

Ni–NiO–Ni tunnel junctions for terahertz and infrared detection

Philip C. D. Hobbs, Robert B. Laibowitz, and Frank R. Libsch

We present complete experimental determinations of the tunnel barrier parameters (two barrier heights, junction area, dielectric constant, and extrinsic series resistance) as a function of temperature for submicrometer Ni–NiO–Ni thin-film tunnel junctions, showing that when the temperature-invariant parameters are forced to be consistent, good-quality fits are obtained between I – V curves and the Simmons equation for this very-low-barrier system (measured $\phi \approx 0.20$ eV). A splitting of ≈ 10 meV in the barrier heights due to the different processing histories of the upper and lower electrodes is clearly shown, with the upper interface having a lower barrier, consistent with the increased effect of the image potential at a sharper material interface. It is believed that this is the first barrier height measurement with sufficient resolution for this effect to be seen. A fabrication technique that produces high yields and consistent junction behavior is presented as well as the preliminary results of inelastic tunneling spectroscopy at 4 K that show a prominent peak at ~ 59 meV, shifted slightly with respect to the expected transverse optic phonon excitation in bulk NiO but consistent with other surface-sensitive experiments. We discuss the implications of these results for the design of efficient detectors for terahertz and IR radiation. © 2005 Optical Society of America

OCIS codes: 240.7040, 160.3900, 040.3060, 040.5160, 130.0250, 190.4360.

1. Introduction

Antenna-coupled tunnel junction (ACTJ) devices are of significant interest as fast detectors of terahertz and IR radiation.^{1,2} They consist of a metal–insulator–metal (MIM) tunnel junction coupled to a thin-film metal antenna. The Ni–NiO–Ni junction is especially useful because of its products with very low resistance–area (RA) products [as low as $1 \Omega(\mu\text{m}^2)$]. The RA product is closely related to the RC time constant of the junction, and at $1 \Omega(\mu\text{m}^2)$ the intrinsic time constant is near 3×10^{-14} s. Wilke *et al.*³ and Fumeaux *et al.*⁴ have demonstrated high-quality Ni–NiO–Ni junctions with low resistance–area products and have shown that these junctions coupled to thin-film metal antennas can be used as IR detectors and frequency mixers in the $10 \mu\text{m}$ band. More recently, Fumeaux, Boreman, and co-workers^{5,6} have extended similar results to the $3.39 \mu\text{m}$ region

and the visible, although with low sensitivity. Both Wilke *et al.* and Fumeaux *et al.* have made very small devices, estimated to be 0.056 and $0.01 \mu\text{m}^2$, respectively. However, both of these area estimates are in doubt, since both quote only the projected area of the crossing of two rather tall and narrow thin-film wires. This procedure takes account of only the horizontal surface of the bottom metal, whereas, since there is no planarization step between metal depositions, the top metal layer would touch the sidewalls of the bottom metal; since the line dimensions in Wilke *et al.* were 240 nm wide and 220 nm tall and those in Fumeaux *et al.*, 100 nm wide by 200 nm tall, these sidewall areas amount to 180% and 400% of the quoted (projected) area, respectively.

In this paper we present a novel application of a fabrication method for building repeatable devices with a high yield as well as characterization data showing I – V curves and tunneling spectroscopy across the range from 4.2 to 300 K with a preliminary extension to 423 K . In addition we show extracted junction parameters (area, barrier thickness, barrier height, dielectric constant, and extrinsic series resistance) that enable these junctions to be characterized with high accuracy; this parameterization enables future device designs. Fitting these parameters requires modifications of the standard Simmons-

P. C. D. Hobbs and F. R. Libsch are with IBM T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598. R. B. Laibowitz is with the Department of Electrical Engineering, Columbia University, New York, 10027.

Received 21 January 2005; revised manuscript received 5 April 2005; accepted 2 May 2005.

0003-6935/05/326813-10\$15.00/0

© 2005 Optical Society of America

equation approach,⁷ which has allowed what we believe to be the first definite observation of barrier-height splitting in a single-metal MIM junction.

Most tunnel barrier fitting attempts in the literature are based on the two classic 1963 papers of Simmons, treating symmetric⁸ and asymmetric⁹ barriers. Apart from two errors in the second paper (pointed out in Appendix B), these papers present the correct zero-temperature WKB quantum treatment of one-dimensional (1-D) tunneling between two Fermi levels where the details of crystal orientation, disorder, insulator band structure, and interface morphology are unknown, as they usually are in thin-film devices, and magnetic effects are not dominant. This theory required computations difficult in 1963, so both papers rely on very rough analytical approximations necessary for hand calculation. These approximations are suitable for low-accuracy work when the barrier heights are large, but we found that at the Ni–NiO barrier height, shown below to be ≈ 0.2 eV, these approximations lead to significant inaccuracies and even to unphysical results such as large negative currents flowing at a positive bias. In addition, since we need to extract several parameters from the fitted curves, it is vital that the fits be as close as possible: Low-resistance junctions have I – V curves that are nearly straight lines, which makes the fits very sensitive to poor-quality approximations. Good fits require many data points and accurate fit functions, which means a lot of computation. Hence we also present a computationally more efficient algorithm than brute force numerical integration of the Simmons equation.

A. Infrared and Terahertz Detection

In discussions of optical processes in metals and semiconductors, we typically use an approximate material model consisting of an ensemble of single-electron states. At low frequencies we use a mean-field theory that takes into account the collective response of the electrons. In the terahertz and IR regions both the collective and the single-particle pictures become problematic. Nonetheless certain metals (particularly the free-electron metals Cu, Ag, and Au) still make good antennas at frequencies as high as 200 THz, although since their dielectric constants have large negative real parts and small imaginary parts (so that $k \gg n$, where $\tilde{n} = n + ik$ is the complex refractive index), they are not well described by a normal conductor model, which predicts $n \approx k$.

For IR and terahertz detection the tunnel junction coupled to a metal antenna can be regarded as a circuit element. Ni–NiO–Ni junctions have been shown to be fast enough to rectify currents at optical frequencies,¹⁰ and their rectifying properties are thought to be essentially invariant from dc to visible wavelengths.^{11,12} An antenna exhibits a radiation resistance R_A , which forms a voltage divider with the load resistance. If the load is a tunnel junction of differential resistance $r_j = R_A$ and negligible capacitance, the small-signal dc responsivity to received

optical power P_{opt} is

$$\mathfrak{R} \equiv \frac{I_{\text{dc}}}{P_{\text{opt}}} = \frac{r_j}{4} \frac{d^2 I}{dV^2}. \quad (1)$$

One factor of 2 comes from averaging \sin^2 over a cycle and the other from 2! in the Maclaurin series. This ratio is the responsivity, which is the appropriate figure of merit for an optical detector. Because numerical second derivatives are so noisy, it is difficult to measure this accurately, so the best procedure is to fit the I – V curve by using the Simmons function and compute the derivatives from the fit. (Alternatively one can use a sinusoidal excitation with harmonic detection to get direct measurements of derivatives, as in Wilke *et al.*) The devices that we discuss here are significantly larger ($0.4 \mu\text{m}^2$) than those cited above. There are three reasons for this choice. First, our research is aimed at traveling-wave devices for the $1.5 \mu\text{m}$ band, so we can relax the requirement for very small devices since the lumped-constant RC product is not a fundamental limit. Second, to gain confidence in the physical significance of the fitted junction parameters, it is important to know the junction area accurately, which is much easier with larger junctions and thinner metal layers. Third, we think that it is probably impossible to achieve a technologically useful electromigration lifetime with the smallest ($0.01 \mu\text{m}^2$) junctions. High-efficiency detection requires good impedance matching to antennas, whose impedances are almost always in the 30–300 Ω range. Ni–NiO–Ni junctions exhibit maximum sensitivity at bias voltages of 80–100 mV, so the current density in a 100 Ω junction of $0.01 \mu\text{m}^2$ would be approximately 10^7 A/cm², which is an order of magnitude above the commonly quoted limit of 10^6 A/cm² for upper-level wiring on integrated circuits.

B. Junction Parameter Fitting

The fitting problem is to minimize the mean-squared discrepancy between a measured I – V curve and a curve calculated from the junction area A , barrier heights ϕ_1 and ϕ_2 , barrier thickness s , dielectric constant K , and extrinsic series resistance R_s . Note that R_s is not the same as R_0 , the differential resistance at zero bias, which was discussed by Wilke *et al.* and which is

$$R_0 = R_s + \left. \frac{\partial V_j}{\partial I} \right|_{I=0}, \quad (2)$$

where V_j is the voltage across the intrinsic junction, i.e., after correction for the extrinsic voltage drop IR_s .

At $T = 0$ the WKB current density calculation is based on Eq. (3) of Ref. 9:

$$J = J_s \exp \left\{ \left(\bar{\phi} + \frac{eV_{21}}{2} \right) \exp \left[-A \left(\bar{\phi} + \frac{eV_{21}}{2} \right)^{1/2} \right] - \left(\bar{\phi} - \frac{eV_{21}}{2} \right) \exp \left[-A \left(\bar{\phi} - \frac{eV_{21}}{2} \right)^{1/2} \right] \right\}, \quad (3)$$

where $V_{21} \equiv V_2 - V_1$ is the junction bias voltage measured from side 2 to side 1 and J_s and A are given by

$$J_s = \frac{e}{h(\beta\Delta s)^2}, \quad (4)$$

$$A = \frac{4\pi\beta\Delta s}{h}(2m)^{1/2}. \quad (5)$$

The quantity $\bar{\phi}$ is the mean value of the barrier height between the classical turning points of the motion, and the adjustment factor $\beta \approx 0.98$ is given by Eq. (A4).

In Appendices A and B we discuss this further and give a method for evaluating $\bar{\phi}$ efficiently, which makes J computationally inexpensive, except for evaluation of β . Fortunately β is such a weak function of the junction parameters that for curve-fitting purposes a value of 0.98 can be used during most of the fit process, with only the last iterations needing the full numerical integration of Eqs. (3) and (A4).

The fit procedure adopted here is as follows. (A program written by one of the authors was used for calculation.) This is not a complete algorithm, but no essential aspect is omitted:

(1) Initialization

- (a) Assign a range to each variable (the N variables to be optimized are given nonzero ranges).
- (b) Choose random starting values within each range: $N + 1$ sets are needed by the Nelder-Mead method.^{13,14}

(2) Preliminary Iteration

- (a) Find the intrinsic junction bias voltage V_{21} from the measured bias by subtracting IR_s , where I is the measured tunnel current. Assuming the residuals to be small, this ensures a self-consistent solution without a subiteration being necessary to find the R_s correction at each point on each iteration.
- (b) Use Eq. (3) at each value of V_{21} to find I from the junction parameters.
- (c) Compute the mean-square residual.
- (d) Generate the next set of parameters according to the optimization method chosen, or exit if converged.

(3) Final Iteration

With the preliminary values of the parameters, generate new guesses in their neighborhood and iterate again but this time use the numerical solution of Eq. (A4) for β . This makes each iteration much slower, but often only a few iterations are needed, and this polishing is necessary only once—exploration of the parameter space can be done with $\beta = 0.98$. (In the present work the main effect of setting $\beta = 0.98$ is an underestimation of R_s .)

In general the mean-square residual function exhibits multiple minima, and this tendency becomes worse as the temperature increases, the measurements become noisier, or the accuracy of the current density calculation deteriorates. Unfortunately there is no guarantee that the global minimum at a given temperature corresponds to the real device parameters, but forcing the geometric parameters to be constant with temperature does lead to reliable results, as will be shown.

2. Experimental Procedure

A. Device Fabrication

The devices used here were fabricated by a germanium shadow-mask technique¹⁵ similar to that of Jackel *et al.*,¹⁶ which uses a three-layer structure consisting of a 600 nm poly(methyl methacrylate) (PMMA) bottom layer, a 50 nm germanium middle layer, and a 600 nm PMMA top layer. The desired pattern is written on the top layer with electron-beam lithography. (A modified process allowing the use of 248 nm optical lithography is under development.) Reactive-ion etching is used to transfer the pattern to the Ge layer, and an oxygen ash step then removes the top layer and undercuts the bottom layer, leaving a suspended Ge shadow mask in the form of slots with bridges. Nickel is then put down in two steps, in thicknesses of 40 ± 5 nm each, by using angled evaporation near room temperature. After the first metal level the sample is removed from the evaporator, and the NiO barrier is formed by oxidizing the nickel briefly in an O_2 plasma similar to that used to ash the PMMA but at a much lower rf power. The second Ni layer is then deposited on top, evaporating from a different angle, and the Ge is lifted off in hot acetone. Over several fabrication runs, by using two-level electron-beam lithography with lift-off, we found that ambient or thermal oxidation of the Ni produces unstable junctions with poor repeatability and low yield, but the shadow-mask patterning and O_2 plasma produce repeatability of $\pm 10\%$ or better in junction resistance, with yields of $> 80\%$. Furthermore shadow-mask fabrication affords excellent control of the junction areas, since the two metal layers are defined by a single layer of lithography and the overlap depends only on the PMMA thickness and the angle of deposition; with a 600 nm PMMA layer a 5° rotation about a 40° nominal deposition angle shifts the metal edge by less than 100 nm.

Figure 1 shows an electron micrograph of one of the junctions used in this work, and Fig. 2 shows a similar junction with the Ge bridge still in place. The obvious sharpness of the shadows allows a good junction area measurement by pixel counting. The geometric area of this junction is measured to be $0.40 \pm 0.02 \mu\text{m}^2$, of which the correction for the sloping sidewall area contributes less than $0.01 \mu\text{m}^2$. All junctions are fabricated on the same chip with a single level of electron-beam lithography, so all their overlap areas are closely similar. The lumps on the metal are due to nucleation occurring at sites of mi-

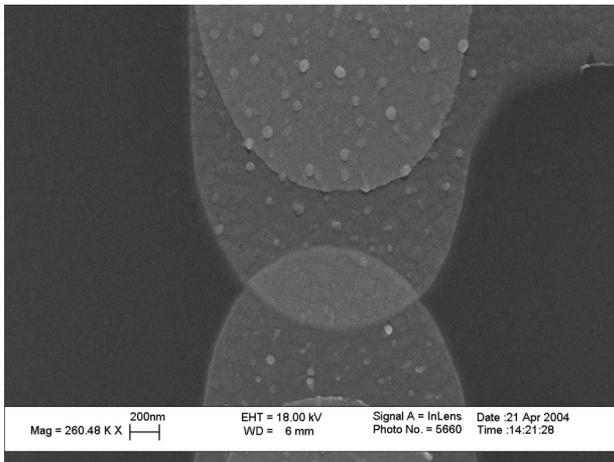


Fig. 1. Scanning electron micrograph of a Ni-NiO-Ni tunnel junction fabricated by the Ge shadow-mask technique. The junction area is $0.40 \pm 0.02 \mu\text{m}^2$, and each metal layer is $50 \pm 5 \text{ nm}$ thick.

nor surface damage to the thin SiO_2 underlayer that seems to have occurred during photoresist ashing. This was verified by scanning electron microscopy pictures of a similarly prepared sample before metal deposition.

B. I - V Curve Tracing

Room-temperature I - V measurements were carried out with a standard four-wire probe station. For low-temperature measurements the device was wire-bonded to a header and placed in a helium cryostat. Current and voltage were measured with a four-wire apparatus. Figure 3 shows four I - V curves from a single test device over the range of 4–295 K. The nonlinearity is about four times larger at low temperature. Figure 4 shows a typical fitted I - V curve together with the fit residuals plotted as $10(I_{\text{exp}} - I_{\text{fit}})$. Apart from the very small hooks visible at the edges (possibly due to joule heating) the fit residual is limited by measurement noise. In all these curves a

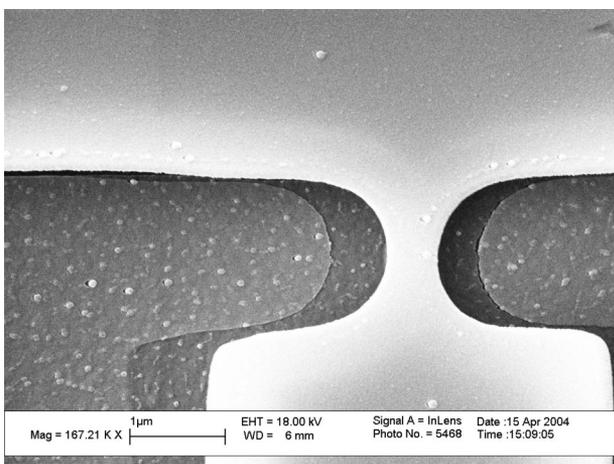


Fig. 2. Suspended germanium shadow mask of the type used in this work.

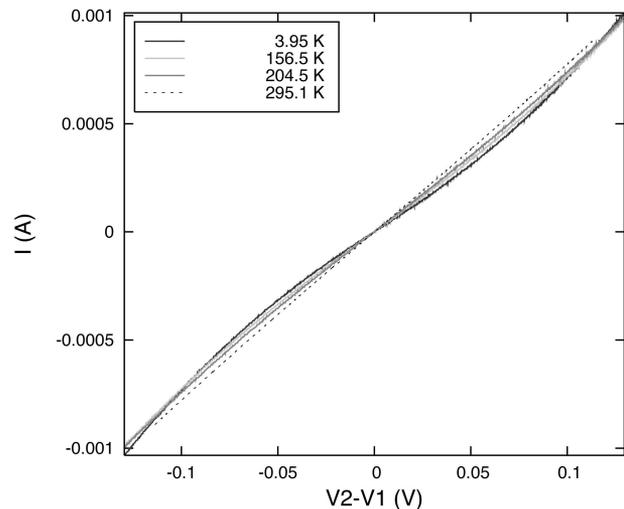


Fig. 3. Some of the I - V curves for a junction similar to that in Fig. 2 at various temperatures. As the temperature rises, the curves become straighter.

positive V_{21} means that the top metal $M2$ is positive with respect to the bottom metal $M1$. Although the device tested here did not survive long enough to be tested above 300 K (which required moving to a different apparatus), another similar device produced fitted barriers near 0.18 eV at 423 K, which continued the trend of barrier height versus T .

C. Parameter Extraction

The approximately 20 junctions measured had qualitatively a similar junction resistance and nonlinearity, and the yield was very good, $\sim 80\%$. For the wide temperature range measurements shown here we used a single junction, so it is appropriate to constrain these fits to make the geometric properties (e.g., the junction area and barrier thickness) constant over T . Besides internal logical consistency this allows an important check: When the geometric pa-

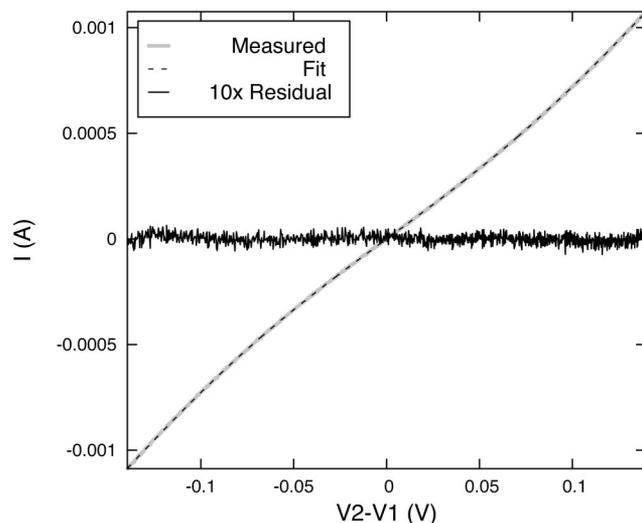


Fig. 4. Typical fitted I - V curve with residual $10(I_{\text{exp}} - I_{\text{fit}})$. This is the 156.5 V curve of Fig. 3.

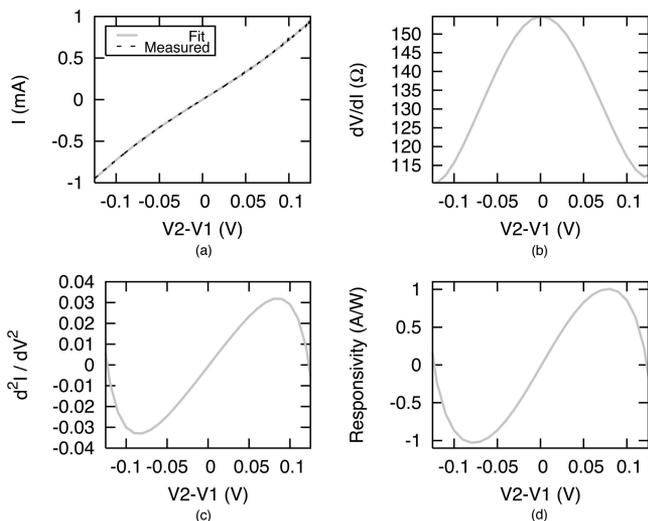


Fig. 5. Parameter extraction from I - V data: $T = 156.5$ K.

rameters are fixed, does the scatter in the other fitted parameters improve or deteriorate?

Figure 5 shows a typical fitted experimental I - V curve measured at 156.5 K. Figure 5(a) shows experimental and fit curves plotted together; Fig. 5(b) shows the differential resistance $r_J = dV/dI$; Fig. 5(c) shows the second derivative d^2I/dV^2 of the fit function; Fig. 5(d) is the responsivity calculated from the fit function by using Eq. (1). Here V_{21} is the voltage measured from $M2$ to $M1$ as above. Two runs of data are shown that lie on top of each other to within the noise level. The responsivity is quite respectable for an IR detector, even though the curves appear nearly straight.

Figure 6 shows a summary of fitted data from several temperatures, obtained by using unconstrained Nelder-Mead downhill simplex optimization on all the junction parameters (junction area A , oxide dielectric constant K , oxide thickness s , barrier heights ϕ_1 , ϕ_2 , and extrinsic series resistance R_s). Most of the points lie on well-defined trajectories but with significant scatter and a few outliers. The barrier heights show no consistent trend except that ϕ_2 is always less than ϕ_1 . Geometric parameters are almost temperature independent, and the dielectric constant of insulating metal oxides is typically a weak function of temperature as well. None of these shows a clear trend. Fixing the values $A = 0.3 \mu\text{m}^2$, $s = 2.5$ nm, and $K = 8.4$ produced the improved fits of Fig. 7. The outliers have disappeared, the scatter around the trend curves has decreased significantly, and the fit residuals have not increased, all of which support the hypothesis that these parameters represent the true physical parameters of the junction. It is interesting that the fitted area in Fig. 6(a) is $\sim 75\%$ of the geometric area, suggesting that the tunneling is occurring rather uniformly over the overlap area. Aluminum-oxide junctions typically behave as though their active areas are a small fraction of the geometric area, suggesting that the tunnel current is dominated by thin spots in the barrier.¹⁷ This is phys-

ically reasonable, because the wave function dies off much more steeply in the junction region with typical Al_2O_3 barrier heights of a few electron volts than with the 0.2 eV barriers that we measure in Ni-NiO. This effect may also be partly responsible for the improved consistency of these results compared with I - V fits in high-barrier systems.

Barrier heights ϕ_1 and ϕ_2 decrease monotonically with increasing T , slowly at first but then with a slope, $d\phi/dT \approx -0.15$ meV/K, somewhat steeper than a pure kT reduction, which would have a slope of -0.086 meV/K. The barriers decrease by ~ 24 meV from 4 to 300 K, which is approximately kT , and the asymmetry declines by almost half. The approximately quadratic variation of $\phi(T)$ is as expected. Thermal excitation increases the occupation number of states above the Fermi level and correspondingly depletes those below; at low temperatures the two nearly cancel, causing the linear term in $\phi(T)$ to be small. The reduced barrier asymmetry is also qualitatively reasonable. The barrier area integral is evaluated between the classical turning points of the motion. The turning points move toward the barrier peak as the energy increases, so higher-energy states that extend farther into the oxide will see less asymmetry from the composition gradient at the lower interface.

The metal resistance curve in Fig. 7(b) is reasonable as well. When it is compared with the scaled resistance of a nearby metal line (also shown), R_s rises more steeply with T , looking much more like the tabulated curve for bulk Ni (Ref. 18) plus an offset, as shown. More work is needed to determine whether the steeply rising part of R_s versus T is a real metal resistance or a fit artifact, due to, e.g., the effects of finite temperature on the tunneling probability.

The origin of the extrinsic series resistance R_s is a somewhat vexing issue. The data of Fumeaux *et al.* and of Wilke *et al.* agree with the resistances and responsivities obtained here, although only qualitative comparisons can be made because of the lack of tabulated data points for fitting and the large uncertainty in their junction areas (factors of 3 and 5, respectively) due to neglect of the sidewall contributions. Wilke *et al.* claimed that the relatively small change that they saw in the zero-bias differential resistance R_0 versus temperature cannot be due to the thin-film structure of the metal, but Fig. 7(b) shows that the resistance of our thin-film Ni line (which was of uniform cross section) alone has a similar temperature dependence ($R_{77\text{ K}}/R_{300\text{ K}} = 0.58$ to Wilke's value value of ($R_{77\text{ K}}/R_{273\text{ K}} = 0.65$ for R_0). This is quite different from the behavior of bulk Ni, despite their claim that a film much thicker than the 6.7 nm electron mean free path should have bulk properties. The discrepancy is probably due to the disorder and columnar structure of unannealed thin films, but merely saying this leaves us little closer to the answer.

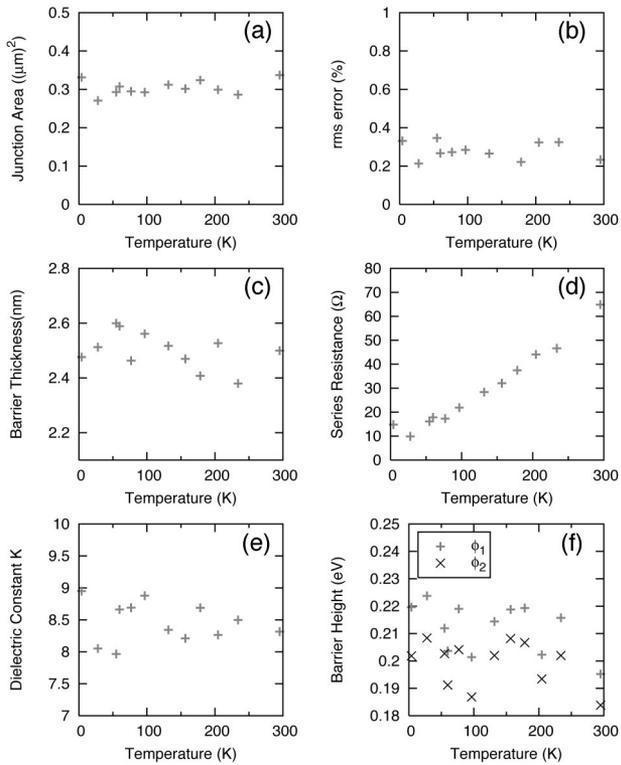


Fig. 6. Tunnel junction parameters extracted from the tested junction with all parameters free. Note the general consistency of (a) the area, (c) barrier thickness, and (e) dielectric constant. These figures show scatter but no clear trend.

D. Barrier Asymmetry

Since the top and bottom metal layers are nominally identical, one might expect the junctions to be highly symmetric electrically, but Fig. 7(a) shows that this is not so; the barrier heights differ by as much as 20 meV. We conjecture that this is due to a composition gradient in the lower Ni–NiO interface, owing to interdiffusion in the oxidation step. This interdiffusion is not present in the upper interface because *M2* is evaporated on top of the oxide near room temperature, so there is no oxygen plasma bombardment of the metal at the upper interface. Apart from the small asymmetry in the *I–V* curve that results, the barrier-height inequality is not significant for detector operation, but this mechanism may provide a way to control the barrier heights to optimize responsivity. It is interesting that the lower barrier is on the side with the more abrupt junction. This is consistent with the image potential being stronger there because of the more nearly ideal metal surface. We believe that this is the first barrier-height measurement in any material system with enough resolution to show this small splitting.

E. Calculated Detector Responsivity

These devices are potentially good detectors for IR and terahertz signals. Figure 8 is a plot of responsivity versus bias voltage for the same temperatures as in Fig. 3, when the extracted junction parameters from the constrained fits are used. If the barrier

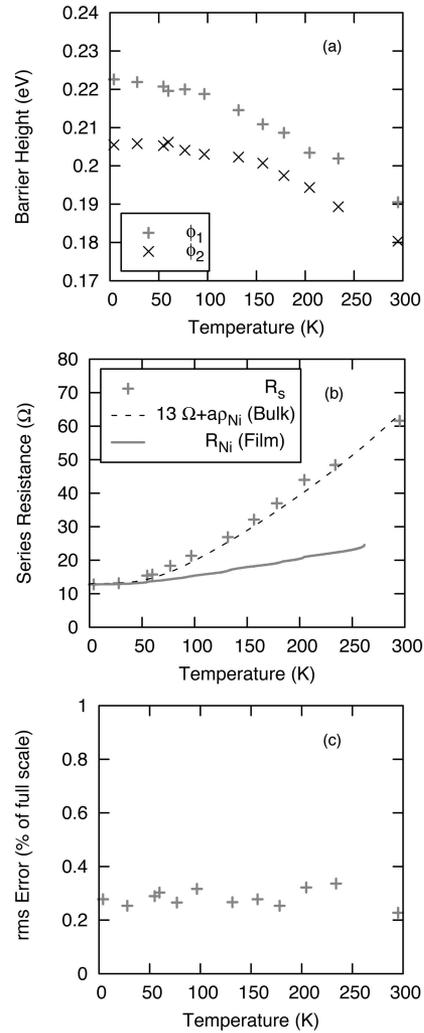


Fig. 7. Junction parameters extracted from the same data set as before but with *A*, *s*, and *K* fixed. Thickness, 2.50 nm; area, 0.3 μm²; *K*, 8.4. Note the reduced scatter and the resemblance of *R_s* to the resistivity of bulk Ni.

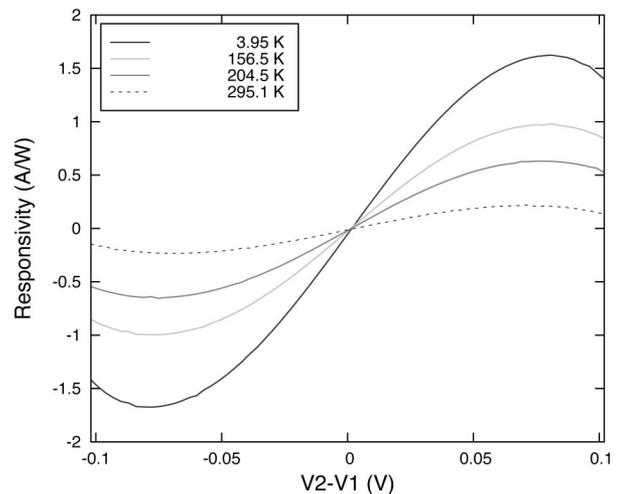


Fig. 8. Calculated responsivity of the test junction as a detector versus *T*. The low responsivity at 298 K (≈ 0.2 A/W) is due to increased metal resistance and reduced barrier heights, which can both be addressed in device design.

heights can be kept near 0.21 eV at room temperature and the metal resistance reduced, room-temperature responsivity values, $\mathfrak{R} > 1 \text{ A/W}$, may result. Barrier heights are commonly tuned by hundreds of milli-electron-volts in the gate metallization integrated circuit processes of the complementary metal-oxide semiconductor (CMOS), so this is not expected to be very difficult; the ability to adjust such an important parameter will be needed in a technologically useful process.

F. Junction Time Constant

Besides the intrinsic speed of the tunneling itself, the RC time constant of the junction may limit its response bandwidth, as is common in semiconductor detectors. Since the fits give a good value for the K seen by the tunneling electrons, we can readily estimate that the capacitance per square meter is

$$C/A \approx \epsilon_0(K/s) = 0.03F/m^2 \quad (6)$$

so that at an RA product of $1 \Omega(\mu\text{m})^2$ the junction time constant is $RC \approx 30 \text{ fs}$. This is only a rough estimate, since on these length scales it is not certain that the plate separation is the same as s in the tunneling equation. Still, since a 3 dB loss at 200 THz corresponds to a time constant of 0.8 fs, it is clear that efficient ACTJ detectors in the near IR will have to use reactive or traveling-wave techniques to reduce the effects of capacitance.

3. Inelastic Tunneling Spectroscopy

In extrapolating from dc measurements to junction properties at 10^{14} Hz , it is important to be certain that tunneling is the dominant conduction mechanism. The results in the sections above show that this is the case, owing to the very close fits to the Simmons equation, but other lines of evidence are reassuring. Optical detection is one of these, and another is inelastic tunneling spectroscopy, which is performed by taking the second derivative of the I - V curve by adding a small sine-wave component to the applied voltage and detecting its second harmonic. Nickel oxide has a complicated phonon spectrum, since its unit cell has a small rhombohedral distortion and significantly noncubic behavior due to magnetic coupling effects, as reported by Chung *et al.*¹⁹ Figure 9 shows a preliminary inelastic tunneling spectrum of a junction on the same chip as that used for the parameter extraction. The spectrum shows a pronounced peak near 59 mV with a shoulder on the side nearest 0 V. The peak position differs by some 9 mV from the transverse-optic phonon dispersion curves of Chung *et al.* but, as is interesting, agrees much better with a peak found in the room-temperature scanning tunneling microscope data of Olejniczak and Bieniecki,²⁰ who found a large peak near 59 mV with the NiO surface negatively biased with respect to the tungsten scanning tunneling microscope tip.

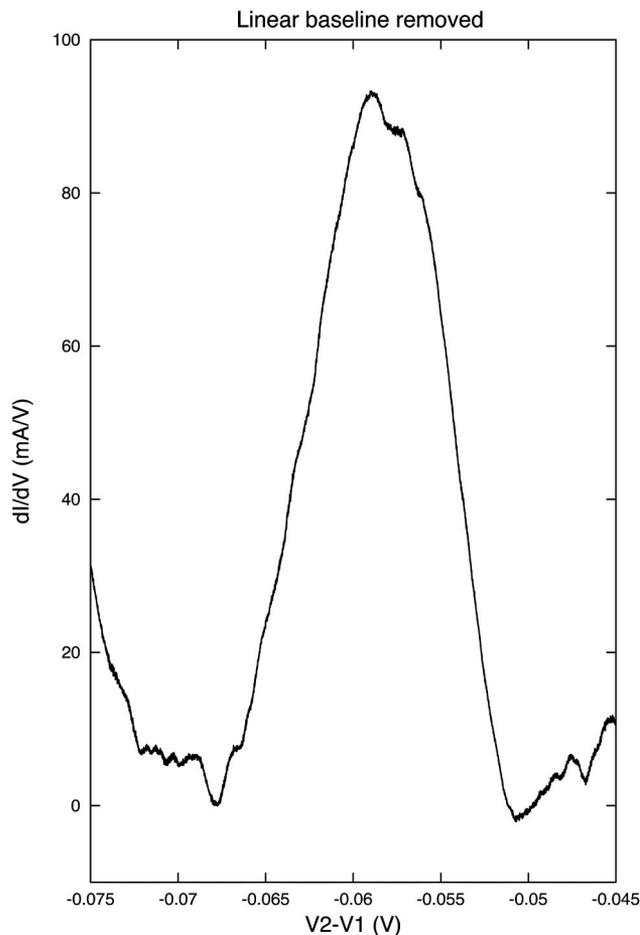


Fig. 9. Inelastic tunneling spectra at 4 K of a sample similar to the one used in the parameter extractions, showing what appear to be transverse-optic phonon peaks of NiO.

4. Discussion

A. Fit Accuracy

There is great improvement in the scatter of the fitted barrier heights and extrinsic series resistances when the consistency conditions of A , s , and K are applied. In general faulty constraints increase the fit residuals and the scatter of the fitted values. The low scatter suggests that the values used are close to correct. These values are taken as an approximate average of the free fits of Fig. 6 with the outliers disregarded. A more systematic procedure would be to use two nested iterations, with the outer loop adjusting the constraints to minimize the total rms residual over all T values and the inner loop optimizing the remaining free parameters for each T . Although such a program is not difficult to write, its run time is prohibitive unless further significant improvements in efficiency are available. To achieve full confidence in the fitted results, it is necessary to perform these fits over a larger set of junctions, measure R_s independently, and seek independent experimental support, e.g., from spin polarization and optical detection. Further work in optical detection with these devices is under way.

B. Systematic Errors

The major sources of systematic error in this work are believed to be joule heating and electromigration. A structure of characteristic size a cooled by three-dimensional conduction into a substrate of thermal conductivity α will have a junction-to-ambient thermal resistance of approximately

$$\Theta_{JA} \sim \frac{1}{a\alpha}, \quad (7)$$

which for a 1 μm junction on doped silicon is 10^4 K/W at room temperature [$\alpha \approx 100$ W/(m K)] (Ref. 21) and much less at low temperature, reaching 200 K/W at 20 K [$\alpha \approx 5$ kW/(m K)]. Thus the room-temperature ΔT for a 100 Ω junction with 100 mV of bias is

$$\Delta T \sim \frac{(0.1 \text{ V})^2}{100 \Omega} (10^4 \text{ K/W}) = 1 \text{ K}, \quad (8)$$

corresponding to a resistance change of a few tenths of a percent in the heated region, which is itself a small part of the total extrinsic resistance. The effect is also at least an order of magnitude less at low temperature ($T < 100$ K) because of the rapid increase of α . The effects of interfacial thermal resistance are not thought to be significant, although they are harder to estimate, because their measured values tend to be nearly proportional to T at moderately low temperature ($10 \text{ K} < T < 100 \text{ K}$) and almost temperature invariant above 150 K,²² whereas phonon-scattering theory predicts a $1/T^3$ behavior,²³ and the fitted parameters of the device tested are nearly temperature independent below 77 K. Solid–solid interfaces have much lower interfacial resistances than liquid–solid interfaces, because the acoustic impedance mismatches are much smaller and phonon scatter much less diffuse. For the Ni–Si interface the calculated interfacial thermal resistance is²³ $\approx 10/T^3$ cm² K/W or $\approx 10^9/T^3$ K/W for a 1 μm^2 junction. This is insignificant above ~ 30 K for the present devices.

Electromigration is a somewhat more serious problem, potentially leading to the formation of voids and shorts. Since the junctions used here are planar overlaps, the current density is highest in the region outside the junction area. Thus the expected major effect of electromigration would be a systematic change in R_s as the run proceeds. This is not believed to be a significant short-term effect for junction biases below 120 mV, because the I – V curves from several runs at a given temperature are identical within the measurement noise. Many of these junctions undergo irreversible breakdown when the bias voltage is greater than 0.2 V, and their characteristics change slowly with time at lower voltages. Each of the experimental data sets consists of two I – V runs at each temperature, and the room-temperature I – V curves are closely similar before and after the low-temperature runs.

Finally the current density at 1 mA in 0.3 μm^2 is 3×10^5 A/cm², and much higher with smaller junctions, so current crowding effects and associated non-linear resistances cannot be ruled out.

5. Conclusion: Implications for Optical Detection

Having a repeatable process for making small Ni–NiO–Ni junctions is a prerequisite for making these devices into technologically useful optical detectors. In this paper we have demonstrated such a process and have shown that the tunneling behavior of the junctions is very well modeled by the (complete) Simmons equation, which gives a theoretical and practical underpinning to future device designs. Work is under way to make efficient antenna-coupled tunnel junction detectors in the 1.55 μm optical communication bands. The major potential advantages of ACTJ detectors at those wavelengths are their very high speed and that they replace III–V devices, requiring only metal, oxide, and dielectric.

We have found that the low-temperature responsivity of these junctions is much better than that at room temperature because of extrinsic series resistance and the decline in the effective barrier heights from ~ 0.22 eV at 4 K to ~ 0.19 eV at 295 K. It may be possible to tune the composition of the metal and oxide so as to raise the barrier height slightly, and it is certainly possible to reduce the metal resistance by using thicker films, higher-conductivity metals, and mask layouts that reduce the lead length. With these improvements, responsivity values of >1 A/W may well be attainable at room temperature. At 1.55 μm , unit quantum efficiency corresponds to $\mathfrak{R} = 1.25$ A/W, so it may be possible to make efficient near-IR detectors provided the poorly understood details of the rectification process do not preclude it. To realize this goal, however, the junction capacitance must be compensated for, either by a traveling-wave design or a reactive matching approach. Further theoretical investigation is needed to elucidate the physics of terahertz and IR rectification in MIM junctions.

Appendix A: Simmons Equations for the Very-Low-Barrier Case

The WKB expression for the tunneling probability of an electron of energy E through a 1-D barrier is²⁴

$$D(E_x) = \exp \left\{ -\frac{4\pi}{h} (2m)^{1/2} \int_{s_1}^{s_2} [\eta + \phi(x) - E_x]^{1/2} dx \right\}, \quad (A1)$$

where η is the Fermi energy. The square root in the integral is inconvenient to handle. A useful approximation is

$$\int_{s_1}^{s_2} f^{1/2}(x) dx \approx \sqrt{\bar{f}} \Delta s, \quad (A2)$$

which leads to an alternative expression for the tun-

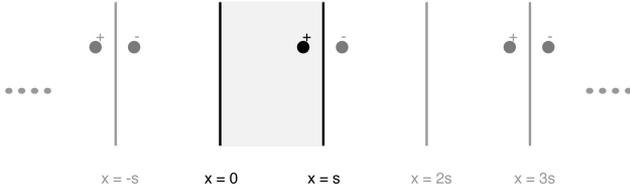


Fig. 10. One-dimensional image potential geometry.

neling probability:

$$D(E_x) = \exp\left\{-\frac{4\pi\beta\Delta s}{h}[2m(\eta + \bar{\phi} - E_x)]^{1/2}\right\}, \quad (\text{A3})$$

where $\bar{\phi}$ is the mean value of the total potential over the interval Δs separating the two classical turning points, s_1 and s_2 , at which the potential function crosses the Fermi levels of the electrodes. The adjustment factor β is given by

$$\beta = 1 - \frac{1}{8\bar{f}^2\Delta s} \int_{s_1}^{s_2} [f(x) - \bar{f}]^2 dx, \quad (\text{A4})$$

where $f(x)$ is the expression inside the square root in the integrand of Eq. (A1). It is nearly constant and within a few percent of unity. If β is computed exactly, Eq. (A3) is closely equivalent to Eq. (A1), but β can also be set constant for more approximate work.

These expressions apply at $T = 0$; this is a good approximation for high barriers but a much poorer one where $\phi \approx 8$ kT. Significant improvements in accuracy at very low RA products require computing the tunnel current at a finite temperature.

Appendix B: Indefinite Integral of the Image Potential

In the 1-D case the total potential is the sum of the trapezoidal potential because of the applied bias voltage and the tunneling barriers, plus the image potential, which is the attraction felt by a charge carrier for a conductive surface. Because of the hall-of-mirrors geometry (see Fig. 10), there is an infinite sum over the images of images. In the perfect-conductor limit the potential V_{img} is given by [Eq. (32) of Ref. 8]

$$V_{\text{img}}(x) = \frac{-e^2}{4\pi\epsilon_0 K} \left[\frac{1}{2x} + \sum_{n=1}^{\infty} \frac{ns}{(ns)^2 - x^2} - \frac{1}{ns} \right], \quad (\text{B1})$$

where the metal surfaces are at $x = 0$ and $x = s$ and K is the relative dielectric constant. [Eqs. (21) and (22) of Ref. 9 are erroneous.] This sum has a rather distressingly slow convergence: To achieve double-precision accuracy (a 48 bit significand), almost 60,000 terms are necessary. In the expression for the tunnel current we require the integral of the total potential between the classical turning points of the motion (where the potential crosses the Fermi level).

Since these points move as the junction bias is changed, we require the indefinite integral of the image potential:

$$F_{\text{img}}(x) = \int^x V_{\text{img}}(x') dx'. \quad (\text{B2})$$

The integrand is symmetric with respect to $s/2$, so

$$F_{\text{img}}(x) = 2F_{\text{img}}\left(\frac{s}{2}\right) - F_{\text{img}}(s-x), \quad \left\{x > \frac{s}{2}\right\}. \quad (\text{B3})$$

For $x \in [0, s/2]$ the sum is uniformly convergent and its terms are regular, so we can decompose it in partial fractions,

$$F_{\text{img}}(x) = -\frac{a}{2} \ln x - \frac{\alpha}{2} \left[r\left(\frac{x}{s}\right) - r\left(-\frac{x}{s}\right) \right], \quad (\text{B4})$$

where $\alpha = e^2/(4\pi\epsilon_0 K)$ and r is given by

$$r(y) = \frac{1}{2} \int^y dy' \sum_{n=1}^{\infty} \left(\frac{1}{n-y'} - \frac{1}{n} \right). \quad (\text{B5})$$

Taking the binomial expansion of the first term and collecting terms, Eq. (B5) becomes

$$r(y) = \frac{1}{2} \int^y dy' \sum_{j=1}^{\infty} (-y')^j \sum_{n=1}^{\infty} \binom{j}{n} \frac{1}{n^{j+1}}. \quad (\text{B6})$$

Expressing the last sum in terms of the Riemann ζ function, we can rewrite the image potential integral as

$$F_{\text{img}}(x) = \left(-\frac{e^2}{4\pi\epsilon_0 K} \right) \left\{ \frac{1}{2} \ln \left[\frac{x(1+x/s)}{1-x/s} \right] - \frac{x}{s} + \sum_{m=0}^{\infty} \left(\frac{x}{s} \right)^{2m+1} \frac{[\zeta(2m+1) - 1]}{2m+1} \right\}, \quad \left\{x < \frac{s}{2}\right\}. \quad (\text{B7})$$

The advantage of this more complicated formula over Eq. (B2) is the very rapid convergence of the infinite sum. Because $\zeta(m) - 1 \sim 2^{-m}$ for large m , the m th term of the sum makes a maximum relative contribution of approximately $0.5^{4m+1}/m$, so that double-precision accuracy needs at most 11 terms.

The authors thank George A. Keefe, John R. Kirtley, and Roger H. Koch for samples and helpful discussions and Niranjana Ruiz, S. J. Chey, and Dennis Martinez for device fabrication.

References

1. A. Sanchez, C. F. Davis, Jr., K. C. Liu, and A. Javan, "The MOM tunneling diode: theoretical estimate of its performance

- at microwave and infrared frequencies," *J. Appl. Phys.* **49**, 5270–5277 (1978).
2. B. Michael Kale, "Electron tunneling devices in optics," *Opt. Eng.* **24**, 267–274 (1985).
 3. I. Wilke, Y. Oppliger, W. Herrmann, and F. K. Kneubuehl, "Nanometer thin-film Ni–NiO–Ni diodes for 30 THz radiation," *Appl. Phys. A* **58**, 329–341 (1994).
 4. C. Fumeaux, W. Herrmann, H. Rothuizen, P. De Natale, and F. K. Kneubuehl, "Mixing of 30 THz laser radiation with nanometer thin-film Ni–NiO–Ni diodes and integrated bow-tie antennas," *Appl. Phys. B* **63**, 135–140 (1996).
 5. C. Fumeaux, M. Gritz, I. Codreanu, W. Schaich, F. Gonzalez, and G. Boreman, "Measurement of the resonant lengths of infrared dipole antennas," *Infrared Phys. Technol.* **41**, 271–281 (2000).
 6. C. Fumeaux, J. Alda, and G. Boreman, "Lithographic antennas at visible frequencies," *Opt. Lett.* **24**, 1629–1631 (1999).
 7. L. S. Dorneles, D. M. Schaefer, M. Carara, and L. F. Schelp, "The use of Simmons' equation to quantify the insulating barrier parameters in Al/AlO_x/Al tunnel junctions," *Appl. Phys. Lett.* **82**, 2832–2834 (2003).
 8. J. G. Simmons, "Generalized formula for the Electric Tunnel Effect between similar electrodes separated by a thin insulating film," *J. Appl. Phys.* **34**, 1793–1803 (1963).
 9. J. G. Simmons, "Electric tunnel effect between dissimilar electrodes separated by a thin insulating film," *J. Appl. Phys.* **34**, 2581–2590 (1963).
 10. J. G. Small, G. M. Elchinger, A. Javan, A. Sanchez, F. J. Bachner, and D. L. Smythe, "ac electron tunneling at infrared frequencies: thin-film MOM diode structure with broadband characteristics," *Appl. Phys. Lett.* **24**, 275–279 (1974).
 11. M. Heiblum, S. Wang, J. R. Whinnery, and T. K. Gustafson, "Characteristics of integrated MOM junctions at dc and at optical frequencies," *IEEE J. Quantum Electron.* **QE-14**, 159–169 (1978).
 12. S. K. Masalmeh, H. K. E. Stadermann, and J. Korving, "Mixing and rectification properties of MIM diodes," *Physica B* **218**, 56–59 (1996).
 13. J. A. Nelder and R. Mead, "A simple method for function minimization," *Comput. J.* **7**, 308–313 (1965).
 14. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University, 1988), Section 10.4.
 15. G. A. Keefe, IBM T. J. Watson Research Center, Yorktown Heights, N.Y. 10598 (personal communication).
 16. L. D. Jackel, R. E. Howard, E. L. Hu, D. M. Tennant, and P. Grabbe, "50-nm silicon structures fabricated with trilevel electron beam resist and reactive-ion etching," *Appl. Phys. Lett.* **39**, 268–270 (1981).
 17. J. D. R. Buchanan, T. P. A. Hase, B. K. Tanner, N. D. Hughes, and R. J. Hicken, "Determination of the thickness of Al₂O₃ barriers in magnetic tunnel junctions," *Appl. Phys. Lett.* **81**, 751 (2002).
 18. D. Lide, ed., *CRC Handbook of Chemistry and Physics*, 81st ed. (CRC Press, 2000), pp. 12–46.
 19. E. M. L. Chung, D. M. Paul, G. Balakrishnan, M. R. Lees, A. Ivanov, and M. Yethiraj, "Role of electronic correlations on the phonon modes of MnO and NiO," *Phys. Rev. B* **68**, 140–146 (2003).
 20. W. Olejniczak and M. Bieniecki, "Fine structure in differential conductance of oxidized nickel observed in a room temperature stm experiment," *Solid State Commun.* **101**, 877–882 (1997).
 21. D. Lide, ed., *CRC Handbook of Chemistry and Physics*, 81st ed. (CRC Press, 2000), pp. 12–200.
 22. E. Gmelin, M. Asen-Palmer, M. Reuther, and R. Villar, "Thermal boundary resistance of mechanical contacts between solids at subambient temperatures," *J. Phys. D* **32**, R19–R43 (1999).
 23. E. T. Swartz and R. O. Pohl, "Thermal boundary resistance," *Rev. Mod. Phys.* **61**, 605–667 (1989).
 24. Ref. 8, Eq. (1).