# NEGATIVE RESISTANCE

Resistance is the measure of allowed current passing through a component when a terminal voltage is applied. Ohm's Law defines that the resistance is the ratio of the applied voltage to the resulting current,

$$R = \frac{V}{I} \tag{1}$$

This simple $I$–$V$ relationship is shown in Fig. 1(a). Sometimes a component deviates from having a linear relationship, resulting in a nonlinear curve. An example of a rectifier is shown in Fig. 1(b). Such resistance is referred to as being nonlinear, as opposed to linear. It is also helpful to distinguish the static resistance ($V/I$) from the dynamic resistance or differential resistance (the slope of the $I$–$V$ curve as $dV/dI$). So in a linear resistor, the static resistance is the same as the dynamic resistance ($V/I = dV/dI$), whereas in a nonlinear resistor, $V/I \neq dV/dI$.

It is understood that the term negative resistance refers to the dynamic characteristic, that is, the slope of the $I$–$V$ curve ($dV/dI$) is negative. Notice that in the example of a common negative resistance [Fig. 1(c)], only $dV/dI$ is negative, but the static resistance is always positive. That is why negative resistance and negative differential resistance are used interchangeably. Negative resistance can also be classified into two shapes: (1) N-shape negative resistance, as shown in Fig. 1(c), and (2) S-shape negative resistance, as shown in Fig. 1(d). The S-shape negative resistance is typical behavior of a switch.

Most of the devices exhibiting negative resistance have two terminals, but there are also devices with three or more terminals. An example is a thyristor. The function of the extra terminal is to control the shape of the $I$–$V$ characteristics between the two terminals that carry the majority of the current. Specifically, for the S-shape curve shown in Fig. 1(d),

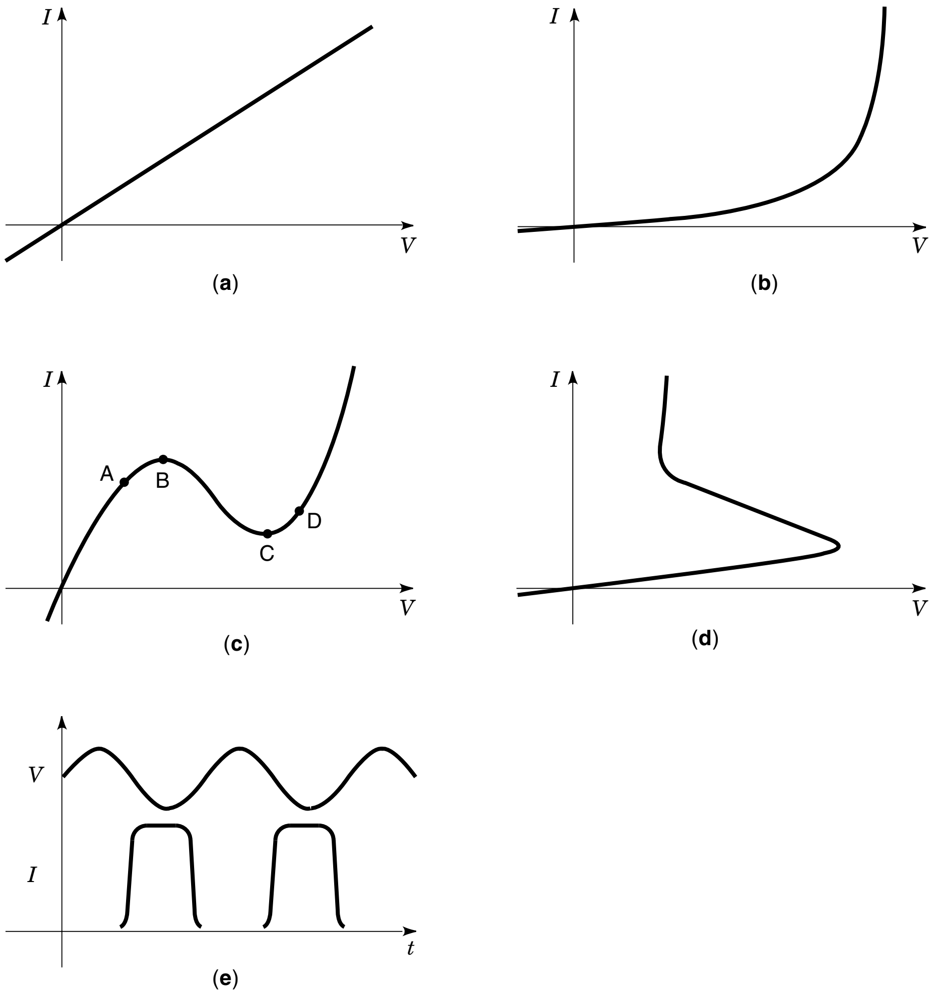


**Figure 1.** $I$–$V$ relationship showing (a) linear resistance, (b) nonlinear resistance, (c) N-shape differential negative resistance, and (d) S-shape negative differential resistance. (e) In a transit-time device, negative resistance comes from the phase difference between terminal voltage and current.

the triggering (breakover) voltage or current before negative resistance sets in can be varied.

There is yet another group of devices where the negative resistance does not come from a negative slope in the $I$–$V$ curve. In these devices, when an ac voltage is applied, the resulting ac current is not in phase with the applied voltage. The majority of this phase delay comes from the transit time of charge which is generated within the device and subsequently travels to the boundaries of the devices and to the terminals. For this reason, these devices are called transit-time devices. The phase relationship between the applied voltage and the resulting current is shown in Fig. 1(e). The unique feature is that when the small-signal voltage is positive, the small-signal current is negative, giving rise to negative ac power absorbed. So this type of negative resistance comes from $\vec{V}/\vec{I}$ being negative, as opposed to $dV/dI$ being negative. It is worthwhile to mention that both inductor and capacitor cause a phase shift between the ac voltage and current. The crucial requirement here is that in order to have net ac power gain, a phase shift of more than 90° is necessary, which is not achievable with either a capacitor or an inductor.

In the following sections, the negative resistance devices originating from $dV/dI$ being negative are discussed, followed by their applications. Finally, similar discussions on the transit-time devices are presented.

## TUNNEL DIODE

Tunneling is a quantum mechanical process whereby electrons with insufficient energy to surmount a thin energy barrier tunnel directly through it. In a semiconductor $p$–$n$ junction, significant tunneling occurs when both sides of the junction are heavily doped so that the depletion region (energy barrier) is thin. Also, since the tunneling process requires conservation of energy, the tunnel diode (Esaki diode) is fabricated such that both sides of the junction are degenerately doped so that states of equal energy exist on both sides of the junction (in the conduction and valence bands). Figure 2(a) illustrates the situation under low forward bias where the bottom of the conduction band on the $n$-side and the top of the valence band on the $p$-side overlap and significant tunneling occurs. At higher bias, the overlap becomes smaller and the tunneling current decreases, thereby producing the negative differential resistance (NDR) feature mentioned earlier. Eventually, at still higher forward voltages, when the overlap disappears as in Fig. 2(b), the normal diffusion current becomes dominant and the current increases with bias once again. A useful figure of merit for NDR devices is the peak-to-valley ratio, which is a measure of the ratio of the peak current obtained just before the onset of NDR [point B in Fig. 1(c)] to the minimum current at the termination of the

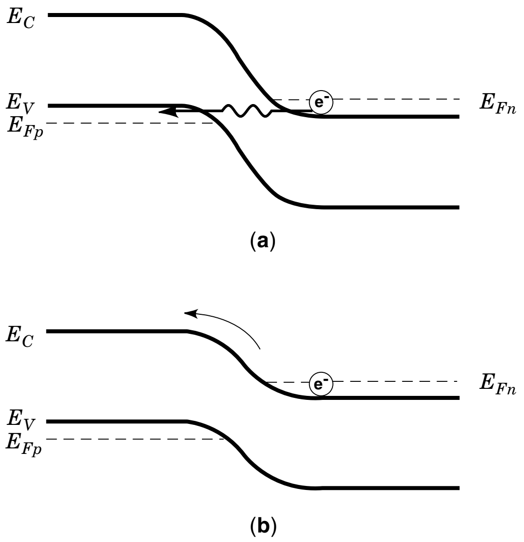


**Figure 2.** Band diagram showing how forward bias voltage affects the tunneling current in a tunnel diode. In (a), the junction is slightly forward biased so that there is still an overlap of states at the same energy in the conduction and valence bands. In (b), the forward bias is sufficient to remove the overlap and the current consists of carriers which surmount the barrier (diffusion current).

NDR region [point C in Fig. 1(c)]. Peak-to-valley ratios of 3 to 20 are typical for semiconductor tunnel diodes.

## RESONANT TUNNELING STRUCTURES

A semiconductor quantum well is a structure formed when materials with either a conduction or valence band discontinuity are joined such that the discontinuity creates a potential well. For example, a well formed in the conduction band is accomplished by sandwiching a layer of high electron affinity material in between layers of lower electron affinity material. The conduction band discontinuity serves as the barrier in the well. In the case of a resonant tunneling structure, the quantum well exists between two very thin, low electron affinity layers that have a large discontinuity and serve as the well barrier. For example, in the GaAs/AlGaAs system, as shown in the inset of Fig. 3, the GaAs serves as the quantum well and the outside contact regions, while the barriers are made from AlGaAs. As the well is made thin, the continuum of states in the conduction band become discrete levels of subbands according to the following expression:

$$E_n - E_{\mathrm{C}} = \frac{h^2 n^2}{8m^* W^2}, \qquad n = 1, 2, 3\ldots \tag{2}$$

where $E_{\mathrm{C}}$ is the conduction band edge, $h$ is Planck's constant, $m^*$ is the effective mass, and $W$ is the thickness of the well. The band diagram of Fig. 3 shows how significant current only flows when $E_C$ of the contact region aligns with the first subband energy. As the bias is further increased, the tunnel current diminishes, leading to a decrease in total current and thus an NDR region in the $I$–$V$ characteristic. Depending on the depth of the well (height of the barriers), tunneling through higher subband energies can also occur, leading to additional NDR regions.

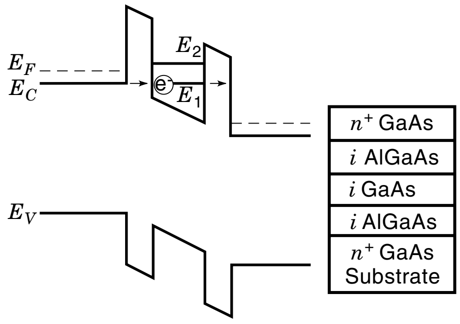


**Figure 3.** Band diagram of a resonant tunneling diode under bias conditions such that the conduction band edge aligns with the first subband level, $E_1$, causing a tunneling current to flow. The layer structure is shown in the inset.

A resonant tunneling bipolar transistor (RTBT) results when the quantum well is placed within the emitter or base of a bipolar transistor. A resonant tunneling hot electron transistor (RHET) has a similar structure without the emitter–base $p$–$n$ junction. The advantages of such three-terminal structures are that the negative resistance is tunable by the base bias, and the output is isolated from the input.

## SEMICONDUCTOR SWITCHES

A thyristor is a four-layer Si device consisting of an $n$–$p$–$n$–$p$ structure. The structure is called a silicon controlled rectifier (SCR) when a gate contact is made to the middle $p$ region, as shown in the inset of Fig. 4. As positive voltage is applied to the anode, a small reverse bias current flows until either the depletion regions of the central junctions join (a condition known as punchthrough), or avalanche multiplication occurs at the reverse biased junction. Beyond this "breakover voltage" ($V_{\mathrm{BO}}$), a large current flows and the voltage across the device drops. An S-shape negative resistance

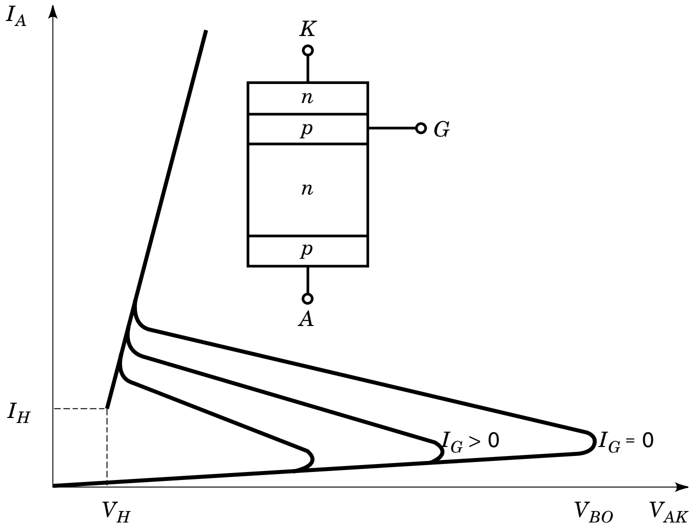


**Figure 4.** SCR current–voltage characteristics illustrating how positive gate current lowers the breakover voltage. The device structure is shown in the inset.

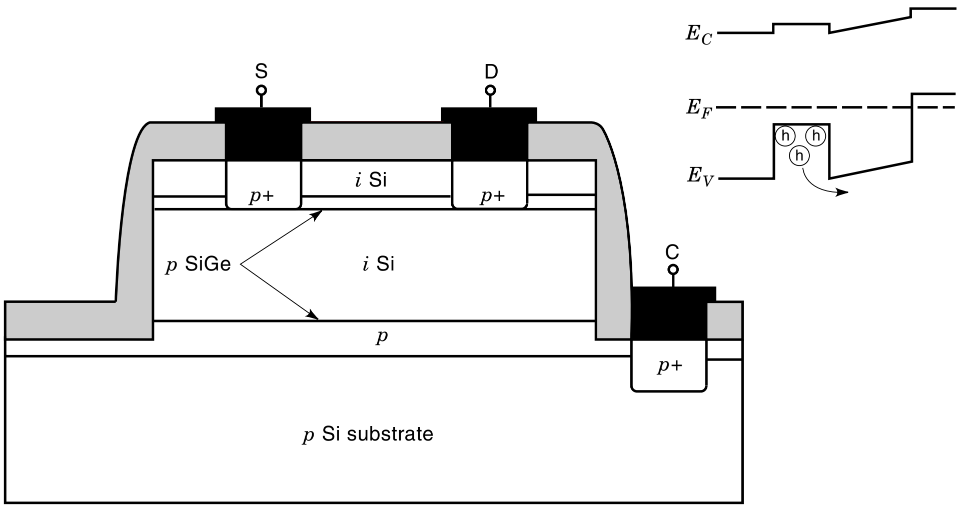


**Figure 5.** Charge injection transistor (CHINT) device structure for the Si/SiGe materials system. The top channel consists of small bandgap SiGe separated from the collector by an undoped Si barrier region. The inset schematically exhibits the device band diagram in the direction perpendicular to the channel.

region arises in this device as it switches from the high-impedance forward-blocking state to the low-impedance on state. With positive gate current, the breakover voltage is lowered, which is the mechanism used to trigger the device into the on state. Once the SCR is switched on, an applied bias lower than the holding voltage ($V_h$) will turn the device off.

Other semiconductor switches which possess S-shape negative resistance regions similar to the thyristor are the metal insulator semiconductor switch (MISS) and the planar doped barrier switch (1).

## GUNN EFFECT AND THE TRANSFERRED ELECTRON DEVICE

The transferred electron device directly exploits a particular material property of semiconductors like GaAs and InP, which have lower mobility satellite valleys to which electrons transfer at high applied electric fields. A uniformly doped length of these materials display normal positive ohmic characteristics for low fields, but for applied voltages that cause the electric field to exceed some critical field a reduction in the current results, as more electrons are transferred to the low mobility satellite valley. The critical electric field for GaAs is 3.2 kV/cm and 10.5 kV/cm for InP. This process of intervalley scattering can create instabilities within the semiconducting material with sufficient applied bias, which gives rise to microwave oscillations. The occurrence of these oscillations is known as the Gunn effect, named for its first observer.

The transferred electron device (TED) or Gunn diode is simply composed of a length of material with two contacts on the ends. Planar structures grown on semiinsulating substrates are also possible. The essence of operation of the TED lies with the generation of a charge accumulation region within the device. Once an adequate bias is applied across the length of the device, an accumulation region will form in an area where there may be a crystal defect or a difference in doping—something that causes the electric field to be slightly larger than in the rest of the device. Once the field in this region is sufficiently large to possess a negative differential mobility, a space charge instability will form and grow (either as a simple accumulation of electrons or a dipole consisting of electrons and ionized donors). Once the field in this unstable region rises above the point where it possesses a negative differential mobility, the dipole ceases to grow (matures) and continues to travel along the length of the device with the saturated drift velocity ($v_{sat}$). The result is charge pulses that arrive at the anode with a period of $L/v_{sat}$ where $L$ is the length of the device. The transferred electron device is widely used as oscillator for 1 GHz to 100 GHz applications.

## REAL-SPACE TRANSFER DEVICES

Real-space transfer is similar to the Gunn effect described above, except that electrons are transferred to a lower mobility material in real space at high fields rather than being transferred in momentum space. Unlike the Gunn effect, however, either electrons or holes can be transferred. In its most common implementation, a heterojunction barrier (e.g., AlGaAs/GaAs) and modulation doping form the basis of the structure. At low fields, carriers confined in the small bandgap material flow from one contact to the other exhibiting normal ohmic behavior. At sufficiently high bias, some carriers gain enough energy to surmount the barrier into the low-mobility, high-bandgap material. This transfer lowers the effective mobility and decreases the total current, which gives rise to an NDR region.

The charge injection transistor (CHINT) or negative resistance field effect transistor (NERFET) is a three-terminal de-

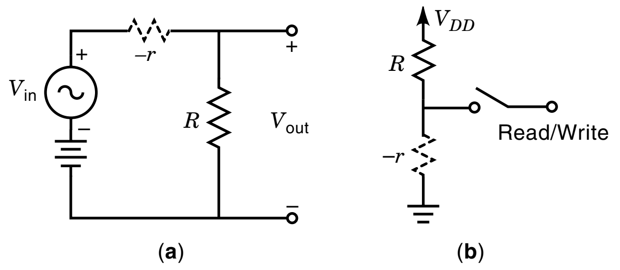


**Figure 6.** A simple negative resistance amplifier is shown in (a), while (b) exhibits the use of a bistable negative resistance device as a memory element.

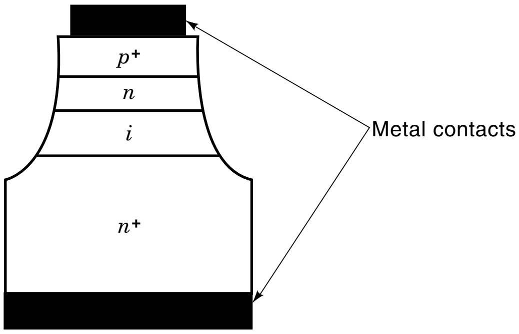


**Figure 7.** Structure of an IMPATT diode.

vice utilizing real-space transfer. The structure used in the Si/SiGe system is illustrated in Fig. 5. A lightly doped channel of small gap material (SiGe) exists between the source and drain. A barrier between the collector and the channel is formed by a layer of undoped Si, which has a larger bandgap than the SiGe. The inset of Fig. 5 illustrates the band structure of the CHINT device perpendicular to the channel. As voltage is applied between the source and drain, the carriers present in the channel produce a current flow just as in a normal FET. As the bias increases and the field across the channel increases, some carriers gain sufficient energy to surmount the barrier and are collected by the collector. This siphoning of carriers by the collector causes the total drain current to decrease with increasing drain voltage producing an NDR characteristic.

## APPLICATIONS

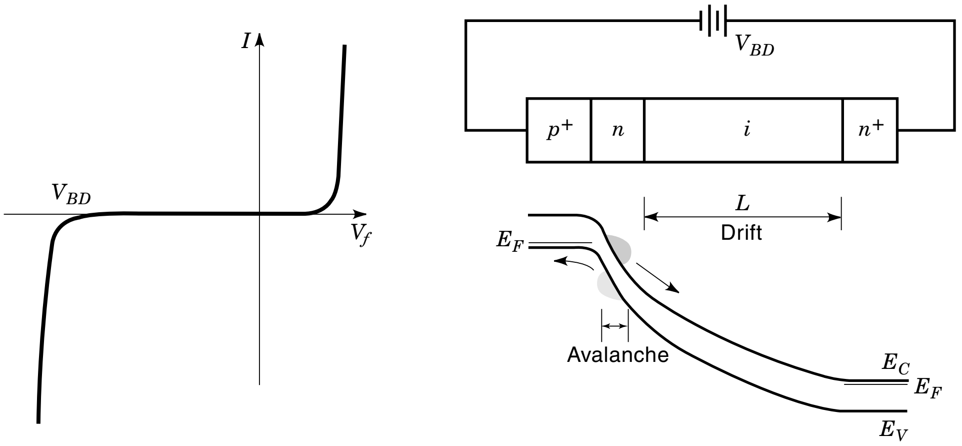
The most common uses of negative resistance are for circuit applications in oscillators, amplifiers, memory, and active filters. An oscillator is made by combining a negative resistance device together with a tuned *RLC* circuit so that the net resistance becomes zero. (The majority of *R* comes from parasitics.) The frequency of the oscillations is determined by the *LC* of the circuit. The series connection of a normal resistance and a negative resistance provides a simple example of a negative resistance amplifier. By connecting a negative resistance $(-r)$ in series with a normal resistor $(R)$ as in Fig. 6(a), and taking the output across the normal resistor, the voltage ratio becomes:

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{R}{R + (-r)} \tag{3}$$

which produces a voltage gain. Active filters for use on monolithic microwave integrated circuits (MMIC) are greatly improved with the use of a negative resistance to compensate for losses in inductors, transmission lines, and lossy dielectrics. In these filters, the negative resistance is combined with the $L$ and $C$ components to increase the overall quality factor $(Q)$, which leads to a sharper cutoff at the frequency band edges. Memory circuits using negative resistance exploit the fact that these devices exhibit bistable behavior. Figure 6(b) shows an elemental memory circuit with a negative resistance device loaded with a resistor $R$ from a supply voltage $V_{DD}$. The $I$–$V$ curve of Fig. 1(c) shows the two stable states, A and D, possible for a given loadline. By forcing the voltage across the negative resistance device high or low, the circuit is forced to one or the other of the stable states.

## TRANSIT-TIME DEVICES

In a transit-time device, there are two important mechanisms. The first is the generation and build-up of charge within the device, followed by transit of these charges across the length of the device. Although the transit of charge is quite similar for different transit-time devices, the generation or injection of charge is quite different among them. These mechanisms include avalanche multiplication, thermionic emission, and tunneling. The most popular device among them is the IMPATT (impact-ionization-avalanche transit-time) diode. This discussion will be confined to the operation of the IMPATT diode. Other transit-time devices are BARITT (barrier-injection transit-time) diode, TRAPATT (trapped-plasma avalanche-triggered transit) diode, DOVATT (double-velocity avalanche transit-time) diode, MITATT (mixed-tunnel-avalanche transit-time) diode, DOVETT (double-velocity transit-time) diode, TUNNETT (tunnel-injection transit-time) diode, and QWITT (quantum-well-injection transit-time) diode. Brief discussions on each of these can be found in Ref. (1).



(a)

(b)

**Figure 8.** (a) The $I$–$V$ curve of an IMPATT diode does not process a negative slope. (b) Schematic diagram showing where the charge is generated (in this case by avalanche multiplication), and the region of charge transit.

An IMPATT diode can be realized by different physical structures. All of these are variations of a *p–n* junction. The most common is called the Read diode and it is shown in Fig. 7. Common semiconductor materials are Si and GaAs. Here the *p–n* junction provides a high field for avalanche multiplication, and the large intrinsic layer is the region where transit time is originated. The dc characteristics are shown in Fig. 8(a). Notice that in the $I$–$V$ curve, there is no region where the slope is negative. During operation, the diode is reverse biased with a dc value near the breakdown ($V_{\mathrm{BD}}$), so that a small ac signal will drive it into avalanche multiplication. Referring again to Fig. 1(e), the delay of current with respect to the voltage comes from two components. The first is the time it requires to build up the charge internally. The second is the transit time across this intrinsic region. This charge build-up time is a characteristic of avalanche multiplication, and it is absent in other injection mechanisms such as tunnelling and thermionic emission. That is why the IMPATT diode is more efficient in power generation compared with, for example, a BARITT diode. After the charge is generated, it traverses the intrinsic region with saturation velocity ($v_{\mathrm{sat}}$). During the time of this transit, there is a continuous terminal current of magnitude

$$J = \frac{Q_{\mathrm{A}} v_{\mathrm{sat}}}{L} \tag{4}$$

where $Q_{\mathrm{A}}$ is the charge per area (cross-sectional), and $L$ is the region of the intrinsic layer. The magnitude of $Q_{\mathrm{A}}$ is related to the magnitude and frequency of the ac signal and detailed derivation is beyond the scope of this article. Assuming that the transit time dominates the phase delay, the frequency of operation is given by

$$f = \frac{V_{\mathrm{sat}}}{2L} \tag{5}$$

since the duration of the current pulse corresponds to the transit time of the charge packet, and this current pulse is roughly half of the cycle.

The main application of a transit-time diode is microwave generation in the 3 GHz to 300 GHz range. The transit-time devices are the most efficient microwave oscillators in this frequency range. Usually a transit-time device is a discrete component and it is mounted in a resonator cavity. When a dc bias is applied to the device, an ac output is produced. Applications of these oscillators are in radar systems and alarm systems. In high-power oscillators, a good heat sink is crucial.

## BIBLIOGRAPHY

1. K. Ng, *Complete Guide to Semiconductor Devices,* New York: McGraw-Hill, 1995.

Clifford A. King
Kwok K. Ng
Lucent Technologies

**NERVOUS SYSTEM.** See Neural nets based on biology.

**NETWORK ANALYSIS.** See Network equations.