Resonant Tunneling and Its Prospects in Si/ZnS and Si/CaF₂ Heterostructures

by
Robert M. Steinhoff

Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degrees of Bachelor of Science in Electrical Science and Engineering and Master of Engineering in Electrical Engineering and Computer Science at the Massachusetts Institute of Technology
Feb 7, 1997
Copyright © 1997 by Robert M. Steinhoff. All rights reserved.

The author hereby grants to M.I.T. permission to reproduce distribute publicly paper and electronic copies of this thesis and to grant others the right to do so.

2/6/97

Author

Department of Electrical Engineering and Computer Science
Feb 7, 1997

2/6/97

Certified by
Qing Hu
Thesis Supervisor

2/6/97

Accepted by
F. R. Morgenthaler
Chairman, Department Committee on Graduate Theses

MAR 2 1 1997
Resonant Tunneling and Its Prospects in Si/ZnS and Si/CaF₂ Heterostructures

by

Robert M. Steinhoff

Submitted to the
Department of Electrical Engineering and Computer Science

Feb 7, 1997

In Partial Fulfillment of the Requirements for the Degrees of Bachelor of Science in Electrical Science and Engineering and Master of Engineering in Electrical Engineering and Computer Science

Abstract:

The theory of tunneling and resonant tunneling is reviewed. Experimental justification of this theory was provided by an InP-based RTD with room-temperature PVR ≈ 12. Its valley current was well-modeled as the sum of thermionic emission through the second resonance and a temperature-independent scattering/defect component. ZnS and CaF₂ were experimentally considered as epitaxial barrier materials for a Si-based RTD. Temperature-dependent J-V measurements were utilized to determine thermal barrier heights to ZnS. The ZnS/As(1ML)/Si(100) conduction band edge offset and the Al/ZnS(100) barrier height were found to be 1.1±0.1 and 1.02±0.04 eV, respectively. Al/SiO₂/a-Si/ZnS/As(1ML)/Si(100) double-barrier devices were fabricated but did not show NDR. Although post-oxidation RTA was performed on this structure, TEM revealed that the Si quantum well did not crystallize. Large current densities through single-barrier Si/CaF₂/Si(111) samples and TEM of these samples demonstrated the lack of a quality CaF₂ barrier. A working RTD was not fabricated in either of these Si-based materials systems.

Thesis Supervisor: Qing Hu
Title: Associate Professor of Electrical Engineering and Computer Science
Table of Contents

1 Introduction...................................................................................................... 4

2 Theoretical Background...................................................................................... 5
   2.1 Globally Coherent Transport................................................................. 6
       2.1.1 Single-Barrier Devices................................................................. 8
       2.1.2 Double-Barrier Devices............................................................... 11
   2.2 Sequential Tunneling.............................................................................. 13
   2.3 Resonance Broadening.......................................................................... 15
   2.4 RTD Circuit Models and Instabilities................................................... 18

3 InP-Based RTD.................................................................................................. 20

4 Si-Based Materials........................................................................................... 25
   4.1 ZnS......................................................................................................... 27
   4.2 CaF₂...................................................................................................... 38

5 Al/SiO₂/a-Si/ZnS/As(1ML)/Si Double-Barrier Device.................................... 43

6 Conclusions and Future Work........................................................................ 45

Acknowledgments.............................................................................................. 47

Appendix: Semiclassical Space Charge Theory.............................................. 48

References......................................................................................................... 50
1 Introduction

Recent innovations in molecular beam epitaxy (MBE) and scanning probe technology allow for the growth of atomically-precise heterostructures and hence of electrical devices with dimensions comparable to the electron wavelength (~ nm). Interesting quantum interference effects naturally arise in such small devices. In particular, the resonant tunneling diode (RTD), first proposed and demonstrated (at 77 K) by Chang, Esaki, and Tsu [1] in 1974, relies upon the wave nature of the electron to pass a particular band of resonant energies, exactly analogous to the way a half-wave optical filter is used to pass light of a particular band of frequencies.

This energy filtration property of the RTD gives it a current-voltage (J-V) characteristic in which the current first increases (as the applied voltage increases from zero) to a “peak”, and then decreases to a minimum in the “valley”. Further peaks may be observed at higher resonances depending upon the structure. A figure of merit for an RTD is its “peak-to-valley” current ratio (PVR) which has been measured at room temperature as high as 52 [2]. The RTD’s signature J-V curve is one of the few macroscopic phenomena which require the wavelike properties of matter to explain, and resonant tunneling is therefore an extremely exciting and fast-paced topic of fundamental physics research.

Aside from its educational benefits, there are many practical applications for the RTD. Its negative differential-resistance (NDR) allows for the compact production of high-frequency (> 100 GHz) oscillators and fast-switching (< 5 ps) bistable devices. In comparison to Esaki tunnel diodes (p'/n+ junctions which also display NDR), RTDs, with their lower capacitance and higher peak current densities, operate roughly 50-100x faster [3]. Numerous applications for RTDs in analog-to-digital conversion (ADC) circuits, high-speed/high-density static random-access memories (SRAMs), and digital arithmetic logic units (ALUs) are currently being explored.

At the moment, RTDs showing reproducible and significant NDR have been manufactured only with compound semiconductors such as GaAs and InP, not with Si. This absence is attributable to a variety of factors dominated by an immature Si-based heterojunction technology and the large Si electron (and hole) effective mass. The lack of
a marketable Si-based RTD\(^1\) is unfortunate because Si, due to its economical advantages over compound semiconductors, is presently the material of choice for the integrated-circuit (IC) industry. The advent of a manufacturable RTD on Si would indeed be a technological breakthrough, as it would allow for easy and economical integration of RTDs into the current IC production processes.

There are many Si lattice-matched materials available with which to build an RTD. These include ZnS, ZnMgS, ZnSSe, GaAsN, InAsN, AlAsN, GaPN, InPN, AlPN, CaF\(_2\), SiGeC, CoSi\(_2\), CeO\(_2\) and other rare earth metal oxides. Chemical bonding at the interfaces between these materials and Si, however, is not well understood and few observations of silicon-based resonant tunneling have been reported. Triple-barrier devices consisting of Si emitter, CaF\(_2\) barriers, CoSi\(_2\) wells, and CoSi\(_2\) collector (PVR \(\approx 2\) at room temperature) [4], and double-barrier devices consisting of Si barriers with SiGe emitter, collector, and well (1.2 at 77K) [5, 6], are the extent of previously-reported Si-based RTDs.

In this dissertation, we will consider the use of CaF\(_2\) and ZnS as potential barrier materials for a Si-based RTD, whose advent will be motivated by experimental evidence of resonant tunneling in an InP-based RTD. First, however, the principles of the RTD must be reviewed. The following section is intended to stand alone as a theoretical review of tunneling and resonant tunneling, and much of the discussion therein is not necessary for the subsequent experimental analysis.

\section{Theoretical Background}

The solid-state theorist's goal, as it relates to electrical engineering, is to predict the steady-state charge distribution and current densities within a device when external voltages are applied to it. To isolate the device from its surroundings (e.g., batteries and leads), these external voltages are assumed to establish Fermi levels in fully-thermalized electron reservoirs coupled to the device considered. The full many-body Hamiltonian

\[^1\] A "Si-based" RTD refers to one which may be fabricated upon a Si substrate without a lattice-transforming buffer layer.
for the device, including scattering, is far too difficult to work with and a number of simplifications must be made to make the problem tractable.

### 2.1 Globally Coherent Transport

The simplest practical model one can assume neglects scattering processes entirely. All electron-electron interactions, as well as certain correlation effects such as image-force barrier lowering, are then approximated by a single-electron Hartree potential (exchange effects can be partially accounted for by using the full Hartree-Fock Hamiltonian or the simplifications described in the introduction to [7]). That is, we self-consistently treat non-interacting electrons as if they moved in the potential created by their own expected charge distribution. Then, we may solve the “globally coherent” scattering problem for a single electron emitted at a particular energy from a particular reservoir and superpose the effects of all electrons weighted by their emission rate from the reservoirs. This is the standard first-order approach for analyzing microscopic devices and is described more fully in any solid-state device text [8, 9].

For analyzing devices which are inherently 3-dimensional, transverse (perpendicular to the direction of transport) translational symmetry is typically assumed throughout the device (i.e., the transverse lattice period is the same in all materials). This forces the transverse Bloch wavevector $k_T$ to be conserved throughout the device, and allows one to calculate the transmission probability $T_e$ (subscript denotes “emitter”), for a given $k_T$, one-dimensionally (in the longitudinal, z, direction). Further, it is often assumed that, in the energy range of interest, the transmission probability is independent of $k_T$ so that the sum over the transverse modes may be performed immediately in the effective mass approximation. Using the Fermi-Dirac distribution with temperature $T$ (to be distinguished from transmission probabilities by its lack of sub/superscripts) and emitter chemical potential $\mu_e$, one obtains for the current sourced from the emitter due to a single band extremum [8]:

\[ I = \sum \frac{1}{2} e \frac{d \rho}{dz} \ln \left( \frac{\rho + 1}{\rho - 1} \right) \]
\[ J_e = \frac{q}{A} \sum_k \int dE.; \rho_c(E_z)[1 + \exp\left(\frac{E_z - \mu_e}{kT}\right)]^{-1} \nu_e(E_z) T_e(E_z) \]

\[ = \frac{q}{(2\pi)^2} \int dE.; \rho_c(E_z) \int d^2k_z [1 + \exp\left(\frac{E_z - \mu_e}{kT}\right)]^{-1} \nu_e(E_z) T_e(E_z) \]

\[ = \frac{2\pi q m_T kT}{h^2} \int dE.; \rho_c(E_z) \nu_e(E_z) \ln[1 + \exp\left(-\frac{E_z - \mu_e}{kT}\right)] T_e(E_z) = \int dE.; j_e(E_z) T_e(E_z) \tag{1} \]

where \( m_T = (m \cdot m_v) \) is the transverse mass and the integral is performed over longitudinal energies above the emitter band edge (or \( j_e T_e \) can be considered zero for \( E_z \) below the band edge). The longitudinal density of states \( \rho_c \) and the “attempt frequency” \( \nu_e \) (rigorously, this is defined as the device-approaching component of the probability current measured infinitesimally close to the device; classically, this can be thought of as the average number of bounces on the barrier per second) in the emitter have been introduced. The final equality in eqn (1) defines the incident current density per longitudinal energy unit, \( j_e \). One must, of course, subtract from \( j_e \) the collector-sourced current (and sum over all band extrema) to determine the total current.

For a bulk, flatband emitter (denoted by the superscript “0”), \( \nu_c \rho_c E \rho_c = \frac{2}{h} \) (a factor of 2 for spin is included in the density of states) and

\[ j_e^0(E_z) = \frac{4\pi q m_T kT}{h^3} \ln[1 + \exp\left(-\frac{E_z - \mu_e}{kT}\right)]. \tag{2} \]

This is plotted in Fig. 1. When band bending within the emitter is considered, this current density is redistributed with sharp peaks appearing at the quantized longitudinal energy levels of the accumulated states (scattering within the emitter and the coupling of these states to the device broadens these peaks, as will be discussed in §2.3). Since \( j_e \) gives the incident current density at the emitter/device interface, \( T_e \) is to be calculated without band bending in the emitter. At this point in our calculations, \( T_e \) is an effective transmission probability which may include scattering within the device; it is precisely the probability of an injected electron reaching the bulk collector.
2.1.1 Single-Barrier Devices

For simple single-barrier devices, it will prove helpful to determine an analytical expression for the expected current density. The WKB approximation [10] (valid for slowly-varying potentials) for the wave function within the barrier (where the wavevector is imaginary with magnitude \( \kappa \)) gives

\[
T' \sim \exp(-2 \int_{\text{barrier}} dz \kappa(z)) = \exp(-\frac{2t_b}{\hbar} \sqrt{2m_b \tilde{V}_b})
\]

where \( \tilde{V}_b \) is the average barrier height (relative to the incoming energy) and \( t_b \) is the barrier thickness. In the equality above, we have defined \( m_b \) as an “effective barrier mass” which takes into account both the potential profile within the barrier and any non-
parabolicity of the barrier's band structure (which is significant when the evanescent energy is well within the bandgap, as is often the case). The superscript “1” is here introduced to distinguish this single-barrier coefficient from full-device (e.g., multi-barrier) transmission probabilities. This is the conventional expression for the transmission coefficient for direct tunneling [11], in which electrons tunnel through the entire barrier, in contrast with Fowler-Nordheim tunneling, in which a large uniform (throughout the barrier) electric field $F$ allows electrons to tunnel into a propagating band in the barrier [12]. For Fowler-Nordheim tunneling in the effective mass approximation, the WKB integral gives the same final result as eqn (3) with $\bar{V}_t$ replaced by $4\phi_b/9$ where $\phi_b$ is the classical barrier height ($t_h$ is then the classical thickness, $\phi_b/qF$). Using the free electron mass, $m_0$ (though the effective barrier mass is usually somewhat smaller than this; e.g., $0.30m_0$ in SiO$_2$ [13]), and an average barrier height of 1 eV, one obtains approximately a tenfold reduction in transmission probability for each 2.3 Å of barrier thickness in the regime of direct tunneling.

Using the single-barrier transmission probability (with $T'=1$ for electrons with energy greater than $\phi_b$) in eqn (1), we obtain the results of Table 1. The emitter-sourced current density within each of the two degenerate regimes was derived assuming an

<table>
<thead>
<tr>
<th>Table 1: Single-Barrier Current Components</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Transport Process</strong></td>
</tr>
<tr>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>Degenerate Direct Tunneling</td>
</tr>
<tr>
<td>Degenerate Fowler-Nordheim Tunneling</td>
</tr>
<tr>
<td>Thermionic Emission</td>
</tr>
</tbody>
</table>
infinitely deep emitter Fermi sea (i.e., the conduction band edge is well below the Fermi level); when it is finite, these expressions remain accurate provided that the finite depth is large relative to the characteristic energy (characterizing the exponential reduction in transmission probability with the depth of the electron) for the process. The total emitter-sourced single-barrier current is that corresponding to the appropriate regime of degenerate emission plus the thermionic contribution. The classical barrier height \( \phi_b \) to be used in Table 1 is the maximum barrier potential energy minus the chemical potential in the emitter.

The above expressions give the emitter-sourced current contribution from electrons with energies less than the chemical potential (the degenerate component) and from those with energies greater than the classical barrier height (thermionic emission); the contribution from intermediate energies (“thermionic tunneling through the tip”) should also be determined, but is not so easily evaluated. In this region of energies, the maximum current contribution to the integrand of eqn (1) occurs for electrons with energy \( \frac{4E_Fn^3}{9(kT)^2} \) below the classical barrier height and may be neglected, in comparison to the thermionic emission component, provided that \( kT \ll E_Fn \).

For highly-conductive emitters, the image-force modifies the potential seen by an incident electron of energy \( E \); in calculating the transmission probability, the maximum potential, \( \phi_b \), is reduced by [14]

\[
\Delta \phi_b = \left( \frac{q^3 FT^4}{4\pi \varepsilon_b} \right)^{1/2}
\]

where \( \varepsilon_b \) is the barrier dielectric constant. \( T' \) here enters self-consistently because the transmitted wave, and hence the image in the emitter, only has charge \( T'q \) [15]. Frenkel-Poole effects at the emitter/barrier interface give the same expression without the factor \( T'/4 \) [16].
2.1.2 Double-Barrier Devices

Although the double-barrier transmission probability is a more complicated function of energy, these same globally coherent methods were first used by Chang, Esaki, and Tsu [1] to explain the J-V characteristics of the RTD. As an example, the transmission probability $T^c$ (the superscript "c" denotes coherent) for a simple one-dimensional structure is shown in Fig. 2. The peak in transmission probability (which actually extends to one, as shown in the inset of Fig. 2) occurs because the wave function at this energy can constructively (i.e., the round-trip distance is an integral number of wavelengths) accumulate so much state within the well that the small fraction which does transmit provides the same current as the incident wave. Higher energy resonances exist but are not plotted. Although one needs to self-consistently recalculate the potential profile and transmission probabilities for each applied voltage, it is clear that an increase in applied voltage will lower the transmission resonance relative to the emitting band.

Figure 2. Coherent Electron Transmission Probability and Transmitted Group Delay Through an Unbiased RTD.
edge. When this resonance falls below the band edge, NDR occurs because the average electron transmission probability is reduced as the bias is further increased (assuming the second resonance is much higher in energy).

Also plotted is the transmitted electron group delay within the device, $\tau_d$, which, analogous to the calculation of the group velocity of a wave packet, is given by

$$\tau_d = \hbar \frac{\partial \delta(E)}{\partial E}$$

where $\delta$ is the phase of the scattering wave function (with energy $E$) after transmission (immediately after the final barrier) [17]. For this calculation, the phases of all incident (meaning the reflected component is neglected) wave functions must be chosen to coincide (for different $E$) at the beginning of the device (the onset of the first barrier); this phase matching simply forces to zero the group delay of the incoming wave to reach the beginning of the device so that no propagation delay within the bulk emitter is included in the total group delay. Physically, $\tau_d$ represents the expected amount of time that a transmitted electron wave packet (of infinitesimal energy width) “interacts” with the device. This method for determining the group delay can be generalized to determine electron “dwell times” in particular regions (e.g., the barriers only) of the device using perturbation methods [18], but the extended method is computationally much more difficult and would add little to our discussion.

In Fig. 2, $\tau_d$ is plotted merely to demonstrate that an electron, especially near resonance, spends a significantly long time (~40 ps) transmitting through the device in comparison to bulk scattering times (on the scale of ps). The assumption of globally coherent transport is invalid. When the electron is near resonance, and is thus allowed to accumulate in the well, it classically reflects many times, back and forth off the barriers, before leaving; it will likely scatter during this long process. Inclusion of scattering is not difficult within the “sequential tunneling” model which we now explore.
2.2 Sequential Tunneling

This model was first proposed by Luryi [19], who considered the transmission through an RTD to be a two-step process; an electron tunnels first from emitter into well, where it may scatter and equilibrate, and second from well into collector (or back into emitter). To explore this possibility, we rely on the results of Bardeen which showed, to first order in time-dependent perturbation theory, that the rate of probability transfer from the state $|e\rangle$ into the state $|d\rangle$ (signifying “device”) across a weakly-transmissive barrier is given by the matrix element of the current operator between corresponding states of the uncoupled system [20, 21]. It is easy to reformulate this matrix element in terms of the attempt frequencies of these states and the electron transmission probability for the single barrier (meaning bulk, flatband “emitter” and “device” regions are artificially imposed as in Fig. 3), $T'$. The rate of emission from state $|e\rangle$ (of energy $E_e$) into state $|d\rangle$ ($E_d$) is then found to be [22]

$$\Gamma_{ed} = \hbar T_{ed} |v_e \omega \delta (E_e - E_d))|. \tag{6}$$

$\delta$ tends to the Dirac delta function (with approximate width $\frac{\hbar}{t}$; the requirement that $|e\rangle$ not be significantly depleted after $t$ places constraints on the breadths of the densities of states involved) as the time of emission ($t$) becomes large. The total rate of emission from state $|e\rangle$ into all “device” states, $\{ |d\rangle \}$, is then

$$\Gamma_{e\{d\}} = \sum_d \Gamma_{ed} = \omega \hbar \sum_d T_{ed} |v_d \omega \delta (E_e - E_d)) = \omega \hbar T_{ed} |v_d \omega \frac{\rho_{dc}}{2} = \omega \tilde{T}_{e\{d\}}. \tag{7}$$

In the third equality, we have again assumed a common transverse lattice so that the transmission probability is non-zero only when the $k_T$ (and the spin, which has forced the division by 2) of $|e\rangle$ equals that of $|d\rangle$. The final equality defines $\tilde{T}_{e\{d\}}$. When bulk, flatband material is on the “device” side of the barrier, the reciprocal relation between $\rho_e$ and $v$ gives

$$\tilde{T}_{e\{d\}} = T_{ed}^{-1} = T_{de}^{-1}. \tag{8} \text{ (for bulk, flatband “device” region)}$$

as is required (since the bulk, flatband condition simply asserts that there are no subsequent reflections and interference effects; all transmitted electrons are collected).
The second equality in eqn (8) follows for a spin-independent Hamiltonian from time-reversal symmetry [8].

![Diagram](image)

**Figure 3. Diagram for the Determination of the Electron Transfer Rate Between Barrier-Isolated Regions from Eqns (6) and (7).**

In general, even when the "device" contains subsequent reflections (e.g., a quantum well), \( \tilde{T}_e(d) \) is the effective transmission probability from the state \( |e\rangle \) into empty "device" states. This differs from the full-device transmission probability, \( T_e \), in that the "device" states may not be empty during steady-state mono-energetic injection from the emitter. For example, in a quantum well with one important resonant state \( |d\rangle \) coupled to emitter and collector ("c"), the fraction of electrons stored in the well which transmits to the collector is \( \frac{\Gamma_{d(c)}}{\Gamma_{d(c)} + \Gamma_{d(e)}} \). Since this quantity is usually very close to unity when the device is biased, transport is limited by emission into the well and subsequent scattering is irrelevant; results of the globally coherent model are then accurate. When this quantity is not very close to unity, however, there is a clear distinction between \( T_e \) and \( \tilde{T}_e(d) \):

\[
T_e = \frac{\Gamma_{d(c)} \tilde{T}_e(d)}{\Gamma_{d(c)} + \Gamma_{d(e)}} = \left( \frac{h \Gamma_{d(c)} \Gamma_{d(e)}}{\Gamma_{d(c)} + \Gamma_{d(e)}} \right) \frac{\rho_{de}}{2}. \tag{9}
\]

Eqn (9) shows that the transmission probability through an RTD mimics the density of states in the well. Assuming that other quantities in eqn (1) vary negligibly while \( \rho_{de} \) is significant (where it integrates to 2 because of our treatment of spin),
\[ J_e = \frac{2\pi q m k T \Gamma_{d(c)} \Gamma_{d(e)}}{h(\Gamma_{d(c)} + \Gamma_{d(e)})} \rho_{e}(E_R) \nu_{e}(E_R) \ln[1 + \exp(-\frac{E_R - \mu_e}{kT})], \]  

(10)

and measurement of current probes the density of states in the emitter at the resonant energy \( E_R \). The RTD therefore turns off when biased such that \( E_R \) lies in the emitter energy gap, where the density of states is zero. Of course, in the above expression, \( \nu_e \) and the \( \{ \Gamma \} \) depend on the applied bias.

The usefulness of the sequential tunneling model lies in its ability to reduce a problem of multi-barrier transmission into one of single-barrier transfer rates specified by the densities of states in barrier-isolated regions, as expressed in eqns (6) and (7). Electron scattering may then be formally introduced via scattering transfer rates within the barrier-isolated regions. From these rate (and continuity) equations, the dynamic and steady-state properties of multi-barrier devices are easily and intuitively evaluated.

### 2.3 Resonance Broadening

For use of eqn (9), the broadening which occurs to an approximately discrete density of states must be quantified. From first-order perturbation theory, a state of unperturbed energy \( E_0 \) which depopulates at a rate \( \Gamma \) has an approximately Lorentzian density of states given by [23]

\[ \rho(E) = \frac{h\Gamma/2\pi}{(E - E_0)^2 + (h\Gamma/2)^2}. \]  

(11)

The total depopulation rate may be computed as the sum of the depopulation rates due to different independent (and weak) broadening processes.

We have already written in eqn (7) an expression for the magnitude of one broadening process: the transmissive barrier itself. For double-barrier coherent transmission, eqns (9) and (11) give

\[ T_e^{\Gamma} = \left( \frac{h\Gamma_{d(c)} \Gamma_{d(e)}}{\Gamma_{d(c)} + \Gamma_{d(e)}} \right) \frac{h(\Gamma_{d(c)} + \Gamma_{d(e)})/2\pi}{(E - E_0)^2 + (h(\Gamma_{d(c)} + \Gamma_{d(e)})/2)^2} = \frac{h^2 \Gamma_{d(c)} \Gamma_{d(e)}}{(E - E_0)^2 + (h(\Gamma_{d(c)} + \Gamma_{d(e)})/2)^2}, \]  

(12)
exactly as obtained in [9] by directly solving the coherent Schrödinger equation.

To quantify the resonance broadening due to barrier transmission, consider a flatband region of width \( w \), whose states couple through a barrier to bulk, flatband material. In the effective mass approximation, without barrier coupling, the approximate eigenstate energies are \( E_n = \frac{\hbar^2 \pi^2 n^2}{2m_w^2} \) (for all positive integers \( n \)). From eqns (7) and (8),

\[
h\Gamma_b = h\nu T' = \frac{\hbar^2 k}{2m_w w} T' = \frac{E_n}{\pi n} T' = \frac{(E_n - E_{n-1})}{2\pi(1 - \frac{1}{2n})} T',
\]

which shows that the broadening roughly equals the uncoupled energy level separation multiplied by the transmission probability. In general, since broadening in a maximally coupled system \( (T' \sim 1) \) is expected to wash away the uncoupled spectrum, and must therefore be the same magnitude as the uncoupled energy level separation, we heuristically conclude that \textit{resonance broadening due to a transmissive barrier roughly equals the uncoupled resonance separation multiplied by the probability of transmission}. Of course, for a quantum well, there are two barriers and the broadening due to each should be added, as was done in eqn (12).

The effects of scattering processes on a resonance may be easily estimated through knowledge of the average carrier mobility \( \mu \). This is classically \( \frac{q\langle\tau\rangle}{m_c} \) where \( m_c \) is the conductivity effective mass \( \left( \frac{1}{m_v} + \frac{1}{m_o} + \frac{1}{m_o} \right)^{-1} \) if isotropic and \( \langle\tau\rangle \) is the average carrier lifetime. The scattering rate \( \Gamma_s \) is roughly \( \frac{1}{\langle\tau\rangle} \) and the energy broadening due to scattering can be estimated as

\[
h\Gamma_s = \frac{h\nu}{\mu m_c}.
\]

For typical mobilities of 1000 cm²/Vs and \( m_c = m_o \), \( h\Gamma_s = 1.2 \) meV; broadening due to scattering is typically much larger than the broadening due to the barriers. Therefore, on the basis of resonance broadening alone (i.e., even if \( k_r \) were conserved), RTD valley current is typically dominated not by the finite transmissivity of the barriers, but by incoherent scattering processes.
From first principles, it is extremely difficult to quantify the perturbative effects of scattering processes on the measured current. This requires a detailed knowledge of the wave functions and their scattering cross-sections for particular impurities, defects, and phonons. Chevoir and Vinter have pursued this and found, as we have done heuristically, that the valley current of an RTD is almost exclusively due to incoherent scattering processes \[24\]. In addition to scattering processes in the well, their treatment includes something which we have not yet considered: non-$k_T$-conserving single-barrier transmission. It is important to realize that, although crystalline defects and interface roughness may only cause slight scattering among $k_T$ through a single barrier, the transmission enhancement for the small amount which does scatter into resonant states (only the longitudinal density of states in the well is sharply peaked; even elastic processes, which transfer longitudinal energy into transverse energy by non-conservation of $k_T$, can scatter an off-resonance electron into a resonant state) may contribute significantly to the valley current.

In the case of an amorphous barrier with low transmissivity (meaning that the electron-barrier interaction is strong) separating two flatband crystalline regions (ignoring practical difficulties for the moment), one might expect the transmitted electron (for an incident electron of a particular energy and $k_T$) to have $k_T$ components spread almost uniformly (but only into states propagating away from the barrier, and likely peaked at the incident $k_T$) over a single energy surface (corresponding to the incident energy) of the Brillouin zone. The observation of Fowler-Nordheim oscillations through amorphous SiO$_2$ and their excellent fit with theoretical one-dimensional (i.e., $k_T$-conserving) models, however, seem to indicate that $k_T$ is actually well-conserved through the single barrier \[25, 26\]. It would be very interesting and enlightening to model an amorphous barrier and quantify the $k_T$ scattering which is expected in transmission. Of course, the requirement of multi-dimensional modeling, the lack of long-range order, and the lack of understanding of short-range order in amorphous materials makes this a very difficult task; it has not been attempted to date. This author’s personal belief is that the thermally-grown oxides through which Fowler-Nordheim oscillations have been observed are actually quite ordered, and not literally “amorphous”.

Whatever the cause, the experimental work on Fowler-Nordheim oscillations indicates that scattering processes in thermally-grown SiO$_2$ which do not conserve $k_T$ are not significant and that SiO$_2$-barrier RTDs have promise. Of course, an RTD with amorphous barriers is likely to also have an amorphous well, which would cause intolerable resonance broadening ($\sim$1.2 eV) for typical amorphous mobilities ($\sim$1 cm$^2$/Vs)
used in eqn (14). Although these difficulties exist for amorphous RTDs, and all sharp
detail is expected to be significantly smeared, noticeable inflections (but not clear NDR)
in the low-temperature J-V characteristics of amorphous SiO₂/Si/SiO₂ RTDs have been
observed [27]. Other reports of NDR in amorphous systems [28] appear, on close
inspection by this author, to be defect related. The considerations presented here and the
lack of positive experimental results cast strong doubt on the existence of practical
applications for non-crystalline RTDs (unless a method can be found for quality
overgrowth of crystalline Si on thin SiO₂); a lattice-matched materials system is preferred.

2.4 RTD Circuit Models and Instabilities

Within the picture of sequential tunneling, it is easy to obtain a dynamic model for
resonant tunneling. When biased such that backflow from collector into well and from
well into emitter are negligible, the total RTD current density $J$ is just the charge density
within the well, $\sigma$, multiplied by the emission rate out of the well, $\Gamma$ (eliminating
subscripts from what was $\Gamma_{d(c)}$ in our previous explicit notation). With charge injected
into the well from the emitter at the rate $J_e$, we may write:

$$\frac{1}{\Gamma} \frac{d}{dt} J = \frac{1}{\Gamma} \frac{d}{dt} (\sigma \Gamma) = \frac{d}{dt} \sigma - J \frac{d}{dt} \left(\frac{1}{\Gamma}\right) = J_e - J \frac{d}{dt} \left(\frac{1}{\Gamma}\right).$$

(15)

It now seems reasonable to consider $J_e$ and $\Gamma$ to be explicit functions of the applied
voltage with no dynamic contribution from transient charge in the well. This assumption
is generally not satisfied however; charge accumulation in the well does indeed have an
effect on these parameters, as is evidenced by the hysteresis and bistability which exist in
the J-V characteristics of all RTDs. The hysteresis present in the experimental data of
Fig. 7 (the valley voltage is always higher when sweeping from 0 to 1 V than when
sweeping from 1 to 0 V) is caused by charging of the well. In the positive-going sweep,
at a point of bistability, more electrons reside in the well than do so on the negative-going
sweep, since electrons must transmit through and rest in the resonant state. Therefore,
there is a smaller potential drop between emitter and well in the positive-going sweep,
raising the resonant energy relative to emitting energies from that observed in the negative-going sweep.

For the moment, we neglect these charging effects (which can ultimately be ignored when we consider only small signals) and use the DC characteristics, \( J_e(V) \) and \( \Gamma(V) \), to obtain the large-signal model from eqn (15):

\[
(1 + \frac{dV}{dt} \frac{d}{dV} \frac{1}{\Gamma(V)}) J = J_e(V) - \frac{1}{\Gamma(V)} \frac{d}{dt} J. \tag{16}
\]

This leads immediately to the small-signal model (in which lower case letters represent the differentials of the corresponding large signals),

\[
v = R_0 j + \frac{R_o}{\Gamma} \frac{d}{dt} i, \tag{17}
\]

where \( \Gamma \) and \( R_0^{-1} = \frac{d}{dV} J_e \) are evaluated at the operating point. The well therefore exhibits the series combination of resistance and inductance, with the inductance lag caused by trapped charge in the well. The full small-signal model must also include any bulk and contact series resistance, \( R_s \), and the depletion capacitance, \( C_D \), as shown in Fig. 4. Within the tank circuit, the ratio of impedance magnitudes (for radial frequency \( \omega \)) is \( 1: \omega |R_o|C_D : \omega |R_s|C_D(\omega / \Gamma) \) capacitor : resistor : inductor. Thus, the inductive impedance is significant only for frequencies comparable to or greater than \( \Gamma (~10 \text{ Thz}) \) [17].

This small-signal equivalent circuit applies for all values of \( R_o \), but is most interesting in the NDR region, where both the resistor and the inductor assume negative values. Differentially, the circuit is then active and instabilities arise. It is not hard to see that natural frequencies with exponentially increasing time dependence exist throughout the entire NDR region, and we must therefore rely on the large-signal model for circuit analysis in this region. Here, positive feedback (negative resistance) transfers charge back and forth (within the tank circuit) from the well (inductor) to the depletion regions (capacitor), and creates the bizarre wave-plateau behavior displayed in Fig. 7 between 0.3 and 0.5 V. The details of this plateau are not easily explained, but it is clear from these
arguments that the local instability of the system will create macroscopic current oscillations, significantly rectified by the nonlinearity of these devices. Dynamic theoretical models within the NDR region for a particular structure found ongoing 5 THz oscillations whose time-averaged current matched fairly well with the experimental plateau [29].

3 InP-Based RTD

The theory presented above has been verified by a wealth of experiments on resonant tunneling heterostructures [3] afforded by modern advances in III-V MBE and metalorganic chemical vapor deposition (MOCVD). We have examined a pseudomorphic (i.e., the InP lattice is preserved throughout the device with tensile AlAs and compressive InAs) InP(100)-based RTD whose (symmetric) material cross-section and energy-band diagram are shown in Fig. 5. Non-parabolic band-structure effects are strong in this structure, and we must monitor the conduction band minima at Γ, X, and L [14]. The minima in Fig. 5 correspond to the strained materials [30]. Barriers and well were not intentionally doped. Emitter and collector were degenerately doped n-type (1x10^{18} Si/cm^3; the Fermi level is approximately 89 meV above the bulk InGaAs conduction band edge) with 2 nm of undoped buffer layers grown adjacent to the barriers. Transmission electron microscopy (TEM) was performed on this structure (a different growth with nominally the same structure) and revealed the sharp interfaces in Fig. 6.

Room-temperature J-V characteristics of a (1.4 μm)^2 device are shown in Fig. 7. Qualitatively, these characteristics were expected. Many formal theoretical methods exist which include electron scattering (density matrices [31], Wigner functions [32], Green’s functions [33]) and attempt to quantitatively predict the current density within such a structure, but these are very sensitive to a number of unknown parameters. First, the magnitude of the injected current for all biases depends strongly on the carrier density within the emitter and any quantized surface accumulation. Second, as mentioned in §2.3, the valley current is typically dominated by 3-D effects and phonon scattering, whose magnitude may be adjusted to match experimental data. Finally, one must contend
with non-parabolic band structure effects and wavefunction connection rules across interfaces. For this particular structure, treatment of the barriers is particularly difficult (and rigorously impossible) in the effective mass approximation because the conduction band extremum at $X$ is lower in energy than that at $\Gamma$ (i.e., $X$ states are less evanescent and coupling between emitter $\Gamma$ states and evanescent barrier $X$ states becomes important for thick barriers [34, 35]). A 10-band tight-binding Hamiltonian was used to calculate the two lowest (in energy) electron resonances shown in Fig. 5 [36]. The excited resonance energy is much lower (by ~0.5 eV) than that which would be predicted by the effective mass approximation (at $\Gamma$).
mainly because of the non-parabolicity of the InAs band structure and the small InAs electron effective mass ($0.02m_0$).

Theoretical $J$-$V$ calculations will not be pursued here. Instead, we focus on characterizing the thermally-activated valley current from the temperature-dependence of the measured current, shown in Fig. 8. The strong dependence of the valley current on temperature raised the PVR from 3.7 at 200°C to 20.5 at -75°C. The magnitude of the current before its peak was limited partially by contact/series resistance and did not depend significantly on temperature because of the degenerate emitter doping.

In the valley region, one can extract an activation energy from the Arrhenius plot in Fig. 9. If the model of eqn (10) were correct (that is, if the valley current were attributable to thermionic emission through a sharp resonance), this log-scale plot would appear linear for activation energies much greater than the thermal energies ($kT$) considered. For the fits of Fig. 9, we assumed that the chemical potential was approximately constant over the temperature range of interest. If the activation energy were much greater than the thermal energies considered and the chemical potential varied...
linearly with temperature, it is easy to see from eqn (10) that the activation energy measured would be the difference between the excited resonant energy and the chemical potential extrapolated to zero temperature. In addition to this activated current, the fits in Fig. 9 include a temperature-independent current (of the same magnitude as the activated current) which is attributed to scattering and defect transport of electrons from the emitter Fermi sea.

When the activation energy becomes comparable to or smaller than the thermal energies considered, this fit becomes sensitive to the prefactor (prefactor here refers to the temperature-dependence outside of the logarithm in the fit expression: $T'$ in Fig. 9). For our data, they were comparable and the fit activation energies (for the voltages shown in Fig. 9) dropped from 72-102 meV with no prefactor ($T^0$) to 44-75 meV with the prefactor $T'$ to 8-49 meV with the prefactor $T^0$ (which would be more applicable if the valley-current-contributing transmission probability were approximately constant over the thermal energy spread as was assumed for the thermionic emission expression in

\[ T^0 \]
Table 1). All of these fits, with three adjustable parameters, matched the data very well, but we henceforth consider only the $T'$ model because of its foundations in eqn (10).

The order of magnitude of eqn (10) and the excited resonance energy in Fig. 5 are consistent with the experimental current densities and activation energies shown in Fig. 9. The activation energies are significantly (> 2x) larger than the optical phonon energies of the materials considered, and the possibility of significant phonon-assisted transport in the thermally-activated current component can be ruled out. Another indication that the thermally-activated current component is due to emission through the excited resonance is the activation energy's decrease with increasing bias. The excited resonance energy is reduced (relative to the emitter Fermi level) by slightly less than half of the $q$-multiplied voltage increase because slightly less than half of the voltage increase appears across each barrier (with the remainder dropped in the increased collector depletion region). Self-consistent simulations [36] for these biases agree well with the obtained activation energies and we conclude that the thermally-activated valley current is due to thermionic emission through the excited resonance.
4 Si-Based Materials

Since processes which contribute to the valley current in RTDs are thermally-activated and low-temperature measurements are often required to reduce this contribution, it is important to characterize the conductivity of bulk Si at low temperatures. Compared to lightly-doped III-V materials whose donor levels are typically between 1 and 10 meV below the conduction band edge, donor levels in lightly-doped Si are fairly deep. To be specific, ground state donor levels lie approximately 51 meV and 54 meV for Sb and P, respectively, under the Si conduction band edge in the limit of light doping [37, 38]. As the dopant density increases, however, the reduction in potential energy of the extended conduction band states (due to the background of ionized donors of concentration $N$) drops the above activation energies by approximately $[4.3 \times 10^{-8} \text{ eV cm}] N^{1/3}$. This first-order model predicts the degeneracy, at all temperatures, of Sb-doped Si for $N > 1.7 \times 10^{18} \text{ cm}^3$ [39]. Of course, localized donor levels always lie lower in energy than the extended conduction band states, and higher-order effects must be considered for doping densities of this magnitude. Also, for heavily-doped samples, consideration of conductive subbands which arise about the isolated donor levels (due to interactions of the localized electron states at different donor sites) is required.

To investigate and quantify the carrier freeze-out which occurs in heavily-doped ($\sim 10^{18} \text{ Sb/cm}^3$) Si, Hall measurements [14] were performed on Van der Pauw structures [40]. Plotted in Fig. 10 at various temperatures are the obtained conductivities and Hall mobilities.

For the moment, let us assume that conduction only occurs in the Si conduction band, with negligible contribution from donor subbands. Then, the average carrier mobility, $\mu$, in the effective mass approximation is given by [41]

$$\mu = \mu_H \frac{\langle \tau \rangle^2}{\left\langle \tau^2 \right\rangle} \frac{3m_e m_h m_z}{m_e^2 (m_e + m_h + m_z)},$$

where the subscript “H” denotes the Hall mobility. In heavily-doped samples, ionized impurity scattering is likely to determine the carrier lifetime, for which
Figure 10. Experimental Conductivity and Hall Mobility in N+ (Sb-doped) Si Substrates.

\[ \frac{\langle \tau \rangle^2}{\langle \tau^2 \rangle} = \frac{512}{315\pi} \] [41]. Using this and the masses of Si, the carrier density and mobility (\(\mu=0.598\mu_\parallel\)) were calculated for Fig. 11. The room-temperature mobility is much smaller than that observed in the purest of Si (1500 cm²/Vs) but is not significantly below previously reported values for heavily-doped Si (200 cm²/Vs for 2x10^18 cm⁻³) in which ionized impurity scattering dominates electron transport [14]. The sharp knee which occurs near 25 K indicates a donor activation energy of approximately 2 meV (more quantitative numerical methods which rely on some assumed density of donor states give roughly the same value). For temperatures below approximately 15 K, with thermal energies small enough to resolve donor-state details, the leveling off of the conductivity in Fig. 10 is caused by conduction within donor subbands.

For future low-temperature characterization of RTDs grown on similarly-doped substrates, Fig. 10 should be used to ensure that bulk series resistance is not limiting the measured current.
Figure 11. Carrier Density and Mobility in N+ (Sb-doped) Si Substrates Calculated Neglecting Conduction Within Donor Subbands.

4.1 ZnS

Although ZnS is lattice-matched to Si, the ZnS/Si system has not been thoroughly explored. At this stage of ZnS/Si(100) development, bonding complications at the Si surface are overcome using an As monolayer to join the Si (100) surface to the Zn atoms. Vicinal substrates with a 4° off-cut towards [0 1 1] are typically used to provide a double-stepped Si surface which allows for the growth of single-domain (to an approximately 50-100 nm extent [42]) ZnS. Details of the ZnS growth, which was performed by faculty and students at Texas A&M University (TAMU), and the interface chemistry can be found in [42, 43].
Briefly, 2000 Å Si buffer layers (not intentionally doped, but significant background As pressure was always present) were e-beam evaporated onto 1-10 Ω-cm (phosphorous-doped) Si substrates (cleaned by chemical oxidation with subsequent 870°C removal of oxide) in 10^-11 T ultrahigh vacuum (UHV) at a substrate temperature of 620°C. An anneal at 850°C produced the double-stepped Si surface, as confirmed by in situ reflection high-energy electron diffraction (RHEED). After the substrate was cooled to room temperature, more than 1 ML of As was deposited from an effusion cell. The excess (beyond 1 ML) As was then removed with a 600°C anneal. Approximately 2000 Å of ZnS was epitaxially evaporated (~0.3 Å/s) from a crystalline ZnS source (in an effusion cell) at a substrate temperature of 50°C. Samples were subsequently annealed at 320°C (for ~10 min) and removed from vacuum.

2000 Å of Al or Au was then e-beam evaporated in HV (~10^-7 T). Metal covered the entire wafer except for thin rings which isolated 15, 150, and 1500 μm diameter circular devices. In the case of Al, rings were etched in Al-leach; in the case of Au, a lift-off process was used to remove the rings. The large area metal contact outside of these circles, which was effectively shorted to the backside (confirmed by checking that the measured current scaled with area), was used as the collector for electrical measurements.

Prior to this research, the band edge offsets for this Al/ZnS/As(1ML)/Si structure were unknown. Photoelectron spectroscopy has been used to measure the conduction band edge offset between ZnS and non-As-terminated Si(111) as 1.7±0.2 eV, though the quality of this epitaxy was poor [44]. The substrate orientation and the presence of As affect the interface dipole, and this offset is likely different for As-passivated ZnS/Si(100) structures. Since a large barrier height (relative to kT) is required for RTD NDR, it is important that the barrier height for the As-terminated structures be quantified. Temperature-dependent J-V characterization was used for this purpose.

Fig. 12 displays the currents measured for the Al samples at selected temperatures. The magnitude of the voltage was swept from 0 to 2 to 0 V (arrows indicate the sweep direction) at a rate of ~1 V/min. The lower temperature J-V curves, with their very low current magnitudes (150 μm diameter devices were tested), are dominated by noise and displacement current (primarily charging traps in the ZnS, as will be discussed). The Arrhenius plot in Fig. 13 was constructed from these data (except where arrows indicate the sweep direction, plots correspond to the |voltage|-increasing sweeps) to determine the activation energy, if one existed, for conduction. At the lower
Figure 12. Current Densities Through Al/ZnS/As(1ML)/Si Diodes at Different Temperatures.

Figure 13. Arrhenius Plot for Determination of Activation Energy in Al/ZnS/As(1ML)/Si Diodes.
temperatures, where these curves flatten out, the current measured was presumably
defect-assisted transport (e.g., via grain boundaries or traps in the ZnS) with a relatively
small thermal activation energy. At the higher temperatures, the log-linearity of these data
indicate a thermally-activated process with a well-defined activation energy; this is plotted
in Fig. 14 as a function of the applied bias. The thermionic emission prefactor $T^2$ (see
Table 1) was used for these fits, but its use is not well-founded (among other
simplifications, the model which produced it assumed unity transmission for all
classically-transmitted electrons and neglected backscattering within the barrier) beyond
an order-of-magnitude calculation; the obtained activation energies are only accurate to
$kT \approx 40$ meV which, entering in the exponential, determines the order of magnitude of
the current. Since the detail within Fig. 14 at small gate voltages is finer than this, it will
not be analyzed.

To relate the activation energy to a band diagram using the thermionic emission
model of Table 1, as was done for Fig. 15, the chemical potential must be obtained as a
function of voltage and temperature (since the extrapolated zero-temperature barrier height

![Diagram](image)
is measured; see §3). For a metal (with a Fermi sea much deeper than $kT$), the chemical potential is independent of both of these parameters; the value obtained from Fig. 14 for negative biases, $1.02\pm0.04$ eV, is the Al/ZnS barrier height. For a non-degenerate semiconductor with fully-ionized donors, however, charge neutrality in the bulk forces $E_c - \mu \over kT$ to remain constant as the temperature changes, and the extrapolated $\mu$, to $T=0$ (at $T=0$, of course, the donors are not fully-ionized, but we extrapolate from data obtained at higher, fully-ionizing temperatures), is therefore approximately $E_c$ (this can also be seen from Fig. 18). Therefore, the activation energy measured for positive gate voltages, 1.00 eV, is approximately the difference between the maximum barrier potential and the bulk Si conduction band edge.

![Figure 15. Zero-Bias Energy Band Diagram for Al/ZnS/As(1ML)/Si(100) Structure.](image)

Then, to determine the ZnS/Si conduction band edge offset, the amount of band-bending within the Si must be known. Capacitance-voltage (C-V) measurements, shown in Fig. 16, were made to help quantify this, but displayed much hysteresis and noise in
Figure 16. Measured Capacitance of Al/ZnS/As(1ML)/Si Diode Demonstrating Noise and Hysteresis.

Re-measuring a particular device (error bars extending one standard deviation above and below the average capacitance are shown). The hysteresis signifies field-driven transfer of charge within the ZnS (i.e., mobile ions or trapped electrons in the ZnS, not at the ZnS/Si interface which would be marked by the opposite hysteresis [14]), which may be the cause for the noise. Electron traps in the ZnS may also account for the lack of image-force barrier lowering in Fig. 14 which, from eqn (4), should be roughly 50 meV when 2 V are applied across the ZnS. If there were traps in the ZnS, and the ZnS charged negatively when current flowed, the electric field at the emitter/barrier interface (and hence the barrier reduction) would have been smaller than that which would have been predicted neglecting traps. In fact, the constant activation energy for voltages greater than approximately 0.3 V indicates that there is no electric field at the emitter/barrier interface (as pictured in the band diagrams in Fig. 14) and, therefore, that this activation energy corresponds to the flatband barrier height; 1.00±0.04 eV is the ZnS/Si conduction band edge offset.

This value should be compared with that which would be obtained neglecting traps in the ZnS. For this purpose, ideal theoretical MIS C-V curves (numerically
calculated from the equations in the Appendix) are shown alongside the experimental data in Fig. 17. The depletion capacitance indicates a doping density of $\sim 5 \times 10^{14} \text{ cm}^{-3}$. The $C-V$ stretch-out (from the $5 \times 10^{14} \text{ cm}^{-3}$ theoretical curve to the experimental data) is common and can be attributed to traps in the Si, in the ZnS, or at their interface [14]. Fig. 18 shows the theoretical variation of the chemical potential at the ZnS/Si interface with the applied voltage, and indicates that approximately 0.2 eV of band-bending exists within the Si at a bias of 2 V (if the flatband voltage is approximately zero). $C-V$ stretch-out reduces this figure, so that the extracted ZnS/Si conduction band edge offset neglecting ZnS traps is between 1.0 and 1.2 eV. This range roughly contains that obtained in the preceding paragraph, and we conclude that the ZnS/Si conduction band edge offset is 1.1±0.1 eV.

![Figure 17. Comparison of Measured Al/ZnS/As(1ML)/Si Capacitance with Theory.](image)

The symmetry (between positive and negative biases) in the $J-V$ characteristics and the Ohmic behavior of $A^*$ (in Fig. 14) make one suspect that the conductivity of bulk ZnS, not interfacial barriers, limited the current in these measurements. Indeed, at room
Figure 18. Theoretical MIS Capacitance and Semiconductor Fermi Level Corresponding to Al/ZnS/As(1ML)/Si Structure.

temperature, with typical electron mobilities of 165 cm²/Vs, the mean free path of an electron in pure ZnS is roughly 60 Å [14]; backscattering is significant through 2000 Å of ZnS and drift/diffusion through the ZnS must be considered. The quasi-Fermi level for drift/diffusion through the ZnS can be established either by the (Fermi level of the) emitter or by impurity states in the ZnS (e.g., donor levels). In the first case, the obtained activation energy is still the Schottky barrier height (but the prefactor differs from that of Table 1 since the current is proportional to the gradient of the Fermi level in the ZnS [14]). In the second case, the obtained activation energy is that for bulk ZnS conduction.

In an attempt to eliminate the second possibility, the Au electrode MIS diodes were tested. Previous work has determined Al/ZnS and Au/ZnS Schottky barrier heights to be approximately 0.8 eV (whose difference from the 1.0 eV we have obtained could be attributed to a difference in surface preparation) and 2.0 eV, respectively [45]. If the interface properties of these materials were limiting the current, one would expect to
measure significantly less current through the Au devices than through the Al devices for negative gate voltages.

This large reduction in current was indeed observed (~20x). Additionally, Au devices degraded (as shown in Fig. 19) when current densities (in either direction) on the order of $10^{-6}$ A/cm$^2$ were applied. Although the current measured under positive gate biases was not noticeably affected by the $10^{-6}$ A/cm$^2$ stress, that measured under negative biases (which drove ~20x less current than positive biases) was significantly (~3x) reduced by the stress. As will be argued further, the current measured under negative biases was primarily due to defects which were annealed and removed ("degraded" may be a misnomer) by the stress.

We focus then on the degraded device, and display the measured temperature-dependent $J$-$V$ data in Figs. 20 and 21. This highly-asymmetric characteristic is consistent with the model of Schottky emission. In Fig. 21, a well-defined activation energy exists only for positive gate voltages greater than approximately 0.6 V, for which Fig. 22 (the inset shows both sweeps and significant hysteresis) indicates a barrier height
increasing temperature from 100°C to 175°C (in 25°C steps)

Figure 20. Current Densities Through Degraded Au/ZnS/As(1ML)/Si Diodes at Different Temperatures.

solid lines are positive gate biases (from 0.2 V to 2.0 V in 0.2 V steps); dashed lines are negative gate biases

Figure 21. Arrhenius Plot for Determination of Activation Energy in Degraded Au/ZnS/As(1ML)/Si Diodes.
of approximately 1.0 eV, which is consistent with that obtained for Al electrodes. For negative biases, defect transport processes dominate any bulk emission over the large Au/ZnS barrier height until temperatures much greater than those measured here are reached. Fig. 21 indicates that these defect processes do not have a single well-defined activation energy.

Figure 22. Extracted Activation Energies in Degraded Au/ZnS/As(1 ML)/Si Diodes.

It would be consistent with our data for the bulk ZnS conductivity to have an activation energy of 1.0 eV (due perhaps to a deep donor level) and for the Al/ZnS barrier height to be somewhat larger than this (but not nearly as high as that of Au/ZnS, for which the interface truly limits the current). The previously-quoted 0.8 eV Al/ZnS barrier height [45], however, makes this scenario doubtful. Similarly, the ZnS/Si conduction band edge offset may be larger than 1.2 eV. The Al/ZnS or ZnS/Si barrier heights are not likely to be smaller than 1.0 eV because this would require an impurity density of states (within the ZnS) more complicated than a set of donor levels (the ZnS must charge significantly negative near the Al or Si). The most straightforward model is that of Fig.
15. The approximately 1 eV ZnS barrier heights to Al and Si, being much larger than \( kT \) at room temperature, are encouraging for the future room-temperature operation of ZnS/Si RTDs.

4.2 CaF\(_2\)

The CaF\(_2\)/Si system, though lattice-matched, is unruly. Previous theoretical calculations have shown that the dipole moment of the terminated CaF\(_2\) (100) surface diverges and that a perfect CaF\(_2\)/Si(100) interface is prohibited. The CaF\(_2\) (111) surface, however, does not have this problem and high-quality CaF\(_2\) growth has been demonstrated on Si(111) substrates [46].

We have fabricated and characterized single-barrier CaF\(_2\) devices on Si(111). Growth of these samples was done in an UHV (7x10\(^{-11}\) T) system on HF-cleaned (followed by 900°C H-desorption in UHV) 0.08-0.018 \( \mu\)m (Sb-doped) Si(111) substrates. CaF\(_2\) layers were congruently evaporated from a crystalline CaF\(_2\) source (effusion cell temperature \(~1300°C\)) onto a 100 \( \AA \) e-beam evaporated (undoped) Si buffer layer. Further Si was deposited and the wafer was removed from UHV. Substrate temperature was maintained at 600°C during deposition of both Si and CaF\(_2\). Immediately after a dilute HF etch (to remove any surface oxide), Al metallization was performed exactly as in §4.1. In some cases, a CF\(_4\):O\(_2\) reactive ion etch (RIE) was performed to ensure that spreading resistance through the Si cap was negligible. Four thickness combinations were grown and tested: (nominally) 50/50, 200/50, 200/200, and 200 \( \AA \) Si/2000 \( \AA \) CaF\(_2\).

We discovered a room-temperature reaction between Al and CaF\(_2\), as evidenced by visible pitting of the Al surface on the 50/50 structures after metallization. The HF etch prior to metallization appears to be a factor in this reaction, and the chemistry involved is presently being investigated. The samples with 200 \( \AA \) Si buffer layers were grown to avoid this issue.

The band diagram for the 200/50 structure is shown in Fig. 23. The 2.3 eV CaF\(_2\)/Si(111) conduction band edge offset was determined elsewhere with X-ray
photoelectron spectroscopy (XPS) [47]. This value has been found to increase by as much as 1 eV as the CaF₂ deposition temperature was changed [48], and is not precisely known for our method of growth. Previously measured Al Schottky barrier heights depend sensitively on the semiconductor surface preparation and vary from about 0.69 [49] to 0.76 eV [50] on n-type Si(111).

Fig. 24 shows the current measured through the tunneling CaF₂ structures. These currents are many orders of magnitude larger than Table 1 would predict for degenerate direct tunneling (for barrier heights greater than roughly 1eV). Because of these large currents and the pitting of the Al surface, Pt contacts (e-beam evaporated through a shadow mask immediately after a dilute HF etch) were made to the 200/50 and 200/200 structures to determine whether the Al electrically-affected the barrier. J-V characteristics for the Pt samples were highly asymmetric with the polarity expected for a Pt/Si Schottky barrier. Of these two metals, Pt displays the higher barrier height (~0.84 eV [51]) to n-type Si(111) for a given Si surface preparation; its Schottky characteristic was therefore more evident. For a Pt gate voltage of -1 V (reverse bias), the current density (~0.1 A/cm²) was still many orders of magnitude higher than that
which would be expected for a quality CaF$_2$ barrier. We conclude that our method of growth produces an electrically-leaky barrier, regardless of the gate metal.

Nonuniform CaF$_2$ growth was the suspected cause for these large currents and TEM was performed on the 200/50 sample (after deposition of Al) to investigate this possibility. Low-resolution TEM (Fig. 25) showed rough polycrystalline CaF$_2$ and Si layers. At higher resolutions (Figs. 26 and 27), there appeared to be significant reaction and intermixing between the CaF$_2$ and Si; an interface was not well-defined. Dense and periodic stacking faults were evident (Fig. 27) in the overgrown Si. The quality of growth was indeed poor.

For 200/2000 devices, small currents (~5x10$^{-9}$ A/cm$^2$ at 25°C and ~5x10$^{-6}$ A/cm$^2$ at 200°C with a ±2 V bias) indicated that the bulk CaF$_2$ was insulating. It therefore appears that, for our method of growth, an extended (> 200 Å) conducting transition region between Si and CaF$_2$ is present. This transition region must be characterized further and, if possible, eliminated. C-V measurements of the 200/2000 sample were not
Figure 25. (112) Cross-Sectional TEM Bright-Field Micrograph of Tunneling CaF$_2$(111) Structure.

Figure 26. High-Resolution (112) Cross-Sectional TEM Bright-Field Micrograph of Tunneling CaF$_2$(111) Structure.
useful for quantifying CaF₂/Si interface properties because of the high substrate doping density (C-V was flat at 4.2x10⁸ F/cm² which indicated an approximate 1400 Å CaF₂ thickness).

The interfacial reaction discovered between CaF₂ and Si must be understood before double-barrier devices can be manufactured in this system.
Epitaxial growth of Si on ZnS is a very delicate process, but has been demonstrated. In doing so, since exposed ZnS is volatile at the high temperatures (~600°C) required for epitaxial growth of Si, an amorphous Si cap (~30 Å) was deposited at 200°C and subsequently crystallized (verified by \textit{in situ} RHEED) at ~650°C. Further deposition of Si proceeded epitaxially at 600°C. TEM and electrical measurements have demonstrated single-domain (to an approximately 20 nm extent) Si overgrowth by this procedure in the TAMU system [52].

Unknown factors, however, are presently causing the evaporation of the (supposedly) buried ZnS layer during the Si crystallization anneal and a fully-crystalline Si/ZnS/Si/ZnS/Si device has not been fabricated. In an attempt to overcome this problem, and produce a second barrier, a-Si(50Å)/ZnS(50Å)/As(1ML)/Si(100) samples were subjected to room-temperature ultraviolet (UV) radiation-promoted O\textsubscript{3} oxidation. For this process, O\textsubscript{2} was backfilled into a HV (10\textsuperscript{6} T) chamber to atmospheric pressure, and UV radiation from a Hg lamp, positioned in front of the wafer, converted O\textsubscript{2} into reactive O\textsubscript{3} (and other O species). This UV-O\textsubscript{3} treatment both removes C contamination [53] and oxidizes the wafer [54]. The oxide thickness obtained by this procedure is self-limiting at ~10 Å on crystalline Si [54]. After removal from vacuum and rapid thermal anneal (RTA), which ramped the temperature (~50°C/s) to 650, 700, or 750°C (with no hold time) in forming gas, Al gates were deposited on the oxidized surface as in §4.1. It was hoped that the RTA would crystallize the Si well without evaporating the underlying ZnS. Our hopes relied on some special unfounded effect of the overlying oxide to preferentially reduce the mobility of Zn and S atoms while the Si crystallized.

A TEM cross section of the oxidized sample is shown in Fig. 28. There was no noticeable difference between TEM photos obtained for samples of different RTA temperatures; the anneals did not crystallize the well as desired. The unexpectedly high (approximately 10x more than previously grown samples) density of ZnS stacking faults is intolerable for RTDs. The cause for this low-quality growth is believed to be due to contaminants (e.g., C), whose presence on the Si surface both nucleates stacking faults and makes the ZnS more susceptible to evaporation during the Si crystallization anneal.

\textit{J-V} characteristics for positive gate voltages (which correspond to electron emission from the Si; NDR is expected if the ZnS barrier and the Si well were single-
crystal) are shown in Fig. 29. The non-oxidized sample was etched in dilute HF immediately prior to metallization. Two oxidized/unannealed samples are plotted for reference; one had in situ Al deposited on it immediately after oxidation, without exposure to air, and the other was exposed to air between oxidation and metallization. Contamination and further oxidation which occurs upon exposure to air likely caused the 10x difference in current between these two samples. All annealed samples were exposed to air prior to metallization, and we therefore compare the four solid curves in Fig. 29. The significant (> 100x) decline in current with increasing RTA temperature is attributed to high-temperature H-passivation of dangling bonds in the amorphous materials. H is well-known for its ability to reduce the density of states in the energy gap of amorphous Si [55].

Although a RTD of this structure seems feasible, significant process and MBE development are required for its successful completion. First, the quality of the deposited ZnS must be enhanced, and the density of stacking faults reduced. Second, further post-oxidation RTA studies must be performed to find, if it exists, the temperature window
over which the ZnS remains stable and the Si crystallizes. It seems that this task would more likely, or as likely, be accomplished without the oxide, since it has previously been accomplished without the oxide by the TAMU group. Pursuit of a fully-crystalline Si/ZnS/Si/ZnS/Si RTD appears more worthwhile.

6 Conclusions and Future Work

Although attempts at fabricating an RTD on Si were unsuccessful, much has been learned in the process. We have taken preliminary steps in the characterization of two lattice-matched Si-based systems which have the potential for providing a Si-based RTD. For the first of these systems, ZnS/Si(100), temperature-dependent $J-V$ measurements have quantified crucial barrier heights and found them to be sufficiently large for room-
temperature RTD operation. We are confident that present difficulties with the growth of Si on ZnS will be overcome and that a crystalline Si/ZnS/Si/ZnS/Si(100) RTD will soon be manufactured.

For the second system, CaF$_2$/Si(111), an interfacial reaction between CaF$_2$ and Si has been identified but is not well understood at this time. Questions of interface chemistry must be answered so that higher-quality heterostructures can be manufactured. This will require an extensive effort in single-barrier structural characterization, and particularly TEM, as growth parameters are adjusted. TEM should also be used to ensure that processing (e.g., the HF etch or metallization) after Si/CaF$_2$/Si growth does not have a deleterious effect on the CaF$_2$/Si interfaces. Although $J$-$V$ measurements on single-barrier samples provide less (and less direct) structural information than TEM, the magnitude of the current provides an immediate indication of the quality of the barrier; in the interests of economy and efficiency, $J$-$V$ measurements should be performed on future CaF$_2$ samples, with comparative TEM then applied discriminantly to those sets of samples which show an unexpectedly large variation in current.

RTDs will inevitably be manufactured in these materials, but much time and research is still required to identify the necessary growth conditions.
Acknowledgments

I would like to thank all of my coworkers at Texas Instruments for their technical assistance. In particular, my company supervisor, Dr. Alan Seabaugh, has been the guiding hand in my education and provided invaluable support for me and my work. I can not overstate my gratitude towards Dr. Berinder Brar for his theoretical assistance and for providing many of the programs used to collect and analyze data. Dr. Roger Lake never failed to resolve my theoretical dilemmas. Discussions with Dr. Glen Wilk have been very enlightening. Dr. Gerhard Klimeck has always been helpful and supportive. The technical expertise of Pamela Stickney, Marshall Howell, and Robert Aldert have been crucial to the completion of this work. TEM preparation and microscopy was performed entirely by Phil Basham and Babak Amirhekmat, whose excellent and expedient work is greatly appreciated.

I am indebted to Professor Wiley Kirk, director of the ZnS program at the Center for Nanostructure Materials & Quantum Device Fabrication at Texas A&M University, for providing ZnS samples for my characterization and analysis. Dr. X. Joe Zhou’s discussions and correspondences provided invaluable information on the chemistry of the ZnS/Si system. Great thanks must also be paid to Dr. Shan Jiang, whose parallel and consistent characterization of devices contacted at Texas A&M ensured the accuracy of this research.

This work was all made possible and directed towards an academic end by my M.I.T. thesis advisor, Professor Qing Hu.
Appendix: Semiclassical Space Charge Theory

All calculations of chemical potential and capacitance within this dissertation were performed by using the well-known semiclassical equations of Seiwatz and Green [56] for the charge density (per volume), \( \rho \), and the sheet charge density (per area), \( \sigma \), in a semi-infinite semiconductor:

\[
\frac{d\sigma}{dz} = \rho = \sum_v qN_v f_{3/2} \left( \frac{E_v - \mu}{kT} \right) - \sum_e qN_e f_{3/2} \left( \frac{E_e - \mu}{kT} \right) - \sum_{A} qN_A (1 + g_A \exp \left( \frac{E_A - U}{kT} \right))^{-1} + \sum_{D} qN_D (1 + g_D \exp \left( \frac{E_D - U}{kT} \right))^{-1}
\]

\[
\sigma^2 = \frac{\sum_v N_v \left( f_{3/2} \left( \frac{E_v - \mu}{kT} \right) - f_{3/2} \left( \frac{E_v - \mu}{kT} \right) \right) + \sum_e N_e \left( f_{3/2} \left( \frac{E_e - \mu}{kT} \right) - f_{3/2} \left( \frac{E_e - \mu}{kT} \right) \right) + \sum_{A} N_A \ln \left( \frac{g_A + \exp \left( \frac{E_A - U}{kT} \right)}{g_A + \exp \left( \frac{E_A - U}{kT} \right)} \right)^{-1} + \sum_{D} N_D \ln \left( \frac{g_D + \exp \left( \frac{E_D - U}{kT} \right)}{g_D + \exp \left( \frac{E_D - U}{kT} \right)} \right)^{-1}
\]

where Laplace’s equation, \( \frac{d^2E}{dz^2} = \frac{\rho}{\varepsilon} \), was used in evaluating \( \sigma \). The terms in both of these expressions are ordered to account for holes, electrons (with sums over band extrema), ionized acceptors, and ionized donors (with sums over the impurity levels), respectively, with \( \{E\} \) being their associated energies.

\[
f_{v}(\lambda) = \frac{1}{\Gamma(n)} \int_0^\infty x^{n-1} \frac{d^2}{dx^2} \left( 1 + \exp(x - \lambda) \right) \lambda^{n-1} d\lambda \rightarrow \begin{cases} \exp(\lambda) & \text{for } \lambda \ll 0 \\ \frac{\lambda^n}{\Gamma(n+1)} & \text{for } \lambda \gg 0 \end{cases}
\]

is the Fermi-Dirac integral,

\[
N = \frac{2(2\pi m_e kT)^{3/2}}{\hbar^3}
\]

for electrons or holes \( (m_\omega^3 = m_m m_e) \), and \( \{g\} \) are the neutral (i.e., the number of ways in which the donor/acceptor can be neutralized) degeneracies of the impurity levels (assumed 2 for all calculations). The superscript “b” denotes bulk quantities which are determined by charge neutrality (i.e., \( \rho^b = 0 \)). The semiconductor capacitance \( C \) (per unit area) may then be calculated:

\[
C = \frac{q}{\varepsilon} \frac{d\sigma}{dE} = q \frac{d\sigma}{dz} \frac{dz}{dE} = q \rho \left( \frac{q\sigma}{\varepsilon} \right)^{-1} = \frac{\varepsilon \rho}{\sigma}.
\]

This gives the low-frequency capacitance; holes, electrons, and impurities respond to the signal.
For an uncompensated (e.g., donors, but no acceptors) semiconductor with a clear majority carrier (electrons), the high-frequency capacitance can be obtained from the same expression by simply omitting the terms in $\rho$ and $\sigma^2$ which correspond to the minority carrier. This provides an approximation for the capacitance in which only majority carriers and impurity states respond to the (small, high-frequency) signal.
References

[36] These results were obtained with the Texas Instruments Nanotechnology Engineering Modeling (NEMO) program.