TUNNEL DEVICES

Tunnel devices take advantage of the wavelike properties of charge carriers in implementing a desired function. The term *tunneling* refers to the movement of matter from one side of an energetic barrier to the other even though it does not possess sufficient energy to overcome the potential barrier according to the laws of classical mechanics. Instead, the quantum mechanical wave function of the particle penetrates inside the barrier and extends into the medium on the far side. Therefore some probability exists for the particle to be on the other side. For these wave properties of matter to be manifested, the particle mass and the tunneling distance must be small.

Tunnel devices typically possess nonlinear, nonmonotonic current-voltage characteristics and are characterized by very rapid fundamental propagation times. They find application in such diverse areas as microwave oscillators; multiple-level logic, switches, memory elements; and lasers. They are examples of functional devices whose underlying physical mechanisms are exploited and applied to meet a sophisticated demand simply. Often they replace a set of many interconnected devices in the form of a single device, which performs the desired function more naturally. In addition to playing an important role in modern electronics and optoelectronics, tunneling devices have played an important role in twentieth century science by offering direct, macroscopic evidence of microscopic quantum mechanical phenomena.

TUNNELING CONCEPT

The quantum mechanical concept of tunneling may be illustrated by a simple example. A particle with energy E impinges from the left on a potential barrier of height V, as illustrated in Fig. 1. The solution to the Schrödinger wave equation to the left of the barrier consists of traveling waves (the incident wave travels to the right and any reflected component travels to the left). Because the particle energy is stated to be less than the barrier height, the wave vector inside the barrier is complex. The corresponding solution is a sum of decaying exponentials. To the right of the barrier, the solution is a wave traveling to the right. Continuity of the wave function and its derivative at the boundary specifies the overall solution to within a multiplicative constant. There is a nonzero probability of finding the particle on the far side of the barrier.

Associated with each particular choice of the energy E of the incident particle is a transmission coefficient $T(E)$, defined as the ratio of the current density of particles transmitted

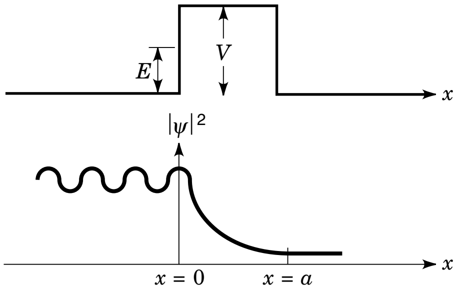


Figure 1. A particle with energy E impinges from the left on a potential barrier of height V and width a. In the case considered, $E < V$, and the wavefunction inside the barrier takes the form of a decaying exponential. The wavefunction penetrates all the way through the finite barrier and emerges in the form of a traveling wave on the far side, and it is possible for the tunneling probability to be appreciable.

J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*. Copyright © 1999 John Wiley & Sons, Inc.

through the barrier to the current density of particles with energy E impinging on the barrier. Because the wavefunction penetrating into the barrier decays exponentially, T will also decay exponentially with increasing barrier width and also with increasing difference between the barrier height and the energy of the incident particle. If T is much less than unity, it may be approximated as

$$T \sim 4\left(\frac{E}{V}\right)\left(1-\frac{E}{V}\right)\exp\left[-\sqrt{\frac{8m(V-E)}{\hbar^2}}a\right] \qquad (1)$$

where m is the particle mass and $\hbar^2$ is Planck's constant divided by 2π. In Fig. 2 the transmission coefficient is shown as a function of barrier width for a typical electron at room temperature with energy 40 *meV* and effective mass $m_e = 0.05\ m_0$. For this realistic range of parameters, the barrier width should be on the order of nanometers for T to be appreciable.

FUNCTIONAL DEVICES

Before the advent of tunnel devices, electronic devices (e.g., diodes, transistors) were dominated by physical mechanisms which gave rise to a monotonic dependence of outputs on inputs.

In 1965, J. A. Morton (1) of Bell Labs popularized the term *functional devices,* a family of which tunneling devices are natural members. Morton described traditional electronic circuits as consisting of vast numbers of interconnected transistors and other devices with simple, monotonic relationships between inputs and outputs. The equations that describe these relationships are mathematical approximations which arise out of physical interactions within matter. The relative simplicity and monotonicity of the resulting equations allows for representation using classical network equations employed in circuit function synthesis.

Morton argued that substantial inroads could be made by abandoning classical circuit concepts and exploiting instead the most basic interactions between energy and matter. Such functional devices would be designed to perform a desired function as simply as possible. The aim would be to reduce greatly the number of elements and process steps per function:

> . . . the aim of electronics should be not simply to reproduce physically the narrow elegance of classical circuit theory; rather, it should be to perform needed system functions as directly, as simply, and as economically as possible from the most relevant knowledge of energy–matter interactions (1).

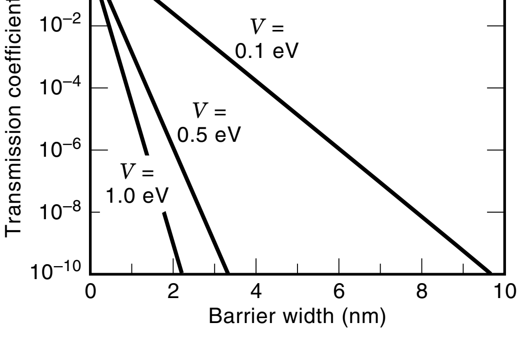


Figure 2. The transmission coefficient for a typical room temperature electron (E = 40 meV) through a finite potential barrier of various widths (a) and heights (V).

Despite such arguments for a natural and elegant approach to device innovation, in the 1960s functional devices had only gained acceptance in fulfilling niche applications. The slow rate of incorporating the devices at that time is the result of at least two important factors. First, it had been possible until that time and has been possible until recently to continue to extract additional functionalities by making larger, denser circuits by interconnecting simple *black boxes* of traditional circuit elements. Technological inertia contributed to the success of this brute force approach to a point. It is believed, however, that present-day technology is approaching the practical limits of simple-minded miniaturization and densification, and therefore a more elegant and fundamental approach is necessary. A related factor is human inertia, wherein designers who employ electronic devices in creating circuits can manage complexity by conceptualizing device behavior in terms of sets of monotonic curves. They are not trained to manipulate the complicated hypersurfaces which mathematically describe the characteristics of functional devices.

As the demand for sophisticated functionality from a small number of densely packed devices grows and the limitations of traditional device and circuit approaches become more apparent, however, there will necessarily be a cultural shift in the area of circuit design. The complexities of functional devices will be recognized as a source of opportunities and challenges.

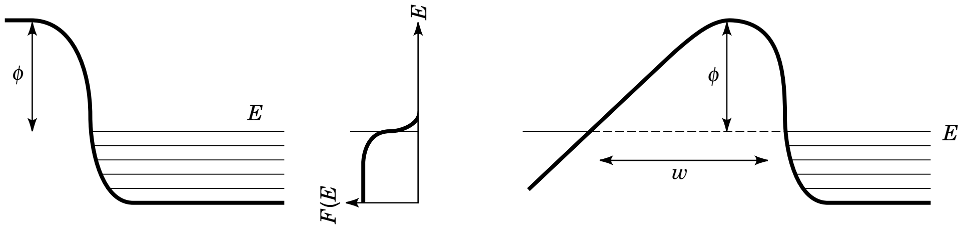
TUNNELING: EARLY DEVELOPMENTS

Through the development of the theory discussed following, Fowler and Nordheim (2) are generally credited with one of the early triumphs of the quantum theory: explaining the ejection of electrons from a cold metal in vacuo subjected to a high electric field. They invoked the *new wave mechanics of Schrodinger* to show how electrons could tunnel through a sufficiently thin energetic barrier and escape into the vacuum.

Although Fowler and Nordheim were indeed the first to apply the Schrodinger mechanics specifically to the metal-vacuum system, it was Oppenheimer (3), in a work concerning the ionization of hydrogen atoms via the tunneling process, who stated

> . . . [the] pulling of electrons out of metal by [large] fields . . . is probably to be accounted for in this way.

The history of the study of electron emission from metals is the background against which the advances attributed to Fowler and Nordheim may be understood. Schottky (4) studied the escape of electrons from a conductor via the process of thermionic emission, a purely classical process wherein a fraction of the electrons in the metal have sufficient energy to



(**a**) Thermionic emission (**b**) Field emission

Figure 3. Mechanisms of thermionic and field emission. In the case of (a) thermionic emission, a fraction of the electrons have sufficient energy to escape classically. In the case of (b) field emission, electrons may not have sufficient energy to escape classically, but may tunnel out quantum mechanically for small enough potential height and barrier width.

overcome the metal-vacuum barrier (Fig. 3). At higher temperatures, the average energy of electrons and the breadth of their statistical energetic distribution is increased, leading to a strong temperature dependence of the resulting current. When a weak field is applied to the metal, the Schottky formula describes the temperature T and electric field F dependence of the thermionic emission current:

$$J(F,T) = A_R T^2 \exp\left[-\frac{\phi - (e^3F)^{1/2}}{k_B T}\right] \qquad (2)$$

where A_R is a material-dependent constant, ϕ is the metal-vacuum barrier height in the absence of an applied field, and $(e^3F)^{1/2}$ accounts for the lowering of the barrier height brought about by the application of the field. In the case where no field is applied, Eq. (2) reduces to the Richardson–Laue–Dushman equation, often called the Richardson equation, from which the Richardson constant A_R derives its name.

Thermionic emission theory did not adequately explain the behavior of strong currents which could be obtained at low temperatures if very high electric fields were applied. A number of experiments had shown that the current was independent of temperature over a broad temperature range. This led to attempts to distinguish between electrons *pulled out by fields* and those *of a thermionic character* in a way which Fowler and Nordheim thought artificial. A new empirical relationship was proposed (5), which was correct under high fields and low temperatures (*field emission*) and also at high temperatures (*thermionic emission*) but untested in the transitional region between these mechanisms and not motivated by fundamental considerations.

Fowler and Nordheim posited that the escape of electrons from a cold conductor under application of a sufficiently high electric field could be explained by a quantum mechanical description: electrons could tunnel through the field-deformed potential energy barrier to generate a current of field-emitted electrons [Fig. 1 (b)]. They invoked de Broglie's description wherein electrons could be viewed as having wavelike properties, with a characteristic wavelength λ given by

$$\lambda = \frac{h}{(2mE)^{1/2}} \qquad (3)$$

where h is Planck's constant, m is the electron mass, and E is the electron energy. The wavelike properties of the electron allow it to pass through an energetic barrier at an appreciable rate if the barrier is lowered below the level E_F in a distance comparable with the electron wavelength λ.

Fowler and Nordheim treated the matter of field-emission tunneling by solving the Schrodinger wave equation on either side of the barrier with appropriate boundary conditions. They obtained an expression for the quantum mechanical transmission of electrons through the barrier as a function of electron energy and linked this with the rate of electrons impinging on the barrier as a function of energy. They obtained an electron field-emission tunneling current density J of

$$J(T,F) = AF^2 \exp\left(-\frac{B\phi^{3/2}}{F}\right) \qquad (4)$$

where F is the electric field strength, $\phi$is the conductor work function, and A and B are weaker functions of F and ϕ. Thus a plot of $\ln(J/F^2)$ versus $1/F$, the Fowler-Nordheim plot, is predicted to be a straight line, a fact which is borne out experimentally. The temperature-independence of the measured field-emission current is also predicted and explained by the Fowler–Nordheim theory.

THE TUNNEL, OR ESAKI, DIODE

In 1958, Leo Esaki (6) observed a negative differential resistance in the forward current-voltage characteristic of a Zener diode (Fig. 4). The devices under study were germanium *p-n* junctions with very heavy dopant concentrations on the order of 10^{19} cm^{-3}. From capacitance measurements Esaki found that the junction width was approximately 150 Å. He accounted for the observed current-voltage characteristic in terms of tunneling of electrons in the conduction band of the heavily doped *n*-side through the narrow junction (whose width was comparable to the electron wavelength, making the tunneling probability appreciable) into the valence band of the heavily doped *p*-type contact, and through the analogous process for holes.

To gain an understanding of the operation of the Esaki diode, one must consider the transmission coefficient for electrons and holes through the barrier as a function of their allowed energies and also the availability of states from which and into which to tunnel. Fermi's golden rule gives the transition rate from initial state i to final state m as

$$W_{i\to m} = \frac{2\pi}{\hbar} M_{im}\rho(E_m) \qquad (5)$$

where M_{im} is called the matrix element for the transition and in this case is directly related to the tunneling transmission coefficient. The density of states $\rho(E_m)$ describes the availability of states into which the carriers may tunnel. To obtain the net tunneling current in a particular direction, the difference between $W_{c\to v}$ and $W_{v\to c}$will be considered.

This concept and its consequences are illustrated schematically in Fig. 5. At zero bias [Fig. 5(a)], *n*-side electrons above

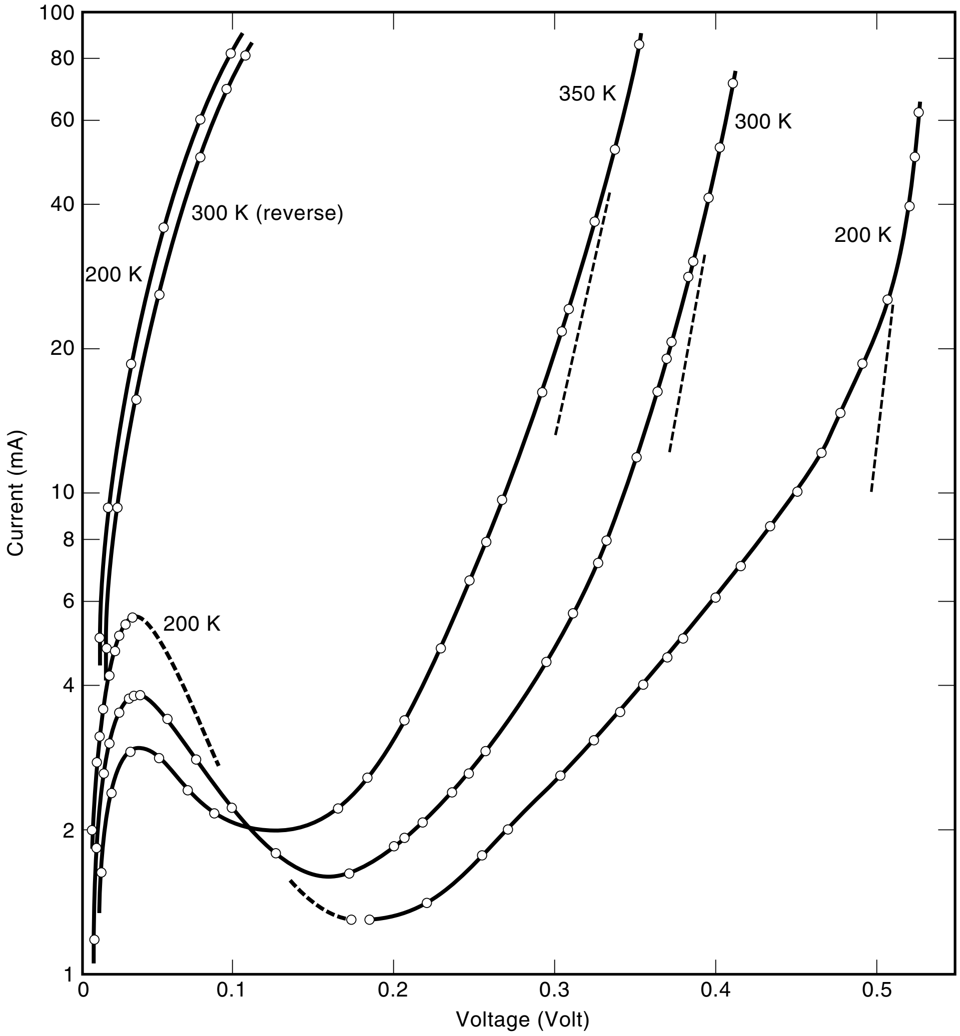


Figure 4. *I*–*V* characteristic of the tunnel diode of Leo Esaki's seminal work. The negative differential resistance characteristic—explained with the aid of Figure 5—provides evidence of the importance of the tunneling mechanism and forms the basis for device applications of tunneling.

the Fermi level can tunnel into vacant states on the *p*-side. However, since they do so at an equal rate in the opposite direction (and the same argument applies for holes), there is no net current. As a small forward bias is applied [Fig. 5(b)], electrons on the *n*-side become energetically aligned with unoccupied states on the *p*-side. As the bias is increased further [Fig. 5(c)], more of the electrons lie opposite the forbidden band on the *p*-side, so that tunneling (in this simple model) is not possible. At even higher biases [Fig. 5(c)], classical drift-diffusion processes dominate the *I*–*V* characteristic, and the

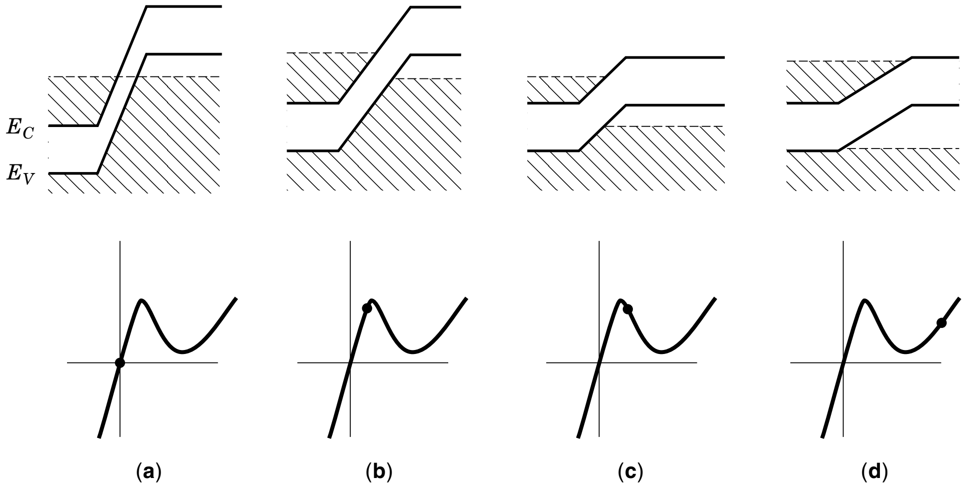


Figure 5. Schematic portrayal of the mechanism of Esaki diode negative resistance. Quantum mechanical effects dominate the current at low forward bias (b): electrons and holes tunnel through the forbidden zone into the opposite band. As the bias is increased (c), fewer states are available into which carriers may tunnel, and the current decreases. The classical diode current takes over at higher biases (d).

diodes begin to obey the usual Shockley equation. It is essential that both sides of the junction be degenerately doped (i.e., that the Fermi level lie within the conduction band in the n-type contact and within the valence band in the *p*-type contact).

Esaki was a cowinner of the 1973 Nobel Prize for Physics for his experimental discovery regarding tunneling phenomena in semiconductors.

Excess Current

In many tunnel diode applications a large ratio of peak current to valley current is required. For this reason, the *excess current,* the value of the current in the valley region of the I–V characteristic, where tunneling current is expected to drop to zero and before standard thermionic emission current takes over, is of practical significance. A number of hypotheses were put forth to explain this observation. Mechanisms whereby tunneling carriers could lose energy through photon, phonon, plasmon, or Auger processes were suggested but were not sufficiently important to explain the observed excess tunneling current. Starting from the hypothesis put forth by Esaki that electrons could not tunnel completely through the energy gap but only part of the way, making use of states in the energy gap, Chynoweth et al. (7) developed and experimentally corroborated a model for the excess current.

Desired Properties

One of the most prominent applications of the Esaki diode is as a high-speed component in oscillator circuits and switches. The preservation of the diode's negative resistance at high frequencies makes it a candidate for such applications. The switching speed is determined by the current available for charging the junction capacitance. To achieve high-speed performance, low capacitance is desired, and sufficient current must be supplied by the diode to charge the junction capacitance. Therefore one simple figure of merit is the ratio of the peak current to the junction capacitance. Another important figure of merit is the ratio of the peak current to the valley current, known as the peak-to-valley ratio (P/V), which is related to the current gain obtainable. Maximizing the peak-to-valley ratio in Esaki diodes represents a compromise, primarily in the doping level. At lower (though still degenerate) dopings, the peak current is small because there is only a narrow energy range over which conduction-band electrons see unoccupied valence-band states (and analogously for holes in the valence band). At higher dopings, the density of band-gap states increases (as described above), and the valley current increases. The maximal P/V is found for some intermediate concentration. In either case, the requisite doping level is near the maximum level which can be activated in the semiconductor, typically around 10^{19} cm^{-3}.

Although a remarkable device and one which provided a satisfying early example of engineering in the quantum domain, the Esaki diode exhibits some intrinsic properties which limit its usefulness to certain regimes and application areas. Most importantly, the degenerate doping levels required to achieve a reasonable peak current give rise to a large shunt capacitance which limits high-speed performance and necessitates presenting the device with impedances properly matched to the capacitive reactance of the diode.

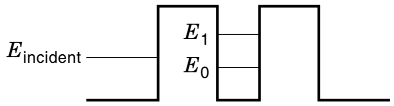


Figure 6. Resonant transmission in a double barrier system. The alignment of the incident particle energy relative to the energies of barrier-confined states determines the rate of transmission through the system.

THE RESONANT TUNNELING DIODE

These fundamental limitations on the performance of the Esaki diode, taken together with the promising prospect that it demonstrated for devices based on tunneling, motivated the development of a structure whose performance was not fundamentally linked to heavy doping. This was first sought and realized in the form of the resonant (intraband) tunneling diode.

The history of resonant tunneling precedes the perception of its need in device implementations. The concept originally elaborated by Bohm (8) is illustrated in Fig. 6. The system of double barriers is characterized by a set of quantized energy states. If an incident particle impinges with energy equal to one of these bound-state energies, it is resonantly transmitted. If it differs substantially, it is resonantly reflected.

The first suggestions for resonant-tunneling devices were made by Davis and Hosack (9) and Ioganson (10). Esaki and Tsu (11) proposed a superlattice implementation of the same basic concept and anticipated explicitly the negative differential resistance resulting from the strong energy dispersive effects which may arise in such a structure if the critical dimensions are on the order of the electron wavelength. In 1973 they extended their theoretical considerations (12) to the case of a multiple-barrier superlattice as opposed to a theoretically infinite one. In 1974, a superlattice was implemented using molecular beam epitaxy (13), with 45 Å GaAs wells clad by 40 Å AlAs barriers. The negative differential conductance characteristic of Fig. 7 was reported.

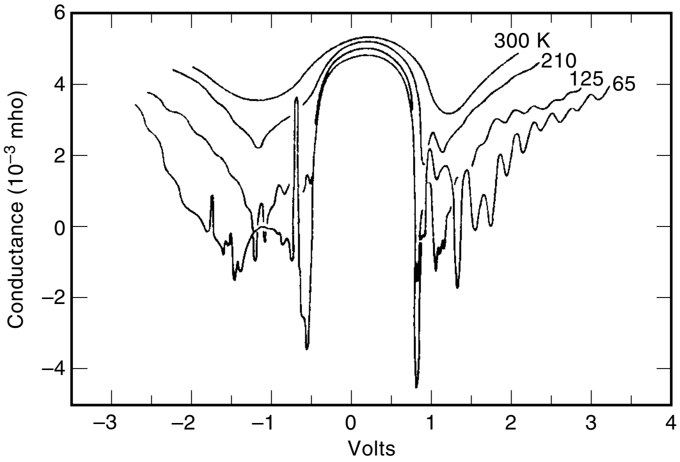


Figure 7. Differential conductance of the first resonant intraband tunneling structure of Esaki and Chang, 1974 (14).

The mechanism of the intraband resonant tunneling diode may be illustrated (Fig. 8) by considering one pair of barriers. The same principles apply in determining the conductance features of Fig. 7, the more complicated structure giving rise to the more intricate observed features. At zero bias, the energy of conduction-band electrons in the emitter is less than that of quantum-confined electrons between the barriers [Fig. 8(a)] and they are not resonantly transmitted. As the bias is increased, these energies become aligned [Fig. 8(b)]. The states at this same energy in the collector are almost completely unoccupied, so that resonant transmission is achieved, and the conductance is increased. As bias is further increased, however, the energies of the electrons in the injector and inside the well are out of resonance and the conductance is reduced [Fig. 8(c)].

It was not until the observation of fast intrinsic response times by Sollner et al. (14) and soon thereafter of room temperature negative differential resistance (15,16) that the field of superlattices and quantum wells began to grow rapidly. Sollner et al. obtained a P/V ratio of 6 : 1 at low temperatures, and although the negative differential resistance characteristic was not manifested at room temperature, the effect of resonant tunneling was nevertheless apparent in the room temperature differential conductance characteristic. In addition, Sollner et al. reported one of the first experimental manifestations of the anticipated high-speed response of the room temperature device (RTD). The authors concluded that the charge transport mechanisms are characterized by a time of 6×10^{-14} s. The room temperature NDR of Shewchuck et al. (17) was one of many incremental steps of progress in the direction of acceptably high room temperature peak-to-valley ratios which came with gradual technological improvements in molecular beam epitaxy. In particular, very thin (~1 nm) high barrier layers were eventually obtained with precise thickness control and uniformity. Double-barrier RTDs operating at room temperature have been achieved with peak-to-valley ratios as high as 50 : 1 at 300 K (17). With the benefit of such incremental technological progress, the experimentally observed fundamental oscillation frequency has improved approximately linearly with time.

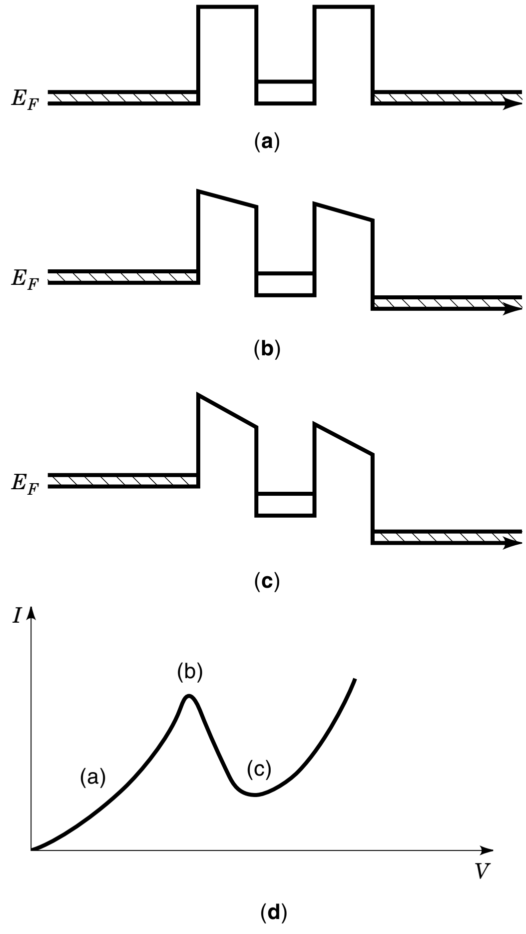


Figure 8. Mechanism of negative differential conductance in resonant tunneling diodes. At zero bias (a), the electron energy is less than that of the confined barrier states. Under increased bias (b), the incident and confined energies become aligned, and the states at this same energy in the collector are almost completely unoccupied, so that resonant transmission is achieved. As bias is further increased (c), electrons in the injector and the double barrier fall out of resonance.

RESONANT INTERBAND TUNNELING

It is remarkable that, with the benefit of the high quality atomic-layer engineering made possible by molecular beam epitaxy, resonant tunneling diodes achieved room temperature peak-to-valley ratios no better than those of the original Esaki diodes of thirty years earlier which used much less sophisticated material engineering techniques. Had the abrupt interfaces and high doping of modern epitaxial crystal growth techniques been possible at that time, the Esaki diode would likely have provided still more competitive performance.

On the other hand, the RTD held the clear advantage of a much lower capacitance and more manageable technological challenges. In the light of these observations, in 1989 Sweeny and Xu proposed (18) a structure operating on both interband and resonant tunneling principles with the objective of preserving the attractive features of each one. Their resonant interband tunnel diode concept was an otherwise ordinary *p-n* diode with quantum wells in the conduction and valence bands. Thus, although it was a bipolar interband tunneling device like Esaki's, it incorporated the resonance features of the RTD through the use of coupled quantum wells. The device did not rely on heavy doping to ensure tunneling, but instead took advantage of quantum wells (grown or induced) and exploited the resonance tunneling phenomenon. The operation of one such device is illustrated in Fig. 9. Regions I and IV of Fig. 9 have opposite doping and need not be degenerate. The well regions II and III have a lower band gap and are doped the same as their higher band-gap outer neighbors. As in previous tunneling devices, the barrier must be sufficiently thin that there is significant interpenetration of the carrier wave functions in II and III. Using band-gap engineering, high carrier concentrations may be achieved in the wells without requiring degenerate doping at any point in the structure. When a bias is applied so that the conduction-band density of states in the *n*-type quantum well (II) is energetically aligned with the valence-band density of states in the *p*-type quantum well (III), resonant transmission of carriers through the barrier occurs. As the bias is further increased, no states are available for tunneling transmission until the

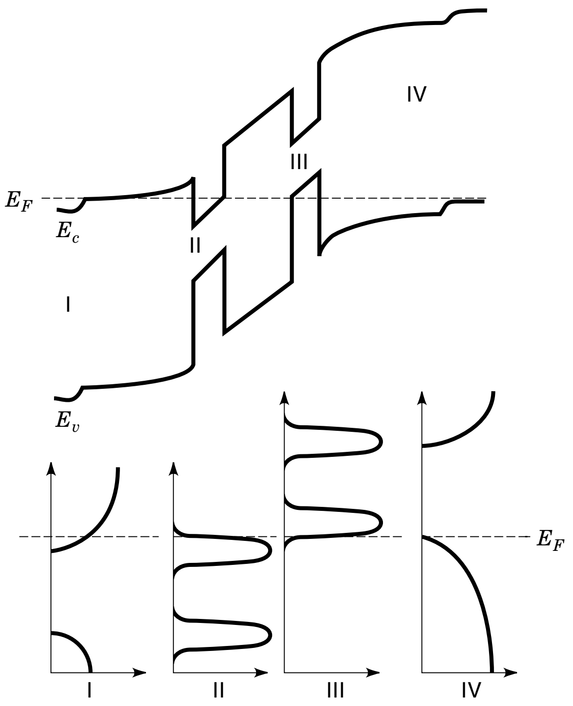


Figure 9. Resonant Interband Tunneling device operation. Resonant transmission occurs between the conduction to the valence bands, rather than within a single band as in earlier intraband tunneling devices.

conduction band of (II) and (III) are aligned. These possibilities and the resulting *I–V* characteristic, are illustrated in Fig. 10. The type-II heterostructural version in their proposal was later suggested independently and demonstrated the same year (19). This very first implementation of the RIT achieved a room temperature *P/V* ratio of 20. Within four years, a room temperature peak-to-valley current ratio of greater than 100 had been demonstrated by Xu et al. (20).

The two alternate implementations of the resonant interband tunneling concept work similarly (18). The polytype heterostructural implementation exploits the fact that the conduction band of one material is below the valence band of the other in a type-II heterostructure. Resonant intraband tunneling could be realized in such a device [Fig. 11(a)] using very low doping, enabling ultrahigh-speed performance not limited by significant contact capacitances. Another realization, a modulation-doped homostructure [Fig. 11(b)], also allows resonant intraband tunneling with a minimum of material doping.

SINGLE-ELECTRON TUNNELING: EFFECTS AND DEVICES

The preceding discussion centered around collective transport of many electrons through a system. Tunneling of individual electrons—known as single-electron tunneling—is difficult to observe and control, since thermal fluctuations in electron energy (of order kT) are typically larger than the Coulomb energy change of the system.

Substantial progress has nevertheless been made in this area (21–24). The possibility of observing single-electron tunneling in very small systems was noted around the same time that Esaki observed the effects of macroscopic tunneling electron tunneling in semiconductors. To observe single-electron tunneling, it is not sufficient simply for the system to be small and to have a large effective capacitance, in turn giving rise to a large charging energy (much greater than kT). It must also be well isolated (electromagnetically decoupled) from the environment such that the electron is essentially localized within the system. This localization condition may be expressed in the requirement that the tunneling resistance (impedance) of the system be much greater than the quantum resistance.

Only recently—with the aid of technological advances and further important progress (25)—has broad interest been generated in this problem, and a wide range of investigations begun into single-electron tunneling effects. If the conditions

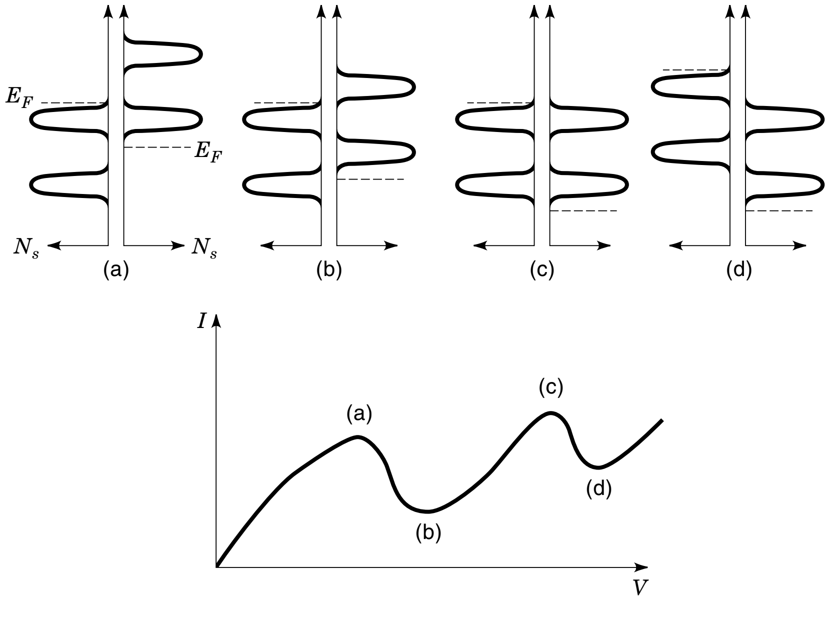


Figure 10. Mechanism of *I–V* characteristic of resonant intraband tunneling diode. The alignment of quantum confined energetic states in the conduction and valence bands determines the rate of resonant transmission.

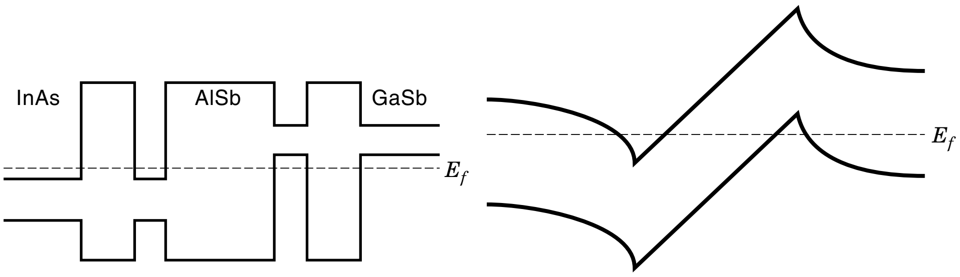


Figure 11. (a) Polytype heterostructure implementation of the resonant interband tunneling diode. (b) Modulation-doped implementation of the resonant interband tunneling diode.

described above are met, the effect may be observed in a number of ways. Because of the discrete nature of electrons, there exists a staircase relation between system charge and voltage, so that conductance peaks may be observed at specified voltages. A Coulomb staircase I–V characteristic (25)—a dramatic manifestation of the effect of single-electron charging—arises in suitable structures in which one tunnel barrier is more strongly transmitting than the other (26). One application area already demonstrated is in the use of controlled Coulomb blockade effects in realizing accurate current standards: by cycling tunneling barrier heights, individual electrons can be made to pass through the confined system at the applied frequency—resulting in an “electronic turnstile” (27)—and producing a current $I = ef$ (where e is the electronic charge and f the frequency of modulation).

The introduction of further tunneling junctions and more complex connections and coupling provide a rich variety of externally observable single-electron tunneling phenomena. Foreseeable applications in conventional electronics include memory cells, D/A converters, and sensitive analog transistors.

TUNNELING IN OPTOELECTRONICS

Laser technology has also benefited from innovative solutions to problems posed by new applications. Conventional semiconductor lasers are bipolar devices which rely on band-to-band transitions between the conduction and valence bands. The energy associated with these interband transitions and consequently the energy of the emitted photons is largely determined by the properties of the semiconductor material. By introducing quantum wells in which spatially confined electrons and holes have ground-state energies above the semiconductor bandgap, it is possible to tailor somewhat the energy of photon emission.

For a number of applications, mid-wave or long-wavelength lasers with photon energies ranging from 2 to 12 μm are desired. One solution which reduces the dependence on material choice takes advantage of transitions in quantum wells within a particular band, typically the conduction band. In these intraband devices, the photon emission energy may be selected by careful tailoring of well and barrier widths.

One fundamental requirement in lasers is that of population inversion. If photons are emitted during the stimulated transition from state 2 to state 1, then the population of state 2 must exceed that of state 1. Two dominant approaches have been adopted in achieving population inversion in intraband lasers, both of them based on tunneling.

The group of Capasso (28) realized the first intrasubband laser. In this device, population inversion within the conduction band was achieved via very careful design of the active region, which consists of sets of wells and barriers for injection, relaxation, and removal of carriers. Electrons are injected by resonant tunneling into one of the higher states of the active region quantum wells. By simultaneously making the lower state depopulation mechanisms resonant with other phonon and tunneling processes, the lifetime of the lower state is made less than that of the upper state, and population inversion may be achieved. Using this tunneling-based mechanism, room temperature quantum cascade lasers have been achieved (29).

Another approach to achieving population inversion was proposed by Yang and Xu (30). As illustrated in Fig. 12, intraband tunneling or simply intraband transport may be used to inject carriers into the upper state (from I-II in the figure), and interband tunneling to remove carriers from the lower state (from II-III) to invert the populations in the first two states of the conduction band. The structure is designed to prevent tunneling out of the upper state. Low-temperature operation of a device incorporating this concept was demonstrated by Yang et al. in 1997 (31).

TUNNEL DEVICE MODELING

Since the seminal work of Fowler and Nordheim (2), a variety of advances have been made in the accuracy of models of tunnel device operation (32). An element common to these developments is the use of the effective mass approximation. In this approach, widely employed throughout condensed matter

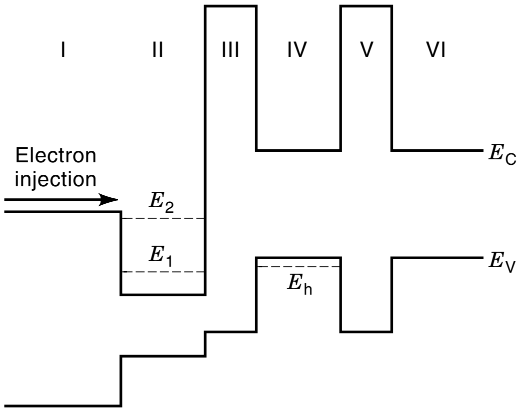


Figure 12. Illustration of resonant interband tunneling assisted population inversion. Carriers are injected into the upper state (from I-II) and are removed from the lower state (from II-III) via interband tunneling. The population in the first two states of the conduction band is thereby inverted.

physics, an effective mass term (not generally equal to the true physical mass of the carrier) is used to account for the effects of the atomic potentials which the charge carrier encounters during its trajectory through a solid medium. The simplifications involved transform the modeling of interactions in a solid from what would be a daunting task, accounting individually for a large, complex set of potentials, into potentially tractable problems. The effective mass is used in an approximate Schrodinger equation, and the validity of the results vary according to structure and the region within a given structure.

The Bohm approximation (or the golden rule of Fermi) is usually invoked to quantify tunneling currents. Tunneling may be viewed as the scattering of an electron in an electric field in which the scattering potential is usually invariant in the transverse direction, so that the transverse momentum vector is conserved. (In more complex devices, one may take advantage of the transverse direction in further enhancing functionality.)

Although a number of simplifications have already been invoked to this point, further approximations are typically employed in modeling interband tunnel devices. In this case, the coupling between conduction and valence bands provides the central mechanism for device operation, and a singleband effective mass approximation does not yield physically correct results.

In such devices, interband coupling is strongest in the tunneling region. To a first-order approximation, the current-voltage relationship is obtained from a coupled-band effective mass equation. If spin is taken into consideration, there are two conduction-band contributions and six valence-band contributions. Within this eight-band framework, the $\vec{k} \cdot \vec{p}$ of Kane (33) is the most commonly employed and may also be the most exact. Altarelli (34) provides a review of the $\vec{k} \cdot \vec{p}$ treatment.

To gain physical insight into the problem without onerous computation, a two-band model is often used. By symmetry considerations of the Bloch functions, the conduction band and light-hole bands are most strongly coupled and must be retained. The other bands of the eight-band Luttinger–Kohn Hamiltonian (29) may be removed if the effective masses of the bands remaining are adjusted to ensure that their dispersion relationships agree reasonably with the known band structure over the energy of interest [for a review, see Datta (36)]. Even this simplified two-band model provides coupling between the differential equations associated with different bands.

OVERVIEW, ASSESSMENT, AND OUTLOOK

Digital functional devices based on carrier tunneling take advantage of nonmonotonic I–V characteristics and fast intrinsic response times. Applications include high-speed analog-to-digital converters, parity bit generators, and multiple-valued logic elements. Three terminal devices, such as the resonant tunneling bipolar transistor proposed by Capasso and Kiehl in 1985 (37) have also been developed. This particular device, which uses a quantum well in the p-type base layer, exhibits a series of peaks in its collector current as a function of base-emitter voltage. For this device, applications include multiple-valued logic, parity generators, analog-to-digital converters, and multiple state memory, all implemented more naturally than using a collection of two-terminal devices. Three-terminal unipolar devices based on tunneling include the resonant hot electron transistor (RHET) of Yokoyama et al. (38) and the quantum wire transistor proposed in 1985 by Luryi and Capasso (39).

High-speed analog devices typically exploit the high-frequency negative differential conductance obtained from tunneling devices. Two-terminal oscillators are perhaps the simplest device examples, which take advantage of the fact that the negative differential conductance persists on time scales as short as the lifetime of the electron in the resonant state between the barriers of a resonant tunneling diode. RTDs may also be used as efficient mixers by exploiting the rapid variation in dynamic conductance with voltage near the negative differential resistance portion of the I–V characteristic.

RTDs find applications in switching when they are biased with a source resistance larger than the magnitude of their negative differential resistance. A stable bias point is no longer achievable in the NDR region, and switching occurs between the stable points outside of this region. Switching times as short as 2 ps have been measured (40,41).

The preceding list of applications covers those which could be thought of as niche roles for tunneling devices. They fulfill a specific role often very effectively but typically in isolation from general circuit applications. It has been argued that tunneling devices have a much more important role to play in the future.

Exponential improvements in circuit speed have been enabled by an exponential downward trend in minimum device geometries and switching power. This downscaling cannot continue indefinitely. Before fundamental physical limits are reached, ICs based on transistors will be rendered impossible or exorbitant by a combination of problems related to device technology, interconnection, noise, and reliability. A saturation in circuit density improvements is likely to imply a saturation in the historically downward trend in cost per bit or function.

Three-terminal devices based on tunneling would provide a means to continue this downward scaling and in fact to exploit it to the fullest. However, as argued previously, tunneling devices technologies will not gain acceptance if they cannot penetrate the culture of circuit design—if they do not become accessible to their users. This necessitates coordination between device creators and device users in matching physics with function, as per Morton's vision of 1965.

Another possible trend is a further extrapolation of Morton's functional device concept and of the desirability of device miniaturization. The inspiration is taken from biological systems and biochemical reactions and interactions, which possess the desired characteristics of being based on tunneling and, therefore are very fast; of being intrinsically multifunctional; and of possessing, in some instances, the potential to implement a *local learning function.* At least two approaches have been witnessed on this front. First, simple tunneling processes have been incorporated into otherwise standard bipolar transistors (42) and MOS transistors (43). In the bipolar transistor, the large ratio of electron tunneling transmission to hole transmission (due to the large effective mass disparity in compound semiconductors) yields improved emitted injection efficiency of homojunction BJTs with a more easily fabricated layer structure than traditional heterojunction BJTs. In the MOS case, tunneling into a floating gate structure enables functionality analogous to learning. A second approach, more revolutionary and therefore less mature in approach, involves developing devices which perform some of the basic electronic functions using human-engineered molecules, a field known as (bio)molecular electronics. Both approaches use tunneling, as a consequence of miniaturization and as a means for expanding and exploring new functionalities.

BIBLIOGRAPHY

1. J. A. Morton, From physics to function, *IEEE Spectrum,* **2**: 62–66, 1965.
2. R. H. Fowler and L. Nordheim, Electron emission in intense electric fields, *Proc. R. Soc. Lond.,* **119**: 173–181, 1928.
3. J. R. Oppenheimer, Three notes on the quantum theory of aperiodic effects, *Phys. Rev.,* **13**: 66–81, 1928.
4. W. Schottky, Uber den einfluss von structurwirkungen, besonders der Thomsonchen bildkrfat, auf die electronenemission der metalle, *Physikalische Zeitschrift,* **15**: 872–878, 1914.
5. R. A. Millikan and C. C. Lauritsen, Relations of field currents to thermionic currents, *Proc. Natl. Acad. Sci.,* **14**: 45–49, 1928.
6. L. Esaki, New phenomenon in narrow germanium p-n junctions, *Phys. Rev.,* **109**: 603–604, 1958.
7. A. G. Chynoweth, W. L. Feldmann, and R. A. Logan, Excess tunnel current in silicon Esaki junctions, *Phys. Rev.,* **121**: 684–694, 1961.
8. D. Bohn, *Quantum Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1951.
9. R. H. Davis and H. H. Hosack, Double barrier in thin-film triodes, *J. Appl. Phys.,* **34**: 864–866, 1963.
10. L. V. Iogansen, The possibility of resonance transmission of electrons in crystals through a system of barriers, *Soviet Phys. JETP,* **18**: 146–150, 1964.
11. L. Esaki and R. Tsu, Superlattice and negative differential conductivity in semiconductors, *IBM J. Res. Develop.,* **14** (1): 61–65, 1970.
12. L. Esaki and R. Tsu, Tunneling in a finite superlattice, *Appl. Phys. Lett.,* **22**: 562–564, 1973.
13. L. Esaki and L. L. Chang, New transport phenomenon in a semiconductor "superlattice," *Phys. Rev. Lett.,* **33**: 495–498, 1974.
14. T. C. L. G. Sollner et al., Resonant tunneling through quantum wells at frequencies up to 2.5 Thz. *Appl. Phys. Lett.,* **43**: 588–590, 1983.
15. T. J. Shewchuk et al., Resonant tunneling oscillations in a GaAs-$Al_xGa_{1-x}As$ heterostructure at room temperature, *Appl. Phys. Lett.,* **46**: 508–510, 1985.
16. M. Tsuchiya, H. Sakaki, and J. Yoshino, Room temperature observation of differential negative resistance in AlAs/GaAs/AlAs resonant tunneling diode, *Jpn. J. Appl. Phys.,* **24**: L466–L468, 1985.
17. J. Smet, T. P. E. Broekaert, and C. G. Fonstand, Peak-to-valley current ratios as high as 50 : 1 at room temperature in pseudomorphic $In_{0.53}Ga_{0.47}As$/AlAs/InAs resonant tunneling diodes, *J. Appl. Phys.,* **71**: 2475–2477, 1992.
18. M. Sweeny and J. Xu, Resonant interband tunnel diodes, *Appl. Phys. Lett.,* **54**: 546–548, 1989.
19. J. R. Soderstrom, D. H. Chow, and T. C. McGill, New negative differential resistance device based on resonant interband tunneling, *Appl. Phys. Lett.,* **55**: 1094–1096, 1989.
20. D. J. Day et al., Experimental demonstration of resonant interband tunnel diode with room temperature peak-to-valley current ratio over 100, *J. Appl. Phys.,* **73**: 1542–1544, 1993.
21. C. J. Gorter, *Physica,* **17**: 777, 1951.
22. C. A. Neugebauer and M. B. Webb, *J. Appl. Phys.,* **33**: 74, 1962.
23. I. Giaever and H. R. Zeller, *Phys. Rev. Lett.,* **20**: 1504, 1968.
24. J. Lambe and R. C. Jaklevic, *Phys. Rev. Lett.,* **22**: 1371, 1969.
25. K. K. Likharev, *IBM J. Res. Dev.,* **32**: 144, 1988.
26. U. Meirav and E. B. Foxman, *Semiconductor Sci. Technol.,* **11**: 255–284, 1996.
27. Y. Nagamune et al., *Appl. Phys. Lett.,* **64**: 2379, 1994.
28. J. Faist et al., Quantum cascade laser: Temperature dependence of the performance characteristics and high T_0 operation, *Appl. Phys. Lett.,* **65**: 2901–2903, 1994.
29. J. Faist et al., High power mid-infrared ($\lambda \sim 5$ μm) quantum cascade lasers operating above room temperature, *Appl. Phys. Lett.,* **68**: 3680–3682, 1996.
30. R. Q. Yang and J. M. Xu, Population inversion through resonant interband tunneling, *Appl. Phys. Lett.,* **59**: 181–183, 1991.
31. R. Q. Yang et al., High power mid-infrared interband cascade lasers band on type-II quantum wells, *Appl. Phys. Lett.,* **71**: 2409–2411, 1997.
32. E. E. Mendez, J. Nocera, and W. I. Wang, Conservation of momentum and its consequences in interband resonant tunneling, *Phys. Rev. B,* **45**: 3910–3913, 1992.
33. E. O. Kane, Band structure of indium antimonide, *J. Phys. Chem. Solids,* **1**: 249–261, 1957.
34. M. Altarelli, in G. Allen et al., *Heterojunction and Semiconductor Superlattices,* Berlin: Springer-Verlag, 1986.
35. J. M. Luttinger and W. Kohn, Motion of electrons and holes in perturbed periodic fields, *Phys. Rev.* **97**: 869–883, 1955.
36. S. Datta, *Quantum Phenomena,* Reading, MA: Addison-Wesley, 1989.
37. F. Capasso and R. A. Kiehl, Resonant tunnelling transistor with quantum well base and high-energy injection: A new negative differential resistance device. *J. Appl. Phys.,* **58**: 1366, 1985.
38. N. Yokoyama et al., A new functional resonant tunnelling hot electron transistor (RHET). *Jpn. J. Appl. Phys.,* **24**: L853, 1985.
39. S. Luryi and F. Capasso, Resonant tunnelling of two-dimensional electrons through a quantum wire: A negative transconductance device, *Appl. Phys. Lett.,* **47**: 1347, 1985.
40. J. F. Whitaker et al., Picosecond switching time measurement of a resonant tunneling diode, *Appl. Phys. Lett.,* **53**: 385–387, 1988.
41. S. K. Diamond et al., Resonant tunneling diodes for switching applications, *Appl. Phys. Lett.,* **54**: 153–155, 1989.
42. J. Xu and M. Shur, A tunneling emitter bipolar transistor, *IEEE Electron Device Lett.,* **EDL-7**: 416–418, 1986.
43. C. Diorio et al., A single-transistor silicon synapse, *IEEE Trans. Electron Devices,* **43**: 1972–1980, 1996.

Reading List

C. B. Duke, *Tunneling in Solids,* New York: Academic, 1969.

H. C. Okean, Tunnel diodes, in *Semiconductors and Semimetals,* Vol. 7, Part B, ed. R. K. Willardson and A. C. Beer, New York: Academic, 1966.

D. K. Roy, *Tunnelling and Negative Resistance Phenomena in Semiconductors,* New York, NY: Pergamon, 1977.

EDWARD H. SARGENT
J. M. XU
University of Toronto

TUNNEL DIELECTRICS, MANUFACTURING. See Gate and tunnel dielectrics, manufacturing aspects.

TUNNEL DIODE OSCILLATORS. See Harmonic oscillators, circuits.

TUNNEL DIODES. See Tunnel devices.

TUNNELING DIODES. See Zener effect.

TURBINES FOR HYDROELECTRIC POWER. See Hydroelectric power stations.

TURBINES, HYDRAULIC. See Hydraulic turbines.

TURBINES, STEAM. See Steam turbines.

TURBINES, WIND. See Wind turbines.