J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*

QUANTUM DEVICES

Most existing semiconductor devices, such as bipolar and field effect transistors, can be treated as classical systems of carriers near equilibrium, where quantum effects manifest themselves as minor corrections that may limit device reliability. During the 1980s and 1990s considerable research interest and effort were expended on semiconductor structures that could exploit quantum effects to perform circuit functions. Although, to date none of these structures has evolved beyond the laboratory, continuing interest has been maintained because of two major factors.

First, it is generally recognized that microelectronics will soon cease to improve at the device level. Evolution toward ever-faster devices is driven by the minimal size of lithographic features. Smaller devices yield faster transit times at lower operating voltages and currents, leading to higher maximum frequencies at lower power per device (1). At the time of writing, the Semiconductor Industry Association roadmap predicts continuous device improvement until 2015, at which point the minimum lithographic size would fall below 1000 Å and the DRAM size would reach 16 or 64 Gb. The roadmap and its implications are discussed by a number of authors in Ref. 1. In addition to technological hurdles, this evolution faces rapidly escalating fabrication costs. However, even if both the technological and economic constraints are overcome, new operational concepts will be needed in the foreseeable future, because the minimum channel length L of a standard scaled silicon metal-oxide semiconductor field effect transistor (Si MOSFET) bottoms out slightly below $L \approx 0.05$ μm. More complicated device layouts, such as double-gate devices, or cryogenic operation may yield another device generation or two, but probably not more.

The second factor driving research into quantum devices springs from semiconductor bandgap engineering. Modern epitaxy of III–V semiconductors, silicon, and silicon-based alloys provides exceptional control over layer thickness, doping, and composition. It is now possible to specify regions of carrier localization and transport, tailor electric fields and potential barriers, and make use of size and charge quantization, the latter pushing device research toward single electronics, in which the transfer of one electron suffices to control a device.

Higher speed is often cited as an advantage of quantum devices, but speed is rarely the prime consideration. Even though quantum mechanical tunneling is an intrinsically fast process, this does not typically translate into device performance improvement because of other delays, such as RC time constants of bias electrodes. Frequently, a more significant advantage is higher functionality, made possible either by strong, tunable current-voltage nonlinearities or by unusual electrode symmetries of quantum devices. Examples covered in this article include multistate memory and logic implementations using reduced device counts. Quantum tunneling also plays a key role in the recently developed quantum cascade lasers, while the use of quantum dots embedded in the active medium has opened up new horizons in low-threshold lasers.

After a simple introduction to quantization effects and quantum mechanical tunneling, this article covers a number of basic quantum device structures and operating principles. Various proposed applications—ranging from memories and logic circuits to novel lasers—are then presented. A brief overview of the prospects of quantum devices serves as the conclusion.

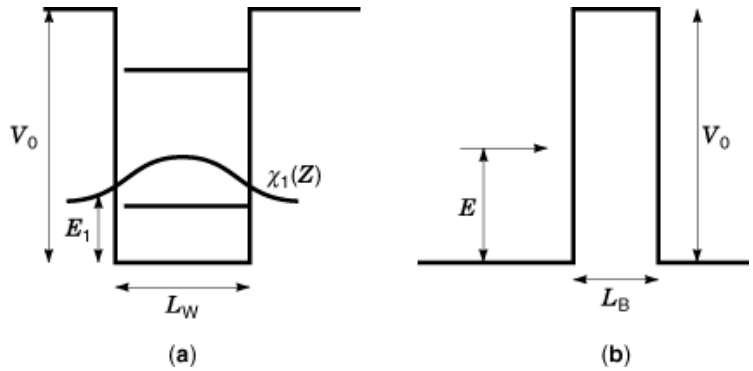

Fig. 1. (a) Finite quantum well potential diagram, showing the wavefunction $\chi_1(z)$ of the the lowest discrete level E_1. (b) A particle of energy E incident on a single barrier.

Quantum Mechanical Density Of States and Tunneling

Quantum devices are based on two consequences of the Schrödinger equation that have no classical analog. First, if a particle is confined by some potential V (r) on a scale comparable to its de Broglie wavelength, the particle's momentum $\mathbf{k}$ is quantized. The continuous energy spectrum E $(\mathbf{k}) = {}^2\mathbf{k}^2$ / $2m*$ of free motion ($m*$ is the effective mass) is broken up into subbands E_n $(\mathbf{k})$. Second, as long as the confining potential V $(\mathbf{r})$ is not infinite, the particle has a finite probability of penetrating the classically forbidden region. These effects are most easily illustrated in the case of one-dimensional (1-D) motion in a finite potential well of width L_W and barrier height V_a shown in Fig. 1(a). The 1-D Schrödinger equation for the wavefunction $\chi(z)$ can be written as follows:

$$\left[\frac{-\hbar^2 d^2}{2m^* dz^2} + V(z)\right]\chi(z) = E\chi(z) \qquad (1)$$

where is the reduced Planck's constant. Solving this equation in all three regions and imposing continuity conditions on $\chi(z)$ and $d\chi/dz$, one obtains energy levels E_n and the corresponding $\chi_\mathrm{n}(z)$. The normalized χ_n (z) give the probability of finding the particle at some coordinate z_0 by$P(z_0) = |\chi(_0 z_0|^2$. In an infinite potential well, the eigenfunctions $\chi_\mathrm{n}(z)$ must go to zero at $|z| = L_\mathrm{W}/2$ and the energy levels are given by

$$E_n = \frac{\hbar^2 \pi^2 n^2}{2m^* L_\mathrm{w}^2} \qquad (2)$$

where$n$ is an integer. In the finite potential well case of Fig. 1(a), the well contains a finite number of energy levels E_n. The corresponding wavefunctions $\chi_\mathrm{n}(z)$ penetrate into the potential barriers according to

$$\chi_n(z) \sim e^{-\kappa_n |z|} \qquad (3)$$

where $\kappa_\mathrm{n} = [2m*(V_0 - E_\mathrm{n})/^2]^{1/2}$ [the other mathematically possible solution, $\chi_n(z) \sim e^{K}{}_n{}^{|z|}$, diverges as $|z| \rightarrow \infty$ and can be excluded]. Although the barrier penetration is exponentially decreasing, Eq. (3) implies that a carrier in the state characterized by $\chi_\mathrm{n}(z)$ has a finite probability of being found in the barrier region $|z| > L_\mathrm{W}/2$.

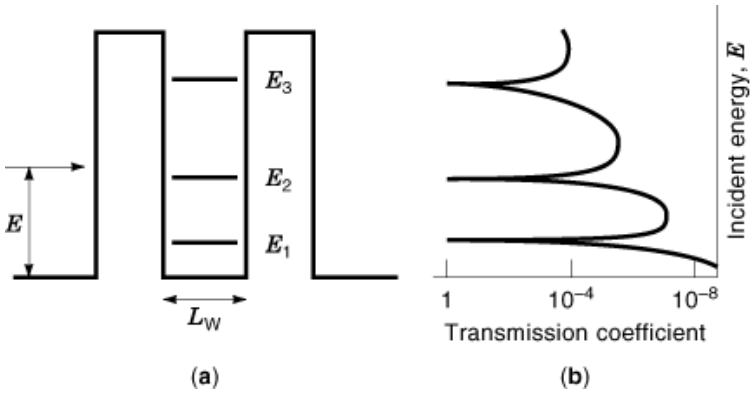

Fig. 2. (a) A particle of energy E incident on a double-barrier potential, showing quantized levels. (b) The corresponding 1-D transmission coefficient $T(E)$

A particle of kinetic energy E incident on a 1-D potential barrier of finite height V_0 and width L_B, shown in Fig. 1(b), illustrates the concept of tunneling. Classically, if $E < V_0$ the particle is always reflected, but barrier penetration analogous to Eq. (3) ensures a finite transmission probability $T(E)$ that depends on V_0 and L_B. By solving the Schrödinger equation and imposing the continuity conditions at the barrier boundaries, one can solve for the reflection $R(E)$ and transmission $T(E)$ probabilities (2), with $R + T = 1$. This problem is treated in all textbooks on quantum mechanics. For a particularly thorough discussion, see Ref. 2. For incident energy E such that $e^{-KL}B \ll 1$, where $\kappa = [2m^*(V_0 - E)/^2]^{1/2}$, one obtains

$$T(E) \approx e^{-2\kappa L_B} \tag{4}$$

Hence the transmission probability $T(E)$ for $E < V_0$ is exponentially small but nonzero, and it increases with incident energy E. This monotonic behavior changes drastically when the same particle is incident on two potential barriers separated by a well of width L_W, shown in Fig. 2(a). The double-barrier transmission $T(E)$ can be obtained (3) by repeated application of Eq. (1), but intuitively the energy levels and wavefunctions in the quantum well should coincide approximately with Fig. 1(a). Then, semiclassically (2), a particle occupying one of the energy levels E_n oscillates between the barriers with velocity $U_z = k_z/m^*$ and, in effect, is incident on a barrier twice in each period of oscillation. Every incidence involves some probability $T(E_n)$ of tunneling out of the double-barrier potential, making the levels E_n metastable.

If the energy E of a particle incident from the left does not coincide with one of the levels E_n, the total $T(E)$ for tunneling through the double-barrier potential is given by the product of the individual transmission probabilities of the first (emitter) and second (collector) barriers, $T(E) = T_E T_C$. On the other hand, if the incident energy matches one of the energy levels E_n, the amplitude of the wavefunction builds up in the well as in a Fabry–Perot resonator, and the resulting $T(E)$

$$T(E = E_n) = \frac{4T_E T_C}{(T_E + T_C)^2} \tag{5}$$

can reach unity (3). Hence, $T(E)$ is a sharply peaked function of incident energy, shown in Fig. 2(b).

Many quantum devices rely on this highly nonlinear transmission probability, which leads to a nonlinear $I(V)$ characteristic, including regions of negative differential resistance (*NDR*). The required confining potentials are provided by semiconductor heterostructures. Real incident carriers typically have some energy

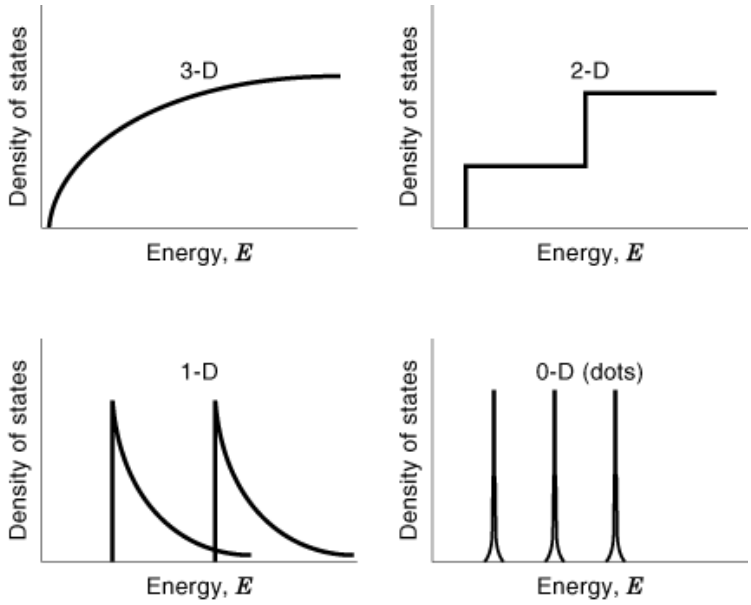

Fig. 3. Schematic densities of states for free three-dimensional motion (a); two-dimensional motion, as in a quantum well (b); one-dimensional motion, as in a quantum wire (c); and discrete spectrum of a fully confined quantum dot (d).

distribution, and the total current involves an average over this distribution. Crucially, real devices are hardly ever one-dimensional in the ideal sense of Eq. (1) (2,3,4,5) because of other spatial degrees of freedom. Carriers may be confined by potentials in one or more dimensions, while remaining free to move in the others. With the exception of full three-dimensional confinement—whether in an atom or a quantum dot—the energy spectrum is not a collection of discrete levels E_n. Instead, the carriers occupy quantized states in various subbands, and for the simplest case of isotropic constant effective mass$m*$ the appropriate densities of states can be derived by counting the states up to some energy E (4). The results follow and are illustrated schematically in Fig. 3 for the three relevant situations of no confinement at (free 3-D motion), confinement by a potential in one direction (free 2-D motion in a quantum well), and confinement by a potential in two directions (free 1-D motion in a quantum wire):

$$\text{3-D: } g^{\text{3-D}}(E) = m^{*3/2}(2E)^{1/2}/\pi^2\hbar^3 \tag{6a}$$

$$\text{2-D: } g^{\text{2-D}}(E) = m^*/\pi\hbar^2 \tag{6b}$$

$$\text{1-D: } g^{\text{1-D}}(E) = (m^*/2E)^{1/2}/\pi\hbar \tag{6c}$$

A typical semiconductor implementation of Fig. 2(a) has the double-barrier potential along the epitaxial direction $V(z)$, with free transverse motion. The total wavefunction Ψ (**r**) of an electron in one of the quantum well

levels $\chi_n(z)$ depends on in-plane momentum $\mathbf{k}_\perp$ and can be written as

$$\Psi_{n,\mathbf{k}_\perp}(\mathbf{r}) = N\chi_n(z)e^{i\mathbf{k}_\perp\cdot\mathbf{r}_\perp} \tag{7}$$

where N is a normalization factor. The corresponding total energy is given by

$$E = E_n + \frac{\hbar^2\mathbf{k}_\perp^2}{2m^*} \tag{8}$$

The density of states in the well is given by Eq. (6b). Tunneling carriers are incident from an emitter reservoir outside the double-barrier potential, where the density of states can be taken as 3-D. As long as the factorization of the wavefunction into in-plane and tunneling-direction components remains valid, transverse degrees of freedom do not complicate the situation unduly. Transverse momentum $\mathbf{k}_\perp$ remains a constant of motion as the carrier tunnels through the 2-D subbands E_n with a transmission probability$T(E_z)$ that depends on the energy of motion along the tunneling direction only, $E_z = {}^2k^2{}_3/2m^*$. The total tunneling current density$J$ can be computed by integrating over the electron distribution in the emitter reservoir:

$$J = \frac{e}{2\pi\hbar}\int N(E_z)T(E_z)\,dE_z \tag{9}$$

where $N(E_z)dE_z$ is the number of electrons with the same E_Z per unit area. Scattering by phonons, impurities, or other carriers will relax $\mathbf{k}_\perp$ conservation (inelastic processes like phonon emission also enter into energy conservation). If scattering is strong, the simple dependence of $T(E_z)$ on energy E_z given by Eq. (2) is replaced by a more complex dependence on both E_z and the total energy E of the tunneling carrier (5).

Effects associated with scattering and limited phase coherence significantly alter the idealized, sharply peaked $I(V)$ characteristic predicted by Fig. 2(b). The huge peak-to-valley ratios (*PVRs*) have not been experimentally observed, even at low temperatures. For this reason, the sequential tunneling model (6), in which transport is described by carriers tunneling into the quantized density of states in the well followed by uncorrelated tunneling out to the collector, has become a standard. The $I(V)$ nonlinearities arise from E and $\mathbf{k}_\perp$ conservation without recourse to near-unity transmission coefficients of the double-barrier potential in the coherent limit. This model and its relevance to optimizing resonant tunneling structures for device applications, such as maximizing PVR or peak current densities, are discussed in the next section.

Double-Barrier Resonant Tunneling Structures

The first experimental resonant tunneling (*RT*) data were reported by Chang, Esaki, and Tsu 7, on a GaAs/AlGaAs device. Since then, improvements in material quality and device design have led to *RT* diodes with very sharp low-temperature $I(V)$ characteristics (8), as illustrated in Fig. 4. The $I(V)$ curve exhibits strong *NDR*, with PVR reaching ~30. Still, the valley current is much larger than predicted by coherent tunneling theory because of nonresonant processes.

The sequential tunneling model (6) is illustrated in Fig. 5,using the *n*-GaAs/AlGaAs *RT* structure of Fig. 4 as an example. At $V = 0$, E_1 in the well lies above the emitter E_F, so that E and $\mathbf{k}_\perp$ -conserving tunneling is impossible. As V increases, E_1 is lowered with respect to the emitter. Resonant tunneling becomes possible when E_1 aligns with the occupied states in the emitter. The subset of emitter carriers that can tunnel into the well conserving both E and $\mathbf{k}_\perp$, often known as the supply function, can be estimated by the simple geometrical overlap between the occupied states in the emitter and the available states in the well—it is shown in boldface

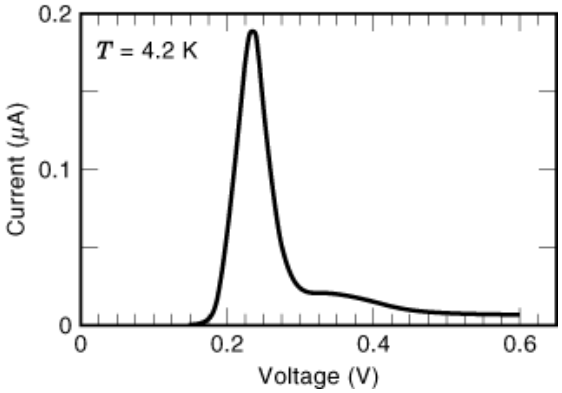

Fig. 4. $I(V)$ characteristics of a GaAs/AlGaAs double-barrier RT structure at $T = 4.2$ K. After Zaslavsky et al. 8.

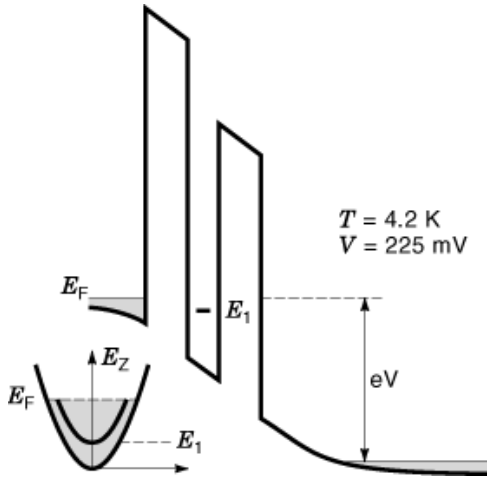

Fig. 5. Self-consistent potential distribution in a GaAs/AlGaAs RT structure under bias $V = 0.225$ V at $T = 4.2$ K. The supply function is obtained from the geometric overlap between the $E(\mathbf{k}_\perp)$ dispersions in the emitter and well, as shown at lower left. After Zaslavsky et al. 8.

in Fig. 5. Larger V lowers E_1 below the occupied states in the emitter, at which point the supply function drops (ideally to zero). The current density into the well due to E and $\mathbf{k}_\perp$ conserving tunneling is then

$$J = \frac{e}{2\pi\hbar} N(V) T_{\mathrm{E}}(V) \qquad (10)$$

Other current components, like direct tunneling into the collector and phonon-, impurity-, or interface roughness-assisted tunneling, must be added to Eq. (10). In addition, at sufficiently high V tunneling through the second subband E_2 becomes possible. Nonresonant components contribute to the valley current: for example, the strong phonon coupling in GaAs leads to a phonon-assisted replica peak when E_1 is biased below the emitter by the GaAs optical phonon energy $\omega_{\mathrm{opt}} = 36$ meV (9). The quantitative modeling of nonresonant current components usually relies on adjustable parameters (10). This is unfortunate, because the valley current plays an important role in the minimum power dissipation of RT devices.

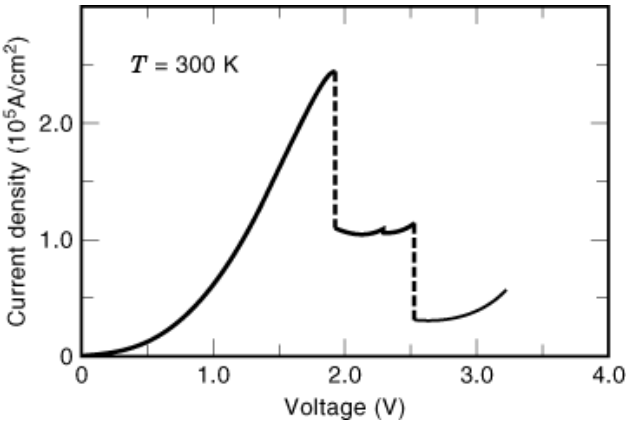

Fig. 6. $I(V)$ characteristics of an $In_{0.53}Ga_{0.47}As/AlAs$ *RT* diode, showing high peak-to-valley ratio and peak current density at $T = 300$ K. Device area is μm^2. Figure courtesy of E. R. Brown, 1995.

The resonant current of Eq. (10) depends sensitively on the alignment of E_1 with the occupied emitter states. Yet in a standard, two-terminal *RT* diode, this alignment can only be controlled by the applied bias V. If electric fields in the emitter and collector regions are ignored, one immediately obtains that $V(^n)_P = 2E_n$ in a symmetric structure. However, in real devices, potential drops in the emitter and collector regions are significant, especially in high-frequency *RT* diodes, where low emitter-collector capacitance is often achieved by a large undoped collector spacer (11). A self-consistent calculation of the potential, as shown in Fig. 5, is therefore necessary to predict $V_P{}^{+++}$. An additional complication is the dynamically stored charge density σ_W in the well as the tunneling current flows. The effect of σ_W is to screen the first barrier and hence reduce the bias-induced lowering of the 2-D subbands. In symmetric *RT* structures,σ_W is small, but if the collector barrier is made larger, σ_W becomes significant. In highly asymmetric devices, the electrostatic feedback due to σ_W can cause a hysteretic loop to appear in the $I(V)$, with V_P depending on the direction of the bias sweep (12). At least in principle, a bistable $I(V)$ makes a two-state semiconductor memory possible in a single device.

Double-barrier GaAs/AlGaAs structures have been very useful in clarifying the relevant physics, but their $I(V)$ characteristics are less suitable for devices. First, the sharp *NDR* needs to survive at room temperature, where the valley current is supplemented by thermionic emission over the barriers and temperature-assisted tunneling through higher-lying subbands. These PVR-degrading effects can be reduced by designing higher barriers and increasing the subband separation. In contrast, for high-speed operation, one needs to maximize J_P, since high currents necessary for rapid charging of the various capacitances—ideally $J_P \geq 10^5$ A/cm^2. The use of very narrow and high barriers is therefore indicated, as well as the lightest possible effective mass m^* to increase subband separation. Limitations of the GaAs/AlGaAs system have kept the fastest double-barrier *RT* oscillators (13) with high $J_P \approx 10^5$ A/cm^2 at a room-temperature PVR of only 3. Better PVR and J_P have been obtained in InGaAs/AlAs *RT* diodes: a device with $J_P > 10^5$ A/cm^2 and PVR $\sim$8 is shown in Fig. 6 (14). This *RT* structure included a large undoped collector spacer region to reduce the emitter–collector capacitance, hence the high V_P. Because of the sharp *NDR*, the biasing circuit becomes unstable above V_P, leading characteristic jumps in the $I(V)$ (15).

Another variant of two-terminal *RT* devices uses polytype GaSb/AlSb/InAs heterostructures with a staggered bandgap alignment (16). A schematic polytype band diagram is shown in Fig. 7. The current is due to holes tunneling from the GaSb emitter into the InAs well—a geometrical evaluation of the supply function requires inverting the emitter dispersion in Fig. 5 (17). The polytype structure represents an *RT* version of the Esaki tunnel diode. Its advantage lies in the bandgap blocking beyond V_P, where the emitter states line up

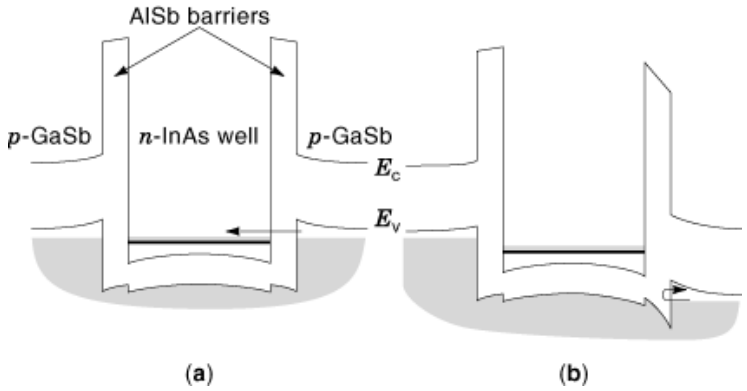

Fig. 7. (a) Potential diagram of a polytype GaSb/AlSb/InAs *RT* structure near flatband and (b) under large bias $V > V_P$. Arrows indicate the tunneling current of GaSb holes into the *n*-InAs quantum well. In (b) resonant tunneling is blocked by the InAs bandgap. After Beresford et al. 18.

with the InAs bandgap. Very good PVR has been achieved in polytype structures, albeit at modest peak current densities, and *RT* designs with very wide wells $L_W \approx 1000$ Å are realizable without compromising PVR (18).

Silicon-based *RT* devices have been fabricated in SiGe/Si heterostructures (19). Unfortunately, the available barrier height is rather small, $V_0 \approx 200$ meV. Because of low V_0 and relatively heavym^*, no room temperature *NDR* has been observed in SiGe/Si devices to date, although PVR ∼4 has been observed at cryogenic temperatures (20). Consequently, even though these devices have been employed for spectroscopy of anisotropic hole dispersions (21) and strain relaxation (22), the prospects of their integration into mainstream technology appear remote.

In addition to epitaxial double-barrier structures, lateral tunneling devices have been fabricated by gating a modulation-doped 2-D electron gas (2DEG) heterostructure. By applying a gate potential V_G with respect to the 2DEG, electrons can be electrostatically depleted underneath the surface gates, creating the double-barrier potential. The advantages include: excellent electronic properties of the 2DEG; tunability of V_G-induced barriers; and planar device layout. The main drawback is the relative weakness of the barriers. Hence, the $I(V)$ of lateral *RT* structures produced by electrostatic gating show *NDR* at low temperatures only (23).

Two-terminal *RT* structures are useful for oscillators and frequency multipliers but ill-suited for more general circuitry. The addition of a third terminal to control the $I(V)$, either with a small current as in a bipolar transistor or a gate voltage V_G, has been attempted in a number of schemes.

Current-controlled three-terminal *RT* devices involve a separate contact to the quantum well that can source or sink a "base" current large enough to alter the alignment of E_n and the emitter states (24). If the device is biased close to V_P, a small base current can lower E_n below the emitter, giving rise to negative transconductance. A significant constraint on such devices is the effective base resistance. To have significant 2-D subband separation and hence strong *NDR* in the $I(V)$, the quantum well width L_W must be small. But the lateral base resistance is inversely proportional toL_W. Setting the benchmark for a high-speed device at 1 ps, the R_B C time delays due to emitter-well or well-collector capacitance are

$$R_B C \approx \epsilon_s L^2 R_s / L_B \qquad (11)$$

where ε_S is the dielectric constant, L is the lateral extent of the device, L_B is the emitter or collector barrier thickness, and R_S is the sheet resistance of the base. The resulting $R_S \leq 1\,\text{k}\Omega$ per square constraint is difficult

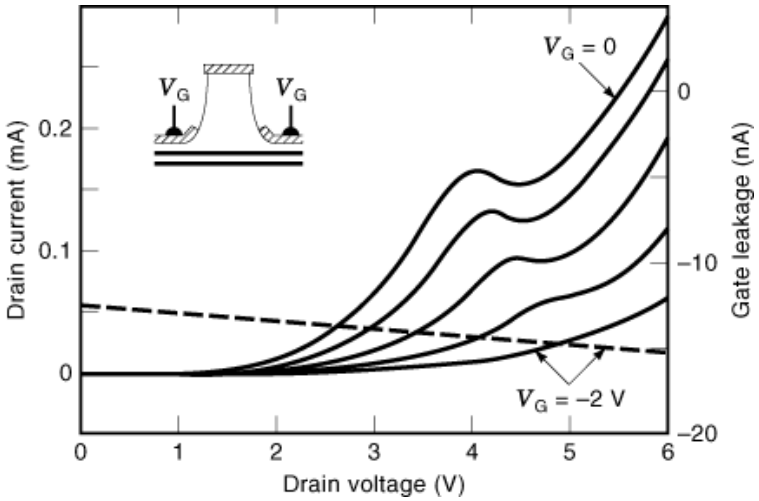

Fig. 8. Room temperature $I(V, V_G)$ characteristics of a three-terminal Schottky-gated GaAs/AlGaAs *RT* device (shown schematically in the inset) of 10 μm $\times$ 0.7 μm stripe geometry for $V_G = 0$ to 2.0 V in 0.5 V increments. Dashed line shows gate leakage for $V_G = 2.0$ V. After Kolagunta et al. (28).

to meet in narrow $L_W \approx 100$ Å quantum wells and, moreover, impurity scattering in the heavily doped well can wash out the *NDR*.

An alternative route to a three-terminal *RT* structure is voltage control by a sidewall gate. The vertical pillar geometry of epitaxially grown *RT* structures makes the fabrication rather difficult. Self-aligned *p*-type implantation with the top metal contact of an *n*-type *RT* diode serving as a mask has been used to control the *RT* current by reverse-biasing the resulting in-plane *p*–*n* junction (25). In addition to leakage currents, lateral straggle of the implantation becomes a problem for deep submicron devices. Alternatively, a metal Schottky gate can be deposited on the sidewalls of the *RT* diode pillar (26,27). By employing an undercut *RT* pillar profile (see inset of Fig. 8), room temperature three-terminal operation has been reported (28): the $I(V, V_G)$ curve of a GaAs/AlGaAs *RT* stripe geometry device is shown in Fig. 8. Gate control is achieved with reasonably small gate leakage, but note that the side-gating geometry of Fig. 8 sacrifices the effective transconductance g_m unless the pillar diameter is extremely narrow, resulting in formidable fabrication difficulties.

A device that combines gate control with a planar layout is the double electron layer tunneling transistor (29), where resonant tunneling occurs between 2DEGs in an epitaxially grown GaAs/AlGaAs double quantum well structure (see inset of Fig. 9). The source and drain make separate electrical contact to the two 2DEGs, so the source-drain current is determined by inter-2DEG tunneling (30). In this 2-D–2-D tunneling geometry, conservation of E and $\mathbf{k}_\perp$ requires precise alignment of the 2-D subbands in different wells. Drain voltage V_D changes the subband alignment; as a result, a two-terminal source-drain $I_D(I_D)$ curve exhibits a resonant current peak followed by an *NDR* region. Third-terminal control is provided by either a surface gate or a backgate, which can modulate the subband energy in either the upper or the lower well. A set of $I_D(V_D, V_G)$ curves showing gate control of the peak position, is shown in Fig. 9 for a surface-gated device at low temperature. Available PVR in this device reaches $\sim$10 at $T = 77$ K; InGaAs-containing heterostructures may yield room temperature operation. The separate contacting of the 2DEGs necessitates the fabrication of depletion gates both above and below the double quantum well active region [and hence flip-chip processing (29)]. By combining two devices in series, one with a backgate and one with a surface gate, unipolar complementary memories are achievable (31).

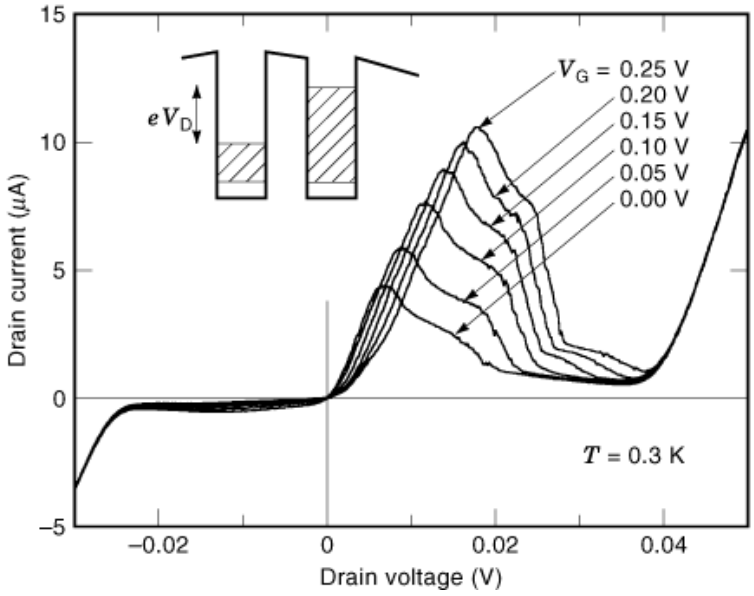

Fig. 9. Electrical $I(V_D, V_C$, characteristics of the double electron layer tunneling transistor, showing surface gate control of the current peak position, at low T. Inset shows the band diagram of the double quantum well structure: resonant tunneling is only possible when the energies of the subbands in both wells coincide. Figure courtesy of J. Simmons, 1998.

A long-proposed three-terminal *RT* structure is shown in Fig. 10 (32). The original double barrier is grown epitaxially; then an angled interface is etched, and an AlGaAs gate insulator is deposited, followed by a gate electrode. A positive gate bias V_G induces 2DEG in the undoped GaAs layers as in a standard FET. In the well, 1-D subband quantization E'_1 arises from the double-barrier potential $V(z)$ combined with the gate-induced $V(x)$. A potential difference between the 2DEGs above and below the double-barrier potential will produce a tunneling current subject to the usual E and $\mathbf{k}_\perp$ conservation, with the conserved $\mathbf{k}_\perp = k_y$ corresponding to free motion along the quantum wire. The gate bias V_G controls the emitter 2DEG density and hence the magnitude of the *RT* current. Interestingly, V_G can also be used to tune V_P because the fringing electric field shifts $E^{\text{prime;}}{}_1$ with respect to E_F for the same source-drain bias V. As a result, $g_m < 0$ can be achieved. If the 2DEG depletion in the collector region is ignored, the electric field distribution can be solved by conformal mapping techniques and g_m can be extracted (29). In real devices, depletion of the 2DEG in the collector reduces the gate control of E'_1 and hence the transconductance.

The geometry in Fig. 10 makes the device difficult to fabricate and only proof-of-concept prototypes have been achieved by cleaved edge regrowth (33). Low-temperature $I(V, V_G)$ curves of the resulting device (34) and the corresponding negative g_m are shown in Fig. 11. Room temperature operation and fabrication of such devices by technological means are yet to be reported.

In addition to the severe fabrication problems faced by three-terminal *RT* devices, it is not clear that their negative g_m can be usefully applied for computation. Although it has been suggested that such devices can perform complementary functions (35), no *RT* transistor circuit analogous to a CMOS inverter has been demonstrated to date. In a complementary CMOS transistor pair, the current is due to carriers of opposite polarity, and the drains of the two transistors can be connected. It might appear that both transistors in the CMOS pair can be directly replaced by a pair *RT* devices with negative g_m . But in *RT* devices, the current depends on the alignment of the emitter and quantum well densities of states and hence on the emitter bias V_E. The effective emitter bias V_E on the second *RT* device in the pair would itself vary between a high and a

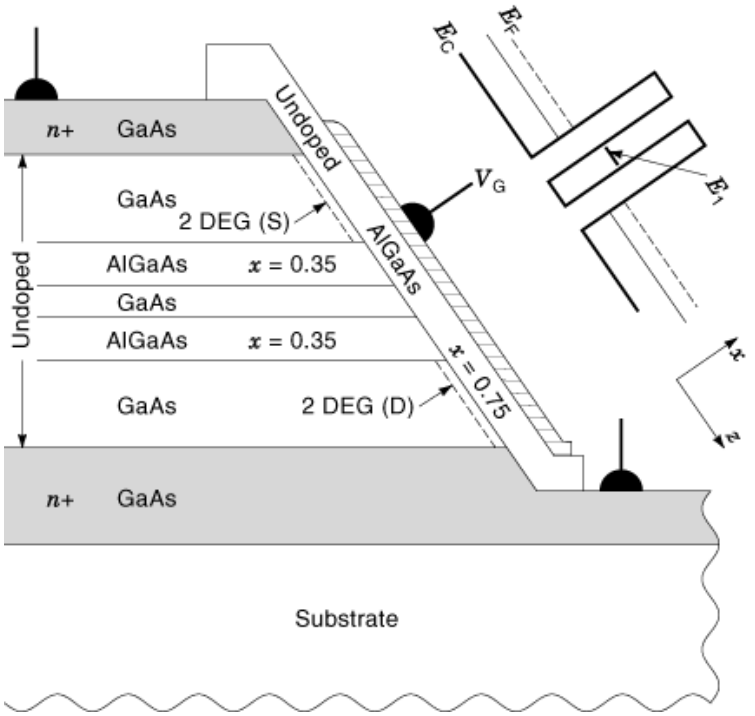

Fig. 10. Schematic cross section of a gated 2-D *RT* transistor together with the band diagram. The current is carried by 2-D electrons tunneling through 1-D quantum wire subbands. Both the 2-D electron gas density and the relative alignment of the wire subbands E_n with the emitter 2-D electron gas can be controlled by the gate. After Luryi and Capasso (32).

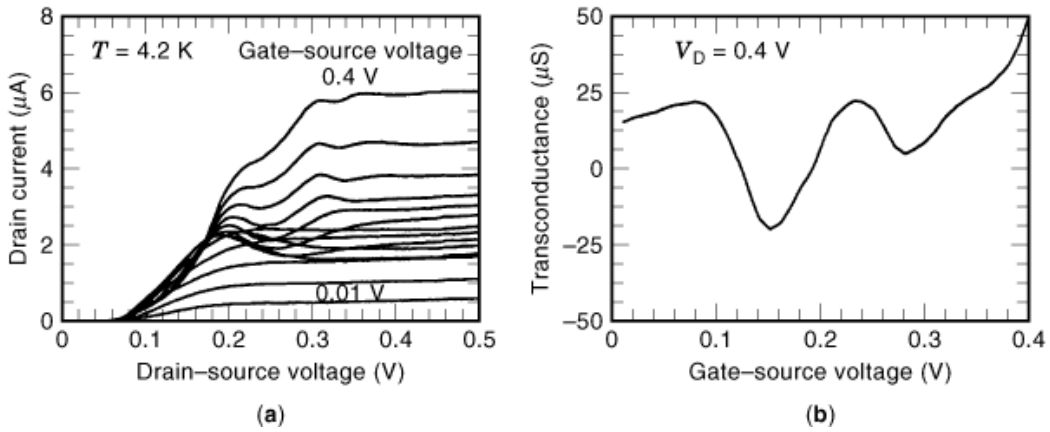

Fig. 11. (a) Three-terminal $I(V_D, V_G)$ at $T = 4.2$ K of a gated 2-D *RT* transistor produced by cleaved edge overgrowth. Gate bias V_G is changed in 0.03 V steps. (b) Corresponding transconductance at $V_D = 0.4$ V. Device width is 300 μm. After Kurdak et al. (34).

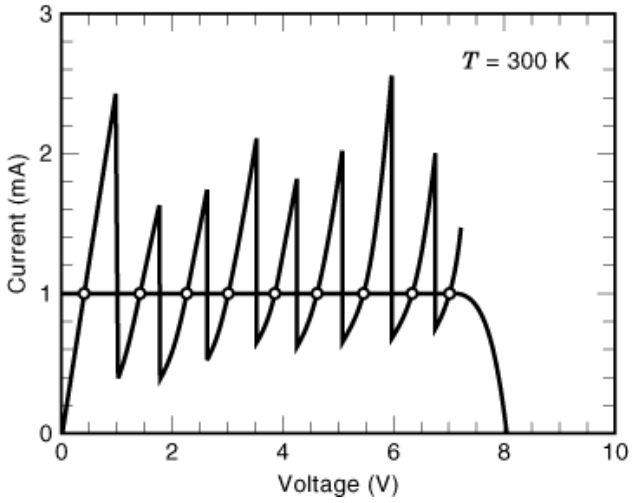

Fig. 12. Room-temperature $I(V)$ of a cascaded RT structure with $N = 8$ diodes in series. Device area is 32 μm^2. Superimposed in the $I(V)$ curve is the load line when the device is biased by an FET current source, with stable points indicated by open circles. After Seabaugh et al. (36) (© 1992 IEEE).

low state, rather than remaining at a constant potential. In CMOS inverter language, in an RT pair the gate voltage V_{IN} would be referenced to V_{OUT}, rather than V_{DD}, making it difficult to design a useful circuit (4).

Cascaded Double-Barrier Structures and Superlattices

Many applications of RT devices require a multipeak $I(_V)$ with peaks of the same magnitude and regularly spaced in voltage. Neither condition is met by a standard RT structure: the subband separation $(E_n - E_{n-1})$ changes with n, so the peak voltages $V_P{}^{(n)}$ are not evenly spaced, and the peak currents increase rapidly with n as the emitter barrier height drops. However, a multipeak $I(V)$ curve can be obtained from a cascaded RT structure with n double-barrier potentials separated by doped cladding regions. Current flows once the RT diodes are biased above threshold, $V \geq NV_{th}$. As V is increased, one of the diodes will be biased beyond V_P, suddenly presenting a high resistance to the biasing circuit, so the $I(V)$ exhibits NDR. If V is increased further, current continuity requires that almost all of the increase drop over the off-resonance diode, until it begins to conduct through the higher subband E_2. This process then is repeated with other diodes, for a total of n evenly spaced peaks in the $I(V)$, as shown in Fig. 12 (36). This type of device can be used for a multistate memory.

If the doped cladding regions are removed and the subbands in different wells are allowed to interact, the result is a superlattice (SL) of period $d = L_B + L_W$ shown in Fig. 13(a). Consider the wavefunction $\psi(z)$ along the SL direction. If the barriers are infinitely high, we have isolated quantum wells with the usual quantized levels E_n described by wavefunctions $\chi_n{}^{(m)}(z)$, where m labels the quantum well. If the barriers are finite, the $\chi_n{}^{(m)}(z)$ wavefunctions penetrate into the barriers according to Eq. (3), and the previously degenerate levels will broaden into minibands Δ_n. According to the Bloch theorem, a state in the nth miniband of a superlattice can be described by linear combinations of wavefunctions periodic in the SL period d, multiplied by a plane wave (37)

$$\Psi^{kz}(z) = \sum_{m=1}^{N} e^{ik_z md} \varphi_n^{(m)}(z) \qquad (12)$$

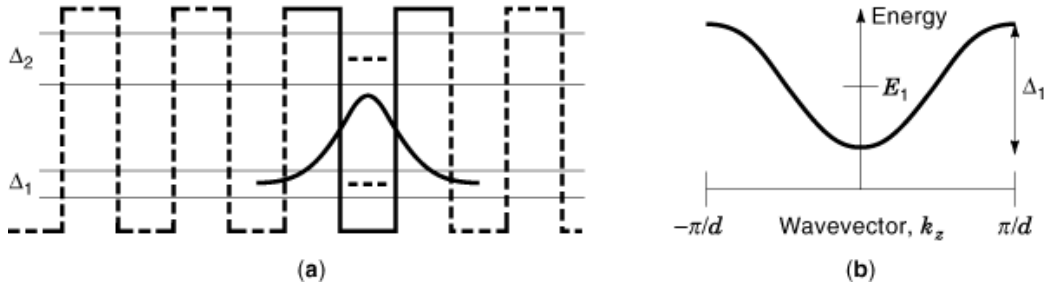

Fig. 13. (a) Superlattice potential diagram, showing the broadening of energy levels E_n into minibands of width Δ_n. The superlattice period is $d = L_W + L_B$. Dotted lines mark the two lowest energy levels confined by the mth potential well (bold line), wavefunction $\chi_j{}^m(z)$ is also shown. The potential $V'_0(Z)$ used in the calculations of miniband dispersion is shown by a dashed line—it includes all wells other than the mth. (b) Model superlattice miniband dispersion.

As long as $\Delta_n \ll (E_n - E_{n-1})$, $\varphi^{(m)}{}_n(z)$ are to a good approximation built up from combinations of $\chi_n{}^{(m)}(z)$. For some range of barrier parameters V_0 and L_B, only adjacent wells matter (this is known as the tight-binding approximation), and the problem simplifies drastically. The dispersion $E(k_z)$ for motion along the SL axis becomes

$$E_n^{SL}(k_z) = E_n + S_n + 2T_n cos(k_z d) \tag{13}$$

where the shift integral S_n is defined as

$$S_n \equiv \int \chi_n^{(m)}(z) V_0'(z) \chi_n^{(m)}(z)\, dz \tag{14}$$

and the transfer integral T_n as

$$T_n \equiv \int \chi_n^{(m)}(z) V_0'(z) \chi_n^{(m+1)}(z)\, dz \tag{15}$$

The potential $V'_0(z)$ employed in the calculation of the shift and transfer integrals includes all potential wells other than the mth [see Fig. 13(a)]. The width of the nth miniband $\Delta_n = 4T_n$ and the allowed values of k_z obtain from periodic boundary conditions on Eq. (13):$k_z = 2\pi P/Nd$, where $P = 0, 1, 2, \ldots (N - 1)$.

The dispersion for motion along the SL direction is plotted in Fig. 13(b). It is evident that the effective mass along the SL, $m^* = {}^2(k_z{}^{-1}\partial E/\partial k_z)^{-1}$, is a strongly varying function of k_z: starting with a "band-edge" value $m^*_{SL} \equiv m^* k_z = \pi/2d$, and becomes negative thereafter.

If a small electric field $\mathscr{E}$ is applied along the SL direction and there is no scattering, the semiclassical equation of motion $(\partial k_z/\partial t) = e\,\mathscr{E}$ implies that k_z changes linearly with time. Since $V(k_z)$ is periodic, carriers execute the so-called Bloch oscillations. It was the pioneering suggestion of Esaki and Tsu to employ semiconductor heterostructures to create the required SL potential that opened the modern era of heterostructure bandgap engineering (38). In that celebrated paper, the effects of a finite scattering time τ on the average drift velocity V_D of electrons propagating in a 1-D superlattice with dispersion given by Eq. (13) was evaluated

classically:

$$v_{\mathrm{D}} = \int_{t=0}^{\infty} e^{-t/\tau} a(t)\, dt \qquad (16)$$

where $a\ (t) \equiv a[k_{\mathrm{z}}(t))]$ is the acceleration of the miniband electron. The result for v_{D} in terms of E τ, SL period d, and $\mathrm{m}^*_{\mathrm{SL}}$ was

$$v_{\mathrm{D}} = \frac{\hbar}{m^*_{\mathrm{SL}} d} \frac{\xi}{1+\xi^2} \qquad (17)$$

where $\xi = e\mathscr{E}\ \tau \mathrm{d}/$ The average drift velocity peaks at $\xi = 1$, that is when the electric field $\varepsilon = /\ \mathscr{E}\tau d$. Beyond this point, increasing $\mathscr{E}$ results in a lower v_{D} because more and more carriers reach the negative-mass region. As a result, the SL $I(V)$ characteristic should exhibit *NDR*. Experimentally, because of Zener tunneling between minibands and electric field domain formation due to space-charge instabilities (39), only relatively weak $I(V)$ nonlinearities have been observed (40).

If the electric field $\mathscr{E}$ is sufficiently large, the nth miniband breaks up into a set of discrete levels, known as the Wannier–Stark ladder of states shown in Fig. 14(a) (41,42). As soon as the extent of the Wannier–Stark wavefunctions falls below N periods, they no longer reach from one end of the SL to another. Scattering becomes necessary for dc current to flow, and the current will remain small until $\mathscr{E}$ brings into resonance Wannier–Stark states in adjacent wells. At these sharply defined values of $\mathscr{E}$, $e\ \mathscr{E}_{\mathrm{j}}$, $d = E_{\mathrm{j}} - E_1, j = 2, 3 \ldots$, the current can flow by sequential tunneling between different Wannier–Stark states in adjacent wells, followed by relaxation to a lower-lying state [see Fig. 14(b)]. The $I(V)$ curve should then exhibit peaks at $V = Ne\mathscr{E}_j\ d$, followed by *NDR* regions.

A particularly interesting process is photon emission in the regime where$\mathscr{E} > \mathscr{E}$, which was proposed by Kazarinov and Suris decades ago as a voltage-tunable laser (43). The scheme is shown in Fig. 14(c) and the photon energy is$\omega = e(\mathscr{E} - \mathscr{E}_j)d$, tunable in the infrared by the applied voltage. The problem with this exciting possibility, as with the Esaki–Tsu *NDR* at low,$\mathscr{E}$ is maintaining a uniform electric field $\mathscr{E}$ throughout the SL. Devices that operate in the *NDR* regions of their $I(V)$ characteristics are particularly susceptible to the electric field breaking up into high- and low-field domains (39). For this reason, voltage-controlled lasing illustrated in Fig. 14(c) has not been observed, and it is not clear whether it can be observed even in principle.

On the other hand, the alignment provided by $\mathscr{E}_j$ between different Wannier–Stark states in adjacent wells can also provide a lasing medium, provided that at least some fraction of the $E_2 \rightarrow E_1$ relaxation is radiative [see Fig. 14(b)]. The voltage tunability is now lost, since $\omega = (E_2 - E_1)$, but now the device need not operate in the *NDR* region. Infrared lasing in a conceptually similar device—the quantum cascade laser (QCL)—based on intersubband transitions in a modified SL structure has been achieved recently (44). A more detailed discussion of the QCL follows.

Resonant Tunneling Nanostructures and Coulomb Blockade

If a double-barrier *RT* structure is etched into a very narrow pillar or biased to a narrow effective size by a lateral gate, new quantum effects come into play. The more obvious is lateral size quantization in the quantum well. Lateral confining potentials are weak compared to the narrow wells available by epitaxy, so lateral quantization will be much weaker: to a good approximation, each of the 2-D subbands E_{n} in the well will give rise to a series of fully quantized states E_{nm}, where m labels the states of the lateral potential $V(x, y)$. Tunneling into these discrete levels can be treated within the usual sequential formalism (6,45) but with a new effect. The

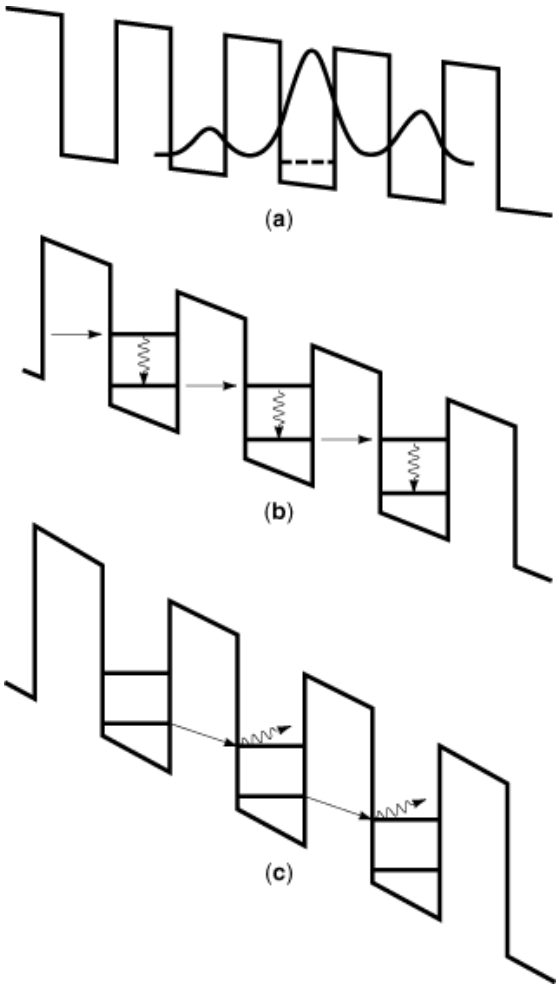

Fig. 14. (a) Schematic band diagram of a biased superlattice, in which the minibands break up into Wannier–Stark states. The lowest-lying Wannier–Start state E_1 in the mth well is shown by the dotted line, together with the corresponding schematic wavefunction for $\Delta_1/e\mathscr{E}d \approx 3$, where Δ_1 is the miniband energy width. (b) Sequential tunneling through a superlattice when excited and ground states in adjacent wells are aligned, resulting in a current peak. Vertical lines represent intersubband relaxation. (c) Photon-assisted sequential tunneling tunable by the electric field.

charging energy U required to transfer even a single electron into the well becomes significant for small devices. If the charging energy is ignored, the situation is shown in Fig. 15(a). Since the lateral confining potential $V(x, y)$ changes between the emitter and well, $\mathbf{k}_\perp$ is no longer a conserved quantity. As the bias V lowers E_{11} below E_F in the emitter, tunneling through this single state becomes possible—this defines the threshold V_{th}. At higher V, additional tunneling channels open up. The resulting $I(V)$ will exhibit a rising staircase of steplike features (46), their strength depending on the transmission $T_E(V)$ and also on the degeneracy of the E_{1m} states. No *NDR* is expected in the $I(V)$ because $\mathbf{k}_\perp$ conservation no longer impedes tunneling through higher-lying E_{1m} states when E_{11} drops below the occupied states in the emitter.

This picture of tunneling into a quantum dot would be unpromising from the device standpoint, if not for the charging energy $U = e^2/2C_W$ associated with the tunneling of a single electron into the well (C_W is

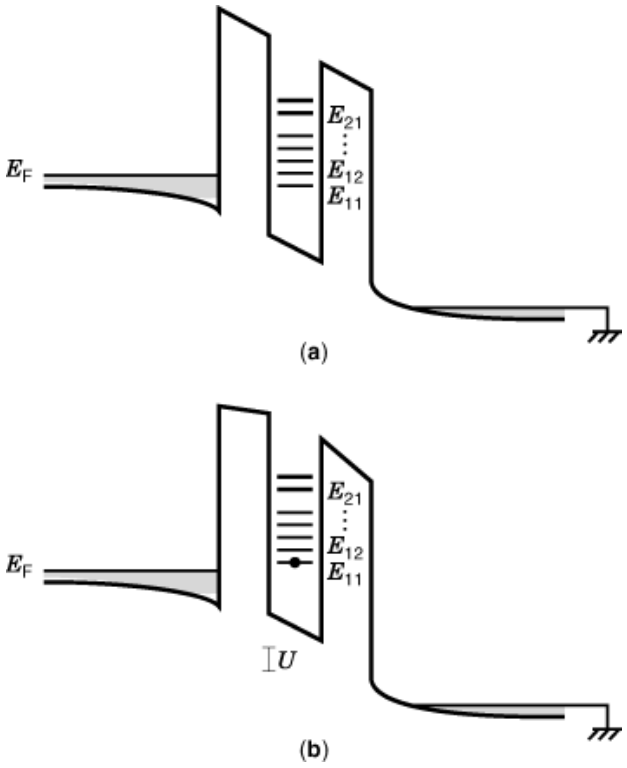

Fig. 15. (a) Schematic band diagram of a double-barrier RT nanostructure with lateral quantization. The levels E_{nm} arise from the quantization of the nth 2-D subband E_n into discrete quantum dot states. (b) Coulomb blockade regime. A single electron tunneling into the well changes the emitter–well alignment by $U \approx e^2/2C_W$, where C_W is the dot capacitance. Tunneling into the well is cut off if U raises E_{11} above the occupied states in the emitter, as shown.

the effective capacitance of the dot). This U can appreciably alter the alignment of E_{nm} with emitter E_F, as illustrated in Fig. 15(b). A simple, geometric estimate of the capacitance is $C_W \approx \varepsilon_S L^2/L_B$, where L_B is the effective barrier thickness. For current to flow, at least one electron must tunnel into the dot. So, V_{th} shifts proportionally to the single-electron charging energy U. The shift to other steplike features depends on the average occupation of the dot, which is determined by the transmission ratio T_E/T_C of the emitter and collector barriers (46). If $T_E/T_C \ll 1$, the occupation of the dot by more than one electron at a time is rare and all the steplike features in the $I(V)$ corresponding to additional tunneling channels shift together. On the other hand, if $T_E/T_C \gg 1$, each available level is occupied most of the time, so the opening of every additional channel requires sufficient biasing to overcome the charging energy—the so-called Coulomb blockade. Such devices have been studied to probe the energy spectra of quantum dots with and without electron–electron interactions (46,47).

It is the charging energy required to change the electron occupation that makes RT nanostructures promising for devices. Consider a third gate electrode that can change the potential between the quantum dot and the emitter but that is sufficiently isolated from the dot to prevent any possibility of electron transfer from the gate. Then, if the device is biased by V_E near a voltage step corresponding to the addition of another electron to the dot, a small change in V_G can tune the occupation of the dot, resulting in a single-electron transistor. Because of fabrication difficulties, gate control of single-electron tunneling is easier in the planar geometry, with the dot and gate electrode defined by electrostatic metal gates deposited on top of a high-mobility 2DEG. A

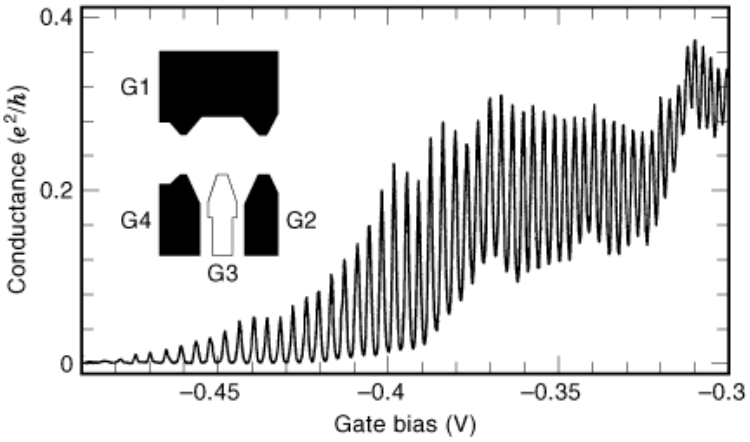

Fig. 16. Conductance of a small 2-D electron gas island as a function of gate bias V_G, at $T = 20$ mK. Inset shows a top view of the island geometry. Gates G1, G2, and G4 are biased into deep depletion, forming a 2DEG island weakly coupled to 2DEG electrodes. V_G is applied to the gate G3. Changes in the electron occupation of the island produce regular spikes in the conductance. Figure courtesy of C. J. B. Ford, 1996.

top view of the gated structure is shown in the inset of Fig. 16: the outside gates are biased into deep depletion, forming a small island of 2DEG connected to the reservoirs by tunneling barriers. The island is quite large, so the lateral size quantization is negligible. As a result, the energy spectrum is entirely defined by the charging energy U. The gate electrode can alter the effective size and capacitance of the island, so the $I(V)$ characteristic as a function of V_G should show regularly spaced steps corresponding to the adding of electrons to the island. At low temperatures, very regular conductance ($g = \partial I/\partial V$) peaks have been observed in such structures (48), an example is shown in Fig. 16.

Precise single-electron control over the occupation of small quantum dots or islands has led to many proposals of logic and memory circuits based on single-electron transistors (*SETs*) (49). To some extent, single electron devices can be considered the logical endpoint of miniaturization-driven semiconductor technology. In fact, standard silicon floating-gate memories have been made small enough for the single-electron charging of the gate to control the transistor threshold, resulting in "classical" single-electron devices (50,51). The main difficulty for large-scale circuitry of this type is posed by the extremely stringent fabrication requirements, especially at noncryogenic temperatures. Currently, *SET* characteristics (like the data in Fig. 18) are measured at low temperature, to ensure the condition $U = e^2/2C \gg_W kT$. Clearly, device sizes will need to be reduced by orders of magnitude before higher temperature operation can be contemplated. Even at $T = 77$ K, the charging energy must be certainly larger than 10 meV. This requires a capacitance $C_W < 10^{-17}$ F, a very stringent condition. It is imperative to avoid parallel capacitance due to leads or other electrodes, since even a simple thin wire has an intrinsic capacitance of about 10^{-16} F/μm. It is also not clear that semiconductor *SET* realizations have any advantages over metal tunnel junctions for most proposed devices: the first observation of Coulomb-blockade phenomena (52) and the first *SET* with voltage gain (53) both employed small Al tunnel junction capacitors. One specific application for which the *SET* appears promising is the construction of precision current standards. In a gated 2DEG island, by sequentially lowering and raising the emitter and collector barriers in the Coulomb-blockade regime, the transfer of one electron per cycle of barrier biasing can be achieved (54). If the barriers are cycled at a frequency f, the current is given by $I = ef_W$, making for a very precise current source. It is anticipated that such a device may provide a new metrological current standard, although single-electron transfer along a chain of small metallic islands may prove a more successful implementation (55).

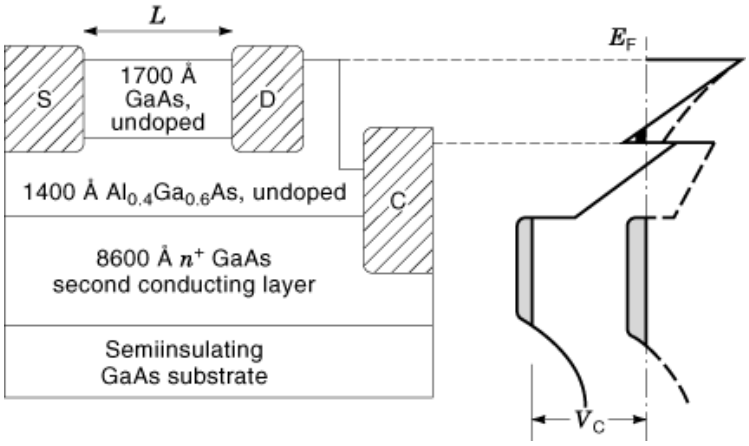

Fig. 17. Schematic cross section and energy band diagram of a three-terminal GaAs/AlGaAs RST transistor. In this version, there are no electrons in the channel at flatband—the channel density $n(x)$ is induced by collector bias V_c. If instead the channel consists of a modulation-doped quantum well, $n(x)$ can be nonzero at $V_c = 0$

Real-Space Transfer Structures

Another approach to exploit a highly nonlinear $I(V)$ involves the integration of the real-space transfer (RST) mechanism with a third terminal. The idea behind RST is the transfer of carriers between semiconductor layers of high and low mobility (56), first demonstrated in two-terminal modulation-doped multiquantum well GaAs/AlGaAs heterostructures (57). When the electric field along the GaAs wells is small, electrons remain there, and the source–drain $I(V_D)$ is ohmic. However, as the field is increased, the power input into the electron distribution exceeds the energy loss into the lattice, and electrons heat up to some effective temperature T_e. At high T_e, there is partial transfer over the heterostructure barrier V_0 to the AlGaAs layers, where the mobility is much lower, giving rise to *NDR* in the two-terminal $I(V_D)$. The analogy to the Gunn effect is obvious.

A realistic treatment of electron heating in an RST structure involves the formation of longitudinal electric field domains, redistribution of carriers both vertically and laterally, self-consistent electric fields in the transfer direction, and quantum mechanical reflections at the barriers. The complexity of the problem compels the use of Monte Carlo techniques (58). Like two-terminal resonant tunneling diodes, two-terminal RST structures are potentially useful as high-frequency oscillators, but the available PVR is rather small. What makes RST structures considerably more interesting as devices is the possibility of extracting the hot carriers via a collector electrode, resulting in a three-terminal RST transistor (59).

Figure 17 shows a schematic cross section and band diagram of such a device. The source and drain contacts are to a high-mobility GaA? channel, whereas the collector is separated from the channel by a large AlGaAs barrier. As V_D is increased, a drain current I_D begins to flow, and the carriers heat up to some effective temperature $T_e(V)_D$. This temperature determines the RST current injected over or tunneling through the collector barrier, giving rise to I_C. Unlike the two-terminal device, here the RST current is removed from the drain current loop, leading to very strong *NDR* in the $I_D(V_D)$ curve, with room temperature PVR reaching 160 in GaAs/AlGaAs devices (60), and up to 10^5 in pseudomorphic InGaAs/GaAs RST transistors (61).

There has been recent progress in Si/SiGe RST transistors, which are more promising for integration into mainstream technology. The drain $I_D(V_D)$ and collector I_C (V_D) characteristics of a p-Si/SiGe RST transistor at room temperature 62 are shown in Fig. 18. Here SiGe layers comprise the channel and collector regions, separated by a 3000 Å Si barrier. Negative $V_?$ induces a hole density in the channel, whereas V_D heats the holes. As V_C increases, the drain characteristic exhibits *NDR*, with PVR slightly exceeding two. Even though the PVR

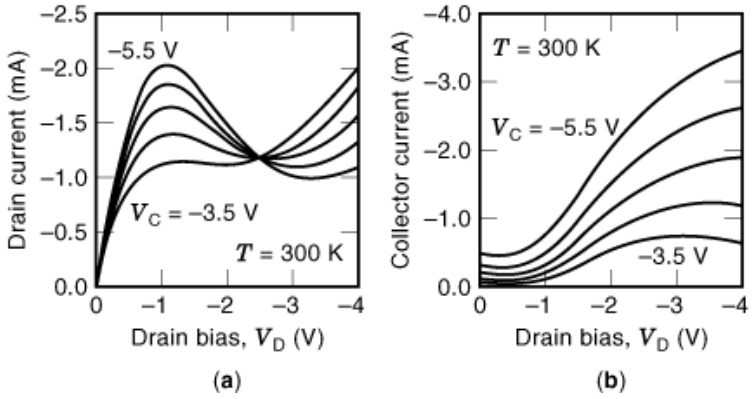

Fig. 18. Room temperature Si/SiGe RST transistor (a) drain and (b) collector current as a function of drain bias V_D for constant $V_C = -3.5$ to -5.5 V in -0.5 V increments. Source–drain channel length $L = 0.5$ μm, device width is 40 μm. After Mastrapasqua et al. (62).

is greatly inferior to that available in III–V devices, it is sufficient to implement a single-device exclusive-OR gate with a 10 dB on/off ratio at $T = 300$ K and a 65 dB on/off ratio at $T = 77$ K. For 0.5 μm source–drain separation, this device had a current-gain cutoff frequency $f_T = 6$ GHz

Like resonant tunneling devices, RST transistors provide *NDR* characteristics in the source–drain circuit. Further, since the source and drain contacts of an RST transistor are fully symmetric, these devices have additional logic functionality. A single RST transistor can perform an exclusive-OR function, because the collector current I_C flows if source and drain are at different logic values, regardless of which is "high."

Resonant Hot-Electron and Bipolar Transistors

As we have seen, three-terminal *RT* structures with direct modulation of the resonant current are difficult to fabricate. An alternative approach is the incorporation of a double-barrier *RT* potential into the emitter of a bipolar or a hot-electron transistor (63). A schematic band diagram of the resonant hot-electron transistor (*RHET*) is shown in Fig. 19. In a hot-electron transistor (64), carriers are injected from the emitter, traverse the base without scattering, and surmount the collector barrier V_C. If the base is narrow, a large fraction α_T of injected carriers arrives at the collector, giving a current gain $\beta = \alpha_T/(1 - \alpha_T)$. Of course, the R_BC time constant of Eq. (11) still applies, so there is a trade-off between high gain (narrow base) and fast operation (low R_B and hence a wider base)—state-of-the-art results on InGaAs devices have yielded $\beta \approx 10$ at room temperature (65). In the *RHET*, this current gain is combined with the resonant emitter $I_E(V_{BE})$. Consider the collector current $I_C(V_{BE})$ at some fixed V_{BC}. At small V_{BE} the emitter *RT* structure is below threshold, and $I_E \approx I_C$ is negligible. At larger V_{BE}, a resonant current flows through the emitter, injecting hot carriers into the base. Finally, as V_{BE} biases the emitter *RT* diode beyond V_P, the emitter current drops. The corresponding PVR in I_C will approximately reproduce the PVR of the emitter diode. Peak-to-valley ratios of approximately 10 have been reported in the $I_C(V_{BE})$ characteristics of *RHETs* at $T = 300$ K (65)

Similar characteristics can be obtained by inserting a double-barrier or cascaded *RT* diode in the emitter-base junction in an *n–p–n* bipolar transistor (66). Here emitter bias V_{BE} divides between the *RT* diode in the emitter and the emitter–base *n–P* junction to maintain current continuity. As long as V_{BE} is less than the *n–p* junction built-in voltage V_{bi}, I_E, increases as in a conventional bipolar transistor. Beyond flatband, $V_{BE} \geq V_{bi}$, additional V_{BE} drops over the *RT* diode, and I_E exhibits one or more *NDR* regions, which are reproduced in

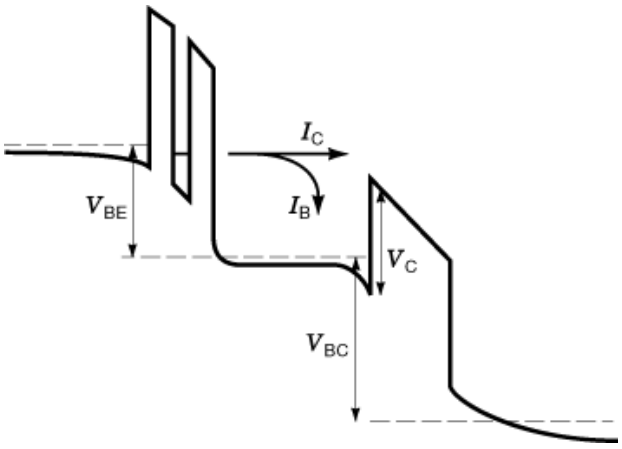

Fig. 19. Schematic band diagram of a resonant hot-electron transistor. The resonant $I(V)$ of the *RT* diode in the emitter–base junction is replicated in the collector current I_C, leading to a peaked $I_C(V_{BE})$ characteristic.

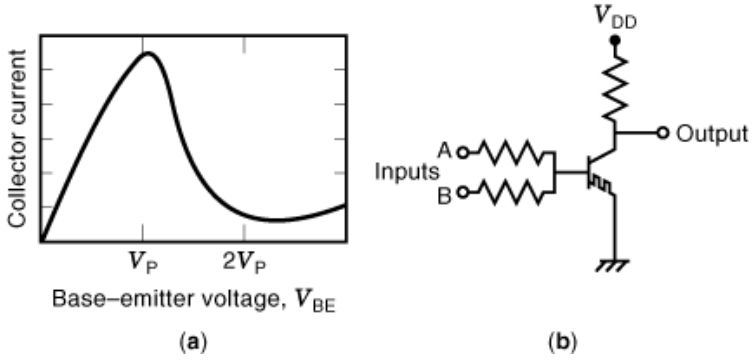

Fig. 20. (a) Schematic $I_C(V_{BE})$ of a resonant hot-electron transistor in the common-emitter configuration. The *RHET* exhibits negative transconductance for $V_{BE} > V_P$ with high PVR. (b) Exclusive-NOR circuit using a single *RHET*.

the I_C. The multipeak I_C characteristic of a bipolar transistor with two *RT* diodes in the emitter has been used as a frequency multiplier: driving the base with an ac signal of frequency f, signals at $3f$ (for sawtooth input) and $5f$ (for sinusoidal input) were generated with reasonable conversion efficiency (67).

Like RST transistors, resonant hot-electron and bipolar transistors exhibit higher logic functionality, illustrated schematically in Fig. 20. Given a common-emitter $I_C(V_{BE})$ characteristic with reasonable PVR, the output I_C can be high when $V_{BE} = V_{high} < V_P$, but low when $V_{BE} = 0$ or $2\ V_{high}$. Thus, a single device with two base contacts has exclusive- OR functionality, as shown in Fig. 20(b). Room temperature operation with reasonable V_{OUT} voltage swing has been demonstrated (65). In addition to the resistor network, a drawback of these designs is the finite PVR that causes power dissipation and reduced noise margin when both base inputs are high.

Similar logic functionality has been demonstrated in a silicon-based multiemitter floating-base bipolar transistor (68). A schematic diagram of an *n–p–n* device with two emitter contacts is shown in the inset of Fig. 21. The controlling base current is supplied by a reverse-biased emitter contact, where the current flows by interband tunneling, as in a backward diode. Consider $I_C(V_{E2}, V_C)$ when $V_{E1} = 0$ and V_C is biased high. Voltage V_{E2} forward-biases one of the emitter junctions and reverse-biases the other. A small tunneling current

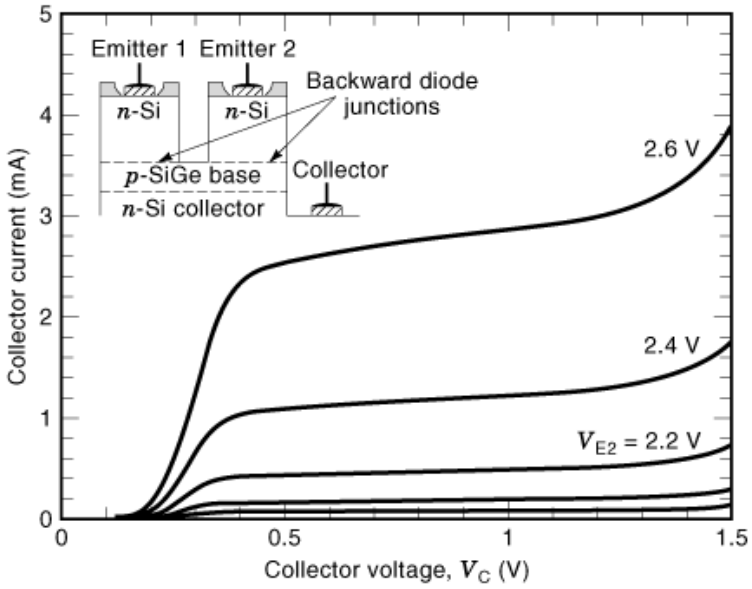

Fig. 21. Room temperature $I_C(V_{E2}, V_C)$ of a Si/SiGe multiemitter heterojunction bipolar transistor with a floating base, with inset showing a schematic device diagram. The biasing on the two emitters can be swapped without affecting the transistor characteristics.

flows in the reverse-biased junction, but this is precisely the base current needed to cause electron injection in the forward-biased emitter. The injected electrons reach the collector, and standard transistor operation, including high current gain, is preserved: room-temperature I_C V_{E2} V_C of an *n–p–n* Si/SiGe/Si floating-base transistor with gain $\beta \approx 400$ is shown in Fig. 1 (69). Since the emitter contacts are symmetric, a double emitter device possesses exclusive- OR functionality. I_C is large when one of the emitters is biased high and the other is grounded, but when both emitters are grounded or high $I_C \approx 0$. A further advantage is simplified fabrication, because no base contact metallization is required; a possible disadvantage is the higher effective base contact resistance—there is, at present, no reliable quantitative model of interband tunneling as a function of emitter–base junction doping.

Quantum Device Oscillators

An obvious application of any device with a strong *NDR* characteristic is an oscillator. Resonant tunneling diodes excel as solid-state high-frequency oscillators, because they are relatively easy to fabricate and exhibit reasonable output power with high maximum oscillation freencies f_{MAX} compared to competing microwave tunnel and transit-time diodes.

The inset of Fig. 22 shows a simple equivalent circuit of a two-terminal diode oscillator with an *NDR IV* characteristic, one that has been used for tunnel diodes. The real part of the equivalent circuit impedance R_{eq} is given by

$$R_{eq} = R_S + \frac{-R_D}{1 + (\omega R_D C_D)^2} \qquad (18)$$

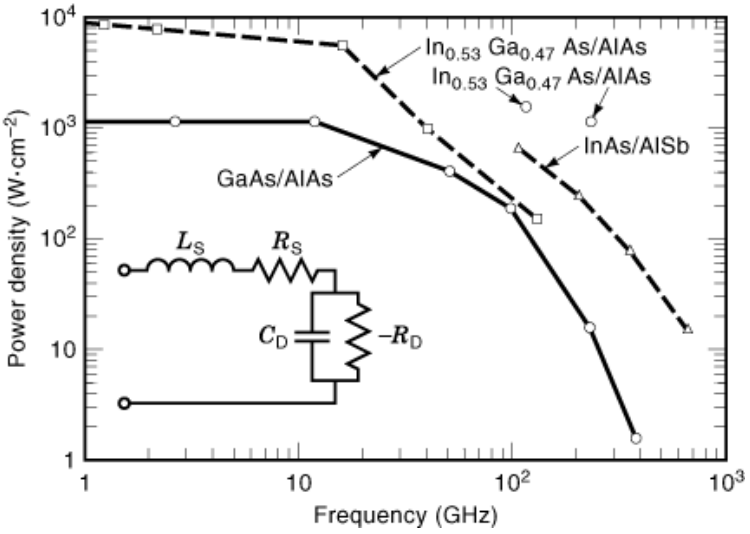

Fig. 22. Comparison of RT oscillators fabricated in different material systems. The InAs/AlSb oscillator should reach $f_{max} \approx 1$ THz. Inset shows the simple equivalent circuit of a two-terminal tunnel diode oscillator, including the parasitics. Graph courtesy of E. R. Brown, 1996.

where $-R_D = (V_V - V_P/I_V - I_P)$ is the negative diode resistance, C_D is the diode capacitance, and R_S is the series lead resistance. For steady-state oscillation, R_{eq} must be negative, so the cutoff frequency f_{MAX} is

$$f_{max} = \frac{1}{2\pi R_D C_D}\sqrt{\frac{R_D}{R_S} - 1} \qquad (19)$$

To increase f_{MAX}, series resistance R_S and diode capacitance C_D should be small.

Equations (16) and (17) have been employed in the design of RT diode oscillators with empirical parameters (e.g., taking for C_D the emitter-collector capacitance) and extended to include collector transit-time effects (15). However, the equivalent circuit of Fig. 22 is physically unsatisfactory. The current flowing in an RT diode depends on the alignment of the emitter and the 2-D subband in the well, with the tunneling currents into and out of the well balancing in steady state, $J_{IN} = J_{OUT}$. The main difficulty lies in the unknown energy distribution of the dynamically stored charge density σ_W, which makes it impossible to describe J_{OUT} as a unique function of the potential difference V_C between the well and the collector. It is possible to derive a small-signal dynamic model if one assumes that carriers equilibrate in the well (70). For realistic RT diodes, one finds that the key parameter for high speed is the lifetime of the metastable state, which should be minimized by making the collector barrier as transparent as possible while keeping the sharpness of the 2-D quantization sufficient for NDR in the $I(V)$ characteristic.

Figure 22 summarizes experimentally measured, room temperature oscillator performance of high-speed RT oscillators fabricated in different material systems: GaAs/AlAs, InGaAs/AlAs, and InAs/AlSb (15). While the power density in GaAs/AlAs devices is limited by their relatively low PVR at $T = 300$ K, In GaAs/AlAs RT oscillators exhibit good output power, while InAs/AlSb devices show promise for submillimeter wave ($f > 300$ GHz) performance and hold the record for solid-state oscillator frequency at 712 GHz (71). No other solid-state sources generate coherent power at submillimeter fundamental frequencies. One possible application of such devices is for low-noise local oscillators in radioastronomy.

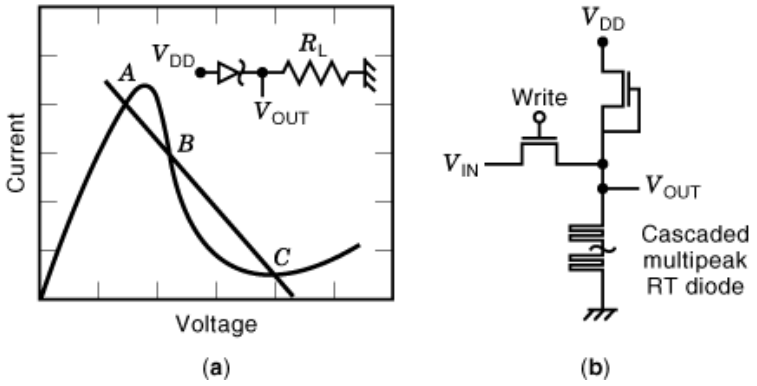

Fig. 23. (a) Schematic bistable memory made from an RT diode in series with a load resistor R_L. Output voltage $V_{OUT} = IR_L$ depends on whether the circuit is at point A or C ; point B is unstable. (b) Schematic diagram of a multistate memory constructed by biasing a cascaded RT device with a multipeak $I(V)$, like that in Fig. 12, with a constant current. After Seabaugh et al. (36) (© 1992 IEEE).

Quantum Device Memories

Several approaches in constructing memory circuits from quantum devices have been pursued. Single-device memories can be constructed from a bistable RT diode (12) or an ordinary RT diode in series with a load resistor R_L, which for proper R_L yields two stable bias points, as shown in Fig. 23(a). Voltage pulses can be used to change the memory state. The drawback of such memories is that at least one of the states corresponds to high current through the RT diode, resulting in prohibitive power dissipation. There have been attempts to increase the functionality of the RT memory in a cascaded RT structure with a multipeak $I(V)$ characteristic, as in Fig. 12, biased by a constant current source, as in Fig. 23(b). Given N peaks in the $I(V)$, the output node V_{OUT} can be at any of the$(N + 1)$ stable voltage points. Switching between V_{OUT} states is performed by setting an input voltage via a momentarily enabled write line. This type of multistate memory also dissipates a great deal of power, unless the PVR is very high. Ultimately, the quantifiable advantage of a multistate memory is the reduction of the number of elements necessary to store the same information by a factor of $\log_2(N + 1)$ for an $(N + 1)$-state device replacing a binary flip-flop.

A different approach is the series connection of two NDR devices, which can be RT diodes, RST transistors, or any other NDR device. If the total applied bias V_{DD} exceeds twice the peak voltage V_P, the voltage division between the devices becomes unstable. One of the two devices takes on most of the bias, thereby determining the voltage of the middle node V_{OUT}. This is illustrated by the load-line construction in Fig. 24: operating points A and C are stable, but B is unstable. Switching between the two states can be accomplished by changing the middle node bias via an additional electrode. The current flowing through the two NDR devices when $V_{DD} > 2V_P$ depends on the valley current. If the PVR of the devices is large, the current will be small regardless of whether the circuit is in state A or$C$.

A schematic memory constructed from two RT diodes in series with an additional control electrode separated from the middle node by a tunnel barrier is shown in Fig. 25(a). After their original demonstration at $T = 77$ K using InGaAs/AlAs/InP $RHETs$ (72), such devices have been fabricated in the polytype InAs/AlSb/GaSb system (73), which provides good PVR at $T = 300$ K. To switch V_{OUT}, a voltage V_{IN} is applied to the subcollector electrode, inducing a tunneling current between the middle node and the subcollector. When the subcollector current reaches I_P, V_{OUT} switches, yielding hysteresis in the V_{OUT} versus V_{IN} characteristic, shown schematically in Fig. 25(b).

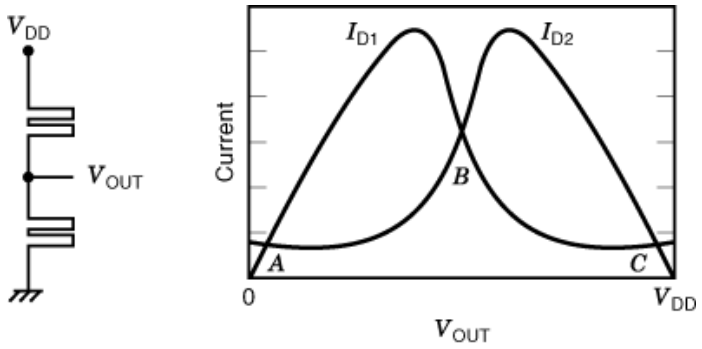

Fig. 24. Graphical construction for determining the operating points of a circuit consisting of two identical *NDR* devices in series. Points *A* and *C* are stable; point *B* is unstable.

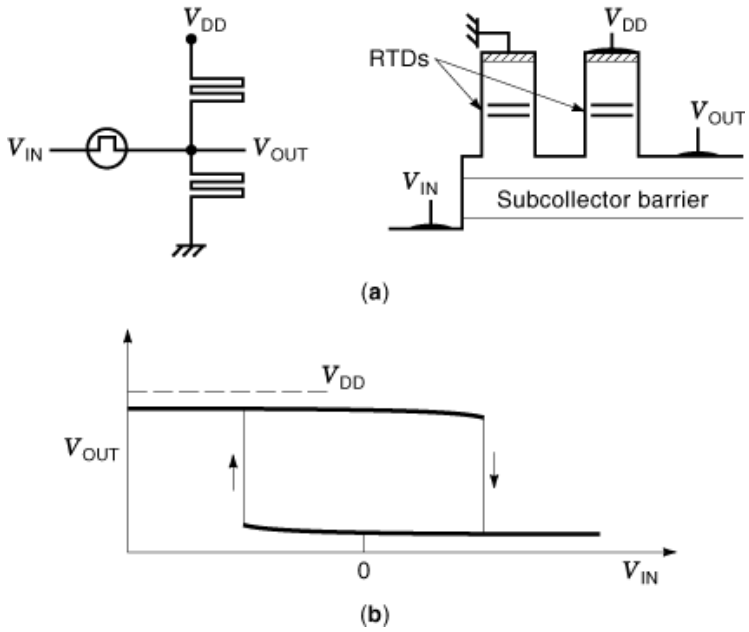

Fig. 25. (a) Circuit and cross-sectional diagrams of a memory based on two *RT* diodes in series with an additional subcollector terminal to control the voltage V_{OUT} of the middle node. After Shen et al. (73) (© 1995 IEEE). (b) Schematic input–output characteristics: V_{OUT} voltage swing (vertical extent of the loop) depends on the diode characteristics, increasing with PVR; V_{IN} switch points (horizontal extent of the loop) depend on the subcollector $I(V_{OUT}-V_{IN})$ two-terminal characteristic.

Recently, a tunneling-based random access memory cell combining two *RT* diodes in series with heterostructure FETs fabricated from the same epitaxial heterostructure grown on InP was reported to operate at room temperature (74). Despite poor PVR, this design achieved relatively low power consumption by employing *RT* diodes with very low peak current densities, while compensating for the low current drive of the diodes with an additional HFET.

Memory cells based on two *RT* diodes or *RHETs* in series, along the lines of Fig. 25, are smaller than standard CMOS designs. The remaining issue for large-scale memory arrays is power dissipation. Because a

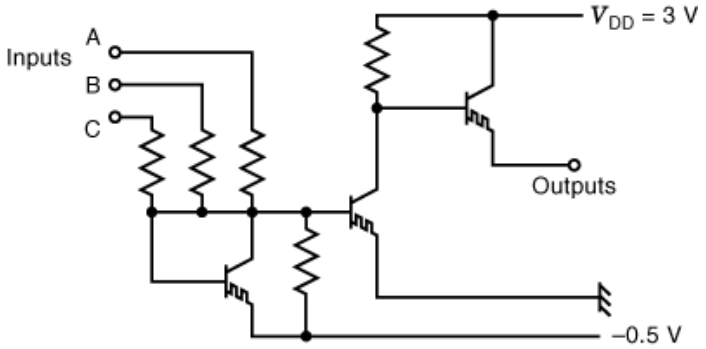

Fig. 26. A three-input majority logic gate implemented with three resonant hot-electron transistors and a resistor summing network. The last *RHET* before the output node is larger in area to increase I_C and, hence, the current drive of the logic gate. After Takatsu et al. (75) (© 1992 IEEE).

reasonable I_P is needed to charge up the interconnect capacitance (unless additional transistors are used (74), which increases the cell area), the relevant figure of merit is PVR. Polytype InAs/GaSb/AlSb *RT* diodes allowed for PVR of ∼20 at $T = 300$ K (73), but much higher PVR appears necessary for practical devices.

Quantum Device Logic

Quantum devices for logic elements have been proposed and, in some cases, demonstrated by a number of groups. In particular, the compact exclusive-NOR (XNOR) functionality of *RHETs* has been employed in the design of elementary logic components, such as latches and full adders (75). A typical building block in such designs is the three-input majority logic gate, shown in Fig. 26, which uses three *RHETs*. By using a four-resistor summing network connected to the emitter–base diode of the first *RHET*, the operating point lies below V_P in the $I_C(V_{BE})$ characteristic if none or one of the inputs is high and above V_P if two or three inputs are high. The second *RHET* senses whether the output of the first is above or below V_P. The third *RHET*, which is larger, increases the output current drive. By combining this majority logic gate with two XNOR gates made of two *RHETs* each, a full adder operating at $T = 77$ K was demonstrated (75). Room temperature operation of a hybrid full adder incorporating bipolar transistors with and without *RT* diodes in the emitter–base junction has also been reported (76). Such designs accomplish the required logic function with a reduced number of transistors, but at the expense of additional resistors. The impact of all these resistors on switching speed and propagation delay has not been characterized to date. Also, the integration of these circuits with conventional silicon technology is problematic, while the possibility of a stand-alone quantum device logic circuitry built in III–V semiconductors competing with the ever-advancing silicon CMOS logic is extremely remote.

Integration of high-functionality devices with conventional logic circuitry is considerably easier when they are built in Si/SiGe heterostructures. Both the Si/SiGe RST transistor of Fig. 18 and the multiemitter floating-base Si/SiGe HBT of Fig. 21 combine higher logic functionality with silicon technology compatibility. Further, in addition to the exclusive- OR function, these devices provide even higher logic functionality if the number of input terminals is increased. For example, three input terminals permit a single-device implementation of an ORNAND gate. Depending on whether the control input is high or low, the output current behaves as either a NAND or an OR function of the other two inputs (77). This added logic appears especially attractive for BiCMOS circuitry, where bipolar transistors are selectively added to CMOS logic blocks, typically to increase current drive. However, epitaxial deposition of pseudomorphic SiGe layers for the active regions obviously requires additional fabrication steps and reduces the thermal budget available for subsequent processing. The trade-off

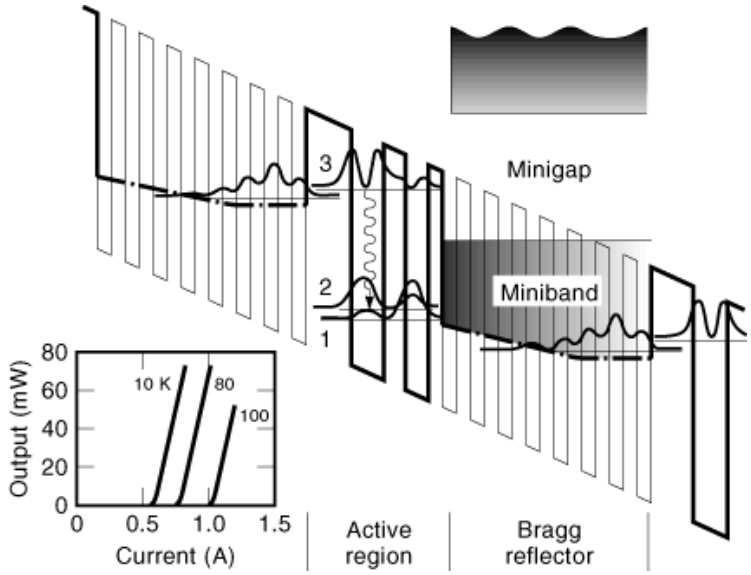

Fig. 27. Schematic conduction band diagram of the quantum cascade laser at a field $+\!+_z = 8.5 \times 10^4$ V/cm. The radiative $E_3 \rightarrow E_2$ transition in the coupled-quantum-well active region is shown by the wavy line. Bold lines indicate the squared moduli $|\chi(z)^2|$) of the subband wavefunctions in the active region. Note that the lower two states in the coupled quantum well, E_2 and E_1, line up with the SL miniband, whereas the E_3 state lines up with the SL minigap. The peak optical-power output from a single facet versus injection current for this laser at various heat sink temperatures T is shown at lower left (pulsed mode operation). At $T = 100$ K, the threshold current density $J_{\text{th}} = 3 \times 10^3$ A/cm^2. After Faist et al. (78).

between the added fabrication complexity and the area savings due to the higher functionality will decide the future of silicon-based quantum devices.

Optical Quantum Devices: Quantum Cascade and Quantum-Dot Lasers

A significant area where quantum devices are about to make their mark is solid-state laser sources in the mid-infrared ($\lambda = 4$ μm to 12 μm) and ultra-low-threshold quantum dot lasers.

The recently developed quantum cascade laser (QCL) (44,78) combines resonant tunneling and superlattice miniband spectrum in a device structure that makes full use of heterostructure bandgap engineering. Figure 27 shows a partial band diagram of the QCL together with its output characteristics. The entire QCL structure is composed of 25 stages of an InGaAs/AlInAs coupled-quantum-well active region followed by a doped superlattice reflector. The active region is designed for the following 2-D subband structure under bias: a higher-lying E_3 subband with a wavefunction $|\chi_3(z)|^2$ concentrated in the first well and two lower-lying subbands E_2 and E_1 concentrated in the first and second well, respectively. The radiative transition is $E_3 \rightarrow E_2$, so the laser output energy is $h\omega = E_3 - E_2$. This radiative transition has to compete with other $E_3 \rightarrow E_2$ relaxation mechanisms, mostly optical phonon emission. However, because $\omega \gg \omega_{\text{opt}}$, phonon-assisted $E_3 \rightarrow E_2$ relaxation requires large in-plane momentum transfer and is slow. On the other hand, since $\omega_{\text{opt}} \approx (E_2 - E_1)$, phonon-assisted $E_2 \rightarrow E_1$ relaxation is very fast. The superlattice downstream of the active region completes the set of conditions necessary for population inversion between E_3 and E_2, as it blocks direct tunneling out of the E_3 level but allows efficient tunneling out of E_1 into an SL miniband and then into the E_3 level in the

subsequent active region. There the process is repeated, until the electron cascades down all 25 stages and is collected in the doped optical cladding layers that sandwich the QCL.

Finally and crucially, the layers near the middle of the SL region are doped in the 10^{17} cm^{-3} range to provide carriers for injection into the active regions and ensure overall charge neutrality under operating conditions. The role of the SL regions is best appreciated by comparing the QCL structure of Fig. 27 with the conceptually similar structure of Fig. 14(c). In Fig. 14(c), a constant electric field in an undoped SL would be impossible to maintain in the presence of significant current, and tunneling from the E_2 level into the continuum would work against population inversion. The doped SL region maintains charge neutrality and prevents nonradiative tunneling out of the upper level. Note that since effective Bragg reflection requires very accurate grading of layer widths in the SL regions, this elegant approach places stringent demands on band structure modeling and epitaxial layer control.

The lasing characteristics shown in Fig. 27 at lower left corresponds to a $\lambda \approx 4.5$ μm laser with cleaved facets operated in pulsed mode, but continuous mode operation at $T = 140$ K and pulsed operation at room temperature has recently been reported in an optimized QCL structure (79). The power output is quite high, but the threshold current density J_{th} increases rapidly with temperature. Recent theory of gain in QCL pointed to the importance of hot-electron effects in the presence of subband nonparabolicity (80). Not only do electrons tunnel into E_3 with a considerable spread in energy of transverse motion, but those that relax nonradiatively to the E_2 subband are initially very hot, since $(\omega - \omega_{opt}) \approx 3000$ K for the $\lambda = 4.5$ μm transition. Because of nonparabolity, the gain depends on the difference between hot-electron distributions in these subbands. Calculations show that these distributions are radically different in the limits of low and high sheet-carrier concentrations n_D per QCL period. For low $n \ll 10^{11} cm^{-2}$, the rate of electron–electron collisions is low, the dominant scattering process is optical–phonon emission within the same subband, and the resulting electronic distribution decreases toward the bottom of the E_2 subband, as if the effective temperature were negative. In this regime, the calculated peak gain is substantial even at $T = 300$ K, but this regime has not yet been realized experimentally. Instead, QCL structures thus far have focused on the high n_D limit, where the peak gain is lower. Implementing the low-concentration regime appears to be a promising strategy for maximizing QCL performance.

Another interesting new development is the quantum dot (QD) laser, where fully confined semiconductor quantum dots placed in a matrix of a wider-bandgap semiconductor serve as the active medium (81). The advantages of a QD laser arise from the discrete, atomic-like density of states (see Fig. 3), which makes it easier to achieve population inversion, reducing the threshold current for lasing. Further, given sufficiently small dots with energy separation larger than kT at room temperature, QD lasers exhibit excellent high-temperature performance. Even though the proposals of temperature-insensitive QD lasers are not new (82), the technological fabrication of QD media with desired dot size, density, and size uniformity only became possible with advances in epitaxial self-assembly. It turns out, for carefully selected growth conditions, that the transition from planar to islanded growth of strained layers (such as InAs on GaAs) can be controlled to yield dots of remarkably consistent shape, size, and even lateral ordering. When these dots are overgrown with GaAs, the result is a dense QD array in a wider bandgap matrix. In these dots, the density of states is discrete, and electron-hole interactions are not screened, resulting in very high material gain compared to standard quantum-well lasers. As a result, lasers fabricated with such a QD active medium were shown to have the predicted low-threshold and temperature stability properties (83). Further, control over the QD size allows for some tuning of the lasing frequency, extending GaAs-based lasers to the technologically important 1.3 μm and 1.55 μm wavelengths.

Two basic device geometries have been applied to QD lasers: stripe geometry lasers with the light propagating along the QD plane, with the Fabry–Perot cavity formed by standard cleaved mirrors, and vertical cavity QD lasers with distributed Bragg reflector mirrors. The first approach allows the fabrication of high-power lasers with ultralow threshold current density, which also greatly improves device reliability by reducing dislocation growth and suppressing mirror overheating due to nonradiative surface recombination at the mirrors.

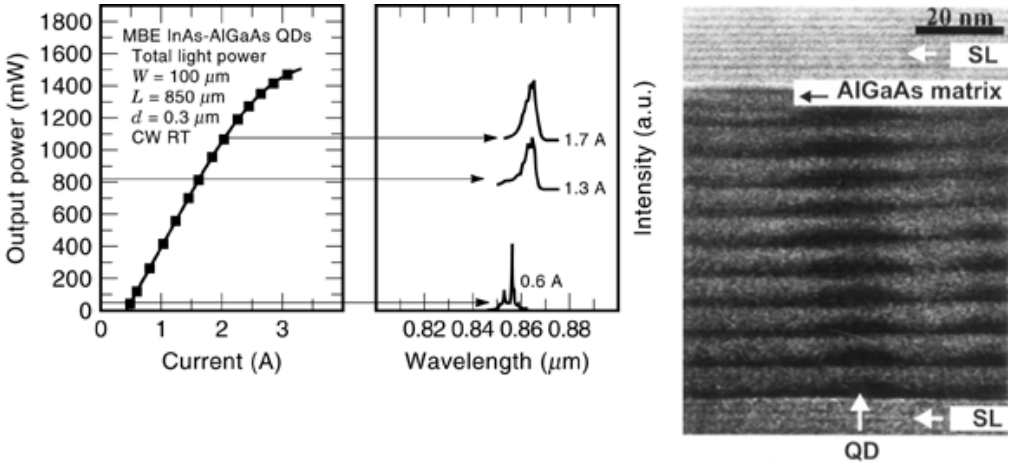

Fig. 28. High-power continuous-wave operation of an edge-emitting quantum dot laster at $T = 300$ K, together with a transmission electron micrograph of ordered InGaAs quantum dots in the active region. Figure courtesy of N. N. Ledentsov, 1998.

High-power, continuous wave, room temperature operation of a QD laser is shown in Fig. 28, together with a transmission electron micrograph of the QD active region (84). In the other geometry, light is emitted in the vertical direction, with the cavity formed by multilayer distributed Bragg reflector mirrors. Not only does this geometry promise lasers with ultralow total currents, but lasers based on single QD can be potentially realized. In both geometries, the ultrahigh gain available in the dots together with efficient QD carrier confinement relaxes the constraints on optical waveguiding, promising, for example, vertical cavity QD lasers even with moderate Bragg mirror reflectivity. The absence of exciton heating and screening, as well as much reduced phonon-asisted nonradiative relaxation in quantum dots (85), distinguishes QD laser physics from conventional semiconductor lasers sufficiently to warrant its discussion as a new quantum device, rather than an improved conventional laser.

Future Of Quantum Devices

This article has reviewed some of the recent research in the area of quantum devices. It is clear that even though many of these devices are quite successful according to some benchmarks, none has found large-scale commercial application to date. To be sure, quantum device research has achieved much progress. Fascinating new physics has been discovered, exemplified by the fractional quantum Hall effect (86). The basic effects relevant to electronic devices, such as tunneling in heterostructures, ballistic transport, and charge injection across potential barriers are now available as robust and reproducible phenomena. Yet only in the relatively distant future—at very small device dimensions L, cryogenic temperatures T, or whatever other design criteria future technology may require—might quantum devices offer a sufficient advantage for mainstream analog or digital electronics. On the other hand, they have significant potential if they can be integrated with mainstream semiconductor devices or, in the case of quantum dot lasers, if they can improve the performance of an existing device without incurring major fabrication overhead. Finally, there are niche applications where quantum devices appear ready to take over; a classic example is the quantum cascade laser, which promises a source in the range of wavelengths where the competition—lead-salt lasers—is relatively inefficient.

Evolution in microelectronics has been associated with the progress in two areas: miniaturization of devices driven by advances in lithography and ion implantation doping, and bandgap engineering made possible by modern epitaxy. Of these two areas, the first definitely has had a greater impact in the commercial arena, whereas the second has been supplying the device physics field with new systems to explore. These roles may well be reversed in the future. Development of new and exotic lithographic techniques with nanometer resolution will set the stage for the exploration of various physical effects in mesoscopic devices, whereas epitaxially grown devices, particularly heterojunction transistors integrated with optoelectronic elements, will be gaining commercial ground. When and whether this role reversal will take place will be determined perhaps as much by economic as by technical factors.

The logic of industrial evolution will provide new paths for a qualitative improvement of system components, other than the traditional path of a steady reduction in fine-line feature size. Miniaturization progress faces diminishing returns in the future, when the speeds of integrated circuits and the device packing densities will be limited primarily by the delays and power dissipation in the interconnects rather than individual transistors. Further progress may then require circuit operation at cryogenic temperatures or heavy reliance on high-bandwidth optical and electronic interconnects. Implementation of optical interconnects within the context of silicon microelectronics requires hybrid-material systems with islands of foreign heterostructures grown or grafted on Si substrates. In this scenario, the current noncompetitiveness of quantum devices could give way to novel devices serving as small, highly functional application-specific components that add significant value to main blocks of microelectronic circuitry.

BIBLIOGRAPHY

1. S. Luryi J. Xu A. Zaslavsky (eds.) *Future Trends in Microelectronics: Off the Beaten Path*, New York: Wiley-Interscience, 1999.
2. C. Cohen-Tannoudji B. Diu F. Laloë *Quantum Mechanics*, vol. **1**, Chapter I, New York: Wiley-Interscience, 1977.
3. B. Ricco M. Ya. Azbel Physics of resonant tunneling: The one-dimensional double-barrier case, *Phys. Rev. B*, **29**: 1970, 1984.
4. S. Luryi A. Zaslavsky Quantum-effect and hot-electron devices, in S. Sze (ed), *Modern Semiconductor Device Physics*,New York: Wiley-Interscience, 1998, pp. 253-342.
5. S. V. Meshkov Tunneling of electrons from a two-dimensional channel into the bulk, *Zh. Eksp. Teor. Fiz*, **91**: 2252, 1986, *] Sov. Phys. JETP*, **64**: 1337, 1986].
6. S. Luryi Frequency limit of double-barrier resonant-tunneling oscillators, *Appl. Phys. Lett.*, **47**: 490, 1985.
7. L. L. Chang L. Esaki R. Tsu Resonant tunneling in semiconductor double barriers, *Appl. Phys. Lett.*, **24**: 593, 1974.
8. A. Zaslavsky, *et al.* Magnetotunneling in double-barrier heterostructures, *Phys. Rev. B*, **40**: 9829, 1989.
9. V. J. Goldman D. C. Tsui J. E. Cunningham Evidence for LO-phonon-emission-assisted tunneling in double-barrier heterostructures, *Phys. Rev. B*, **36**: 7635, 1987.
10. N. S. Wingreen K. W. Jacobsen J. W. Wilkins Resonant tunneling with electron-phonon interaction: An exactly solvable model, *Phys. Rev. Lett.*, **61**: 1396, 1988; F. Chevoir and B. Vinter, Calculation of phonon-assisted tunneling and valley current in a double-barrier diode, *Appl. Phys. Lett.*, **55**: 1859, 1989.
11. S. K. Diamond *et al.* Fabrication of 200-GHz $f_{\max}$ resonant-tunneling diodes for integrated circuit and microwave applications, *IEEE Electron. Dev. Lett.*, **10**: 104, 1989.
12. V. J. Goldman D. C. Tsui J. E. Cunnigham Observation of intrinsic bistability in resonant-tunneling structures, *Phys. Rev. Lett.*, **58**: 1256, 1987.
13. E. Ozbay D. M. Bloom 110-GHz monlithic resonant-tunneling-diode trigger circuit, *IEEE Electron. Dev. Lett.*, **12**: 480, 1991.
14. E. R. Brown *et al.* A quasioptical resonant-tunneling-diode oscillator operating above 200 GHz, *IEEE Trans. Microw. Theory. Tech.*, **41**: 720, 1993.
15. E. R. Brown High-speed resonant-tunneling diodes, in N. G. Einspruch and W. R. Frensley (eds), *Heterostructures and Quantum Devices*, San Diego: Academic Press, 1994.

16. M. Sweeny J. Xu Resonant interband tunneling diodes, *Appl. Phys. Lett.*, **54**: 546, 1989; J. R. Söderstrom, D. H. Chow, and T. C. McGill, New negative differential resistance device based on resonant interband tunneling, *Appl. Phys. Lett.*, **55**: 1094, 1989.
17. E. E. Mendez J. Nocera W. I. Wang Conservation of momentum, and its consequences, in interband resonant tunneling, *Phys. Rev. B*, **45**: 3910, 1992.
18. R. Beresford *et al.* Resonant interband tunneling through a 110 nm InAs quantum well, *Appl. Phys. Lett.*, **56**: 551, 1990.
19. H. C. Liu *et al.* Resonant tunneling in $Si/Si_{1-x}Ge_x$ double-barrier structures, *Appl. Phys. Lett.*, **52**: 1809, 1988.
20. G. Schuberth *et al.* Resonant tunneling of holes in Si/Si_{1-x} Ge_x quantum-well structures, *Phys. Rev. B*, **43**: 2280, 1991.
21. U. Gennser *et al.* Probing band structure anisotropy in quantum wells via magnetotunneling, *Phys. Rev. Lett.*, **67**: 3828, 1991.
22. A. Zaslavsky *et al.* Strain relaxation in silicon-germanium microstructures observed by resonant tunneling spectroscopy, *Appl. Phys. Lett.*, **67**: 3921, 1995; C. D. Akyüz, et al., Inhomogeneous strain in individual quantum dots probed by transport measurements, *Appl. Phys. Lett.*, **72**: 1739, 1998.
23. S. Y. Chou *et al.* Observation of electron resonant tunneling in a lateral dual-gate resonant tunneling field-effect transistor, *Appl. Phys. Lett.*, **55**: 176, 1989; K. Ismail, D. A. Antoniadis, and H. I. Smith, Lateral resonant tunneling in a double- barrier field effect transistor, *Appl. Phys. Lett.*, **55**: 589, 1989.
24. M. A. Reed, *et al.* Realization of a three-terminal resonant tunneling device: The bipolar quantum resonant tunneling transistor, *Appl. Phys. Lett.*, **54**: 1034, 1989.
25. C. J. Goodings, *et al.* Variable-area resonant tunneling diodes using implanted in-plane gates, *J. Appl. Phys.*, **76**: 1276, 1994.
26. T. K. Woodward T. C. McGill R. D. Burnham Experimental realization of a resonant tunneling transistor, *Appl. Phys. Lett.*, **50**: 451, 1987.
27. M. Dellow Gated resonant tunneling devices, *Electron. Lett.*, **27**: 134, 1991; P. Guéret, et al., Confinement and single-electron tunneling in Schottky-gated, laterally squeezed double-barrier quantum-well heterostructures, *Phys. Rev. Lett.*, **68**: 1896, 1992.
28. V. R. Kolagunta *et al.* Self-aligned sidewall-gated resonant tunneling transistors, *Appl. Phys. Lett.*, **69**: 374, 1996.
29. J. A. Simmons *et al.* Unipolar complementary bistable memories using gate-controlled negative differential resistance in a 2D-2D quantum tunneling transistor, *1997 Tech. Dig. IEEE Int. Electron. Devices Meeting*, Washington, DC, 1997, p. 755.
30. J. P. Eisenstein L. N. Pfeiffer K. W. West Field-induced resonant tunneling between parallel 2-dimensional electron-systems, *Appl. Phys. Lett.*, **58**: 1497, 1991.
31. J. S. Moon, *et al.* Unipolar complementary circuits using double electron layer tunneling transistors, *Appl. Phys. Lett.*, **74**: 314, 1999.
32. S. Luryi F. Capasso Resonant tunneling of two dimensional electrons through a quantum wire, *Appl. Phys. Lett.*, **47**: 1347, 1985; erratum ibid., **48**: 1693, 1986.
33. L. N. Pfeiffer *et al.* Formation of a high-quality two-dimen- sional electron gas on cleaved GaAs, *Appl. Phys. Lett.*, **56**: 1697, 1990.
34. C. Kurdak *et al.* Surface resonant tunneling transistor: A new negative transconductance device, *Appl. Phys. Lett.*, **64**: 610, 1994.
35. F. Capasso K. Mohammed A. Y. Cho Resonant tunneling through double barriers, perpendicular quantum transport phenomena in superlattices, and their device applications, *IEEE J. Quantum Electron.*, **QE-22**: 1853, 1986.
36. A. C. Seabaugh Y. C. Kao H. T. Yuan Nine-state reso- nant tunneling diode memory, *IEEE Electron. Dev. Lett.*, **13**: 479, 1992.
37. For a compelete discussion see G. Bastard *Wave Mechanics Applied to Semiconductor Heterostructures*, Chapter I, New York: Wiley, 1988.
38. L. Esaki R. Tsu Superlattice and negative differential conductivity in semiconductors, *IBM J. Res. Develop.*, **14**: 61, 1970.
39. L. Esaki L. L. Chang New transport phenomenon in a semiconductor "superlattice," *Phys. Rev. Lett.*, **33**: 495, 1974; K. K. Choi et al., Periodic negative conductance by sequential resonant tunneling through an expanding high-field superlattice domain, *Phys. Rev. B*, **35**: 4172, 1987.

40. A. Sibille *et al.* Observation of Esaki-Tsu negative differential velocity in GaAs/AlAs superlattices, *Phys. Rev. Lett.*, **64**: 52, 1990; H. T. Grahn Electrical transport in narrow-miniband semiconductor superlattices, *Phys. Rev. B*, **43**: 12094, 1991.
41. H. M. James Electronic states in perturbed periodic systems, *Phys. Rev.*, **76**: 1611, 1949.
42. The extent of Wannier-Stark wavefunctions over a finite number of SL periods allows for their observation by photocurrent measurements, see E. E. Mendez, F. Agulló-Rueda, and J. M. Hong, Stark localization in GaAs-GaAlAs superlattices under an electric field, *Phys. Rev. Lett.*, **60**: 2426, 1988.
43. R. Kazarinov R. Suris Possibility of the amplification of electromagnetic waves in a semiconductor with a superlattice, *Fiz. Tekh. Poluprovodn.*, **5**: 797, 1971 *[Sov. Phys. Semicond.*, **5**: 707, 1971.]
44. J. Faist *et al.* Quantum cascade laser, *Science*, **264**: 533, 1994.
45. H. C. Liu G. C. Ayers Resonant tunneling through one-, two-, and three-dimensionally confined quantum wells, *J. Appl. Phys.*, **65**: 4908, 1989.
46. B. Su V. J. Goldman J. E. Cunningham Observation of single-electron charging in double-barrier heterostructures, *Science*, **255**: 313, 1992; Single-electron tunneling in nanometer-scale double-barrier heterostructure devices, Phys. Rev. B, **46**: 7644, 1992.
47. T. Schmidt *et al.* Quantum-dot ground states in a magnetic field studied by single-electron tunneling spectroscopy on double-barrier heterostructures, *Phys. Rev. B*, **51**: 5570, 1995.
48. U. Meirav M. A. Kastner S. J. Wind Single-electron charging and periodic conductance resonances in GaAs nanostructures, *Phys. Rev. Lett.*, **65**: 771, 1990; L. P. Kouwenhoven et al., Single electron charging effects In semiconductor quantum dots, Z. Phys. B., **85**: 367, 1991.
49. An extensive discussion is available in H. Grabert, M. H. Devoret, (eds.) *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures*, New York: Plenum Press, 1992. The chapter by D. V. Averin and K. K. Likharev is devoted to device applications.
50. K. Yano *et al.* Room-Temperature single-electron memory, *IEEE Trans. Electron. Devices I*, **41**: 1628, 1994; K. Yano et al., 128 Mb early prototype for gigascale single-electron memories, *IEEE Int. Solid-State Circuits Conf.*, 1998, pp. 344–345.
51. A. Nakajima *et al.* Room temperature operation of Si single-electron memory with self-aligned floating dot gate, *Appl. Phys. Lett.*, **70**: 1742, 1997.
52. T. A. Fulton G. J. Dolan Observation of single-electron charging effects in small junctions, *Phys. Rev. Lett.*, **59**: 109, 1987.
53. G. Zimmerli R. L. Kautz J. M. Martinis Voltage gain in the single-electron transistor, *Appl. Phys. Lett.*, **61**: 2616, 1992.
54. L. P. Kouwenhoven *et al.* Quantized current in a quantum-dot turnstile using oscillating tunnel barriers, *Phys. Rev. Lett.*, **67**: 1626, 1991.
55. H. Pothier *et al.* Single-electron pump based on charging effects, *Europhys. Lett.*, **17**: 249, 1992; J. M. Martinis, M. Nahum, and H. D. Jensen, Metrological accuracy of the electron pump, *Phys. Rev. Lett.*, **72**: 904, 1994.
56. Z. S. Gribnikov Negative differential conductivity in a multilayer heterostructure, *Fiz. Tekh. Poluprovodn.*, **6**: 1380, 1972 *[Sov. Phys. Semicond.*, **6**: 1204, 1973].
57. M. Keever *et al.* Measurements of hot-electron conduction and real-space transfer in GaAs/Al_x Ga_{1-x} As heterojunction layers, *Appl. Phys. Lett.*, **38**: 36, 1981.
58. K. Hess *et al.* Negative differential resistance through real-space electron transfer, *Appl. Phys. Lett.*, **35**: 469, 1979.
59. A. Kastalsky S. Luryi Novel real-space hot-electron transfer devices, *IEEE Electron. Device Lett.*, **4**: 334, 1983; S. Luryi et al., Charge injection transistor based on real-space hot-electron transfer, *IEEE Trans. Electron. Devices* **31**: 832, 1984.
60. A. Kastalsky *et al.* Negative-resistance field-effect transistor grown by organometallic chemical vapor deposition, *Solid State Electron.*, **29**: 1073, 1986.
61. C. L. Wu *et al.* Very strong negative differential resistance real-space transfer transistor using a mulitple δ -doping GaAs/InGaAs pseudomorphic heterostructure, *Appl. Phys. Lett.*, **66**: 739, 1995.
62. M. Mastrapasqua *et al.* Charge injection transistor and logic elements in Si/Si_{1-x} Ge_x heterostructures, in S. Luryi, J. Xu, and A. Zaslavsky (eds.), *Future Trends in Microelectronics: Reflections on the Road to Nanotechnology*, Dordrecht: Kluwer, 1996, pp. 377–384.

63. N. Yokoyama *et al.* A new functional resonant-tunneling hot electron transistor (*RHET*), *Jn. J. Appl. Phys.*, **24**: L-853, 1985; N. Yokoyama et al., Resonant-tunneling hot electron transistor (*RHET*), *Solid State Electron.*, **31**: 577, 1988.
64. M. Heiblum I. Anderson C. M. Knoedler DC performance of ballistic tunneling hot-electron-transfer amplifiers, *Appl. Phys. Lett.*, **49**: 207, 1986.
65. T. S. Moise *et al.* Room-temperature operation of a resonant-tunneling hot-electron transistor based integrated circuit, *IEEE Electron. Dev. Lett.*, **14**: 441, 1993.
66. For a review of resonant tunneling bipolar transistor research see F. Capasso, S. Sen, and F. Beltram, Quantum-effect devices, in S. M. Sze (ed.), *High-Speed Semiconductor Devices*, New York: Wiley, 1990, pp. 465–520.
67. S. Sen *et al.* Multiple state resonant tunneling bipolar transistor operating at room temperature and its application as a frequency multiplier, *IEEE Electron. Dev. Lett.*, **9**: 533, 1988.
68. Z. S. Gribnikov S. Luryi Article comprising a bipolar transistor with a floating base, US patent 5,461,245, filed August, 1994; K. Imamura et al., Proposal and demonstration of multi-emitter HBT's, *Electron. Lett.*, **30**: 459, 1994.
69. A. Zaslavsky *et al.* Multiemitter Si/Ge_x Si_{1-x} heterojunction bipolar transistor with no base contact and enhanced logic functionality, *IEEE Electron. Dev. Lett.*, **18**: 453, 1997.
70. This model is developed in S. Sze (ed.), *Modern Semiconductor Device Physics*, New York: Wiley-Interscience, 1998, pp. 306–310.
71. E. R. Brown *et al.* Oscillations up to 712 GHz in InAs/AlSb resonant-tunneling diodes, *Appl. Phys. Lett.*, **58**: 2291, 1991.
72. T. Mori *et al.* A static random access memory cell using a double-emitter resonant-tunneling hot electron transistor for gigabit-plus memory applications, *Jpn. J. Appl. Phys.*, **33**: 790, 1994.
73. J. Shen *et al.* Static random access memories based on resonant interband tunneling diodes in the InAs/GaSb/AlSb material system, *IEEE Electron. Dev. Lett.*, **16**: 178, 1995.
74. J. P. A. van der Wagt A. C. Seabaugh E. A. Beam, III RTD/HFET low standby power SRAM gain cell, *IEEE Electron. Dev. Lett.*, **19**: 7, 1998.
75. M. Takatsu *et al.* Logic circuits using resonant-tunneling hot-electron transistors (*RHET*'s), *IEEE J. Solid-State Circuits*, **27**: 1428, 1992; N. Yokoyama et al., Resonant hot electron transistors, in: J. Shah, (ed.), *Hot Carriers in Semiconductor Nanostructures: Physics and Applications*, San Diego: Academic Press, 1992, pp. 443–467.
76. A. C. Seabaugh M. A. Reed Resonant-tunneling transistors, in N. G. Einspruch and W. R. Frensley (eds.), *Heterostructures and Quantum Devices*, San Diego: Academic Press, 1994.
77. S. Luryi *et al.* Charge injection logic, *Appl. Phys. Lett.*, **57**: 1787, 1990.
78. J. Faist *et al.* Vertical transition quantum cascade laser with Bragg confined excited state, *Appl. Phys. Lett.*, **66**: 538, 1995.
79. J. Faist *et al.* High power mid-infrared ($\lambda \sim 5$ μm) quantum cascade lasers operating above room temperature, *Appl. Phys. Lett.*, **68**: 3680, 1996; C. Gmachl et al., Continuous-wave and high-power pulsed operation of index-coupled distributed feedback quantum cascade laser at $\lambda \approx 8.5$ μm, *Appl. Phys. Lett.*, **72**: 1430, 1998.
80. V. B. Gorfinkel S. Luryi B. Gelmont Theory of gain spectra for quantum cascade lasers and temperature dependence of their characteristics at low and moderate carrier concentrations: *IEEE J. Quantum Electron.*, **32**: 1995, 1996.
81. N. N. Ledentsov *et al.* Three-dimensional arrays of self-ordered quantum dots for laser applications: *Microelectron. J.*, **28**: 915, 1997.
82. Y. Arakawa H. Sakaki *Appl. Phys. Lett.*, **40**: 939, 1982.
83. N. Kirstaedter *et al.* Low threshold, large T_0 injection laser emission from (InGa)As quantum dots, *Electron. Lett.*, **30**: 1416, 1994.
84. M. V. Maximov *et al.* High-power continuous-wave operation of a InGaAs/AlGaAs quantum dot laser, *J. Appl. Phys.*, **83**: 5561, 1998.
85. H. Benisty C. M. Sotomayor Torres C. Weisbuch Intrinsic mechanism for the poor luminescence properties of quantum-box systems, *Phys. Rev. B*, **44**: 10945, 1991.
86. D. C. Tsui H. L. Störmer A. C. Gossard Two-dimensional magnetotransport in the extreme quantum limit, *Phys. Rev. Lett.*, **48**: 1559, 1982.

A. ZASLAVSKY
Brown University