AC Behavior of Resonant-Tunneling Diodes in the Scattering-Dominated Regime

by

John Paul Mattia

S.B., S.M., E.E., Massachusetts Institute of Technology

Submitted to the
Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1996

Author____________________________________
Department of Electrical Engineering and Computer Science
September 6, 1996

Certified by____________________________________
Alan L. McWhorter
Professor of Electrical Engineering
Thesis Supervisor

Accepted by____________________________________
Frederic R. Morgenthaler
Chairman, Committee on Graduate Students

AC Behavior of Resonant-Tunneling Diodes in the Scattering-Dominated Regime

by

John Paul Mattia

Submitted to the Department of Electrical Engineering and Computer Science
on September 6, 1996, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The behavior of a resonant tunneling diode (RTD) whose electron transport is dominated by scattering is explored. Measurements from 45 MHz to 26 GHz of the small-signal ac behavior of such a device, as well as a model to explain the static and dynamic behavior, are presented. After the parasitics are extracted from the measured results, the intrinsic RTD response is plotted for frequencies of 45 MHz to 4 GHz at 16 biases. The real part of the admittance, the conductance, exhibits a rolloff at a bias-dependent frequency. Since the imaginary part of the admittance, the susceptance, is nearly that of a capacitor, the deviation from a purely capacitive susceptance is shown. This deviation exhibits a peak at a frequency which corresponds to the rolloff in the conductance. The dc and ac behavior are described theoretically with a rate-equation model. In this model, the RTD is divided into three regions. The current between two states in adjacent regions is proportional to the probability that the initial state is occupied and the final state is empty. The transition coefficient for this transport process is calculated using a semi-classical model for scattering. In order to evaluate the agreement of the model with the data, a numerical analysis is performed using three methods. The first method uses the specified growth parameters for the RTD layers, which yields a predicted ac response that is qualitatively similar to the actual response. It also yields a simulated current peak that is 2.3 times that of the measured peak. In the second method, the barriers are adjusted so that the simulated and measured peak current is the same. The simulated ac behavior in this case is in better agreement with the measured behavior. The final method incorporates the conductance at three frequencies, which yields values for several physical parameters in the RTD as well as the the best agreement with the ac data.

Thesis Supervisor: Alan L. McWhorter
Title: Professor of Electrical Engineering

Acknowledgments

I began writing a version of the acknowledgements which, I felt, gave proper thanks to the people who made direct or indirect contributions to this work. However, I quickly realized that I was on the road to being the first person in history to have an acknowledgements section which was longer than the body of the thesis. Although such a thesis would probably be more entertaining and achieve much greater readership, I hope that my mentors, colleagues, friends and family will accept this abbreviated version instead.

I am extremely indebted to my advisor, Prof. Alan McWhorter. It is rare that a student has the opportunity to be mentored by someone with this level of scientific insight, knowledge, and experience, who also happens to have the wisdom from steering the Solid State Division at Lincoln Laboratory over 30 years. He has patiently instructed me on many scientific (and managerial) matters, and his insight continues to amaze me. I am also grateful to him for originally suggesting the rate-equation approach to me. His retirement will be great loss for MIT and Lincoln, and I wish him well in his future endeavors.

I am grateful to Profs. Qing Hu and Terry Orlando for agreeing to be readers for the thesis and providing additional guidance for this work. I have also very grateful for my past experiences teaching under Prof. Hu.

I would like to thank the people of Lincoln Laboratory, who provided me with unique opportunities during my time in graduate school and as an employee. I am also grateful to Lincoln for my funding, and in particular I thank Dave Shaver for continuing my funding at a critical time. I have benefited by several conversations with Prof. Benjamin Lax, who was instrumental in motivating me to return to MIT for graduate school.

Group 84 at Lincoln Laboratory was an integral part of this thesis. Chris Parker, Alex Mcintosh, Rick Mathews, and Paul Cociani showed me the ropes in the microwave lab, while demonstrating terrific senses of humor. I am grateful to Karen Molvar, for her expert processing of the devices used in this work and cheerful attitude. I thank Elliott Brown for originally getting me started in RTDs. This work has also benefited by conversations with Al Murphy, Mark Hollis, and Bob Calawa.

Several people in other groups at Lincoln provided me with their valuable time, equipment or terrific advice. Paul Maki grew the world-class RTDs measured in this work, and his MBE made the experiment in this work possible. George Young, Lincoln's resident UNIX wizard, provided answers to my many computer questions, and Roy McMorran provided HP computer access. Bor-Yeu Tsaur, Alan Strauss, George Turner, Hong Choi, and Steve Eglash have also been encouraging and helpful over the years.

My transition to the microwave world was aided by Jim Bales, who was the senior graduate student when I first arrived in the microwave lab. Jim is one of those people who knows something about everything, and if you don't believe me, you can challenge him in the College Bowl tournement. I must express my gratitude to him for proofreading critical parts of this document, and I owe him thanks (with compounded interest) for omitting his name from my masters thesis.

I have benefited from my association with many students at MIT. Originally, I met Jay Damask, Farzana Khatri, Farhan Rana, Thalia Rich, Krishna Shenoy, and Mark Somerville while I was TAing. Teaching graduate students around here is a dangerous task, because they are quite sharp. In order to maintain the dignity a TA requires to teach, it was necessary for me to employ the take-them-to-the-Muddy-Charles strategy, which was surprisingly effective, especially in Jay's case. Either nobody noticed my tactics, or they encouraged them, or they forgave me for them, but their friendship was valuable in any case.

It would be a dull graduate education if all of my time was spent on the thesis. Several people in Ashdown House conspired to make sure this didn't happen. Some of the major conspirators were: Tara Arthur, Juan Bruno, Eddie Chalom, Yonald Chery, Kath Holden, Pam Ku, Tom Lee, Cedric Logan, Elliot Mack, John Matz, Kathy Misovec, John Shirlaw, Richard Stone and Pat Walton. S. S. Papa Rao should be given a special commendation for the daunting task of living with me, and I hope he recovers from the experience. It took me a while to figure out that Vernon and Beth Ingram are actually the senior graduate students in the dorm, and simply use humor, wisdom and great food to disguise themselves as the housemasters. Sorry for blowing your cover.

Many people have made indirectly contributions to this work. Jen Lloyd, Joe Lutsky, Biz Williamson and Jeff Gross have provided great hikes. Andrew Muenz, Roy Briere, Jonathan Root, Sue Paul, Adam Dershowitz, Pete Andrews, Debbie Lew, and Dave Ernst have provided humor and good times on demand. I am also thank Adam for doing his part to reduce homelessness by providing me with housing during the last part of this work.

I must express my sincere gratitude to Thomas Margolis for sending many songs that were joyful yet excoriating, Prof. William Silver for a monk's instruction in the six foot staff as well as the art of circling a pole, Peter Snyder for his race's part in merrily destroying the Ring, Robert Hackett for leaving his dwarven elbow out of my ribs because I've seen the structural damage it can cause. I'd also like to thank Miller for the secret recipe for Glug. I thank Eric Roth for some of my early training; I regret he won't see the results.

A few people have shared more than the average amount of my good and bad times, and it is simply not possible to give them the thanks they deserve. I would not be the same without the sharp wit of Judy Chen attacking me when I least expect it. Ken Yee has created an art form out of observation that is both extremely humorous and cynical. Raj Aggarwal graciously allowed me an exemption to the Yankee Rule of Texas and admitted that (perhaps) I'm much better as a live Yankee. I am also indebted to Mike Polley for strongly suggesting that she remove this statement from her acknowledgments. This work would not have been possible without your constant encouragement and good humor. Thank you all, I wish you well in everything you do, and I hope that I will have the chance to repay some of the debt that I owe you.

It is said that the acorn does not fall far from the tree, and I am blessed to have a family with such large branches and great roots. My father, mother and sister have provided love and support over the years, not to mention a golden opportunity to hone my debate skills over the dinner table. My grandmother Helen and godmother Aunt Terry have provided encouraging words at perfect moments during my education.

Last, but certainly not least, I thank Laura Adams for her love and support. Anyone who could tolerate such large doses of me (not for the faint of heart, I assure you) and still be supportive, motivating, and willing to contribute expert corrections to my thesis is a likely candidate for canonization. I am grateful that she was there, especially when Bear got out of control.

Contents

1

Introduction

In this thesis, we explore the dynamic behavior of the resonant tunneling diode (RTD) when the transport through the device is dominated by scattering. This type of transport occurs in many practical RTDs, especially those that are operating at room temperature. Many proposed applications [1, 2], such as millimeter-wave oscillators [3], analog circuits [4], and binary [5, 6] or multiple-value [7, 8, 9] logic circuits, will utilize the speed of the RTD. Therefore, it is important to have an accurate dynamic model for scattering-dominated transport.

We present a rate-equation model for the RTD that is both accurate and computationally efficient, and provides insight into the operation of the device. The model can be used to predict the dc and ac behavior of an RTD before its fabrication or it can be used to extract physical parameters from an RTD after fabrication. We demonstrate that it may also be used to predict the ac response from the measured dc behavior, which is useful because measuring the static behavior requires much less experimental effort. The predictions of the model were compared to the results of an ac measurement, which used well-established admittance techniques to find the intrinsic response of the RTD.

Our presentation will begin in this chapter with a review of some of the relevant literature. We will first review some prior work on the static properties of the RTD and then some of the work on the dynamic behavior. The literature has become quite large as the RTD field has matured, so this review is necessarily restricted to an abbreviated list of papers that illustrate prior work.

In the second chapter, we describe our measurement of the ac response of the RTD. Our devices were specifically designed to avoid deficiencies in previously published work, and the manufacture of these devices is described in Section 2.1. The next section discusses the apparatus used to measure the small-signal admittance at frequencies from 45 MHz to 26 GHz. Section 2.3 describes the method used to remove parasitics from the admittance, which allowed the intrinsic admittance of the RTD to be found. The final section of Chapter 2 shows the intrinsic ac response at 16 different biases for a representative device.

Chapter 3 is devoted to the theoretical modeling of the RTD. In Section 3.1, we present a model for transport in the scattering-dominated limit. In this limit, electron transport occurs only with the aid of a scattering event. We derive expressions for the current and charge for the static case and examine the variation in these quantities for the sinusoidal steady state. Finally, a method is shown for calculating the time dependence of the of an electron wavefunction in a scattering-dominated RTD.

In Chapter 4, we compare the data presented in Chapter 2 with the model presented in Chapter 3. We implemented a computer program to numerically solve Poisson's and Schrödinger's equations in a self-consistent manner. The details of the numerical techniques are presented first. Next, we examine some of the simulated properties of our manufactured RTD. The numerical analysis can be performed using differing amounts of knowledge of the device, and we show three methods of performing this analysis. The first method uses neither ac nor dc data and yields qualitative agreement with the experiment. The second method, which incorporates only dc data, gives reasonable agreement between the model and experiment. The third incorporates both ac and dc data, and this method yields the best agreement between the model and the experiment.

Finally, in Chapter 5, a summary of conclusions is presented as well as some rec-

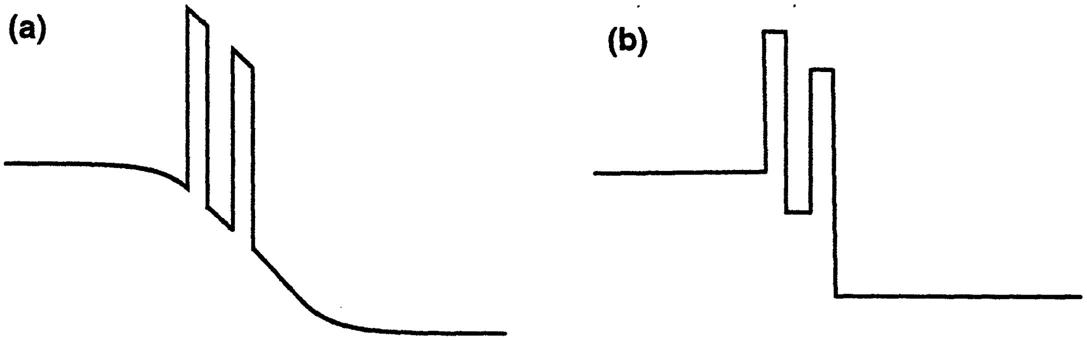

Figure 1-1. Potential profile of a typical RTD with an applied bias. (a) Actual profile. (b) Staircase approximation to the actual profile, which simplifies analytical calculations considerably.

ommendations for refinements to this work. We note that several detailed calculations have been placed in the appendices, in order to not interrupt the flow of the main text.

1.1 Review of Previous Work

In the simplest terms, an RTD is an electronic device with a potential profile that has two closely spaced barriers, as shown schematically in Fig. 1-1. In this system, there is an electronic state where much of the electron wavefunction is confined between the barriers. The energy of this state is denoted as the resonant energy. At zero bias, the resonant energy is presumed to be above the conduction band edge of the cathode E_{C1}. As the bias is increased, the resonant energy is lowered, and the current increases. At still higher biases, the resonant energy drops below E_{C1}, and the current decreases.*

---

*We will see in Chapter 4 that the current actually begins decreasing while the resonant energy is still above E_{C1}.

We will now review some of the literature pertaining to this device. We will first discuss the static properties and then the dynamic properties.

Static RTD Models

A large number of publications have focused on the static properties of the RTD, and most of this work is concerned with finding the I-V characteristic for a given structure. The basic approach for calculating the I-V characteristic in the absence of scattering was first presented by Tsu and Esaki [10]. The Tsu-Esaki approach breaks the calculation of the I-V characteristic into two parts. The first part of the calculation considers an incident electron wavepacket of energy E, for which the quantum-mechanical transmission amplitude $t(E_\ell)$ is calculated, where E_ℓ is the longitudinal part of E. Then the total current J_{RTD} is calculated by integrating the product of three factors: the velocity in the tunneling direction, the difference between occupation factors f_c and f_a on the cathode and anode sides of the RTD, respectively, and transmission probability at each value of E_ℓ. Thus, the current is written as

$$J_{RTD} = -\frac{e}{4\pi^3\hbar}\int_0^\infty dk_l \int d^2k_t \,\frac{\partial E}{\partial k_l}\,[f_c(E) - f_a(E)]\, t^* t\,, \tag{1.1}$$

where k_l is the longitudinal momentum, and d^2k_t is the differential element of the two orthogonal transverse momenta, with the integral being taken over the entire transverse momentum space. When using Eq. (1.1), the bulk of the computational effort arises from finding the transmission probability.

A common method of computing $|t|^2$ employs the transfer matrix, which is often abbreviated as the T matrix. This approach for solving the RTD problem was described by Ricco and Azbel [11]. We will examine this technique in some detail, since it was used extensively for calculations in this thesis.

The actual RTD potential profile, shown schematically in Fig. 1-1(a), is often

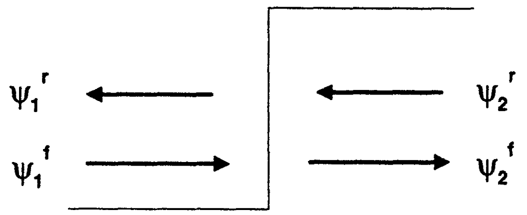

Figure 1-2. Forward- and reverse-traveling wavefunctions used with the T matrix.

approximated using the so-called "staircase approximation", shown schematically in Fig. 1-1(b). This approximation is made in analytical calculations to simplify the expressions. It is also made for numerical calculations, which use several hundred stair steps. The numerical calculation of the wavefunctions may be refined by dividing the potential profile into finer stair steps and then utilizing the T matrix at each boundary.

Consider forward- and reverse-traveling wavefunctions ψ_1^f and ψ_1^r on the left side of the boundary shown in Fig. 1-2, as well as the corresponding wavefunctions ψ_2^f and ψ_2^r on the right side of the boundary. The T matrix is defined so that the right- and left-side wavefunctions can be related at the boundary by matrix multiplication,

$$\begin{bmatrix} \psi_1^f \\ \psi_1^r \end{bmatrix} = \begin{bmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{bmatrix} \begin{bmatrix} \psi_2^f \\ \psi_2^r \end{bmatrix} . \tag{1.2}$$

The elements of the matrix are easily found by matching the wavefunction and the particle current at the boundary.

Ricco and Azbel use their analysis to describe the transmission through the double-barrier structure. The transmission amplitude t has a resonance, which may be expanded into the Breit-Wigner form

$$t \propto \frac{i\Gamma/2}{(E - E_R) + i\Gamma/2} . \tag{1.3}$$

This transmission amplitude has a peak at energy E_R and a full width at half maximum (FWHM) in energy of Γ. They show that Γ is inversely proportional to the escape time τ_0 of an electron placed in the well. Fundamentally, $\tau_0 = \hbar/\Gamma$.

In a more general fashion, Gerjuoy and Coon [12] noted that the transmission properties and time delays of resonant tunneling are well-understood phenomena from scattering theory, and they applied the S-matrix formalism. They also demonstrated the relation of the transmission width and electron lifetime in the well. Their model subdivided the process of tunneling through the double-barrier potential into two parts: tunneling into the well from the cathode, and tunneling out of the well into the anode. There is a lifetime associated with each of these two processes, which combine to form the overall lifetime.

The I-V characteristics calculated from the above models often achieve a qualitative agreement with experimentally obtained I-V characteristics. In order to obtain quantitative agreement, more sophisticated modeling must be performed. For example, scattering in the well must be included. In addition, an RTD under bias has an accumulation layer between the cathode contact and the first barrier that electrons encounter. It is not an easy task to add the effect of this layer to an analytical model, which is one reason why the staircase approximation is often made. The algebra in solving staircase problems yields closed-form solutions in terms of exponential functions, but this approach is not practical when including the accumulation layer. Two authors addressed this problem with numerical calculations [13, 14]. Their calculations self-consistently solve both Schrödinger's and Poisson's equations. The results show that the staircase approximations change the shape of the calculated I-V characteristic and tend to underestimate the applied bias at which the peak current occurs.

Also, the accumulation layer forms a potential well, so it often has one or more quantized states. It has been experimentally shown that such states can cause a kink in the I-V characteristic at lower biases [15]. These states have been probed by both photoluminescence (PL) [15] and magnetotunneling [16].

The methods for calculating $|t|^2$ discussed to this point yield results that are poor approximations for the negative-differential resistance (NDR) region. Once the RTD has been biased sufficiently, there is no way for an incident electron to conserve total energy and transverse momentum while tunneling through the resonant level. Thus, the conventional tunneling current becomes vanishingly small, and other mechanisms become responsible for the "excess current." The figure of merit for the RTD is the peak-to-valley-current ratio (PVCR), whose inverse roughly indicates the amount of excess current relative to the conventional tunneling current.

One way of calculating the excess current is by introducing an optical-absorption-like parameter. It has been shown [17, 18] that an absorption parameter enhances current at biases higher than the peak. Obviously, the amount of absorption will depend on the underlying physical mechanisms which are responsible for the current flow, and we will briefly examine three such mechanisms: disorder-assisted tunneling, alternate-valley tunneling and phonon-assisted tunneling.

Disorder-assisted tunneling [19] is an elastic-scattering mechanism due to interface roughness in the sample. Transverse momentum need not be conserved in such structures and can open channels to the resonant level. Reference [19] considered structures with terraces, and showed that this tunneling mechanism both changes the shape of the I-V characteristic and contributes to the valley current.

Alternate-valley tunneling is a process in which conduction band minima other than the lowest minimum play a role. This mechanism occurs in a typical GaAs/AlAs double-barrier structure, where most of the incident electrons have **k** values near the Γ point. Although the AlAs presents a large barrier at the Γ point, it does not do so at the X point. The effective-mass approach to solving for the transmission coefficient would only use the barrier presented at the Γ point, but a more complete calculation [20] shows that there are transmission resonances associated with the X valleys.

Phonon-assisted tunneling has been treated by Wingreen *et al.* [21]. They began with a Hamiltonian that included an electron-phonon interaction and calculated the transmission function and current. For an Einstein band of phonons (a reasonable model for optical phonons, the dominant scattering mechanism in III-V compounds) the transmission function has sidebands separated by the phonon energy. These sidebands create new tunneling channels for electrons to flow once the resonant energy drops below the cathode conduction band edge.

Small-Signal Response of the RTD

The development of high-quality RTD structures has yielded impressive high-speed results: Whitaker *et al.* [22] demonstrated a switching time of 2 ps, and Brown *et al.* [3] demonstrated a maximum oscillation frequency of 712 GHz. As a consequence, there has been a large amount of experimental and theoretical work aimed at understanding the speed limitations of the RTD. We shall discuss only small-signal ac measurements and modeling, since that is the focus of this thesis.

Several authors have published measurements of the ac response. Many workers [23, 24, 25, 26, 27, 28] use standard microwave reflection techniques, and a few use standard LCR bridges [29, 30, 31, 32]. There is also one publication [33] that used a free-electron laser to probe the RTD at terahertz frequencies. Unfortunately, there are several methods of presenting the data, and the above publications contain the following plots: S_{11}-ω, Y-ω, Z-ω, C-V, and G-V, where S_{11} is the single-port scattering parameter, and Y,Z,C,and G are admittance, impedance, capacitance and conductance, respectively. The differing forms of data presentation has made it difficult to quantitatively compare the different measurements. Qualitatively, however, the previously published measurements are consistent with our initial measurements [34]. We have chosen to present our data in terms of the admittance. This choice will simplify the task of plotting since the admittance remains finite at the peak voltage where the dc conductance goes to zero.

Small-signal ac measurements on RTDs are often a difficult task because the intrinsic speed of the devices interferes with their characterization. Devices with high PVCRs tend to oscillate or exhibit switching behavior in the NDR region, which eliminates the possibility of a direct admittance measurement at those biases. Thus, there is a tradeoff between the PVCR and the ease of measurement. Several previously reported admittance measurements have been performed either on devices that had high PVCRs but exhibited switching or oscillatory behavior in the NDR region [35, 36, 37], or on devices that were free of switching or oscillatory behavior, but lacked an NDR region [38].

The literature on ac modeling of the RTD is extensive. Some authors [31] simply feed their data into a CAD program and the program outputs an equivalent circuit which closely matches the observed behavior. Other authors [23, 26, 27, 28] used an equivalent circuit model from the Esaki diode. Such approaches lack an underlying physical basis which is desirable for the purposes of this thesis.

Other publications build more physical models of the RTD dynamics using differing approaches. Frensley performed a numerical calculation that utilized the Wigner function [39]. This approach incorporates the quantum statistics of the device. However, there is a large amount of the computation time associated with the Wigner function, which makes it unsuitable for circuit-level simulation.

Brown *et al.* [40] presented a model which began with the following assumption: when the bias applied to an RTD is changed abruptly in time, the RTD current does not change instantaneously and approaches its new value exponentially in time. Such behavior can be modeled by an inductance in series with a resistance. This approach yields a simple dynamic model that yields qualitative agreement with measurements of the microwave oscillation power. The admittance behavior is determined by the transmission width of the RTD, which is roughly constant with bias. As we shall see, the speed of scattering-dominated transport is in fact bias dependent, so this model will not suit our purpose.

Another approach, by Hu and Stapleton [41, 42], began by assuming that the electrons in the well have a hot-electron distribution. They solved for the transmission with an absorption parameter [17, 18], and assumed that "absorbed" electrons are left in the well. They then assumed that the Fermi level and temperature of the electrons in the well are uniquely determined by energy and particle conservation. The authors used the result to predict how the charge storage will change with bias, but did not compare their predictions with any experimental data.

In another approach, Fu and Dudley [43] used linear response theory [44] to predict the ac behavior, and derived an ac equivalent of the dc Landauer-Buttiker [45] formula. They predict an inductive response, but do not compare their results with experimental data. Jauho *et al.* [46, 47] use a nonequilibrium Green's function technique (the Keldysh formalism) to solve for the ac response. This technique includes a method for coupling electrons and phonons. They predict that an RTD which is biased in the NDR region will have a conductance which is negative at lower frequencies and positive at higher frequencies. We will see experimental evidence for this behavior in the following chapter, but this theoretical approach is unnecessarily complicated to explain the results for the scattering-dominated case.

Comparing the models of Hu and Stapleton, Fu and Dudley, and Jauho *et al.* with the RTD behavior measured here is not a simple matter. Such an undertaking has not been addressed in the present work.

2

Experimental Results

In this chapter, the results of a small-signal ac measurement on an RTD will be presented. The goal of the experiment was to characterize the admittance of an RTD using well-established microwave measurement techniques. By using these techniques, we minimized the uncertainty in the interpretation of the data.

In Section 2.1, we will discuss the competing goals in the design of the RTD, and we will describe the devices that were fabricated for the experiment. In Section 2.2, we will describe the microwave scattering measurement as well as the typical data that it yields. In Section 2.3, we will examine how the scattering data can be used to separate the response of the RTD from that of the parasitics. Finally in Section 2.4, we show the intrinsic RTD response for a representative device. We will also examine the change in the intrinsic response with bias.

2.1 Device Design and Fabrication

In order to measure the response of the RTD, a fabrication process was designed to produce devices that could be probed at high frequencies. The structure contained an RTD as well as the mechanical support necessary for a coplanar probe. The center wire was connected directly to the top of the RTD mesa, and the two ground strips were bent slightly downward and connected to the RTD substrate. This approach simplified the design, since no on-chip transmission lines were needed to connect the probe to the device. Also, this approach simplified the task of modeling the parasitics.

The design of the RTD took a number of considerations into account. Ideally, we would like the RTD to have three properties: excellent PVCR, high speed, and stability in the NDR region. Unfortunately, devices with high PVCRs are hard to stabilize in the NDR region. Also, faster devices are harder to measure since parasitics play a larger role in the response. We addressed these issues by designing the devices to have sufficiently low PVCR to be stable in the NDR region. We also designed the devices so that they were slow enough to be easily measured. Both of these decisions required an RTD with relatively thick barriers.

The design of the wafer is shown schematically in Fig. 2-1(a) and (b). The layers were grown by gas-source molecular beam epitaxy on n^+-InP substrates. A 500-nm-thick heavily doped (n-type, $N_D = 2 \times 10^{18}$ cm^{-3} Si) $In_{0.53}Ga_{0.47}As$ layer was grown first, followed by a 100-nm-thick lightly doped (n-type, $N_D = 2 \times 10^{16}$ cm^{-3} Si) $In_{0.53}Ga_{0.47}As$ spacer layer. The double-barrier structure was then grown and consisted of nominally 4.5-nm-thick undoped AlAs barriers and a 5.5-nm-thick undoped $In_{0.53}Ga_{0.47}As$ quantum well. Above the double-barrier structure, a 10-nm-thick lightly doped (n-type, $N_D = 2 \times 10^{16}$ cm^{-3} Si) $In_{0.53}Ga_{0.47}As$ spacer layer was deposited and a final 400-nm-thick layer of heavily doped (n-type, $N_D = 2 \times 10^{18}$ cm^{-3} Si) $In_{0.53}Ga_{0.47}As$. We note that the layers shown in Fig. 2-1 are specified with their nominal thicknesses. Because of strain, the AlAs barrier thicknesses are approximately 4.1 nm [48]. We also note that the grading from heavy doping in the ohmic regions to light doping in the spacer layers occurs over approximately 10 nm. These graded-doping regions occur outside of the layers shown in Fig. 2-1(a), so that the spacer layers are guaranteed to be lightly doped for at least the indicated width.

The lightly doped spacer layers are present to improve the electrical characteristics of the device. Lighter doping reduces the number of dopants near the double-barrier structure. Dopants in the proximity of the barrier would locally perturb the barrier

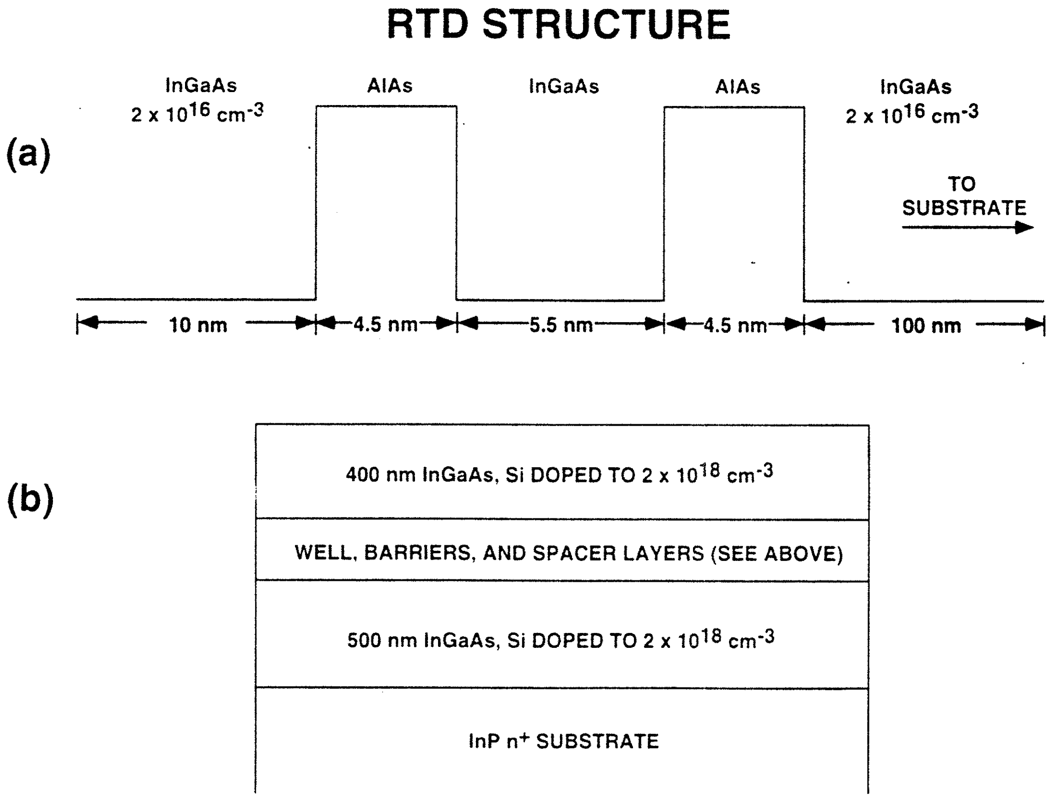

Figure 2-1. Schematic of layers grown for the RTD structure. (a) The double-barrier structure and spacer layers, and (b) the ohmic regions.

potential. It has been shown [19] these potential variations give rise to a larger valley current.

The 100-nm spacer layer is lightly doped for a second purpose as well. For the data shown in this work, we will bias the structure depicted in Fig. 2-1(b) so that the top is grounded and the bottom is at a positive potential. For the biases that will interest us, the 10-nm spacer layer will contain an accumulation layer and the 100-nm spacer layer will be depleted. By lightly doping this 100-nm spacer layer, we lengthen the size of the depletion layer. Thus, the depletion capacitance is reduced. We will find it useful when measuring the admittance of the RTD to have as small a depletion-layer capacitance as possible.

Once the above layer structure was grown, the remainder of the fabrication process was designed to provide contacts and mechanical support for the coplanar probes. The support structure was designed to introduce as few parasitic electrical components as was feasible.

Figure 2-2 shows a schematic cross section of the completed design. A layer of Pd-Ge-Au was deposited to serve as a mask for the mesa definition and as an ohmic contact to the completed mesa. The final height of the mesa was approximately 800 nm. After the mesa was defined, a commercial dielectric called Futurex was deposited and wet etched until the underlying metal became exposed. Ti-Au-Ni was deposited and patterned by a lift-off. A wet etch then defined the underlying dielectric. Finally, Pd-Ge-Au was deposited and patterned to provide contact to the bulk regions.

The mask set defined several hundred square mesas that had nominal sizes of 2, 4, 8, 16, and 24 μm on a side. The top metal pad had dimensions of 30×36 μm, and the Pd-Ge-Au field metal was horizontally separated from the top metal pad by approximately 2 μm.

The current-voltage (I-V) curve and the dc differential conductance of a typical device at room temperature are shown in Fig. 2-3. The peak-to-valley current ratio

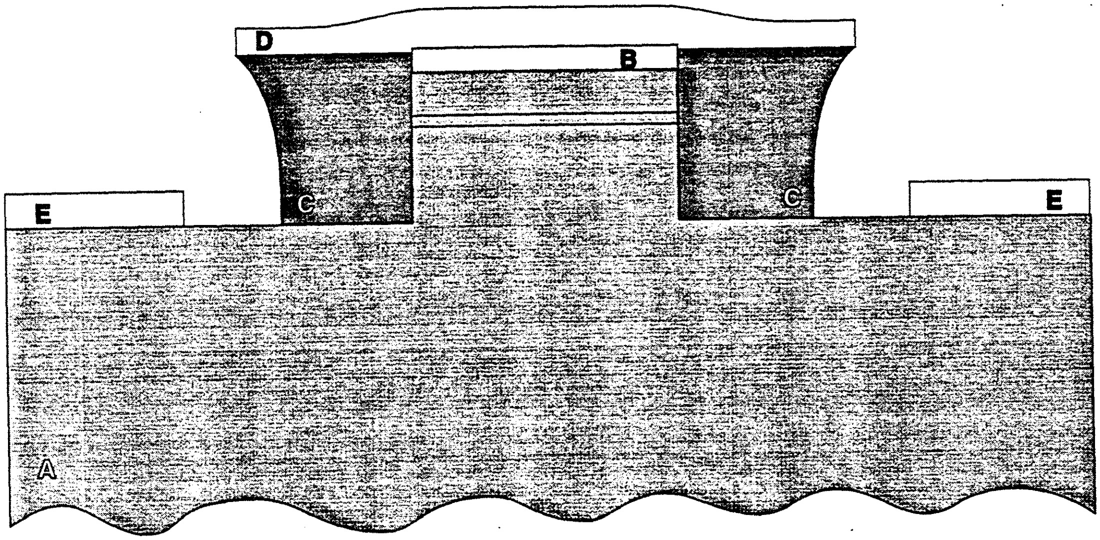

Figure 2-2. Cross section of the finished RTD structure. The layer materials are indicated as follows: A. $In_{0.53}Ga_{0.47}As$ bulk and mesa; B. Pd-Ge-Au metal, which served as mesa etch mask and the top contact; C. Futurex dielectric, which is the mechanical suppport for the mesa probe pad; D. Ti-Ni-Au metal, which served as an etch mask for the Futurex dielectric and as the microwave probe pad; E. Pd-Ge-Au, which was the electrical contact to the bulk regions.

is 2.81. Since the I-V is smooth and continuous in the NDR region, the device is free of oscillations and exhibits no switching behavior.

2.2 Reflection Measurement of the Mesa

We performed a reflection measurement in order to find the ac response. The measurement used well-established wafer probing techniques that were suitable for microwave frequencies. The apparatus for finding the microwave response is shown schematically in Fig. 2-4. The ac signal and the dc bias were delivered to the RTD structure by a coplanar microwave probe. The reflection coefficient S_{11} [49] was measured using a network analyzer (an HP8510) at frequencies ranging from 45 MHz to 26 GHz. Once S_{11} was known, a transformation was applied to find the admittance, which is the ac quantity we shall be considering throughout this thesis. The network

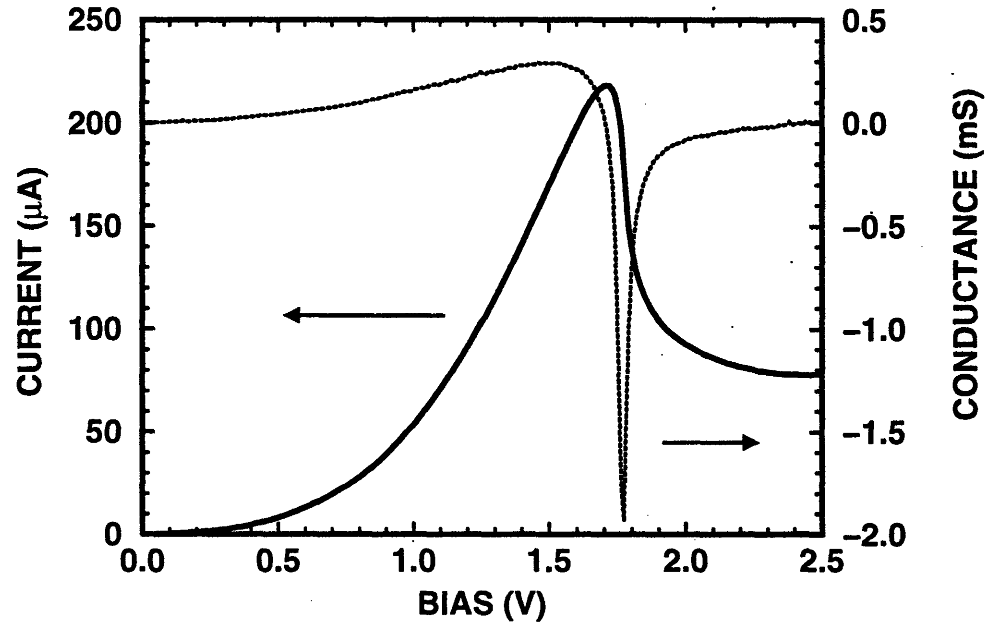

Figure 2-3. I-V curve and dc differential conductance for a typical fabricated RTD. The ground potential reference is the metal labeled B in Fig. 2-2. The device area was 256 μm^2.

analyzer was calibrated using traditional coplanar probing techniques on a commercial substrate. The small-signal ac measurements were performed with an incident power of -30 dBm, which corresponds to an rf voltage of approximately 7 mV. A semiconductor parameter analyzer (SPA), the HP4145, provided the bias for the network analyzer.

Since we chose an RTD with relatively thick barriers, the device presents a high impedance termination to the transmission line. As a consequence, there is a large amount of scatter in the data because the network analyzer is resolving reflection coefficients that are close to 1. The data can be improved by choosing devices which have a large area, and hence a lower mismatch. However, devices which are too large will exhibit switching behavior in the NDR region. (Such switching behavior can be deduced from a discontinuity in the I-V data.) In light of these considerations, most of our admittance measurements were performed on devices which were nominally 16 μm on a side. The measurements on devices of this size showed no switching

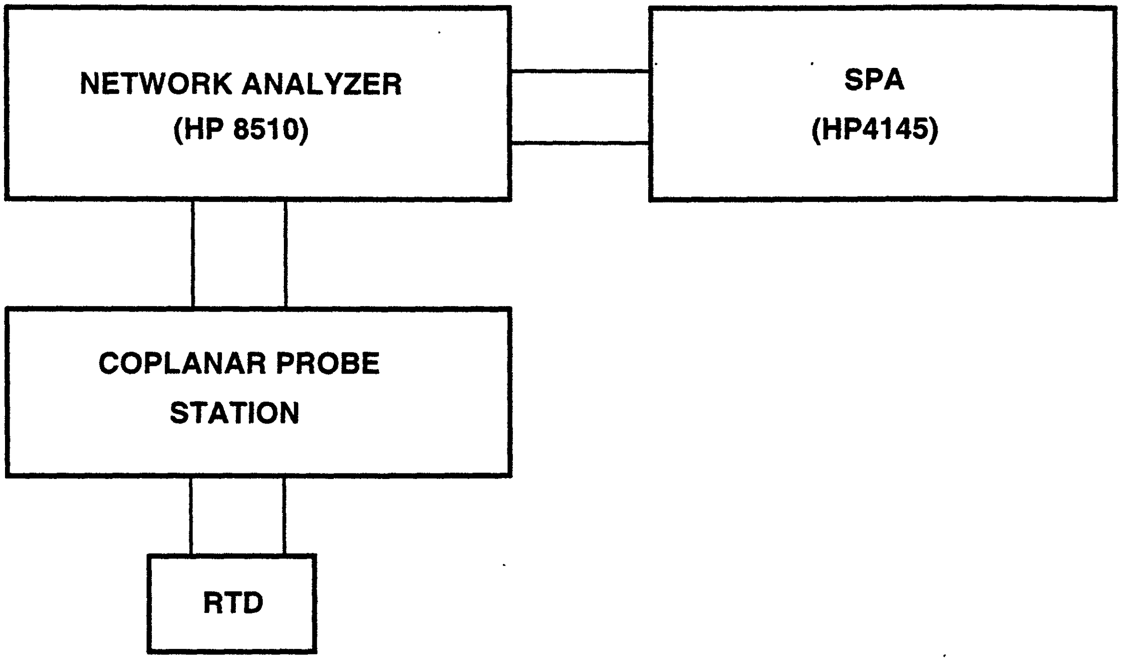

Figure 2-4. Schematic of experimental apparatus used to measure the reflection coefficient S_{11}. The semiconductor parameter analyzer (SPA) sets up a bias for the network analyzer. Both the network analyzer and the SPA are computer controlled.

behavior, and they yielded some improvement in the amount of scatter in comparison to the smaller size devices.

We employed averaging to further reduce the scatter in the data, which substantially lengthened the time required to perform a measurement. It became necessary to automate the entire measurement process, so that many frequency points and biases could be measured and recorded. At a minimum, each device was measured at 30 bias points, and at least 3 scans of 200 frequencies were performed. For the plots shown in this work, a minimum of 512 points were averaged at each measured frequency. In order to measure this volume of data, each device required about two hours to characterize.

The most reliable data were taken from a total of 10 devices which were in roughly the same location on the wafer. These devices were quite similiar in their dc and ac

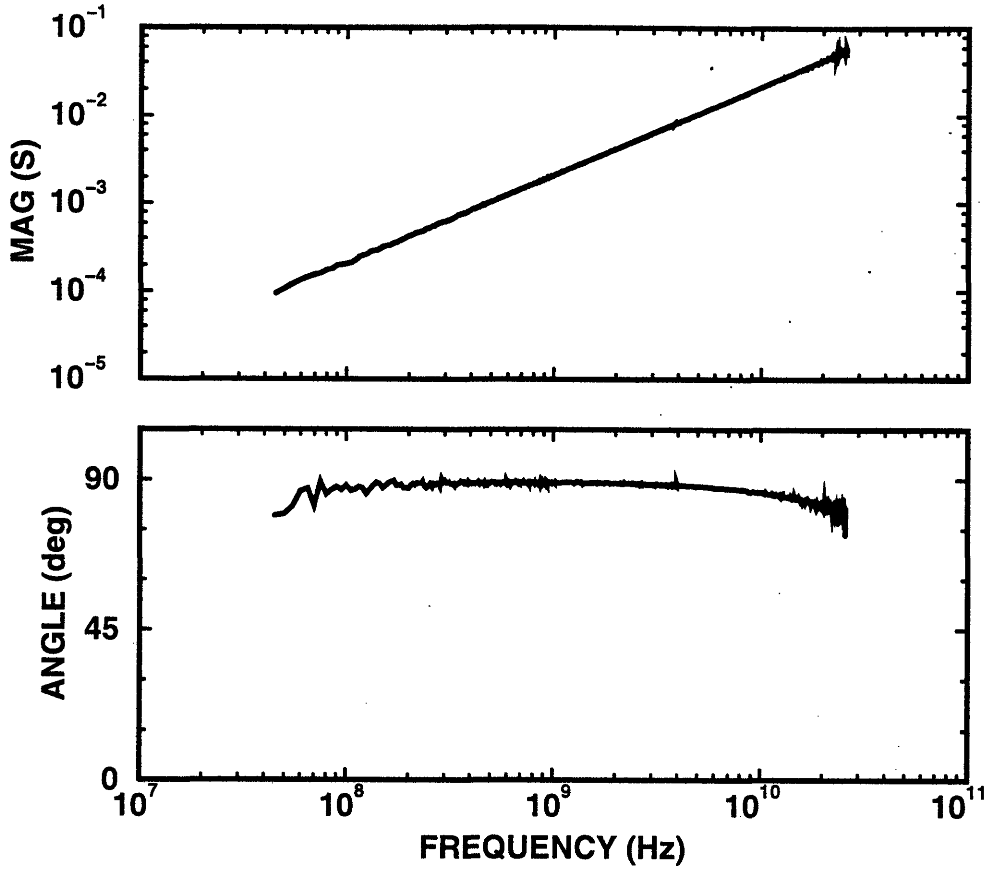

Figure 2-5. Magnitude and phase of the admittance vs frequency for the representative RTD and its associated parasitics at a bias of 0.2 V. These data and the data shown in Fig. 2-3 are from the same device.

electrical characteristics. For brevity and consistency, we will focus on only one particular device throughout this work. The I-V curve for this representative device was already shown in Fig. 2-3.

A plot of the admittance at a bias of 0.2 V for the representative device is shown in Fig. 2-5. As can be seen from the figure, the magnitude varies roughly linearly with frequency. Also the phase is close to 90 degrees for most of the frequencies measured. For other biases, the behavior above 10 GHz was qualitatively similiar; we will discuss this behavior in the next section. For frequencies less than 5 GHz, there were significant differences; this behavior is the focus of the remainder of this thesis.

2.3 Characterization of the Parasitics.

In order to extract the intrinsic RTD admittance from the measured reflection coefficient, it was necessary to characterize the electrical parasitics. We began this characterization by identifying the parasitic circuit topology associated with the mesa structure. Next, we made reasonable assumptions about the low-bias, high-frequency behavior of the RTD. We then used the admittance data under these conditions to extract the values of the parasitic elements. Finally, we assumed that the parasitics did not change at higher biases, so that the intrinsic RTD admittance could be extracted.

In Fig. 2-6(a), we show a cross section of the RTD mesa structure, indicating the physical origin of the parasitic elements. The conductance of the bend at the bottom of the mesa is represented by G_{m1} and the conductance of the mesa by G_{m2}. The Futurex dielectric is modeled by a distributed capacitance C', and the resistance in the substrate is modeled by the distributed resistance R'. The dominant circuit elements are shown in Fig. 2-6(b), where Y_R is the intrinsic admittance of the RTD, and $G_m = G_{m1}G_{m2}/(G_{m1}+G_{m2})$. We have modeled the distributed resistance and capacitance as a transmission line of length ℓ. The characteristic admittance of this line is $Y_o = \sqrt{j\omega C'/R'}$. Using this model, we expect a measurement of the mesa structure to have the response

$$Y_m = Y_o \frac{Y_L + Y_o \tanh \gamma\ell}{Y_o + Y_L \tanh \gamma\ell}, \tag{2.1}$$

where $\gamma = \sqrt{j\omega C'R'}$ is the complex propagation constant of the transmission line, and $Y_L = Y_R G_m/(Y_R + G_m)$ is the load at the end of the transmission line. Using estimates of the distributed elements, we found that the quantity $\gamma\ell$ is small enough so that the hyperbolic function can be approximated by the first term of a Taylor series expansion. This yields the result

$$Y_m \approx \frac{Y_R G_m + j\omega C_p(Y_R + G_m)}{Y_R(1 + G_m R_p) + G_m}, \tag{2.2}$$

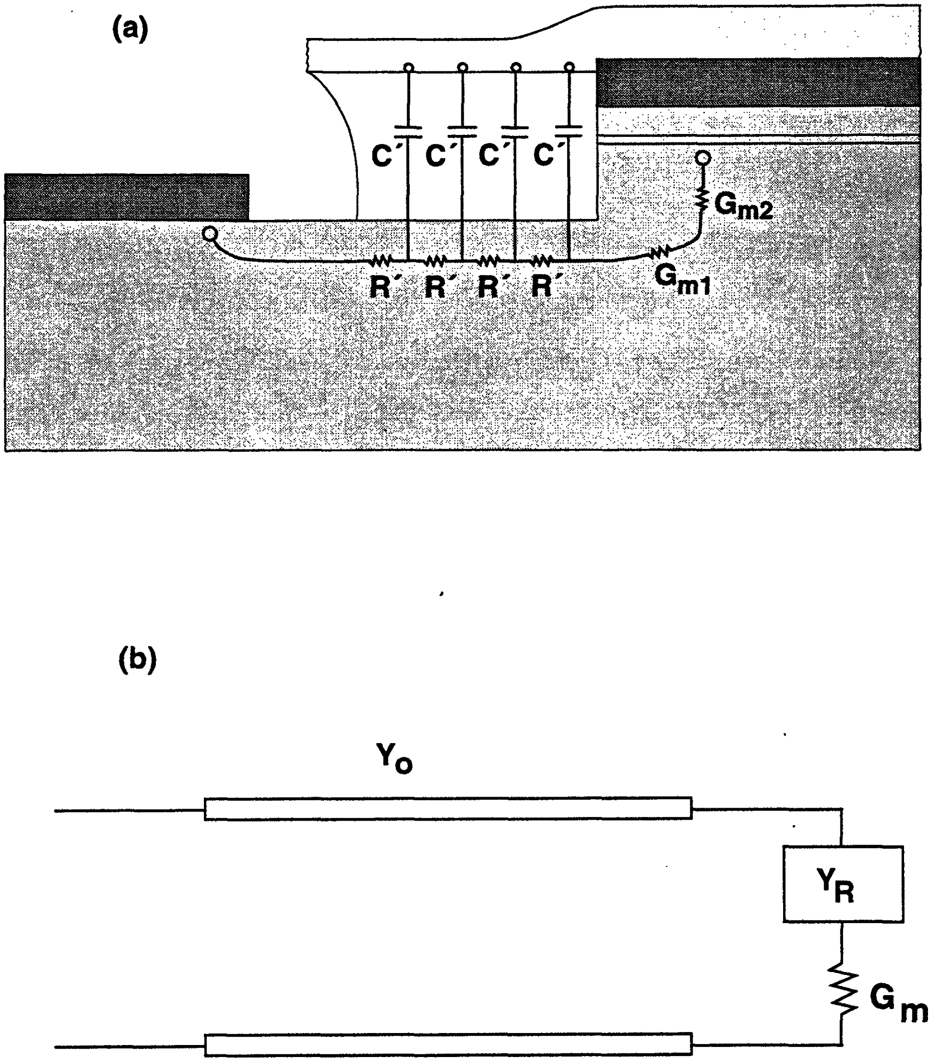

Figure 2-6. Parasitic circuit model for the fabricated structure. (a) Partial cross section, which shows the distributed capacitance C' in the Futurex dielectric, distributed resistance R' in the bulk semiconductor, conductance G_{m1} of the bend at the bottom of the mesa, and the mesa conductance G_{m2}, and (b) simplified circuit diagram of the RTD and distributed parasitics, where $G_m = G_{m1}G_{m2}/(G_{m1}+G_{m2})$, $Y_o = \sqrt{j\omega C'/R'}$, and Y_R is the intrinsic admittance of the RTD.

where $C_p = C'\ell$ and $R_p = R'\ell$. We note that C_p is physically the pad capacitance from the Futurex dielectric and that R_p is the resistance from the mesa to the field metal. Estimates of the parasitics show that the product $G_m R_p$ is much less than 1. Therefore, the measured admittance will be approximated as

$$Y_m \approx \frac{Y_R G_m + j\omega C_p (Y_R + G_m)}{Y_R + G_m} \,. \tag{2.3}$$

Once the values of C_p and G_m are extracted, we solve this equation to find the intrinsic admittance of the RTD, since Y_R is the only remaining unknown.

In order to find the values of C_p and G_m from the admittance data, we made some assumptions about the admittance of the RTD at low biases. The space-charge capacitance C_d can be accurately calculated for biases greater than approximately 0.2 V, since at these biases the 100-nm spacer layer is completely depleted and the depletion layer extends into the heavily doped layer. Under this condition, the position of the depletion edge and therefore the value of C_d become insensitive to changes in bias. (Since the capacitance associated with the barriers is large, C_d is dominated by the depletion-layer capacitance.) From the data shown in Fig. 2-3, it can be deduced that the dc conductance of the RTD is small compared with ωC_d for biases less than 0.5 and frequencies greater than 100 MHz. We shall therefore assume that the space-charge layer capacitance C_d dominates the RTD admittance at these biases and frequencies. Thus, the RTD admittance is modeled as $Y_R = j\omega C_d$, and the response of the mesa structure is approximately

$$Y_m = \frac{j\omega(C_d + C_p)G_m - \omega^2 C_d C_p}{G_m + j\omega C_d} = \frac{j\omega C_d G_m}{G_m + j\omega C_d} + j\omega C_p \,. \tag{2.4}$$

Since the measured angle shown in Fig. 2-5 is close to 90 degrees for all frequencies less than 26 GHz, we surmise that terms involving $\omega^2 C_d^2$ may be neglected when compared to G_m^2. As we shall see, the extracted values of the parasitics are consistent

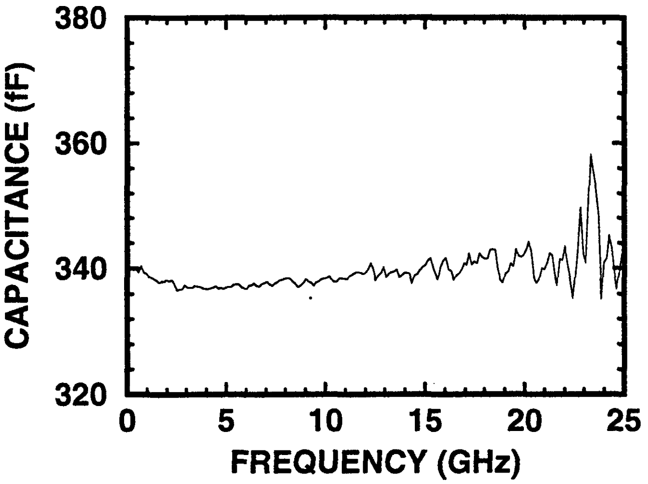

Figure 2-7. Extracted capacitance $C_p + C_d$ vs frequency for a bias of 0.2 V. The sum of the space-charge layer capacitance C_d and the pad capacitance C_p is found by dividing the imaginary part of the measured admittance by the angular frequency.

with this approximation. The real and imaginary parts are then given by

$$Y_m \approx \frac{\omega^2 C_d^2}{G_m} + j\omega(C_d + C_p)\,. \tag{2.5}$$

The value of C_p is extracted first. As can be seen from Eq. (2.5), the sum of C_d and C_p is found by dividing the imaginary part of the admittance by the angular frequency, i.e., $\mathrm{Im}\{Y_m\}/\omega$. A plot of this function for the representative device at 0.2 V is shown in Fig. 2-7. The average value is approximately 340 fF, and most of the data are within 2% of this value for frequencies greater than 4 GHz.

We measured the capacitance from several devices in this manner, and the results are shown in Fig. 2-8. The error bars shown in the plot represent the variation in the measured capacitance from a set of 25 devices: 10 devices with mesas nominally 16 μm on a side, and 5 devices each with mesas of 8, 4 and 2 μm on a side. (The actual sizes are slightly smaller because of processing considerations.) The size of the metal pad on top of the RTD mesa was the same in all cases. As can be seen from the figure, the measured capacitance of the 2- and 4-μm devices is approximately the same. For

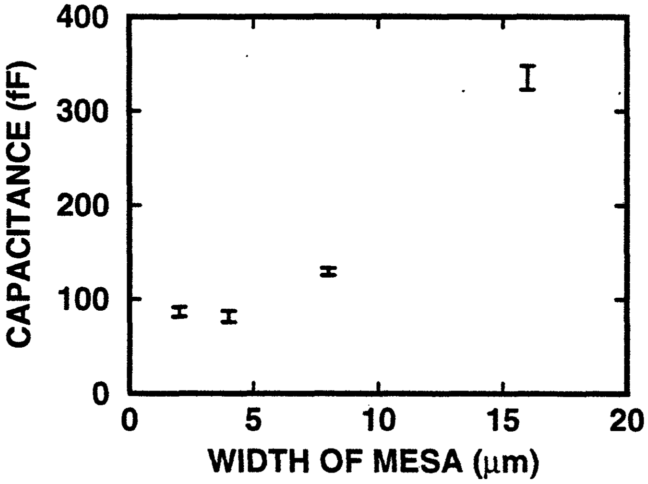

Figure 2-8. Extracted low-bias capacitance $C_p + C_d$ from several device sizes. The smallest devices have a negligible contribution from the space-charge layer capacitance C_d and are therefore dominated by the pad capacitance C_p.

devices of this size, the space-charge layer capacitance C_d is a very small contribution to the overall capacitance, so the measured capacitance is composed almost entirely of the pad capacitance C_p. We used the value C_p from the smallest devices, and scaled it by the ratio of Futurex areas for the larger devices, which resulted in a value of C_p =66 fF for the 16-μm devices.

Once the value of C_p was determined, we used the real part of the admittance to extract G_m. This conductance is given by $G_m = \omega^2 C_d^2/\mathrm{Re}\{Y_m\}$. A plot of this calculation for the representative device appears in Fig. 2-9. From the plot, we see that most of the data above 4 GHz lies within 8% of the value 0.22 S. The variation among the 16 μm devices was less than 10%.

It is now seen that the terms involving G_m^2 are approximately 50 times larger than those involving $\omega^2 C_d^2$ at 26 GHz. Thus, the approximation $\omega^2 C_d^2 \ll G_m^2$ used to obtain Eq. (2.5) from Eq. (2.3) is justified. We tested the accuracy of the low-bias modeling by plotting the difference between the expected and actual values of the admittance, as shown in Fig. 2-10. The actual value of the admittance was taken from the data shown in Fig. 2-5, and the expected value was calculated from Eq. (2.5).

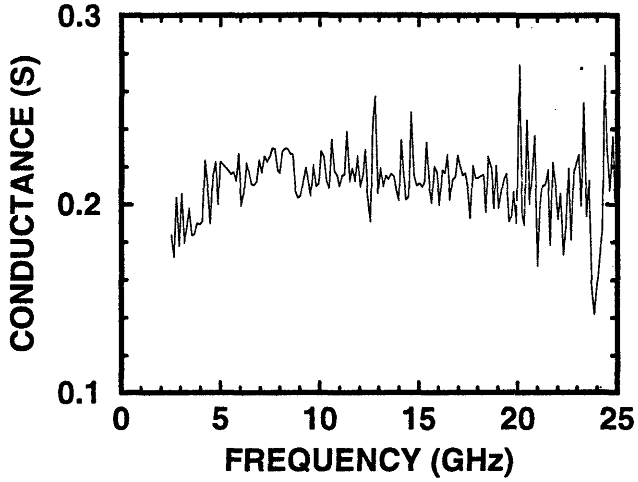

Figure 2-9. Extracted parasitic conductance G_m vs frequency for a bias of 0.2 V. The parasitic conductance is found from the real part of the admittance and the extracted capacitance.

The small difference indicates good agreement, although there is some disagreement at the lowest and highest frequencies. The small disagreement at low frequencies is most likely due to the fact that we have completely neglected to include a low-bias conductance for the RTD. Although this conductance is small, it is not zero. There is a small disagreement at higher frequencies; we will examine only frequencies less than 4 GHz for the remainder of this experimental work so that we are assured to be within the range of validity of the parasitic model.

Now that the parasitics are characterized, Eq. (2.3) can be solved for the RTD admittance to yield

$$Y_R = \frac{G_m(j\omega C_p - Y_m)}{Y_m - G_m - j\omega C_p}. \tag{2.6}$$

At this point, the separation of the intrinsic RTD admittance from the parasitics is complete.

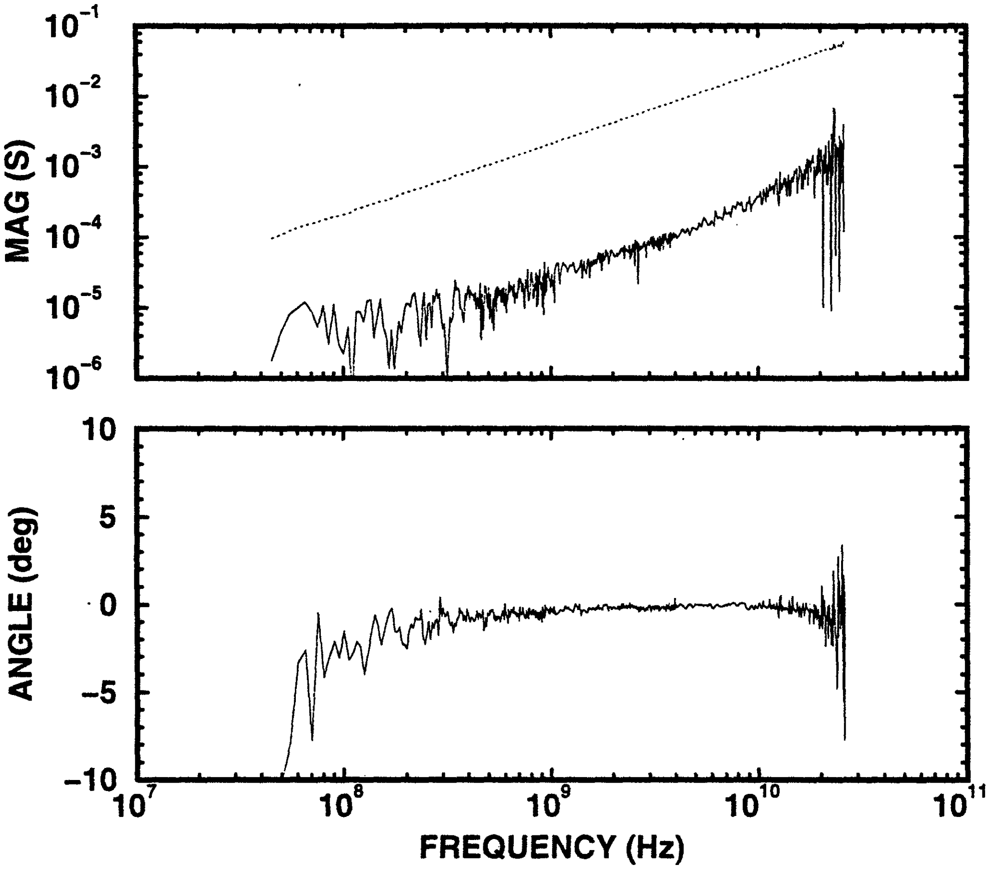

Figure 2-10. Comparison of calculated parasitic model with data. The dotted lines are the measured response. The solid lines are the difference between the measured and calculated responses.

2.4 Intrinsic Admittance of the RTD.

This section will examine the dependence of the intrinsic admittance of the RTD Y_R on frequency. This admittance was extracted using Eq. (2.6). When plotting the admittance, we will decompose Y_R into its real part, the conductance G_R, and its imaginary part, the susceptance B_R. In principle, it is redundant to examine both the real and imaginary parts, since they are related by the Kramers-Kronig relations. However, examining both parts increases intuitive understanding of the behavior of the device.

The extracted RTD conductance G_R for the biases 0.75, 1.0, 1.2, and 1.4 V is shown in Fig. 2-11 on page 36. This group of conductance plots have the same qualitative behavior. Both the low- and high-frequency conductances are monotonically increasing with bias. Each plot has a rolloff at about 800 MHz, although the rolloff is seen to occur at higher frequencies for the higher biases.

The extracted susceptance B_R for a bias of 0.75 V is shown in Fig. 2-12 on page 37. The susceptance appears to be linear in frequency, which would indicate that it is dominated by a capacitance. In fact, the susceptance is not perfectly linear. We shall define the capacitance C_s as the average of the quantity $B_R(\omega)/\omega$ for 10 angular frequencies between $2\pi \times 3.9$ GHz and $2\pi \times 4.0$ GHz. This average would be the value of the capacitance if the susceptance were perfectly linear in frequency. Table 2.1 (also on page 37) shows the tabulated values of C_s for several biases. We then define the "excess" susceptance B_x as $B_x = B_R - \omega C_s$, which is the deviation from a linear susceptance.

The excess susceptances are plotted in Fig. 2-13 on page 38 for the biases 0.75, 1.0, 1.2, and 1.4 V. Since C_s was calculated using the susceptance around 4 GHz, B_x drops to zero near that frequency. The data in Fig. 2-13 show that there is a small deviation from a purely capacitive susceptance. For these biases, the deviation is negative, and the minimum susceptance occurs at a frequency which is slowly increasing with bias.

In the next group of biases, 1.60, 1.65, 1.67 and 1.69 V, shown in Fig. 2-14 on page 39, the extracted low-frequency conductance drops off dramatically. This behavior is expected, since the device is biased closer to the current peak, where the dc conductance drops to zero. The rolloff frequency stays roughly constant for the biases 1.60, 1.65, and 1.67 V. At 1.69 V the extracted conductance is roughly constant with frequency. The excess susceptances for these biases are shown in Fig. 2-15 on page 40. The biases 1.60, 1.65, and 1.67 V have an excess negative susceptance, and the bias 1.69 V has no significant deviation from a capacitive susceptance.

In the third set of plots, for biases of 1.70, 1.71, 1.72, and 1.73 V, shown in Fig. 2-16 on page 41, the dc conductance moves from the positive-differential resistance

(PDR) to the NDR region. The rolloffs occur at significantly lower frequencies at these biases than at lower biases. Also noteworthy is the fact that the high-frequency conductance is still positive. The excess susceptances for these biases have become positive, as can be seen in Fig. 2-17 on page 42.

For the final set of plots, for biases of 1.76, 1.78, 1.80, and 1.84 V, shown in Fig. 2-18 on page 43, the dc conductance is also in the NDR region. Like the previous set of biases, the rolloff frequencies are significantly less than in the PDR region, as is shown on these expanded frequency scales. The rolloff begins at a frequency lower than 45 MHz, which is the low-frequency limit of the network analyzer. The excess susceptance, shown in Fig. 2-19 on page 44, has a decreasing peak value with bias. The peak frequency is increasing slightly with bias.

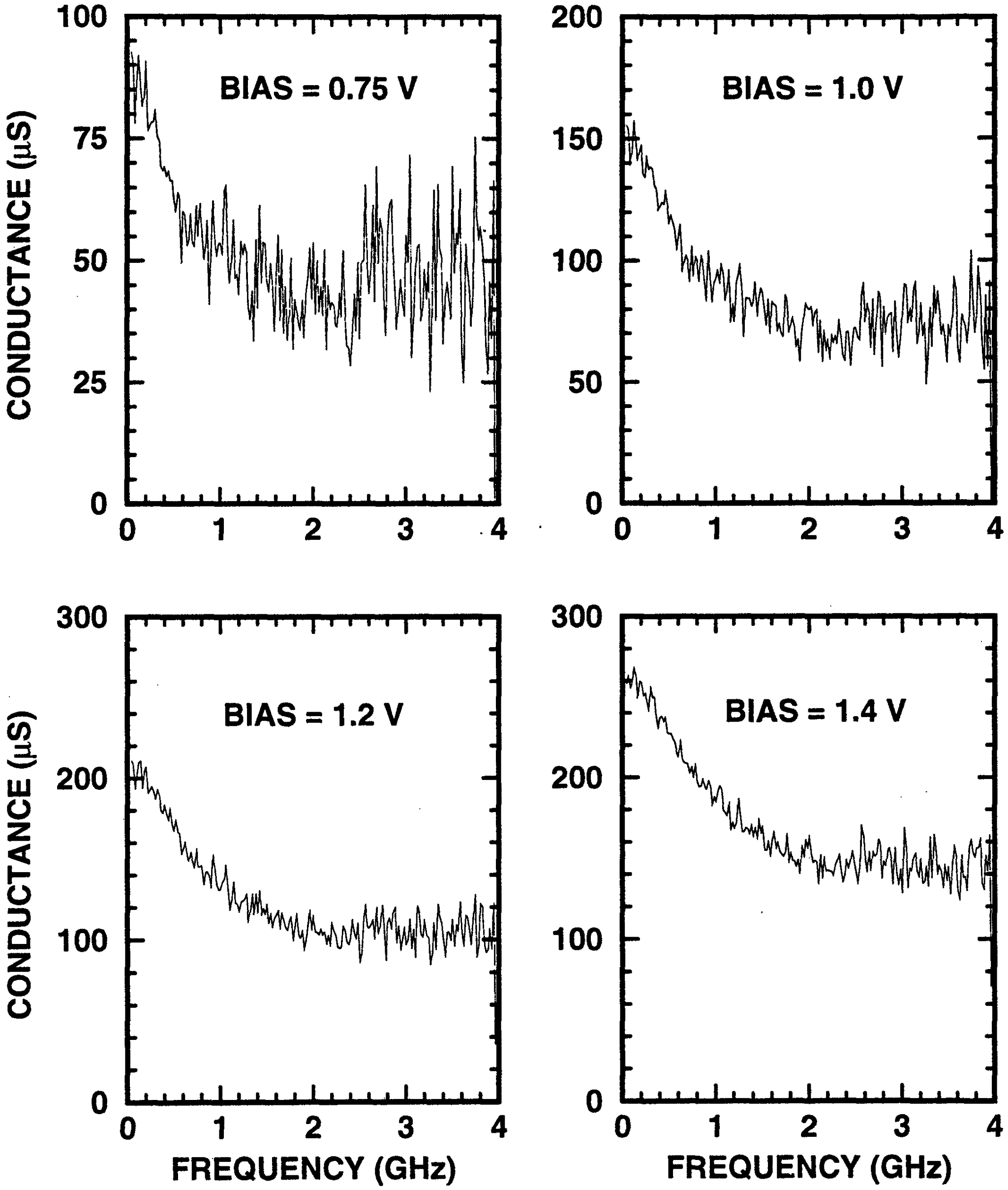

Figure 2-11. Extracted conductance G_R for the RTD at biases of 0.75, 1.0, 1.2, and 1.4 V as indicated.

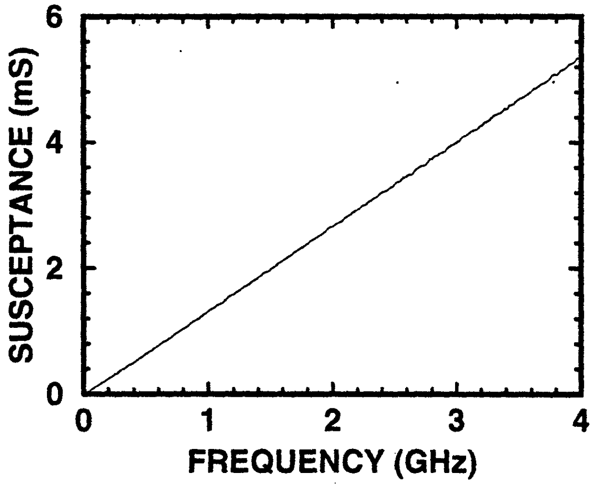

Figure 2-12. Extracted susceptance B_R for the RTD at a bias of 0.75 V.

BIAS (V)	C_s (fF)	BIAS (V)	C_s (fF)
0.75	213	1.70	202
1.00	208	1.71	202
1.20	205	1.72	202
1.40	203	1.73	203
1.60	201	1.76	204
1.65	201	1.78	203
1.67	202	1.80	203
1.69	202	1.84	203

Table 2.1. Calculated values for C_s, the dominant capacitance in the RTD susceptance B_R. The capacitance C_s is calculated by averaging B_R/ω for 10 frequencies between 3.9 and 4.0 GHz.

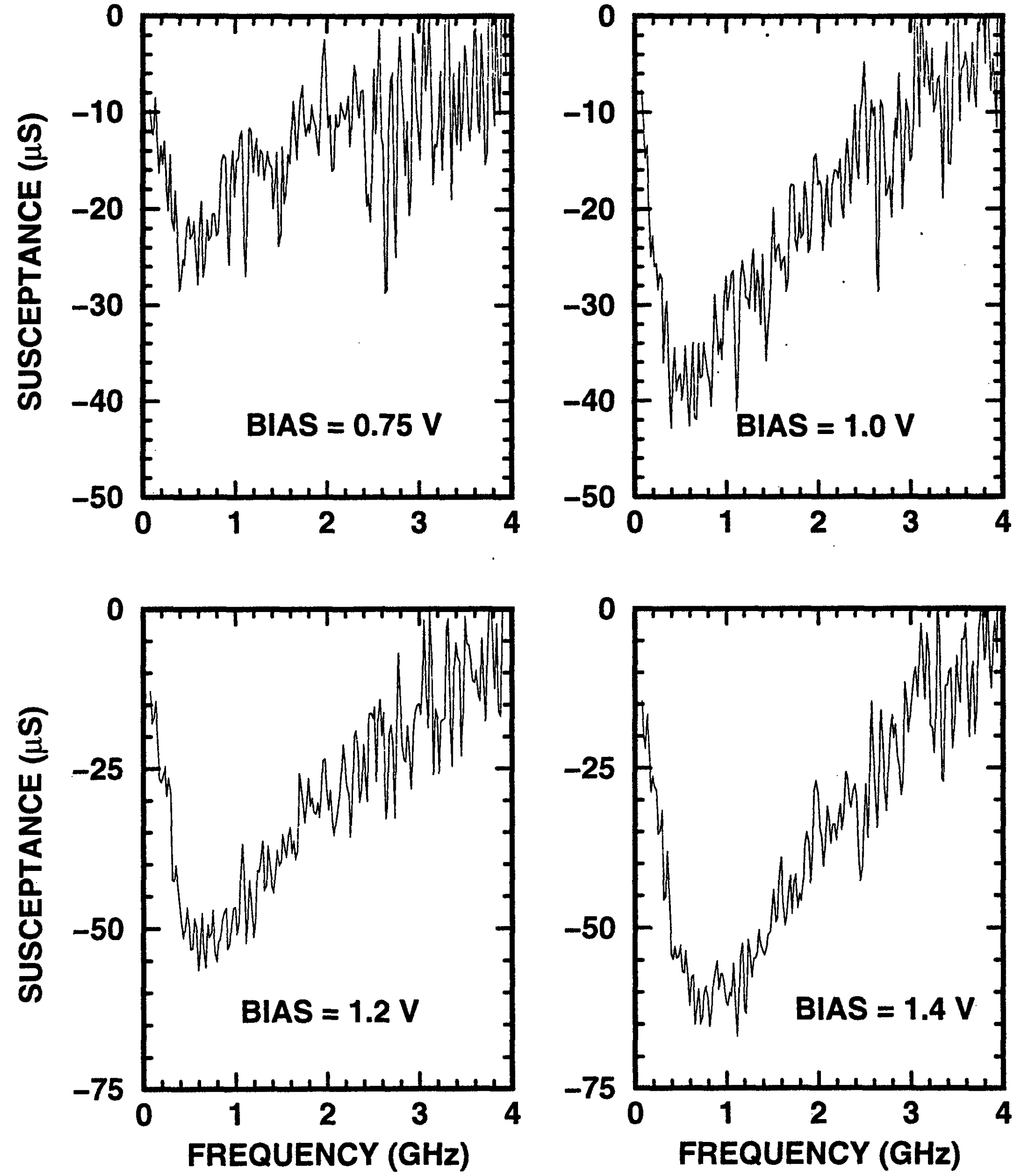

Figure 2-13. Excess susceptance $B_x = B_R - \omega C_s$ for the RTD at biases of 0.75, 1.0, 1.2, and 1.4 V. The excess susceptance represents the deviation of the RTD susceptance B_R from that of the capacitor C_s calculated by averaging B_R/ω in the 3.9 to 4.0 GHz range. The values used for C_s are shown in Table 2.1.

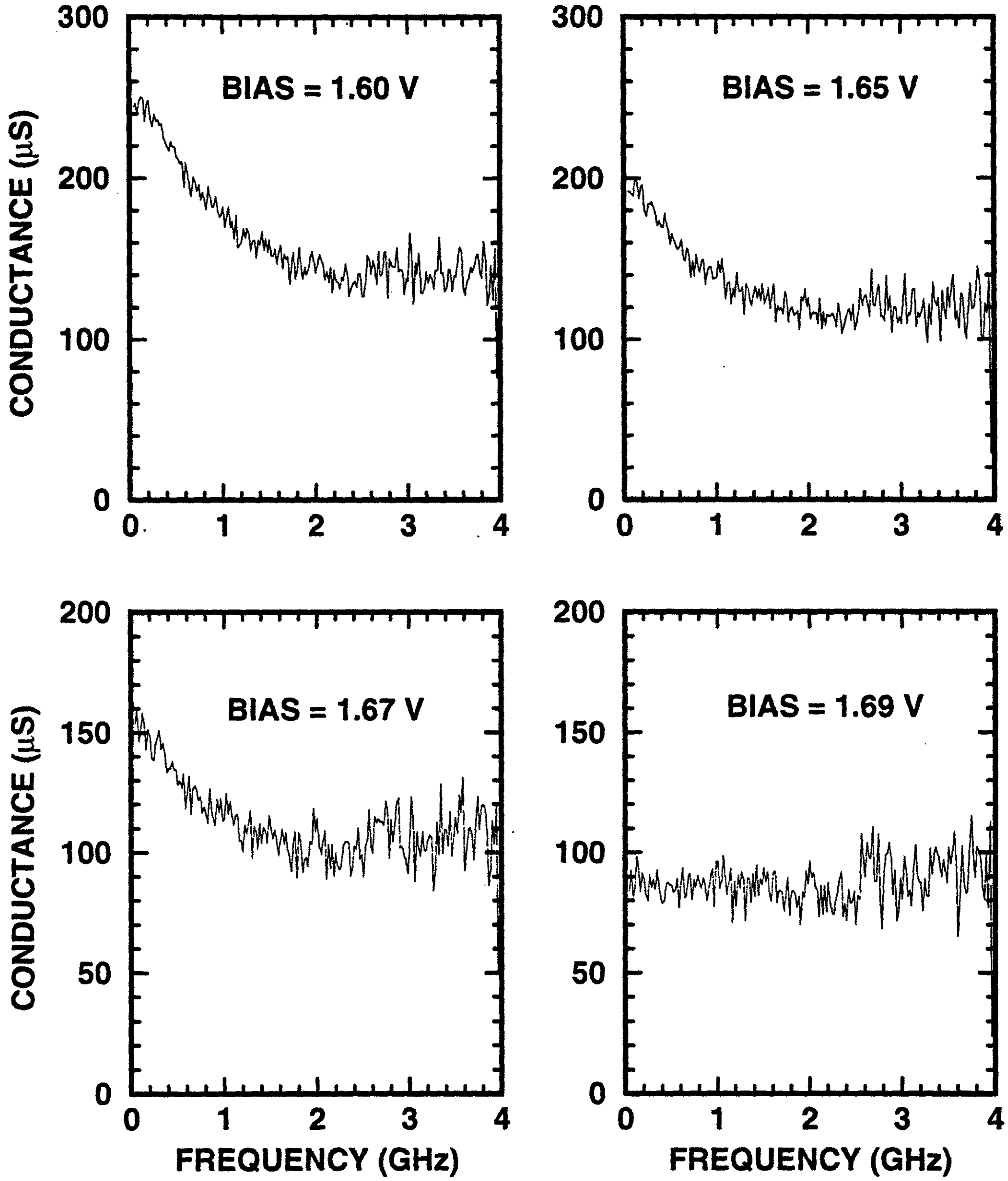

Figure 2-14. Extracted conductance G_R for the RTD at biases of 1.60, 1.65, 1.67, and 1.69 V as indicated.

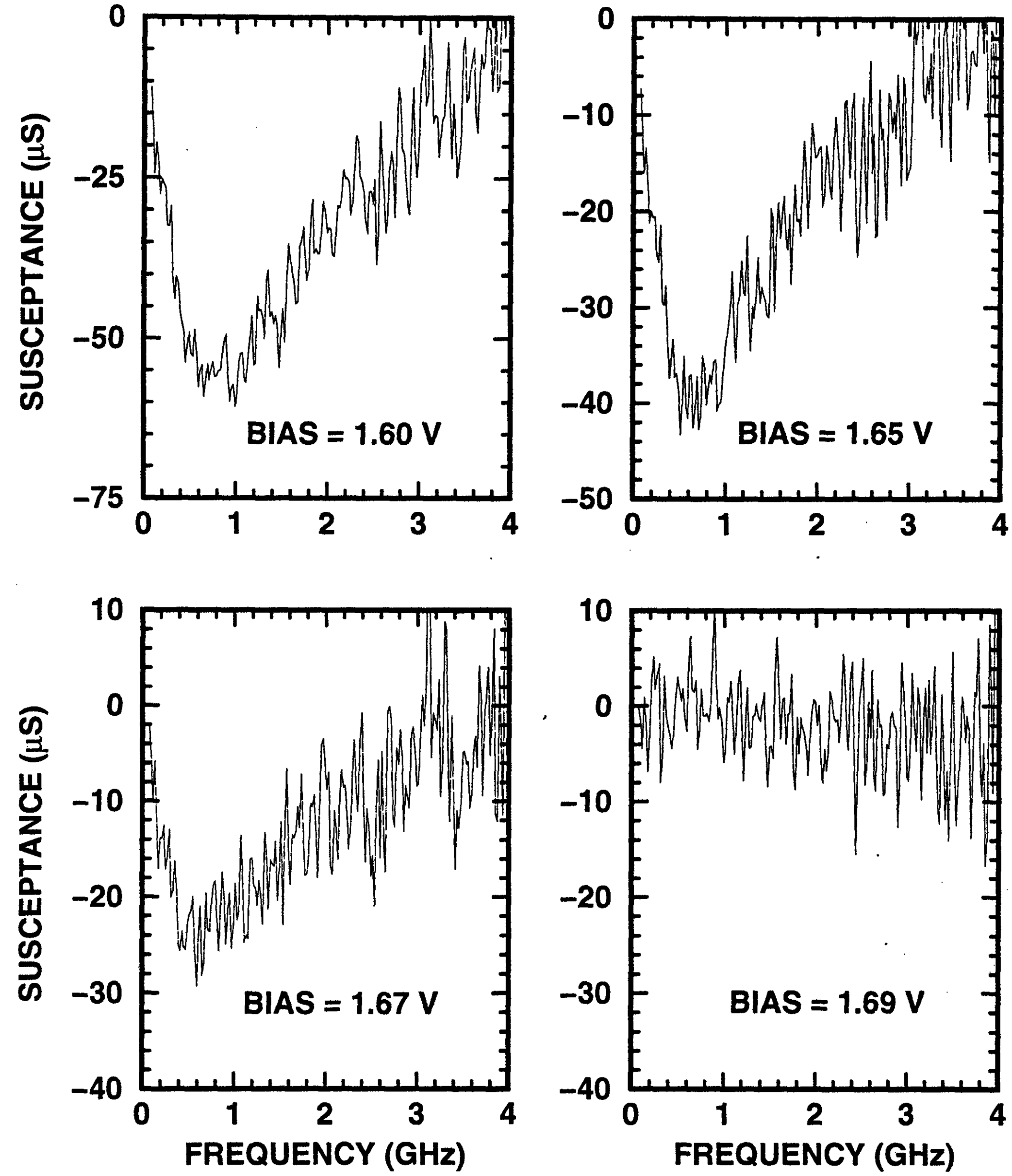

Figure 2-15. Excess susceptance $B_x = B_R - \omega C_s$ for the RTD at biases of 1.60, 1.65, 1.67, and 1.69 V. The excess susceptance represents the deviation of the RTD susceptance B_R from that of the capacitor C_s calculated by averaging B_R/ω in the 3.9 to 4.0 GHz range. The values used for C_s are shown in Table 2.1.

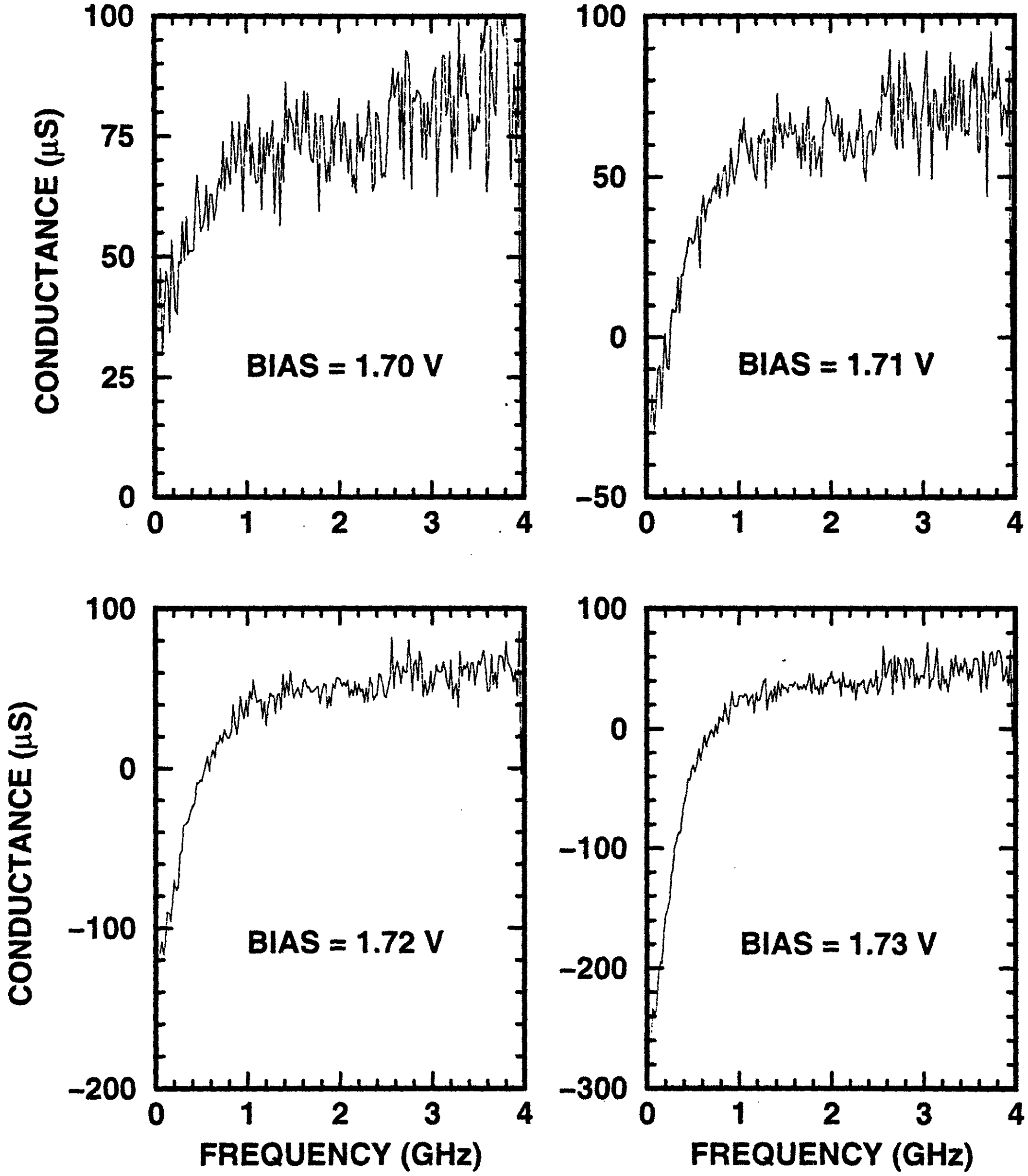

Figure 2-16. Extracted conductance G_R for the RTD at biases of 1.70, 1.71, 1.72, and 1.73 V as indicated.

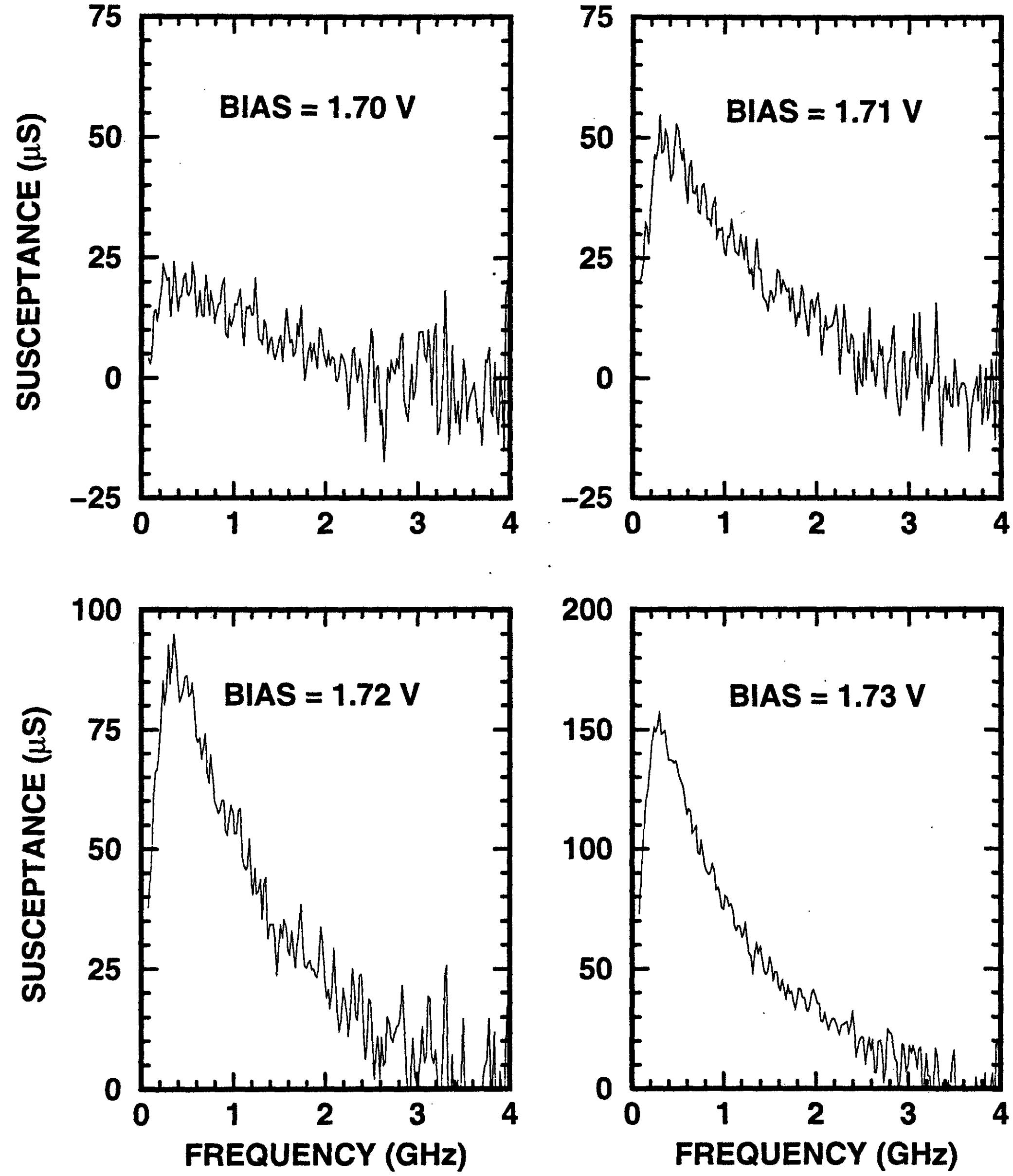

Figure 2-17. Excess susceptance $B_x = B_R - \omega C_s$ for the RTD at biases of 1.70, 1.71, 1.72, and 1.73 V. The excess susceptance represents the deviation of the RTD susceptance B_R from that of the capacitor C_s calculated by averaging B_R/ω in the 3.9 to 4.0 GHz range. The values used for C_s are shown in Table 2.1.

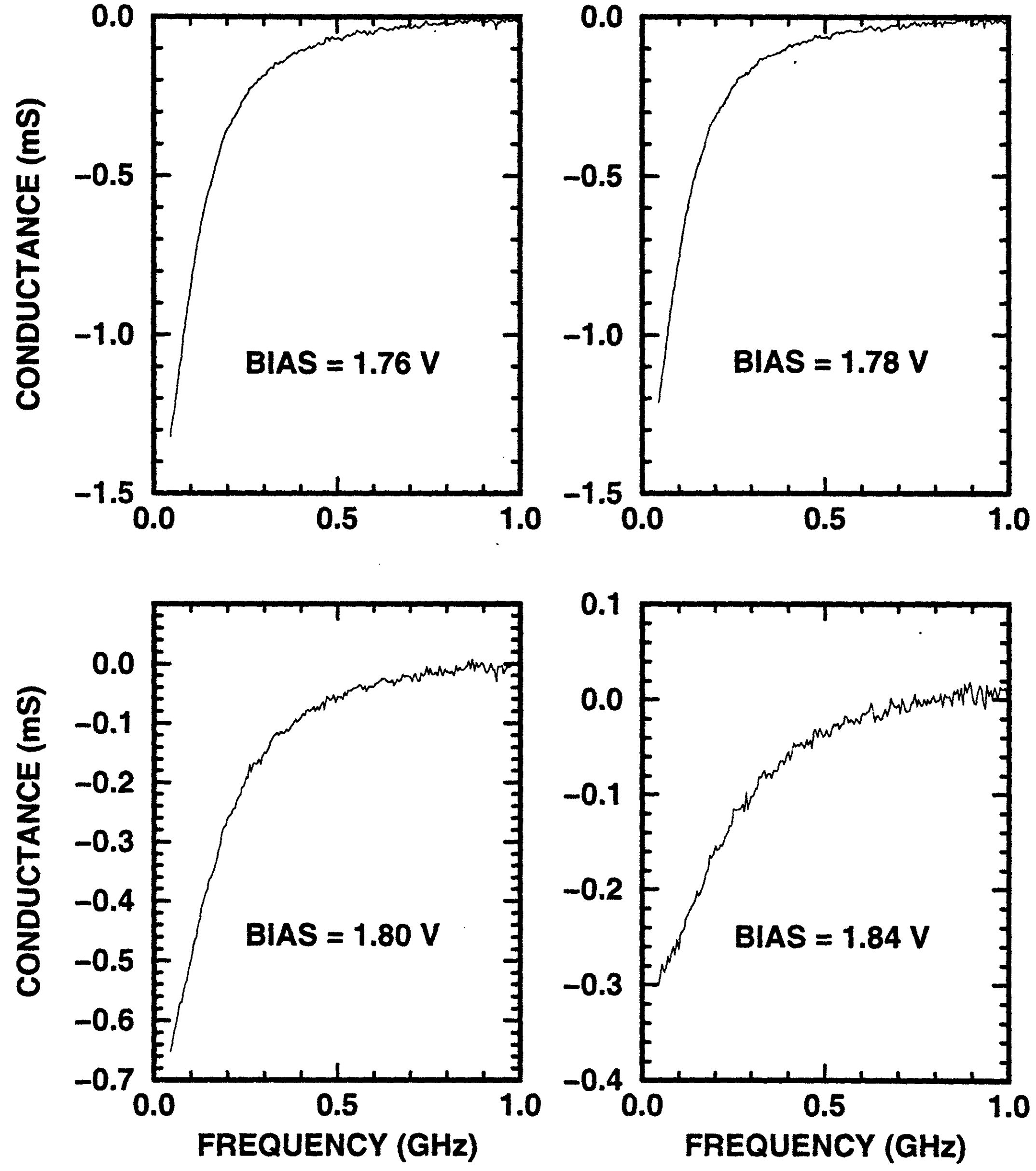

Figure 2-18. Extracted conductance G_R for the RTD at biases of 1.76, 1.78, 1.80, and 1.84 V as indicated. Note that the scale of the frequency axis is different from the previous plots.

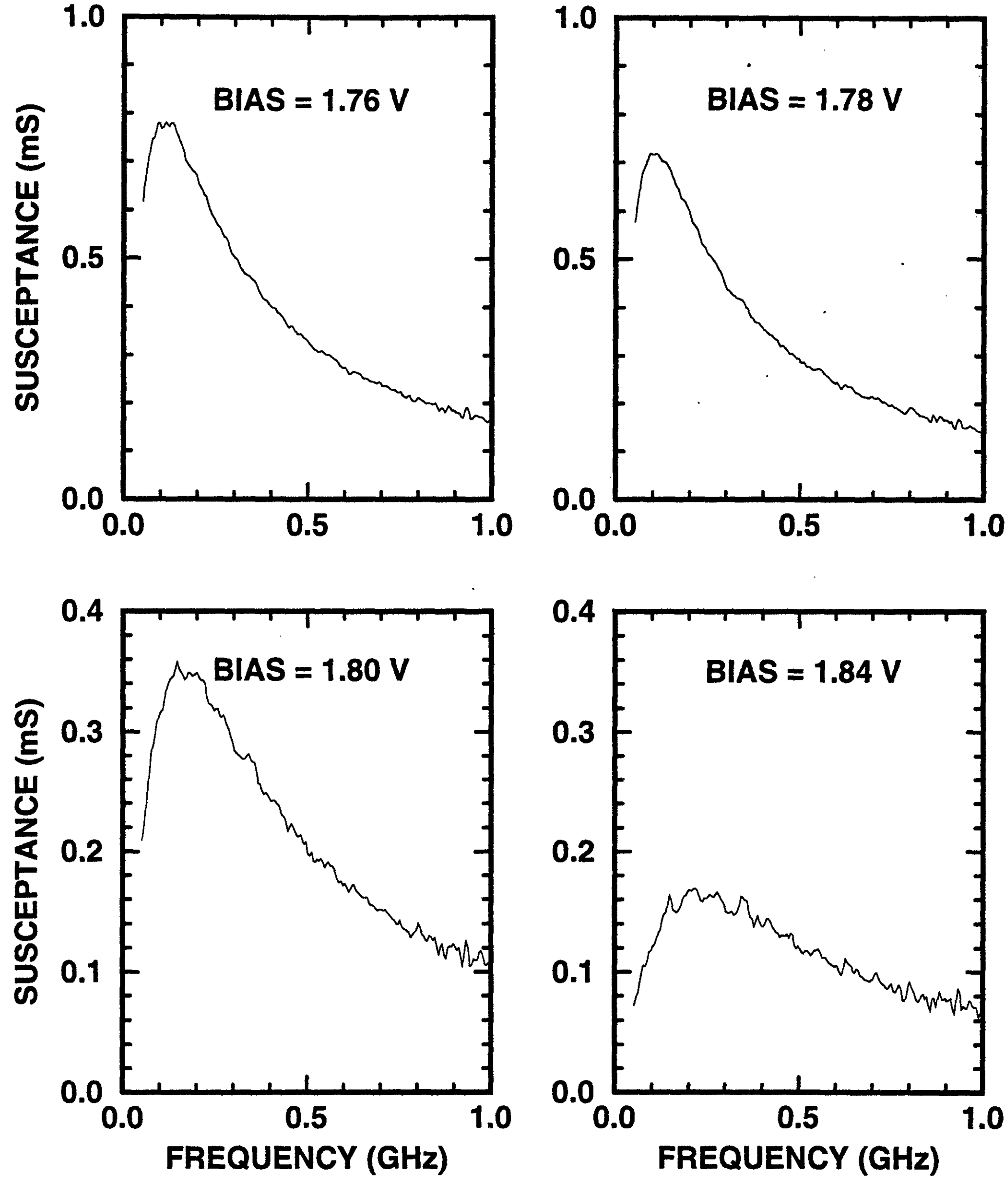

Figure 2-19. Excess susceptance $B_x = B_R - \omega C_s$ for the RTD at biases of 1.76, 1.78, 1.80, and 1.84 V. The excess susceptance represents the deviation of the RTD susceptance B_R from that of the capacitor C_s calculated by averaging B_R/ω in the 3.9 to 4.0 GHz range. The values used for C_s are shown in Table 2.1. Note that the scale of the frequency axis is different from the previous plots.

3

Rate-Equation Model of the RTD

In this chapter, we describe a method of modeling the RTD transport with rate equations. This method will enable us to explain the measured RTD behavior presented in Chapter 2, as will be shown in Chapter 4.

We will begin by formulating the rate equations for an RTD. Once the rate equations are established, expressions for the RTD current and charge in the well will be derived. By examining the small-signal change in the current and charge, we will arrive at a equivalent circuit for small-signal excitation. In order to complete the description of the RTD, we will describe a method for calculating the electron-transfer rates using the Fourier response of the Schrödinger solution to the RTD.

3.1 Formulation of the Rate Equations

In this section, we will derive expressions for the charge flow and storage in the RTD. Under a small bias, the RTD has a band bending like that shown schematically in Fig. 3-1. We have defined the regions so that region 1 is the cathode, region 2 is the well, and region 3 is the anode. The x direction is perpendicular to the plane of the well, while the y and z directions are in the plane of the well. For simplicity, we will be considering bias conditions such that only the lowest quasibound level plays a significant role in conduction through the double-barrier structure.

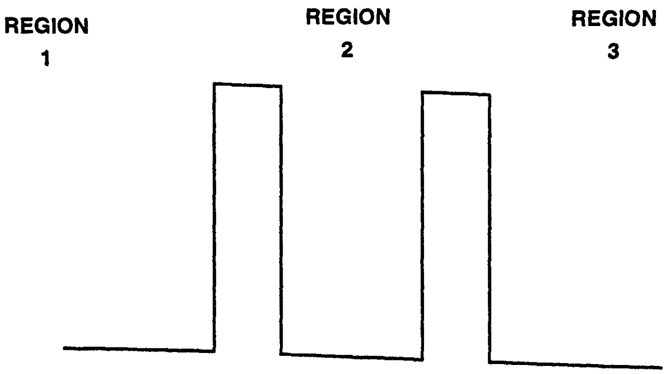

Figure 3-1. Schematic of the RTD band structure. The cathode, well, and anode are defined to be regions 1, 2, and 3, respectively. The x direction is perpendicular to the plane of the well, and the y and z directions are in the plane of the well.

First, we will formulate electron-transport rates between two states in different regions. This approach will yield undetermined coefficients, and the relation of these coefficients will be examined in the second part.

Formulation of the Transfer Rates

It will be assumed that the RTD barriers are thick enough so that the quasibound lifetime in the well is much greater than the mean scattering time. It will further be assumed that the scattering time is fast enough to create a thermalized energy distribution in the well, so that the carriers in the well are described by a separate quasi-Fermi level. Since the tunneling time through the device is much greater than the scattering time, a negligible number of electrons traverse coherently from region 1 to region 3. All electron transport proceeds by first traversing from region 1 to 2, and then from region 2 to 3. This type of transport is often referred to as sequential tunneling [50].

The basic method of calculating the rates in our model is as follows:

> The mean rate of electron transfer from an initial to final state is proportional to the probability that an initial state is occupied, as well as the probability that a final state is empty.

This bilinear approach is analogous to the Shockley-Read-Hall model for electron traps [51, 52], which is used to describe the behavior of bulk and surface traps [49]. However, this approach will yield results for the RTD which are substantially different from those of bulk and surface traps.

We first write the rate between a state in region 1 and a state in region 2. Consider an eigenstate which is primarily localized inside of region 1. There is a wavefunction associated with this eigenstate with a uniquely defined wavevector $\mathbf{k}_1$ in a region far away from the double-barrier structure, where the conduction band does not vary in space. The wavefunction can be decomposed into forward- and reverse-traveling waves, which are denoted $\mathbf{k}_1$ and $-\mathbf{k}_1$, respectively. For the purpose of calculations involving the density of states, we will assume that these traveling-wave states are normalized in a large box of length L_R in the x direction and area A_R in the y and z directions. There is also a quasibound state that is primarily localized in the well, which will be denoted by its transverse wavevector $\mathbf{k}_2$. The well state with momentum $-\mathbf{k}_2$ is degenerate with $\mathbf{k}_2$. These states are normalized in the volume with large area A_R and the entire x direction in length (since the state is bounded in x). In accordance with our basic method, the rate $r_{\mathbf{k}_1 \to \mathbf{k}_2}$ at which electrons make the transition from the traveling-wave state $\mathbf{k}_1$ to the state $\mathbf{k}_2$ is then written as

$$r_{\mathbf{k}_1 \to \mathbf{k}_2} = c_{12}(\mathbf{k}_1, \mathbf{k}_2) f_1 \left(1 - f_2\right) , \tag{3.1}$$

where f_1 and f_2 are the mean occupation factors of the states $\mathbf{k}_1$ and $\mathbf{k}_2$, respectively, and $c_{12}(\mathbf{k}_1, \mathbf{k}_2)$ is the transition coefficient for a particular $\mathbf{k}_1$ and $\mathbf{k}_2$. We shall write the reverse rate $r_{-\mathbf{k}_2 \to -\mathbf{k}_1}$ similarly as

$$r_{-\mathbf{k}_2 \to -\mathbf{k}_1} = c_{21}(-\mathbf{k}_1, -\mathbf{k}_2) f_2 \left(1 - f_1\right) , \tag{3.2}$$

where $c_{21}(-\mathbf{k}_1, -\mathbf{k}_2)$ is the transition coefficient for the reverse process.

The bilinear expressions for the transition rates between regions 2 and 3 is entirely analogous to those for regions 1 and 2. There is a transition rate between the state primarily localized in region 2, $\mathbf{k}_2$, and an eigenstate which is primarily localized inside of region 3. We denote the latter eigenstate by a uniquely defined wavevector $\mathbf{k}_3$ in a region far away from the double-barrier structure. Again, we decompose the wavefunction into forward- and reverse-traveling waves, denoted by $\mathbf{k}_3$ and $-\mathbf{k}_3$, respectively. Like $\mathbf{k}_1$, $\mathbf{k}_3$ is considered to be normalized in a large box of length L_R in the x direction and area A_R in the y and z. We write the rate of transition from $\mathbf{k}_2$ to $\mathbf{k}_3$, $r_{\mathbf{k}_2 \to \mathbf{k}_3}$, and from $-\mathbf{k}_3$ to $-\mathbf{k}_2$, $r_{-\mathbf{k}_3 \to -\mathbf{k}_2}$, by mapping 1→2 and 2→3 in Eqs. (3.1) and (3.2). This mapping yields the results

$$r_{\mathbf{k}_2 \to \mathbf{k}_3} = c_{23}(\mathbf{k}_2, \mathbf{k}_3) f_2 \, (1 - f_3) \tag{3.3}$$

$$r_{-\mathbf{k}_3 \to -\mathbf{k}_2} = c_{32}(-\mathbf{k}_2, -\mathbf{k}_3) f_3 \, (1 - f_2) \ , \tag{3.4}$$

where f_3 is the mean occupation number of the state $\mathbf{k}_3$, and $c_{23}(\mathbf{k}_2, \mathbf{k}_3)$ and $c_{32}(-\mathbf{k}_2, -\mathbf{k}_3)$ are the transition coefficients for the forward and reverse processes, respectively.

Relationship of the coefficients of proportionality.

The proportionality constants that we have introduced are not independent of each other. We now derive the relation that $c_{12}(\mathbf{k}_1, \mathbf{k}_2) = c_{21}(-\mathbf{k}_1, -\mathbf{k}_2)$ when the energies of the states $\mathbf{k}_1$ and $\mathbf{k}_2$ are the same [53].

Since transition rates are proportional to the squared magnitude of the scattering matrix element [54], it will suffice to examine the scattering matrix amplitude. The matrix element S_{21} describes the transition from $\mathbf{k}_1$ to $\mathbf{k}_2$,

$$S_{21} = \langle \mathbf{k}_2 | \, S \, | \mathbf{k}_1 \rangle \ , \tag{3.5}$$

where S is the scattering matrix operator. The time-reversal operator $\hat{T}$, which is anti-linear and unitary, can be inserted into the above definition to obtain

$$\begin{aligned}
S_{21} &=< \mathbf{k}_2|\hat{T}^\dagger\hat{T}S\hat{T}^\dagger\hat{T}|\mathbf{k}_1 > \\
&=< \mathbf{k}_2|\hat{T}^\dagger S^\dagger\hat{T}|\mathbf{k}_1 > \\
&=< \mathbf{k}_2|\hat{T}^\dagger S^\dagger|{\mathbf{k}_1}^R > \\
&= \left(< {\mathbf{k}_2}^R|S^\dagger|{\mathbf{k}_1}^R >\right)^* \\
&=< {\mathbf{k}_1}^R|S|{\mathbf{k}_2}^R > \\
&= S_{1^R 2^R}\,,
\end{aligned} \tag{3.6}$$

where the superscript dagger indicates the adjoint operator, $|{\mathbf{k}_1}^R>$ and $|{\mathbf{k}_2}^R>$ are the time-reversed states of $|\mathbf{k}_1>$ and $|\mathbf{k}_2>$, respectively, and $S_{1^R 2^R}$ is the scattering matrix element between $|{\mathbf{k}_1}^R>$ and $|{\mathbf{k}_2}^R>$. In the above manipulation, we have assumed that there are no spin-dependent effects. This assumption allowed us to use the relation $\hat{T}S\hat{T}^\dagger = S^\dagger$ since the Hamiltonian that connects the states $\mathbf{k}_1$ and $\mathbf{k}_2$ is then symmetric under time reversal. Comparing the first and last lines of Eq. (3.6) shows that the scattering amplitude between the states $|\mathbf{k}_1>$ and $|\mathbf{k}_2>$ is the same as that for $|{\mathbf{k}_1}^R>$ and $|{\mathbf{k}_2}^R>$. However, the time-reversed states $|{\mathbf{k}_1}^R>$ and $|{\mathbf{k}_2}^R>$ are simply the states with reversed momenta, i.e., $|-\mathbf{k}_1>$ and $|-\mathbf{k}_2>$, as well as time-reversed spins. As long as there are no spin-dependent effects present, the spin states are degenerate, and the Hamiltonian does not mix states of different spin. Thus, after summing over the spin states, we find that the transition rates between the states $|\mathbf{k}_1>$ and $|\mathbf{k}_2>$ is the same as that for $|-\mathbf{k}_1>$ and $|-\mathbf{k}_2>$, from which we conclude $c_{12}(\mathbf{k}_1,\mathbf{k}_2) = c_{21}(-\mathbf{k}_1,-\mathbf{k}_2)$. Furthermore, a similar derivation shows that $c_{23}(\mathbf{k}_2,\mathbf{k}_3) = c_{32}(-\mathbf{k}_2,-\mathbf{k}_3)$.[‡]

[‡] For states of unequal energies, more general relations may be proved:

$$\frac{c_{12}(\mathbf{k}_1,\mathbf{k}_2)}{c_{21}(\mathbf{k}_1,\mathbf{k}_2)} = e^{(E_1-E_2)/kT}$$

$$\frac{c_{23}(\mathbf{k}_2,\mathbf{k}_3)}{c_{32}(\mathbf{k}_2,\mathbf{k}_3)} = e^{(E_2-E_3)/kT}\,,$$

where E_n is the energy of state $\mathbf{k}_n$. The derivations in the remainder of this chapter deliberately delay the use of the equality of the E_n.

3.2 Derivation of the RTD Current and Charge in the Well

The rates in Eqs. (3.1) and (3.2) and the equality of the forward and reverse coefficients are combined to find the net transition rate between the states $\mathbf{k}_1$ and $\mathbf{k}_2$, which is

$$\begin{aligned} r_{\mathbf{k}_1\to\mathbf{k}_2} - r_{-\mathbf{k}_2\to-\mathbf{k}_1} &= c_{12}\left[f_1(1-f_2) - e^{(E_2-E_1)/kT} f_2(1-f_1)\right] \\ &= c_{12} f_1(1-f_2)\left[1 - e^{(F_2-F_1)/kT}\right] , \end{aligned} \tag{3.7}$$

where E_1 and E_2 are the energies of the states $\mathbf{k}_1$ and $\mathbf{k}_2$, respectively, and F_1 and F_2 are the quasi-Fermi energies of the regions 1 and 2, respectively. Similarly, the net transition rate between the states $\mathbf{k}_2$ and $\mathbf{k}_3$ is given by

$$r_{\mathbf{k}_2\to\mathbf{k}_3} - r_{-\mathbf{k}_3\to-\mathbf{k}_2} = c_{23} f_2(1-f_3)\left[1 - e^{(F_3-F_2)/kT}\right] , \tag{3.8}$$

where E_3 is the energy of the state $\mathbf{k}_3$, and F_3 is the quasi-Fermi energy of region 3. We note that the net transition rates between two regions are zero when their quasi-Fermi energies are equal, and this result is physically sensible since the quasi-Fermi energies are equal only when the two states have achieved equilibrium with each other. Equations (3.7) and (3.8) are the rates between two states, so we sum over all pairs of states in order to find the current between the two regions. For the current density between regions 1 and 2, J_{12}, this yields

$$\begin{aligned} J_{12} &= -2qV_R \iint \frac{d^3k_1}{8\pi^3}\,\frac{d^2k_2}{4\pi^2}\left(r_{\mathbf{k}_1\to\mathbf{k}_2} - r_{-\mathbf{k}_2\to-\mathbf{k}_1}\right) \\ &= -2qV_R \iint \frac{d^3k_1}{8\pi^3}\,\frac{d^2k_2}{4\pi^2}\, c_{12} f_1(1-f_2)\left[1 - e^{(F_2-F_1)/kT}\right] , \end{aligned} \tag{3.9}$$

and similarly for the current density between regions 2 and 3, J_{23},

$$J_{23} = -2qV_R \iint \frac{d^3k_3}{8\pi^3} \frac{d^2k_2}{4\pi^2} c_{23} \, f_2(1 - f_3) \left[1 - e^{(F_3 - F_2)/kT}\right] , \tag{3.10}$$

where $V_R = L_R A_R$ is the volume of normalization. The integration is performed over all values of tranverse k's, i.e., $-\infty$ to ∞ and positive values of longitudinal k, but the limits have been omitted from Eqs. (3.9) and (3.10) for simplicity. The factor of 2 in front of the integrals accounts for the fact that there are two spin states for each value of $\mathbf{k}_2$, and we have assumed that there is no coupling between states of different spin in different regions because there are no spin-dependent effects present.

We will make a number of definitions in order to convert Eqs. (3.9) and (3.10) to integrals over energies. For parabolic bands, the energy of an electron in the cathode E_1 can be decomposed into its longitudinal and transverse parts, $E_{\ell 1}$ and E_{t1}, respectively, so that $E_1 = E_{\ell 1} + E_{t1}$. The components are given by

$$E_{\ell 1} = \frac{\hbar^2 k_{x1}^2}{2m^*} + E_{C1} \tag{3.11}$$

$$E_{t1} = \frac{\hbar^2}{2m^*} \left(k_{y1}^2 + k_{z1}^2\right) , \tag{3.12}$$

where $E_{C1}(x)$ is the energy of the conduction band edge in the cathode, and m^* is the effective mass, which is assumed to be isotropic. Also, the energy of an electron in the anode E_3 can be decomposed into its longitudinal and transverse parts, $E_{\ell 3}$ and E_{t3}, respectively, so that $E_3 = E_{\ell 3} + E_{t3}$. These components are given by

$$E_{\ell 3} = \frac{\hbar^2 k_{x3}^2}{2m^*} + E_{C3} \tag{3.13}$$

$$E_{t3} = \frac{\hbar^2}{2m^*} \left(k_{y3}^2 + k_{z3}^2\right) , \tag{3.14}$$

where $E_{C3}(x)$ is the energy of the conduction band edge in the anode. The energy of an electron in the well E_2 can be decomposed into its transverse component E_{t2}

and longitudinal component E_R so that $E_2 = E_{t2} + E_R$. The transverse component is defined by

$$E_{t2} = \frac{\hbar^2}{2m^*}\left(k_{y2}^2 + k_{z2}^2\right) . \tag{3.15}$$

The longitudinal component E_R is the resonant energy in the well. Since the electron is not completely confined to the well, there is some width in energy associated with this resonance, which will be discussed in later sections. In the scattering-dominated regime, the width is an insignificant fraction of the thermal energy kT, so we have neglected it in the calculation of E_2. We also define the energy of the conduction band edge in this region to be $E_{C2}(x)$. Using these energy definitions, the mean occupation numbers are written as

$$f_1 = \frac{1}{1+\exp\left[(E_{\ell 1} + E_{t1} - F_1)/kT\right]} \tag{3.16}$$

$$f_2 = \frac{1}{1+\exp\left[(E_R + E_{t2} - F_2)/kT\right]} \tag{3.17}$$

$$f_3 = \frac{1}{1+\exp\left[(E_{\ell 3} + E_{t3} - F_3)/kT\right]} . \tag{3.18}$$

We now consider the form of the coefficients c_{12} and c_{23}. We expect that there will be a sharply peaked resonance in the well with longitudinal energy E_R. For our model, we shall assume that the transitions into the well are within an energy width of this resonance. In making this restriction, we are neglecting optical-phonon processes. When such processes are significant, the I-V curve has a secondary peak [21]; we conclude that these processes can be neglected since we do not see such a peak in Fig. 2-3. Also, we will insist that the transverse wavevector is conserved. With these restrictions, the coefficients c_{12} and c_{23} then take the form

$$c_{12} = w_{12}(E_{\ell 1}, E_R)\, 4\pi^2 A_R^{-1}\, \delta(\mathbf{k}_{t1} - \mathbf{k}_{t2}) \tag{3.19}$$

$$c_{23} = w_{23}(E_{\ell 3}, E_R)\, 4\pi^2 A_R^{-1}\, \delta(\mathbf{k}_{t2} - \mathbf{k}_{t3}) \,, \tag{3.20}$$

where the transverse wavevector is $\mathbf{k}_{tn} = \hat{y} k_{yn} + \hat{z} k_{zn}$, and δ is the Dirac δ function. The quantities $w_{12}(E_{\ell 1}, E_R)$ and $w_{23}(E_{\ell 3}, E_R)$ are functions only of the longitudinal energies.

We use Eqs. (3.19) and (3.20) to integrate out $\mathbf{k}_{t1}$ and $\mathbf{k}_{t3}$, and then convert the integral over wavevectors to an integral over energy, which leaves

$$J_{12} = -\frac{q m^* L_R}{\pi \hbar^2} \int_0^\infty dE_{t2}\, (1 - f_2) \left[1 - e^{(F_2 - F_1)/kT}\right] \int_{E_{C1}}^\infty \frac{dE_{\ell 1}}{2\pi\hbar v_{\ell 1}}\, w_{12}(E_{\ell 1}, E_R) f_1 \tag{3.21}$$

$$J_{23} = -\frac{q m^* L_R}{\pi \hbar^2} \int_0^\infty dE_{t2}\, f_2 \left[1 - e^{(F_3 - F_2)/kT}\right] \int_{E_{C3}}^\infty \frac{dE_{\ell 3}}{2\pi\hbar v_{\ell 3}}\, w_{23}(E_{\ell 3}, E_R)(1 - f_3) \,, \tag{3.22}$$

where $v_{\ell n} = \sqrt{2(E_{\ell n} - E_{Cn})/m^*}$ is the longitudinal velocity in region n.

For thick-barrier RTDs, the width in energy of the quasibound state is very narrow compared to kT. Since we expect that electron transfer will occur only within a few widths of the energy peak, the functions $w_{12}(E_{\ell 1}, E_R)$ and $w_{23}(E_{\ell 3}, E_R)$ are highly peaked and are changing much more rapidly than the Fermi factors for the temperatures in which we are interested. Therefore, we shall evaluate f_1 at $E_{\ell 1} = E_R$ and f_3 at $E_{\ell 1} = E_R$, and then remove them from the integrals. We will denote the Fermi factors so evaluated as $\underline{f_1} = f_1(E_{\ell 1} = E_R, E_t)$ and $\underline{f_3} = f_3(E_{\ell 3} = E_R, E_t)$. The current densities then become

$$J_{12} \approx -\frac{qm^*}{\pi\hbar^2}\,\nu_1 \int_0^\infty dE_{t2}\,(1-f_2)\underline{f_1}\left[1-e^{(F_2-F_1)/kT}\right] \tag{3.23}$$

$$J_{23} \approx -\frac{qm^*}{\pi\hbar^2}\,\nu_3 \int_0^\infty dE_{t2}\,f_2\left(1-\underline{f_3}\right)\left[1-e^{(F_3-F_2)/kT}\right], \tag{3.24}$$

where

$$\nu_1 = L_R \int_{E_{C1}}^\infty \frac{dE_{\ell 1}}{2\pi\hbar v_{\ell 1}}\, w_{12}(E_{\ell 1}, E_R) \tag{3.25}$$

$$\nu_3 = L_R \int_{E_{C3}}^\infty \frac{dE_{\ell 3}}{2\pi\hbar v_{\ell 3}}\, w_{23}(E_{\ell 3}, E_R)\,. \tag{3.26}$$

The quantities ν_1 and ν_3 have units of inverse time, and they are a rough indication of the speed of the device, as will be seen later. We will refer to them as transfer rates. A straightforward calculation shows that the integrands in Eqs. (3.23) and (3.24) can be expressed as*

$$(1-f_2)\underline{f_1}\left[1-e^{(F_2-F_1)/kT}\right] = \underline{f_1} - f_2 \tag{3.27}$$

$$(1-\underline{f_3})f_2\left[1-e^{(F_3-F_2)/kT}\right] = f_2 - \underline{f_3}\,. \tag{3.28}$$

These simplified integrands may be integrated analytically over E_{t2}, using the integral shown in Appendix B. The expressions for the current densities then become

$$J_{12} = -\frac{qm^*kT\nu_1}{\pi\hbar^2}\ln\frac{1-f_{2R}}{1-f_{1R}} \tag{3.29}$$

$$J_{23} = -\frac{qm^*kT\nu_3}{\pi\hbar^2}\ln\frac{1-f_{3R}}{1-f_{2R}}, \tag{3.30}$$

---

*Note that this manipulation would not be needed if the exponentials of differences in energy had been left in Eqs. (3.7) and (3.8).

where we have defined

$$f_{nR} = \frac{1}{1 + e^{(E_R - F_n)/kT}}\,. \tag{3.31}$$

Equations (3.29) and (3.30) are the final expressions for the current densities. These equations have a necessary physical property. When $F_1 > F_2 > F_3$, the current is always negative, so the electron current is flowing from left to right in Fig. 3-1. This property ensures that power is being dissipated.

The calculation of the charge per unit area in the well proceeds by integrating the occupied states in the well,

$$Q_w = -2q \int \frac{d^2k_2}{4\pi^2} f_2 = -\frac{qm^*}{\pi\hbar^2} \int_0^\infty dE_{t2} f_2\,. \tag{3.32}$$

Again using the results in Appendix B, this integral evaluates to

$$Q_w = -\frac{qm^*kT}{\pi\hbar^2} \ln \frac{1}{1 - f_{2R}}\,. \tag{3.33}$$

Since the Fermi factor f_{2R} is always between 0 and 1, the quantity Q_w is always negative.

3.3 Small-Signal Variation about Steady State

We will now examine the behavior of the device when a small-signal sinusoid is superimposed on the bias. Our goal is to derive an equivalent-circuit model which will describe the response to this small-signal excitation. The excitation must be small enough so that the RTD response is linear in the input.

The derivation begins from the current continuity condition, in order to ensure charge conservation,

$$\frac{\partial Q_w}{\partial t} = J_{12} - J_{23} \,. \tag{3.34}$$

Each quantity can be decomposed into a steady-state component that is independent of time, and a time-dependent component that contains the small-signal variation. The quantities in Eq. (3.34) are decomposed as

$$J_{12} = J_{12}^o + \delta J_{12} \tag{3.35}$$

$$J_{23} = J_{23}^o + \delta J_{23} \tag{3.36}$$

$$Q_w = Q_w^o + \delta Q_w \,, \tag{3.37}$$

where the superscript o denotes the steady-state value and the δ prefix denotes the small-signal variation. By setting the small-signal variation to zero, the current continuity condition yields $J_{12}^o = J_{23}^o$. This equation reflects the physical requirement that the current crossing the left barrier must cross the right barrier in steady state, in order that the charge in the well remains constant. Since the steady-state quantities cancel, the small-signal quantities must obey

$$\frac{\partial}{\partial t}\delta Q_w = \delta J_{12} - \delta J_{23} \,. \tag{3.38}$$

Thus, there must be strict conservation of the incremental charge as well as the large-signal charge.

By linearizing Eqs. (3.29), (3.30), and (3.33), we find analytical expressions for the

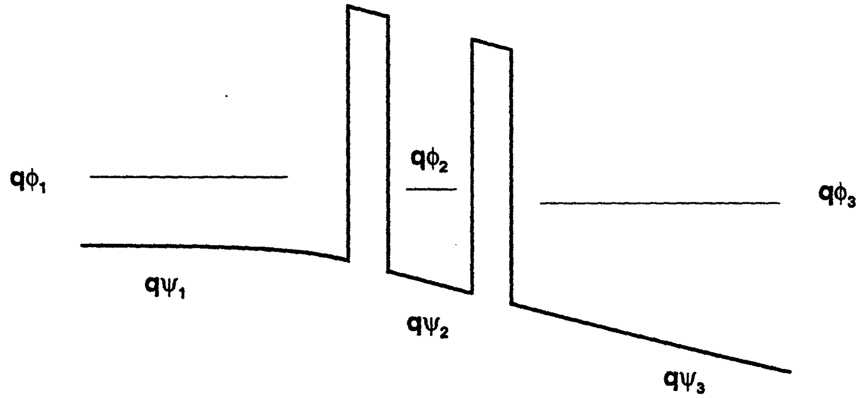

Figure 3-2. Location of the potentials used in the equivalent circuit, shown on an energy diagram. The quantities ψ_1 and ψ_3 represent the electron potential in the quasineutral region of the cathode and anode, respectively. The quantity ψ_2 is defined to be the electron potential at the point in the well where $E_R - E_{C2}$ equals its equilibrium value; this point is roughly at the center of the well. The quantities ϕ_1, ϕ_2, and ϕ_3 are the quasi-Fermi potentials in regions 1, 2, and 3, respectively.

incremental quantities δJ_{12}, δJ_{23} and δQ_w. This manipulation yields the expressions

$$\delta J_{12} = -\frac{qm^*kT}{\pi\hbar^2}\left[\nu_1\frac{\delta f_{1R}}{1-f_{1R}} - \nu_1\frac{\delta f_{2R}}{1-f_{2R}} + \delta\nu_1 \ln\frac{1-f_{2R}}{1-f_{1R}}\right] \tag{3.39}$$

$$\delta J_{23} = -\frac{qm^*kT}{\pi\hbar^2}\left[\nu_3\frac{\delta f_{2R}}{1-f_{2R}} - \nu_3\frac{\delta f_{3R}}{1-f_{3R}} + \delta\nu_3 \ln\frac{1-f_{3R}}{1-f_{2R}}\right] \tag{3.40}$$

$$\delta Q_w = -\frac{qm^*kT}{\pi\hbar^2}\frac{\delta f_{2R}}{1-f_{2R}}, \tag{3.41}$$

where $\delta\nu_1$ and $\delta\nu_3$ represent the small-signal variation in the transfer rates associated with regions 1 and 3, respectively, and δf_{nR} is the small-signal variation in the mean occupation number in region n.

Before calculating δf_{nR} for the three regions, we will convert the independent variables in the Fermi factors f_{nR} from energies to potentials. Later, this will allow us to relate the results to an equivalent circuit in a more straightforward manner. The potentials are schematically shown on the energy band diagram in Fig. 3-2. We define ψ_1 and ψ_3 as the electron potentials in the quasineutral region of the cathode and anode, respectively, and $\phi_n = -F_n/q$ as the quasi-Fermi potential in region n. The electron potential of the quasibound state is denoted $\psi_R = -E_R/q$. Let x_w be the point in the well where $E_R - E_{C2}$ remains equal to its equilibrium value. We then define $\psi_2 = \psi_2(x_w)$ to be the electron potential in region 2. This definition of x_w will prove useful later, since it will force the two incremental quantities $\delta\psi_2(x_w)$ and $\delta\psi_R$ to be equal. The point at which this condition is met is close to the center of the well; we could simply define $\psi_2(x_w)$ to be the electron potential at the center of the well without introducing much error into later calculations.

These definitions transform the Fermi factors into

$$f_{nR} = \frac{1}{1 + e^{q(\phi_n - \psi_R)/kT}} . \tag{3.42}$$

As detailed in Appendix B, the linearized variation in f_{nR} is then

$$\begin{aligned} \delta f_{nR} &= \frac{q}{kT} f_{nR}(1 - f_{nR})(\delta\psi_R - \delta\phi_n) \\ &= \frac{q}{kT} f_{nR}(1 - f_{nR})(\delta\psi_2 - \delta\phi_n) . \end{aligned} \tag{3.43}$$

The incremental variation in ν_1 and ν_3 involve taking the derivative of the transition probability rates w_{12} and w_{23}. The expressions for the these rates will turn out to be algebraically messy, and the derivatives will be more so. For now, we will work with the symbolic derivatives. Then, the incremental quantities become

$$\delta\nu_1 = \frac{\partial\nu_1}{\partial\psi_{21}}(\delta\psi_2 - \delta\psi_1) \tag{3.44}$$

$$\delta\nu_3 = \frac{\partial\nu_3}{\partial\psi_{32}}(\delta\psi_3 - \delta\psi_2) , \tag{3.45}$$

where ψ_{nm} is a shorthand for $\psi_n - \psi_m$. In writing these expressions for $\delta\nu_1$ and $\delta\nu_3$, we have used the fact that the transfer rates are a function only of the band structure. In so doing, we have neglected any effect that changes in the quasi-Fermi level might have on the transfer rates. Physically, this approximation consists of neglecting the direct influence of electron-electron scattering in the well on the transfer rates, which is suitably accurate for our purpose. This approximation will again be used when we calculate the transfer rates.

After substituting the above expressions into Eqs. (3.39) through (3.41), the incremental variations become

$$\delta J_{12} = -\frac{q^2 m^* \nu_1}{\pi \hbar^2}\left[f_{1R}(\delta\psi_2 - \delta\phi_1) - f_{2R}(\delta\psi_2 - \delta\phi_2) + \eta_1(\delta\psi_2 - \delta\psi_1)\right] \tag{3.46}$$

$$\delta J_{23} = -\frac{q^2 m^* \nu_3}{\pi \hbar^2}\left[f_{2R}(\delta\psi_2 - \delta\phi_2) - f_{3R}(\delta\psi_2 - \delta\phi_3) + \eta_3(\delta\psi_3 - \delta\psi_2)\right] \tag{3.47}$$

$$\delta Q_w = -\frac{q^2 m^*}{\pi \hbar^2} f_{2R}(\delta\psi_R - \delta\phi_2)\,, \tag{3.48}$$

where the dimensionless quantities η_1 and η_3 have been defined as

$$\eta_1 = \frac{kT}{q\nu_1}\left(\frac{\partial \nu_1}{\partial \psi_{21}}\right)\ln\frac{1-f_{2R}}{1-f_{1R}} \tag{3.49}$$

$$\eta_3 = \frac{kT}{q\nu_3}\left(\frac{\partial \nu_3}{\partial \psi_{32}}\right)\ln\frac{1-f_{3R}}{1-f_{2R}}\,. \tag{3.50}$$

In the quasineutral cathode and anode regions, there is a large electron population and a great deal of scattering. As a consequence, the relaxation time is much smaller than the time scale of the applied variation for frequencies used in Chapter 2. The variations $\delta\phi_1$ and $\delta\psi_1$ are therefore approximately equal, as are the variations $\delta\phi_3$ and $\delta\psi_3$. These equalities allow us to manipulate the incremental quantity δJ_{12} to yield an equivalent circuit for the charge flowing across the left barrier:

$$\begin{aligned}
\delta J_{12} &= -\frac{q^2 m^* \nu_1}{\pi \hbar^2}\left[f_{1R}(\delta\psi_2 - \delta\phi_1) - f_{2R}(\delta\psi_2 - \delta\phi_1 + \delta\phi_1 - \delta\phi_2) + \eta_1(\delta\psi_2 - \delta\psi_1)\right] \\
&= -\frac{q^2 m^* \nu_1}{\pi \hbar^2}\left[f_{1R}(\delta\psi_2 - \delta\psi_1) - f_{2R}(\delta\psi_2 - \delta\psi_1 + \delta\phi_1 - \delta\phi_2) + \eta_1(\delta\psi_2 - \delta\psi_1)\right] \\
&= G_1(\delta\phi_1 - \delta\phi_2) + G_\alpha(\delta\psi_1 - \delta\psi_2)\,,
\end{aligned} \tag{3.51}$$

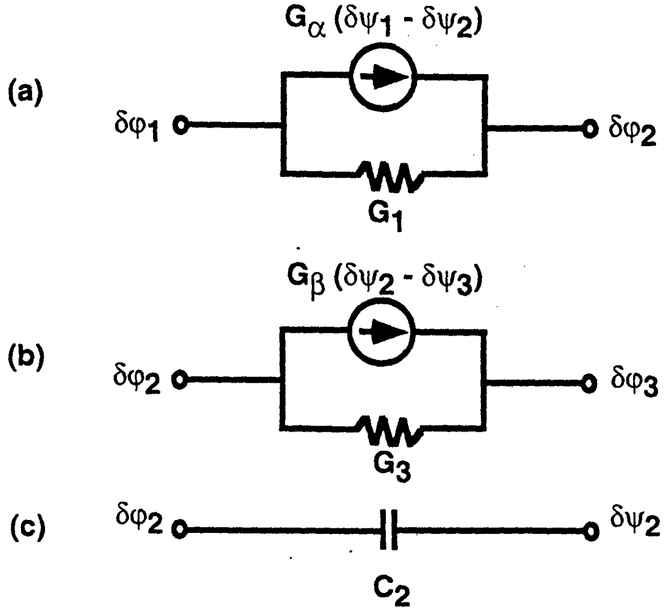

Figure 3-3. Equivalent circuit models for small-signal excitation for (a) the current across the left barrier, (b) the current across the right barrier, and (c) the charge being stored in the well.

where

$$G_1 = \frac{q^2 m^* \nu_1}{\pi \hbar^2}\, f_{2R} \tag{3.52}$$

$$G_\alpha = \frac{q^2 m^* \nu_1}{\pi \hbar^2}\, (f_{1R} - f_{2R} + \eta_1)\ . \tag{3.53}$$

The first component of the current flowing between the cathode and well is proportional to the difference in their quasi-Fermi potentials. This component is physically behaving as a conductance (or equivalently, a resistor) would. The second component of the current is proportional to two other potentials in the circuit, so it is behaving like a transconductance. This transconductance arises because the current tunneling through the left barrier changes when the potential profile changes. Since the potential profile is directly controlled by the electron potentials and not the quasi-Fermi potentials, the additional current occurs as a transconductance in the equivalent circuit. This circuit is shown schematically in Fig. 3-3(a).

Similarly, the equivalent circuit for the current across the right barrier becomes

$$\delta J_{23} = G_3(\delta\phi_2 - \delta\phi_3) + G_\beta(\delta\psi_2 - \delta\psi_3)\,, \tag{3.54}$$

where

$$G_3 = \frac{q^2 m^* \nu_3}{\pi\hbar^2}\, f_{2R} \tag{3.55}$$

$$G_\beta = \frac{q^2 m^* \nu_3}{\pi\hbar^2}\, (f_{3R} - f_{2R} + \eta_3)\,. \tag{3.56}$$

Again, the first component of the current is proportional to the difference in their quasi-Fermi potentials, which is behaving as a conductance. The second component is behaving as a transconductance, and is analogous to the transconductance which arose for the left barrier. The equivalent circuit for this current is shown schematically in Fig. 3-3(b).

In sinusoidal steady state, the change in the incremental well charge becomes

$$\frac{\partial}{\partial t}\delta Q_w = j\omega C_2(\delta\phi_2 - \delta\psi_2) \tag{3.57}$$

where

$$C_2 = \frac{q^2 m^*}{\pi\hbar^2} f_{2R}\,. \tag{3.58}$$

Any change between the electron potential and quasi-Fermi potential in the well results in a change in the charge storage. The charge storage in the well is modeled by a capacitor in Fig. 3-3(c), which is physically sensible.

We now use the incremental-current-continuity equation, Eq. (3.38), to complete the RTD equivalent circuit. Using the expressions in Eqs. (3.51), (3.54) and (3.57) to eliminate δJ_{12}, δJ_{23}, and δQ_w, we find

$$j\omega C_2(\delta\phi_2 - \delta\psi_2) = \\ [G_1(\delta\phi_1 - \delta\phi_2) + G_\alpha(\delta\psi_1 - \delta\psi_2)] - [G_3(\delta\phi_2 - \delta\phi_3) + G_\beta(\delta\psi_2 - \delta\psi_3)] \ . \quad (3.59)$$

This equation is actually enforcing Kirchoff's current law at the node labeled $\delta\phi_2$ in Fig. 3-4(a). Therefore, the equivalent circuit shown in Fig. 3-4(a) is the small-signal equivalent circuit model of the RTD. To complete the model, we add two capacitances. The first capacitance C_1 is the capacitance across the left barrier and represents field storage in that region. The second capacitance C_3 represents the right barrier as well as the depletion region. This complete equivalent circuit is shown in Fig. 3-4(b).

When actually performing calculations of the RTD admittance, it should be noted that there is a useful transformation for the equivalent circuit shown in Fig. 3-4(b). If the current sources were not present, the subcircuit consisting of the nodes $\delta\phi_1$, $\delta\phi_2$, $\delta\phi_3$, and $\delta\psi_2$ could be simplified using a Y-Δ transformation. A modified version of this transformation must be used in order to incorporate the current sources. We have derived the modified Y-Δ transformation and have described its use in Appendix C.

3.4 Calculation of the Transfer Rates

In this section, a method to calculate the transfer rates ν_1 and ν_3 is presented. The calculation will be broken into two parts. First, a technique for finding the time evolution of a wavefunction is presented. This method is useful for determining the time response of the wavefunction at a spatial point given the time behavior at another spatial point. In the second part of this section, a phenomenological method to account for phonon scattering will be introduced. This will allow us to calculate expressions for the transfer rate between two regions.

For brevity, we will present detailed derivations for the transfer only between regions 1 and 2 in this section, which will yield an expression for ν_1. Of course,

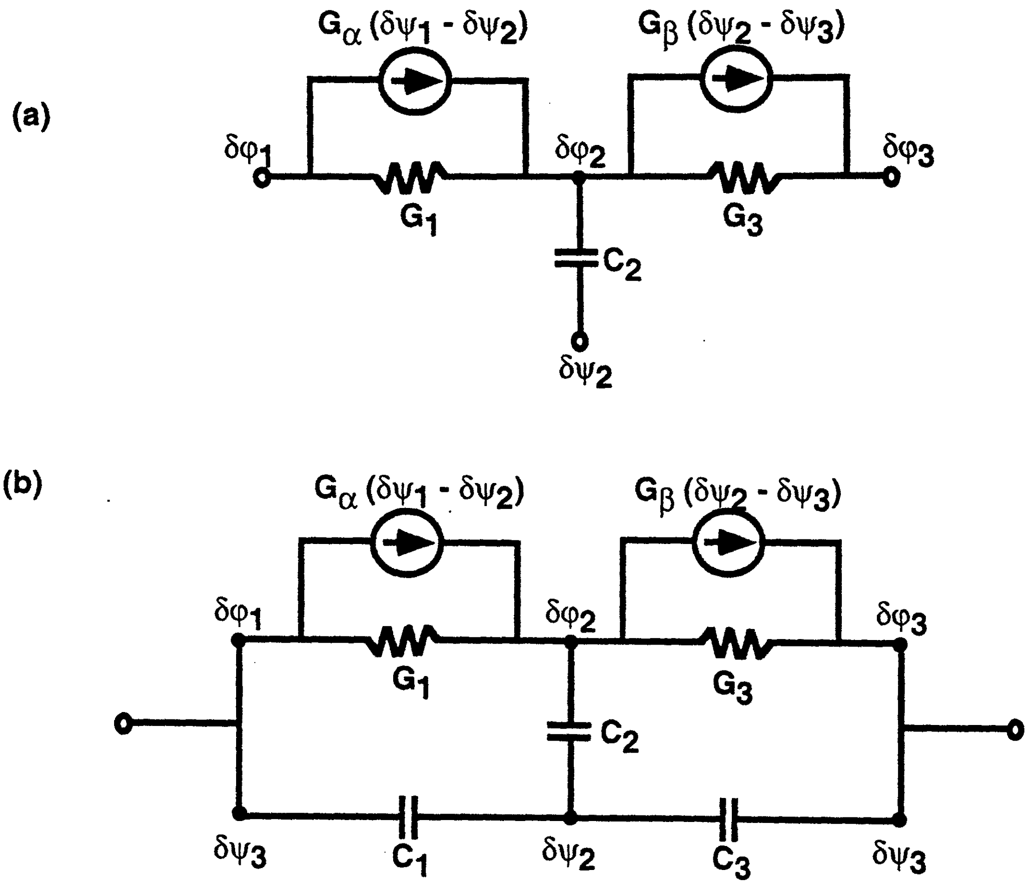

Figure 3-4. Equivalent-circuit model of the RTD for small-signal excitation. (a) The RTD equivalent-circuit model and (b) the RTD equivalent-circuit model with barrier and depletion region capacitances.

similar reasoning applies to the transfer of electrons between regions 2 and 3, and we will simply write out important results where appropriate.

3.4.1 Solution of a Resonant Wavefunction Using the Fourier-Transformed Schrödinger Equation

Solutions for the quantum-mechanical behavior of electrons in double-barrier structures have been published elsewhere [11, 12, 55], and those results will not be replicated here. These methods of solving Schrödinger's equation do not explicitly model the transient properties of the wavefunction that are needed to solve for the transfer rate ν_1. In order to find these properties, we will solve the Fourier transform of Schrödinger's equation, which will provide insight into the time behavior of the wavefunction.

First, we will outline the traditional solution of the Schrödinger equation for the double-barrier problem. Next, we will examine the general properties of the Fourier transform of the Schrödinger equation. It will be seen that the traditional solutions can be used to find explicit solutions of the time evolution of the wavefunction. We will then solve the specific problem of finding the time evolution of an electron placed in the cathode.

Outline of the Traditional Quantum-Mechanical Solution to the RTD

The traditional solution begins with the time-dependent Schrödinger equation*

$$\frac{\mathbf{p}^2}{2m^*}\Psi_T(\mathbf{x},t) + V_R(\mathbf{x})\Psi_T(\mathbf{x},t) = i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{x},t)\,, \tag{3.60}$$

where $V_R(\mathbf{x})$ is the potential energy profile of the RTD and appears schematically in Fig. 3-5. For potential profiles that are independent of time, the wavefunction solution is assumed to be of the form $\Psi_T(\mathbf{x},t) = \Psi(\mathbf{x})T(t)$ so that the Schrödinger equation is separable. It is then found that the time-dependent part is $T(t) = \exp(-iEt/\hbar)$, where E is now the energy of the eigenstate. This solution is the sinusoidal steady state of the wavefunction. The corresponding spatial part obeys the time-independent Schrödinger equation,

$$\frac{\mathbf{p}^2}{2m^*}\Psi(\mathbf{x},E) + \left[V_R(\mathbf{x},E) - E\right]\Psi(\mathbf{x},E) = 0\,. \tag{3.61}$$

---

*In this work, quantum-mechanical quantities are used in conjunction with i and $\exp(-i\omega t)$, but admittances are used in conjunction with j and $\exp(j\omega t)$. This unfortunate clash of notation arises from the differing conventions in the literature.

Mathematically, since the solution Ψ depends on the parameter E in the differential equation, we shall regard Ψ as a function of E as well as x. Once the equation is solved, the transmission amplitude is found by dividing the outgoing wave in the anode by the incident wave in the cathode.

Let the components of Ψ be designated as Ψ_1 in the cathode, Ψ_2 in the well, and Ψ_3 in the anode. We will use the superscripts f and r to denote the forward- and reverse-traveling components, as shown in Fig. 3-5. Thus, the individual wavefunction in a region n is given by $\Psi_n = \Psi_n^f + \Psi_n^r$. With this notation, the transmission amplitude t is

$$t = \frac{\Psi_3^f(\mathbf{x}_3, E)}{\Psi_1^f(\mathbf{x}_1, E)}, \tag{3.62}$$

where $\mathbf{x}_1$ and $\mathbf{x}_3$ are points located in the cathode and anode, respectively. For systems with a resonance, the transmission amplitude is often cast into the Breit-Wigner form,

$$t = \frac{\Psi_3^f}{\Psi_1^f} = t_0 \frac{i\Gamma/2}{(E - E_R) + i\Gamma/2}. \tag{3.63}$$

This form of t is valid for energies near the resonance. Mathematically, the Breit-Wigner form is equivalent to expanding the transmission amplitude in a Laurent series about a pole [12]. This leads to a transmission amplitude that is a highly peaked function at energy E_R, with a FWHM in energy of Γ. The t_0 factor is sometimes written as a weak function of energy or is often approximated as a constant over the range of the energy peak. We shall treat t_0 as a constant.

Fourier Approach to the Solution of Schrödinger's Equation

As with the traditional solution, we start with the time-dependent Schrödinger equation, Eq. (3.60). In this case however, we first apply a Fourier transform, which

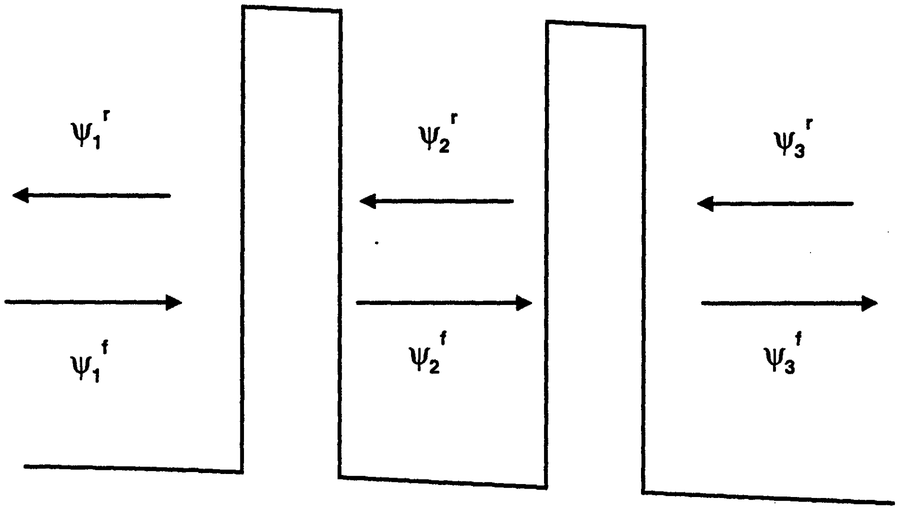

Figure 3-5. Definitions of the wavefunction components for the double-barrier structure.

yields a frequency-dependent Schrödinger equation,

$$\frac{\mathbf{p}^2}{2m^*}\Psi(\mathbf{x},\omega) + [V_R(\mathbf{x}) - \hbar\omega]\,\Psi(\mathbf{x},\omega) = 0\,, \tag{3.64}$$

where ω is the Fourier-transformed time variable. In a purely mathematical sense, this equation is equivalent to the time-independent Schrödinger equation, which is seen by mapping $E \to \hbar\omega$ in Eq. (3.61). As a consequence, we can find solutions to Eq. (3.64) by mapping $E \to \hbar\omega$ in the solutions of Eq. (3.61). Taking advantage of this useful property, we simply perform the mapping in the Breit-Wigner solution to find

$$t = \frac{\Psi_3^f(\mathbf{x}_3,\omega)}{\Psi_1^f(\mathbf{x}_1,\omega)} = t_0\,\frac{i\Gamma/2}{\hbar(\omega-\omega_R)+i\Gamma/2}\,, \tag{3.65}$$

where $\hbar\omega_R = E_R$. It is worth emphasizing the physical difference between this equation and the previous Breit-Wigner solution. The previous solution in Eq. (3.63) is the ratio of the incident and outgoing waves in sinusoidal steady state, while the present solution in Eq. (3.65) is the ratio of the Fourier-transformed waves, which can be used to obtain information about the time evolution. In the Fourier scheme, the meaning of the transmission amplitude is much like a response function of a circuit. The incident wave at one point plays the role of the circuit excitation, and the outgoing wave at another plays the role of the circuit output. The transmission coefficient between these two points in space is playing the role of the Fourier transform of the impulse response.

We again consider the case of sinusoidal excitation, in order to relate this Fourier-transform approach to the traditional RTD tunneling. The incident-excitation function in region 1 is varying sinusoidally in time, so there is a frequency-dependent factor* which is given by $2\pi\delta(\omega - \omega_1)$, where ω_1 is the frequency of oscillation. The excitation function (which is the incident wave) is then

$$\Psi_1^f(\mathbf{x}_1, \omega) = \Psi_1^+(\mathbf{x}_1, \omega)\, 2\pi\delta(\omega - \omega_1)\,, \tag{3.66}$$

where $\Psi_1^+(\mathbf{x}, \omega)$ describes the spatial dependence of the wavefunction in the cathode, which is normalized in the box L_R. The output function (which is the transmitted wave) is the multiplication of the excitation (which is the incident wave) with the frequency response function (which is the transmission amplitude) given in Eq. (3.65). This multiplication yields

$$\Psi_3^f(\mathbf{x}_3, \omega) = \Psi_1^+(\mathbf{x}_1, \omega)\, t_0 \frac{i\Gamma/2}{\hbar(\omega - \omega_R) + i\Gamma/2}\, 2\pi\delta(\omega - \omega_1) \tag{3.67}$$

for the output function. The δ function makes the inverse transform of this equation

---

*We are using the Fourier-transform convention

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{i\omega t} dt\,.$$

easy to perform, and it is given by

$$\Psi_3^f(\mathrm{x}_3, t) = \Psi_1^+(\mathrm{x}_1, \omega_1)\, t_0 \, \frac{i\Gamma/2}{\hbar(\omega_1 - \omega_R) + i\Gamma/2}\, e^{-i\omega_1 t} \,. \tag{3.68}$$

This expression is precisely what we would have found if we had calculated the response in the sinusoidal steady state from the outset.

Calculation of the Wavefunction Evolution in the RTD Well

The Fourier technique is well suited to calculating how an electron in the cathode evolves into a one-dimensional RTD well over time. We must first find the frequency response function, which is done by finding the solutions of the time-independent Schrödinger equation in the three regions and then mapping $E \rightarrow \hbar\omega$ as before. Since the wavefunction in the well has both forward and reverse traveling wave components, there are both forward and reverse frequency response functions. They are defined as follows:

$$A = \frac{\Psi_2^f(x_2, \omega)}{\Psi_1^f(x_1, \omega)} \qquad \text{forward frequency response function}$$

$$B = \frac{\Psi_2^r(x_2, \omega)}{\Psi_1^f(x_1, \omega)} \qquad \text{reverse frequency response function} \,.$$

The nature of the double-barrier structure is such that only a narrow range of energies has any significant probability amplitude in the well, so it is expected that A and B are highly peaked functions. Unfortunately, the analytical solutions for A and B are intractable when using realistic potential profiles that include the accumulation and depletion layers. We will instead examine numerical solutions.

A potential profile appears in Fig. 3-6 for the representative device presented in Chapter 2. (The I-V characteristic of this device appears in Fig. 2-3.) The profile was found by solving Poisson's equation for an applied bias of 1.0 V. The solutions to the Schrödinger equation were found using T-matrices [11], which is a conventional

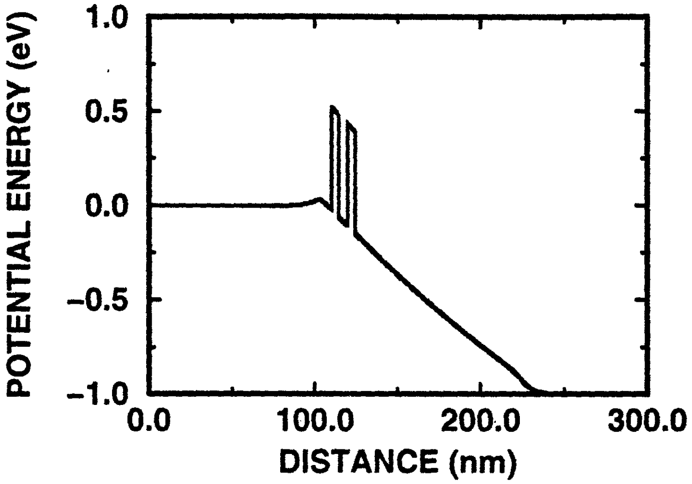

Figure 3-6. Band structure for the representative device at a bias of 1.0 V. The barrier heights have been shortened for clarity.

numerical technique for this problem. Briefly, the potential profile is approximated by many small regions of width Δx, each having a constant potential. For each region of size Δx, the constant potential is given by the average value of the true potential over that region. This is often called a staircase approximation. The procedure to calculate the transmission of the wavefunction between any two adjacent regions is then straightforward and can easily be iterated for all the regions.

Using this technique, we have solved the time-independent Schrödinger equation and have plotted the magnitude for $A(E)$ for x_2 in the center of the well in Fig. 3-7(a). We have chosen x_1 far enough away from the double-barrier structure that a change in x_1 results only in a phase change for $A(E)$. The peak magnitude has been normalized to one from its original value of approximately 210; a wavefunction which is more confined to the well will have a higher peak value. The abscissa has been plotted in units of the FHWM Γ_1. As we now demonstrate, a Breit-Wigner function is an excellent approximation to data in Fig. 3-7(a). A plot of the difference between the true magnitude and the Breit-Wigner magnitude over a range of 10 Γ_1 appears in Fig. 3-7(b). As can be seen from the figure, the Breit-Wigner approximation slightly

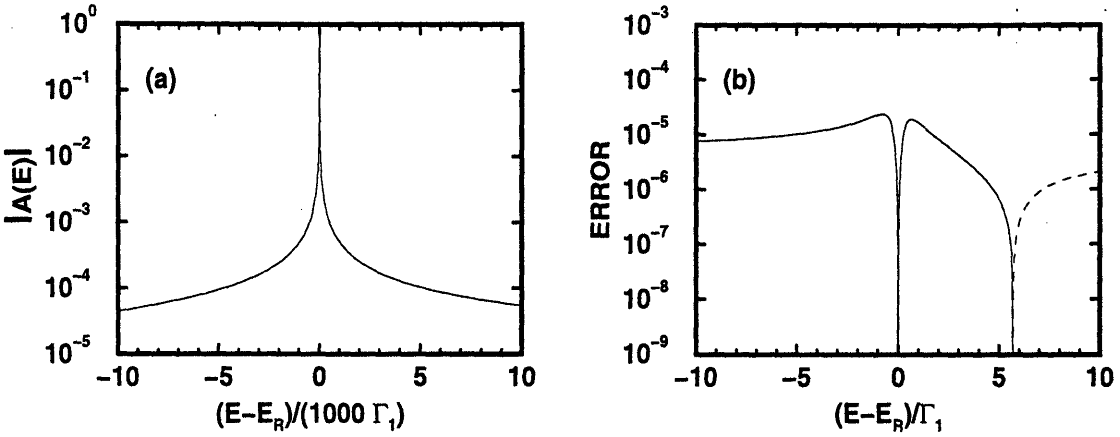

Figure 3-7. (a) Calculated value of the magnitude of $A(E) = \Psi_2^f(x_2, E)/\Psi_1^f(x_1, E)$ using T-matrices, and (b) difference between a Breit-Wigner function and the data shown in (a). The solid and dashed lines indicate energies where the Breit-Wigner function overestimates and underestimates the true function, respectively. Traditionally, $A(E)$ would be related to the transmission amplitude from cathode to well. In this work, we show that $A(E)$ yields the frequency response of a forward-traveling wave in the well to an impulse in the cathode.

overestimates the true value for energies between -10 and 5.7 Γ_1, and then it underestimates the true value for energies above 5.7 Γ_1. These trends continue for the range not shown in the figure: The Breit-Wigner approximation overestimates the true value between -10,000 and -10 Γ_1 with an error of less than 10^{-5}; it underestimates the true value between 10 and 10,000 Γ_1, also with an error of less 10^{-5}. Later, we will be concerned with only the square magnitude of this function, so the amount of error over this range will be quite insignificant.

The calculated value for $B(E)$ at the center of the well is plotted Fig. 3-8(a). The peak value is approximately the same as that for $A(E)$, which occurs because the quasibound level is almost a stationary state. We have chosen to normalize this curve using exactly the peak value for $A(E)$. Again, the Breit-Wigner function is an excellent approximation to the true function, and the error is seen in Fig. 3-8(b). As shown in the figure, there is a larger amount of error at the peak. The larger error

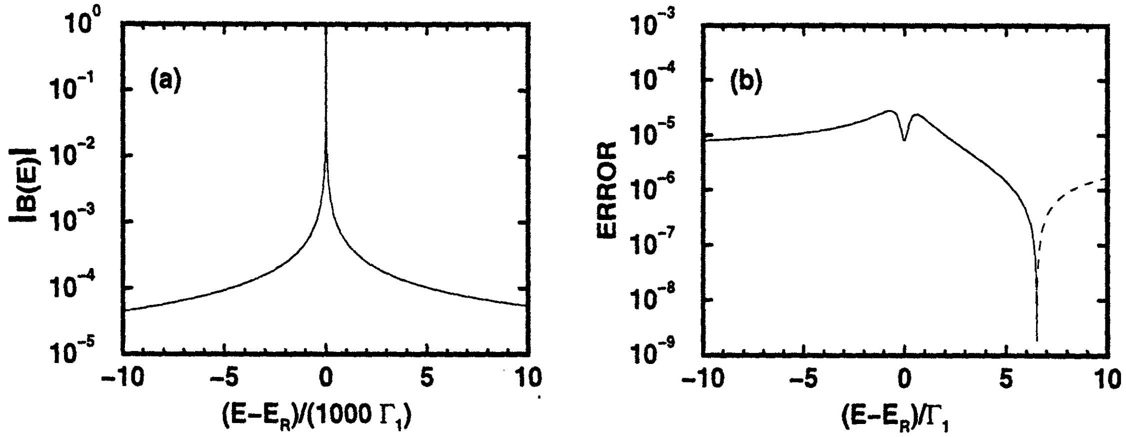

Figure 3-8. (a) Calculated value of the magnitude of $B(E) = \Psi_2^r(x_2, E)/\Psi_1^f(x_1, E)$ using T-matrices, and (b) difference between a Breit-Wigner function and the data shown in (a). The solid and dashed lines indicate energies where the Breit-Wigner function overestimates and underestimates the true function, respectively.

arises from our choice of normalization and is a measure of the equality in magnitudes between the forward and reverse traveling waves. Still, the error is always less than 10^{-4}, and is less than 10^{-5} for less than -7 Γ_1 or greater than 2 Γ_1. As before, these trends continue outside the range of the figure, to energies of $\pm$10,000 Γ_1.

In order to further enhance the confidence in the Breit-Wigner approximation, we will examine the phase associated with $A(E)$ and $B(E)$. Figures 3-9(a) and 3-10(a) on pages 77 and 78, respectively, show the phase calculated from the T-matrices for a range of 20 Γ_1 for $A(E)$ and $B(E)$, respectively. This range is where the phase is changing most rapidly. The differences between the phase calculated by the T-matrices and by the Breit-Wigner function is shown in Figs. 3-9(b) and 3-10(b), from which it is seen that the errors are quite small. (Note the ordinate scale change between the latter two figures.) The errors in the phase over a range of 20,000 Γ_1 are shown in 3-9(c) and 3-10(c). Even over this large range, the worst error in the phase is less than 5 degrees.

Since we are expecting only a narrow range of energies to have significant probability amplitude in the well, it is not too surprising that the Breit-Wigner function is a good approximation. Mathematically, the approximation is again tantamount to writing a Laurent series about the pole nearest to the excitation. Thus, we use the form shown in Eq. 3.63, and map the energies to frequencies to write

$$A(x_2,\omega_1) \approx A_o(x_2)\,\frac{i\Gamma_1/2}{\hbar(\omega_1-\omega_R)+i\Gamma_1/2} \tag{3.69}$$

$$B(x_2,\omega_1) \approx B_o(x_2)\,\frac{i\Gamma_1/2}{\hbar(\omega_1-\omega_R)+i\Gamma_1/2}\,, \tag{3.70}$$

where $\hbar\omega_1 = E$ is the frequency of excitation, ω_R is the frequency of the peak, x_2 is the x coordinate in the well, A_o and B_o are the calculated peak values of A and B, respectively, and are approximated as constants. We have suppressed the dependence on x_1 since its value only contributes a phase when x_1 is chosen far enough away from the double-barrier structure. Taking the inverse Fourier transforms of A and B is a straightforward task and yields

$$A(x_2,t) = A_o(x_2)\,\frac{\Gamma_1}{2\hbar}\,e^{-\Gamma_1 t/2\hbar}e^{-i\omega_R t}\,\theta(t) \tag{3.71}$$

$$B(x_2,t) = B_o(x_2)\,\frac{\Gamma_1}{2\hbar}\,e^{-\Gamma_1 t/2\hbar}e^{-i\omega_R t}\,\theta(t)\,, \tag{3.72}$$

where $\theta(t)$ is the Heaviside unit-step function

$$\theta(t) = \begin{cases} 0 & t<0 \\ 1 & t\geq 0\,. \end{cases} \tag{3.73}$$

Equations (3.71) and (3.72) are equivalent to the impulse response of the traveling waves in the well.*

To calculate the transition rate between region 1 and 2, we must find the response

---

*This result was derived from the approximation of a one-pole frequency response. Such an approach implicitly neglects the barrier penetration time and the transit time across the well.

in the well once an excitation in the cathode has been "turned on" at $t = 0$. This excitation is written as

$$\Psi_1^f(x_1, t) = \Psi_1^f(x_1, \omega_1) e^{-i\omega_1 t}\theta(t)\,, \tag{3.74}$$

where $\Psi_1^f(x_1, \omega_1)$ contains the spatial information about the wavefunction. To compute $\Psi_2^f(x_2, \omega)$, we would multiply the frequency response function, Eq. (3.69), by the Fourier transform of the excitation. This computation is equivalent to convolving the impulse response, Eq. (3.71), with the time-domain excitation, Eq. (3.74). The convolution results in the expression

$$\Psi_2^f = A_o(x_2) \cdot \frac{i\Gamma_1/2}{\hbar(\omega_1 - \omega_R) + i\Gamma_1/2} \left[1 - e^{-\Gamma_1 t/2\hbar} e^{-i(\omega_R - \omega_1)t}\right] e^{-i\omega_1 t}\theta(t)\,, \tag{3.75}$$

where again we have suppressed the dependence on x_1 since its value contributes only a phase when x_1 is chosen far enough away from the double-barrier structure. Similarly for the reverse-traveling wave, the time evolution is

$$\Psi_2^r = B_o(x_2) \cdot \frac{i\Gamma_1/2}{\hbar(\omega_1 - \omega_R) + i\Gamma_1/2} \left[1 - e^{-\Gamma_1 t/2\hbar} e^{-i(\omega_R - \omega_1)t}\right] e^{-i\omega_1 t}\theta(t)\,. \tag{3.76}$$

Thus, when an electron is prepared in the cathode state, its wavefunction gradually leaks into the well as a ringing exponential. The time scale of the exponential varies inversely with the width of the transmission. This result is similar to that in Ref. [12].

The reasoning which lead to Eqs. (3.75) and (3.76) can also be applied to the transition from anode to well. For this case, the transmission from the incident anode state to the forward and reverse well states will also be described by Breit-Wigner functions. These functions will have the same resonant frequency and FWHM. We convolve the impulse response functions with a step excitation in an anode state,

$$\Psi_3^r(x_3,t) = \Psi_{x3}^r(x_3,\omega_3)e^{-i\omega_3 t}\theta(t)\,, \tag{3.77}$$

where $\Psi_{x3}^r(x_3,\omega_3)$ describes the spatial extent of the wavefunction in the anode, ω_3 is the frequency of excitation, and x_3 is a point located in the anode. Convolving this excitation with the impulse response results in the time evolution

$$\Psi_2^r = C_o(x_2)\cdot\frac{i\Gamma_1/2}{\hbar(\omega_3-\omega_R)+i\Gamma_1/2}\left[1-e^{-\Gamma_1 t/2\hbar}e^{-i(\omega_R-\omega_3)t}\right]e^{-i\omega_3 t}\theta(t) \tag{3.78}$$

$$\Psi_2^f = D_o(x_2)\cdot\frac{i\Gamma_1/2}{\hbar(\omega_3-\omega_R)+i\Gamma_1/2}\left[1-e^{-\Gamma_1 t/2\hbar}e^{-i(\omega_R-\omega_3)t}\right]e^{-i\omega_3 t}\theta(t)\,, \tag{3.79}$$

where $C_o(x_2)$ and $D_o(x_2)$ are the maximum values of the forward and reverse traveling waves. Like the case for x_1, we have suppressed the dependence on x_3 since its value contributes only a phase when x_3 is sufficiently far from the double-barrier structure.

3.4.2 Calculation of the Transfer Rates ν_1 and ν_3 in the Presence of Phonon Collisions

When calculating the transfer rates ν_1 and ν_3, we will use a phenomenological approach to account for the scattering. We regard phonon collisions as a Poisson process, with a mean collision time τ. From the time $t = 0$ that an electron is placed in the cathode, the probability that it suffers a collision in the time interval between t and $t + dt$ is $\exp(-t/\tau)\,dt/\tau$. Then, for electrons in a given initial state, $\exp(-t/\tau)\,dt/\tau^2$ is the average rate of collisions that occur between t and $t + dt$ if the state is refilled instantaneously after each collision. Phenomenologically, we regard the collision as a measurement; the phase information about the wavefunction is lost, and we determine the position of the electron. The probability that an electron is measured between x_2 and $x_2 + dx_2$ when this collision occurs is $|\Psi_2(x_2,t)|^2\, dx_2$, where $\Psi_2(x_2,t) = \Psi_2^f(x_2,t) + \Psi_2^r(x_2,t)$, with the latter given by Eqs. (3.75) and (3.76). Therefore, the probability per second of a collision having occurred within the bounds of the well is

$$w_{12} = \frac{1}{\tau} \int_0^\infty \frac{e^{-t/\tau}}{\tau} \int_{\text{well}} |\Psi(x_2,t)|^2\, dx_2\, dt\,. \tag{3.80}$$

This approach enables us to calculate the longitudinal transfer rate.*

From Eqs. (3.75) and (3.76), the squared magnitude of the wavefunction in the well is

$$\begin{aligned} |\Psi_2(x_2,t)|^2 &= \left|\Psi_2^f(x_2,t) + \Psi_2^r(x_2,t)\right|^2 \\ &= |A_o(x_2) + B_o(x_2)|^2 \frac{\Gamma_1^2/4}{\hbar^2(\omega_1 - \omega_R)^2 + \Gamma_1^2/4}\, \gamma_1(t)\,, \end{aligned} \tag{3.81}$$

where

$$\gamma_1(t) = 1 + e^{-\Gamma_1 t/\hbar} - 2e^{-\Gamma_1 t/2\hbar} \cos\left[(\omega_1 - \omega_R)t\right]\,. \tag{3.82}$$

Using this expression in the scattering integral, Eq. (3.80), we find that the integral over time can be performed analytically and yields

$$w_{12} = \frac{\Gamma_1 \Lambda_1}{4\hbar} \cdot \frac{\Gamma_s \Gamma_1}{\hbar^2(\omega_1 - \omega_R)^2 + (\Gamma_s + \Gamma_1/2)^2} \cdot \frac{2\Gamma_s + \Gamma_1}{\Gamma_s + \Gamma_1}\,, \tag{3.83}$$

where

$$\Lambda_1 = \int_{\text{well}} |A_o(x_2) + B_o(x_2)|^2\, dx_2\,, \tag{3.84}$$

---

*We note that the scattering and coherent tunneling are being treated on unequal footing. The coherent tunneling is solved assuming no interference from a scattering event, but the scattering event terminates the tunneling.

and the scattering width Γ_s is defined to be $\hbar/\tau$. The longitudinal transfer rate now incorporates scattering, with a form similiar to that of Stone and Lee [17].

The above expression for w_{12} enables us to calculate the rate from the definition of ν_1. Substituting Eq. (3.83) into Eq. (3.25), we find

$$\nu_1 = \frac{2\Gamma_s + \Gamma_1}{\Gamma_s + \Gamma_1} \int_{E_{C1}}^{\infty} \frac{dE_1 \Lambda_1 \sqrt{m^*}}{8\pi\hbar^2 \sqrt{2(E_1 - E_{C1})}} \cdot \frac{\Gamma_s \, \Gamma_1^2}{(E_1 - E_R)^2 + (\Gamma_s + \Gamma_1/2)^2}, \tag{3.85}$$

where we have mapped the frequency to energy $E_1 = \hbar\omega_1$. The integral over E_1 can be solved analytically to yield

$$\nu_1 = \frac{\Lambda_1 \sqrt{m^*}}{16\hbar^2} \cdot \frac{\Gamma_1^2 \Gamma_s}{E_{\Gamma_1} \left[E_{\Gamma_1} - (E_R - E_{C1})\right]^{1/2}} \cdot \frac{2\Gamma_s + \Gamma_1}{\Gamma_s + \Gamma_1}, \tag{3.86}$$

where

$$E_{\Gamma_1} = \sqrt{(E_R - E_{C1})^2 + (\Gamma_s + \Gamma_1/2)^2}. \tag{3.87}$$

We will examine the properties of this expression for ν_1 when we are comparing the theory with the experimental results in the next chapter.

As a final note, we write down the transfer rate from well to anode as

$$\nu_3 = \frac{\Lambda_3 \sqrt{m^*}}{16\hbar^2} \cdot \frac{\Gamma_1^2 \Gamma_s}{E_{\Gamma_3} \left[E_{\Gamma_3} - (E_R - E_{C3})\right]^{1/2}} \cdot \frac{2\Gamma_s + \Gamma_1}{\Gamma_s + \Gamma_1}, \tag{3.88}$$

where

$$E_{\Gamma_3} = \sqrt{(E_R - E_{C3})^2 + (\Gamma_s + \Gamma_1/2)^2} \tag{3.89}$$

$$\Lambda_3 = \int_{\text{well}} \left|C_o(x_2) + D_o(x_2)\right|^2 dx_2 \,. \tag{3.90}$$

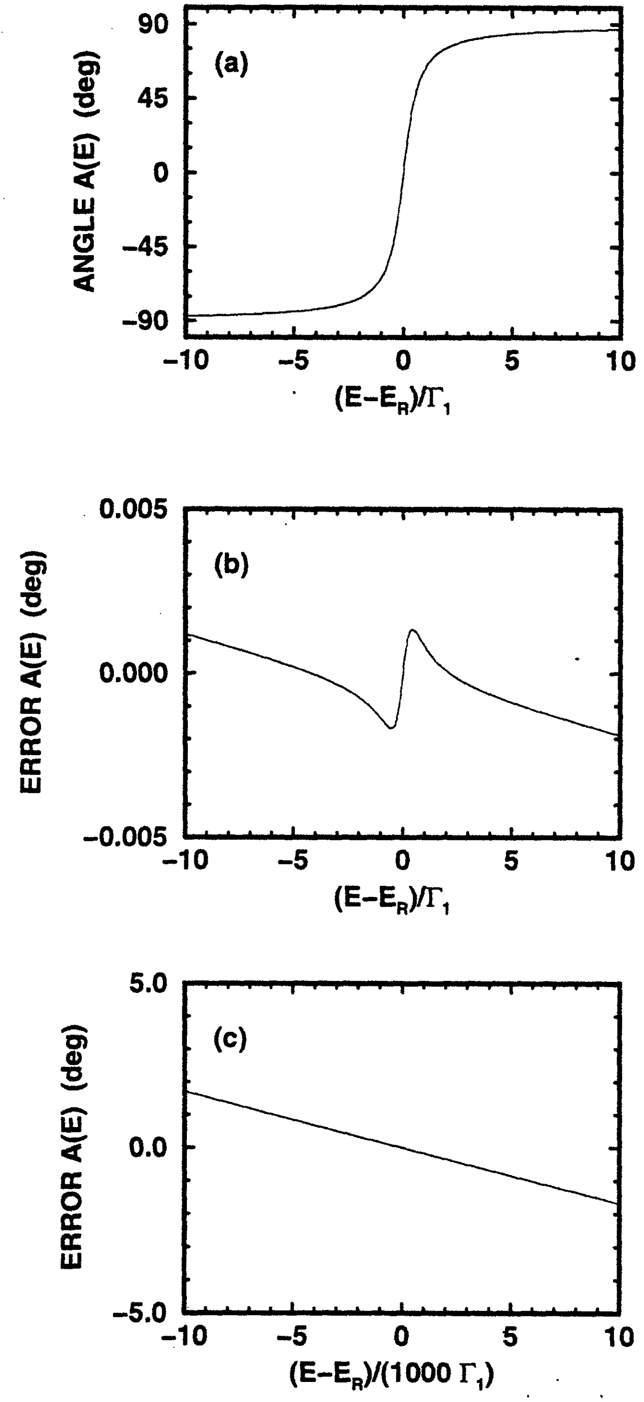

Figure 3-9. (a) Phase of $A(E) = \Psi_2^f/\Psi_1^f$. (b) Difference between a Breit-Wigner function and the data shown in (a) for a range of 20 Γ_1, and (c) difference for a range of 20,000 Γ_1. The error introduced by the Breit-Wigner approximation is small.

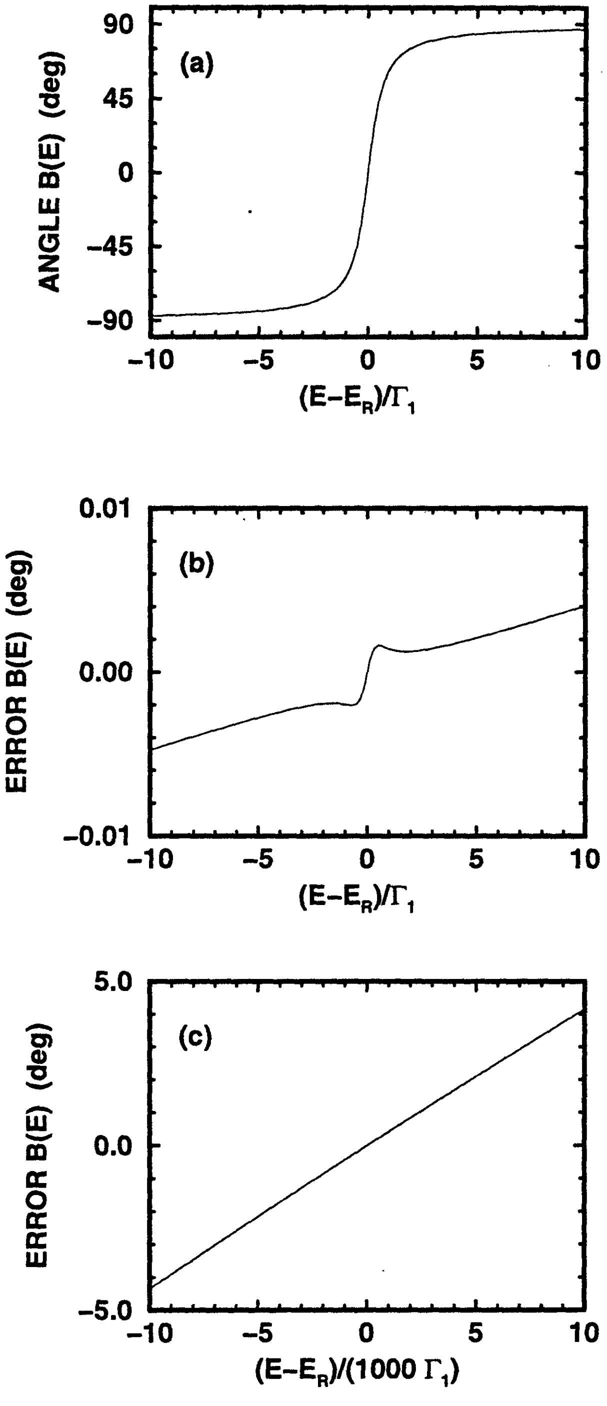

Figure 3-10. (a) Phase of $B(E) = \Psi_2^r/\Psi_1^f$. (b) Difference between a Breit-Wigner function and the data shown in (a) for a range of 20 Γ_1, and (c) difference for a range of 20,000 Γ_1. The error introduced by the Breit-Wigner approximation is again small.

4

Comparison of Theory and Experiment

In this chapter, we will evaluate the agreement between the experimental data in Chapter 2 and the theoretical model in Chapter 3. In order to perform the comparison, a computer program was written to calculate the theoretical behavior. The program self-consistently calculates the band structure and quantum-mechanical behavior with greater accuracy than would be feasible analytically.

The numerical solutions in this chapter will be used to explain the observed behavior of the RTD. However, it will be shown that for some purposes, the results are accurate enough that the program could be used as a device simulation tool during the design phase of a fabrication process. Alternatively, the program could be used to predict ac behavior for fabricated devices, using only the I-V curve and knowledge of the fabrication, which is useful since the experimental effort involved in obtaining an I-V characteristic is much less than that of the high-frequency ac behavior.

We will begin by describing the components of the simulator. We will then compare the theoretical and measured results for both the static and dynamic behavior. By fitting the conductance at three frequencies, we will show that values for the physical parameters introduced in the previous chapter can be extracted.

4.1 Brief Description of Numerical Techniques

The numerical calculation has three components to solve the distinct parts of the problem. One component solves Poisson's equation, another solves Schrödinger's equation, and the third calculates the expressions from the RTD model in the previous chapter. The program uses well-established numerical methods to simplify the task of programming, so we will give only a rough description of the numerical techniques here. (A detailed and lucid discussion of these techniques may be found in Press, *et al.* [56].)

The first part of the program is the Poisson solver. The electrostatic potential is solved throughout the RTD using Poisson's equation

$$\nabla^2\psi = -\frac{\rho(\psi)}{\epsilon}, \tag{4.1}$$

where ϵ is the permittivity, and ρ is the charge, which is a function of the potential (and implicitly the quasi-Fermi level). The doping level is high enough that the charge must be calculated using degenerate statistics. The numerical solution to Poisson's equation was accomplished using a finite difference scheme. This involves choosing a set of discrete points in space x_m that are separated by a distance Δx. The potential at each discrete point is then related to the potential at the two adjacent points by

$$\frac{\psi(x_{m+1}) - 2\psi(x_m) + \psi(x_{m-1})}{\Delta x^2} = \frac{q}{\epsilon}(n - N_D^+), \tag{4.2}$$

where n and N_D^+ are the density of electrons and ionized donors, respectively. For a grid of M points, there are $M-2$ equations based on Eq. (4.2), which involve M unknown potentials. The 2 remaining unknowns are given by the potentials at the end of the grid, where the potential is presumably easy to calculate since it is far away from the double-barrier structure. The M equations are solved using a Newtonian iteration algorithm [56]. For the solutions shown in this chapter, Δx is 0.1 nm and M is greater than 2500, which provide an excellent approximation to the band structure.

After Poisson's equation is solved, the potential solution is used to find the solutions to Schrödinger's equation. To calculate the wavefunction, we employed the T-matrix approach, which was described on page 13. For a particular energy, the program calculates the wavefunction which is incident from the cathode and another from the anode. Once this wavefunction is known throughout the structure, the program calculates the forward and reverse frequency response functions. At this point, it iterates until it finds the resonant energy in the response functions. Once the response function at the peak is found, the program solves for the FWHM by explicity searching for an energy which has a response of half the peak. The FWHM is used to calculate the rates ν_1 and ν_3, which in turn determine the quasi-Fermi level in the well from the condition that $J_{12} = J_{23}$. The charge densities associated with the quasibound state and any accumulation-layer states are added into ρ and Poisson's equation is solved again. This process continues until the potential and wavefunctions converge to their final values.

In calculating the wavefunctions, there is a difficult issue regarding the potential energy and effective mass in the barriers. The conduction band minimum occurs at the X point in the AlAs barriers, but at the Γ point in the $In_{0.53}Ga_{0.47}As$ cathode, well and anode. Experimentally, it has been shown [20] that tunneling through a barrier is best described by the X-point barrier height for thick layers and by the Γ-point barrier height for thinner layers. Also, there has been a large amount of work dedicated to the issue of heterojunctions with materials of differing mass (see [57, 58], for example). We have found that using the Γ-point barrier height and the effective mass of the $In_{0.53}Ga_{0.47}As$ throughout the structure will yield suitably accurate results for the purposes of this thesis.

The third component of the numerical solution begins by computing the rates ν_1 and ν_3 from Eqs. (3.86) and (3.88) in Section 3.4. The rates were calculated with an assumed scattering time of 0.1 ps, which corresponds to a scattering width Γ_s of 6.5 meV. At applied voltages near the current peak, the calculated rates depend only on the square root of Γ_s, which means that they are not strongly dependent on this assumption.

Once the rates are known, the dc current and charge may be calculated according to the method in Section 3.2 and the ac response may be calculated from the expressions presented in Section 3.3. Since our calculation accounted for only a single resonance and elastic tunneling, the valley current will not be modeled well. However, enough of this region will be simulated to provide insight into the ac behavior for the NDR region.

4.2 Comparison of the Theoretical and Measured Current

In this section, we present three calculations of the dc current. The first calculation was performed using the values for the RTD structure given in Section 2.1. Comparing the calculated and actual current for these values will aid us in evaluating the program as a predictive simulation tool. In the second calculation, the barrier thicknesses were adjusted so that the simulated peak current matched the measured value. The third calculation was the same in all respects to the second, except that the well width was reduced by 0.1 nm in order to further test the sensitivity of the I-V to the structure parameters.

The first calculation using the structure parameters specified in the growth of the RTD would be the approach of a designer in the initial phases of device design. At that stage, a designer would be finalizing a fabrication process and would want a CAD tool that could simulate the electrical behavior of a device for a given set of growth parameters.

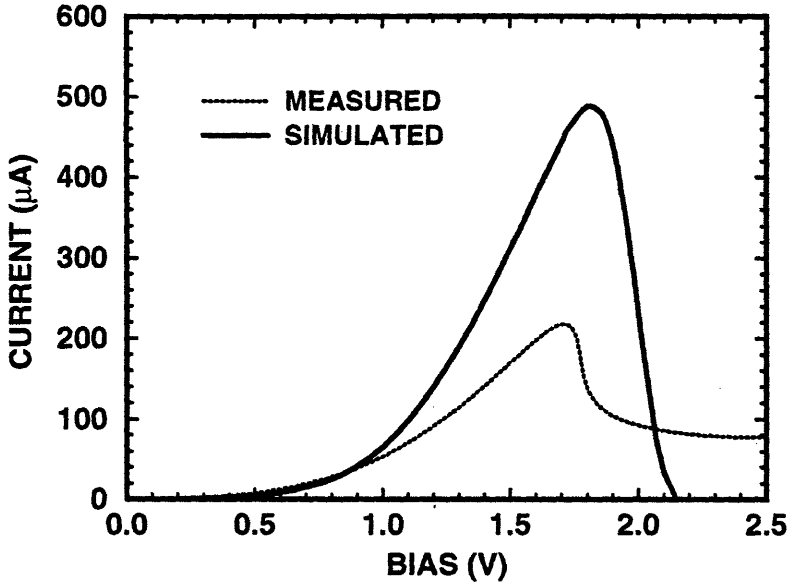

Figure 4-1. Comparison of the calculated (solid line) and measured (dotted line) I-V characteristic. The calculation used device parameters specified in Fig. 2-1, except for those listed in Table 4.1. Due to strain, the barrier thicknesses listed in the table are thinner than those in Fig. 2-1. Also the doping of the 10-nm spacer layer was changed to reflect the small length of this region.

The parameters for this simulation are taken from Fig. 2-1, with the exception of those summarized in Table 4.1. The final two parameters in this table differ from those given in Fig. 2-1. The value of the barrier thicknesses was chosen to reflect the strain, which was briefly discussed on page 20. Also, the doping of the 10-nm spacer layer was changed from its nominal value of 2×10^{16} cm^{-3} . For this doping, we would expect to find an average of only one dopant atom in a cube that is 37 nm on a side. Thus, it is clear that specifying a continuous doping is a poor approximation for this 10 nm region. We have instead modified the value of continuous doping by using the following crude approximation: Consider the expected location of the dopant that is nearest to the first RTD barrier on the cathode side. Since it is improbable to find any dopants in the 10-nm spacer layer, we approximate it as being devoid of dopants. Then the dopant nearest to the first barrier is in the region with a doping of 2×10^{18} cm^{-3}. The expectation value for the location of the nearest dopant atom is approximately 4 nm from the boundary between the high- and low-doped regions. A

Relative effective mass	0.037
Relative permittivity	13.7
Barrier height (Γ-band offset)	2.24 eV
Barrier thicknesses	4.1 nm
10-nm spacer layer doping	3.6×10^{17} cm^{-3}

Table 4.1. Parameters for the calculated current shown in Fig. 4-1; parameters not listed are given in Fig. 2-1. The parameters for the simulation results in Fig. 4-2 are the same, except that the barrier thicknesses were changed to 4.5 nm.

cube which is $10+4=14$ nm on a side containing only one dopant would correspond to a doping level of 3.6×10^{17} cm^{-3}. This is the doping used for the 10-nm spacer layer.

The calculated I-V characteristic for the above parameters appears as a solid line in Fig. 4-1. The I-V from the actual device (originally shown in Fig. 2-3) is shown as a dotted line for comparison. As can be seen from the figure, the simulated current is approximately 2.3 times the measured current, and the simulated peak voltage occurs 0.1 V higher than the measured peak voltage. We note that the overall shape of the curve is similar except for the valley region, which is to be expected. This degree of prediction is quite reasonable for our RTD since the thicknesses of the strained layers are difficult to predict with great accuracy.

In the second simulation, we adjusted the barrier thicknesses so that the calculated and actual peak currents were the same. This yielded barrier thicknesses of 4.5 nm. In all other respects, the parameters for this simulation and the previous one are the same. Unfortunately, for these size barriers, we are at the limit of the numerical accuracy of the computer. For some biases, it was found that the quantum-mechanical currents for the forward and reverse directions were not equal and could differ by 10%. However, this level of numerical inaccuracy does not introduce a large amount of error, so we will examine the behavior of the simulation.

The result, shown in Fig. 4-2, is that the shape of the simulated I-V is quite

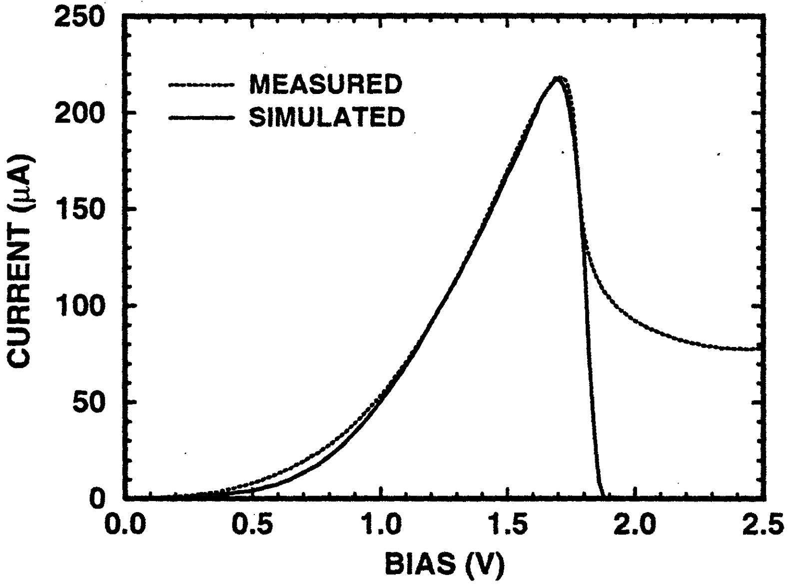

Figure 4-2. Comparison of the calculated (solid line) and measured (dotted line) I-V characteristic. The barriers for this simulation have been adjusted so that the calculated and measured peak current agree. All other structure parameters are the same as those in Fig. 4-1. Note that the peak voltage agrees quite well by fitting only the current.

similar to the measured I-V and that the peak value occurs at a applied potential 20 mV less than the actual potential. The largest disagreement between the two curves (excluding the valley current, which has not been included in the model) occurs at biases near 0.7 V. Further simulations suggest that the I-V for these biases are sensitive to differing values of the doping in the 10-nm spacer layer.

In order to test the sensitivity of the fit to the well width, we have performed a third calculation. This simulated current is calculated from the same parameters as the second simulation except for the well width. The well width was decreased by 0.1 nm, which is much less than thickness of a monolayer. The results are shown in Fig. 4-3. The I-V characteristic is noticeably off from the value of the peak, demonstrating that the fit is sensitive to the values of the well width.

We now examine some of the quantities that were calculated in the simulation shown in Fig. 4-2. The calculated rates ν_1 and ν_3 are shown by the solid lines in Fig. 4-4(a) and (b), respectively. (The points with error bars were extracted from

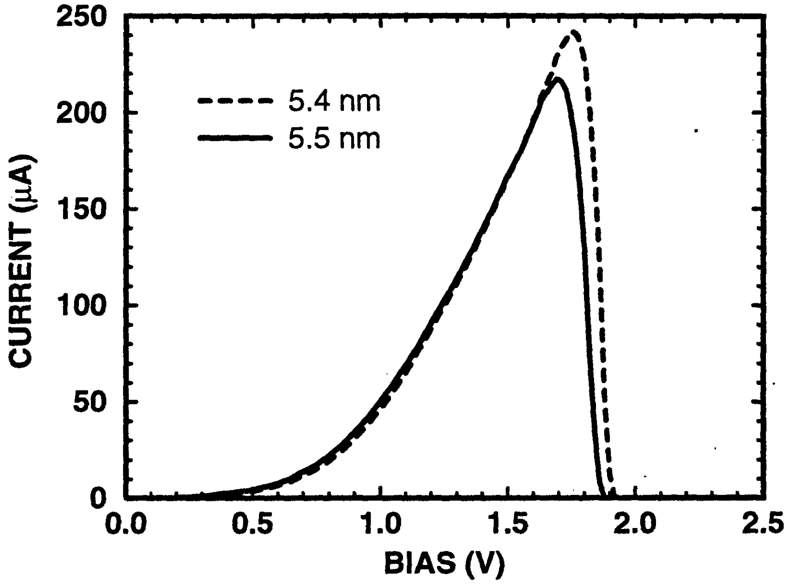

Figure 4-3. Effect on the calculated I-V characteristic of a small change in the well width. Solid line is the simulated I-V for a well width of 5.5 nm (from Fig. 4-2), and dashed line is for a well width of 5.4 nm.

the ac data and will be discussed in the following section.) The rate ν_1 is seen to be decreasing, with a slight local maximum centered at the peak voltage V_p. The rate ν_3 is monotonically increasing. The calculated difference between the quasi-Fermi levels in the cathode and well, $F_1 - F_2$, is shown by the solid line in Fig. 4-5. (Again, the points with error bars were extracted from the ac data and will be discussed later.) The quasi-Fermi level in the well drops precipitously near the peak voltage. This behavior is sensible; the rate at which electrons are transfered from the cathode also drops precipitously, so F_2 will be closer to F_3.

4.3 Comparison of Theoretical and Experimental AC Behavior

In this section, we examine the ac behavior of the RTD. We will first examine the predicted ac behavior using only the as-specified growth parameters, and we will

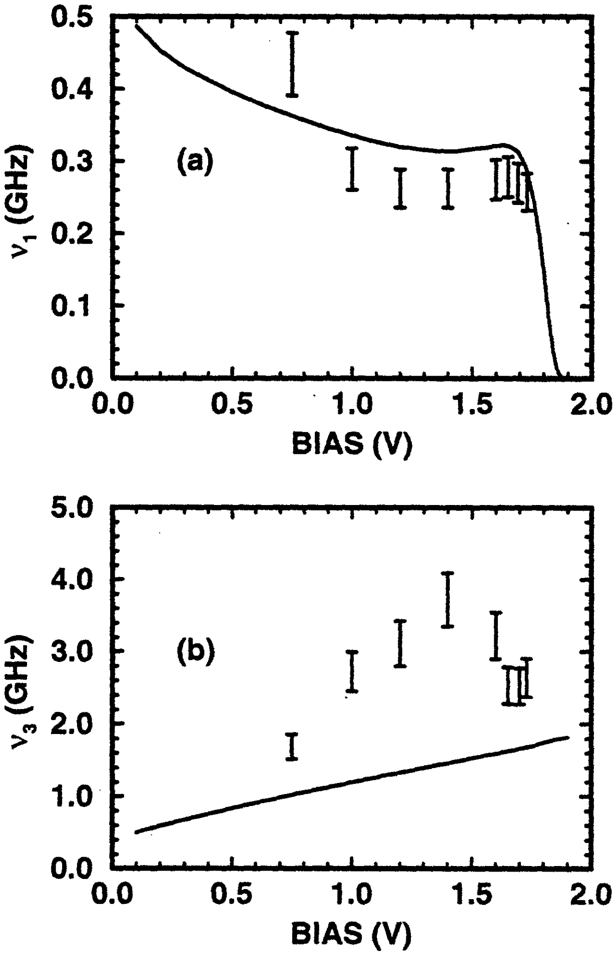

Figure 4-4. Electron transfer rates across (a) the left barrier, ν_1, and (b) the right barrier, ν_3. Solid lines were calculated for the simulation shown in Fig. 4-2, and the points with error bars were extracted from the ac data.

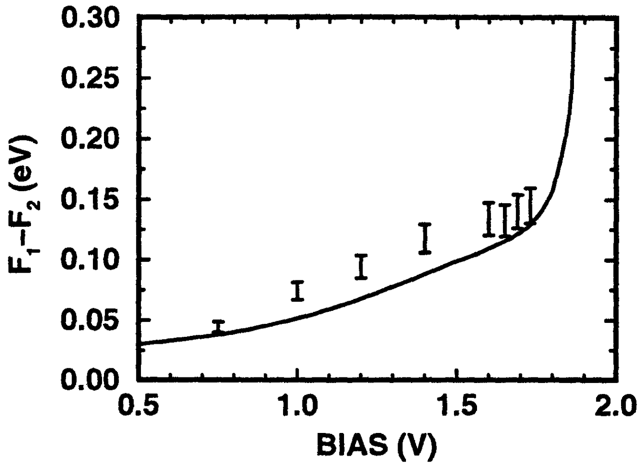

Figure 4-5. Difference between the quasi-Fermi levels in the cathode and well. Solid line was calculated for the simulation shown in Fig. 4-2, and the points with error bars were extracted from the ac data.

then examine the ac behavior that is predicted by fitting the calculated peak value of the dc current with the measured value. The latter approach could potentially save considerable experimental effort by predicting the ac response using only a dc measurement, which is extremely easy to perform in a laboratory.

We will first examine the ac response associated with the simulation shown in Fig. 4-1. In Fig. 4-6 and 4-7, we have plotted the conductance G_R vs frequency for biases of 0.75, 1.40, 1.80, and 1.88 V. These plots show several similar features to the measured plots presented in Section 2.4. In the PDR region, the conductance exhibits a rolloff to a nonzero value. Also, the rolloff frequency is lower at higher biases. In the NDR region, the conductance is negative at low frequencies and becomes positive at higher frequencies.

Figure 4-7 shows the "excess" susceptance vs frequency. (As detailed on page 34, the excess susceptance $B_R - \omega C_s$ represents the deviation of the RTD susceptance B_R from that of the capacitor C_s calculated by averaging B_R/ω in the 3.9 to 4.0 GHz range.) These plots also behave much like those presented in Section 2.4. The excess susceptance is negative in the PDR region but becomes positive in the NDR region.

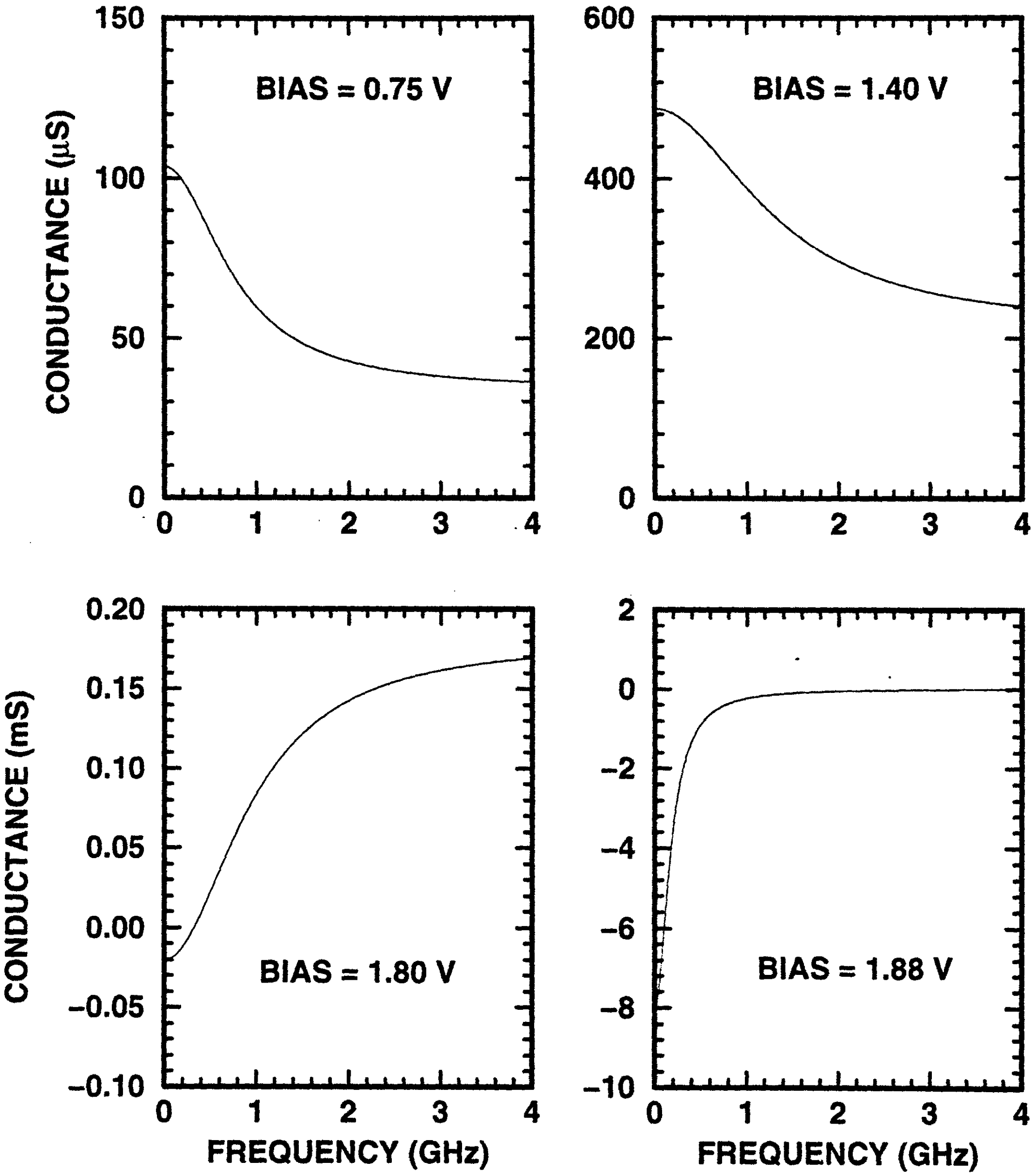

Figure 4-6. Simulated conductance G_R for the RTD at biases of 0.75, 1.40, 1.80, and 1.88 V as indicated. This ac response corresponds to the simulation shown in Fig. 4-1.

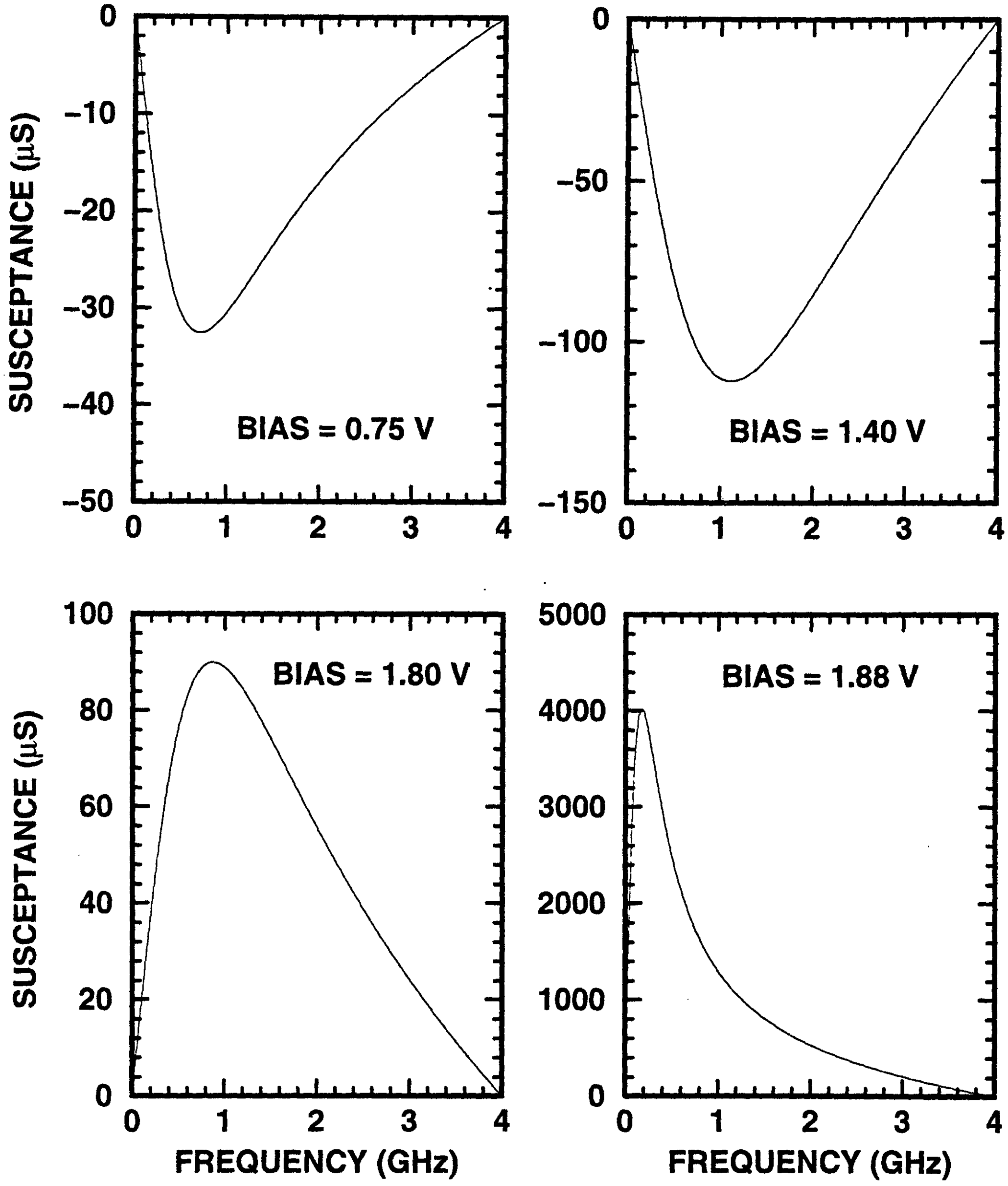

Figure 4-7. Simulated excess susceptance $B_x = B_R - \omega C_s$ for the RTD at biases of 0.75, 1.40, 1.80, and 1.88 V as indicated. The excess susceptance represents the deviation of the RTD susceptance B_R from that of the capacitor C_s calculated by averaging B_R/ω in the 3.9 to 4.0 GHz range. This ac response corresponds to the simulation shown in Fig. 4-1.

The frequency of the extremum is lower at higher biases.

Next we examine the ac response associated with the simulation shown in Fig. 4-2. This simulated response is in better quantitative agreement with the data, as will be seen in Figs. 4-8 through 4-13. Figures 4-8, 4-10, and 4-12 show conductance vs frequency, and Figs. 4-9, 4-11, and 4-13 show the excess susceptance vs frequency. Each one of these plots has three curves. The first is the intrinsic-admittance data and is shown as thin solid lines. These data were taken from the representative device, and were already presented at the end of Section 2.4. The second set of curves, which appears as dotted lines, is the calculated ac response corresponding to the simulation shown in Fig. 4-2. We denote it as the ac simulation. (As mentioned in the last section, this simulation encountered numerical accuracy problems for some biases.) The third set of data will be denoted as the "extracted" response and appears as a dashed line.

The extracted response is essentially a fitted curve. The fit was achieved at each bias by using the measured dc current as well as the measured conductance at three frequencies: dc, one point in the conductance rolloff, and high frequency (averaged in the neighborhood of 4 GHz). As always, the dc conductance is the same as the slope of the I-V characteristic. From these measured points, we can extract values for the unknown parameters in the ac model presented in Chapter 3.

For a single bias, the model has six unknown values: the quasi-Fermi level F_2 in the well, the rates ν_1 and ν_3, and the derivative of these three quantities with respect to bias. In this case, the derivatives are treated as unknowns since we are extracting F_2, ν_1, and ν_3 at only one bias. In order to solve for these six parameters, we need six conditions:

1. J_{12} must equal the measured dc current density.

2. $J_{12} = J_{23}$, which relates F_2 to the rates ν_1 and ν_3.

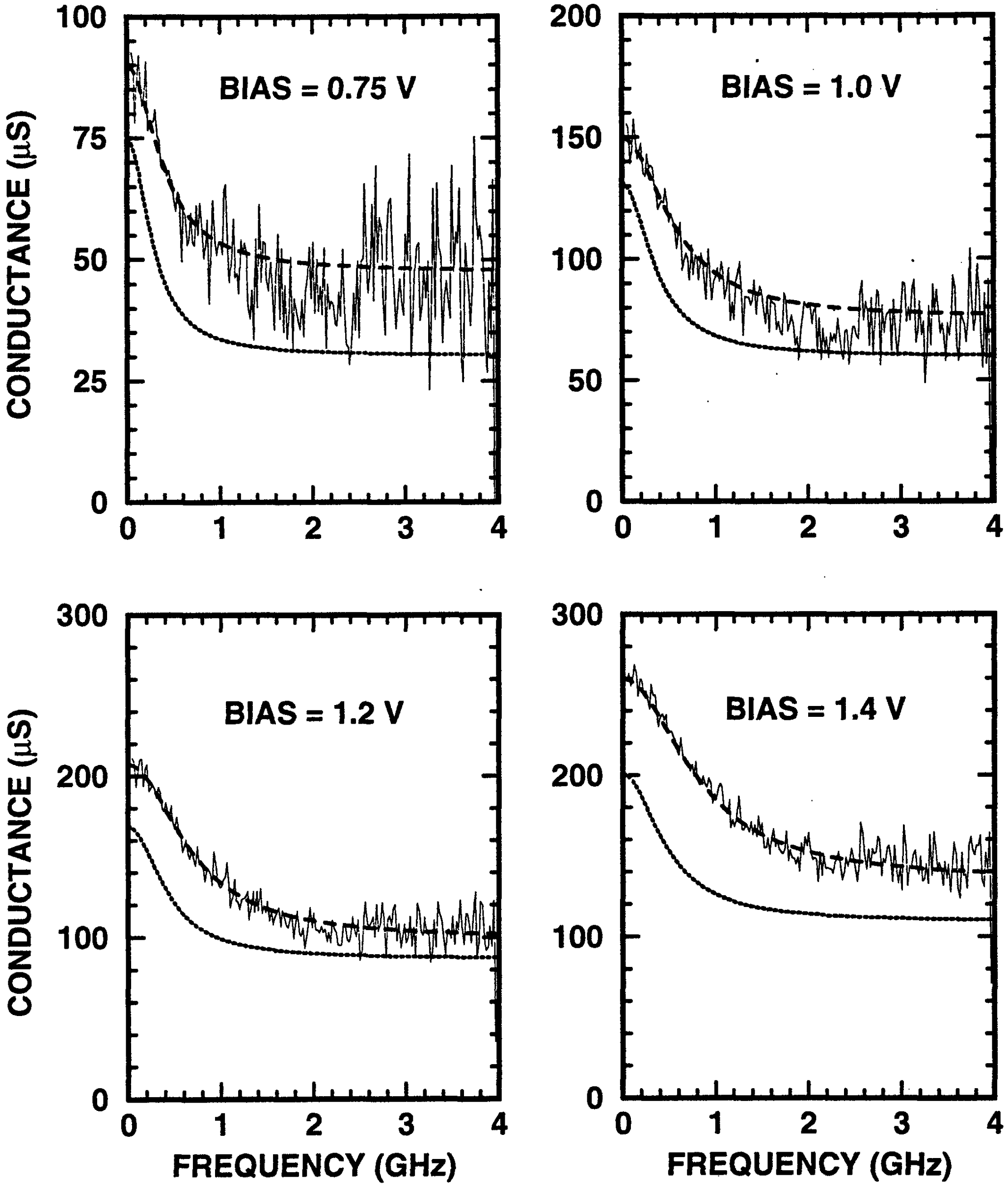

Figure 4-8. Measured and simulated conductance G_R for the RTD at biases of 0.75, 1.00, 1.20, and 1.40 V as indicated. The lines are drawn as follows: solid is measured data from Fig. 2-11, dotted is the ac response for the simulation shown in Fig. 4-2, and dashed is fitted using measured conductance at three frequencies (see text) as well as the measured dc current.

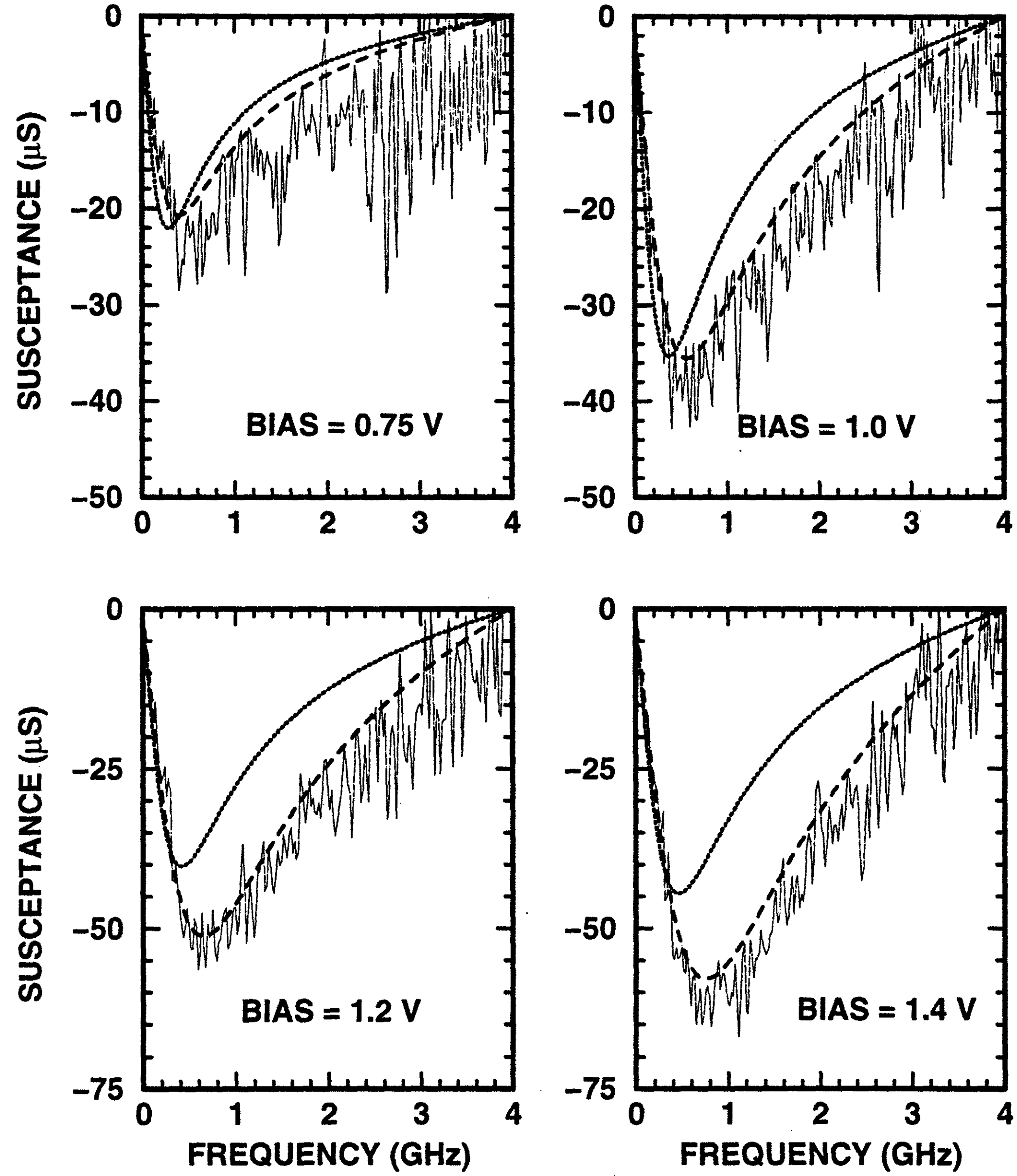

Figure 4-9. Measured and simulated excess susceptance $B_x = B_R - \omega C_s$ for the RTD at biases of 0.75, 1.00, 1.20, and 1.40 V as indicated. The excess susceptance represents the deviation of the RTD susceptance B_R from that of the capacitor C_s calculated by averaging B_R/ω in the 3.9 to 4.0 GHz range. The lines are drawn as follows: solid is measured data from Fig. 2-13, dotted is the ac response for the simulation shown in Fig. 4-2, and dashed is fitted using measured conductance at three frequencies (see text) as well as the measured dc current.

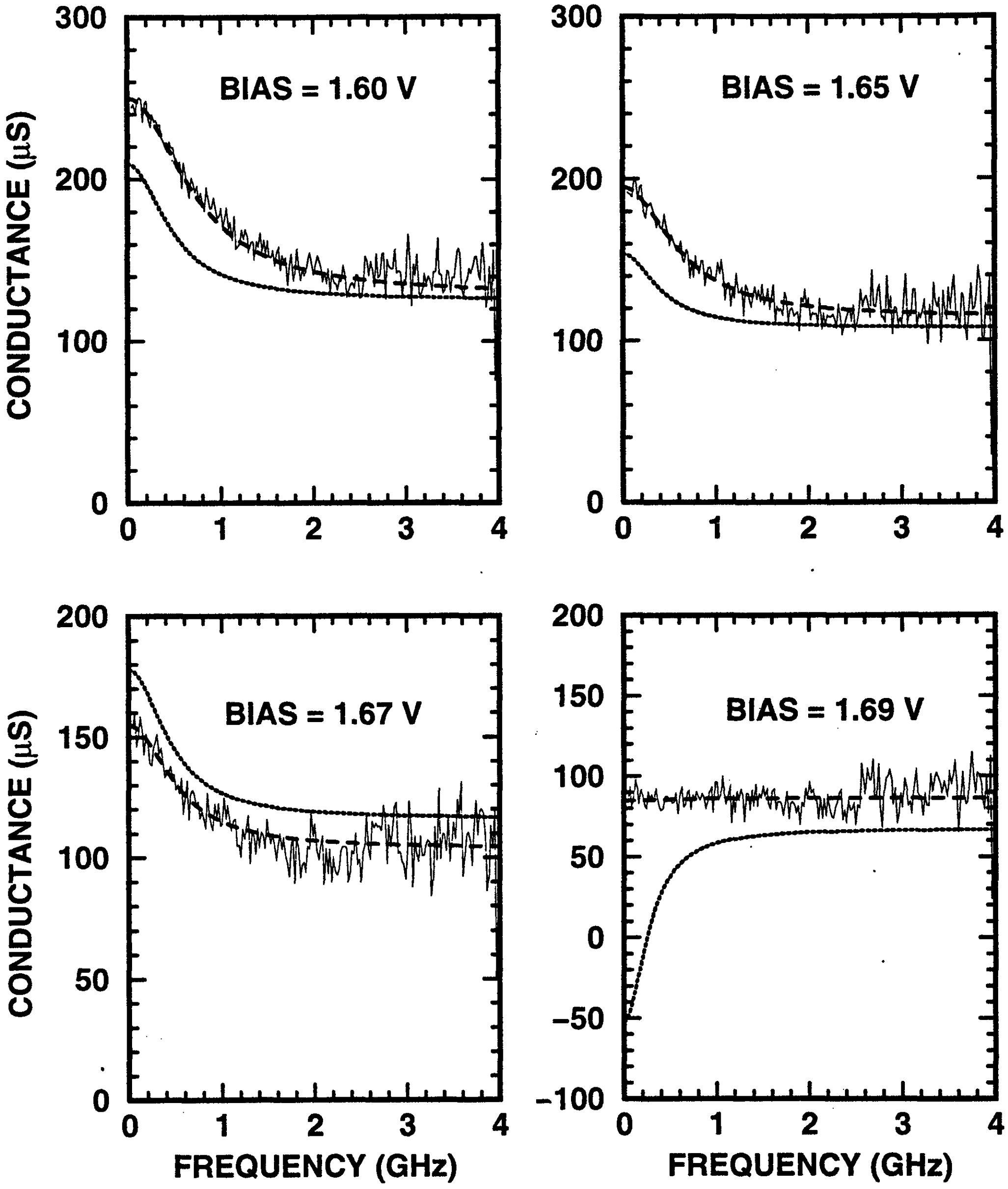

Figure 4-10. Measured and simulated conductance G_R for the RTD at biases of 1.60, 1.65, 1.67, and 1.69 V as indicated. The lines are drawn as follows: solid is measured data from Fig. 2-14, dotted is the ac response for the simulation shown in Fig. 4-2, and dashed is fitted using measured conductance at three frequencies (see text) as well as the measured dc current.

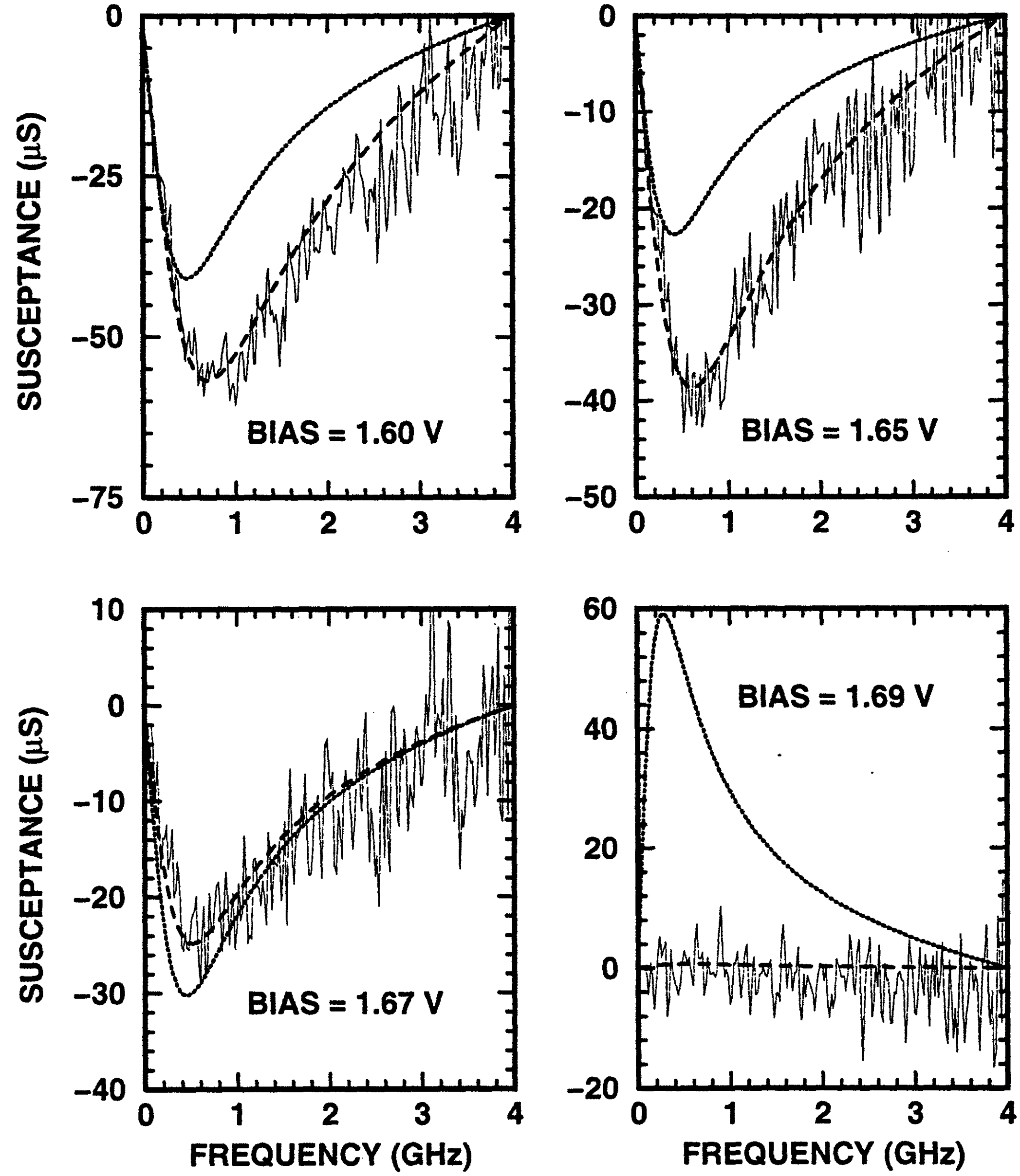

Figure 4-11. Measured and simulated excess susceptance $B_x = B_R - \omega C_s$ for the RTD at biases of 1.60, 1.65, 1.67, and 1.69 V as indicated. The excess susceptance represents the deviation of the RTD susceptance B_R from that of the capacitor C_s calculated by averaging B_R/ω in the 3.9 to 4.0 GHz range. The lines are drawn as follows: solid is measured data from Fig. 2-15, dotted is the ac response for the simulation shown in Fig. 4-2, and dashed is fitted using measured conductance at three frequencies (see text) as well as the measured dc current.

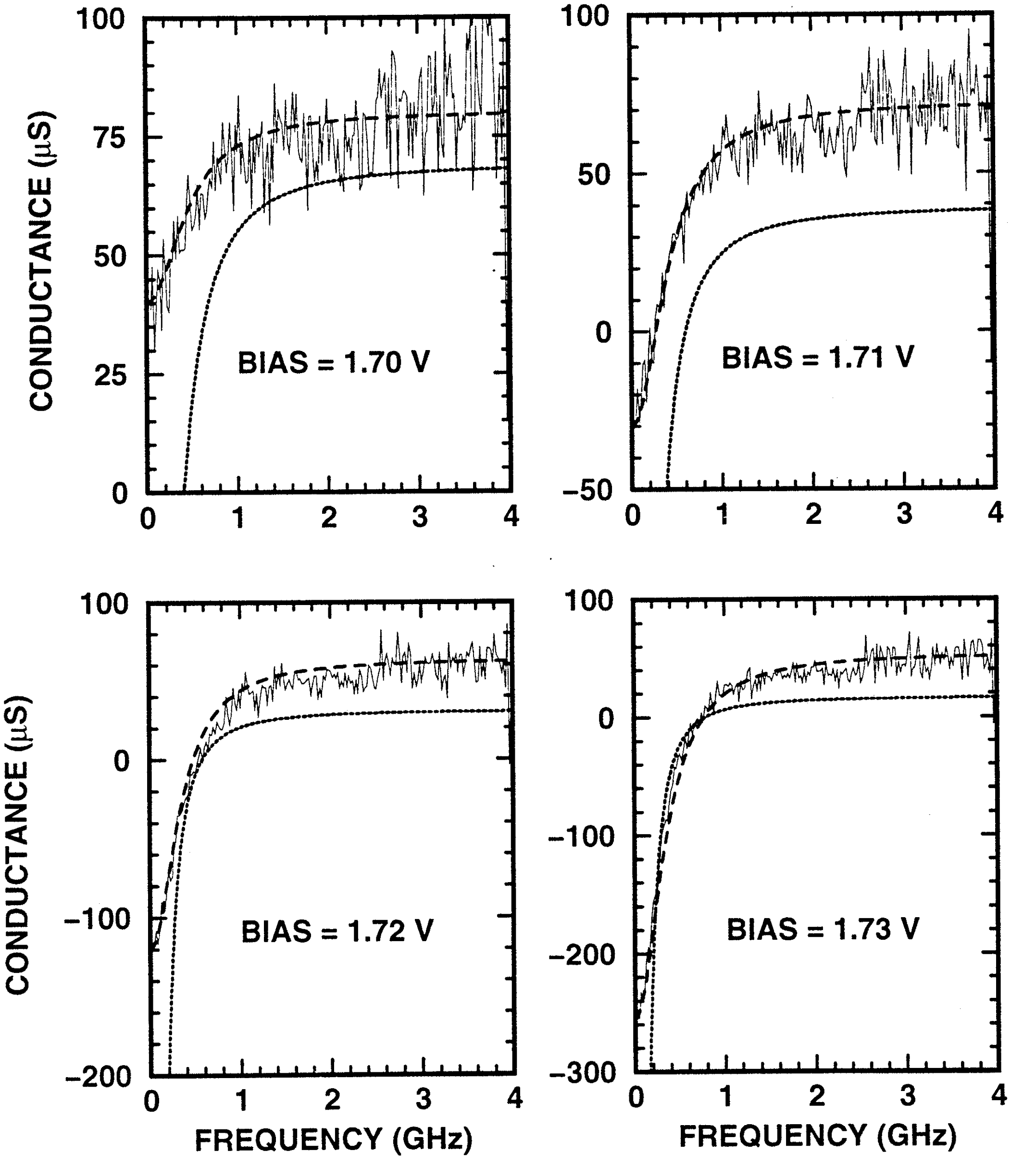

Figure 4-12. Measured and simulated conductance G_R for the RTD at biases of 1.70, 1.71, 1.72, and 1.73 V as indicated. The lines are drawn as follows: solid is measured data from Fig. 2-16, dotted is the ac response for the simulation shown in Fig. 4-2, and dashed is fitted using measured conductance at three frequencies (see text) as well as the measured dc current.

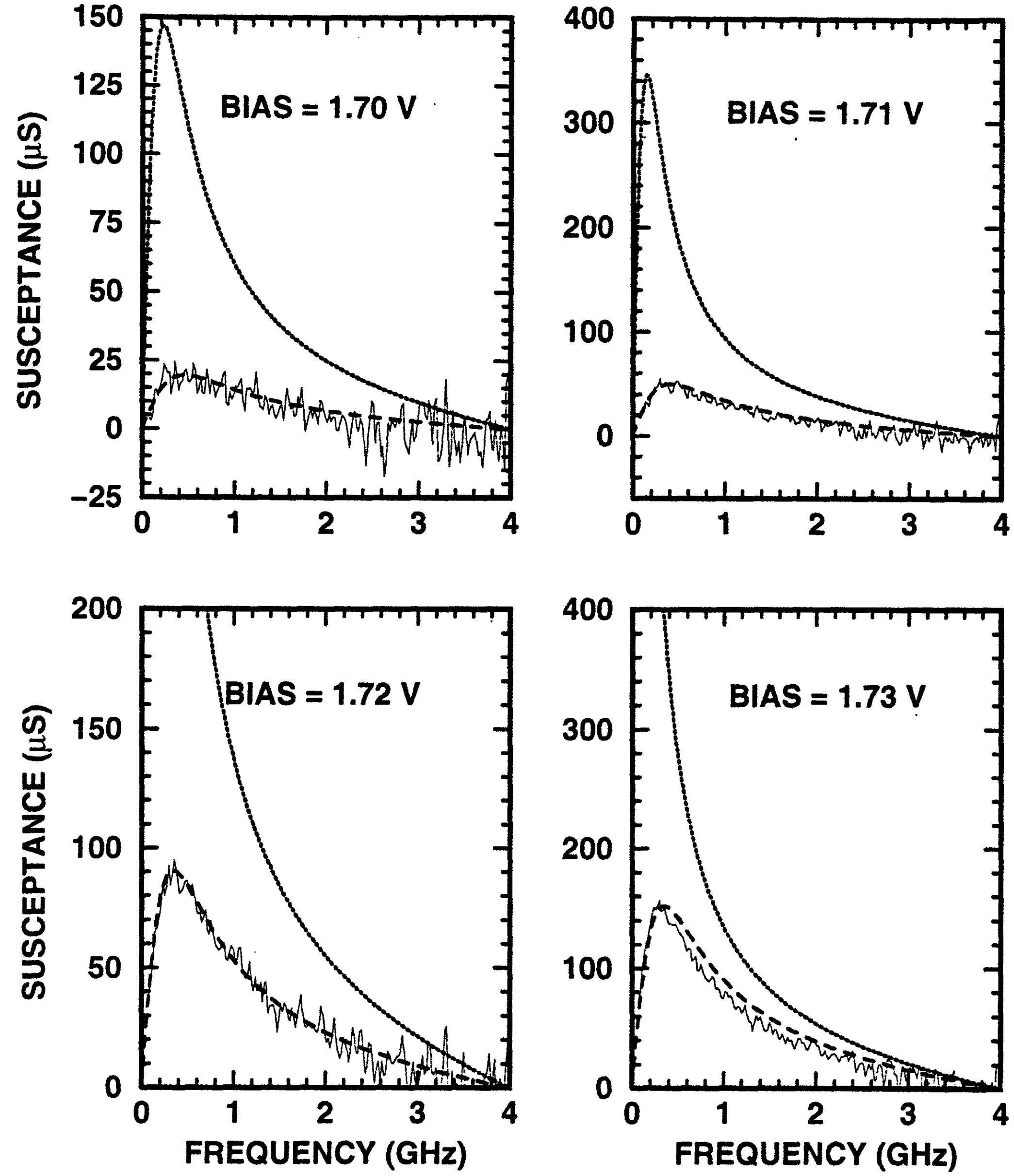

Figure 4-13. Measured and simulated excess susceptance $B_x = B_R - \omega C_s$ for the RTD at biases of 1.70, 1.71, 1.72, and 1.73 V as indicated. The excess susceptance represents the deviation of the RTD susceptance B_R from that of the capacitor C_s calculated by averaging B_R/ω in the 3.9 to 4.0 GHz range. The lines are drawn as follows: solid is measured data from Fig. 2-17, dotted is the ac response for the simulation shown in Fig. 4-2, and dashed is fitted using measured conductance at three frequencies (see text) as well as the measured dc current.

3. $\delta J_{12}/\delta V$ must equal the dc value of the conductance, where δJ_{12} and δV are the incremental change in the current J_{12} and applied voltage, respectively.

4. $\delta J_{23}/\delta V$ must also equal the dc value of the conductance, where δJ_{23} is the incremental change in the current J_{23}.

5. The measured value of the conductance at high frequency must equal the value calculated from the circuit shown in Fig. 3-4(b).

6. The measured value of the conductance at one frequency in the rolloff must equal the value calculated from the circuit shown in Fig. 3-4(b).

These six conditions allow us to find the six unknown parameters. Once the parameters are known, the ac behavior is calculated and is drawn on the plots using a dashed line. We note that the extraction uses information only from the real part of the admittance over the range of 0 to 4 GHz, which does not guarantee a good fit to the imaginary part.*

Figure 4-8 contains plots of conductance G_R vs frequency for the biases 0.75, 1.00, 1.20, and 1.40 V. The simulated behavior is qualitatively similar to the measured behavior and predicts the rolloff quite reasonably. However, the simulation underestimates the actual value of the conductance. The reason for the discrepancy is evident if we compare the simulated and actual I-V characteristic in Fig. 4-2. Although the simulated and actual I-V characteristics are close, they do not match exactly, so there is a small offset in the dc value of the conductance. The extracted response achieves a better fit than the simulated response. The excess susceptance B_x vs frequency for these biases appears in Fig. 4-9. As with the conductance, the simulated response has the same qualitative behavior, but is not an exact match.

---

*The real and imaginary parts are linked by the Kramers-Kronig relations. However, since the relations require knowledge of the response over all frequencies (and not just a range of frequencies), it is reassuring to see the good fit.

The extracted behavior of the excess susceptance is well matched to the measured response even though only the conductance was used for the fitting.

Figure 4-10 shows the conductance vs bias for the next set of biases, which are 1.60, 1.65, 1.67, and 1.69 V. In this set of plots, the simulated low-frequency conductance makes the transition from an initially positive value to a negative one. This behavior is sensible since the simulated voltage peak V_p occurs at a bias that is approximately 20 mV lower than the actual peak. In order to improve the fit between simulated and actual behavior for these biases, we would have to fit V_p as well as the peak current. One way in which this could be accomplished is by changing the doping in the 10-nm spacer layer. The extracted parameters again produce a good fit between the theoretical and measured responses.

The excess susceptance for these biases appears in Fig. 4-11. The behavior in this figure is similar to that seen in the previous one. The simulated behavior goes from that of the PDR region to that of the NDR region, which is evident from a change in sign of the excess susceptance. Again, an improved fit of the position of V_p would improve the fit of the calculated behavior. The extracted parameters produce an ac response that is well matched to the actual response.

The final group of biases that will be examined are 1.70, 1.71, 1.72, and 1.73 V. The conductance appears in Fig. 4-12. Unfortunately, as can be seen by the measured and simulated I-V characteristics in Fig. 4-2, there is a large mismatch in the dc conductance. As a consequence, there is also a large mismatch between the values of the calculated and measured responses, and we have not expanded the scale to see the low-frequency calculated response for all of the plots. The behavior of the two is still qualitatively similar, and we note that the rolloff occurs at lower frequencies for higher biases. We also note the conductance is negative at lower frequencies, but turns positive at higher frequencies. The excess susceptance is shown in Fig. 4-13, and the large mismatch between simulated and actual behavior is evident for these plots as well. For both Figs. 4-12 and 4-13, the extracted parameters provide a good fit to the measured response.

The extrema of the excess susceptance provides a indicator of the speed of the response. The extrema are the minima and maxima in the plot of B_x vs frequency for the PDR and NDR regions, respectively. In Fig. 4-14, we see that the measured extremum frequency grows with increasing bias and then sharply cuts off. The calculated extrema behave similarly.

When solving the six conditions listed on page 91 and 98 for each bias, we found values for some of the physical parameters of our models. These values appear as points with error bars in Figs. 4-4 and 4-5. Fig. 4-4(a) and (b) show the extracted values for ν_1 and ν_3, respectively. There is not perfect agreement between the calculated and extracted values, but there is some level of agreement. The extracted values for the quantity $F_1 - F_2$ are shown in Fig. 4-5. The error bars for Figs. 4-4 and 4-5 were crudely determined by trying different values of the parameter and checking the degree of agreement between theory and experiment.

In summary, the ac behavior calculated from the simulation in Fig. 4-2 is not a perfect match to the measured data, but does replicate the qualitative features. The transition from the PDR to NDR regions in the ac behavior is evident from the calculated low frequency conductance. The conductance of the NDR region is calculated to be negative at low frequencies, but turns positive at higher frequencies. Both the calculated and measured extrema of the excess susceptances increase with bias but then cut off sharply.

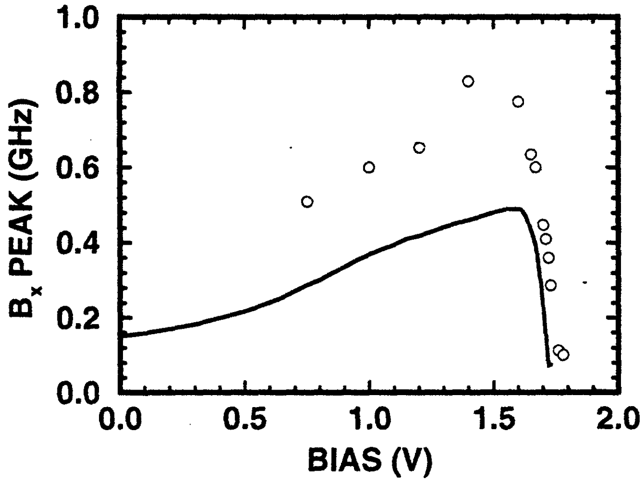

Figure 4-14. Plot of the extrema of the excess susceptance B_x. The extrema are the minima and maxima in the plot of B_x vs frequency for the PDR and NDR regions, respectively. The circles represent the measured values and the solid line represents the calculated values of the extrema.

5

Conclusion

In this final chapter, we summarize the work presented in this document. We will also suggest future avenues for related research.

5.1 Summary

In Chapter 2, we described the manufacture of a device that was specifically designed for ac characterization, which yielded a device with an NDR region that was devoid of switching or oscillatory behavior. The experimental apparatus was able to measure the small-signal admittance behavior from 45 MHz to 26 GHz, and the measurement process was automated to allow for large amounts of averaging. Once the data were obtained, we were able to characterize the parasitic elements and remove their effects from the admittance data. We then examined the intrinsic RTD admittance. In the PDR region, the conductance exhibited a rolloff and remained positive at all frequencies. In the NDR region, the conductance was negative at low frequency and became positve at higher frequencies. The intrinsic susceptance was close to that of a capacitor; the deviation from a capacitive susceptance was negative in the PDR region and positive in the NDR region.

A rate-equation model for the dynamic response of the RTD was presented in Chapter 3. The model is based on the assumption that the rate of electron transfer between regions is bilinear in the number of available states and the number of electrons attempting to make the transition. From this assumption, we were able to calculate the charge in the well and the current across the right and left barriers. The incremental changes in these quantities led to a circuit model for small-signal excitation. In order to calculate the electron transfer rates across the barriers, we derived the impulse response of a quantum-mechanical resonant system. The transfer rate of electrons into the well was then derived from the excitation of the resonant system by an electron in the cathode (or alternatively, the anode). The process of scattering was incorporated semi-classically.

In Chapter 4, we compared the measured data of Chapter 2 to the theoretical model of Chapter 3. To evaluate the model, a computer program was written to self-consistently solve both Poisson's and Schrödinger's equations. We used the program to simulate three structures. The first structure used the anticipated values for the RTD, as a device designer would do before manufacturing a device. This simulation yielded the correct qualitative behavior and a simulated peak current that was a factor of 2.3 times the actual peak current. In the second simulation, we adjusted the barriers so that the simulated and actual peak current were the same. The ac behavior predicted by second simulation was similar to the measured response. The third simulation varied the well width by 0.1 nm, we demonstrated the sensitivity of the solution to the structure parameters.

In addition to the predicted ac response from the second simulation, we also described an algorithm to fit the measured ac response. This algorithm extracted values for the transition rates across the barriers and the quasi-Fermi level in the well plus the derivatives of those three quantities with bias. The extracted values gave good agreement with the measured ac response and were in reasonable accord with the calculated values.

5.2 Future Work

Since only a finite amount of time can be invested in a thesis, several ideas for further investigation had to be laid aside. Also, there are several refinements to the current work that were not pursued. We will mention some of these ideas here.

This work has been concerned with the small-signal dynamic behavior, but the rate-equation approach of Chapter 3 might also be a starting point for the large-signal response. Our preliminary large-signal data [59] demonstrated that the switching time slows down at lower temperature. Hence, one expects that the switching time is dependent on the scattering time.

The scattering time might also be explored by the static behavior as well. Measurements of the I-V vs temperature [59] might well be used to better characterize the scattering time, which was only estimated in Chapter 4.

Now that this work has demonstrated good agreement for the rate-equation of a one-level RTD, a logical extension would be to incorporate higher levels. This extension would have to include interlevel transitions as well as transitions from the well to the cathode and anode. In order to further improve the behavior of the valley current, one would probably need to include inelastic transitions as well.

Finally, the 10-nm spacer layer raised a difficult issue. The continuous doping model is not accurate for extremely small regions which have low doping. Further work that might incorporate stochastic doping models would be needed to improve the approximation used in this work.

A

Bardeen Tunneling Hamiltonian Calculation of the Rates

The transition-rate problem considered in Section 3.4 can be solved using the Bardeen tunneling Hamiltonian [60]. Bardeen's method is useful for calculating the quantum-mechanical current flow between regions that are separated by a significant barrier. The full three-region problem of tunneling through an RTD has been described using Bardeen's method [61, 62]. We present the simpler two-region problem here.

In order to solve the transition-rate problem analytically,* we will make some simplifying assumptions about the potential profile of the RTD. We will use the potential $V(x)$ shown schematically in Fig. A-1(a). Since we are concerned here only with the transition between regions 1 and 2, we have extended the right barrier into region 3. We have ignored the accumulation layer for simplicity, but have retained the electric field in the well and barrier. According to Bardeen's method, we separate this potential into the two potentials, V_1 and V_2, also shown schematically in Fig. A-1(b) and (c), respectively. Using these potentials, we then solve Schrödinger's equation in each region. The current flow is related to the overlap between the wavefunctions

---

*I am grateful to my advisor Prof. Alan McWhorter, who provided me with the original copy of the solution used in this appendix.

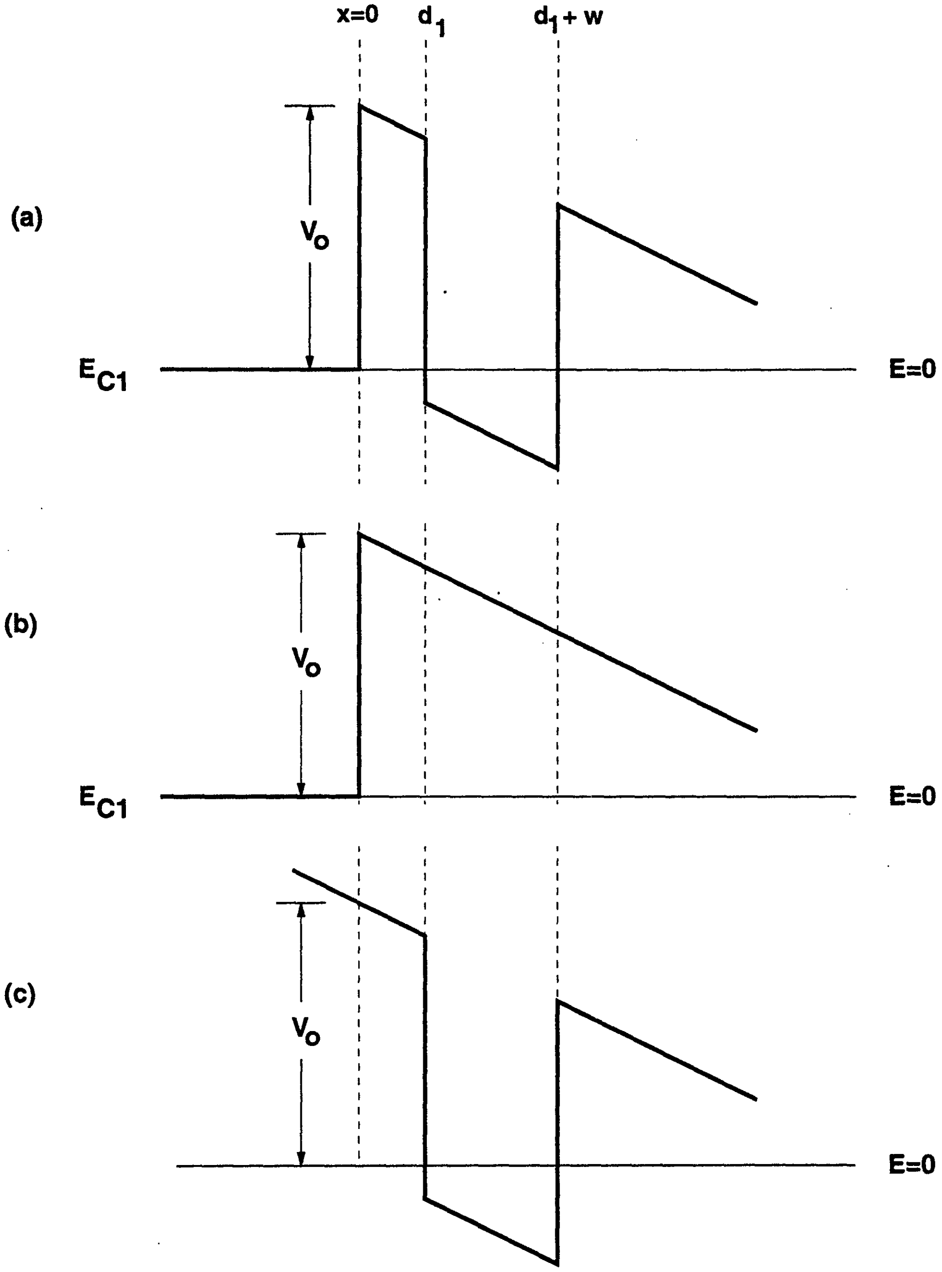

Figure A-1. Schematics of the (a) actual potential profile $V(x)$, and the potential profiles (b) $V_1(x)$ and (c) $V_2(x)$ used to obtain the tunneling Hamilitonian for the two-region problem, as indicated.

from the different regions.

In this appendix, we will first solve for the rate of growth of a state in the well given an electron in some initial state in region 1 and then will calculate an approximate value of the relevant matrix element.

Rate of Growth of the Electron Wavefunction in the Well

The Hamilitonian for the RTD is written as

$$H = \begin{cases} H_1 = -(\hbar^2/2m^*)\nabla^2 + V_1 & x < 0 \\ H_2 = -(\hbar^2/2m^*)\nabla^2 + V_2 & x > 0\,, \end{cases} \tag{A.1}$$

where for simplicity, we have taken m^* to be the same throughout the structure. We note that the sum of the potentials of the two subregions, $V_1 + V_2$, is not equal to the original potential $V(x)$. Assume that only one longitudinal state $\phi_r(x)$ exists in the well. Then the solution to Schrödinger's equation

$$i\hbar\frac{\partial\psi}{\partial t} = H\psi \tag{A.2}$$

can be written in the form

$$\psi = \sum_{\mathbf{k}_1} a_{\mathbf{k}_1}(t)\phi_{k_{t1}}(x)e^{i\mathbf{k}_{t1}\cdot\mathbf{r}-iE_{\mathbf{k}_1}t/\hbar} + \sum_{\mathbf{k}_{t2}} b_{\mathbf{k}_{t2}}(t)\phi_r(x)e^{i\mathbf{k}_{t2}\cdot\mathbf{r}-i(E_r+E_{\mathbf{k}_{t2}})t/\hbar}\,, \tag{A.3}$$

where $\phi_{k_{\ell 1}}(x)e^{i\mathbf{k}_{t1}\cdot\mathbf{r}}$ is a solution to H_1 and $\phi_r(x)e^{i\mathbf{k}_{t2}\cdot\mathbf{r}}$ to H_2. After substituting Eq. (A.3) into Schrödinger's equation, we find

$$\begin{aligned}
&\sum_{\mathbf{k}_1}(i\hbar\dot{a}_{\mathbf{k}_1} + E_{\mathbf{k}_1}a_{\mathbf{k}_1})\phi_{k_{\ell 1}}(x)e^{i\mathbf{k}_{t1}\cdot\mathbf{r}-iE_{\mathbf{k}_1}t/\hbar} \\
&\qquad + \sum_{\mathbf{k}_{t2}}\left[i\hbar\dot{b}_{\mathbf{k}_{t2}} + (E_r + E_{\mathbf{k}_{t2}})b_{\mathbf{k}_{t2}}\right]\phi_r(x)e^{i\mathbf{k}_{t2}\cdot\mathbf{r}-i(E_r+E_{\mathbf{k}_{t2}})t/\hbar} \\
&\quad = \sum_{\mathbf{k}_1} a_{\mathbf{k}_1}e^{-iE_{\mathbf{k}_1}t/\hbar}H\phi_{k_{\ell 1}}e^{i\mathbf{k}_{t1}\cdot\mathbf{r}} + \sum_{\mathbf{k}_{t2}} b_{\mathbf{k}_{t2}}e^{-i(E_r+E_{\mathbf{k}_{t2}})t/\hbar}H\phi_r(x)e^{i\mathbf{k}_{t2}\cdot\mathbf{r}} \,. \qquad \text{(A.4)}
\end{aligned}$$

In the above equations, $E_{\mathbf{k}_1} = E_{k_{\ell 1}} + E_{\mathbf{k}_{t1}}$. Multiplying by $\phi_r^* e^{-i\mathbf{k}_{t2}'\cdot\mathbf{r}}$ and integrating, then dropping the prime, yields

$$\begin{aligned}
i\hbar\dot{b}_{\mathbf{k}_{t2}} &= \sum_{\mathbf{k}_1} a_{\mathbf{k}_1}\langle r|H_\ell|k_{\ell 1}\rangle\,\delta_{\mathbf{k}_{t1},\mathbf{k}_{t2}}e^{i(E_r-E_{k_{\ell 1}})t/\hbar} + b_{\mathbf{k}_{t2}}\langle r|H_\ell|r\rangle \\
&\qquad - \sum_{\mathbf{k}_1} a_{\mathbf{k}_1}E_{k_{\ell 1}}\langle r|k_{\ell 1}\rangle\,\delta_{\mathbf{k}_{t1},\mathbf{k}_{t2}}e^{i(E_r-E_{k_{\ell 1}})t/\hbar} - b_{\mathbf{k}_{t2}}E_r\,, \qquad \text{(A.5)}
\end{aligned}$$

where we have neglected the higher coupling terms with $\dot{a}_{\mathbf{k}_1}$ and have defined

$$H_\ell = \begin{cases} H_{\ell 1} \equiv -\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial x^2} + V_1 & x < 0 \\ H_{\ell 2} \equiv -\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial x^2} + V_2 & x > 0\,. \end{cases} \qquad \text{(A.6)}$$

Assume at $t = 0$ that an electron is in state $\mathbf{k}_{10}$, so that $a_{\mathbf{k}_1}(t = 0) = \delta_{\mathbf{k}_1,\mathbf{k}_{10}}$ and $b_{\mathbf{k}_{t2}}(t = 0) = 0$. Let $k_{\ell 0}$ and $\mathbf{k}_{t0}$ be the longitudinal and transverse components of $\mathbf{k}_{10}$, respectively. Since all $a_{\mathbf{k}_1}$ and $b_{\mathbf{k}_{t2}}$ for which $\mathbf{k}_t \neq \mathbf{k}_{t0}$ will remain zero, we can restrict ourselves to $\mathbf{k}_t = \mathbf{k}_{t0}$ and drop the subscripts 1 and 2, which are no longer needed. Also, since $\langle r|H_\ell|r\rangle \approx \langle r|H_{\ell 2}|r\rangle = E_r$ to first order in the tunneling coefficient, we obtain to first order

$$i\hbar\dot{b}_{t0} = \langle r|H_\ell - H_{\ell 1}|k_{\ell 0}\rangle\, e^{i(E_r-E_{k_{\ell 0}})t/\hbar} \qquad t > 0 \qquad \text{(A.7)}$$

Approximate Solution of the Matrix Element

We define $M_{rk_{\ell 0}} \equiv \langle r|H_\ell - H_{\ell 1}|k_{\ell 0}\rangle$. To evaluate $M_{rk_{\ell 0}}$ we observe that

$$H_\ell - H_{\ell 1} = \begin{cases} -V_0 & \text{between the barriers} \\ 0 & \text{elsewhere} \end{cases} \tag{A.8}$$

If the cathode is of width L_R with infinite barrier at $x = -L_R$, then in the quasi-classical approximation

$$\phi_{k_{\ell 0}}(x) = \begin{cases} A\sin\left[k_{\ell 0}(L_R + x)\right] & -L_R < x < 0 \\ B\sqrt{\frac{\alpha_{\ell 0}(0)}{\alpha_{\ell 0}(x)}}\exp\left[-\int\limits_0^x \alpha_{\ell 0}(x')dx'\right] & x > 0\,, \end{cases} \tag{A.9}$$

where

$$\alpha_{k_{\ell 0}}(x) = \left[2m^*(V_0 - Fx - E_{k_{\ell 0}})/\hbar^2\right]^{1/2} \tag{A.10}$$

$$\int\limits_0^x \alpha_{\ell 0}(x')dx' = \frac{2}{3}\frac{\sqrt{2m*}}{\hbar F}\left[(V_o - E_{k_{\ell 0}})^{3/2} - (V_0 - Fx - E_{k_{\ell 0}})^{3/2}\right] \tag{A.11}$$

Continuity of ψ and $\partial\psi/\partial x$ at $x = 0$ requires

$$B = A\sin k_{\ell 0}L_R \tag{A.12}$$

$$\tan k_{\ell 0}L_R \approx -\frac{k_{\ell 0}L_R}{\alpha_{\ell 0}(0)L_R} \tag{A.13}$$

neglecting $d\alpha_{\ell 0}/dx|_0$. Here $k_{\ell 0}L_R$ is in either the second or fourth quadrant. The density of allowed values of $k_{\ell 0}$ is therefore L_R/π, and for large L_R the normalization of $\phi_{k_{\ell 0}}$ yields

$$A \approx \sqrt{\frac{2}{L_R}} \tag{A.14}$$

$$\begin{aligned} |B| &\approx \sqrt{\frac{2}{L_R}} |\sin k_{\ell 0} L_R| = \sqrt{\frac{2}{L_R}} \frac{|\tan k_{\ell 0} L_R|}{(1+\tan^2 k_{\ell 0} L_R)^{1/2}} \\ &= \sqrt{\frac{2}{L_R}} \frac{k_{\ell 0}}{\left[{k_{\ell 0}}^2 + \alpha_{\ell 0}^2(0)\right]^{1/2}} = \sqrt{\frac{2}{L_R}} \frac{k_{\ell 0}}{(2m^* V_0/\hbar^2)^{1/2}} \, . \end{aligned} \tag{A.15}$$

For the quasi-classical approximation to be valid we must have

$$\left| \frac{d(k^{-1})}{dx} \right| \ll 1 \tag{A.16}$$

or here

$$\frac{m^* F'}{\hbar^2 |k|^3} \ll 1 \tag{A.17}$$

for k real or imaginary. The condition seems to be adequately met. .

For ϕ_r we have in the quasi-classical approximation

$$\phi_r = \begin{cases} C\sqrt{\frac{\hbar^2 k_r(d_1)\alpha_r(d_1)}{2m^* V_0 \alpha_r(x)}} \, \exp\left[-\int\limits_x^{d_1} \alpha_r(x')dx' \right] & x < d_1 \\ \frac{C}{\sqrt{k_r(x)}} \sin\left[\int\limits_{d_1}^{x} k_r(x')dx' + \vartheta_1 \right] & d_1 < x < d_1 + w \\ C\sqrt{\frac{\hbar^2 k_r(d_1+w)\alpha_r(d_1+w)}{2m^* V_0 \alpha_r(x)}} \, \exp\left[-\int\limits_{d_1+w}^{x} \alpha_r(x')dx' \right] & x > d_1 + w \, , \end{cases} \tag{A.18}$$

where

$$k_r(x) = \left[2m^*(E_r + Fx)/\hbar^2\right]^{1/2} \tag{A.19}$$

$$\alpha_r(x) = \left[2m^*(V_0 - E_r - Fx)/\hbar^2\right]^{1/2} \tag{A.20}$$

$$\vartheta_1 = \tan^{-1}\left[\frac{k_r(d_1)}{\alpha_r(d_1)}\right] . \tag{A.21}$$

The allowed values of E_r are given by the condition that

$$\int_{d_1}^{d_1+w} k_r(x)dx + \vartheta_1 + \vartheta_2 =$$

$$\frac{2}{3}\frac{\sqrt{2m^*}}{\hbar F}\left\{[E + F(d_1 + w)]^{3/2} - [E + Fd_1]^{3/2}\right\} + \vartheta_1 + \vartheta_2 = \pi , \tag{A.22}$$

where

$$\vartheta_2 = \tan^{-1}\left[\frac{k_r(d_1 + w)}{\alpha_r(d_1 + w)}\right] . \tag{A.23}$$

Note that if $d\alpha/dx|_{d_1}$ and $d\alpha/dx|_{d_1+w}$ are neglected, we obtain

$$\phi_r(d_1) = \frac{C\sin\vartheta_1}{\sqrt{k_r(d_1)}} = \frac{C\sqrt{k_r(d_1)}}{(2m^*V_0/\hbar^2)^{1/2}} \tag{A.24}$$

$$\phi_r(d_1 + w) = \frac{C\sin\vartheta_2}{\sqrt{k_r(d_1 + w)}} = \frac{C\sqrt{k_r(d_1 + w)}}{(2m^*V_0/\hbar^2)^{1/2}} \tag{A.25}$$

and

$$\left.\frac{d\phi_r}{dx}\right|_{x=d_1} \approx C\sqrt{k_r(d_1)}\;\cos\vartheta_1 = C\sqrt{k_r(d_1)}\;\frac{\alpha_r(d_1)}{(2m^*V_0/\hbar^2)^{1/2}} \tag{A.26}$$

$$\left.\frac{d\phi_r}{dx}\right|_{x=d_1+w} \approx -C\sqrt{k_r(d_1+w)}\;\cos\vartheta_2 = -C\sqrt{k_r(d_1+w)}\;\frac{\alpha_r(d_1+w)}{(2m^*V_0/\hbar^2)^{1/2}} \tag{A.27}$$

so that ψ is continuous and $d\psi/dx$ is approximately continuous at $x = d_1$ and $x = d_1 + w$.

Now using Eqs. (A.8), (A.9), (A.15), and (A.18)-(A.20), we have

$$\begin{aligned} |M_{rk_{\ell0}}| &= V_0 \int_{d_1}^{d_1+w} \phi_r^*(x)\phi_{k_{\ell0}}(x)dx \\ &= V_0 k_{\ell0} C^* \sqrt{\frac{2}{L_R}}\sqrt{\frac{\hbar^2\alpha_{\ell0}(0)}{2m^*V_0}} \\ &\times \int_{d_1}^{d_1+w} \frac{1}{\sqrt{k_r(x)\alpha_{\ell0}(x)}} \sin\left[\int_{d_1}^{x} k_r(x')dx' + \vartheta_1\right] \exp\left[-\int_0^x \alpha_{\ell0}(x')dx'\right] dx\,. \end{aligned} \tag{A.28}$$

Since the exponential cuts off fairly rapidly, we crudely approximate the integral by setting $k_r(x) = k_r(d_1)$ and $\alpha_{\ell0}(x) = \alpha_{\ell0}(d_1)$ for $d_1 < x < d_1 + w$ and letting w become infinite, giving

$$\begin{aligned} |M_{rk_{\ell0}}| &\approx V_0 k_{\ell0}\sqrt{\frac{2}{L_R}}\sqrt{\frac{\hbar^2\alpha_{\ell0}(0)}{2m^*V_0}}\frac{C^*\exp\left[-\int_0^{d_1}\alpha_{\ell0}(x')dx'\right]}{\sqrt{k_r(d_1)\alpha_r(d_1)}} \\ &\times \int_{d_1}^{\infty} \sin\left[k_r(d_1)(x-d_1)+\vartheta_1\right] e^{-\alpha_{\ell0}(d_1)(x-d_1)}dx \end{aligned} \tag{A.29}$$

The latter integral may be evaluated to yield

$$\int_{d_1}^{\infty} \sin\left[k_r(d_1)(x-d_1)+\vartheta_1\right] e^{-\alpha_{\ell 0}(d_1)(x-d_1)}dx = \frac{\alpha_{\ell 0}(d_1)\sin\vartheta_1 + k_r(d_1)\cos\vartheta_1}{\alpha_{\ell 0}^2(d_1)+k_r^2(d_1)} \tag{A.30}$$

We take C to be the value it would have for $k_r(x) = k_r(d_1)$ and $\alpha_r(x) = \alpha_r(d_1)$, namely

$$C = \left[\frac{k_r(d_1)}{\alpha_r^{-1}(d_1)+w/2}\right]^{1/2} \tag{A.31}$$

Since $E_{k_{\ell 0}} \approx E_r$ for the elastic transition, we set $\alpha_{\ell 0} = \alpha_r$ to obtain

$$\begin{aligned} |M_{rk_{\ell 0}}| &\approx V_0\sqrt{\frac{2}{L_R}}\,\frac{2k_r(d_1)k_{\ell 0}\sqrt{\alpha_{\ell 0}(0)\alpha_r(d_1)}}{(2m^*V_0/\hbar^2)^2}\,\frac{\exp\left[-\int_0^{d_1}\alpha_r(x)dx\right]}{\left[\alpha_r^{-1}(d_1)+w/2\right]^{1/2}} \\ &= \frac{4}{L_R\left[w+2\alpha_r^{-1}(d_1)\right]^{1/2}}\,\frac{E_{k_{\ell 0}}^{1/2}(E_r+Fd_1)^{1/2}(V_0-E_r)^{1/4}(V_0-E_r-Fd_1)^{1/4}}{(2m^*/\hbar^2)^{1/2}V_0} \\ &\quad \times \exp\left[-\int_0^{d_1}\alpha_r(x)dx\right] \end{aligned} \tag{A.32}$$

We note that this expression for the matrix element goes to zero when the resonant energy goes to zero, i.e., when the resonant energy goes to the energy of the conduction band edge in region 1.

B

Integrating and Differentiating the Fermi Factors

In this appendix, we give some details about Fermi occupation function that are omitted from the main part of the text. We will use the following notation: the Fermi factor f_ψ as a function of the potentials will be written as

$$f_\psi = \frac{1}{1 + e^{q(\phi-\psi)/kT}}, \tag{B.1}$$

where ϕ and ψ are the quasi-Fermi and electron potentials, respectively. The Fermi factor f_E as a function of the quasi-Fermi and electron energies is

$$f_E = \frac{1}{1 + e^{(E-F)/kT}}, \tag{B.2}$$

where E is the energy, and F is the quasi-Fermi energy.

By differentiating f_ψ, we find

$$\begin{aligned} \frac{\partial f_\psi}{\partial \phi} &= -\frac{\partial f_\psi}{\partial \psi} \\ &= -\frac{q}{kT}\frac{e^{q(\phi-\psi)/kT}}{(1 + e^{q(\phi-\psi)/kT})^2} \\ &= -\frac{q}{kT} f_\psi (1 - f_\psi) . \end{aligned} \tag{B.3}$$

The incremental change in the Fermi occupation factor δf_ψ, which was used for linearizing the current in Section 3.3, is then given by

$$\delta f_\psi = \delta\psi \frac{\partial f_\psi}{\partial \psi} + \delta\phi \frac{\partial f_\psi}{\partial \phi}$$

$$= \frac{q}{kT} f_\psi(1 - f_\psi)(\delta\psi - \delta\phi) \tag{B.4}$$

where $\delta\psi$ and $\delta\phi$ are the incremental changes in the electron and quasi-Fermi potentials, respectively.

Integration of the Fermi Function

The result in Eq. (B.3) makes integration of the Fermi function an easy matter. Since the results in Chapter 3 require the integral of the Fermi function over energy, we recast Eq. (B.3) as

$$\frac{\partial f_E}{\partial E} = -\frac{1}{kT} f_E(1 - f_E)\,. \tag{B.5}$$

We can then perform the integration of f_E as follows:

$$\int f_E dE = kT \int \frac{f_E(1 - f_E)}{kT(1 - f_E)} dE = -kT \int \frac{1}{(1 - f_E)} df_E = kT \ln(1 - f_E)\,. \tag{B.6}$$

C

Modified Y-Δ transformation

In order to calculate the admittance, we transform the topology shown in Fig. C-1(a) to that of Fig. C-1(b). The latter topology simplifies the algebra associated with deriving the linear response. If the current sources were not present in Fig. C-1(a), this would be a Y-Δ (also called a T-π) transformation. However, since current sources do exist, we need to find a modified version of the Y-Δ transformation.

The circuit transformation is found by demanding that the terminal currents for the transformed circuit be equal to those of the original circuit when the same potential is applied to the terminals. It can be proven after some algebra that this

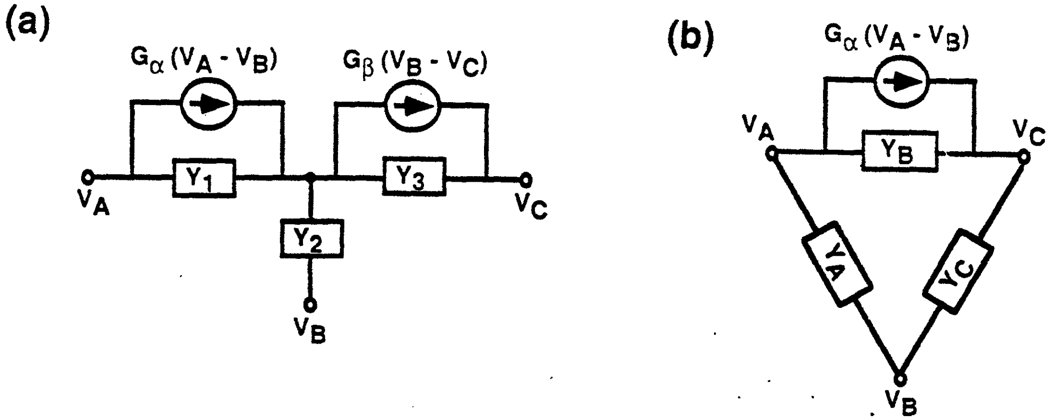

Figure C-1. Modified Y-Δ transformation.

condition is met if we make the following equivalences:

$$\begin{aligned} Y_A &= \frac{Y_2(Y_1+G_\alpha)}{Y_1+Y_2+Y_3} && \text{(C.1)} \\ Y_B &= \frac{Y_1(Y_3+G_\beta)}{Y_1+Y_2+Y_3} && \text{(C.2)} \\ Y_C &= \frac{Y_2(Y_3+G_\beta)}{Y_1+Y_2+Y_3} && \text{(C.3)} \\ G_B &= \frac{(Y_3G_\alpha - Y_1G_\beta)}{Y_1+Y_2+Y_3}. && \text{(C.4)} \end{aligned}$$

Finally, we note that the topology chosen in Fig. C-1(b) is not unique. In particular, the current source can be placed across Y_A or Y_C, but its value will then be different.

Bibliography

[1] See, for example, H. C. Liu and T. C. L. G. Sollner, "High-frequency resonant tunneling diodes," and F. Capasso, F. Beltram, S. Sen, A. Palevski, and A. Y. Cho, "Quantum Electron Devices: Physics and Applications," both in *High Speed Heterostructure Devices* (R. A. Kiehl and T. C. L. G. Sollner, eds.), New York: Academic, 1994, and references therein.

[2] See, for example, R. J. Aggarwal, Ph.D. thesis, Massachusetts Institute of Technology, 1996, and references therein.

[3] E. R. Brown, J. R. Söderström, C. D. Parker, L. J. Mahoney, K. M. Molvar, and T. C. McGill, "Oscillations up to 712 GHz in InAs/AlSb resonant-tunneling diodes," *Appl. Phys. Lett.* **58**, p. 2291, 1991.

[4] L. Yang, S. D. Draving, D. E. Mars, and M. R. Tan, "A 50 GHz broad-band monolithic GaAs/AlAs resonant tunneling diode trigger circuit," *IEEE J. Solid-State Circuits* **29**, p. 585, 1994.

[5] K. L. Lear, K. Yoh, and J. S. Harris, "Monolithic integration of GaAs/AlAs resonant tunnel diode load and GaAs enhancement mode MESFET drivers for tunnel diode FET logic gates," *Proceedings of the 1988 GaAs and Related Compounds Conference*, 1988.

[6] Y. Watanabe, Y. Nakasha, K. Imanishi, and M. Takikawa, "Monolithic integration of InGaAs/InAlAs resonant tunneling diode and HEMT for single-

transistor SRAM application," *Proceedings of the 1992 International Electron Devices Meeting*, 1992.

[7] M.-H. Shieh and H. C. Lin, "A multiple-dimensional multiple-state SRAM cell using resonant tunnling diodes," *IEEE J. Solid-State Circuits* **29**, p. 623, 1994.

[8] S. J. Wei and H. C. Lin, "Multivalued SRAM using resonant tunnling diodes," *IEEE J. Solid-State Circuits* **27**, p. 212, 1992.

[9] Z. X. Yan and M. J. Deen, "A new resonant-tunnel diode-based multivalued memory circuit using a MESFET depletion load," *IEEE J. Solid-State Circuits* **27**, p. 1198, 1992.

[10] R. Tsu and L. Esaki, "Tunneling in a finite superlattice," *Appl. Phys. Lett.* **22**, p. 562, 1973.

[11] B. Ricco and M. Y. Azbel, "Physics of resonant tunneling. The one-dimensional double-barrier case.," *Phys. Rev. B* **29**, p. 1970, 1984.

[12] E. Gerjuoy and D. D. Coon, "Analytic S-matrix considerations and the time delay in resonant tunneling.," *Superlat. Microstruct.* **5**, p. 305, 1989.

[13] M. Cahay, M. McLennan, S. Datta, and M. S. Lundstrom, "Importance of space-charge effects in resonant tunneling devices," *Appl. Phys. Lett.* **50**, p. 612, 1987.

[14] H. Ohnishi, T. Inata, S. Muto, N. Yokoyama, and A. Shibatomi, "Self-consistent analysis of resonant tunneling current," *Appl. Phys. Lett.* **49**, p. 1248, 1986.

[15] J. S. Wu, K. H. Chang, C. P. Lee, C. Y. Chang, D. G. Liu, and D. C. Liou, "Quantum effect in the accumulation layer on field-induced photoluminescence of double-barrier resonant-tunneling structures," *Appl. Phys. Lett.* **59**, p. 87, 1991.

[16] Y. G. Gabato, J. M. Berroir, Y. Guldner, J. P. Vieren, F. Chevoir, and B. Vinter, "Selection-rule breakdown in coherent resonant tunneling in a tilted magnetic field," *Phys. Rev. B* **44**, p. 13795, 1991.

[17] A. D. Stone and P. A. Lee, "Effect of inelastic processes on resonant tunneling in one dimension," *Phys. Rev. Lett.* **54**, p. 1196, 1985.

[18] M. Jonson and A. Grincwajg, "Effect of inelastic scattering on resonant and sequential tunneling in double barrier heterostructures.," *Appl. Phys. Lett.* **51**, p. 1729, 1987.

[19] J. Leo and A. H. MacDonald, "Disorder-assisted tunneling through a double-barrier structure," *Phys. Rev. Lett.* **64**, p. 817, 1990.

[20] K. V. Rousseau, K. L. Wang, and J. N. Schulman, "Γ- and X-state influences on resonant tunneling current in single- and double-barrier GaAs/AlAs structures," *Appl. Phys. Lett.* **54**, p. 1341, 1989.

[21] N. S. Wingreen, K. W. Jacobson, and J. W. Wilkins, "Inelastic scattering in resonant tunneling," *Phys. Rev. B* **40**, p. 11834, 1989.

[22] J. F. Whitaker, G. A. Mourou, T. C. L. G. Sollner, and W. D. Goodhue, "Picosecond switching time measurement of a resonant tunneling diode," *Appl. Phys. Lett.* **53**, p. 385, 1988.

[23] D. Lippens and P. Mounaix, "Small-signal impedance of GaAs-$Al_xGa_{1-x}As$ resonant tunnelling heterostructures at microwave frequency," *Electron. Lett.* **24**, p. 1180, 1988.

[24] J. M. Owens, D. J. Halchin, K. L. Lear, W. S. Lee, and J. S. H. Jr., "Microwave characteristics of MBE grown resonant tunneling devices," *IEEE MTT Digest*, 1989.

[25] S. P. Stapleton, M. J. Deen, E. Berolo, and R. H. S. Hardy, "Experimental study of microwave reflection gain of AlAs/GaAs/AlAs quantum well structures," *Electron. Lett.* **26**, p. 84, 1990.

[26] P. Mounaix, P. Bedu, D. Lippens, and E. Barbier, "Measurement of negative differential conductance to 40 GHz for vertically integrated resonant tunnelling diodes," *Electron. Lett.* **27**, p. 1359, 1991.

[27] M. F. C. Schemmann, H. C. Heyker, J. J. M. Kwaspen, and T. G. van de Roer, "Microwave impedence measurements on resonant tunnelling diodes and network model parameter extraction," *IEE Proc.-H* **138**, p. 248, 1991.

[28] T. Wei, S. Stapleton, and E. Berolo, "Equivalent circuit and capacitance of double barrier resonant tunneling diode," *J. Appl. Phys.* **73**, p. 829, 1993.

[29] J. Genoe, C. V. Hoof, W. V. Roy, J. H. Smet, K. Fobelets, R. P. Mertens, and G. Borghs, "Capacitances in double-barrier tunneling structures," *IEEE Trans. Electron Devices* **38**, p. 2006, 1991.

[30] J. Jo, H. S. Li, Y. W. Chen, and K. L. Wang, "Observation of a large capacitive current in a double barrier resonant tunneling diode at resonance," *Appl. Phys. Lett.* **64**, p. 2276, 1994.

[31] R. E. Miles, G. Millington, R. D. Pollard, D. P. Steenson, J. M. Chamberlain, and M. Henini, "Accurate equivalent circuit model of resonant tunneling diodes," *Electron. Lett.* **27**, p. 427, 1991.

[32] J. P. Mattia, S. M. thesis, Massachusetts Institute of Technology, 1992.

[33] J. S. Scott, J. P. Kaminski, M. Wanke, and S. J. Allen, "Terahertz frequency response of an $In_{0.53}Ga_{0.47}As$/AlAs resonant-tunneling diode," *Appl. Phys. Lett.* **64**, p. 1995, 1994.

[34] J. P. Mattia, E. R. Brown, A. R. Calawa, and M. J. Manfra, "Small-signal admittance and switching measurements of the resonant-tunneling diode," *Appl. Phys. Lett.* **63**, p. 521, 1994.

[35] J. M. Gering, D. A. Crim, D. G. Morgan, P. D. Coleman, W. Kopp, and H. Morkoc, "A small-signal equivalent-circuit model for GaAs-$\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ resonant tunneling heterostructures at microwave frequencies," *J. Appl. Phys.* **61**, p. 271, 1987.

[36] D. Lippens and P. Mounaix, "Small-signal impedance of GaAs-$\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ resonant tunnelling heterostructures at microwave frequency," *Electron. Lett.* **24**, p. 1180, 1988.

[37] R. E. Miles, G. Millington, R. D. Pollard, D. P. Steenson, J. M. Chamberlain, and M. Henini, "Accurate equivalent circuit model of resonant tunneling diodes," *Electron. Lett.* **27**, p. 427, 1991.

[38] J. Genoe, C. V. Hoof, W. V. Roy, J. H. Smet, K. Fobelets, R. P. Mertens, and G. Borghs, "Capacitances in double-barrier tunneling structures," *IEEE Trans. Electron Devices* **38**, p. 2006, 1991.

[39] W. R. Frensley, "Quantum transport calculation of the small-signal response of a resonant tunneling diode," *Appl. Phys. Lett.* **51**, p. 448, 1987.

[40] E. R. Brown, C. D. Parker, and T. C. L. G. Sollner, "Effect of quasibound-state lifetime on the oscillation power of resonant tunneling diodes," *Appl. Phys. Lett.* **54**, p. 934, 1989.

[41] Y. Hu and S. P. Stapleton, "Capacitance of a resonant tunneling diode," *Japan. J. Appl. Phys.* **31**, p. 23, 1992.

[42] Y. Hu and S. P. Stapleton, "Double-barrier resonant tunneling transport model," *IEEE J. Quantum Electron.* **29**, p. 327, 1993.

[43] Y. Fu and S. C. Dudley, "Quantum inductance with linear response theory," *Phys. Rev. Lett.* **70**, p. 65, 1992.

[44] See, for example, W. Jones and N. H. March, *Theoretical Solid State Physics.* New York: Wiley, 1973.

[45] R. Landauer, "Electron transport in open and closed systems," *Z. Phys. B* **68**, p. 217, 1987, and M. Büttiker, "Four terminal phase-coherent conductance," *Phys. Rev. Lett.* **57**, p. 1761, 1986.

[46] A.-P. Jauho, N. S. Wingreen, and Y. Meir, "Time-dependent transport in interacting and noninteracting resonant-tunneling systems," *Phys. Rev. B* **50**, p. 5528, 1994.

[47] A.-P. Jauho, N. S. Wingreen, and Y. Meir, "Time-dependent transport in mesoscopic systems: general formalism and applications," *Semicond. Sci. Technol.* **9**, p. 926, 1994.

[48] Paul Maki, private communication.

[49] See, for example, S. Sze, *Physics of Semiconductor Devices.* New York: Wiley, 1981.

[50] S. Luryi, "Frequency limit of double-barrier resonant-tunneling oscillators," *Appl. Phys. Lett.* **47**, p. 490, 1985.

[51] W. Shockley and W. T. Read, "Statistics of the recombination of holes and electrons," *Phys. Rev.* **87**, p. 835, 1952.

[52] R. N. Hall, "Electron-hole recombination in germanium," *Phys. Rev.* **87**, p. 387, 1952.

[53] The work in this section benefited from several conversations with my colleague Farhan Rana, to whom I am grateful.

[54] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics, Non-Relativistic Theory.* Addison-Wesley, 1958.

[55] E. O. Kane, "Basic Concepts of Tunneling," in *Tunneling in Solids* (E. Burstein and D. Lundqvist, eds.) New York: Plenum, 1969.

[56] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C.* London: Cambridge University Press, 1988.

[57] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures.* New York: Wiley, 1988.

[58] C. B. Duke, "Tunneling in Solids," in *Solid State Physics* (F. Seitz, D. Turnbull, and H. Ehrenreich, eds.) Vol. 10, New York: Academic, 1969.

[59] J. P. Mattia, to be published.

[60] J. Bardeen, "Tunneling from a many-particle point of view," *Phys. Rev. Lett.* **61**, p. 57, 1961.

[61] H. C. Liu and G. C. Aers, "Resonant tunneling through one-, two-, and three-dimensionally confined quantum wells," *J. Appl. Phys.* **65**, p. 4908, 1989.

[62] S. P. Apell, "Photoinduced resonant tunneling treated by an extended transfer Hamiltonian method," *Phys. Rev. B* **45**, p. 6757, 1992.