity. Figure 7 illustrates the behavior of both the "building-up" and the "dying-out" modes, according to the magnitude of the triggering kick. This type

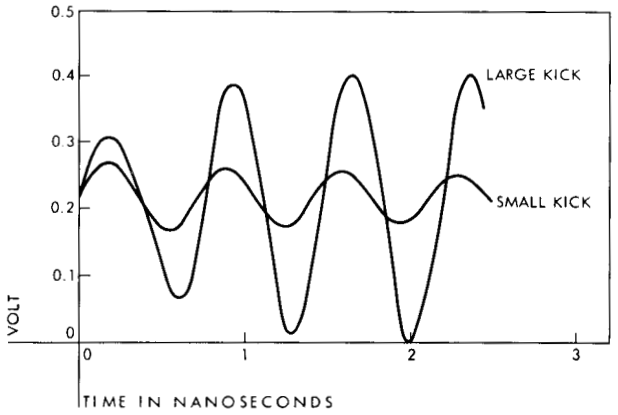


Figure 7 **Transient $v(t)$ of the "Class C oscillation" at large and small triggering kicks.**

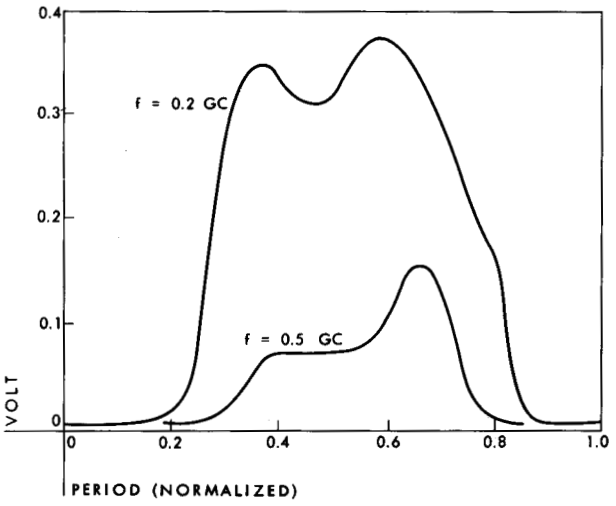
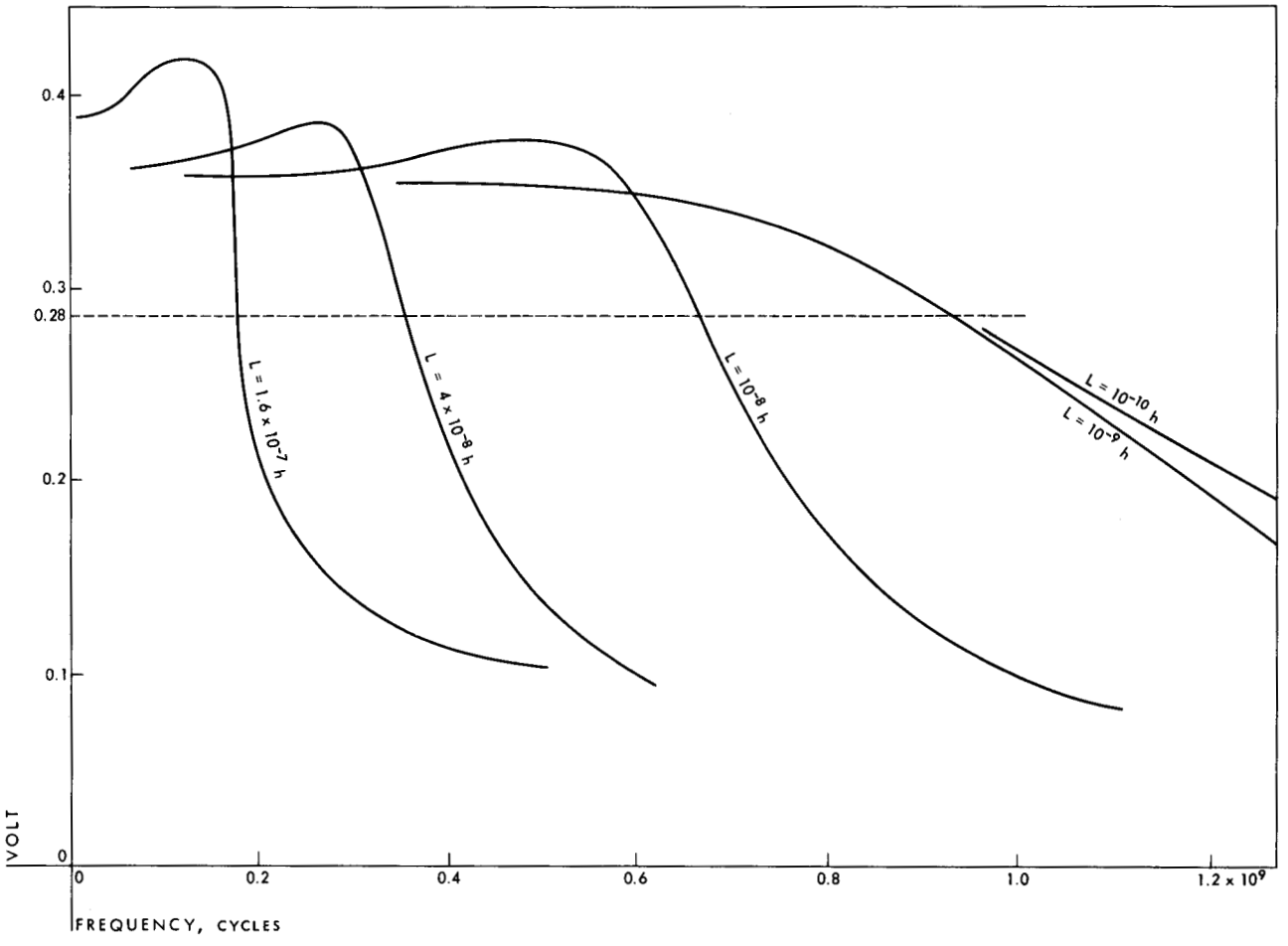


Figure 8 **Output pulse shape $v_1 - v_0$ of Goto's twin diode circuit at rf exciting frequencies 200 mc and 500 mc for $L = 40$ nh and $R = 10$ ohms.**

of oscillation was first found experimentally by R. F. Rutz,[4] who called it "Class C oscillation". This phenomenon was later studied on a mathematical basis by Flatto and Miranker.[5]

Figure 9 **Maximum output pulse voltage at the initial cycle versus the rf exciting frequency at $L = 160$ 40, 10, 1 and 0.1 nh. ($R = 10$ ohms).**



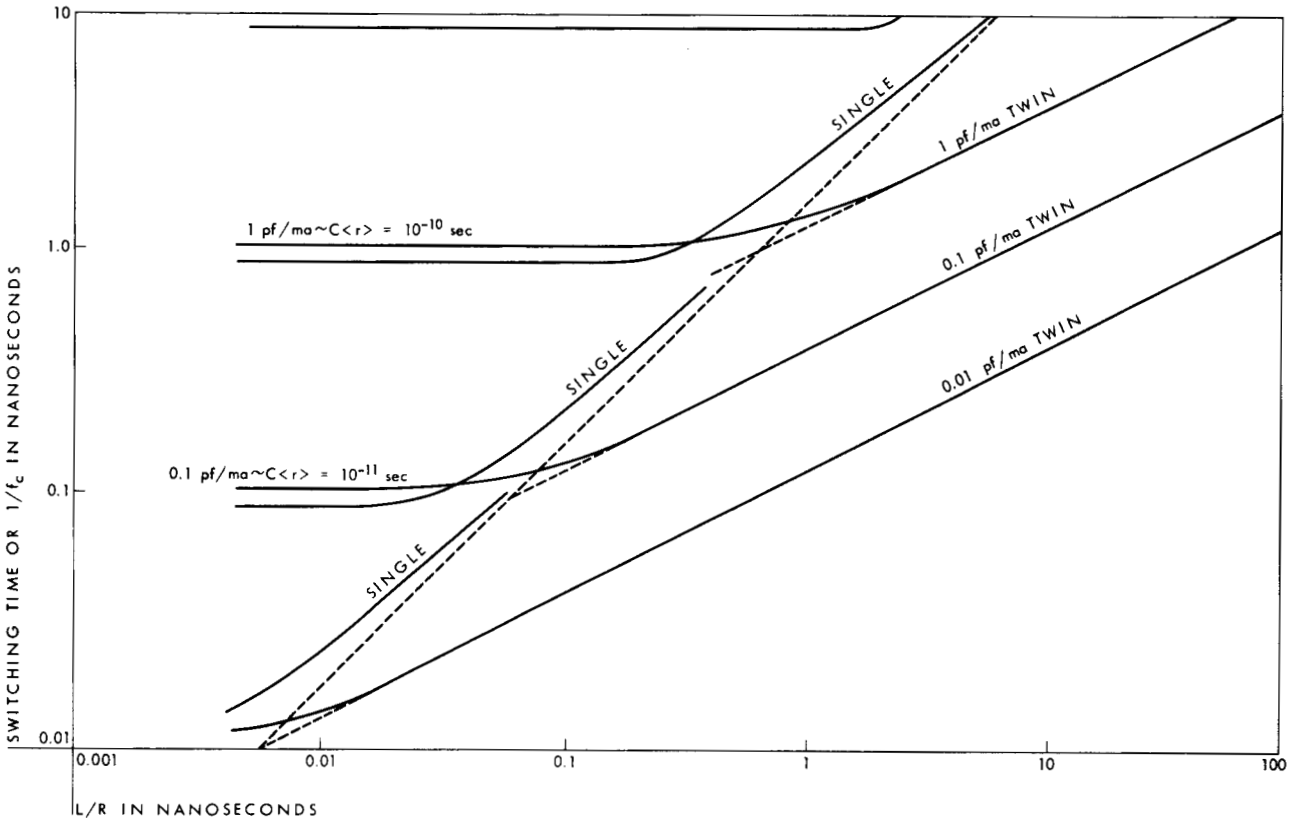


Figure 10 **Intrinsic switching time for single diode and reciprocal cut-off frequency for twin diodes versus L/R at various figure-of-merit diodes.**

•*Goto twin switching diodes*

The fundamental circuit of the Goto's twin-diode may be drawn as shown in Fig. 1(b).[6] The purpose of our present analysis is to obtain a cut-off frequency f_c in this rf excitation scheme by means of the numerical integration of the nonlinear simultaneous equations (3). One might expect f_c to depend not only on the figure of merit of the device but also on R and L in the external circuit. The definition of such a cut-off frequency and the triggering condition also may admit of some discussion.

First of all, let us assume that the two diodes are exactly identical and the external circuit is also symmetrical. If the current level of either one, for instance, $j(v_2)$, is raised by 10% peak current, that is, 0.1 ma, by triggering, one can get a pulse-like repetition output voltage $v_1 - v_2$ at the middle point between both diodes, as shown in Fig. 8. The larger pulse and the smaller one represent the cases of $R = 10$ ohms, $L = 4 \times 10^{-8}$ h, $f = 2 \times 10^8$ sec^{-1} and the same values of R and L, $f = 5 \times 10^8$ sec^{-1}, respectively, where the time scale is normalized in the Figure. The output voltage $v_1 - v_2$ at the initial cycle was then plotted. The wave shape was not much different, even after a number of cycles, unless the frequency is extremely high. Small variations of the triggering depth and the external resistance R (including the series resistance R_s) did not greatly affect the general switching performance, so the above-mentioned depth (10%) and resistance (10 ohms) are kept constant in the following analysis.

We will now determine the cut-off frequency f_c over the wide range of the inductance L from 10^{-10} through 10^{-6} h. Figure 9 shows the maximum of the difference $v_1 - v_2$ at the initial cycle versus the exciting frequency f for various values L. The cut-off frequency may be determined by the requirement that $v_1 - v_2$ should be greater than a certain level in order to transfer properly the signal pulse to the next stage. Here, let us tentatively define the cut-off frequency f_c as that frequency at which $v_1 - v_2$ is reduced to 0.707 of its low-frequency value. The other requirement for f_c may be that the maximum of the output pulse should come at the proper time, for example, before the lapse of two-thirds of one cycle, which was verified to be satisfied under our definition. The results are illustrated with curves indicated as "twin", where $1/f_c$ was plotted against L/R on the logarithmic graph of Fig. 10. Each curve consists of two regions, flat and slope. In the former region, f_c is determined by the device's figure of merit $c\langle r\rangle$ or c/I_p (pf/ma), while, in the slope

region, f_c is proportional to R/L. Furthermore, it should be mentioned that $1/f_c$ is approximately equal to the intrinsic switching time in the single diode flip-flop at the device-limited condition (flat region) and also f_c is just about half the frequency for oscillation given by formula (6)

$$\left(f_c \sim 0.5 \times \frac{1}{2\pi\sqrt{Lc}}\sqrt{1+gR}\right)$$

at the circuit-limited condition (slope region).

Conclusion

The $c\langle r\rangle$ time constant or c/I_p of the device figure of merit may be determined by only the impurity concentrations of both p and n sides for a certain semiconductor material if the ideal sharp step-junction is assumed. This, of course, depends very much on metallurgical processes (heavily doped crystal growing, alloying, et cetera) in diode fabrication. The highest value in germanium tunnel diodes of practical use may be 10^{-11} sec (0.1 pf/ma). For several promising compound semiconductors, InSb, InAs, GaSb and GaAs, the limit would seem to be considerably extended. Though the present analysis is based on the germanium tunnel diode, it can easily be extended, with minor modifications, to other materials. The relationship between the $c\langle r\rangle$ product and c/I_p does depend on the material.

Now we can answer the questions raised at the beginning of this paper. If we can get a germanium tunnel diode of $I_p = 1$ ma, $c_0 = 0.1$ pf and $R_s = 10$ ohms ($c\langle r\rangle \sim 10^{-11}$ sec), the power output dissipated in the load resistance $R_1 = 10$ ohms is approximately 10 μw at 10 Gc, from Fig. 5. With the same tunnel diode, 0.1 nsec switching may be predicted with the inductance less than 10 nh at $R = 500$ ohms, from Fig. 10.[7]

Of course, many complications occur in actual computer switching circuits, especially because of coupling one to the next, the distributed nature of both the circuit and the device, the tolerance of the diode characteristics, et cetera. Even so, it might be worth while to analyze the performance of each unit carefully and to determine whether the circuit switching time is device-limited or circuit-limited. Finally, the author would like to point out the desirability of choosing tunnel diodes of which the figure of merit is well matched with the system requirement.

Acknowledgment

The author would like to thank R. A. Willoughby and H. Serenson for carrying out the numerical integration[8] with the IBM 7090 and G. R. Gunther-Mohr, R. F. Rutz, C. Lanza, S. Triebwasser and other colleagues in this laboratory for stimulating discussions.

References and footnotes

1. L. Esaki, *Phys. Rev.* **109**, 603 (1958).
2. M. Schuller and W. W. Gärtner, *Proc. IRE*, **49**, 1268 (1961).
3. E. O. Kane, *J. Appl. Phys.* **32**, 83 (1961).
4. R. F. Rutz, private communication.
5. L. Flatto and W. L. Miranker, "Limit Cycle Studies for Circuits Containing One Tunnel Diode," IBM Research Report RC-551.
6. E. Goto et al., *IRE Trans. Elect. Computers* **EC-9**, 25 (1960).
7. In less than 1 nsec switching, we might expect that a certain delay effect will come in because of some time constant in the excess current (private communication from P. J. Price and F. F. Fang) and also because of some minority carrier effects in the diffusion current. In such a case, this discussion should be modified.
8. H. Serenson and R. A. Willoughby, "Numerical Integration of a Tunnel Diode Circuit," IBM Research Report RC-413.

Received December 5, 1961