Indirect Electron Tunneling

Michael J. Enz

School of Physics and Astronomy
University of Minnesota
Minneapolis, MN 55455

Indirect electron tunneling is observed in germanium diodes by detecting small perturbations in their I-V characteristics at low temperatures that occur when the applied potential equals the energy of emitted phonons. Frequencies of the observed phonons are calculated to a comparable degree of accuracy ($\pm 0.3E12$ Hz) with neutron scattering data. Electronic differentiation, used to closely inspect changes in the diodes’ I-V curves, is compared with numerical differentiation using both a cubic spline method and a least squares fit to a high order polynomial.

Introduction

Conduction occurs in semiconductors when electrons are excited across an energy gap, from the valence band to the conduction band. However, in certain silicon and germanium tunneling diodes, the minimum energy in the conduction band does not occur at the same point in k-space (momentum-space) as the maximum in the valence band [1]. Therefore, electrons indirectly tunnel by emitting a phonon to conserve momentum. This results in increased current through the diode when the applied voltage exactly equals the energy of the emitted phonon [2].

The I-V characteristics for three germanium diodes are measured at various temperatures and electronically differentiated to closely inspect perturbations due to phonon emission. At a temperature of 4.2 K, the second derivatives have distinct peaks centered at the energies of emitted phonons, allowing identification of the particles using neutron scattering data from Brockhouse [3]. Numerical differentiation is also explored as an alternative to electronic differentiation using both a cubic spline interpolation method and a least squares fit to a high order polynomial.

Theory

In most semiconductor devices, current flows when electrons are excited across a forbidden energy gap, from the valence band to the conduction band. When no voltage is applied across a p-n junction, electrons fill up to the Fermi energy level. Tunneling diodes are heavily doped p-n junctions that have a Fermi energy level in the valence band on the P side.
of the junction and the conduction band on the N side [4,5]. Current flows when electrons "tunnel" across the junction, with only a small change in energy from photon emission [1]. When no voltage is applied, tunneling occurs with the same probability in either direction so no net current flows through the device. When the p-n junction is weakly biased, tunneling favors one direction and a net current is created.

In silicon and germanium, electrons change energy bands indirectly since the minimum distance between the valence and conduction bands in momentum-space does not correspond to strictly a change in energy. Electrons must tunnel by a small change in energy from a photon emission, accompanied by a change of momentum [1]. Conservation of momentum requires the emission of a phonon, which is a quantum of acoustic radiation with energy $h\nu$, where $h$ is Plank’s constant and $\nu$ is its frequency. When the energy applied to the tunneling diode (eV) is exactly equal to the discrete energy of a phonon, there is an increase in the current through the diode, corresponding to a "kink" in the I-V curve [6]. However, this is only noticeable at low temperatures, since electrons have a thermal energy of $k_B T$, or 25.8 meV, at room temperature, which is sufficient energy for emission of phonons of various frequencies.

Experimental Methods

Figure 1: Diagram of summing amplifier circuit used for measuring and electronically differentiating the I-V characteristic of a tunnel diode. The circuit applies a 0-60 mV potential across the diode, with a 1 mV AC signal added from the lock-in amplifier’s reference to allow differentiation.
A summing, inverting amplifier, which adds DC and AC signals with separate gains, is used to ramp the potential difference across the weakly biased range of the diode while a superimposed AC signal remains constant (fig. 1). A programmable HP E3631A is used for the DC voltage source, with a 0-6.1V full operating range and a resolution of 3mV. This provides a 0-60 mV DC voltage range across the diode. Note that the diode is forward biased, accommodating a negative output from the op-amp ($V_{\text{out}}$ in fig. 1). The voltage across the diode ($V_d$) is sampled with a Philips PM2525 multimeter in DC mode, assuming the input to the pre-amp is grounded. In fact, the Stanford Research 570 current pre-amp has a specified maximum input impedance of 1$\Omega$ when operated in the lowest gain mode. Therefore, the assumption introduces an error in the voltage across the diode less than the current through the diode in magnitude, which is a non-trivial task to correct since one cannot add a resistor in series with the diode while preserving the constant amplitude of the AC signal. The pre-amp converts the current to a voltage, allowing measurement with another Philips PM2525 multimeter in DC mode.

Electronic differentiation is done with a lock-in amplifier, using its reference sine wave (1kHz) as the change in voltage and measuring the corresponding change in current at that frequency. A steady 1mV AC signal across the diode is the smallest reference at which accurate measurements could be made. The pre-amp’s output is sent into a Stanford Research SR830 lock-in amplifier which multiplies the signal by the original reference, and measures the resulting voltage. It does the same procedure 90 degrees out of phase, which allows calculation of a phase independent, signal amplitude for only the reference frequency that is directly proportional to the first derivative of the I-V characteristic. While less intuitive, by multiplying the signal with the second harmonic frequency of the reference signal the lock-in amplifier can sample a voltage proportional to the second derivative.

The entire op-amp circuit is enclosed in an aluminum box to shield against noise. The box contains 5 BNC connectors for DC input, AC input, +15V, -15V, and $V_{\text{out}}$ (as specified in fig. 1). The diode is enclosed in a separate aluminum box during room temperature data runs. It is dipped in a glass-lined dewer during liquid nitrogen data runs. For liquid helium data acquisition, the diode is placed at one end of a hollow, 5/16" diameter by 1.4m long, stainless steel tube, which is dipped into a storage container. The tube is stuffed with cotton at four places to avoid oscillations in the liquid level in the tube.
The entire experiment is controlled with a C program in the LabWindows environment on a Pentium based computer running the Windows NT 4.0 operating system. Using a GPIB bus, the computer controls both multimeters, the DC power supply, and the lock-in amplifier. The data collection process consists of repeatedly setting the DC power supply and sampling both multimeters and the first two harmonics from the lock-in amplifier. Delay time is introduced after switching harmonics on the lock-in amplifier and after setting the DC voltage to allow the measured signals to settle. The same voltage point may be sampled repeatedly. Data runs consist of two averages for the derivatives, at 250 evenly spaced voltage values over the diode’s operating range.

The setup was tested using a 100Ω resistor in place of the diode at 77K. The resistance, calculated from the first derivative, varied by only 0.02Ω over the voltage range, while the second derivative was identically zero as expected.

Analysis was completed on an Intel Pentium computer running Windows NT 4.0, using Microsoft Excel. Numerical algorithms (discussed later) were implemented in C++ using the Borland Ver. 5.0 compiler on a Pentium Pro computer running Windows NT 4.0.

Results

The I-V characteristic was sampled for three Germanium tunnel diodes (American Microsemiconductor models IN2712, TD9, and TD-261A) at room temperature, liquid nitrogen, and liquid helium temperatures. The I-V curves are smooth in all cases, with a slight drop in current as the temperature is decreased, due to the decrease in thermal energy available for emission of phonons (fig. 2a). Furthermore, the decrease in tunneling frequency at lower temperatures corresponds to an increase in resistance from 0 to 45mV, which is evident in the plot of dI/dV versus voltage (where dI/dV is R\(^{-1}\) within a first order approximation) (fig. 2b). The second derivative plot shows a dramatic change with temperature (fig 2c). As the temperature is lowered, the magnitude of the curvature is lowered due to the spreading of the I-V curve over a larger voltage range. Since low energies are involved, thermal smearing still hides the quantum phonon emissions even at liquid nitrogen temperatures. However, at 4.2K there are distinct peaks located at the energies of the emitted phonons in electron volts.

These peaks occur due to an increase in current when the applied energy across the diode equals the energy of the emitted phonon. Using the relation eV = hν (as discussed
Figure 2: The I-V characteristic for a tunnel diode, as well as the first two derivatives as a function of voltage, are shown for three temperatures. The plot of $d^2I/dV^2$ shows peaks at locations where the applied potential difference equals the energy of phonons emitted to conserve momentum during indirect tunneling.
earlier) one may compute the frequencies of the phonons. Table 1 shows this calculation for every peak in the second derivative plots for all three diodes. The resulting frequencies are compared with neutron scattering data from Brockhouse and Iyengar to determine the identity of the observed phonons [3].

<table>
<thead>
<tr>
<th>Phonon Energy (meV)</th>
<th>Frequency (10^12 Hz)</th>
<th>Freq. (10^12 Hz) from Brockhouse Data</th>
<th>Phonon Name and direction</th>
<th>Diode Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.79 ± 0.41</td>
<td>2.12 ± 0.10</td>
<td>2.45 ± 0.15</td>
<td>TA[100] and TA[111]</td>
<td>TD9, IN3712</td>
</tr>
<tr>
<td>9.13 ± 0.57</td>
<td>2.20 ± 0.14</td>
<td>1.95 ± 0.10</td>
<td></td>
<td>TD9, TD261A</td>
</tr>
<tr>
<td>9.31 ± 0.78</td>
<td>2.25 ± 0.19</td>
<td>2.45 ± 0.15</td>
<td></td>
<td>TD9, IN3712</td>
</tr>
<tr>
<td>28.72 ± 1.09</td>
<td>6.94 ± 0.26</td>
<td>6.90 ± 0.40</td>
<td>L[100]</td>
<td>TD9</td>
</tr>
<tr>
<td>29.80 ± 1.22</td>
<td>7.20 ± 0.29</td>
<td>7.40 ± 0.30</td>
<td>LO[111]</td>
<td>TD261A, IN3712</td>
</tr>
<tr>
<td>30.34 ± 1.38</td>
<td>7.33 ± 0.33</td>
<td></td>
<td></td>
<td>TD261A, IN3712</td>
</tr>
<tr>
<td>34.00 ± 0.89</td>
<td>8.21 ± 0.21</td>
<td>8.40 ± 0.30</td>
<td>TO[111]</td>
<td>IN3712</td>
</tr>
<tr>
<td>35.38 ± 1.59</td>
<td>8.54 ± 0.38</td>
<td></td>
<td></td>
<td>TD261A</td>
</tr>
<tr>
<td>36.76 ± 1.37</td>
<td>8.88 ± 0.33</td>
<td>9.00 ± 0.30</td>
<td>O[q=0]</td>
<td>TD9, TD261A</td>
</tr>
<tr>
<td>38.91 ± 1.64</td>
<td>9.40 ± 0.40</td>
<td></td>
<td></td>
<td>TD9, TD261A</td>
</tr>
</tbody>
</table>

Table 1: The energies and frequencies of phonons observed during indirect tunneling in germanium at 4.2K are compared with frequencies measured with a neutron scattering experiment [3]. Phonon names are transverse optical (TO), longitudinal optical (LO), transverse acoustical (TA), and optical (O) and directions are specified by their wave vector, \( \mathbf{q} \).

Phonons are emitted from harmonic oscillations of crystal lattices in various directions that must be tested separately in neutron scattering experiments [7]. Indirect tunneling experiments do not allow measurement of phonon direction, but the expected value is included in table 1 for comparison. Also, the germanium and silicon crystal structures are the same so the expected phonon frequencies of silicon are simply scaled by a constant factor (~1.75) due to the change in atomic mass and spacing [7].

The error in the calculated energies for the phonons is due in part to the input impedance of the current pre-amp. This contribution is less than or equal to the current through the diode, providing an uncertainty from 0 to 1.1 meV in the phonon energy. There is an additional uncertainty involved in identifying the locations of peaks in the second derivative plot, which is estimated. For this analysis, phonon emissions are assumed to occur at the local maxima of the graphs, but the overlapping of nearby peaks may result in a much more complex plot. In particular, the results indicate that the TA[100] and TA[111] combine to form a single bump directly between the locations of the individual peaks, so these phonons cannot be resolved independently. In general the results correspond well with the neutron scattering data from Brockhouse which is included in table 1 for comparison.
Numerical Analysis

Numerical differentiation of the I-V characteristic is explored as an alternative for pinpointing the locations of phonon emissions. Numerical differentiation only requires measurement of the voltage and current values, which could be done with improved accuracy by placing a resistor in series with the diode and calculating the voltage drop independent of the pre-amp’s input impedance.

A cubic spline interpolation method was attempted first, which fits a unique third degree polynomial between consecutive data points (termed a zone in this paper) (fig. 3) [8]. The method constrains the function and its first two derivatives to be continuous at zone boundaries, which requires solving the equations,

\[ y_i = A + B x_i + C x_i^2 + D x_i^3 \]
\[ y'_i = B + 2C x_i + 3D x_i^2 \]
\[ y''_i = 2C + 6D x_i \]
evaluated at the edges of the zone of interest.

Only six equations are provided with the known variables \( x[i], x[i+1], y[i], \) and \( y[i+1], \) while there are 8 unknown quantities: \( A, B, C, D, y'[i], y'[i+1], y''[i], \) and \( y''[i+1], \) corresponding to the unknown polynomial coefficients and derivatives. However, the situation could be worse, because with \( n \) data points there are \( n-1 \) zones in the entire function. This equates to \( 6(n-1) \) equations and \( 6(n-1)+2 \) unknowns since the derivatives \( y' \) and \( y'' \) must be equal at the borders of adjacent zones. The extra 2 unknowns arise since the left most boundary and right most boundary are not repeated.

To solve this tridiagonal system of equations two values must be assumed. A logical choice is to assume the value of \( y[1]' \) (the left most value in the entire domain) and \( y''[n]. \) Then, starting in the left most zone, solve for \( y'[2] \) in terms of only one unknown variable such as \( y'[2] = AA*y'[2] + BB, \) where \( AA \) and \( BB \) are constants. Since the first and second
derivatives are equal at zone edges, in the zone to the right \( y''[2] \) can be eliminated from the equations, and the exact same procedure may be used to tabulate the constants AA and BB in every zone. Once the right most zone is reached, the value for \( y''[n] \) is assumed, which allows direct calculation of the coefficients for the cubic polynomial in that zone. Furthermore, this allows calculation of \( y''[n-1] \) which can be used in the same procedure in the zone to the left to calculate the coefficients.

This algorithm works well for smooth input, but oscillations occur when random noise is present in the data. This becomes a greater problem as the x data values become closely spaced, so the implementation includes an averaging system. For example, with two averages the program performs the cubic spline interpolation using every other data point and averages the results with the cubic spline on the remaining data points.

![Interpolated Second Derivative of Current vs Voltage at 4K](image)

**Figure 4:** The second derivative of a diode's I-V characteristic using the cubic spline method is compared with the electronically differentiated signal using a lock-in amplifier. Without averaging oscillations range from -100 to 100, generally alternating every point (corresponding to every computation zone in the method). The averaging damps extrema to the range -0.1 to -0.5, but the oscillation frequency depends on the number of averages which is unacceptable.

The resulting interpolation method with 20 averages provides a nearly exact fit of the I-V characteristic as well as the first derivative. While the resulting second derivative is a continuous function as prescribed by the method, it is completely obscured by small errors in interpolation that manifest themselves as oscillations (fig 4). Averaging creates another oscillation with a period related to the number of averages. This dependence on a parameter of the numerical method is unacceptable.

Another method for numerical differentiation utilizes a least squares fit of a high order polynomial to the entire data set [9]. This minimizes the goodness of fit parameter, \( \chi^2 = \sum [ y_i \)
\[ - \sum a_j x_i^j \], where \((x_i, y_i)\) are known data points and \(a_j\) are the unknown coefficients which can be found by solving the linear system of equations:

\[ \sum_i y_i / \sigma_i^2 = \sum_i \frac{1}{\sigma_i^2} [A + Bx_i + Cx_i^2 ] \]

\[ \sum_i y_i x_i / \sigma_i^2 = \sum_i \frac{x_i}{\sigma_i^2} [A + Bx_i + Cx_i^2 ] \]

\[ \sum_i y_i x_i^2 / \sigma_i^2 = \sum_i \frac{x_i^2}{\sigma_i^2} [A + Bx_i + Cx_i^2 ] \]

and so on.

The \(i\) index ranges over the number of data points sampled, the constants \(A, B, C,\) etc define the fitting polynomial, and \(\sigma_i\) is the uncertainty in \(y_i\). By treating these equations as matrices, the row matrix of unknown coefficients, \(\Gamma\), is given as \(\Gamma = \beta \alpha^{-1}\) where \(\beta\) is the row matrix given by the left hand side of the above equations and \(\alpha\) is the square matrix where the right hand side equals \(\Gamma \alpha\). The inverse of a matrix is taken with a Gauss-Jordan elimination technique, which converts \(\alpha\) to the unity matrix (\(\varepsilon\)) and performs the same operations on \(\varepsilon\). Since the \(\alpha^{-1} = \varepsilon / \alpha\), this method results in \(\varepsilon\) being converted to \(\alpha^{-1}\).

**Figure 5:** Plot of the resulting second derivative using a least squares method to fit an 11th order polynomial to the I-V data and exactly differentiating the resulting function. It is normalized from the range \(-0.47\) to \(-0.22\). The numerical differentiation displaces the peaks which is unsuitable for calculating phonon frequencies.

The current versus voltage data (250 points) for a 4.2K experimental run was fit to an 11th order polynomial assuming \(\sigma_i\) was constant, and the resulting function was exactly differentiated twice. The original function and the first derivative are fit perfectly when “eyeballing” the plots. Figure 5 compares the resulting numerical differentiation for the
second derivative with the electronically differentiated signal, which shows a weakness of the method. In particular, the function is greatly smoothed even though the second derivative is represented by a ninth order polynomial. The technique is unacceptable for calculating the energies of phonon emissions, since the maxima of the numerically differentiated signal are slightly displaced from their actual positions as shown by the electronic differentiation.

Conclusions

By measuring the I-V characteristic for several germanium tunnel diodes, the quantum mechanical indirect tunneling effect is witnessed. Indirect tunneling is observed as an increase in current through the diode (which manifests itself as a peak in the second derivative) when the applied potential exactly equals the energy of the emitted phonon. Using electronic differentiation, the locations of the peaks are used to calculate the frequencies of the phonons. This data is consistent with the results of a neutron scattering experiment by Brockhouse and Iyengar, which allows identification of the phonons in question [3].

Both a cubic spline interpolation method and a least squares fitting method were implemented in order to test numerical differentiation of the I-V characteristic. Numerical differentiation would allow a simplified experimental setup with lower uncertainty in the measured voltages. Neither method performs satisfactorily, for the spline method oscillates with the random noise in the data and the least squares method smooths the second derivative excessively. Therefore, electronic differentiation is deemed a critical component of the experimental method.

References